

Analysis 3

Notes taken by Yeheli Fomberg <yp@campus.technion.ac.il>

Based on lectures by Roy Meshulam and notes by Ran Kiri and Or Shalit
and notes by Michael Polyak and Multivariable Mathematics by Theodore Shifrin

Winter 2025

These notes are not endorsed by the lecturers. I have revised them outside lectures to incorporate supplementary explanations, clarifications, and material for fun. While I have strived for accuracy, any errors or misinterpretations are most likely mine.

Contents

1	Introduction to Topology in Euclidean Spaces	5
1.1	Norms and metrics	5
1.2	Sequences	6
1.3	Definitions	6
1.4	Continuous functions	7
1.5	Convex sets	8
2	Differentiability	9
2.1	Definitions	9
2.2	Properties of the operator norm	13
2.3	Continuous and differentiable partial derivatives	14
3	Least Squares and Gradient Descent	17
3.1	The standard method	17
3.2	Least squares	17
3.3	Gradient descent	18
4	Taylor's Theorem	19
4.1	Higher order partial derivatives	19
4.2	Multi-index notation	19
4.3	Taylor's theorem	20
5	Classifying Critical Points	21
5.1	Preliminary defintions	21
5.2	Classification of critical points	22
6	The Inverse Function Theorem	24
6.1	Preliminaries	24
6.2	The inverse function theorem	24
7	Newton–Raphson Method	26
7.1	The Newton–Raphson Method	26
8	The Open Mapping Theorem	28
8.1	The open mapping theorem	28
8.2	Constrained optimization	29
9	The Implicit Function Theorem	34
9.1	The implicit function theorem	34
9.2	Differentiation of implicit functions	35
10	Finding Roots of Polynomials	37
10.1	Finding roots of polynomials	37
11	Manifolds	38
11.1	C^1 -manifolds	38
11.2	The torus	41
12	The Tangent Space	42
12.1	The tangent space	42

13 Manifolds With Boundary	47
13.1 Manifolds with boundary	47
13.2 The tangent space at the boundary	48
14 Length of a Path	49
14.1 Length of a path	49
14.2 Arc length parameterization	50
15 Line Integrals	52
15.1 Type one line integrals	52
15.2 Type two line integrals	54
16 Integration	56
16.1 Integration over a box	56
16.2 Fubini's theorem	57
17 Zero Volume	61
17.1 Zero volume	61
17.2 Measure zero and integration	62
17.3 Integration over Jordan measurable sets	63
18 Improper integrals	64
18.1 Improper integrals	64
19 Change of variables	66
19.1 The change of variables theorem	66
19.2 Example of polar coordiantes	68
19.3 Example of cylindrical coordiantes	68
19.4 Example of spherical coordiantes	69
19.5 Practice	71
20 k-dimensional volume	73
20.1 Intuition	73
20.2 k -dimensional volume	74
21 Surface integrals	79
21.1 Surface area	79
21.2 Surface integral	79
22 Differential forms	82
22.1 Type two line integrals	82
22.2 1-differential forms	83
22.3 k -forms and the wedge product	85
22.4 Pullback of forms	88
22.5 Motivation for surface integrals and flux	90
22.6 Surface interals of m -forms	92
23 The exterior derivative and orientation	96
23.1 The exterior derivative	96
23.2 Green's theorem	98
23.3 The winding number	101
23.4 Orientation	102
23.5 The interior product and induced orientation	104

23.6 Surface area element	107
24 Generalized Stokes theorem	109
24.1 Partition of unity	109
24.2 Stokes' theorem	110
25 Applications to physics	112
25.1 Classical Stokes' theorem	112
25.2 Geometric intuition for curl	113
25.3 Classical divergence theorem	114
25.4 Geometric intuition for divergence	115
25.5 Conservative fields	115
26 The Hodge star operator	118
26.1 A different view on differential operators	118
26.2 The general Hodge star operator	119

1 Introduction to Topology in Euclidean Spaces

1.1 Norms and metrics

Definition 1.1 (The n -dimensional Euclidean space). The space we will work with in this course, the n -dimensional Euclidean space is defined as follows:

$$\mathbb{R}^n := \left\{ (x_1, x_2, \dots, x_d) \mid \begin{matrix} 1 \leq i \leq d \\ x_i \in \mathbb{R} \end{matrix} \right\}.$$

Definition 1.2 (Norm). A Norm on $\mathbb{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \rightarrow [0, \infty]$ such that for all $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$,

- (1) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality);
- (2) $\|rx\| = r\|x\|$ (absolute homogeneity);
- (3) if $\|x\| = 0$ then $x = 0$ (positive definiteness).

Definition 1.3 (Euclidean norm). The Euclidean norm is defined as:

$$\|x\| = \|x\|_2 := \sqrt{\sum_{i=1}^d x_i^2}.$$

This norm is also called the *standard norm* on $\mathbb{R}^n$, and it is the norm we will use throughout this course.

Definition 1.4 (L^p norm). Another example of a norm, which is slightly less important for the purposes of this course, but is still very relevant in functional analysis is the L^p norm, given by

$$\|x\| = \|x\|_p := \sqrt[p]{\sum_{i=1}^d x_i^p}.$$

Definition 1.5 (Equivalent norms). Suppose that p and q are two norms on $\mathbb{R}^n$. Then p and q are called equivalent if there exist positive real constants c and C such that for every vector $x \in \mathbb{R}^n$ we have

$$cq(x) \leq p(x) \leq Cq(x).$$

Remark 1.1. Equivalence of norms is an equivalence relation. In $\mathbb{R}^n$ all norms are equivalent.

Definition 1.6 (Metric). A metric on $\mathbb{R}^n$ is a function $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty]$ such that for all $x, y, z \in \mathbb{R}^n$,

- (1) $d(x, y) = 0$ if and only if $x = y$;
- (2) $d(x, y) = d(y, x)$ (symmetry);
- (3) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

Definition 1.7 (Euclidean metric). The Euclidean metric is defined as:

$$d_2(P_1, P_2) := \|P_1 - P_2\|_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Remark 1.2. The Euclidean metric is induced by the Euclidean norm, which means that it can be expressed using the Euclidean norm. Indeed $d_2(P_1, P_2) = \|P_1 - P_2\|_2$. We can similarly induce the L^p metric using L^p norms.

Remark 1.3. We can go one step further, and notice that the Euclidean norm is induced by the standard inner product defined on $\mathbb{R}^n$ given by

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i, \quad \forall x, y \in \mathbb{R}^n.$$

On the other hand, not all L^p norms are induced by an inner product. The only L^p norm which is induced by an inner product is the L^2 norm.

Remark 1.4. Similar to Definition 1.5, there exist multiple notions of equivalence of metrics. However, unlike norms, not all metrics on $\mathbb{R}^n$ are equivalent, but all metrics induced by norms, are equivalent.

Definition 1.8 (Angle between vectors). From the Cauchy–Schwartz inequality we have that for all $x, y \in \mathbb{R}^n$

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} \leq 1$$

which allows us to define

$$\angle(x, y) := \arccos \left(\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} \right).$$

Remark 1.5. Recall that if $x, y \in \mathbb{R}^n$ and $\theta = \angle(x, y)$ then as this may be useful sometimes.

$$\begin{aligned} \|x + y\| &= \sqrt{\|x\|^2 + \|y\|^2 + 2\|x\|\|y\|\cos(\theta)} \\ \|x - y\| &= \sqrt{\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos(\theta)} \end{aligned}$$

as this may be useful sometimes.

1.2 Sequences

Definition 1.9 (Convergence). Up until now we didn't have many problems using subscript for indexes of sequences, but now since we have coordinates we must denote a sequence in another similar way called ^{superscript} as such $(x^n)_{n=1}^\infty$. We define convergence of such sequences in the following way:

$$\lim_{n \rightarrow \infty} x^n = x \quad \Longleftrightarrow \quad \forall i \left(\lim_{n \rightarrow \infty} x_i^n = x_i \right).$$

Definition 1.10 (Cauchy sequence). A sequence $(x^n)_{n=1}^\infty$ is called a Cauchy sequence if and only if

$$\lim_{n, m \rightarrow \infty} \|x^n - x^m\| = 0.$$

1.3 Definitions

Definition 1.11 (Complete metric space). A complete metric space is a metric space M such that every Cauchy sequence in M converges to some limit in M .

Definition 1.12 (Open set). An open set in a Euclidean space is a subset U such that for any $x \in U$ exists $\varepsilon > 0$ such that any y such that any $y \in B_\varepsilon(x)$ satisfies $y \in U$.

Definition 1.13 (Neighbourhood). In a topological space X a space a neighbourhood of a point $x \in X$ is a subset such that there exists an open set U such that $p \in U \subset V$.

Definition 1.14 (Closed set). A closed set E in a Euclidean space X is a subset of X such that:

$$(x^n)_{n=1}^\infty \subseteq E \quad x^n \xrightarrow{n \rightarrow \infty} x \implies x \in E.$$

Definition 1.15 (Compact set). A set $X \subset \mathbb{R}^n$ is called compact if every open cover of X has a finite subcover.

Definition 1.16 (Sequentially compact set). A set $X \subset \mathbb{R}^n$ is called sequentially compact if every sequence $(x^n)_{n=1}^\infty \subset X$ has a subsequence $(x^{n_k})_{k=1}^\infty$ that converges to a point x in X .

Remark 1.6. In Euclidean spaces, being sequentially compact is equivalent to being compact.

In the following definitions, let E be a subset of a Euclidean space.

Definition 1.17 (Closure). The closure of a set $E \subseteq \mathbb{R}^n$ is can be

$$\text{Cl}(E) = \left\{ x \mid \exists (x^n) \text{ s.t. } x^n \xrightarrow{n \rightarrow \infty} x \right\}$$

Proposition 1.1. The closure of $E \subseteq \mathbb{R}^n$ is the smallest closed subset of $\mathbb{R}^n$ containing E .

Definition 1.18 (Interior). The interior of E is defined as:

$$\text{Int}(E) = \left\{ x \mid \exists r > 0 \text{ s.t. } B_r(x) \subseteq E \right\}$$

Definition 1.19 (Boundary). The boundary of E is defined as:

$$\partial E = \left\{ x \in \mathbb{R}^n \mid \forall r > 0 \exists x \in E \wedge \exists y \in E^c \text{ s.t. } y, z \in B_r(x) \right\}$$

1.4 Continuous functions

Definition 1.20 (Continuous function). A function $f: \underset{\subseteq \mathbb{R}^n}{A} \rightarrow \underset{\subseteq \mathbb{R}^m}{A}$ is said to be continuous at $x \in A$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $\|x - y\| < \delta$

$$\|f(y) - f(x)\| < \varepsilon$$

and we say that a function is continuous on X if it is continuous for every $x \in X$.

Remark 1.7. An important, equivalent, more general definition for continuity is that if f is a function from A to B then if U is an open set in B implies $f^{-1}(U)$ is an open set we say that f is continuous on A .

Definition 1.21 (Uniformly continuous function). A function $f: \underset{\subseteq \mathbb{R}^n}{A} \rightarrow \underset{\subseteq \mathbb{R}^m}{A}$ is said uniformly continuous on $X \subseteq A$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $\|x - y\| < \delta$

$$\|f(x) - f(y)\| < \varepsilon.$$

Definition 1.22 (Connectedness). A set $A \subset \mathbb{R}^n$ is said to be connected if it cannot be written as a union of two or more disjoint nonempty open subsets. A set that is not connected is said to be disconnected.

Example 1.1. Any interval $I \subset \mathbb{R}$ is connected.

Example 1.2. The set $A = (0, 1) \cup (2, 3) \subset \mathbb{R}$ is disconnected because it can be written as the union of $(0, 1)$ and $(2, 3)$.

Definition 1.23 (Path). A path from $x \in X$ to $y \in X$ is a continuous function $f: [0, 1] \rightarrow X$ such that $f(0) = x$ and $f(1) = y$.

Definition 1.24 (Path connectedness). A set $A \subset \mathbb{R}^n$ is said to be path connected if there exists a path between any two points $x, y \in A$.

1.5 Convex sets

Definition 1.25 (Convex set). A set $C \subset \mathbb{R}^n$ is called convex if and only if

$$\forall x, y \in C \quad \forall t \in [0, 1] \quad tx + (1 - t)y = y + t(x - y) \in C$$

Definition 1.26 (Convex combination). Given a finite set of points $\{x_i\}_{i=1}^n \subset \mathbb{R}^n$, a convex combination of these points is a point of the form

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

where $\{\alpha_i\}_{i=1}^n$ satisfy $\alpha_i \geq 0$ for all $1 \leq i \leq n$ and also

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1.$$

Remark 1.8. Every convex combination of two point lies on the line segment that connects them.

Definition 1.27 (Convex hull). Let $A \subset \mathbb{R}^n$ be a set. The convex hull of A is defined as

$$\text{Conv}(A) = \bigcap_{\substack{A \subset S \\ S \text{ convex}}} S.$$

Remark 1.9. This is the smallest convex set containing A .

Proposition 1.2. Let $A \subset \mathbb{R}^n$ be convex. Then every convex combination of elements of A is contained in A .

Proof. to be added. □

Proposition 1.3. Let $A \subset \mathbb{R}^n$. Then the set of all convex combinations of points in A is equal to the set $\text{Conv}(A)$.

Proof. to be added. □

Theorem 1.4 (Carathéodory's theorem). Let $A \subset \mathbb{R}^n$, and let $x \in \text{Conv}(A)$. Then x is a linear combination of at most $n + 1$ points in A .

Proof. to be added. □

Definition 1.28 (Convex function). Let $A \subset \mathbb{R}^n$ be convex. Then the function $f: A \rightarrow \mathbb{R}$ is said to be convex if for any $x, y \in A$ and $0 \leq \lambda \leq 1$ it satisfies:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

That is, for any two points in A , for any point z on the line segment connecting them, we have $f(z) \leq z$.

Remark 1.10. If the inequality above is also a strong inequality, the function is said to be strictly convex.

Theorem 1.5 (Jensen's inequality). Let f be a convex function, and let $\{\alpha_i\}_{i=1}^n$ be coefficients of some convex combination. Then Jensen's inequality is satisfied:

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i).$$

2 Differentiability

2.1 Definitions

Definition 2.1 (Linear map). Let $A \in \mathbb{R}^{m \times n}$. We define the linear map $T_A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ as such:

$$T(x) := Ax.$$

Definition 2.2 (Hilbert–Schmidt norm). Let $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{m \times n}$ denote the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. We can make this space into a normed space by defining the Hilbert–Schmidt norm as follows

$$\|T\|_{\text{HS}} := \sqrt{\text{tr}(A^t A)} = \sqrt{\sum_{i,j} (a_{ij})^2}.$$

Remark 2.1. The Hilbert–Schmidt norm is induced by the inner product $\langle A, B \rangle = \text{tr}(B^T A)$ for real matrices. Generally, the inner product is given by $\langle A, B \rangle = \text{tr}(B^*, A)$ for bounded operators on Hilbert spaces.

Definition 2.3 (Operator norm). Let $T: V \rightarrow W$ be a linear transformation between inner product spaces, we define the operator norm to be:

$$\|T\|_{\text{op}} = \|T\| = \sup_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V}$$

and T is said to be bounded if $\|T\| < \infty$.

Proposition 2.1. Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear map. Then

$$\|T\|_{\text{op}} \leq \|T\|_{\text{HS}} = \sqrt{\sum_{i,j} a_{ij}^2} < \infty.$$

Proof. By definition since the i th entry of the vector Tx is $\sum_{j=1}^n a_{ij}x_j$ we get

$$\|Tx\|^2 = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}x_j \right)^2.$$

Then from the Cauchy–Schwartz inequality we get

$$\|Tx\|^2 = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}x_j \right)^2 \leq \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}^2 \right) \left(\sum_{j=1}^n x_j^2 \right) \leq \left(\sum_{i,j} a_{ij}^2 \right) \cdot \|x\|^2.$$

Then we can take the root on both sides and get

$$\|Tx\| \leq \sqrt{\left(\sum_{i,j} a_{ij}^2 \right)} \cdot \|x\| = \|T\|_{\text{HS}} \cdot \|x\| < \infty$$

and since this holds for all x we get $\|T\|_{\text{op}} \leq \|T\|_{\text{HS}} \cdot \|x\| < \infty$ which completes the proof. $\square$

Remark 2.2. In general, the Hilbert–Schmidt norm is defined only for bounded operators, but in the case of linear maps $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ we get that T is automaticall bounded, since it attains a maximum on S^{n-1} as T is continuous and S^{n-1} is compact.

Corollary 2.2. *We now obtain that*

$$\|T(x) - T(y)\| \leq \|T\|_{\text{op}} \|x - y\|.$$

This means that T is continuous and even Lipschitz continuous on its domain.

Remark 2.3. As norms on $\mathbb{R}^n \times \mathbb{R}^n \cong \mathbb{R}^{n \times n}$, the Hilbert–Schmidt norm and the operator norms are equivalent.

Definition 2.4 (Affine function). An affine function is a function of the form:

$$T_{A,b}(x) = Ax + b$$

such that $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Definition 2.5 (Differentiable function). Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f: U \rightarrow \mathbb{R}^m$. Given $a \in U$ we say that f is differentiable at a if there exists a linear transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

$$\lim_{h \rightarrow 0} \frac{f(a+h) - (f(a) + Th)}{\|h\|} = 0.$$

We say that f is differentiable if f is differentiable at every $a \in U$.

Remark 2.4. Note that since U is open there exists a δ small enough such that $B(a, \delta) \subseteq U$.

Remark 2.5. We can also say that f is differentiable at a if there exists an “error” function $\epsilon(h)$ defined around 0 such that $f(a+h) = f(a) + Th + \epsilon(h)$ and

$$\lim_{h \rightarrow 0} \frac{\epsilon(h)}{\|h\|} = 0.$$

If $\epsilon(h)$ satisfies the above function we say that $\epsilon(h) = o(\|h\|)$.

Remark 2.6. We denote T from Definition 2.5 by $df(a)$. It is also known as the differential of f at a .

Remark 2.7. The operator df is a linear map $df: \mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. This means that $df(a)$ is a linear map $df(a): \mathbb{R}^n \rightarrow \mathbb{R}^m$ (as we know), and therefore $df(a)(h) \in \mathbb{R}^m$.

Remark 2.8. Since $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, we can consider the Euclidean spaces with their standard inner product, and we denote the matrix representation of the differential (this depends on the basis we choose for the Euclidean spaces, usually we choose the standard basis), also known as the total derivative of f at a , by $Df(a)$.

Proposition 2.3 (The chain rule). *Let f be differentiable at a , and suppose that g is differentiable at $b := f(a)$. Then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a))f'(a) \text{ or } d(g \circ f)(a) = dg(b) \circ df(a).$$

Remark 2.9. The chain rule may not seem very intuitive when stated with formulas. Notice that all we say, is that if there exists two functions, f differentiable at a , g differentiable at $f(a)$, then the differential of $g \circ f$ is given by applying the differential of f at a , and then applying on that the differential of g at $f(a)$. That should be intuitive.

Example 2.1. We want to compute the differential of the function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $x \mapsto \|x\|$. First, we notice that it is the composition of the functions

$$h(x) = \langle x, x \rangle = x^T I x$$

$$g(t) = \sqrt{t}$$

so it may help to compute the differential of quadratic forms $q(x) = x^T Q x$ for $Q \in \mathbb{R}^{n \times n}$. Recall that any quadratic form corresponds uniquely to a symmetric matrix so that $q(x) = x^T A x$ where $A \in \mathbb{R}^n$ is symmetric. Then since A is symmetric we have

$$q(x+h) - q(x) = x^T A h + h^T A x + h^T A h = 2x^T A h + h^T A h$$

and clearly $h^T A h \in o(\|h\|)$ and $2x^T A h$ is linear in h so $dq(x)(h) = 2x^T A h$ and in particular for $h(x) = x^T I x$ we get $dh(x) = 2x^T$. We know that $dg(x) = 1/2\sqrt{x}$ and so from the chain rule

$$d(g \circ h)(x) = dg(h(x)) \circ dh(x) = \frac{1}{2\sqrt{x^T x}} \circ 2x^T = \frac{x^T}{\|x\|}$$

so $df(x) = x^T / \|x\|$.

Definition 2.6 (Directional derivative). Let $f: U \rightarrow \mathbb{R}$ with $U \subset \mathbb{R}^n$, let $a \in U$ and let $0 \neq v \in \mathbb{R}^n$. Then, if the following limit exists:

$$D_v f(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}$$

we say that f is differentiable in the direction of v , and $D_v f(a)$ is called the directional derivative of f at a .

Definition 2.7 (Differentiation along a curve). Let $\gamma: (a, b) \rightarrow \mathbb{R}^n$ so that $\gamma(t_0) = x_0 \in U$ for $t_0 \in (a, b)$ and $U \subset \mathbb{R}^n$. If the derivative

$$\frac{d(f \circ \gamma)}{dt}(t_0)$$

exists, we say that it is the derivative of f along the curve γ . This is a generalization of directional derivatives.

Remark 2.10. Note that if $\mathcal{E} = \{e_1, \dots, e_n\}$ is the standard basis for $\mathbb{R}^n$ then we may denote the directional derivative of f at a in the direction of e_i in one of the following ways:

$$D_{e_i} f(a) = D_i f(a) = \frac{\partial f}{\partial e_i}(a) = f_{x_i}(a).$$

These are also called the partial derivatives of f .

Proposition 2.4. If $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ for is differentiable at a then:

$$D_v f(a) = Df(a)v$$

for every $v \in \mathbb{R}^n$.

Remark 2.11. The proof is using the fact that:

$$f(a+h) = f(a) + Df(a)h + o(h).$$

An important corollary of this proposition is

$$D_v f(a) = Df(a)v = \sum_i^n \frac{\partial f}{\partial x_i}(a) v_i.$$

Definition 2.8 (Gradient). For a linear function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, the differential at a is a map $df(a): \mathbb{R}^n \rightarrow \mathbb{R}$. Therefore from the Riesz representation theorem there exists a vector $v_a \in \mathbb{R}^n$ so that $df(a)(h) = \langle v_a, h \rangle$ for all $h \in \mathbb{R}^n$. We define $\nabla f(a) = v_a$. This is called to gradient of f at a .

Remark 2.12. In orthonormal coordinates, the gradient of f at a point is the transpose of the total derivative at that point.

$$\nabla f(a) = (Df(a))^T = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \cdots & \frac{\partial f}{\partial x_n}(a) \end{pmatrix}^T = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$

Remark 2.13. Notice that we get the biggest value for $\|v\| = 1$ when

$$v = \frac{\nabla f(a)}{\|\nabla f(a)\|}$$

which implies that the direction of the gradient is the direction of the steepest ascent. Conversely, the direction orthogonal to it, is the direction of least change. For instance, we saw in Example 2.1 that the total derivative of the norm at a point is in the direction of the unit vector that connects the origin and that point going outwards. Thus, the gradient is also a vector in that direction and this fits with our intuition that this is the direction of maximal change.

Proposition 2.5. Let $f: U \rightarrow \mathbb{R}$ for $U \subset \mathbb{R}^n$ be differentiable in U . Then if f has a local minimum or maximum at $a \in U$ then $\nabla f(a) = 0$.

Remark 2.14. This is simply a reiteration of Fermat's theorem for the n -dimensional case.

Proposition 2.6. A linear function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $a \in U$ if and only if $f_i = f \circ \pi_i$ is differentiable for every $1 \leq i \leq n$ where π_i is the projection function on the i th coordinate.

Proof. Suppose that f_i is differentiable for every $1 \leq i \leq n$ then we can define a new function:

$$Tv := \begin{pmatrix} Df_1(a)v \\ Df_2(a)v \\ \vdots \\ Df_m(a)v \end{pmatrix}$$

And see that indeed:

$$\|f(a+h) - f(a) - Th\| \leq \sum_{i=1}^m |f(a+h) - f(a) - Df_i(a)h| = \sum_{i=1}^m |\epsilon_i(h)| = o(h)$$

As wanted. On the other hand, if f is differentiable we can use the chain rule to show that each f_i is differentiable with the desired derivative. $\square$

Definition 2.9 (Jacobian). Let $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that each of its first-order partial derivatives exists. We define the Jacobian of f as such:

$$J_f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^T f_1 \\ \vdots \\ \nabla^T f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

2.2 Properties of the operator norm

Remark 2.15. For convenience we identify any matrix $A \in \mathbb{R}^{n \times n}$ as the operator it defines $T_A: \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Remark 2.16 (Basic properties). Recall that we defined the operator norm as

$$\|A\| = \|A\|_{\text{op}} = \sup \{\|Ax\| : \|x\| = 1\} = \sup \left\{ \frac{\|Ax\|}{\|x\|} : x \neq 0 \right\}.$$

We can prove that $A \mapsto \|A\|$ is a norm, and that $\|A\|$ is the smallest constant that satisfies the inequality $\|Ax\| \leq \|A\|\|x\|$ for all $x \in \mathbb{R}^n$. The operator norm also satisfies

$$\|AB\| \leq \|A\|\|B\|$$

because we have

$$\|(AB)x\| \leq \|A\|Bx\| \leq \|A\|\|B\|\|x\|.$$

Remark 2.17. The following proposition may seem like an insignificant lemma, but it is actually useful and its geometric meaning should also be memorable. It characterizes the invertible matrices as matrices that can't downscale vectors "too much".

Proposition 2.7. *A matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if there exists a constant $c > 0$ such that $\|Ax\| \geq c\|x\|$ for all $x \in \mathbb{R}^n$. Moreover, if A is invertible the constant $c = \|A^{-1}\|^{-1}$ is the biggest constant that satisfies that inequality.*

Proof. Suppose A is invertible, then we have

$$\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\| \cdot \|Ax\|,$$

and because $\|A^{-1}\| \neq 0$ we get that

$$c\|x\| \leq \|Ax\|$$

for $c = \|A^{-1}\|^{-1}$ for all $x \in \mathbb{R}^n$. It is the biggest constant that satisfies this inequality because

$$r\|x\| \leq \|Ax\| \iff \|A^{-1}Ax\| \leq r^{-1} \cdot \|Ax\| \iff r \leq \|A^{-1}\|^{-1} = c.$$

Next suppose there exists $c > 0$ such that $\|Ax\| \geq c\|x\|$ for all $x \in \mathbb{R}^n$. From this we get that

$$Ax = 0 \iff c\|x\| \leq \|Ax\| = 0 \implies c\|x\| = 0 \implies x = 0,$$

which shows that A is invertible. □

Proposition 2.8. *Let $A, B \in \mathbb{R}^{n \times n}$. If A is invertible and $\|A - B\| < \frac{1}{2}\|A^{-1}\|^{-1}$, then B is also invertible, and $\|B^{-1}\| \leq 2\|A^{-1}\|$.*

Proof. We want to somehow get to $B - A$ so we may notice that for all $x \in \mathbb{R}^n$ we have that

$$Bx = Ax - (A - B)x.$$

Thus,

$$\|Bx\| \geq \|Ax\| - \|(A - B)x\| \geq \|A^{-1}\|^{-1}\|x\| - \frac{1}{2}\|A^{-1}\|^{-1}\|x\| = \frac{1}{2}\|A^{-1}\|^{-1}\|x\|.$$

From the previous proposition, we get that B is invertible and that $\|B^{-1}\| \leq 2\|A^{-1}\|$ which completes the proof. □

Definition 2.10 (General linear group). We denote $\text{GL}(\mathbb{R}^n)$ the collection of invertible matrices in $\mathbb{R}^{n \times n}$.

Remark 2.18. It is clear that the set $\text{GL}(\mathbb{R}^n)$ coupled with matrix multiplication forms a group.

Proposition 2.9. *The group $\text{GL}(\mathbb{R}^n)$ is an open set in $\mathbb{R}^{n \times n}$, and the operations of multiplication and inverse are continuous.*

Proof.

(1) **Open Set:** Let $A \in \text{GL}(\mathbb{R}^n)$. According to Theorem 2.8 we have that for $r = \frac{1}{2}\|A^{-1}\|^{-1}$ that $B(A, r) \subseteq \text{GL}(\mathbb{R}^n)$, so $\text{GL}(\mathbb{R}^n)$ is indeed open in $\mathbb{R}^n$.

(2) **Continuity of Multiplication:** To prove that the operator $m: \text{GL}(\mathbb{R}^n) \times \text{GL}(\mathbb{R}^n) \rightarrow \text{GL}(\mathbb{R}^n)$ given by $m(A, B) = AB$ is continuous, let $A_k \xrightarrow{k \rightarrow \infty} A$ and $B_k \xrightarrow{k \rightarrow \infty} B$. Then

$$\|A_k B_k - AB\| = \|A_k B_k - A_k B + A_k B - AB\| \leq \|A_k\| \|B_k - B\| + \|A_k - A\| \|B\| \xrightarrow{k \rightarrow \infty} 0.$$

Notice that we used the fact that a convergent sequence is bounded.

(3) **Continuity of Inverse:** To prove that the operator $A \mapsto A^{-1}$ is continuous let $A \in \text{GL}(\mathbb{R}^n)$, let $\epsilon > 0$ and choose $\delta = \min \left\{ \frac{1}{2}\|A^{-1}\|^{-1}, \frac{1}{2}\|A^{-2}\|^{-1} \right\} \cdot \epsilon$. Now for any B such that $B \in B(A, \delta)$ we get that B is invertible and from Theorem 2.8 we have $\|B^{-1}\| \leq 2\|A^{-1}\|$ hence

$$\|A^{-1} - B^{-1}\| = \|A^{-1}(B - A)B^{-1}\| \leq 2\|A^{-1}\|^2 \delta < \epsilon$$

which shows that taking the inverse of A is a continuous operation and completes the proof. □

2.3 Continuous and differentiable partial derivatives

Recall that a map $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called differentiable if it has an affine approximation. In the case of a single variable functions

$$f(x) = f(a) + f'(a)(x - a) + o(x - a).$$

In the case of multivariate functions, we have

$$f'(a) = J_f(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

If all the partial derivatives are continuous, then f is differentiable at a .

Proposition 2.10. *Let $f: U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a function such that $f = (f_1, \dots, f_m)$. Suppose that for all i, j the partial derivative $\frac{\partial f_i}{\partial x_j}$ is continuous in a neighbourhood of a . Then f is differentiable at a .*

Proof. We know that it suffices to prove that f_i is differentiable for all i , so suppose $f: U \rightarrow \mathbb{R}$ and let $h \in \mathbb{R}^n$ then

$$h = (h_1, \dots, h_n) = \sum_{j=1}^n h_j e_j$$

where e_j are the unit vectors. We then have that

$$f(a+h) - f(a) = \sum_{j=1}^n \left(f\left(a + \sum_{i=1}^j h_i e_i\right) - f\left(a + \sum_{i=1}^{j-1} h_i e_i\right) \right)$$

because this is a telescopic sum, and since the elements in this sum are the partial derivatives $\frac{\partial f}{\partial x_j}$ at the point $a + \sum_{i=1}^{j-1} h_i e_i$ we get that this is equal to

$$\sum_{j=1}^n \frac{\partial f}{\partial x_j} \left(a + \sum_{i=1}^{j-1} h_i e_i \right) h_j + o(h_j) = \nabla f(a) \cdot h + \sum_{j=1}^n \left(\frac{\partial f}{\partial x_j} \left(a + \sum_{i=1}^{j-1} h_i e_i \right) - \frac{\partial f}{\partial x_j}(a) \right) h_j + o(h)$$

and now we only need to show that

$$\sum_{j=1}^n \left(\frac{\partial f}{\partial x_j} \left(a + \sum_{i=1}^{j-1} h_i e_i \right) - \frac{\partial f}{\partial x_j}(a) \right) h_j = o(h)$$

but this follows from the continuity of the partial derivatives in a neighbourhood of a . Thus for $h \rightarrow 0$ we get that f is differentiable at a which completes the proof. $\square$

Remark 2.19. Let $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a function. If $\frac{\partial f_i}{\partial x_j}$ is continuous for all i, j then we say $f \in C^1(U, \mathbb{R}^m)$. This is the collection of the continuously differentiable functions from U to $\mathbb{R}^m$.

Remark 2.20. A function $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuously differentiable if and only if the map $x \mapsto Df(x)$ is continuous from U to the space of linear maps $L(\mathbb{R}^n, \mathbb{R}^m)$ with respect to the operator norm.

Example 2.2. We will show that the function $f: \text{GL}_n(\mathbb{R}) \rightarrow \text{GL}_n(\mathbb{R})$ defined by $f(A) = A^{-1}$ is differentiable. We see that for any matrices $A, H \in \text{GL}_n(\mathbb{R})$ we have

$$f(A+H) - f(A) = (A+H)^{-1} - A^{-1} = (A+H)^{-1}(A - (A+H))A^{-1} = -(A+H)^{-1}HA^{-1}.$$

We need to represent this expression using a linear function in H and an error function that is $o(H)$ so

$$-(A+H)^{-1}HA^{-1} = -A^{-1}HA^{-1} + (A^{-1} - (A+H)^{-1})HA^{-1}.$$

The function $H \mapsto -A^{-1}HA^{-1}$ is linear in H and since we know that $A \mapsto A^{-1}$ is continuous we get

$$\begin{aligned} \left\| \frac{(A^{-1} - (A+H)^{-1})HA^{-1}}{\|H\|} \right\| &\leq \frac{\|f(A) - f(A+H)\| \cdot \|H\| \cdot \|A^{-1}\|}{\|H\|} \\ &= \|f(A) - f(A+H)\| \cdot \|A^{-1}\| \xrightarrow{H \rightarrow 0} 0 \end{aligned}$$

which shows that $(A^{-1} - (A+H)^{-1})HA^{-1} = o(H)$. This means that f is differentiable with

$$Df(A)H = -A^{-1}HA^{-1}.$$

From this formula we see that Df is also continuous so f is even continuously differentiable.

Lemma 2.11. Consider $\det: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$. We have that $d \det(I)(A) = \text{tr}(A)$.

Proof. Let $\epsilon > 0$ and let A be any matrix. Denote the columns of A by A_1, A_2 and so on. We have that

$$\det(I + \epsilon A) = \det(e_1 + \epsilon A_1, e_2 + \epsilon A_2, \dots, e_N + \epsilon A_n).$$

Since the determinant is multilinear (linear with respect to the columns) we get that this is the sum of all 2^n terms, so that

$$\begin{aligned} \det(e_1 + \epsilon A_1, e_2 + \epsilon A_2, \dots, e_N + \epsilon A_n) &= \det(e_1, e_2, \dots, e_n) + \\ &\epsilon (\det(A_1, e_2, \dots, e_n) + \det(e_1, A_2, \dots, e_n) + \dots + \det(e_1, e_2, \dots, A_n)) + o(\epsilon) \end{aligned}$$

because the rest of the terms have at least two perturbed columns

$$\det(\epsilon A_1, \epsilon A_2, e_3, \dots, e_N),$$

which means that their determinant is of the form ϵ^2 times another term, so they are $o(\epsilon^2)$. Notice that

$$\det(e_1, \dots, A_i, \dots, e_n) = A_{ii}.$$

Therefore

$$\det(I + \epsilon A) = \det(I) + \text{tr}(A)\epsilon + o(\epsilon)$$

so since the trace is linear we get $d \det(I)(A) = \text{tr}(A)$. □

Lemma 2.12. For an invertible matrix A , we have $d \det(A)(T) = \det(A) \text{tr}(A^{-1}T)$.

Proof. Notice that

$$\det(X) = \det(AA^{-1}X) = \det(A) \det(A^{-1}X).$$

Therefore if we differentiate both sides at I we get

$$d \det(I)(X) = \det(A) d \det(I)(A^{-1}X)$$

so in particular by using Theorem 2.11 we get

$$d \det(A)(T) = \det(A) d \det(I)(A^{-1}T) = \det(A) \text{tr}(A^{-1}T).$$

□

Proposition 2.13 (Jacobi's formula). We have that $d \det(A) = \text{tr}(\text{adj } A dA)$.

Proof. If A is invertible we see from Theorem 2.12 with $T = dA$ that

$$d \det(A)(dA) = \det(A) \text{tr}(A^{-1}dA).$$

Using the formula

$$\text{adj}(A) = \det(A)A^{-1}$$

we get that

$$d \det(A)(dA) = \text{tr}(\text{adj } A dA).$$

□

3 Least Squares and Gradient Descent

Let $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a differentiable function. Suppose we are interested in finding a minimum (maybe a local minimum) of f , then there are two methods to solve find it using the tools we have acquired so far.

3.1 The standard method

We can use the fact that if f has a local minimum at a point $a \in U$ then $\nabla f(a) = 0$. We get n equations in n variables given by

$$\begin{aligned} \frac{\partial f}{\partial x_1}(a_1, \dots, a_n) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_n}(a_1, \dots, a_n) &= 0 \end{aligned}$$

This method is good because if we solve these equations, only get a single point where the gradient is 0, and know that there must be a minimum in U , then we can find it relatively easily.

3.2 Least squares

Consider the function $f(x) = \|Ax - b\|^2$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. When the rank of A is m we can solve the system of linear equations

$$(*) \quad Ax = b$$

and the minimum will be 0. This is also the case when the rank of A is lower than m . However the problem gets way more interesting when $m > n$ because this means that it's possible that there is no solution to $(*)$ so we need to find the solution with the smallest error. This problem is called the “least squares” problem because we are trying to find the least sum of the following squares

$$\|Ax - b\|^2 = \sum_i \left(\sum_j a_{ij}x_j - b_i \right)^2.$$

This problem is also sometimes called “linear regression” and it is very important in all fields of science and engineering. Since this is the simplest case of such a problem it has a closed solution. We notice that

$$f(x) = \langle Ax - b, Ax - b \rangle = \left\langle A^T Ax, Ax, - \right\rangle 2 \langle Ax, b \rangle + \langle b, b \rangle := g(x) + h(x)$$

for

$$\begin{aligned} g(x) &= \left\langle A^T Ax, Ax, = \right\rangle x^T A^T Ax \\ h(x) &= -2 \langle Ax, b \rangle + \langle b, b \rangle \end{aligned}$$

we therefore have that $\nabla f = \nabla g + \nabla h$. Since h is an affine function, its derivative is equal to its linear part:

$$Dh(x): h \mapsto -2 \langle Ah, b, = \rangle - 2b^T Ah \implies \nabla h = -2b^T A.$$

We also have that

$$g(x+h) - g(x) = \left\langle A^T A(x+h), x+h \right\rangle - \left\langle A^T Ax, x \right\rangle = 2 \left\langle A^T Ax, h \right\rangle + \|Ah\|^2$$

and we see that $2\langle A^T Ax, h \rangle$ is linear in h , and that $\|Ah\|^2 \leq \|A\|^2 \|h\|^2 = o(\|h\|)$ so it follows that

$$Dg(x): h \mapsto 2\langle A^T Ax, h \rangle = 2x^T A^T Ah \implies \nabla g(a) = 2x^T A^T A.$$

It now follows that

$$\nabla f(x) = 2x^T A^T A - 2b^T A.$$

To find the minimum of f , we now need to solve the equation

$$\nabla f(x) = 2x^T A^T A - 2b^T A = 0 \quad \text{or} \quad A^T Ax = A^T b.$$

Whenever $m > n$ we have that $\text{rank}(A) = n$ and thus the matrix $A^T A$ is also of rank n so it is invertible. In this case the solution to the equation is given by

$$x = (A^T A)^{-1} A^T b.$$

Remark 3.1. In practice we don't calculate the inverse of $A^T A$ since this is a rather complex numerical task, and instead we find the solution in other ways.

3.3 Gradient descent

In the example above, finding a solution to the system of linear equations $\nabla f(x) = 0$ was easy, however, in practice not all equations are linear, and we can't always find a closed solution.

The gradient descent method uses the fact that the direction of the gradient at a point is the direction of the steepest slope in order to approximate the minimum.

In general, given a function $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ the algorithm goes like this

- (1) Guess a solution x to the minimization problem and calculate $\nabla f(x)$ at x .
- (2) If $\|\nabla f(x)\|$ is 0 or "very close" to 0 then x is the minimum we are looking for and we stop the algorithm.
- (3) Otherwise, we set $x \mapsto x - c\nabla f(x)$. This moves x down the steepest slope, and since we usually set $0 < c < 1$ it doesn't go too far in that direction as well.
- (4) Return to step (2).

The steps above only describe the basic idea, and the implementation is usually more complex. For example, we need to define what "very close" means. We also set a maximal number of iterations so the computation will stop if it doesn't find a minimum.

4 Taylor's Theorem

4.1 Higher order partial derivatives

Theorem 4.1 (Schwarz's theorem). *Let $\Omega \subset \mathbb{R}^n$, let $f: \Omega \rightarrow \mathbb{R}$ and let $p \in \Omega$ be with some neighbourhood $p \in U \subset \Omega$. Suppose that f has continuous second partial derivatives in U , then for all i, j in $\{1, 2, \dots, n\}$*

$$\frac{\partial^2}{\partial x_i \partial x_j} f(p) = \frac{\partial^2}{\partial x_j \partial x_i} f(p).$$

4.2 Multi-index notation

An n -dimensional multi-index is an n -tuple

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

of non-negative integers. Define the following operations on some multi-indices $\alpha, \beta \in \mathbb{N}_0^n$ and a vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$:

Addition and Subtraction

$$\alpha \pm \beta = (\alpha_1 \pm \beta_1, \alpha_2 \pm \beta_2, \dots, \alpha_n \pm \beta_n)$$

Partial Order

$$\alpha \leq \beta \Leftrightarrow \alpha_i \leq \beta_i \quad \forall i \in \{1, \dots, n\}$$

Absolute Value

$$|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n$$

Factorial

$$\alpha! = \alpha_1! \cdot \alpha_2! \cdot \dots \cdot \alpha_n!$$

Binomial Coefficient

$$\binom{\alpha}{\beta} = \binom{\alpha_1}{\beta_1} \binom{\alpha_2}{\beta_2} \dots \binom{\alpha_n}{\beta_n} = \frac{\alpha!}{\beta! (\alpha - \beta)!}$$

Multinomial Coefficient

$$\binom{k}{\alpha} = \frac{k!}{\alpha_1! \alpha_2! \dots \alpha_n!} = \frac{k!}{\alpha!}, \quad k := |\alpha| \in \mathbb{N}_0$$

Power

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$$

Higher Order Partial Derivatives

$$D^\alpha = \partial^\alpha = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \dots \partial_n^{\alpha_n}$$

Remark 4.1. One use for multi-index notation is to describe the generalization of the binomial theorem for multiple variables:

$$(x_1 + \dots + x_n)^k = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha|=k}} \frac{|\alpha|!}{\alpha!} x^\alpha.$$

Definition 4.1 (Multilinear polynomial). A function $p: \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be a multilinear polynomial if

$$p(x) = \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| \leq k}} c_\alpha x^\alpha$$

where c_α are the coefficients of the polynomial. Each function of the form x^α is called a monome. The largest natural k so that there exists a multi-index α of order k so that $c_\alpha \neq 0$ is called the degree of the polynomial. The degree of the zero polynomial is defined to be $-\infty$.

4.3 Taylor's theorem

Theorem 4.2 (Multivariate version of Taylor's theorem). Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a k -times continuously differentiable function at the point $\mathbf{a} \in \mathbb{R}^n$. Then there exist functions $h_\alpha: \mathbb{R}^n \rightarrow \mathbb{R}$, where $|\alpha| = k$, such that:

$$f(\mathbf{x}) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(\mathbf{a})}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha + \underbrace{\sum_{|\alpha|=k} h_\alpha(\mathbf{x}) (\mathbf{x} - \mathbf{a})^\alpha}_{o(\|\mathbf{x} - \mathbf{a}\|^k)},$$

$$\text{and } \lim_{\mathbf{x} \rightarrow \mathbf{a}} h_\alpha(\mathbf{x}) = 0.$$

Remark 4.2 (Structure of the Taylor theorem). In Theorem 4.2 we see that the conditions are almost identical to the conditions we saw in the one-dimensional case, except now we increased the number of variables. The expression in the one-variable case is:

$$f(x) = \sum_{i=0}^k \frac{f^{(i)}(a)}{i!} (x - a)^i + h_k(x)(x - a)^k,$$

where $\lim_{x \rightarrow a} h_k(x) = 0$. The structure here should be more clear, we have the sum which gives us the approximate value up to the k th degree approximation, and we have the remainder part which is $h_k(x)(x - a)^k$. It makes sense to write it in this form since it must be $o(|x - a|^k)$, and then it is also clear why $\lim_{x \rightarrow a} h_k(x) = 0$. In the multivariable version the structure is still identical, and the intuition is the same, except we may have more than a single remainder function. This form of the remainder is called the *Peano form*.

Remark 4.3 (Taylor's theorem in praxis). In praxis, most exercises related to Taylor's theorem, are to find the Taylor expansion of some pathological function like $e^{f \cdot g}$ up to order k around the origin. In this case, it suffices to find the order of e^t , f and g up to order k around the origin, then multiply and compose them naturally, and if for example you end up with something like $x^4 + x^2y^2 + x^2y^3$ plus the remainder, and $k = 4$, you can discard the term x^2y^3 since it is in the family $o(\|(x, y)\|^4)$. This holds for any multiplication or composition of functions, which makes life easier.

5 Classifying Critical Points

5.1 Preliminary definitions

Suppose that $f: U \rightarrow \mathbb{R}$ such that $U \subset \mathbb{R}^n$ is open and $a \in U$.

Definition 5.1 (Local weak minimum). If there exists $r > 0$ such that for any point $x \in B_r(a)$ we have $f(a) \leq f(x)$ then the point is called a local weak minimum.

Remark 5.1. If the inequality is strict it is called a local strong maximum.

Remark 5.2. We define a local weak/strong maximum in the same manner.

Definition 5.2 (Critical point). If f is differentiable at a and $\nabla f(a) = 0$ then a is called a critical point of f .

Definition 5.3 (Saddle point). If a is a critical point of f and for any $r > 0$ exist $x, y \in B_r(a)$ such that $f(x) < f(a) < f(y)$ then a is called a saddle point of f .

Definition 5.4 (Hessian). Let $f \in C^2(U)$. We define the Hessian of f at a to be the following matrix:

$$H_f = H_f(a) = \left[\frac{\partial^2 f(a)}{\partial x_i \partial x_j} \right]_{i,j=1}^n.$$

Notice that by Clairaut's theorem we have that the Hessian is symmetrical and thus it defines a quadratic form as follows:

$$\langle H_f(a)v, v \rangle = \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) v_i v_j.$$

Using Taylor's theorem and the multi-index notation explained earlier we can express the function as such:

$$f(x+h) = f(x) + \sum_{i=1}^n D_i f(x) h_i + \sum_{|\alpha|=2} \frac{D^\alpha f(x) h^\alpha}{\alpha!} + o(\|h\|^2).$$

We can notice that:

$$|\alpha| = 2 \implies \alpha_1 + \dots + \alpha_n = 2,$$

so we can write:

$$\sum_{|\alpha|=2} \frac{D^\alpha f(x) h^\alpha}{\alpha!} = \sum_{i=1}^n \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2}(a) h_i^2 + \sum_{1 \leq i < j \leq n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j,$$

and by Clairaut's theorem:

$$\sum_{|\alpha|=2} \frac{D^\alpha f(x) h^\alpha}{\alpha!} = \sum_{i=1}^n \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2}(a) h_i^2 + \frac{1}{2} \sum_{1 \leq i \neq j \leq n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j = \frac{1}{2} \langle H_f(x)h, h \rangle,$$

so finally we get:

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(x)h, h \rangle + o(\|h\|^2).$$

If assume that x is a critical point of f , we can see that magic happens. We have $\nabla f(x) = 0$ and then:

$$f(x+h) - f(x) = \langle 0, h \rangle + \frac{1}{2} \langle H_f(x)h, h \rangle + o(\|h\|^2) \approx \frac{1}{2} \langle H_f(x)h, h \rangle.$$

Since $H_f(x)$ is symmetrical there exists an orthogonal matrix P so that $A := P^T H_f(x) P = \text{diag}(\lambda_1, \dots, \lambda_n)$, and under the coordinates system corresponding to the eigenvectors of the Hessian we get that:

$$f(x+h) - f(x) \approx \frac{1}{2} \langle Ah, h \rangle = \frac{1}{2} \sum_{i=1}^n \lambda_i h_i^2,$$

and in case $n = 2$ we have:

$$f(x + (a, b)) - f(x) \approx \frac{1}{2} (\lambda_1 a^2 + \lambda_2 b^2)$$

and we can use this formula to determine the type of critical points.

5.2 Classification of critical points

Proposition 5.1. *Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f \in C^2(U)$. Suppose that $a \in U$ is a critical point of f . Then*

- (1) *If $H_f(a)$ is positive definite then a is a local strong minimum.*
- (2) *If $H_f(a)$ is negative definite then a is a local strong maximum.*
- (3) *If $H_f(a)$ is indefinite then a is a saddle point.*

Remark 5.3. In other cases we can't tell anything about the point and it might be of any type.

Remark 5.4. It is easy to remember this classification since the Hessian of $f(x) = x^2$ is (2) which is positive definite and a is a local strong minimum.

Proof. Let a be a critical point of f such that $\nabla f(a) = 0$. We have

$$f(a+h) = f(a) + \frac{1}{2} \langle H_f(a)h, h \rangle + \epsilon(h) \text{ s.t. } \frac{\epsilon(h)}{\|h\|^2} \xrightarrow{h \rightarrow 0} 0.$$

Assume that $H_f(a) > 0$ (is positive definite). We get that

$$0 < c := \min_{\|v\|=1} \langle H_f(a)v, v \rangle = \min_{\|v\| \neq 0} \frac{\langle H_f(a)v, v \rangle}{\|v\|^2}.$$

It follows that for all h we have $\langle H_f(a)h, h \rangle \geq c\|h\|^2$. Let $r > 0$ such that $B_r(a) \subseteq U$ such that for $0 < \|h\| < r$ we have

$$\frac{|\epsilon(h)|}{\|h\|^2} < \frac{1}{2}c \implies |\epsilon(h)| < \frac{c}{2}\|h\|^2.$$

This implies that

$$f(a+h) = f(a) + \frac{1}{2} \langle H_f(a)h, h \rangle + \epsilon(h) \geq f(a) + \frac{c}{2}\|h\|^2 - |\epsilon(h)| > f(a)$$

which means that a is a strong minimum of f . Showing that $H_f(a) < 0$ implies that a is a strong maximum is very similar and thus this part of the proof is omitted.

Assume that $H_f(a)$ is neither positive definite nor negative definite. Let $u, v \in \mathbb{R}^n$ be unit vectors such that

$$\begin{aligned} b &:= \langle H_f(a)u, u \rangle < 0 \\ c &:= \langle H_f(a)v, v \rangle > 0 \end{aligned}$$

and choose $r > 0$ such that

$$0 < \|h\| < r \implies \left| \frac{\epsilon(h)}{\|h\|^2} \right| < \frac{1}{2} \min \{|b|, c\}.$$

For all $0 < t < r$ we get

$$f(a + tu) = f(a) + \frac{1}{2} \langle H_f(a)tu, tu \rangle + \epsilon(tu) \leq f(a) + \frac{b}{2}t^2 + |\epsilon(tu)| < f(a)$$

since for our t we have $|\epsilon(tu)| < \frac{1}{2}|b|t^2$. Similarly we can show that $f(a + tv) > f(a)$ which shows that a is a saddle point which completes the proof. $\square$

Remark 5.5. A simple way to check the sign of the Hessian, is by using Sylvester's criterion. With the notation that Δ_k denotes the determinant of the k th leading minor of a symmetric matrix $A \in \mathbb{R}^{n \times n}$ we have:

- (1) A is positive definite if and only if $\Delta_k > 0$ for all $k = 1, \dots, n$.
- (2) A is negative definite if and only if $(-1)^k \Delta_k > 0$ for all $k = 1, \dots, n$.
- (3) If there exists two minors Δ_k and Δ_{k+1} with negative determinants for some $1 \leq k \leq n$, then A is indefinite.

6 The Inverse Function Theorem

6.1 Preliminaries

Recall that a set $C \subset \mathbb{R}^n$ is called convex if and only if

$$\forall x, y \in C \quad \forall t \in [0, 1] \quad tx + (1 - t)y = y + t(x - y) \in C.$$

Intuitively, we say that C is convex if for every two points $a, b \in C$ the interval connecting the points is a subset of C .

Proposition 6.1. *Let $U \subset \mathbb{R}^n$ be convex and open, let $f: U \rightarrow \mathbb{R}^n$ be differentiable. If $f'(x)$ is bounded by M then,*

$$\forall x, y \in U \quad \|f(x) - f(y)\| \leq M\|x - y\|.$$

Proof. If $f(x) = f(y)$ then we are trivially finished, otherwise we can denote $v = \frac{f(x) - f(y)}{\|f(x) - f(y)\|}$ and the following functions:

$$\begin{aligned} g: \mathbb{R}^n &\rightarrow \mathbb{R} & g(u) &= \langle v, u \rangle \\ h: [0, 1] &\rightarrow U & h(t) &= y + t(x - y) \end{aligned}$$

differentiating gives:

$$Dg = v \quad \text{and} \quad Dh = x - y.$$

We can now define the function:

$$\varphi = g \circ f \circ h$$

and apply Lagrange's theorem to get that there exists $c \in (0, 1)$ such that:

$$\varphi'(c) = \frac{\varphi(1) - \varphi(0)}{1 - 0} = \langle v, f(x) \rangle - \langle v, f(y) \rangle = \|f(x) - f(y)\|.$$

On the other hand, we have:

$$|\varphi'(c)| = \|g'(f \circ h)(f \circ h)'(h)'\| \leq \|v\| \cdot M \cdot \|x - y\|,$$

but we know that $\|v\| = 1$ so we get:

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

as wanted. □

Exercise 6.1. Show that Theorem 6.1 still stands for the $\|\cdot\|_\infty$ norm. Start by finding a vector v such that $\|f(x) - f(y)\|_\infty = \langle v, f(x) - f(y) \rangle$.

Corollary 6.2. *If $f \in C^1$ then f is Lipschitz continuous on any compact subset of its domain.*

6.2 The inverse function theorem

Theorem 6.3 (Inverse function theorem). *Let $U \subset \mathbb{R}^n$ be open, let $f \in C^1(U, \mathbb{R}^n)$ and let $x \in U$. Suppose that $f'(x)$ is invertible. Then there exist open sets $x \in V \subset U$ and $f(x) \in W$ so that $f: V \rightarrow W$ is a bijection. Moreover, f^{-1} is continuously differentiable and we have*

$$(f^{-1})'(f(x)) = [f'(x)]^{-1} \quad \text{or} \quad (f^{-1})'(y) = [f'(f^{-1}(y))]^{-1}.$$

Proof. to be added. □

Remark 6.1. We know that f^{-1} is continuously differentiable since the formula for $(f^{-1})'(y)$ is a composition of f^{-1} and f' which are both continuous functions. We can even generalize and see that if $f \in C^k(U, \mathbb{R}^n)$ then $f^{-1} \in C^k(W, \mathbb{R}^n)$.

Remark 6.2. The geometric intuition for the theorem should be that the function behaves locally at x like an invertible linear map. The intuition for differentiation of f^{-1} is that since $f^{-1} \circ f$ is the identity, the linearization of f^{-1} at $f(x)$ must be the inverse of the linearization of f at x . This is exactly what the formula $(f^{-1})'(f(x)) = [f'(x)]^{-1}$ means. If you are still not convinced, here's an explanation with mathematical signs. If we approximate $f \approx A$ around x and $f^{-1} \approx B$ around $f(x)$ for $A, B \in \mathbb{R}^{n \times n}$ then $f(V) \approx AV$ and since $(f^{-1} \circ f)(V) = V$ we must have $BAV = V$. Thus, since A is invertible we get $B = A^{-1}$. This is exactly the formula $(f^{-1})'(f(x)) = [f'(x)]^{-1}$.

Remark 6.3. Although in the one dimensional case knowing that a differentiable function is locally invertible at any point implies that it is globally invertible, the statement fails in the n -dimensional case.

Example 6.1 (Locally invertible not invertible). Consider the function

$$f(x, y) := \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}.$$

We notice that

$$J_f(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}.$$

By a simple calculation

$$\det(J_f(x, y)) = e^{2x}(\cos^2 y + \sin^2 y) = e^{2x} \neq 0.$$

Since the function is not an injection it is not invertible, but according to the inverse function theorem, for any point a there exists a neighbourhood U_a of a where it is invertible and in particular injective.

7 Newton–Raphson Method

7.1 The Newton–Raphson Method

Let $f_1, \dots, f_n: \mathbb{R}^n \rightarrow \mathbb{R}$ be continuously differentiable functions. Assume we want to solve the following system of equations:

$$(*) \quad \begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}.$$

One real life example where we would want to solve such system of equation might be when we want to find our location using GPS. We can find our exact location using GPS satellites by using a method called trilateration. By calculating our distance from (at least) 3 satellites, and assuming some realistic conditions (e.g. our location is on the sphere of the Earth) we can find our exact location. Suppose satellite i is at location (x_i^1, x_i^2, x_i^3) and our location is (x^1, x^2, x^3) . Moreover suppose we can find the distance from our location to satellite i (r_i) by using some kind of clock and transceiver, we need to solve the following system of equations

$$\begin{cases} f_1(x_1^1, x_1^2, x_1^3) = (x_1^1 - x^1)^2 + (x_1^2 - x^2)^2 + (x_1^3 - x^3)^2 - r_1^2 = 0 \\ f_2(x_2^1, x_2^2, x_2^3) = (x_2^1 - x^1)^2 + (x_2^2 - x^2)^2 + (x_2^3 - x^3)^2 - r_2^2 = 0 \\ \vdots \\ f_n(x_n^1, x_n^2, x_n^3) = (x_n^1 - x^1)^2 + (x_n^2 - x^2)^2 + (x_n^3 - x^3)^2 - r_n^2 = 0 \end{cases}.$$

In order to easily find a solution for these equations later we need to have an initial guess for the solution. One way to obtain a good initial guess is to calculate the solution at some point in time because our location does not change very much even if the calculation takes a few seconds. Using this initial guess (which we may call a^1), we define an approximation sequence a^k by the follow rule

$$(**) \quad a^{k+1} = a^k - [f'(a^k)]^{-1} f(a^k).$$

Remark 7.1. In one dimension we get that

$$a^{k+1} = a^k - \frac{f(a^k)}{f'(a^k)}$$

which is the Newton–Raphson method for functions in a single variable, which we might have see earlier in analysis 1.

Proposition 7.1. Assume that there exists a solution $\bar{x}$ to a system of equations as in $(*)$, such that $f'(\bar{x})$ is invertible. Then there exists $r > 0$ such that for any initial guess $a^1 \in B_r(\bar{x})$ the Newton–Raphson approximation sequence $(**)$ converges to $\bar{x}$.

Remark 7.2. If $f'(\bar{x})$ is not invertible then for any sequence a^n that converges to $\bar{x}$ we get that $\| [f'(a^n)]^{-1} \|$ converges to infinity.

Proof. Since we have

$$[f'(\bar{x})]^{-1} f'(\bar{x}) = I \quad \text{and} \quad f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \epsilon(x - \bar{x})$$

such that $\epsilon(x - \bar{x}) = o(x - \bar{x})$ at a neighbourhood of $\bar{x}$ we can choose $r > 0$ such that for all $x \in B_r(\bar{x})$ we have

$$\| I - [f'(a^n)]^{-1} f'(\bar{x}) \| < \frac{1}{4} \quad \text{and} \quad \| \epsilon(x - \bar{x}) \| \leq \frac{1}{4} \| [f'(x)]^{-1} \|^{-1} \| x - \bar{x} \|.$$

We now get that for all $k \geq 1$

$$\begin{aligned}
 \|a^{k+1} - \bar{x}\| &= \left\| a^k - [f'(a^k)]^{-1} f(a^k) - \bar{x} \right\| \\
 &\leq \left\| a^k - \bar{x} - [f'(a^k)]^{-1} \left(f(\bar{x}) + f'(\bar{x}) (a^k - \bar{x}) \right) \right\| + \left\| [f'(a^k)]^{-1} \epsilon(a^k - \bar{x}) \right\| \\
 &= \left\| \left(I - [f'(a^k)]^{-1} f'(\bar{x}) \right) (a^k - \bar{x}) \right\| + \left\| [f'(a^k)]^{-1} \epsilon(a^k - \bar{x}) \right\|.
 \end{aligned}$$

Thus if $x^k \in B_r(\bar{x})$ then

$$\|a^{k+1} - \bar{x}\| \leq \frac{1}{4} \|a^k - \bar{x}\| + \frac{1}{4} \|a^k - \bar{x}\| = \frac{1}{2} \|a^k - \bar{x}\|.$$

Thus if $a^1 \in B_r(\bar{x})$ then $a^k \in B_r(\bar{x})$ for all k , and thus

$$\|a^{k+1} - \bar{x}\| \leq \frac{1}{2} \|a^k - \bar{x}\| \leq \cdots \leq \frac{1}{2^k} \|a^1 - \bar{x}\| \xrightarrow{k \rightarrow \infty} 0$$

which completes the proof. □

8 The Open Mapping Theorem

8.1 The open mapping theorem

Definition 8.1 (Regular function). Let $U \subset \mathbb{R}^n$. A function $f \in C^1(U, \mathbb{R}^m)$ is said to be regular at a point $a \in U$ if $\text{rank}(J_f(a)) = m$. The function is said to be regular in U if it is regular for any $a \in U$.

Remark 8.1. In Definition 8.1 there is a hidden assumption that $m \leq n$. It is also a very common conventions to say that f is regular at $a \in U$ if $\text{rank}(J_f(a)) = \min(n, m)$. In these notes we use the definition as written above. If $m \leq n$ then one may think of a regular point as a point where the derivative is surjective, and if $n \leq m$ then a regular point (under our conventions this is not a regular point) is a point where the derivative is injective.

Definition 8.2 (Open mapping). A mapping $f: U \rightarrow V$ is called open if for any open set $W \subset U$ the set $f(W) \subset V$ is also open.

Theorem 8.1 (Open mapping theorem). Let $U \subset \mathbb{R}^n$ be an open set, and let $f \in C^1(U, \mathbb{R}^m)$ be regular in U . Then f is an open mapping.

Proof. In the case of $m = n$ the theorem follows directly from the inverse function theorem. Since U is open, and because we know that for any $a \in U$ that $J_f(a)$ is invertible, then we get from the inverse function theorem that there exist open sets $a \in V_a \subset U$ and $f(a) \in W_a = f(V_a)$. We then have that:

$$f(U) = \bigcup_{a \in U} W_a$$

which is a union of open sets therefore an open set. In the case $V \subset U$ is open we can repeat this proof for $f|_V$ so that $f(V)$ is open, which completes this case of the proof. For any open subset $V \subset U$ we can use this exact proof by considering $f|_V$ instead.

Next we must have $m < n$ because if $n < m$ the rank of the Jacobian can't be m . Considering our case, the rank of the Jacobian is m which means it has m linearly independent columns. Without loss of generality we assume they are the first m columns and denote:

$$\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_m)} = \left[\frac{\partial f_i}{\partial x_j} \right]_{(i,j)=1}^m.$$

We now define a new transformation as such:

$$\begin{aligned} F: U &\rightarrow \mathbb{R}^n \\ F(x) &= (f(x), x_{m+1}, \dots, x_n) \end{aligned}$$

we see that the Jacobian of F is of the form

$$J_F = \begin{pmatrix} \frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_m)} & \star \\ 0 & I_{(n-m) \times (n-m)} \end{pmatrix}$$

and we notice that F satisfies the conditions for the inverse function theorem and thus it is an open mapping. We also notice that for every open set $V \subset U$ that:

$$f(V) = \pi(F(V))$$

where π is the projection $\pi: (x_1, \dots, x_n) \mapsto (x_1, \dots, x_m)$. Since it is clear that π is an open mapping we have that f is also an open mapping as a composition of two open mappings which completes the proof. $\square$

Remark 8.2. A general intuition for the theorem is that under the conditions of the theorem, by the inverse function theorem, f^{-1} is almost globally continuous on its domain. This implies intuitively that f should be almost an open mapping. The theorem shows that in fact it is an open mapping.

8.2 Constrained optimization

Proposition 8.2. Let $g \in C^1(U, \mathbb{R}^m)$ where $U \subseteq \mathbb{R}^n$ be regular, let $M := \{x \in U \mid g(x) = 0\}$ and let $f \in C(U) = C(U, \mathbb{R})$. If $a \in M$ and $f(a) \leq f(x)$ for any $x \in M$ or $f(a) \geq f(x)$ for any $x \in M$ then:

$$\nabla f(a) \in \text{span} \{ \nabla g_1(a), \dots, \nabla g_m(a) \}.$$

Remark 8.3. Notice that Theorem 8.2 is useful for finding the minimum and maximum of f on M . Thus if we want to find the minimum of a certain function on a plane M for example, we would first want to describe the plane as the zeros of a regular function g .

Remark 8.4. We can use this proposition to find the minimum/maximum of f on M by solving the following system of equations:

$$\begin{cases} \nabla f(a) = \sum_{i=1}^m \lambda_i \nabla g_i(a) \\ g(a) = 0 \end{cases}.$$

Notice that the first equation gives n equations in $n+m$ variables that are $a_1, \dots, a_n, \lambda_1, \dots, \lambda_m$ and the second equation gives m equations in n variables that are $a_1, \dots, a_n$. Thus we have $n+m$ equations in $n+m$ variables.

Remark 8.5. The variables λ_i for $1 \leq i \leq m$ are called *Lagrange multipliers*.

Example 8.1. Here's a simple example to how we can use Lagrange's multipliers in order to find the distance of a plane from the origin. The function we want to minimize is

$$f(x, y, z) = \|(x, y, z)\|_2 = \sqrt{x^2 + y^2 + z^2}$$

under the constraints of a plane. Let

$$g(x, y, z) = ax + by + cz - d$$

so that $M = \{a \in \mathbb{R}^3 \mid g(a) = ax + by + cz - d = 0\}$ be a parameterization of the plane. We can notice that our task is equivalent to finding the minimum of

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

on that plane. We see that g is indeed regular since $J_g(x, y, z) = (a, b, c) \neq 0$. In order to find the minimum of f on g we will solve the system of equations:

$$\begin{cases} \nabla f(x, y, z) = \sum_{i=1}^m \lambda_i \nabla g_i(x, y, z), \\ g(x, y, z) = 0. \end{cases}.$$

Since g is a function to $\mathbb{R}$ we only have one Lagrange multiplier. We can calculate the respective gradients of the functions:

$$\begin{aligned} \nabla f(x, y, z) &= (2x, 2y, 2z), \\ \nabla g(x, y, z) &= (a, b, c). \end{aligned}$$

The first equation gives $(2x, 2y, 2z) = \lambda(a, b, c)$ which implies $(x, y, z) = \frac{\lambda}{2}(a, b, c)$. Putting that in the second equation gives

$$g\left(\frac{\lambda}{2}(a, b, c)\right) = \frac{\lambda}{2}(a^2 + b^2 + c^2) - d = 0 \implies \lambda = \frac{2d}{a^2 + b^2 + c^2},$$

so the minimum is

$$(x, y, z) = \frac{d}{a^2 + b^2 + c^2}(a, b, c),$$

and the minimal distance of the plane from the origin is:

$$\|(x, y, z)\| = \left\| \frac{d}{a^2 + b^2 + c^2}(a, b, c) \right\| = \frac{d}{\|(a, b, c)\|}.$$

In fact what we have shown so far is not a complete proof because the theorem only implies that the equations hold at a minimum but it doesn't necessarily mean that the point we found is a minimum point. In order to complete the proof we can choose an arbitrary point on the plane $P_0 = (x_0, y_0, z_0)$, denote $R = 2\|P_0\|$ and consider the set:

$$S = \left\{ x \in \mathbb{R}^3 \mid g(x) = 0 \right\} \cap \overline{B_R(0)}.$$

This set is clearly compact and since $P_0 \in S$ we know that it's not empty and thus we know that f has a minimum in that set, and moreover, we know a priori that this minimum must be the global minimum of the function. Since we only found one point satisfying the equations given by the theorem it must be that minimum point and now the proof is complete.

Proof of Theorem 8.2. In order to prove the theorem we will prove its contrapositive. Suppose that

$$\nabla f(a) \notin \text{span} \{ \nabla g_1(a), \dots, \nabla g_m(a) \}.$$

We need to show that a is not a minimum point. Define the function:

$$\begin{aligned} F: U &\rightarrow \mathbb{R}^{m+1} \\ F(x) &= (f(x), g(x)) \end{aligned}$$

the Jacobian will be:

$$J_F(a) = \begin{bmatrix} \nabla f(a) \\ \nabla g_1(a) \\ \vdots \\ \nabla g_m(a) \end{bmatrix}.$$

Since we assumed that $\nabla f(a)$ is linearly independent from the rest of the gradients at a the rank of the Jacobian at a must be $m+1$ and since the rank function is continuous we know that there exists a neighborhood U of a where the Jacobian rank is $m+1$ and thus we can apply the open mapping theorem and get that F is an open mapping in that neighborhood. We know that $F(a) = (f(a), 0)$ is a point in the interior of $F(U)$ so for a sufficiently small epsilon $(f(a) - \varepsilon, 0) \in F(U)$ and thus there exists $x \in U$ so that $F(x) = (f(a) - \varepsilon, 0) = (f(x), g(x))$ which means that

$$(1) \quad g(x) = 0$$

$$(2) \quad f(x) < f(a)$$

and thus a is not a minimum point in M which completes the proof. $\square$

Exercise 8.1.

- (1) Let $A \in \mathbb{R}^{n \times n}$. Show that there exists a vector $\lambda \in \mathbb{R}^n$ so that

$$\|a\|_{\text{op}} = \|\lambda\|_{\infty}.$$

- (2) Find $\|A\|_{\text{op}}$ for

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Solution.

- (1) Recall that

$$\|a\|_{\text{op}} = \sup_{\|x\|=1} \|Ax\|.$$

Since $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is a compact set we know that the maximum must be attained. We will find the maximum of $f(x) = \|Ax\|^2$ under the constraint $g(x) = \|x\| - 1 = 0$. We have that g is regular since as we know from Example 2.1 that

$$Dg(x) = 2x^T \neq 0.$$

Thus, from Theorem 8.2 we get that the maximum is attained at x so that for some $\lambda \in \mathbb{R}$ we have

$$\nabla f(x) = \lambda \nabla g(x) = \lambda 2x^T.$$

Since $\|Ax\|^2 = x^T A^T A x$ we recall from Theorem 8.2 again that $\nabla f(x) = 2x^T A^T A$ so if x is a maximum then

$$2x^T A^T A = \lambda 2x^T \implies x^T A^T A = \lambda x^T \implies A^T A x = \lambda x.$$

This means that the maximum x must be an eigenvector of $A^T A$. Since we know it exists, we get that

$$\|a\|_{\text{op}} = \|\lambda\|_{\infty}$$

where $\lambda \in \mathbb{R}$ is the eigenvalue of roots of eigenvalues of $A^T A$ (these are known as singular values). Since $A^T A$ is symmetric, its eigenvalues are positive and thus applying the root is well defined. Notice that this result matches exactly what we have learned in linear algebra.

- (2) Using (1) we can compute

$$A^T A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The eigenvalues of $A^T A$ are 0 and 2 thus

$$\|A\|_{\text{op}} = \left\| (\sqrt{2}, 0) \right\|_{\infty} = \sqrt{2}.$$

Definition 8.3 (Level set). Let $f: U \rightarrow \mathbb{R}^m$ so that $U \subset \mathbb{R}^n$. A level set of f at some $c \in \text{im}(f)$ is the set

$$L_c(f) := \{(x_1, \dots, x_n) \mid f(x_1, \dots, x_n) = c\}.$$

When the number of independent variables is two, a level set is called a level curve, when it is three, it is called a level surface, for higher values it is called a level hypersurface.

Proposition 8.3. *Let $g: U \rightarrow \mathbb{R}^m$ so that $U \subset \mathbb{R}^n$ be differentiable and let $a \in U$ so that $g(a) = c$. Then the gradient of g at a is either zero, or perpendicular to the tangent space of the level set of g at a .*

Proof. Let $r(t) = (r_1(t), \dots, r_n(t))$ be a curve on the level set $L_c(g)$ with $r(t_0) = a$. Define

$$h(t) := g(r_1(t), \dots, r_n(t)) = c.$$

Using the chain rule we get that

$$\frac{dh}{dt} = \frac{\partial g}{\partial r_1} \bigg|_a \frac{dr_1}{dx_1} \bigg|_{t_0} + \frac{\partial g}{\partial r_2} \bigg|_a \frac{dr_2}{dx_2} \bigg|_{t_0} + \dots + \frac{\partial g}{\partial r_n} \bigg|_a \frac{dr_n}{dx_n} \bigg|_{t_0} = 0.$$

This means that

$$\left(\frac{\partial g}{\partial r_1} \bigg|_a \quad \frac{\partial g}{\partial r_2} \bigg|_a \quad \dots \quad \frac{\partial g}{\partial r_n} \bigg|_a \right) \cdot \left(\frac{dr_1}{dx_1} \bigg|_{t_0} \quad \frac{dr_2}{dx_2} \bigg|_{t_0} \quad \dots \quad \frac{dr_n}{dx_n} \bigg|_{t_0} \right) = 0$$

or in other words $\nabla g|_a \cdot r'(t_0) = 0$. We have shown that the dot product of the gradient g at a and the tangent of any curve that lies on the level set $L_c(g)$ and passes through a is zero which means they are perpendicular or that one of them is zero, which implies that the gradient of g at a is either zero, or perpendicular to the tangent space of the level set of g at a which completes the proof. $\square$

Remark 8.6 (Intuition for lagrange multipliers). As we have seen in Theorem 8.3, the gradient of a multivariate function gives the direction of a vector perpendicular to the tangent space of the level set of the function at that point (consider a paraboloid).

When we are looking for a minimum point x in M we are actually looking for a point x contained in the intersection of M and some level set of f (if x is not in M it is not a minimum in M and if it is not in any level set it is not part of the domain of f), such that the tangent space of g at x is the same (up to scalar multiplication) as the tangent space of the level set of f at x (think about Example 8.1 or a simpler example like $f(x, y) = \|(x, y)\|$ and $g(x, y) = y - ax - b$). If the tangent space of g and f at a are the same direction, this means that the gradient of f at a (which is perpendicular to the tangent space of the level set of f at a) and the gradient of g at a (which is perpendicular to the tangent space of g at a) are parallel, so $\nabla f(a) = \sum_{i=1}^m \lambda_i \nabla g_i(a)$, where λ_i are to make their sizes equal as well.

Do not let Example 8.1 fool you. All of this geometry is only done only in the domain of f and g . This is why their domain must be of the same dimension. In Example 8.1 the geometry (tagent space and level sets calculations) is done in $\mathbb{R}^3$ but when talking about a paraboloid this is done in $\mathbb{R}^2$ not $\mathbb{R}^3$ (try working out the calculations for a paraboloid with some constraint).

As a final note, there is no expectation to understand why the minimum is attained where we say it should be attained when visualizing the recommended examples from the previous paragraphs.

Proposition 8.4 (Stronger version of Theorem 8.2). *Let $F \in C^1(U, \mathbb{R}^m)$ and let $x_0 \in U$ be such that $F(x_0) = 0$ and F is regular at x_0 . Then if x_0 is a local extremum of $f \in C^1(U, \mathbb{R})$ on the hypersurface $M := \{x \in U \mid F(x) = 0\}$, there exists real numbers $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ so that*

$$\nabla f(a) = \sum_{i=1}^m \lambda_i \nabla g_i(a).$$

Remark 8.7. Using Langrange's multipliers, we can show multiple basic cases of famous inequalities like Young's inequality, and using that we can further show Hölder's inequality, and also the Minkowski inequality.

Definition 8.4 (Lagrangian). In the context of constrained optimization, when we want to minimize $f(x)$ subject to a constraint $g(x) = 0$, we define the Lagrangian function by $\mathcal{L}(x, \lambda) := f(x) - \langle \lambda, g(x) \rangle$.

Remark 8.8 (Meaninglessness of the Lagrangian). The Lagrangian is not a new concept, it is simply a convenient way to encode the data of the problem, since the equations in Remark 8.4 are given by the simple equality $\Delta \mathcal{L} = 0$. That is because $\Delta_x \mathcal{L} = \Delta f(x) - \lambda \Delta g(x) = 0$ implies that the gradients of f and g are parallel, and $\Delta_\lambda \mathcal{L} = g(x) = 0$ assures the constraint is satisfied.

Remark 8.9. Note that if we know that f attains a maximum/minimum on a domain $\overline{D}$ for some open $D \subset \mathbb{R}^n$, then one approach is to find the critical points in D by $\Delta f = 0$ and then parameterize ∂D and use lagrange multipliers to find the critical points in $\overline{D} - D = \partial D$.

9 The Implicit Function Theorem

9.1 The implicit function theorem

So far we have talked about explicit functions - functions that take the elements of one set and transform them into an element of the other. In contrast, we can consider a new way of looking at functions using implicit functions. First, we will consider implicit equations.

Definition 9.1 (Implicit equation). An implicit equation is a relation of the form

$$R(x_1, \dots, x_n) = 0$$

where R is a multivariable function. A vector $x \in \mathbb{R}^n$ that satisfies $R(x) = 0$ is called a *solution*.

Remark 9.1. We say that the implicit equation $F(x, y)$ where $(x, y) \in U \subset \mathbb{R}^n \times \mathbb{R}^m$ defines y as an implicit function of x if exist $V_1 \times V_2 \subset U$ such that for any $x \in V_1$ there exists a unique $y \in V_2$ such that $F(x, y) = 0$. Intuitively the function maps each $x \in V_1$ to its corresponding unique y .

Remark 9.2. We denote

$$\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_m)}(x, y) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(x, y) & \cdots & \frac{\partial f_1}{\partial y_m}(x, y) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x, y) & \cdots & \frac{\partial f_m}{\partial y_m}(x, y) \end{pmatrix}.$$

Theorem 9.1 (The implicit function theorem). *Let $U \subset \mathbb{R}^n \times \mathbb{R}^m$ be open, let $f \in C^1(U, \mathbb{R}^m)$, let (a, b) be a solution of $f(x, y) = 0$ and also assume that $\det \frac{\partial(f_1, \dots, f_m)}{\partial(y_1, \dots, y_m)}|_{(a, b)} \neq 0$. Then there exists a neighborhood $(a, b) \in V \times W \subset U$ and a unique function $g \in C^1(V, \mathbb{R}^m)$ such that:*

$$\forall (x, y) \in V \times W \quad f(x, y) = 0 \iff y = g(x)$$

and $g(a) = b$ and $f(x, g(x)) = 0$ for every x near a .

Remark 9.3. The meaning of the theorem is that under these conditions, the set of solutions $f = 0$ is a graph of a continuously differentiable function in n variables. This remark is important, make sure you understand it well.

Remark 9.4. Here is some intuition for the theorem. Suppose we had an implicit equation $F(x_1, \dots, x_n) = 0$, and we wanted to extract k of the variables for some $0 < k < n$ as a function in the rest of the $n - k$ variables. This means that once we give $n - k$ variables value, there is only one solution to $F(x_1, \dots, x_n) = 0$. Since we have $n - (n - k) = k$ variables, we must have k equations, so the 0 in the equation $F(x_1, \dots, x_n) = 0$ must be $0 \in \mathbb{R}^k$. Thus we must have $F: \mathbb{R}^{n-k} \times \mathbb{R}^k \rightarrow \mathbb{R}^k$. Suppose $(a, b) \in \mathbb{R}^{n-k} \times \mathbb{R}^k$ is a solution of $F(x, y) = 0$. Now we notice that the condition $\det \frac{\partial(f_1, \dots, f_m)}{\partial(y_1, \dots, y_m)}|_{(a, b)} \neq 0$ means that any solution (x, y) in a sufficiently small neighbourhood V of (a, b) is determined uniquely by x . Moreover, this condition means that for any x in a neighbourhood of x there exists a y in a neighbourhood of b so that (x, y) is a solution. This implies that there exists a function g in a neighbourhood of a so that $f(x, y) = 0 \iff y = g(x)$ for $(x, y) \in V$.

Example 9.1. Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the function defined by

$$f(x, y) = x^2 + y^2 - 1.$$

Consider the point $(a, b) = (\sqrt{2}/2, \sqrt{2}/2)$. Indeed, it is a solution of $f(x, y)$. We also have that

$$\frac{\partial f}{\partial y}(a, b) = 2y|_{(a,b)} = 2y|_{a,b} = 2b = \sqrt{2} \neq 0.$$

In a sufficiently neighbourhood of (a, b) we notice that we can define the function $g(x) = \sqrt{1 - x^2}$ so that

$$f(x, y) = 0 \iff y = \sqrt{1 - x^2}.$$

Similarly for $(a', b') = (-\sqrt{2}/2, \sqrt{2}/2)$ in a sufficiently neighbourhood of (a', b') we notice that we can define the function $g'(x) = -\sqrt{1 - x^2}$ so that

$$f(x, y) = 0 \iff y = -\sqrt{1 - x^2}.$$

However, this may not be possible for when $y = 0$. Indeed

$$\frac{\partial f}{\partial y}(1, 0) = \frac{\partial f}{\partial y}(-1, 0) = 0$$

and the conditions of the implicit function theorem do not hold, so we can't write y as a function in x . However since

$$\frac{\partial f}{\partial x}(1, 0) \neq 0 \quad \text{or} \quad \frac{\partial f}{\partial x}(-1, 0) \neq 0$$

we see that the conditions hold and indeed we can write x as a function in y in both cases as in $x = \sqrt{1 - y^2}$ or $x = -\sqrt{1 - y^2}$.

Example 9.2. Let $A \in \mathbb{R}^{m \times k}$ and $B \in \mathbb{R}^{m \times m}$. Then we can define the map $F: \mathbb{R}^{k+m} \rightarrow \mathbb{R}^m$ by

$$F(x, y) = Ax + By \in \mathbb{R}^m.$$

It is clear that $D_x F = A$ and $D_y F = B$. The equation $F(x, y) = 0$ is a homogeneous system of equations with m equations and $k + m$ variables. Notice that $(0, 0)$ is the trivial solution of $F(x, y) = 0$. We want to show that we can treat $x_1, \dots, x_k$ as free variables and then compute $y_1, \dots, y_m$ using a linear function in the free variables. If $D_y F = B$ is invertible, then its determinant is not zero so $\det \frac{\partial(f_1, \dots, f_m)}{\partial(y_1, \dots, y_m)}|_{(a,b)} \neq 0$ and the conditions of the implicit function theorem are fulfilled. Indeed, we can see that $y = -B^{-1}Ax$, which is a continuously differentiable function in k variables $g(x) = -B^{-1}Ax$.

9.2 Differentiation of implicit functions

Theorem 9.2 (Implicit function differentiation). *Let $U \subset \mathbb{R}^n \times \mathbb{R}^m$ be an open set, let $f \in C^1(U, \mathbb{R}^m)$, let (a, b) a solution of $f(x, y) = 0$ and also assume that $\det \frac{\partial(f_1, \dots, f_m)}{\partial(y_1, \dots, y_m)}|_{(a,b)} \neq 0$. Then the derivative of the function g from the implicit function theorem is:*

$$Dg(a) = -[D_y f(a, b)]^{-1} D_x f(a, b).$$

Remark 9.5. Notice that this theorem is very useful because it allows us to find $g'(a)$ without knowing what is $g(a)$ itself.

Remark 9.6. Compare this formula with Example 9.2.

Exercise 9.1. Show that there exists an open neighbourhood of $(u_0, v_0) = (1, 0)$ so that we can write the solutions of the system of equations

$$\begin{cases} x^2 + y^2 + u^2 + v^2 = 21 \\ 2x + yu + v = 0 \end{cases}$$

using functions $u(x, y)$ and $v(x, y)$ continuously differentiable around $(x_0, y_0) = (2, -4)$ and find $u_x(2, -4)$ and $v_x(2, -4)$.

Solution. First, we want to find a relation that describes all solutions to this system. We can define $G: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by

$$G(x, y, u, v) = (x^2 + y^2 + u^2 + v^2 - 21, 2x + yu + v)$$

which is smooth and $(x_0, y_0, u_0, v_0) = (2, -4, 1, 0)$ is indeed a solution. Since we want to extract $u(x, y)$ and $v(x, y)$ we need to verify that $\det \frac{\partial(G_1, G_2)}{\partial(u, v)}|_{(2, -4, 1, 0)} \neq 0$. Indeed

$$\frac{\partial(G_1, G_2)}{\partial(u, v)}|_{(2, -4, 1, 0)} = \begin{pmatrix} 2u & 2v \\ y & 1 \end{pmatrix} \xrightarrow{\text{at } (2, -4, 1, 0)} \begin{pmatrix} 2 & 0 \\ -4 & 1 \end{pmatrix}$$

which is invertible so the conditions of the implicit function theorem are satisfied. Thus there exist neighbourhoods $(2, -4) \in V$ and $(1, 0) \in W \subset \mathbb{R}^2$ and a continuously differentiable function $g: V \rightarrow W$ given by

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

so that

$$G(x, y, u, v) = 0 \iff (u, v) = (u(x, y), v(x, y))$$

for all $(x, y, u, v) \in V \times W$. We now use the formula for implicit function differentiation

$$\begin{aligned} Dg(x, y) &= -[D_{u,v}G(x, y, u, v)]^{-1} D_{x,y}G(x, y, u, v) \\ &= -\begin{pmatrix} 2u & 2v \\ y & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2x & 2y \\ 2 & u \end{pmatrix} \\ &= \dots \\ &= \begin{pmatrix} \frac{2x-4v}{2vy-2u} & \frac{2u-2uv}{2vy-2u} \\ \frac{-2xy+4u}{2vy-2u} & \frac{-2y^2+2u^2}{2vy-2u} \end{pmatrix}. \end{aligned}$$

Therefore

$$u_x(x, y) = \frac{2x - 4v(x, y)}{2v(x, y)y - 2u(x, y)} \quad \text{and} \quad v_x(x, y) = \frac{-2xy + 4u(x, y)}{2v(x, y)y - 2u(x, y)}.$$

Remark 9.7. Note that in the previous exercise we found a way to compute all derivatives of $u(x, y)$ and $v(x, y)$, so we can find their Taylor expansion and find approximated solutions.

Proposition 9.3. Let $U \subset \mathbb{R}^n$ be open and let $f \in C^1(Y, \mathbb{R}^m)$. Show that if $n > m$, then f is not injective.

Proof. We can show this by induction of m , but we will only show the base case. Let $m = 1$ and $n > m$. Then if the gradient $\nabla f(x) = 0$ for all x we get that f is constant and in particular not injective. Suppose that there exists x^0 so that $\nabla f(x^0) \neq 0$, then $\frac{\partial f}{\partial x_k}(x^0) \neq 0$ for some k , without loss of generality let $k = n$. We see that the function

$$F(x) = f(x) - f(x^0)$$

satisfies the conditions of the implicit function theorem at x^0 . Thus there exists an interval $x_n^0 \in I \subset \mathbb{R}$ and a neighbourhood $(x_1^0, \dots, x_{n-1}^0) \in V \subset \mathbb{R}^{n-1}$ so that there exists a unique continuously differentiable function $g: V \rightarrow I$ so that

$$F(x) = 0 \iff x_n = g(x_1, \dots, x_{n-1})$$

for all $x \in V \times I$. Therefore, for all $(x_1^0, \dots, x_{n-1}^0) \neq (x_1, \dots, x_{n-1}) \in V$ we get that

$$F(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) = 0$$

which means that

$$f(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) = f(x^0).$$

This shows that f is not injective. □

10 Finding Roots of Polynomials

10.1 Finding roots of polynomials

Definition 10.1 (Simple root). Let $p(x) = \sum_{i=0}^n a_i x^i$ be a polynomial. A root of the polynomial with multiplicity of 1 is called a simple root.

Proposition 10.1. Let $p(x) = \sum_{i=0}^n a_i x^i$ be a polynomial with real coefficients. Then the roots of $p(x)$ are continuously dependent on its coefficients.

Remark 10.1. We want to show that if x_0 is a simple root of a polynomial $P_a(x) = \sum_{i=0}^n a_i x^i$ such that $a = (a_0, \dots, a_n)$ then there exists a neighborhood V of a in $\mathbb{R}^{n+1}$ such that one of the roots of the polynomial

$$P_b(x) = \sum_{i=0}^n b_i x^i$$

is given as a function in the coefficients $g \in C^1(\mathbb{R}^{n+1})$ for all $b \in V$.

Proof. Define the function $f: \mathbb{R}^{n+1} \times \mathbb{R} \rightarrow \mathbb{R}$ by

$$\begin{aligned} f: \mathbb{R}^{n+2} &\rightarrow \mathbb{R} \\ f(b_0, \dots, b_n, x) &= P_b(x) \end{aligned}$$

We notice that the implicit equation $f(b, x) = 0$ describes all the roots of the polynomial $P_b(x)$. Let x_0 be the given simple root of $P_a(x)$ so that we have

$$f(a, x_0) = 0$$

and since x_0 is a simple root we also know that:

$$\left. \frac{df}{dx} \right|_{(a, x_0)} = \left. \frac{dP_a}{dx} \right|_{x_0} \neq 0.$$

Therefore, the conditions for the implicit function theorem are satisfied and we can conclude that there exist neighborhoods V of a and I of x and a function $g \in C^1(V, I)$ such that:

$$\forall (b, x) \in V \times I \quad P_b(x) = 0 \iff x = g(b).$$

□

Remark 10.2. This theorem is of great importance because when using computers and numerical methods to find roots using approximated coefficients we want to know that the roots we find are a good approximation of the exact roots and this theorem shows exactly that.

11 Manifolds

11.1 C^1 -manifolds

Definition 11.1 (C^1 -manifold). Let $k, n \in \mathbb{N}$ such that $k \leq n$. A subset $M \subset \mathbb{R}^n$ is called a C^1 manifold of dimension k (a k -manifold) if for every $a \in M$ there exist open sets $U, V \subset \mathbb{R}^n$ such that $a \in U$ and $V \cap \mathbb{R}^k \times \{0_{n-k}\} \neq \emptyset$ and there exists a regular function $f \in C^1(U, V)$ such that:

$$f(M \cap U) = V \cap \mathbb{R}^k \times \{0_{n-k}\} = \{x \in V \mid x_{k+1} = \cdots = x_n = 0\}.$$

Remark 11.1. Intuitively, a set $M \subset \mathbb{R}^n$ is a C^1 -manifold of dimension k , if locally (at every point), it is similar to $\mathbb{R}^k$. For example, a closed simple curve $\gamma: [0, 1] \rightarrow \mathbb{R}^n$ locally looks like the real line $\mathbb{R}$.

Remark 11.2. Recall that the definition of being regular is that in any point in the domain the derivative at that point is a surjective linear transformation, or in other words, the rank of the Jacobian is equal to its row number.

Remark 11.3. Because f is a transformation between two open sets of the same dimension we know that f is regular if and only if its derivative is invertible at every point.

Definition 11.2 (Diffeomorphism). A function $f \in C^1$ that is invertible and regular is called a diffeomorphism. From the inverse function theorem the inverse of a diffeomorphism is also a diffeomorphism.

Definition 11.3 (Diffeomorphic sets). Two sets $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are called diffeomorphic if there exists a bijective diffeomorphism $f: U \rightarrow V$.

Example 11.1. Diffeomorphic sets can live in Euclidean spaces of different dimension. For example if $A := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ and $B := \{(x, y, z) \in \mathbb{R}^3 \mid z = x^2 + y^2 \text{ and } (x, y) \in A\}$ are diffeomorphic, and the diffeomorphism between them is $g(x, y) = (x, y, x^2 + y^2)$.

Remark 11.4. A 1-dimensional manifold is sometimes called a *differential curve* and a 2-dimensional manifold is sometimes called a *differential surface*.

Remark 11.5. The title of this section is indeed “manifolds” but in fact we are only talking about *embedded manifolds*. That is to say, we are not talking about manifolds in the abstract sense at all, only about those who are specifically embedded in the Euclidean space.

Proposition 11.1. *The following conditions are equivalent:*

- (1) M is a C^1 manifold of dimension k ;
- (2) For every $a \in M$ there exist a neighborhood U_a and a regular function $g \in C^1(U, \mathbb{R}^{n-k})$ such that:

$$M \cap U = g^{-1}(0) := \{x \in U \mid g(x) = 0\};$$

- (3) Up to permutation of the variables, for every $a \in M$ there exists a neighborhood $a \in V \times W$ such that $V \subset \mathbb{R}^k$ and $W \subset \mathbb{R}^{n-k}$ and there exists $h \in C^1(V, \mathbb{R}^{n-k})$ such that:

$$M \cap (V \times W) = \text{graph}(h) = \{(x, h(x)) \mid x \in V\};$$

- (4) For each $a \in M$ there exist a neighborhood U , an open set $V \subset \mathbb{R}^k$ and an injective function $H \in C^1(V, \mathbb{R}^n)$ such that:

- (a) $\text{rank } DH(v) = k$ for all $v \in V$.
- (b) $M \cap U = H(V)$

(c) The function $H^{-1}: H(V) \rightarrow V$ is continuous with respect to the topology on $H(V)$ induced by $\mathbb{R}^n$.

Remark 11.6. The equivalency between (1) and (2) is known as preimage theorem or regular level set theorem. It is helpful to know all of these equivalencies because for example proving that

$$\text{SL}(2) := \left\{ A \in \mathbb{R}^{2 \times 2} \mid \det(A) = 1 \right\}$$

in $(\mathbb{R}^4, \|\cdot\|_{\text{HS}})$ is a manifold of dimension 3 is immediate by the preimage theorem but very difficult by any other definition.

Remark 11.7. The function H from (4) is a diffeomorphism and it is called a *parametrization* or a *chart* of $M \cap U$ at p . Indeed, we also require charts to be compatible, that is, if there exist $H_i: V_i \rightarrow \mathbb{R}^n$ and $H_j: V_j \rightarrow \mathbb{R}^n$ which overlap, then we require the *transition map* $H_i^{-1} \circ H_j$ to be C^1 .

Example 11.2. The graph of the function $f(x) = |x|$ is not a C^1 -manifold. That is because at $(0, 0)$ for all neighbourhoods $(0, 0) \in V \times W$ such that $V, W \subset \mathbb{R}$ there does not exist a function $h \in C^1(V, \mathbb{R})$ such that:

$$M \cap (V \times W) = \text{graph}(h) = \{(x, h(x)) \mid x \in V\}.$$

This may seem obvious, but in fact all the curves given by the functions

$$\begin{aligned} g_1(x, y) &= y^2 - x^2 = 0 \\ g_2(x, y) &= y^2 - x^3 - x^2 = 0 \\ g_3(x, y) &= y^3 - x^2 = 0 \end{aligned}$$

are not C^1 -manifolds, because they all similarly fail to satisfy the property of C^1 -manifolds (locally similar to the graph of a C^1 function) at the points where they are not regular. On the other hand, $g_4(x, y) = x^3 - y^3 = 0$ is not regular at the origin but it is the curve defined by $y = x$ and thus it is a C^1 -manifold.

Example 11.3. Let $S^n = \{x \mid \|x\| = 1\} \subset \mathbb{R}^{n+1}$ be the n -dimensional sphere. We will show that this sphere is a n -dimensional manifold.

According to 2

For any $a \in M$ we can choose the open set $U = \mathbb{R}^{n+1} \setminus \{0\}$ such that $x \in U$. Define the function $g: U \rightarrow \mathbb{R}$ given by

$$g(x) = \left(\sum_{i=1}^{n+1} x_i^2 \right) - 1.$$

We see that

$$\nabla g = 2(x_1, \dots, x_{n+1}),$$

so g is regular and continuously differentiable in U . Finally we check that

$$S^n \cap U = S^n = \{x \in U \mid g(x) = 1\},$$

and this verifies that S^n is an n -dimensional manifold.

According to 3

Let $a \in M$, without loss of generality and allowing permutation of variables we can assume that $a_{n+1} < 0$. Now choose $V = \mathbb{B}_n$ and $W = (-\infty, 0)$. Define:

$$\begin{aligned} h: V &\rightarrow \mathbb{R} \\ h(v_1, \dots, v_n) &= -\sqrt{1 - \sum_{i=1}^n v_i^2} \end{aligned}$$

we can verify that indeed:

$$S^n \cap (\mathbb{B}_n \times (-\infty, 0)) = \{(v, h(v)) \mid v \in V\}$$

According to 4

We will show only the case of $S^n = S^2$. Suppose we want to find a parametrization at the point $(1, 0, 0)$. We define the sets:

$$V = (0, \pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

$$U = \{(x, y, z) \mid x > 0\}$$

and the function:

$$H(\theta, \phi) = (\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta)).$$

We can see that H is indeed a continuously differentiable continuous injection. We see that:

$$DH = J_H = \begin{pmatrix} \cos(\theta) \cos(\phi) & -\sin(\theta) \sin(\phi) \\ \cos(\theta) \sin(\phi) & \sin(\theta) \cos(\phi) \\ -\sin(\theta) & 0 \end{pmatrix}$$

In our case $k = 2$, so seeing that the determinant of the upper is $\sin(\theta) \cos(\theta)$ which is 0 only when $\theta = \frac{\pi}{2}$ but then the matrix is:

$$\begin{pmatrix} 0 & \sin(\phi) \\ 0 & \cos(\phi) \\ -1 & 0 \end{pmatrix}$$

Which is a matrix of rank 2 so we conclude that $\text{rank } DH = 2$ for every value pair (θ, ϕ) in V . Conditions $\langle b \rangle$ and $\langle c \rangle$ also hold after some algebraic manipulation.

Remark 11.8. Every open set $U \subset \mathbb{R}^n$ is a manifold of dimension n and every vector space $V \subset \mathbb{R}^n$ is a manifold of dimension $\dim(V)$.

Definition 11.4 (Continuously differentiable function on a manifold). Let $M \subset \mathbb{R}^n$ be a manifold of dimension k . A function $g: M \rightarrow \mathbb{R}^m$ is called continuously differentiable if for every $a \in M$ and every parametrization $H(V)$ at a we have

$$g \circ H \in C^1(V, \mathbb{R}^m).$$

Exercise 11.1. Show that Definition 11.4 does not depend on the choice of the coordinate system.

Exercise 11.2. Show that Definition 11.4 is equivalent to “There exists a neighborhood $a \in U$ and a function $G \in C^1(U, \mathbb{R}^m)$ such that $G|_{M \cap U} = g|_{M \cap U}$ where g is the C^1 function.”

Proposition 11.2. Let $U \subset \mathbb{R}^n$ be an open set and let $F \in C^1(U, \mathbb{R}^n)$. We define the set

$$M := \{x \in \mathbb{R}^n \mid F(x) = 0\}.$$

Suppose that for all $x \in M$ we have

$$\dim(\ker DF(x)) = k < n.$$

Then M is a k -dimensional C^1 -manifold.

Example 11.4. We will show that the space

$$\mathrm{SO}_n(\mathbb{R}) = \left\{ A \in \mathbb{R}^{n \times n} \mid A^T A = I \right\}$$

is a manifold and find its dimension. First define $F: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ by $F(A) = A^T A - I$, so we can identify $\mathrm{SO}_n(\mathbb{R})$ with the zeros of F . Let $A \in M$, then we know that

$$DF(A)(H) = A^T H + H^T A,$$

so the kernel of $DF(A)$ is all matrices $H \in \mathbb{R}^{n \times n}$ so that

$$A^T H = -H^T A = -(A^T H)^T.$$

Since each solution is of the form $H := (A^T)^{-1} K$ where K is antisymmetric, and A is invertible since $A \in \mathrm{SO}_n(\mathbb{R})$, the kernel is isomorphic to the space of antisymmetric matrices, and so

$$\dim(\ker DF(A)) = \frac{n(n-1)}{2}.$$

Thus from Theorem 11.2 we get that $\mathrm{SO}_n(\mathbb{R})$ is a manifold with dimension $n(n-1)/2$.

11.2 The torus

Example 11.5 (Torus). The torus is a 3d shape resembling a doughnut or a swim ring that also happens to be a very interesting topological object. Denote the big radius of the torus a , and the small radius b . The torus can be described by the equation

$$z^2 + (r - a)^2 = b^2,$$

and this can be seen perhaps more easily by changing to cylindrical coordinates. However, we can also describe it in a more intuitive way using parameterization. For $u, v \in [0, 2\pi]$ define

$$\varphi(u, v) = \begin{pmatrix} \cos v & 0 & -\sin v \\ 0 & 1 & 0 \\ \sin v & 0 & \cos v \end{pmatrix} \begin{pmatrix} a + b \cos u \\ b \sin u \\ 0 \end{pmatrix} = \begin{pmatrix} (a + b \cos u) \cos v \\ b \sin u \\ (a + b \cos u) \sin v \end{pmatrix}.$$

In order to understand the intuition behind this parameterization, we need to notice that the column vector we are multiplying is a parameterization of a circle on the xy -plane around some point a on the x -axis with radius b , and that we are multiplying it by a rotational matrix that rotates around the y -axis, which is exactly what we wanted.

12 The Tangent Space

12.1 The tangent space

Definition 12.1 (Continuous path). A continuous function $f: [a, b] \rightarrow \mathbb{R}^n$ is called a continuous path. If f is also differentiable we say it is a continuously differentiable path. The “1-dimensional” set $f([a, b])$ is called a *curve*.

Remark 12.1. If $f: [a, b] \rightarrow \mathbb{R}^n$ is a continuously differentiable path and $f'(t_0) \neq 0$ then we say that $f'(t_0)$ is the tangent vector of f at t_0 . The map

$$t \mapsto f(t_0) + f'(t_0)(t - t_0)$$

is the only affine map that approximates $t \mapsto f(t)$ with error that is $o(t - t_0)$.

Remark 12.2. Intuitively, we want to think of the tangent space as the collection of all the directions we can move in. It is useful to understand intuitively why this is also what we did so far when we talked about tangent lines and tangent planes. Moreover, if the manifold is an open set, then we can move in all directions, which is why the following definition for a tangent space in $\mathbb{R}^n$ should be intuitive.

Definition 12.2 (Tangent space). For each point $p \in \mathbb{R}^n$ we define the tangent space to $\mathbb{R}^n$ in p as:

$$T_p(\mathbb{R}^n) = \{v_p = (p, v) \mid v \in \mathbb{R}^n\}$$

Remark 12.3. This should be thought of as a copy of $\mathbb{R}^n$ with p as its origin.

Remark 12.4. We denote $(p, v) = v_p$ and sometimes we identify it with v itself. If $V \subset \mathbb{R}^n$ is a vector space, then $V_p = \{(p, v) \mid v \in V\} \subset T_p(\mathbb{R}^n)$, and we can identify it with the affine space $V_p = p + V = \{p + v \mid v \in V\}$. We now proceed to define $T_p(\mathbb{R}^n)$.

Definition 12.3 (Tangent space). Let M be a C^1 manifold of dimension k , and let $p \in M$ be a point on the manifold. The tangent space to M in point p is defined as:

$$T_p(M) = \left\{ p + \gamma'(t_0) \mid \begin{array}{l} C^1 \ni \gamma: (a, b) \rightarrow M \\ \exists t_0 \in (a, b) \quad \gamma(t_0) = p \end{array} \right\},$$

or sometimes, in the nonaffine version

$$T_p(M) = \left\{ \gamma'(t_0) \mid \begin{array}{l} C^1 \ni \gamma: (a, b) \rightarrow M \\ \exists t_0 \in (a, b) \quad \gamma(t_0) = p \end{array} \right\}.$$

Remark 12.5. Note that the definition of the tangent space of $\mathbb{R}^n$ at a point is a special case of the definition of a tangent space at a point on a manifold.

Remark 12.6 (Tangent space criterion of manifolds). The tangent space of a manifold at a point is an affine space, it is a vector space translated from the origin by the base point of the tangent space. The dimension of the tangent space of a manifold at a point, is equal to the dimension of the manifold. This is a very helpful criterion when we try to show that some space is not a manifold.

Example 12.1. Remark 12.6 does not hold when we consider tangent spaces for $E \subset \mathbb{R}^n$ which are not manifolds. For example consider the set $E = \{(x, y) \mid y = |x|\}$. We have that for any point other than the origin the tangent space has one dimension and it is either $y = x$ or $y = -x$. However, although there exist differentiable curves $\gamma: (-\epsilon, \epsilon) \rightarrow E$ so that $\gamma(0) = (0, 0)$ like $\gamma(t) = (t^3, |t|^3)$, all these curves satisfy $\gamma'(0) = 0$ so $T_0(E) = \{0\}$ which is of dimension zero.

Remark 12.7. The differential df of a differentiable map $f: U \rightarrow \mathbb{R}^m$ with $E \subset U$ maps the tangent space of a point $x \in E$ to a tangent space based at $f(x) \in \mathbb{R}^m$. We can see this by noticing that $df(x): \gamma'(0) \mapsto df(x)(\gamma'(0))$ so by definition of the tangent spaces we get $df(x): T_x(E) \rightarrow T_{f(x)}f(E)$. If f is a diffeomorphism of E and $f(E)$ then $df(x)$ is an isomorphism between $T_x(E)$ and $T_{f(x)}f(E)$. In particular, if E is open then $df(x)$ is an isomorphism between $\mathbb{R}^n$ and $T_{f(x)}f(E)$.

Proposition 12.1 (Tangent space computation theorem). *Let M be an m -dimensional C^1 manifold in $\mathbb{R}^n$ and let $x_0 \in M$. Then $T_{x_0}M$ is an m -dimensional subspace of $\mathbb{R}^n$, and if we consider $T_{x_0}M$ as a set of vectors based at x_0 then it is an affine m -dimensional subspace of $\mathbb{R}^n$. If at a neighbourhood W of x_0 the manifold $M \cap W$ is given as a level set $F^{-1}(0) = \{x \in W \mid F(x) = 0\}$ of some $F \in C^1(W, \mathbb{R}^{n-m})$ with $\text{rank } DF(x_0) = n - m$, then $T_{x_0}M = \ker(df(x_0))$ or, in the affine version $T_{x_0}M = x_0 + \ker(df(x_0))$.*

Example 12.2. We want to find the tangent space of the paraboloid $S = \{z = x^2 + y^2/4\}$ at some point (x_0, y_0, z_0) . We notice that S is a level set $F(x, y, z) = 0$ of $F(x, y, z) = x^2 + y^2/4 - z$. We compute

$$\nabla F(x, y, z) = \begin{pmatrix} 2x \\ \frac{1}{2}y \\ -1 \end{pmatrix} \neq 0$$

so by Theorem 12.1 the tangent plane is given by

$$T_{(x_0, y_0, z_0)}(S) = \left\{ (x, y, z) \in \mathbb{R}^3 \mid 2x_0x + \frac{1}{2}y_0y - z = 0 \right\}.$$

or in the affine version as

$$T_{(x_0, y_0, z_0)}(S) = \left\{ (x, y, z) \in \mathbb{R}^3 \mid 2x_0(x - x_0) + \frac{1}{2}y_0(y - y_0) - (z - z_0) = 0 \right\}.$$

Example 12.3. Let $U \subset \mathbb{R}^n$ and let $f \in C^1(U, \mathbb{R})$. The graph

$$\text{graph}(f) = \{(\mathbf{x}, y) \mid y = f(\mathbf{x})\}$$

is a level set $F(\mathbf{x}, y) = 0$ for $F(\mathbf{x}, y) = f(\mathbf{x}) - y$. We compute

$$\nabla F(\mathbf{x}, y) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \\ -1 \end{pmatrix} \neq 0$$

so the tangent plane is given by

$$T_{\mathbf{x}}(\text{graph}(f)) = \frac{\partial f}{\partial x_1}(\mathbf{x}) \cdot (x_1 - \mathbf{x}_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \cdot (x_n - \mathbf{x}_n) - (y - y_0) = 0$$

or equivalently by

$$\frac{\partial f}{\partial x_1}(\mathbf{x}) \cdot (x_1 - \mathbf{x}_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \cdot (x_n - \mathbf{x}_n) = y - y_0.$$

Example 12.4 (Correspondence of SL_n and $\mathfrak{sl}_n$). Let $\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n \times n} \mid \det(A) = 1\}$. This is a level set $\det(A) = 1$. First we compute $T_I \mathrm{SL}_n(\mathbb{R})$. Let $\gamma: (-\epsilon, \epsilon) \rightarrow \mathrm{SL}_n(\mathbb{R})$ be a differentiable curve with $\gamma(0) = I$, and denote $D\gamma(0) = A$ so that $\gamma(t) = I + At + o(t)$. We know from Theorem 2.11 that

$$1 = \det(\gamma(t)) = \det(I + At + o(t)) = \det(I + At + o(t)) = 1 + \mathrm{tr}(A)t + o(t)$$

which implies that $\mathrm{tr}(A) = 0$. This means that $T_I \mathrm{SL}_n(\mathbb{R}) \subseteq \{A \in \mathbb{R}^{n \times n} \mid \mathrm{tr}(A) = 0\} = \mathfrak{sl}_n$. On a not so related note, under the Hilbert–Schmidt inner product we see that $\langle I, A \rangle = \mathrm{tr}(A) = d \det(I)$ which means that $\nabla \det(I) = I$ and indeed $\nabla \det(I)$ is a normal to $T_I \mathrm{SL}_n(\mathbb{R})$ since for any $A \in \mathfrak{sl}_n(A)$ we get $\langle \nabla \det(I), A \rangle = \mathrm{tr}(A) = 0$. On a related note, for every $A \in T_I \mathrm{SL}_n(\mathbb{R})$ we can construct a curve $\gamma(t) = \exp(tA) \in \mathrm{SL}_n(\mathbb{R})$ so that $\gamma(0) = I$ and $\gamma'(0) = A$.

The transition from $\mathrm{SL}_n(\mathbb{R})$ to $T_I \mathrm{SL}_n(\mathbb{R})$ can be thought of the transition from the Lie group $\mathrm{SL}_n(\mathbb{R})$ to the Lie algebra $\mathfrak{sl}_n$.

Remark 12.8. We can produce any Lie algebra by considering the tangent space of some Lie group at its unit. Lie algebras and Lie groups are not part of this course's material.

Example 12.5. We will find the tangent space of the space

$$\mathrm{SO}_n(\mathbb{R}) = \left\{ A \in \mathbb{R}^{n \times n} \mid A^T A = I \right\}$$

at the point I . By Example 11.4 we know that the dimension of SO_n is $n(n-1)/2$ so by Remark 12.6 we get

$$\dim(T_I(\mathrm{SO}_n(\mathbb{R}))) = \frac{n(n-1)}{2}.$$

Let $\gamma: (-\epsilon, \epsilon) \rightarrow \mathrm{SO}_n(\mathbb{R})$ be a continuously differentiable path so that $\gamma(0) = I$. Recall that $F(A) := A^T A - I$. For all $s \in (-\epsilon, \epsilon)$ we have

$$G(s) := F(\gamma(s)) = 0.$$

Thus if we differentiate both sides we get

$$DG(s) = DF(\gamma(s))\gamma'(s) = \gamma(s)^T \gamma'(s) + \gamma'(s)^T \gamma(s) = 0.$$

In particular

$$DG(0) = I^T \gamma'(s) + \gamma'(s)^T = \gamma'(s) + \gamma'(s)^T = 0,$$

therefore any vector in the tangent space must be an antisymmetric matrix. This means that

$$T_I(\mathrm{SO}_n(\mathbb{R})) \subset \left\{ A \in \mathbb{R}^{n \times n} \mid A^T = -A \right\},$$

but since the dimension of both spaces is equal we get an equality.

Definition 12.4 (Tangent bundle). The tangent bundle of a manifold M is given by the union of all tangent spaces

$$T(M) = \bigsqcup_{x \in M} T_x(M) = \{(x, y) \mid x \in M \text{ and } y \in T_x(M)\}.$$

Remark 12.9. There is a natural projection $\pi: TM \rightarrow M$ defined by $\pi(x, v) = x$.

Example 12.6. We know that for any open set $U \subset \mathbb{R}^n$ that $T_x(U) = \mathbb{R}^n$ for any $x \in U$. This shows that $T(U) = U \times \mathbb{R}^n$.

Remark 12.10. The tangent bundle of a smooth n -dimensional manifold is a smooth manifold of dimension n . This fact is not really important in this course, and we will not prove it, but it will be helpful to know this for differential geometry.

Definition 12.5 (Tangent space). Let $H: V \subset \mathbb{R}^k \rightarrow \mathbb{R}^n$ be a parametrization of a manifold M around p , and let $q \in V$ be a point such that $H(q) = p$. Then $T_p(M) = [\text{im}(DH(q))]_p$.

Example 12.7. Consider half a circle and the parametrization:

$$H: \left(\frac{\pi}{2}, \frac{3\pi}{2} \right) \rightarrow \mathbb{R}^2$$

$$H(t) = (\cos(t), \sin(t))$$

Choose $p = (-1, 0) = H(\pi)$. We see that:

$$DH(\pi) = \left(\begin{array}{c} -\sin(t) \\ \cos(t) \end{array} \right)_{t=\pi} = \left(\begin{array}{c} 0 \\ -1 \end{array} \right) : \mathbb{R} \rightarrow \mathbb{R}^2$$

we notice that indeed:

$$\text{im}(DH(\pi)) = \text{span} \left\{ \left(\begin{array}{c} 0 \\ -1 \end{array} \right) \right\}$$

$$\implies T_p(M) = \left\{ p + a \left(\begin{array}{c} 0 \\ -1 \end{array} \right) \mid a \in \mathbb{R} \right\}$$

Remark 12.11. Below are more equivalent definitions for the tangent space.

Definition 12.6 (Tangent space). Let $p \in M$ so that there exists U with $p \in U$. Denote $p = (a, b)$ such that $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^{n-k}$ for which

$$U \cap M = \{ (v, h(v)) \mid v \in V \subseteq \mathbb{R}^k \}.$$

Then we define $T_p(M)$ to be the graph of the affine approximation of h at a .

$$T_p(M) = \{ (v, h(a) + Dh(a)(v - a)) \mid v \in \mathbb{R}^k \} \subset \mathbb{R}^n.$$

Remark 12.12. This definition is a special case of Definition 12.5 where $H: V \rightarrow \mathbb{R}^n$ and $H(v) = (v, h(v))$.

Definition 12.7 (Tangent space). Let $g \in C^1(\mathbb{R}^n, \mathbb{R}^{n-k})$ be a regular function such that $M \cap U = \{x \mid g(x) = 0\}$. Then we define

$$T_p(M) = [\ker Dg(p)]_p = \{x \in \mathbb{R}^n \mid Dg(p)(x - p) = 0\}.$$

Definition 12.8 (Tangent space). Let $f \in C^1(U, \mathbb{R}^n)$ be an invertible and regular function (a diffeomorphism), so that

$$f(M \cap U) = V \times \{0\} \subseteq \mathbb{R}^k \times \mathbb{R}^{n-k}.$$

Then we define

$$T_p(M) = \{p + Df(p)^{-1}(v, 0) \mid v \in \mathbb{R}^k\}.$$

Definition 12.9 (Normal). Let $N \subset \mathbb{R}^3$ be a smooth surface. A vector $0 \neq v \in \mathbb{R}^3$ is said to be normal to M at a point $p \in M$ if it is orthogonal to $T_p(M)$.

Remark 12.13. The tangent plane of a surface at a point is completely determined by the normal vector at that point, and conversely, the normal vector at a point is determined up to sign by the tangent plane.

Remark 12.14. If M is equal to the set of roots of a regular function g in a neighborhood of a then a vector normal to M must be of the form $\alpha \nabla g(p)$. That is because both the normal vector $\mathbf{n}$ and the gradient satisfy

$$T_p(M) = \left\{ x \in \mathbb{R}^3 \mid \nabla g(p) \cdot (x - p) = 0 \right\} = \left\{ x \in \mathbb{R}^3 \mid \nabla \mathbf{n} \cdot (x - p) = 0 \right\}.$$

In $\mathbb{R}^3$ since the normal to a plane is 1-dimensional, the normal vector and the gradient must be colinear.

Example 12.8 (Viviani's curve). Viviani's curve is the set of points satisfying the conditions

$$\begin{aligned} x^2 + y^2 + z^2 &= 1, \\ x^2 + y^2 &= x. \end{aligned}$$

We will show that if C' is Viviani's curve then $C := C' \setminus (1, 0, 0)$ is a 1-dimensional manifold. Define the function $F(x, y, z) = (x^2 + y^2 + z^2 - 1, x^2 + y^2 - x)$. We notice that

$$C = \{(x, y, z) \mid z \neq 0 \text{ and } F(x, y, z) = 0\}.$$

We see that

$$DF(x, y, z) = \begin{pmatrix} 2x & 2y & 2z \\ 2x - 1 & 2y & 0 \end{pmatrix}$$

so $\dim(DF(x, y, z)) \neq 2$ if and only if $z = 0$, and in C we have $z \neq 0$ so $\dim(DF(x, y, z)) = 2$. In particular $\dim(\ker DF(x, y, z)) = 1$ so by Theorem 11.2 we get that C is a 1-dimensional manifold. We will show that C' is not a 1-dimensional manifold.

Assume, by way of contradiction, that C' is a 1-dimensional manifold. Then by Remark 12.6 the dimension of the tangent space at any point, in particular at $(1, 0, 0)$ is 1. Consider the continuously differentiable path $\gamma: \mathbb{R} \rightarrow \mathbb{R}^3$ defined by

$$\gamma(t) = (\cos^2(t), \sin(t) \cos(t), \sin(t)).$$

We can verify that $\gamma(t)$ is on Viviani's curve, and in particular $\gamma(0) = \gamma(\pi) = (1, 0, 0)$. This implies that

$$\gamma'(0), \gamma'(\pi) \in T_{(1,0,0)}(C').$$

However, we see that

$$\gamma'(t) = (-2 \cos(t) \sin(t), \cos(2t), \cos(t)).$$

This means that $\gamma'(0) = (0, 1, 1)$ and $\gamma'(\pi) = (0, 1, -1)$, but these vectors are linearly independent so $\gamma'(0), \gamma'(\pi) \in T_{(1,0,0)}(C')$ is a contradiction, which completes the proof.

13 Manifolds With Boundary

13.1 Manifolds with boundary

Defining manifolds is very convenient, but many interesting sets we want to analyze are not manifolds like any closed interval or half a sphere. That is why in this section we introduce a new object that is almost like a manifold, but slightly different.

Definition 13.1 (Half-space). The half k -dimensional space is defined as:

$$\mathcal{H}_k = \{x \in \mathbb{R}^n \mid x_n \geq 0\}$$

Definition 13.2 (Manifold with boundary). A subset $M \subset \mathbb{R}^n$ is called a C^1 k -manifold with boundary if for every $a \in M$ there exists a neighbourhood $a \in U$, an open set $V \subset \mathbb{R}^k$ and an injection $H \in C^1(V, \mathbb{R}^n)$ such that:

- (1) $\text{rank } DH(v) = k$ for all $v \in V$.
- (2) $M \cap U = H(V \cap \mathcal{H}_k)$.
- (3) $H^{-1}: M \cap U \rightarrow V \cap \mathcal{H}_k$ is continuous with respect to the topology induced by $\mathbb{R}^n$.

The function H is called a *parametrization* or a *chart* of M at a . As in the case of an open manifold, the transition maps must be compatible.

Remark 13.1. In Definition 13.2 we could have replaced the condition $H \in C^1(V, \mathbb{R}^n)$ with $H \in C^1(V \cap \mathcal{H}_k, \mathbb{R}^n)$. Indeed, this definition is more common in practice because it is more convenient.

Definition 13.3 (Boundary and interior). Let $M \subset \mathbb{R}^n$ be a k -manifold with boundary. A point $a \in M$ is called a boundary point if $a = H(b)$ where $b \in \partial \mathcal{H}_k$. We denote the set of the boundary points of M as ∂M . The rest of the points are called interior points and we denote them as $\text{Int}(M)$.

Remark 13.2. The topological definition of boundary is different from the definition of a boundary for a manifold!

Example 13.1. Let $M = [0, 1) \subset \mathbb{R}$ be a 1-manifold with boundary. In this case the topological boundary of M is $\partial_{\text{top}} M = \{0, 1\}$, but the manifold boundary is $\partial M = \{0\}$.

Remark 13.3. The boundary of a manifold with boundary does not depend on its parameterization.

Example 13.2. The following are manifolds with boundary.

- (1) The image of any curve $\gamma: [0, 1] \rightarrow \mathbb{R}^n$ with the boundary being $\gamma(0)$ and $\gamma(1)$ if the curve is not closed.
- (2) Any closed box in $\mathbb{R}^n$.
- (3) A hemisphere.

Remark 13.4. The boundary of a k -dimensional manifold with boundary is a $k-1$ -dimensional manifold.

13.2 The tangent space at the boundary

At Section 12 we saw the multiple definitions of a tangent space to a manifold without boundary. Given a manifold M and a point $p \in M$ we have

$$T_p(M) = \left\{ \gamma'(t_0) \mid \begin{array}{l} C^1 \ni \gamma: (a,b) \rightarrow M \\ \exists t_0 \in (a,b) \quad \gamma(t_0) = p \end{array} \right\}.$$

Next, suppose that M is a manifold with boundary and let $p \in \partial M$. We define the following sets:

$$\begin{aligned} C_p^-(M) &:= \{ \gamma: [0, \epsilon) \mid \gamma \text{ is differentiable with } \gamma(0) = p \}, \\ C_p^+(M) &:= \{ \gamma: (-\epsilon, 0] \mid \gamma \text{ is differentiable with } \gamma(0) = p \}. \end{aligned}$$

Since $p \in \partial M$, there exists a neighbourhood $p \subset U$ of p so that $U \cap M \cong \mathbb{H}$, by some homeomorphism $h: M \rightarrow \mathbb{H}$. We now define the equivalence relation

$$\gamma \sim \alpha \iff (h \circ \gamma)'(0) = (h \circ \alpha)'(0),$$

where the derivatives are by necessity the one-sided derivatives. This essentially tells us that we only care about the derivative of the curve $h \circ \gamma$ at the point $h(p)$, which is the same thing we are saying in the definition of $T_p(M)$.

Definition 13.4 (Tangent space at the boundary). Let M be a manifold with boundary, let $p \in \partial M$, and let $C_p^-(M)$, $C_p^+(M)$, and $\sim$ be as stated above. We then define:

$$T_p(M) := (C_p^-(M) \cup C_p^+(M)) / \sim.$$

Remark 13.5. We can also define the sets:

$$T_p^-(M) := C_p^-(M) / \sim \quad \text{and} \quad T_p^+(M) := C_p^+(M) / \sim.$$

And subsequently,

$$T_p^{\text{outward}}(M) := T_p^+(M) \setminus T_p(\partial M) \quad \text{and} \quad T_p^{\text{inward}}(M) := T_p^-(M) \setminus T_p(\partial M).$$

Remark 13.6. Note that $T_p(\partial M) := T_p^-(M) \cap T_p^+(M)$.

Remark 13.7. Note that in contrast to Section 12, this time we are working with the curves themselves and not the velocity vectors (their derivative at the point of interest). This is a slightly more general, and more correct way to think about the tangent space, since in general manifolds aren't always subsets of $\mathbb{R}^n$. However, it should be clear that both methods are equivalent here.

Example 13.3. Let $M = \{(x, y) \mid x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$ be a 2-dimensional manifold with boundary. We know that for $p \in \text{Int}(M)$ we have $T_p(M) = \mathbb{R}^2$. Let $p \in \partial M = S^1 := \{(x, y) \mid x^2 + y^2 = 1\}$. The curves at $T_p^+(M)$ are “leaving” the circle, so the velocity vector there is “pointing outwards”. Similarly the velocity vectors associated with $T_p^-(M)$ point inwards. Indeed we also have $T_p^-(M) \cap T_p^+(M) = T_p(\partial M) \cong \mathbb{R}$. In contrast, we have that $T_p(M) \cong \mathbb{R}^2$.

14 Length of a Path

14.1 Length of a path

Definition 14.1 (Length of a path). Let $f: [a, b] \rightarrow \mathbb{R}^n$ be a continuous path, and let $P = (t_0 = a < \dots < t_k = b)$. Define:

$$l(f, P) = \sum_{i=1}^k \|f(t_i) - f(t_{i-1})\| \quad \text{and} \quad l(f) = \sup_P l(f, P).$$

If $l(f) < \infty$ we say that f has a length of $l(f)$.

Proposition 14.1. If $f \in C^1([a, b], \mathbb{R}^n)$ then the path f has length.

Proposition 14.2. If $f \in C^1([a, b], \mathbb{R}^n)$ then for every partition P we have

$$l(f, P) \leq \int_a^b \|f'(t)\| dt \quad \text{and} \quad l(f) = \int_a^b \|f'(t)\| dt.$$

Proof. Let $P = \{a < b\}$. By the fundamental theorem of calculus we have that

$$\|f(b) - f(a)\|^2 = \left\| \begin{pmatrix} \int_a^b f'_1(t) dt \\ \int_a^b f'_2(t) dt \\ \vdots \\ \int_a^b f'_n(t) dt \end{pmatrix} \right\|^2 = \sum_{j=1}^n \left(\int_a^b f'_j(t) dt \right)^2 = \sum_{j=1}^n \left(\int_a^b f'_j(s) ds \right) \left(\int_a^b f'_j(t) dt \right).$$

We can now use Cauchy–Schwarz inequality to obtain

$$\begin{aligned} \sum_{j=1}^n \left(\int_a^b f'_j(s) ds \right) \left(\int_a^b f'_j(t) dt \right) &= \int_a^b \sum_{j=1}^n \left(\int_a^b f'_j(s) ds \right) f'_j(t) dt \\ &\leq \int_a^b \underbrace{\left[\sum_{j=1}^n \left(\int_a^b f'_j(s) ds \right)^2 \right]^{\frac{1}{2}}}_{\|f(b) - f(a)\|} \cdot \underbrace{\left[f'_j(t)^2 \right]^{\frac{1}{2}}}_{\|f'(t)\|} dt. \end{aligned}$$

Thus, by dividing by $\|f(b) - f(a)\|$ we get that

$$\|f(b) - f(a)\| \leq \int_a^b \|f'(t)\| dt.$$

This implies that for any two points in a partition $P = \{a = t_0 < t_1 < \dots < t_n = b\}$ we get

$$\|f(t_i) - f(t_{i-1})\| \leq \int_{t_{i-1}}^{t_i} \|f'(t)\| dt \implies l(f, P) \leq \int_a^b \|f'(t)\| dt.$$

This completes the first part of the proof. Now define for $\tau \in [a, b]$ the function

$$L(\tau) = l(f: [a, \tau] \rightarrow \mathbb{R}^n)$$

which is well defined because we know it is finite from what we've just shown. Let $h > 0$, then we see that

$$\frac{f(\tau + h) - f(\tau)}{h} \leq \frac{L(\tau + h) - L(\tau)}{h} \leq \frac{1}{h} \int_{\tau}^{\tau+h} \|f'(t)\| dt \xrightarrow{h \rightarrow 0} \|f'(\tau)\|$$

where the first inequality holds because the length between two points is always longer or equal to their distance, and the second inequality holds from what we showed above. We also notice that the LHS also converges to $\|f'(t)\|$ as $h \rightarrow 0$ which means that L is differentiable at any point and we have

$$L'(t) = \|f'(t)\|.$$

Using the Newton–Leibniz theorem we get

$$L(b) = \underbrace{L(a)}_{=0} + \int_a^b L'(t) dt = \int_a^b \|f'(t)\| dt.$$

□

Remark 14.1. Consider the map $\gamma(t) = (\cos(t), \sin(t))$. Although it defines the same curve over $[0, 2\pi]$ and over $[0, 4\pi]$ (which is the unit circle), the length of this curve is different with each parameterization. We want to define a special type of curve, which has constant length that does not depend on the parameterization. Curves like these are called simple smooth curves.

Definition 14.2 (Simple smooth curve). A simple smooth curve is a set $C \subset \mathbb{R}^n$ that is the image of a function that we call a path $f \in C^1([a, b], \mathbb{R}^n)$ that satisfies:

- (1) The path f is an injection on (a, b) .
- (2) For every $t \in [a, b]$ we have $f'(t) \neq 0$.

Remark 14.2. Condition (2) is sometimes called regularity, and curves that satisfy it are called regular curves. It implies that the sign of f' does not change. This is important since it helps us differentiate between paths that define the same curve but have different orientations.

Exercise 14.1. Show that if f is an injective path then the curve it defines is a C^1 manifold.

Definition 14.3 (Directed curve). A curve is called directed if we differentiate between its beginning and end.

Definition 14.4 (Closed curve). A curve is called closed or a loop if it satisfies $f(a) = f(b)$.

Definition 14.5 (Length of a simple smooth curve). We define the length of a simple smooth curve to be the length of a path that defines it.

Proposition 14.3. *Definition 14.5 is well defined. This means that if two distinct paths define the same smooth simple curve, then their lengths are equal.*

14.2 Arc length parameterization

Let $f: [a, b] \rightarrow \mathbb{R}^n$ be a parameterization of a simple smooth curve C . Consider the function from earlier $L: [a, b] \rightarrow [0, l(C)]$ defined by

$$C^1 \ni L(t) = l(f|_{[a, t]}) \quad \text{with} \quad \forall t: L'(t) = \|f'(t)\| \neq 0.$$

By the inverse function theorem, and since γ is injective, there exists a continuously differentiable function $\sigma: [0, l(C)] \rightarrow [a, b]$ that is an inverse of L such that

$$\sigma'(s) = \frac{1}{L'(L^{-1}(s))} = \frac{1}{\|f'(\sigma(s))\|}. \quad (1)$$

This means we can define a new parameterization $g: [0, l(C)] \rightarrow C \subseteq \mathbb{R}^n$ with

$$g(s) = f(\sigma(s))$$

such that $s = L(t)$ and $t = \sigma(s)$. This means that the function $g(s)$ is the point on the curve such that its distance from the origin $f(a) = g(0)$ (measured on the curve) is exactly s . That is,

$$l(g|_{[0,s]}: [0, s] \rightarrow \mathbb{R}^n) = s.$$

In other words, the parameter of the parameterization is the length of the curve!

We can see this by

$$l(C) = \int_a^b \|f'(t)\| dt = \int_0^{l(C)} \|f'(\sigma(s))\| \sigma'(s) ds = \int_0^{l(C)} \|g'(s)\| ds = \int_0^{l(C)} ds,$$

where the last equality follows from Equation (1)

$$\|g'(s)\| = \|f'(\sigma(s))\| |\sigma'(s)| = 1.$$

Remark 14.3. This also shows that the length of a simple smooth curve is well defined.

Definition 14.6 (Arc length parameterization). A parametrization of a curve g such that $\|g'(s)\| = 1$ is called an arc length parametrization.

Remark 14.4. Notice that we have found a constructive proof for an arc length parameterization of a smooth simple curve. Let $\gamma: [a, b] \rightarrow \mathbb{R}^n$ be the path that parameterizes the curve. Then define the arc length from a to t by the function

$$s(t) = \int_a^t \|\gamma'(u)\| du.$$

Since $\|\gamma'(u)\| > 0$, this gives a strictly increasing function. Since any strictly increasing real-valued function like $h: [a, b] \rightarrow \mathbb{R}$ is invertible we get that $s(t)$ is invertible. Find $t(s) = s^{-1}(s)$ (describe t using the variable s , we will see what this means in the following example). Then the function $\tilde{\gamma}(s) = \gamma(s^{-1}(s))$ is an arc length parameterization. The following example shows this algorithm in action.

Example 14.1. Consider the parameterization of the circle $\gamma: [0, \pi] \rightarrow \mathbb{R}^2$ defined by $\gamma(t) = (\cos(2t), \sin(2t))$. First, we compute $\|\gamma'(t)\|$. The derivative of γ is

$$\gamma'(t) = (-2\sin(2t), 2\cos(2t)),$$

and thus

$$\|\gamma'(t)\| = \sqrt{(-2\sin(2t))^2 + (2\cos(2t))^2} = \sqrt{4(\sin^2(2t) + \cos^2(2t))} = 2.$$

This means that the “speed” of the path is constant, but not 1, so γ is not an arc length parameterization. We need to find $s(t)$.

$$s(t) := \int_0^t \|\gamma'(u)\| du = \int_0^t 2 du = 2t.$$

Now we need to describe t using s . We immediately see that $t(s) = s/2$. Now, according to the algorithm, $\tilde{\gamma}(s) = \gamma(s/2) = (\cos(s), \sin(s))$ is an arc length parameterization.

Exercise 14.2. Find an arc length parameterization for the catenary curve given by

$$y = a \cosh\left(\frac{x}{a}\right)$$

for $a > 0$ and $x \in \mathbb{R}$, in the domain $x \in [0, a \ln(2)]$.

15 Line Integrals

15.1 Type one line integrals

Example 15.1 (Metal spring). Given a parameterization $f: [0, L] \rightarrow \mathbb{R}^3$ of a metal spring given by

$$f(t) = (a \cos(\omega t), a \sin(\omega t), bt) \implies f'(t) = (-\omega a \sin(\omega t), \omega a \cos(\omega t), b),$$

we could calculate the length of the spring by

$$l(C) = l(f) = \int_0^L \|f'(t)\| dt = \int_0^L \sqrt{\omega^2 a^2 \sin^2(\omega t) + \omega^2 a^2 \cos^2(\omega t) + b^2} dt = L \sqrt{\omega^2 a^2 + b^2}.$$

We can notice that $\|f'(t)\| = \sqrt{\omega^2 a^2 + b^2} \neq 1$ so $f(t)$ is not an arc length parameterization as defined in Definition 14.6.

We can define a new parameterization $g: [0, L(\omega^2 a^2 + b^2)^{1/2}] \rightarrow \mathbb{R}^3$ by

$$g(s) = f(\sigma(s)) = f\left(\frac{s}{\sqrt{\omega^2 a^2 + b^2}}\right).$$

We have that

$$\|g'(s)\| = \|f'(\sigma(s))\sigma'(s)\| = \sqrt{\omega^2 a^2 + b^2} \frac{1}{\sqrt{\omega^2 a^2 + b^2}} = 1$$

which means that g is the arc length parameterization for the spring given by f .

Example 15.2. Let f be the function

$$f(t) = \left(\cos t, \sin t, \frac{1}{2}t^2\right) \implies \|f'(t)\| = \|\sin t, -\cos t, t\| = \sqrt{1+t^2}.$$

This means that

$$l(f) = \int_0^L \sqrt{1+t^2} dt = \dots = \frac{1}{2} \left[L\sqrt{1+L^2} + \ln(1+L^2) \right].$$

If we want to find an arc length parameterization we would need to find the inverse function to the the length function of the path, which unlike the previous example is not that easy. However, according to what we have shown in the previous section we know that there exists one.

Definition 15.1 (Type one line intergral). Let $C \subset \mathbb{R}^n$ be a smooth simple curve with parametrization $C^1 \ni f: [a, b] \rightarrow \mathbb{R}^n$. If ρ is a continuous function on C then we define the scalar integral of ρ on C (or type one line integral) as:

$$\int_C \rho ds := \int_a^b \rho(f(t)) \|f'(t)\| dt.$$

Remark 15.1. The term $ds = \|f'(t)\| dt$ is called the *curve length element*.

Remark 15.2. Notice that when $\rho \equiv 1$ we get the length of the curve. We can use this type of integral when we want to sum values over that curve. For example, suppose we have a parameterization f for a wire C , and a function $\rho: C \rightarrow \mathbb{R}$ that gives us the concentration of every point on the wire $x \in C$. Then the scalar integral on ρ will give us the mass of the wire.

Remark 15.3. Recall that the definition for the length of a simple smooth curve had an intuition behind it. We derived it by taking the limit on lengths of more and more refined partitions. We can derive Definition 15.1 in the same way by taking a similar limit on

$$\sum_{i=1}^n \rho(f(t_i)) \cdot \text{length}(f([t_{i-1}, t_i])) = \sum_{i=1}^n \rho(f(t_i)) \cdot \int_{t_{i-1}}^{t_i} \|f'(t)\| dt.$$

Remark 15.4. The scalar integral is not dependent on the parameterization.

Example 15.3. Suppose that the spring from Example 15.1 has a mass concentration function

$$\rho(x, y, z) = z$$

then we can calculate the mass of the spring using a scalar integral.

$$\int_C \rho ds = \int_0^L \rho(f(t)) \|f'(t)\| dt = \int_0^{L\sqrt{\omega^2 a^2 + b^2}} \rho(g(s)) ds = \int_0^{l(C)} \rho(g(s)) ds,$$

where g is the arc length parameterization. Notice how the integral became simpler once we used the arc length parameterization. Since we are using the arc length parameterization we get

$$g(s) = \left(a \cos\left(\frac{\omega s}{\sqrt{\omega^2 a^2 + b^2}}\right), a \sin\frac{\omega s}{\sqrt{\omega^2 a^2 + b^2}}, \frac{bs}{\sqrt{\omega^2 a^2 + b^2}} \right),$$

which implies that $\int_0^{L\sqrt{\omega^2 a^2 + b^2}} \rho(g(t)) ds$ is simply

$$\int_0^{l(C)} \rho(g(t)) ds = \int_0^{l(C)} \frac{bs}{\sqrt{\omega^2 a^2 + b^2}} ds = \frac{b}{\sqrt{\omega^2 a^2 + b^2}} \frac{1}{2} L(\omega^2 a^2 + b^2) = \frac{b}{2} L^2 \sqrt{\omega^2 a^2 + b^2},$$

which is indeed the mass of the spring.

Exercise 15.1. Compute the scalar integral

$$\int_C \sqrt{5 + 5x^2 - 6xy} ds$$

where C is the intersection curve of the cylinder $x^2 + y^2 = 1$ and the plane $z = 2x + 3y$.

Solution. We notice that this is indeed a scalar integral where $\rho(x, y, z) = \sqrt{5 + 5x^2 - 6xy}$. In order to compute the integral we first need to find a parameterization for C (make sure you understand why). Since we are working with a cylinder it is convenient to work with cylindrical coordinates:

$$(x, y, z) = (r \cos(\theta), r \sin(\theta), z),$$

where $z \in \mathbb{R}$ and $\theta \in [0, 2\theta]$ and $r \geq 0$. The first constraint is $x^2 + y^2 = 1$ so

$$x^2 + y^2 = 1 \implies r^2 = 1 \implies r = 1.$$

This means that $x^2 + y^2 = 1$ a constraint equivalent which is equivalent to $r = 1$ in the new coordinates. We also have the constraint $z = 2x + 3y$, which similarly gives:

$$z = 2x + 3y \implies z = 2 \cos(\theta) + 3 \sin(\theta).$$

This means that $z = 2x + 3y$ a constraint equivalent which is equivalent to $z = 2 \cos(\theta) + 3 \sin(\theta)$ in the new coordinates. Therefore our parameterization for our curve is the function $\gamma: [0, 2\pi] \rightarrow \mathbb{R}^3$ defined by

$$\gamma(\theta) := (\cos(\theta), \sin(\theta), 2 \cos(\theta) + 3 \sin(\theta)).$$

Clearly our curve is smooth and simple. Simple computations give

$$\begin{aligned} \gamma'(\theta) &= (-\sin(\theta), \cos(\theta), -2 \sin(\theta) + 3 \cos(\theta)) \\ \|\gamma'(\theta)\| &= \sqrt{\sin^2(\theta) + \cos^2(\theta) + (-2 \sin(\theta) + 3 \cos(\theta))^2} = \sqrt{5 + 5 \cos^2(\theta) - 6 \sin(\theta) \cos(\theta)} \\ f(\gamma(\theta)) &= \sqrt{5 + 5 \cos^2(\theta) - 6 \sin(\theta) \cos(\theta)} \end{aligned}$$

so by definition

$$\int_C \sqrt{5 + 5x^2 - 6xy} ds = \int_0^{2\pi} \sqrt{5 + 5 \cos^2(\theta) - 6 \sin(\theta) \cos(\theta)} d\theta = 15\pi.$$

15.2 Type two line integrals

Definition 15.2 (Vector field). Let $S \subset \mathbb{R}^n$. A function $\mathbf{F}: S \rightarrow \bigcup_{p \in S} T_p(\mathbb{R}^n)$ is called a vector field if for all $p \in S$ we have $\mathbf{F}(p) \in T_p(\mathbb{R}^n)$.

Remark 15.5. A more general definition exists for smooth manifolds but there's no reason to get into it right now. For those who are curious, if $M \subseteq \mathbb{R}^n$ is a smooth manifold, we define a vector field to be of the form $\mathbf{F}: M \rightarrow \bigcup_{p \in M} T_p(M)$ so that for all $p \in M$ we have $\mathbf{F}(p) \in T_p(M)$.

Definition 15.3 (Type two line integral). Let $f: [a, b] \rightarrow \mathbb{R}^n$ be a continuously differentiable parametrization of a directed smooth simple curve C . Let $\mathbf{F}: C \rightarrow \mathbb{R}^n$ be a continuous vector field. The line integral of $\mathbf{F}$ over C (or a type two line integral) is defined to be:

$$\int_C \mathbf{F} \cdot d\vec{r} := \int_C \sum_{i=1}^n F_i dx_i := \int_a^b \mathbf{F}(f(t)) \cdot f'(t) dt.$$

Remark 15.6. If we think about vector field as some physical force like gravity, and consider the force it applies to a point object moving on the curve C at time t and position $f(t)$, it is exactly the projection of $F(f(t))$ on the tangent line at $f(t)$ which is $f'(t)$, and this is why the integral is equal to their dot product.

Remark 15.7. The second part of the equality is summing the scalar integrals on each of the variables separately. We can derive it from the third part by abuse of notation. Let $(x_1(t), \dots, x_n(t))$ be the parameterization of our curve. Then if $f(t) = (x_1(t), x_2(t), \dots, x_n(t))$ we get

$$\int_a^b \mathbf{F}(f(t)) \cdot f'(t) dt = \int_a^b \sum_{i=1}^n F_i(x(t)) x'_i(t) dt = \int_a^b \sum_{i=1}^n F_i(x(t)) \frac{dx_i(t)}{dt} dt = \int_C \sum_{i=1}^n F_i dx_i.$$

We will give an explanation about these dx_i later. For now you may ignore this form of the type two line integral entirely, and may only rely on the third form, since it's the only form we know how to compute.

Remark 15.8. If $g: [a, b] \rightarrow \mathbb{C}$ is a C^1 arc length parameterization of C , then we can define $\mathbf{T}(s) = g'(s)$ and since $\|\mathbf{T}(s)\| = 1$ we get

$$\int_C \sum_{i=1}^n F_i dx_i = \int_a^b \mathbf{F}(g(s)) \cdot g'(s) ds = \int_a^b \mathbf{F}(g(s)) \cdot \mathbf{T}(s) ds.$$

Since the length of $\mathbf{T}$ is the unit length, we sometimes write $\hat{\mathbf{T}}$ instead of $\mathbf{T}$ or $\vec{T}$, which denotes a unit vector in the direction of $\mathbf{T}$. The benefit of writing the integral this way is that we do not need a parameterization of the curve to describe it. Notice that if we consider $\mathbf{F}(g(s)) \cdot \mathbf{T}(s)$ as a scalar function on C , this gives us a way to compute a type two line integral as a type one line integral.

Remark 15.9. In physics, type two line integral is commonly used to calculate the work done by a vector field along a given curve. Thus, it should also be intuitive that unlike a scalar integral, the sign of a type two line integral changes when we change the orientation of the parameterization.

Example 15.4 (Gravitational field). Let the function

$$\mathbf{F}: \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}^3$$

represent the gravitational force field experienced by a point object of mass m located at a position represented by $\mathbf{r}$, due to another point object of mass M located at the origin. We have that

$$\mathbf{F} = G \frac{Mm}{\|\mathbf{r}\|^3} (-\mathbf{r}).$$

Under the assumption that m, M are constant, we can write

$$\mathbf{F} = -K \frac{(x, y, z)}{\|(x, y, z)\|^3}$$

for some constant K . For the sake of simplicity assume $K = 1$. We can calculate the work done by this vector field on m travelling along the curve of the spring from Example 15.1. Recall that

$$f(t) = (a \cos(\omega t), a \sin(\omega t), bt) \quad \text{and} \quad f'(t) = (-\omega a \sin(\omega t), \omega a \cos(\omega t), b).$$

We can see that

$$\|f(t)\| = \|(a \cos(\omega t), a \sin(\omega t), bt)\| = \sqrt{a^2 + b^2 t^2} \quad \text{and} \quad f(t) \cdot f'(t) = b^2 t.$$

Using these calculations we can finally see that

$$\begin{aligned} \int_0^L \mathbf{F}(f(t)) \cdot f'(t) dt &= - \int_0^L \frac{f(t) \cdot f'(t)}{(a^2 + b^2 t^2)^{3/2}} dt = - \int_0^L \frac{b^2 t}{(a^2 + b^2 t^2)^{3/2}} dt \\ &= \frac{1}{\sqrt{a^2 + b^2 t^2}} \Big|_0^L = \frac{1}{\sqrt{a^2 + b^2 L^2}} - \frac{1}{a}. \end{aligned}$$

16 Integration

16.1 Integration over a box

Definition 16.1 (Box). A box in $\mathbb{R}^n$ is a set $R \subset \mathbb{R}^n$ such that:

$$R = \prod_{i=1}^k [a_i, b_i]$$

for real numbers $a_i \leq b_i$.

Remark 16.1. If there exists i such that $a_i = b_i$ then we say that the box is *degenerate*.

Definition 16.2 (Volume of a box). We define the volume of a box to be:

$$\text{Vol}(A) = \prod_{i=1}^k (b_i - a_i)$$

Remark 16.2. Let P be a partition of a box $R \subset \mathbb{R}^n$. We denote $B \sim Q$ if and only if B is a subbox of the partition.

Definition 16.3 (Integration over a box). Let f be a bounded function on a box R . We define the upper and lower Darboux sums with regards to a partition P of R with

$$L(f, P) := \sum_{B \sim R} m_Q \cdot \text{Vol}(Q) \text{ and } U(f, P) := \sum_{B \sim R} M_Q \cdot \text{Vol}(Q)$$

where $M_Q = \sup \{f(x) \mid x \in Q\}$ and $m_Q = \inf \{f(x) \mid x \in Q\}$. We say that f is (Darboux) integrable if:

$$\inf U(f, P) = \sup L(f, P) = I.$$

We denote the integral of f over the box as:

$$I = \int_R f dV = \int_R f(x) dx = \int_R f.$$

Remark 16.3. We can see that for a constant function $f \equiv C$ that

$$\int_R f dV = C \cdot \text{Vol}(R).$$

Definition 16.4 (Refinement). Let P, P' be two partitions of a box R . We say that P' is a refinement of P if for each box $Q' \sim P'$, there exists $Q \sim P$ such that $Q' \subset Q$.

Remark 16.4. Most theorems from analysis 2 about integrals and partitions also apply here, with similar arguments, and are thus omitted.

Proposition 16.1. Let $f: R \rightarrow \mathbb{R}$ be bounded. Then f is integrable on R if and only if for all $\epsilon > 0$ there exists a partition P such that $U(f, P) - L(f, P) < \epsilon$.

Exercise 16.1. Describe the mean value theorem, and the triangle equality for the multidimensional integral.

Remark 16.5. Here are a few uses of the integral that are good to remember.

- (1) Given a set $\Omega \subset \mathbb{R}^n$, the integral $\iiint_{\Omega} 1 dV$ is the n -dimensional volume of Ω .

- (2) Given a set $\Omega \subset \mathbb{R}$ and a mass concentration function $\rho(x, y, z)$, the mass of Ω is described by the following integral

$$m = \iiint_{\Omega} \rho(x, y, z) dV.$$

- (3) We can also use the integral to calculate the moment of inertia of an object, or its center of mass, etc.

Remark 16.6. Recall that in analysis 2 we never actually computed integrals by definition. In the one dimensional case we always used the fundamental theorem of calculus and in the two dimensional case we always used Fubini's theorem. Indeed in the n -dimensional case we are not going to compute integrals by definition.

16.2 Fubini's theorem

Theorem 16.2 (Fubini's theorem). *In 2 dimensions Fubini's theorem states that if f is integrable on $[a, b] \times [c, d] \subset \mathbb{R}^2$ and for all $x \in [a, b]$ the function*

$$F(x) = \int_c^d f(x, y) dy$$

is defined and integrable on $[a, b]$, then

$$\int_R f(x, y) dx dy = \int_a^b F(x) dx = \int_a^b \int_c^d f(x, y) dy dx.$$

Generalizing to n dimensions gives something similar. Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ and suppose all the functions:

$$\begin{aligned} F_1(x_1, \dots, x_{n-1}) &= \int_{a_n}^{b_n} f(x) dx_n \\ F_2(x_1, \dots, x_{n-2}) &= \int_{a_{n-1}}^{b_{n-1}} f(x) dx_{n-1} \int_{a_n}^{b_n} f(x) dx_n \\ &\vdots \\ F_n \equiv I &= \int_{a_1}^{b_1} f(x) dx_1 \cdots \int_{a_{n-1}}^{b_{n-1}} f(x) dx_{n-1} \int_{a_n}^{b_n} f(x) dx_n \end{aligned}$$

are integrable. Then $I = \int_R f dV$.

Remark 16.7. To evaluate the volume of a subspace $\Omega \subset \mathbb{R}^n$ we calculate the integral $\int_{\Omega} f dV$ for the function $f = \mathbb{1}_{\Omega}$.

Definition 16.5 (Simple domain). A simple domain in $\mathbb{R}^2$ is a region (open connected set) $D \subset \mathbb{R}^2$ that can be described as the set of points bounded by two curves and two vertical or horizontal lines. The general form of a vertical simple domain in $\mathbb{R}^2$ bounded by $x = a$, $x = b$ and the curves $g_1(x)$, $g_2(x)$ is

$$D = \left\{ (x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \text{ and } g_1(x) \leq y \leq g_2(x) \right\}.$$

A simple domain in $\mathbb{R}^n$ is defined similarly with a general form

$$D = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid (x_1, \dots, x_{n-1}) \in R \text{ and } g_1(x_1, \dots, x_{n-1}) \leq x_n \leq g_2(x_1, \dots, x_{n-1}) \right\}$$

where R is a simple domain in $\mathbb{R}^{n-1}$.

Remark 16.8. Suppose Ω is a simple domain in $\mathbb{R}^3$. We can use Fubini's theorem and obtain that:

$$\text{Vol}(\Omega) = I = \iiint_{\Omega} f(x, y, z) dx dy dz = \int_a^b dx \int_{y_1(x)}^{y_2(x)} dy \int_{z_1(x, y)}^{z_2(x, y)} f(x, y, z) dz$$

for the relevant boundary functions.

Remark 16.9. Before the next example, recall that the general equation of a plane in $\mathbb{R}^3$ is

$$ax + by + cz = d$$

where $\mathbf{n} = (a, b, c)$ is a vector orthogonal to the plane, and d is the constant determining the plane's offset from the origin. You may also recall the discussion on constrained optimization and think why $\mathbf{n}$ is orthogonal to the plane.

Example 16.1. Let us calculate the volume of $I = \iiint_{\Omega} (xyz) dx dy dz$ where Ω is the space bounded by $x + y + z = 1$ and the x, y, z axes. First we notice that Ω looks like a pyramid. For each z' we can consider the triangle $\Delta z'$ on the $z = z'$ plane with points of the form

$$0 \leq x + y \leq 1 - z'.$$

Since

$$\Delta z' = \left\{ (x, y, z') \in \mathbb{R}^2 \times \{z'\} \mid 0 \leq x \leq 1 - z' \text{ and } 0 \leq y \leq 1 - x - z' \right\},$$

it is a simple domain in $\mathbb{R}^2 \times \{z'\}$, and thus Ω is a simple domain in $\mathbb{R}^3$ given by

$$\Omega = \left\{ (x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \Delta z \text{ and } 0 \leq z \leq 1 \right\}.$$

Thus we have

$$I = \int_0^1 dz \int_0^{1-z} dx \int_0^{1-x-z} (xyz) dy = \frac{1}{720}.$$

Remark 16.10. The following is a more general formulation of Fubini's theorem.

Theorem 16.3 (Fubini's theorem). *Let $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ be closed boxes, and let $f: A \times B \rightarrow \mathbb{R}$ be a function integrable on $A \times B$. For all $x \in A$ we define $g_x: B \rightarrow \mathbb{R}$ with*

$$g_x(y) := f(x, y).$$

If g_x is integrable on B for all x , then the function

$$G(x) := \int_B g_x(y) dy$$

is integrable (on A) and

$$\int_A \left(\int_B g_x(y) dy \right) dx = \int_{A \times B} f(x, y) dx dy.$$

Example 16.2. Consider the n -dimensional unit disc with the ℓ^1 norm:

$$D_n := \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}.$$

We can divide this space into 2^n of equal size, that are isometric to

$$E_n := D_n \cap \{x \in \mathbb{R}^n : x_1, \dots, x_n \geq 0\}.$$

This means that we can find the volume of E_n and multiply it by 2^n to get the result we want.

¹ First, we can notice that

$$\begin{aligned} E_n &= \{x \in \mathbb{R}^n : x_1 \in [0, 1], x_2 \in [0, 1 - x_1], \dots, x_n \in [0, 1 - x_1 - x_2 - \dots - x_{n-1}]\} \\ &= \{(x', x_n) : x_n \in [0, 1] \text{ and } \|x'\|_1 \leq 1 - x_n\} \subset [0, 1]^{n-1} \times [0, 1]. \end{aligned}$$

We now have that

$$\begin{aligned} \text{Vol}(E_n) &= \int_{[0,1]^n} \mathbb{1}_{E_n} dV = \int_0^1 \left(\int_{(1-x_n)E_{n-1}} dx' \right) dx_n = \int_0^1 \text{Vol}((1-x_n)E_{n-1}) dx_n \\ &= \int_0^1 (1-x_n)^{n-1} \text{Vol}(E_{n-1}) dx_n = \frac{\text{Vol}(E_{n-1})}{n}. \end{aligned}$$

Where the second to last equality follows from basic geometrical calculations. By using induction, and given that $\text{Vol}(E_1) = 1$ we get that $\text{Vol}(D_n) = 2^n/n!$.

Remark 16.11 (Different notation, and Fubini with more rigor). Let $Q' \subset \mathbb{R}^n$ and $Q'' \subset \mathbb{R}^m$ be closed boxes and let P' and P'' be their respective partitions, so that $P = P' \times P''$ is a partition of $Q = Q' \times Q'' \subset \mathbb{R}^{n+m}$. Let $f: Q' \times Q'' \rightarrow \mathbb{R}$ be bounded. For $x_0 \in Q'$ we denote $g_{x_0} := f(x_0, *): Q'' \rightarrow \mathbb{R}$, and define $\underline{F}(x) = \sup_{P''} L(g_x, P'')$ and $\overline{F}(x) = \inf_{P''} U(g_x, P'')$ to be the respective lower and upper Darboux sums. If for some $x \in Q'$ the function g_x is integrable on Q'' then we have $\underline{F}(x) = \overline{F}(x) = \int_{Q''} f(x, y) |d^m y|$ and in this case we define $F(x) = \int_{Q''} f(x, y) |d^m y|$. When g_x is not integrable we define $F(x)$ to be some number in $[\underline{F}(x), \overline{F}(x)]$. The set x such that g_x is not integrable is of measure 0 so we can let $F(x)$ to be any value without changing integrals over $F(x)$.

Let $f: Q \rightarrow \mathbb{R}$ be integrable on $Q = Q' \times Q''$. Then for all $x_0 \in Q'$ the expression $\int_{Q''} f(x_0, y) |d^m y|$ exists, except for a set of measure 0 (which could be the empty set), the function $F(x)$ is integrable on Q' , and we have

$$\int_{Q'} F(x) |d^n x| = \int_Q f(x, y) dV.$$

Corollary 16.4. Let $f: Q \rightarrow \mathbb{R}$ be integrable on $Q = Q' \times Q''$, and suppose that the integral $\int_{Q''} f(x_0, y) |d^m y|$ exists for all $x_0 \in Q'$. Then the function $F(x) = \int_{Q''} f(x, y) |d^m y|$ is integrable and we have

$$\int_{Q'} \left(\int_{Q''} f(x, y) |d^m y| \right) |d^n x| = \int_{Q' \times Q''} f(x, y) |d^m y| |d^n x|.$$

Example 16.3. Let $f(x, y) := 1 + x^2 + 2xy$ and $\Omega := [0, 1] \times [1, 2]$. Since $f(x, y)$ is continuous it is integrable over any Jordan measurable set in $\mathbb{R}^2$. Thus by Fubini's theorem we get that

$$\begin{aligned} \int_{\Omega} f(x, y) dV &= \int_0^1 |dx| \left(\int_1^2 (1 + x^2 + 2xy) |dy| \right) = \int_0^1 \left(((1 + x^2)y + xy^2)|_{y=1}^{y=2} \right) |dx| \\ &= \int_0^1 (1 + x^2 + 3x) |dx| = \dots = \frac{17}{6}. \end{aligned}$$

Example 16.4 (Standard simplex area). Suppose we want to compute the volume of the set

$$\Omega := \left\{ x \in \mathbb{R}^n \mid x^1 \leq x^2 \leq \dots \leq x^n \leq 1 \right\}.$$

¹To fully formalize this we need to use the change of variables theorem, which we will see later.

Then by the definition of volume and Fubini's theorem we get that:

$$\begin{aligned}\text{Vol}(\Omega) &= \int_{[0,1]^n} \mathbb{1}_\Omega dV = \int_0^1 \int_0^{x^n} \cdots \int_0^{x^2} 1 |dx^1| \cdots |dx^n| = \int_0^1 \int_0^{x^n} \cdots \int_0^{x^3} x^2 |dx^2| \cdots |dx^n| \\ &= \frac{1}{2} \int_0^1 \int_0^{x^n} \cdots \int_0^{x^4} (x^3)^2 |dx^3| \cdots |dx^n| = \cdots = \frac{1}{(n-1)!} \int_0^1 (x^n)^{n-1} |dx^n| = \frac{1}{n!},\end{aligned}$$

where one must not confuse the coordinates x^i with taking the exponent of x .

17 Zero Volume

17.1 Zero volume

Proposition 17.1. *If $f: R \rightarrow \mathbb{R}^n$ is a continuous function on R then it is integrable.*

Proof. To be added. □

Definition 17.1 (Zero volume). A set $A \subset \mathbb{R}^n$ is said to have zero volume if for each $\varepsilon > 0$ there exist a finite number of boxes $R_1, \dots, R_k$ such that:

$$\sum_{i=1}^k \text{Vol}(R_i) < \varepsilon \quad \text{and} \quad A \subset \bigcup_{i=1}^k \text{Int}(R_i),$$

where $\text{Int}(R_i)$ denotes the topological interior of R_i .

Remark 17.1. Any point $x \in \mathbb{R}^n$ is of volume zero because we can always choose one box of arbitrarily small volume such that x is inside it. It is also easy to prove that a finite union of sets of volume zero is of volumes zero.

Theorem 17.2. *Let $f: R \rightarrow \mathbb{R}^n$ be bounded and suppose that*

$$X = \{x \in R \mid f \text{ is not continuous in } x\}$$

is a set of volume 0. Then f is integrable over R .

Proof. We know that f is bounded so we can denote its bound M , and also we know that for any $\varepsilon > 0$ there exist $R_1, \dots, R_k$ such that:

$$\sum_{i=1}^k \text{Vol}(R_i) < \frac{\varepsilon}{4M} \quad \text{and} \quad X \subset \bigcup_{i=1}^k \text{Int}(R_i).$$

This means we can choose a partition P_X such that for every i we have $R_i \in P_X$ and then we have:

$$U(f, P_X) - L(f, P_X) = \sum_{i=1}^k (M_Q - m_Q) \text{Vol}(R_i) < (M - (-M)) \cdot \frac{\varepsilon}{4M} = \frac{\varepsilon}{2} \quad \text{and} \quad X \subset \bigcup_{i=1}^k \text{Int}(R_i).$$

The set $Y = R \setminus \{R_i\}_i$ is clearly closed and bounded and thus it is integrable. We can choose a partition P_Y of Y such that:

$$U(f, P_Y) - L(f, P_Y) = \sum_{Q \sim P_Y} (M_Q - m_Q) \text{Vol}(Q) < \frac{\varepsilon}{2}$$

and now it is clear that the partition $P = P_X \cup P_Y$ satisfies:

$$U(f, P) - L(f, P) < \varepsilon.$$

□

17.2 Measure zero and integration

Definition 17.2 (Measure zero). We say that $A \subset \mathbb{R}^n$ is of measure zero if there exist $\{R_i\}_{i \in I}$ such that $|I| \leq \aleph_0$ and:

$$\sum_{i \in I} \text{Vol}(R_i) < \varepsilon \quad \text{and} \quad A \subset \bigcup_{i \in I} \text{Int}(R_i).$$

Remark 17.2. A set of volume zero is always of measure zero, but the contrary is not always true. For example the sets of the form $\mathbb{Q}^n \subset \mathbb{R}^n$ are all of measure zero but not of volume zero.

Remark 17.3. A countable union of sets of measure zero is of measure zero.

Theorem 17.3. A compact set $D \subset \mathbb{R}^n$ is of volume zero if and only if it is of measure zero.

Proof. By definition a set of volume zero is always of measure zero.

Suppose that D is of measure zero. Let $\varepsilon > 0$, then there exist a countable collection $\{R_i\}_{i \in I}$ of boxes such that

$$\sum_{i \in I} \text{Vol}(R_i) < \varepsilon \quad \text{and} \quad D \subset \bigcup_{i \in I} \text{Int}(R_i).$$

The set $\{\text{Int}(R_i)\}_{i \in I}$ is an open cover of D , so since D is compact, it has a finite subcover. Thus, there exist $R_1, \dots, R_k$ such that

$$\sum_{i=1}^k \text{Vol}(R_i) < \varepsilon \quad \text{and} \quad D \subset \bigcup_{i=1}^k \text{Int}(R_i)$$

which completes the proof. $\square$

Remark 17.4. A subset of a set of measure zero is of measure zero.

Theorem 17.4 (Lebesgue's theorem). Let $R \subset \mathbb{R}^n$ be a box and let $f: R \rightarrow \mathbb{R}^n$ be bounded. Then f is integrable if and only if the set of discontinuity points of f is of measure zero.

Before we prove this theorem we first need the following two definitions.

Definition 17.3 (Oscillation over a set). The oscillation of a function $f: R \rightarrow \mathbb{R}^n$ over a set $B \subset \mathbb{R}^n$ is defined by

$$\omega(f, B) := \sup \{f(x) \mid x \in R \cap B\} - \inf \{f(x) \mid x \in R \cap B\}.$$

Definition 17.4 (Oscillation at a point). The oscillation of a function $f: R \rightarrow \mathbb{R}^n$ at a point $x \in R$ is defined by

$$\omega(f, x) := \lim_{r \rightarrow 0^+} \omega(f, U_r(x))$$

where $U_r(x) = (x_1 - r, x_1 + r) \times \dots \times (x_n - r, x_n + r)$, is the open ball with radius r and center x with the infinity norm.

Lemma 17.5. A function $f: R \rightarrow \mathbb{R}^n$ is continuous at $x \in R$ if and only if $\omega(f, x) = 0$.

Lemma 17.6. For every $\varepsilon > 0$ the set:

$$W_\varepsilon := \{x \in R \mid \omega(f, x) \geq \varepsilon\}$$

is compact.

Remark 17.5. We also define another set:

$$W := \{x \in R \mid \omega(f, x) \neq 0\}.$$

Proof of Theorem 17.4. To be added $\square$

17.3 Integration over Jordan measurable sets

Definition 17.5 (Jordan measurable set). A set $\Omega \subset \mathbb{R}^n$ is said to be Jordan measurable if it is bounded and $\partial\Omega$ is of measure zero.

Remark 17.6. Since $\partial\Omega$ is compact from Heine–Borel, we could equivalently require it to be of zero volume instead of measure zero. Moreover, a set $\Omega \subset \mathbb{R}^n$ is Jordan measurable if and only if there exists a rectangle $\Omega \subset R \subset \mathbb{R}^n$ such that $\mathbb{1}_\Omega$ is Riemann integrable.

Definition 17.6 (Set with volume). Let $\Omega \subset \mathbb{R}^n$ be a Jordan measurable set. We say that Ω has volume if it is a closure of an open set. We require this because if Ω is closed, we don't have to use improper integrals, and if it is closed but not the closure of an open set, we have cases like $\Omega = \{x\}$ which don't have volume.

Definition 17.7 (Integrability over Jordan measurable sets). Let $\Omega \subset \mathbb{R}^n$ be a set with volume, let $f: R \rightarrow \mathbb{R}^n$ be a function, and assume that $\Omega \subset R$. We say that f is integrable over Ω if $f \cdot \mathbb{1}_\Omega$ is integrable over R and

$$\int_{\Omega} f dV = \int_R f \cdot \mathbb{1}_\Omega dV.$$

Proposition 17.7. *If $f: \Omega \rightarrow \mathbb{R}^n$ is continuous and Ω has volume then f is integrable over Ω .*

Definition 17.8 (Volume of a set). The volume of a Jordan measurable set $A \subset \mathbb{R}^n$ is defined as:

$$\text{Vol}(A) := \int_A 1 dV.$$

Note that Jordan measurable sets that are not sets with volume have volume zero.

Remark 17.7. We notice that this function is well defined because the constant function $f \equiv 1$ is always continuous on any domain that has volume.

Remark 17.8. A Riemann integral on a space with volume has all the basic properties as we defined them on intervals: linearity, additivity, the triangle inequality, and more. Linearity and additivity in particular imply that the set of integrals on some set Q form a vector space.

Theorem 17.8. *The graph of a Riemann integrable function is of measure zero, or volume zero if the function was defined on a compact space. A manifold has measure zero. The image of a manifold by a smooth transformation is of volume zero.*

18 Improper integrals

18.1 Improper integrals

Definition 18.1 (Improper integral). Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f: U \rightarrow \mathbb{R}^+$ be a Riemann integrable function on any compact subset $K \subset U$ with volume. We define

$$\int_U f := \sup \left\{ \int_K f dV \mid K \subset U \text{ and } K \text{ has volume and compact} \right\}.$$

Remark 18.1. In the case $\int_U f < \infty$, we say that f is integrable on U . Otherwise, we say it is not integrable and we write $\int_U f = \infty$.

Proposition 18.1 (Exhaustion). Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f: U \rightarrow \mathbb{R}^+$ be a Riemann integrable function on any compact subset $K \subset U$ with volume. Let $\{K_m\}_{m=1}^\infty$ be a sequence of compact subsets of U with volume such that

- (1) $U = \bigcup_{m=1}^\infty K_m$;
- (2) for all $m \geq 1$ we have $K_m \subset \text{Int}(K_{m+1})$.

Then f is integrable on U if and only if $\lim_{m \rightarrow \infty} \int_{K_m} f < \infty$, and in that case we have $\int_U f = \lim_{m \rightarrow \infty} \int_{K_m} f$.

Proof. It is clear that the sequence $\int_{K_m} f$ is increasing so its limit is either finite or ∞ . We need to show that the limit is finite if and only if f is integrable on U . Since the second direction follows immediately from the definitions, we only prove the second direction. Suppose that $\lim_{m \rightarrow \infty} \int_{K_m} f$ is finite. Let $r < \int_U f$, and let $K \subset U$ be a compact set with volume so that $\int_K f > r$. Then from assumptions (1) and (2) we get that there exists m so that $K \subset \text{Int}(K_m)$, and thus $\int_{K_m} f \geq \int_K f \geq r$. This implies that $\lim_{n \rightarrow \infty} \int_{K_m} f \geq \int_U f$, which shows that $\int_U f$ is finite so f is integrable on U which completes the proof. $\square$

Remark 18.2. For every open set U there exists a sequence $\{K_m\}_{m=1}^\infty$ as in Theorem 18.1.

Remark 18.3. Many times it is more convenient to use a family of compact sets with cardinality $\aleph$ instead of $\aleph_0$. In this case, it can be easily verified that the proposition still holds, and we can compute integrals either by $\lim_{t \rightarrow \infty} \int_{K_t} f$ or by $\lim_{t \rightarrow 0} \int_{K_t} f$ instead.

Remark 18.4. So far we have defined the integral of nonnegative functions. Given a function $f: U \rightarrow \mathbb{R}$ we can expand definition by noticing that it can be expressed using nonnegative function:

$$f = f_+ - f_-$$

where

$$f_+ := \frac{|f(x)| + f(x)}{2} = \max(f(x), 0) \quad \text{and} \quad f_- := \frac{|f(x)| - f(x)}{2} = -\min(f(x), 0).$$

Definition 18.2 (Integrability). Let $U \subseteq \mathbb{R}^n$, and let $f: U \rightarrow \mathbb{R}$ be a Riemann integrable function on all compact sets $K \subset U$ with volume. We say that f is integrable on U if f_+ and f_- are both integrable, and we define:

$$\int_U f := \int_U f_+ - \int_U f_-.$$

Proposition 18.2. Let $U \subseteq \mathbb{R}^n$, and let $f: U \rightarrow \mathbb{R}$ be a Riemann integrable function on all compact sets $K \subset U$ with volume. Then f is integrable on U if and only if $|f|$ is integrable on U .

Proof. Recall that

$$|f| = f_+ + f_-.$$

Suppose that f is integrable (so by definition $f^\pm$ are integrable), then we notice that

$$\sup_K \int_K |f| = \sup_K \left(\int_K f_+ + \int_K f_- \right) \leq \int_U f_+ + \int_U f_-.$$

Thus, integrability of f_+ and f_- implies integrability of $|f|$. Next suppose that $|f|$ is integrable. Then, since $f_+, f_- \leq |f|$, from the comparison test they are both integrable and so f is integrable which completes the proof. $\square$

Proposition 18.3 (Comparison test). *Let $f, g: U \rightarrow \mathbb{R}$ be Riemann integrable functions on all compact $K \subset U$ with volume. Then if f is integrable and $|g| \leq f$ then g is also integrable and we have*

$$\left| \int g \right| \leq \int |g| \leq \int |f| < \infty,$$

and in particular, g is integrable.

Remark 18.5. We can verify that the integral of a continuous function $f: U \rightarrow \mathbb{R}$ on any Jordan measurable set $J \subset U \subset \mathbb{R}^n$ is equal to the improper integral of f on J .

Question. Consider the integral $\int_{\mathbb{R}^2} x dx dy$. The sequence $K_n = [-n, n]^2$ exhausts $\mathbb{R}^2$, and also $\lim_{n \rightarrow \infty} \int_{K_n} x dx dy = 0$. This implies that $\int_{\mathbb{R}^2} x dx dy = 0$ (by Theorem 18.1), however $\int_{\mathbb{R}^2} x dx dy$ is clearly indefinite, so we reached a contradiction. What was the problem in our reasoning?

Answer. In Theorem 18.1 we required $f: U \rightarrow \mathbb{R}^+$. In order to use Theorem 18.1 in the general case $f: U \rightarrow \mathbb{R}$, it is best to first use it on $|f|$, then conclude that $\int f$ converges, which is actually a sufficient condition to use Theorem 18.1.

Exercise 18.1. Let $U = (0, 1)^n$ be the n th dimensional unit sphere. Determine for which values of $p > 0$ the function $f(x) = 1/\|x\|^p$ is integrable on U .

19 Change of variables

19.1 The change of variables theorem

Remark 19.1. Recall the change of variables theorem from analysis 2: If $g \in C^1([a, b], \mathbb{R})$ and $f: [a, b] \rightarrow \mathbb{R}$ is continuous, then

$$\int_{g(a)}^{g(b)} f(y) dy = \int_a^b f(g(x)) \cdot g'(x) dx,$$

where $g(x)$ does not have to be injective, however, when $g'(x) \neq 0$ then it is injective, and strictly monotone. In the case it is strictly decreasing, we get that $g(b) < g(a)$ and $g'(x) < 0$, and when it is strictly increasing, we have $g(a) < g(b)$ and $g'(x) > 0$. In both of these cases we can simply write:

$$\int_{g(I)} f(y) dy = \int_I f(g(x)) \cdot g'(x) dx,$$

where $I = [a, b]$ and integrating over $[a, b]$ is the same as integrating from a to b .

Theorem 19.1 (Change of variables theorem). *Let $\Omega \subseteq \mathbb{R}^n$ be a set with volume, and assume that $\Omega \subseteq U$ where U is open. Let $g \in C^1(U, \mathbb{R}^n)$ be an injection with an invertible derivative at every point $u \in U$. Then, if*

- (1) $g(\Omega)$ has volume;
- (2) $f: g(\Omega) \rightarrow \mathbb{R}$ is integrable;
- (3) $f \circ g: \Omega \rightarrow \mathbb{R}$ is integrable.

Then:

$$\int_{g(\Omega)} f(y) dy = \int_{\Omega} [f \circ g(x)] \cdot |\det(Dg(x))| dx.$$

Remark 19.2. The left hand side of the change of variables equation, is integrating over the “manifold world” (usually g parameterizes a manifold), and the right hand side is integrating over the “parameterizing world”. Since we want to find $\int_{g(\Omega)} f(y) dy$ by integrating over the parameterizing world, it is obvious we must integrate over $f \circ g$, and since g sends an infinitesimal neighbourhood U of x in Ω , to a neighbourhood $g(U)$ of $y = g(x)$ of $|\det Dg(x)|$ times the size of U in $g(\Omega)$, we must match up to that by multiplying $f \circ g(x)$ by that factor, by $|\det Dg(x)|$. The following of the remarks are less significant, but this one you want to remember.

Remark 19.3. This theorem shows that g takes any set of measure zero to a set of measure zero.

Remark 19.4. It suffices to assume that g satisfies the conditions on sets that are not of measure zero.

Remark 19.5. The requirement that $g(\Omega)$ has volume is unnecessary, it can be shown under the assumption that Ω has volume, and it is only there for convenience. The same can be said about the requirement that $f \circ g: \Omega \rightarrow \mathbb{R}$ is integrable.

Remark 19.6. The theorem also holds if Ω is an open set, and the integral is improper.

Exercise 19.1. Find $\text{Vol}(B_r^n)$, where $\text{Vol}(B_r^n)$ is the volume of an n -dimensional unit ball in $\mathbb{R}^n$ with the standard metric (as opposed to Example 16.2).

Solution. Consider the change of variables given by the function $y := g(x): \mathbb{R}^n \rightarrow \mathbb{R}^n$ with $g(x) := rx$. We have that $|\det(Dg(x))| = r^n$. Therefore we have that $\text{Vol}(B_r^n) = r^n \text{Vol}(B_1^n)$.

We shall now prove that $\text{Vol}(B_1^n) = \frac{2\pi}{n} \text{Vol}(B_1^{n-2})$. First, we shall write $(x, y) \in \mathbb{R}^{n-2} \times \mathbb{R}^2$ and use Fubini's theorem with the notation from Remark 16.11 to notice that:

$$\text{Vol}(B_1^n) := \int_{\|(x,y)\|_2^2 \leq 1} 1 dV = \int_{\|x\|^2 + \|y\|^2 \leq 1} 1 dV = \int_{\|y\|^2 \leq 1} \left(\int_{\|x\|^2 \leq 1 - \|y\|^2} 1 |d^{n-2}x| \right) d^2y.$$

The integrand is simply the volume of an $n-2$ -dimensional ball with radius $(1 - \|y\|^2)^{\frac{1}{2}}$, so by the previous result from the change of variables theorem we get that:

$$\text{Vol}(B_1^n) := \int_{\|y\|^2 \leq 1} \left(\int_{\|x\|^2 \leq 1 - \|y\|^2} 1 |d^{n-2}x| \right) d^2y = \text{Vol}(B_1^{n-2}) \cdot \int_{\|y\|^2 \leq 1} (1 - \|y\|^2)^{\frac{1}{2}} d^2y.$$

We can now use the familiar polar coordinates from analysis 2 (we will also see about them again later in this section) and get

$$\int_{\|y\|^2 \leq 1} (1 - \|y\|^2)^{\frac{1}{2}} d^2y = \int_0^{2\pi} \left(\int_0^1 (1 - r^2)^{\frac{n-2}{2}} \cdot r |dr| \right) |d\theta| = \dots = \frac{2\pi}{n}.$$

This completes the proof of the identity $\text{Vol}(B_1^n) = \frac{2\pi}{n} \text{Vol}(B_1^{n-2})$. We know that $\text{Vol}(B_1^1) = 2$ and that $\text{Vol}(B_1^2) = \pi$, so by this recursive relation, and the identity $\text{Vol}(B_r^n) = r^n \text{Vol}(B_1^n)$ we get that formulas

$$\text{Vol}(B_r^{2k}) = \frac{\pi^k}{k!} r^{2k} \quad \text{and} \quad \text{Vol}(B_r^{2k+1}) = \frac{2(k!)(4\pi)^k}{(2k+1)!} r^{2k+1},$$

for any integer $k \in \mathbb{Z}$.

Definition 19.1 (Parallelopete). A parallelopete (or parallelepiped) generated by the vectors $v_1, \dots, v_n \in \mathbb{R}^n$ is the set

$$P(v_1, \dots, v_n) := \left\{ \sum_{i=1}^n a_i v_i : 0 \leq a_i \leq 1 \right\}.$$

When $a \in \mathbb{R}^n$ we also call $a + P(v_1, \dots, v_n)$ a parallelopete, and if $\{v_1, \dots, v_n\}$ are linearly dependent, we say that parallelopete is degenerate.

Exercise 19.2. Let $v_1, v_2 \in \mathbb{R}^2$. Then we have

$$\text{Vol}(a + P(v_1, v_2)) = |\det(v_1 \mid v_2)| = \|v_1\| \cdot \|v_2\| \cdot |\sin \angle(v_1, v_2)|.$$

Exercise 19.3. Let $\Omega_1 \subseteq \mathbb{R}^n$ and $\Omega_2 \subseteq \mathbb{R}^m$ be sets with volume. Then

$$\text{Vol}(\Omega_1 \times \Omega_2) = \text{Vol}(\Omega_1) \cdot \text{Vol}(\Omega_2).$$

Exercise 19.4. Prove the following:

- (1) Let Q be a nondegenerate parallelopete. Then Q is a set with positive volume.
- (2) Let Q be a degenerate parallelopete. Then Q is a set with zero volume.
- (3) Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear map, and let Q be a parallelopete. Then $T(Q)$ is a parallelopete and in particular it has volume.

Proposition 19.2. Let $\Omega \subset \mathbb{R}^n$ be a set with volume, and let $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear map. Then $T(\Omega)$ has volume and

$$\text{Vol}(T(\Omega)) = |\det(T)| \cdot \text{Vol}(\Omega).$$

Proof. To be added. □

Remark 19.7. Given a parallellotope $P(v_1, \dots, v_n) \subset \mathbb{R}^n$, it can be represented as another parallellotope $T(P(e_1, \dots, e_n))$ for some linear map $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$, and $e_1, \dots, e_n$ the unit vectors. Then by the previous proposition we get

$$\text{Vol}(P(v_1, \dots, v_n)) = |\det(T)| \cdot \text{Vol}(P(e_1, \dots, e_n)) = |\det(T)|.$$

This is one of the most important meanings of the determinant.

Remark 19.8. Theorem 19.2 is in fact a special case of the change of variables theorem.

For now we skip the proof of the change of variables theorem...

19.2 Example of polar coordiantes

In this case, the “parameterizing world” is $[0, \infty] \times [0, 2\pi]$, and the “manifold world” is $\mathbb{R}^2$.

Example 19.1 (Polar coordinates). The parameterization $g: [0, \infty] \times [0, 2\pi] \rightarrow \mathbb{R}^2$ is given by

$$g(r, \theta) = (r \cos \theta, r \sin \theta),$$

where $r = \sqrt{x^2 + y^2}$ is the distance of (x, y) from the origin $(0, 0)$, and θ is the angle between the vector (x, y) and the positive direction of the x -axis.

The map g is one to one and C^1 except for a set of measure zero. Moreover, we notice that

$$\det(Dg(x, y)) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \neq 0$$

except for a set of measure zero. Thus, according to the change of variables theorem we have

$$\iint_{g(\Omega)} f(x, y) dx dy = \iint_{\Omega} f(r \cos \theta, r \sin \theta) |r| dr d\theta.$$

Remark 19.9. Planar physical systems with bodies moving around a central point, or phenomena originating from a central point, are often simpler and more intuitive to model using polar coordinates.

19.3 Example of cylindrical coordiantes

In this case, the “parameterizing world” is $[0, \infty] \times [0, 2\pi] \times \mathbb{R}$, and the “manifold world” is $\mathbb{R}^3$.

Example 19.2 (Cylindrical coordinate system). Let $\Omega \subset \mathbb{R}^3$ be the set in the first quadrant bounded by

$$y = x \quad \text{and} \quad y = 2x \quad \text{and} \quad z = x^2 + y^2 \quad \text{and} \quad z = 2(x^2 + y^2)$$

and $z = h$ for some $h > 0$. We want to find the volume of Ω for $x, y, z \geq 0$. We see that the by definition approach gives us the following integral

$$V = \iiint_{\Omega} 1 dx dy dz.$$

This is a good time to move to cylindrical coordinates. The parameterization $g: [0, \infty] \times [0, 2\pi] \times \mathbb{R} \rightarrow \mathbb{R}^3$ is given by

$$g(r, \theta, z) = (r \cos \theta, r \sin \theta, z).$$

The variable r is called the radial distance; it represents the distance of the point (x, y, z) from the z -axis. The variable θ is called the azimuthal angle; it represents the angle between the projection of the point (x, y, z) on the chosen plane (in this case $z = 0$) and the reference direction on the plane (in this case the unit vector $\mathbf{x}$). The variable z is called the axial coordinate (or height); it represents the signed distance from the chosen plane (in this case $z = 0$) to the point (x, y, z) . We have that

$$\det(Dg(r, \theta, z)) = \begin{vmatrix} \cos \theta & -r \sin \theta & 0 \\ \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r > 0$$

except for when the points are exactly on the z -axis, but since we bound Ω with $z < h$ this set is of measure zero and is thus negligible. We need to find the image of Ω under this parameterization. Under the constraints we get that

$$\begin{aligned} y = x &\implies r \sin \theta = r \cos \theta \implies \tan \theta = 1 \implies \theta = \frac{\pi}{4} \\ y = 2x &\implies r \sin \theta = 2r \cos \theta \implies \tan \theta = 2 \implies \theta = \arctan(2) \\ z = x^2 + y^2 &\implies z = r^2 \\ z = 2(x^2 + y^2) &\implies z = 2r^2 \\ z &= h. \end{aligned}$$

Recall that in this parameterization we also have implicit constraints that $r \geq 0$ and $0 \leq \theta \leq 2\pi$. Therefore, the final limits for the variables are:

$$\frac{\pi}{4} \leq \theta \leq \arctan(2) \quad \text{and} \quad \sqrt{\frac{z}{2}} \leq r \leq \sqrt{z} \quad \text{and} \quad 0 \leq z \leq h.$$

Using the change of variables theorem and Fubini's theorem we get that

$$V = \iiint_{\tilde{\Omega}} r dr d\theta dz = \int_{\frac{\pi}{4}}^{\arctan(2)} d\theta \int_0^h dz \int_{\sqrt{\frac{z}{2}}}^{\sqrt{z}} r dr.$$

This integral can be calculated directly:

$$\int_{\frac{\pi}{4}}^{\arctan(2)} d\theta \int_0^h dz \int_{\sqrt{\frac{z}{2}}}^{\sqrt{z}} r dr = \left(\arctan(2) - \frac{\pi}{4} \right) \int_0^h \left(\frac{z}{2} - \frac{z}{4} \right) dz = \left(\arctan(2) - \frac{\pi}{4} \right) \frac{h^2}{8}.$$

Remark 19.10. Cylindrical coordinates are useful when working with solids of revolution.

19.4 Example of spherical coordinates

In this case, the “parameterizing world” is $[0, \infty] \times [0, \pi] \times [0, 2\pi]$, and the “manifold world” is the Euclidean space $\mathbb{R}^3$.

Example 19.3 (Spherical coordinates). Let $P := (x_0, y_0, z_0) \in \mathbb{R}^3$, and let

$$\Omega := \left\{ (x, y, z) \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \leq R^2 \right\}.$$

We want to calculate the volume of Ω . We can parameterize $\mathbb{R}^3$ with origin at (x_0, y_0, z_0) by $\mathbb{R}$ with origin at $(0, 0, 0)$ using the following parameterization:

$$g(\tilde{x}, \tilde{y}, \tilde{z}) = (\tilde{x} - x_0, \tilde{y} - y_0, \tilde{z} - z_0),$$

so that $g(B(P, R)) = B(0, R)$. We notice that

$$\det(Dg(\tilde{x}, \tilde{y}, \tilde{z})) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 \neq 0.$$

By the change of variables theorem we get that

$$\text{Vol}(B(P, R)) = \iiint_{g(B(0, R))} 1 dx dy dz = \iiint_{B(0, R)} 1 d\tilde{x} d\tilde{y} d\tilde{z} = \text{Vol}(B(0, R)).$$

This shows rigorously that the volume of a sphere does not depend on the position of its center. We now use spherical coordinates to calculate the volume of the sphere around the origin. We use the parameterization $g: [0, \infty] \times [0, \pi] \times [0, 2\pi] \rightarrow \mathbb{R}^3$ given by

$$g(r, \theta, \varphi) = (r \cos \varphi \sin \theta, r \sin \varphi \sin \theta, r \cos \theta).$$

The variable r is called the radial distance; it represents the distance of the point from the origin. The variable θ is called the polar angle; it represents the angle between the radial line (the vector from the origin to the given point) and the positive side of the polar axis (in this case z). The variable φ is called the azimuthal angle; it represents the angle of rotation of the radial line around the polar axis (in this case z) relative to the positive side of the x axis.

Computing the absolute value of $\det(Dg(r, \theta, \varphi))$ is a simple matter of calculation:

$$\begin{aligned} |\det Dg(r, \varphi, \theta)| &= \left| \begin{vmatrix} \cos \varphi \sin \theta & -r \sin \varphi \sin \theta & r \cos \varphi \cos \theta \\ \sin \varphi \sin \theta & r \cos \varphi \sin \theta & r \sin \varphi \cos \theta \\ \cos \theta & 0 & -r \sin \theta \end{vmatrix} \right| \\ &= \left| (\cos \theta) \begin{bmatrix} -r^2 \sin \theta \cos \theta \end{bmatrix} - r \sin \theta \begin{bmatrix} r \sin^2 \theta \end{bmatrix} \right| = r^2 \sin \theta. \end{aligned}$$

Note that since $0 \leq \theta \leq \pi$ the value of $r^2 \sin \theta$ is almost always (except for a set of measure zero) positive, which justifies the last transition. It is clear that the image of $B(0, R)$ is the set defined by the constraint $r^2 \leq R^2$ and thus

$$\tilde{\Omega} = \{(r, \theta, \varphi) \mid 0 \leq \theta \leq \pi \text{ and } 0 \leq \varphi \leq 2\pi \text{ and } 0 \leq r \leq R\}.$$

This means that $\tilde{\Omega}$ is a box, so by Fubini's theorem and the change of variables theorem we get that:

$$\text{Vol}(B(P, R)) = \iiint_{\tilde{\Omega}} r^2 \sin \theta dr d\varphi d\theta = \int_0^R r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi = \frac{4}{3} \pi R^3.$$

Remark 19.11. Spherical coordinates are useful when Ω has spherical symmetry around the origin.

Remark 19.12 (Deriving spherical coordinates). Unlike polar coordinates or cylindrical coordinates, it's not as easy to derive the parameterization function for polar coordinates:

$$g(r, \theta, \varphi) = (r \cos \varphi \sin \theta, r \sin \varphi \sin \theta, r \cos \theta).$$

The easiest method to derive it is by noticing that any point $(x, y, z) \in \mathbb{R}^3$ with norm r is simply the point $(0, 0, r)$ rotated θ degrees, without loss of generality around the y axis (inside the plane $y = 0$), then rotated φ degree around the z axis. If we do this process such that the directions of the rotations are determined by the right hand rule, we get the point that is associated with the point (r, θ, φ) in standard spherical coordinates. Therefore the parameterization is given by:

$$\begin{aligned} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= R_z(\varphi)R_y(\theta)(0, 0, r)^T \\ &= \begin{pmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ r \end{pmatrix} \\ &= \begin{pmatrix} r \cos \varphi \sin \theta \\ r \sin \varphi \sin \theta \\ r \cos \theta \end{pmatrix}. \end{aligned}$$

A different way to look at it would be to notice that $r = \sqrt{x^2 + y^2 + z^2}$ and $\cos(\theta) = \frac{z}{r}$, and $\tan(\varphi) = \frac{y}{x}$, and compute the values of (x, y, z) in terms of (r, θ, φ) . Deriving the rotation matrices themselves is also not hard,

Remark 19.13 (Deriving the rotation matrices). An arbitrary rotation matrix $R_z(\theta)$ in $\mathbb{R}^3$ is the matrix that rotates vectors around the z -axis, in the direction determined by the right hand rule (put your thumb in the direction of the positive direction of the z -axis, and then $R_z(\theta)$ rotates your index finger in the direction of your middle finger). Note that $R_z(\theta)$ is the same as the usual rotation matrix from $\mathbb{R}^2$ which sends $(1, 0)$ to $(\cos \theta, \sin \theta)$ and $(0, 1)$ to $(-\sin \theta, \cos \theta)$ and it fixes the z -component of the vectors and so it must be of the form:

$$R_z(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

However, an astute reader would notice that $R_y(\theta)$ which we saw earlier looks slightly different, and not only because we now fix the y -component of the vectors. Indeed, in order to make sure the rotation is in the direction determined by the right hand rule, we must make sure that the determinant is positive (in fact since this matrix is a rotation it must be in $SO(3, \mathbb{R})$ and so it must have determinant 1), so $R_y(\theta)$ is derived in exactly the same way except we multiply by -1 in order to make it a true rotation (and not a reflection).

19.5 Practice

Exercise 19.5. Let Ω be the region bounded between the curves $y = \frac{1}{x}$, $y = \frac{4}{x}$, $y = 3x$, and $y = x$ in the first quadrant of $\mathbb{R}^2$. Find $\int_{\Omega} \frac{y}{x} |dx dy|$.

Solution. We notice that

$$\Omega = \left\{ (x, y) \mid 1 \leq \frac{y}{x} \leq 3 \text{ and } 1 \leq xy \leq 4 \right\},$$

and so our “manifold world” is Ω and our parameterizing world is another set in the real plane $A := [1, 3] \times [1, 4]$. Well, now it’s not very hard to parameterize A using Ω , the parameterization is

$$g(x, y) = (u(x, y), v(x, y)) = \left(xy, \frac{y}{x}\right) \in A.$$

This parameterization however, is exactly the inverse of the parameterization we need. By the change of variables theorem we have

$$\int_{g^{-1}(A)} \frac{y}{x} |dx dy| = \int_A \frac{g_2^{-1}(u, v)}{g_1^{-1}(u, v)} \left| \det \left(Dg^{-1}(u, v) \right) \right| |du dv| = \int_A \frac{y}{x} \left| \det \left(Dg^{-1}(u, v) \right) \right| |du dv|.$$

By the inverse function theorem we have $Dg^{-1}(u, v) = [Dg(x, y)]^{-1}$, this gives us that

$$\left| \det \left(Dg^{-1}(u, v) \right) \right| = \left| \det \left(Dg(x, y) \right) \right|^{-1} = \left| \begin{pmatrix} y & x \\ \frac{-y}{x^2} & \frac{1}{x} \end{pmatrix} \right|^{-1} = \frac{x}{2y} = \frac{1}{2v}.$$

Thus,

$$\int_{\Omega} \frac{y}{x} |dx dy| = \int_A v \cdot \frac{1}{2v} |du dv| = \frac{1}{2} \int_1^3 \int_1^4 1 |du dv| = \frac{6}{2} = 3.$$

Exercise 19.6. Compute the volume of the set

$$\Omega = \left\{ (x, y, z) \in \mathbb{R}^3 \mid z \geq \sqrt{x^2 + y^2} \text{ and } x^2 + y^2 + z^2 \leq 1 \right\}.$$

Solution. Let $g(r, \theta, \varphi) = (r \cos \varphi \sin \theta, r \sin \varphi \sin \theta, r \cos \theta)$ be the parameterization of spherical coordinates. Then the constraints are $r < 1$ and $r \cos \theta \geq r \sin \theta$ or simply $0 \leq \theta \leq \frac{\pi}{4}$. Then, by the change of variables theorem

$$\begin{aligned} \text{Vol}(\Omega) &= \int_0^1 |dr| \int_0^{\pi/4} |d\theta| \int_0^{2\pi} r^2 \sin \theta |d\varphi| = 2\pi \int_0^{\pi/4} \frac{1}{3} \sin \theta |d\theta| \\ &= \frac{2\pi}{3} \left[-\cos \frac{\pi}{4} + \cos 0 \right] = \frac{2\pi}{3} \left[1 - \frac{1}{\sqrt{2}} \right]. \end{aligned}$$

20 k -dimensional volume

20.1 Intuition

In the change of variables theorem we had that under the right conditions

$$\int_{g(\Omega)} f(y) dy = \int_{\Omega} [f \circ g(x)] |\det(Dg(x))| dV(x)$$

and in particular we have that

$$\text{Vol}(g(\Omega)) = \int_{g(\Omega)} 1 dV(y) = \int_{\Omega} |\det(Dg(x))| dV(x).$$

Since we know that locally g acts like a linear map, we know that if we take Ω to be a small enough box, the ratio between the volume of Ω and $g(\Omega)$ (almost a parallelotope) will be as close as we want to $|\det(Dg(x))|$. Intuitively, this means that

$$\lim_{dx_i \rightarrow 0} \text{Vol}(g(P(dx_1, \dots, dx_n))) = |\det(Dg(x))| \text{Vol}(P(dx_1, \dots, dx_n)).$$

In the case of two dimensions, this already looks very much like calculating some area on a surface.

Example 20.1 (Volume of a torus). Recall that in Example 11.5 we saw that one simple parameterization for the torus is

$$\varphi(u, v) = \begin{pmatrix} \cos v & 0 & -\sin v \\ 0 & 1 & 0 \\ \sin v & 0 & \cos v \end{pmatrix} \begin{pmatrix} a + b \cos u \\ b \sin u \\ 0 \end{pmatrix} = \begin{pmatrix} (a + b \cos u) \cos v \\ b \sin u \\ (a + b \cos u) \sin v \end{pmatrix}.$$

We now define a parameterization for the interior of the torus by applying the function

$$g(t, u, v) := \begin{pmatrix} (a + t \cos u) \cos v \\ t \sin u \\ (a + t \cos u) \sin v \end{pmatrix}.$$

on the set $\Omega = [0, b] \times [0, 2\pi] \times [0, 2\pi]$. The interior of the torus is now $\tilde{\Omega} = g(\Omega)$. Since g is injective except on a set of measure zero, and since its derivative is invertible at every point $u \in \Omega$ except for a set of measure zero, g does not violate the conditions for the change of variables theorem and we can continue. We have that

$$Dg = \begin{pmatrix} \cos u \sin v & -t \sin u \cos v & -(a + t \cos u) \sin v \\ \sin u & t \cos u & 0 \\ \cos u \sin v & -t \sin u \sin v & (a + t \cos u) \cos v \end{pmatrix} \implies |\det(Dg)| = \dots = t(a + t \cos u)$$

and $\det(Dg) \neq 0$ except when $t = 0$ or $a = -t \cos u$, but since this is a set of measure zero, it does not affect the calculation. Note that we ignore the absolute value since by the definition of a the expression $t(a + t \cos u)$ is almost always positive. We get that

$$\begin{aligned} \text{Vol}(g(\Omega)) &= \int_{g(\Omega)} 1 dV(x, y, z) = \int_{\Omega} \det(Dg(t, u, v)) dV(t, u, v) \\ &= \int_{[0, b] \times [0, 2\pi] \times [0, 2\pi]} t(a + t \cos u) dV(t, u, v). \end{aligned}$$

Using Fubini's theorem we get that

$$\begin{aligned} \int_{[0,b] \times [0,2\pi] \times [0,2\pi]} t(a + t \cos u) dV(t, u, v) &= \int_0^{2\pi} \int_0^{2\pi} \int_0^b t(a + t \cos u) dt du dv \\ &= \int_0^{2\pi} \int_0^{2\pi} \left(\frac{b^2 a}{2} + \frac{b^3}{3} \cos u \right) du dv = 2\pi^2 b^2 a. \end{aligned}$$

With this, we finished calculating the volume of a torus. We now proceed to define some tools that will help us calculate its surface area.

20.2 k -dimensional volume

Remark 20.1. We want to derive the k -dimensional volume of a k -parallelotope embedded in a higher dimensional Euclidean space $\mathbb{R}^n$. Let $P(v_1, \dots, v_k) \subset \mathbb{R}^n$ be a k -parallelotope where

$$v_i := \begin{pmatrix} \tilde{v}_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$$

and $\tilde{v}_i \in \mathbb{R}^k$ for all $1 \leq i \leq k$. Naturally, we want to claim that

$$\text{Vol}(a + P(v_1, \dots, v_k)) = \text{Vol}(P(\tilde{v}_1, \dots, \tilde{v}_k)),$$

and thus according to Remark 19.7 we get that

$$\text{Vol}(a + P(v_1, \dots, v_k)) = \text{Vol}(P(\tilde{v}_1, \dots, \tilde{v}_k)) = \left| \det [\tilde{v}_1 \mid \dots \mid \tilde{v}_k] \right| = \left| \det \tilde{V} \right|$$

where $\tilde{V} = [\tilde{v}_1 \mid \dots \mid \tilde{v}_k]$ is a $k \times k$ matrix. However, it is of the utmost importance to notice that we got this result precisely because $\tilde{V}$ is the transition matrix from the standard basis of $\mathbb{R}^k$ to the basis $\{\tilde{v}_1 \mid \dots \mid \tilde{v}_k\}$ of $\mathbb{R}^k$. But when given random vectors an arbitrary parallelotope $P(v_1, \dots, v_k)$, each v_i is not necessarily comprised from $\tilde{v}_i$ and $n - k$ more zero coordinates so things become more difficult. However, luckily for us there is a quick general way to compute $\text{Vol}(P)$ by using Gram's matrix. It is very important to know it because most parallelotopes are not given in this extremely convenient way.

Definition 20.1 (Gram matrix). Let $v_1, \dots, v_n$ be vectors in an inner product space (not necessarily of dimension n). If $V = [v_1 \mid \dots \mid v_n]$ is a matrix so that v_i is the i th column, then the Gram matrix (or Gramian) of $v_1, \dots, v_n$ is an $n \times n$ matrix defined by:

$$G(v_1, \dots, v_n) := V^* V = \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{pmatrix},$$

where V^* is the Hermitian transpose of V . The determinant of the Gram matrix is denoted by $\mathcal{G}(v_1, \dots, v_n)$ or simply by $\mathcal{G}$.

Proposition 20.1 (Volume of a parallelotope). *Let $v_1, \dots, v_k \in \mathbb{R}^n$ be k linearly independent vectors in $\mathbb{R}^n$. Then $\text{Vol}(P(v_1, \dots, v_k)) = \sqrt{\mathcal{G}(v_1, \dots, v_k)}$.*

Proof. Note that $\sqrt{\mathcal{G}(v_1, \dots, v_k)}$ is always well defined since $G = V^*V$ is positive-definite, so $\mathcal{G} > 0$. Indeed, if as before

$$v_i := \begin{pmatrix} \tilde{v}_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$$

we denote

$$V = [v_1 \mid \dots \mid v_k] = \begin{pmatrix} \tilde{V} \\ 0 \end{pmatrix} \in \mathbb{R}^{n \times k},$$

and since $\mathcal{G} := \det(G(\tilde{v}_1, \dots, \tilde{v}_k))$ then

$$\sqrt{\mathcal{G}} = \sqrt{\det V^*V} = \sqrt{\det \tilde{V}^*\tilde{V}} = |\det \tilde{V}| = \text{Vol}(P(v_1, \dots, v_k)),$$

where the last transition is by Remark 20.1. In the general case, where we know nothing about $v_1, \dots, v_k$, we first apply the Gram–Schmidt process on these vectors (without normalizing them), to obtain a new set of orthogonal vectors $w_1, \dots, w_k$. Geometrically, it is clear that $\text{Vol}(P(v_1, \dots, v_k)) = \text{Vol}(P(w_1, \dots, w_k))$, and since $w_1, \dots, w_k$ are orthonormal we get the identity $\text{Vol}(P(v_1, \dots, v_k)) = \prod_{i=1}^k \|w_i\|$. With some technical computations, it is possible to show (for example by induction) that $\prod_{i=1}^k \|w_i\|^2 = \mathcal{G}$, and therefore $\text{Vol}(P(v_1, \dots, v_k)) = \sqrt{\mathcal{G}}$, which completes the proof. $\square$

Remark 20.2. The gram matrix $G(v_1, \dots, v_n)$ is not dependent on the choice of some orthonormal basis. That is because if $A \in O(n)$ is an orthogonal matrix that changes one orthogonal basis to another then $\langle Av_i, Av_j \rangle = \langle v_i, A^T Av_j \rangle = \langle v_i, v_j \rangle$.

Definition 20.2 (k -dimensional parallelotope volume). Let $v_1, \dots, v_k \in \mathbb{R}^n$ be some vectors in the Euclidean space such that $k < n$. The k -dimensional volume of the parallelotope $a + P(v_1, \dots, v_k)$ is given by

$$\text{Vol}_k(a + P(v_1, \dots, v_k)) = \sqrt{\det V^*V} = \sqrt{\mathcal{G}},$$

where $V = [v_1 \mid \dots \mid v_k]$, and $\mathcal{G}$ is the determinant of the Gram matrix corresponding to the vectors $v_1, \dots, v_k$ in any orthonormal basis of the subspace $\text{span}\{v_1, \dots, v_k\} \subset \mathbb{R}^n$.

Example 20.2. In the special case $n = 3$ and $k = 2$ we get that for vectors

$$v_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

with $V = [v_1 \mid v_2]$ that the volume is

$$\begin{aligned} \text{Vol}_2(P(v_1, v_2)) &= \sqrt{\det \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{pmatrix}} = \dots \\ &= \sqrt{(x_1y_2 - y_1x_2)^2 + (x_1z_2 - x_2z_1)^2 + (y_1z_2 - y_2z_1)^2} \\ &= \|v_1 \times v_2\| \end{aligned}$$

where $\times$ denotes the vector product defined by

$$v_1 \times v_2 := \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} = \hat{i} \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - \hat{j} \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + \hat{k} \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

Notice that this product results in a vector with length $\|v_1\|\|v_2\|\sin\alpha$ where α is the angle between v_1 and v_2 , with direction orthogonal to the plane spanned by v_1 and v_2 , and its sign is given by the right hand rule.

Using these results we can conclude that the 2-dimensional volume of $P(v_1, v_2)$ is equal to $\|v_1 \times v_2\| = \|v_1\|\|v_2\|\sin\alpha$, which indeed coincides with the classical result from Euclidean geometry.

Exercise 20.1. Given $m = n - 1$ vectors $v_1, \dots, v_m$ in $\mathbb{R}^n$ we define the generalized vector product $N = v_1 \times v_2 \times \dots \times v_m$ to be the unique vector in $\mathbb{R}^n$ such that for all $x \in \mathbb{R}^n$ we have

$$\langle N, x \rangle = \det \begin{pmatrix} | & | & | & | \\ x & v_1 & \cdots & v_m \\ | & | & | & | \end{pmatrix}.$$

- (1) Show that the generalized product is well defined (such a vector exists and is unique). Find the direction of N , and find the explicit coordinates of N .

- (2) Show that

$$\|N\|^2 = \det \begin{pmatrix} | & | & | & | \\ N & v_1 & \cdots & v_m \\ | & | & | & | \end{pmatrix}.$$

- (3) Show that $\text{Vol}_m(P(v_1, \dots, v_m)) = \|v_1 \times v_2 \times \dots \times v_m\|$.

Remark 20.3. Let $T: \mathbb{R}^k \rightarrow \mathbb{R}^n$ be a linear transformation such that $k \leq n$, and let $v_1, \dots, v_k \in \mathbb{R}^k$. Then

$$T(P(v_1, \dots, v_k)) = P(Tv_1, \dots, Tv_k)$$

and

$$\begin{aligned} \text{Vol}_k(T(P(v_1, \dots, v_k))) &= \sqrt{\det(TV)^*TV} = \sqrt{\det V^* \det T^*T \det V} \\ &= \sqrt{\det T^*T} \text{Vol}_k(P(v_1, \dots, v_k)). \end{aligned}$$

Intuitively, this means that a linear transformation $T: \mathbb{R}^k \rightarrow \mathbb{R}^n$ amplifies the k -dimensional volume of a parallelotope by a factor of $\sqrt{\det T^*T}$.

Example 20.3 (Surface area). Let $U \subset \mathbb{R}^2$, let $\varphi: U \rightarrow \mathbb{R}^3$ be a parameterization of some part of a manifold $S = \varphi(U) \subset M$. In particular φ is injective, continuously differentiable, and $\text{rank } D\varphi = 2$. Moreover, for a sufficiently small rectangle R we have that

$$(u, v) + P\left(\begin{pmatrix} \Delta u \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \Delta v \end{pmatrix}\right) \xrightarrow[\varphi]{\text{approx.}} \varphi(u, v) + D\varphi\left(P\left(\begin{pmatrix} \Delta u \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \Delta v \end{pmatrix}\right)\right) := \tilde{R}.$$

Since we have that

$$D\varphi\left(\begin{pmatrix} \Delta u \\ 0 \end{pmatrix}\right) = \varphi_u \Delta u \quad \text{and} \quad D\varphi\left(\begin{pmatrix} 0 \\ \Delta v \end{pmatrix}\right) = \varphi_v \Delta v$$

we get from Example 20.2 that

$$\text{Vol}_2(\tilde{R}) = \|\varphi_u \times \varphi_v\| \Delta v \Delta u$$

so if we denote $R' := \varphi(R)$, we can conclude that

$$\text{Area}(R') \approx \|\varphi_u \times \varphi_v\| \Delta v \Delta u.$$

Definition 20.3 (Surface area). The surface area of some part of a 2-dimensional manifold $S \subset M \subseteq \mathbb{R}^3$ parameterized by $\varphi: U \rightarrow S$ is defined by

$$A(S) := \int_U \sqrt{\mathcal{G}(x)} dV = \int_U \sqrt{\det(D\varphi^*(u, v)D\varphi(u, v))} dudv = \int_U \|\varphi_u \times \varphi_v\| dudv.$$

We will see a more general definition for k -dimensional volume in the following section.

Remark 20.4. In case $\varphi \equiv T$ for some linear transformation T , then $D\varphi = T$, and we have that

$$A(S) = \int_U \sqrt{\det T^*T} dudv = \sqrt{\det T^*T} \text{Vol}_2(U).$$

Example 20.4. Recall that the equation $z = x^2 + y^2$ describes a paraboloid. Therefore, the equation $z = \sqrt{x^2 + y^2}$ behaves similarly but grows linearly. The equation $z = \sqrt{x^2 + y^2}$ describes a cone.

We want to find the surface area of the cone bounded by $z = 1$. We can do this by using the coordinates:

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ z &= r \end{aligned}$$

where $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Then $g: \Omega \rightarrow \mathbb{R}^3$ where $\Omega = [0, 1] \times [0, 2\pi]$ and

$$g(r, \theta) := (r \cos \theta, r \sin \theta, r)$$

is a parameterization of the cone under $z = 1$. Since

$$g_\theta = \begin{pmatrix} -r \sin \theta \\ r \cos \theta \\ 0 \end{pmatrix} \quad \text{and} \quad g_r = \begin{pmatrix} \cos \theta \\ \sin \theta \\ 1 \end{pmatrix}$$

we have $g_r \times g_\theta = r(-\cos \theta, -\sin \theta, 1)$, and thus $\|g_r \times g_\theta\| = \sqrt{2}r$. Similarly we have that

$$G = \begin{pmatrix} \langle g_r, g_r \rangle & \langle g_r, g_\theta \rangle \\ \langle g_\theta, g_r \rangle & \langle g_\theta, g_\theta \rangle \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & r^2 \end{pmatrix},$$

so $\mathcal{G} = 2r^2$. Therefore, by using the definition for surface area we get

$$A(g(\Omega)) := \int_\Omega \sqrt{\mathcal{G}} |dudv| = \int_\Omega \|g_\theta \times g_r\| |dudv| = \int_0^{2\pi} \left(\int_0^1 \sqrt{2}r dr \right) d\theta = \sqrt{2}\pi.$$

Example 20.5. We can similarly compute the surface area of a 2-sphere S^2 in $\mathbb{R}^3$. We can do this by using spherical coordinates:

$$\begin{aligned} x &= \cos \theta \sin \varphi \\ y &= \sin \theta \sin \varphi \\ z &= \cos \varphi \end{aligned}$$

where $0 \leq \varphi \leq \pi$ and $0 \leq \theta \leq 2\pi$. Then $g: \Omega \rightarrow \mathbb{R}^3$ where $\Omega = [0, \pi] \times [0, 2\pi]$ and

$$g(\varphi, \theta) := (\cos \theta \sin \varphi, \sin \theta \sin \varphi, \cos \varphi)$$

is a parameterization of the 2-sphere $g(\Omega) = S^2$. Since

$$g_\varphi = \begin{pmatrix} \cos \theta \cos \varphi \\ \sin \theta \sin \varphi \cos \theta \\ -\sin \theta \end{pmatrix} \quad \text{and} \quad g_\theta = \begin{pmatrix} -\sin \theta \sin \varphi \\ \cos \theta \sin \varphi \\ 0 \end{pmatrix}$$

we have

$$g_\theta \times g_\varphi = -(\cos \theta \sin^2 \varphi, \sin \theta \sin^2 \varphi, \sin \varphi \cos \varphi),$$

and thus $\|g_\theta \times g_\varphi\| = \sin \varphi$. Similarly we have that

$$G = \begin{pmatrix} \langle g_\theta, g_\theta \rangle & \langle g_\theta, g_\varphi \rangle \\ \langle g_\varphi, g_\theta \rangle & \langle g_\varphi, g_\varphi \rangle \end{pmatrix} = \begin{pmatrix} \sin^2 \varphi & 0 \\ 0 & 1 \end{pmatrix},$$

so $\mathcal{G} = \sin^2 \varphi$. Therefore, by using the definition for surface area we get

$$A(S^2) = A(g(\Omega)) := \int_{\Omega} \sqrt{\mathcal{G}} |du dv| = \int_{\Omega} \|g_\theta \times g_\varphi\| |du dv| = \int_0^{2\pi} \left(\int_0^\pi \sin \varphi d\varphi \right) d\theta = 4\pi.$$

21 Surface integrals

21.1 Surface area

Definition 21.1 (k -dimensional surface area). Let $U \subset \mathbb{R}^k$, let $\Omega \subset U$ be a set with volume, let $\varphi: U \rightarrow \mathbb{R}^n$ be a parameterization as in Remark 11.7 (this means $\varphi \in C^1$, injective, with a continuous inverse, and $\text{rank } D\varphi = k$). Denote $S = \varphi(\Omega)$. We define the k -dimensional surface area of S to be

$$A_k(S) := \int_{\Omega} \sqrt{\mathcal{G}(\mathbf{x})} dV = \int_{\Omega} \sqrt{\det(D\varphi^*(\mathbf{x})D\varphi(\mathbf{x}))} dV(\mathbf{x}),$$

where $\mathbf{x} = (x_1, x_2, \dots, x_k)$ are the k variables of φ .

Remark 21.1. We can require φ to be injective, or $\text{rank } D\varphi = k$ everywhere except for a set of points of measure zero.

Remark 21.2. When $n = 3$ and $k = 2$ this definition agrees with Definition 20.3.

Example 21.1 (Length of a curve). In the case of $k = 1$ we have $\varphi: [a, b] \rightarrow \mathbb{R}^n$ such that $\varphi'(t) \neq 0$ for all $t \in [a, b]$, and we get that $A_1(\varphi[a, b])$ is the “1-dimensional surface area of $\varphi[a, b]$ ”. In other words, the length of the curve $\varphi[a, b]$. Indeed, we see that

$$\varphi(t) = \begin{bmatrix} \varphi_1(t) \\ \vdots \\ \varphi_n(t) \end{bmatrix} \quad \text{and} \quad \varphi'(t) = \begin{bmatrix} \varphi'_1(t) \\ \vdots \\ \varphi'_n(t) \end{bmatrix} = D\varphi|_t.$$

From this follows that

$$D\varphi^*(t)D\varphi(t) = \begin{bmatrix} \varphi'_1(t) & \cdots & \varphi'_n(t) \end{bmatrix} \begin{bmatrix} \varphi'_1(t) \\ \vdots \\ \varphi'_n(t) \end{bmatrix} = \sum_{i=1}^n (\varphi'_i(t))^2.$$

Thus, the length of the curve is

$$A_1(\varphi[a, b]) = \int_a^b \sqrt{\sum_{i=1}^n (\varphi'_i(t))^2} dt = \int_a^b \|\varphi'(t)\| dt$$

which is the same result we got in Section 14.1.

Remark 21.3. We now have a mathematical notion of surface area, which means we can now integrate over surfaces.

21.2 Surface integral

Definition 21.2 (Surface integral of a scalar field). Using the notation from Definition 21.1, let $n = 3$ and $k = 2$. Let $f: \varphi(\Omega) = S \rightarrow \mathbb{R}$ be continuous. Then the surface integral of f on S denoted by $\int_S f dA = \int_S f d\sigma = \int_S f(\sigma) d\sigma$ is defined by

$$\int_S f dA := \int_{\Omega} (f \circ \varphi) \cdot \sqrt{\det(D\varphi^*D\varphi)} dV = \int_{\Omega} (f \circ \varphi) \cdot \|\varphi_u \times \varphi_v\| du dv.$$

In general when $k \leq n$ we define the surface integral of f on S in the same way:

$$\int_S f dA := \int_{\Omega} (f \circ g)(x) \cdot \sqrt{\mathcal{G}(x)} dV.$$

Remark 21.4. The notation dA or $d\sigma$ is called the *surface area element*. We don't really know what it is equal to, but we know that after "pulling it back" to Ω it must be $\sqrt{\det(D\varphi^*D\varphi)}dV$ or $\sqrt{\mathcal{G}(x)}dV$ or $\|\varphi_u \times \varphi_v\|dudv$ (all of these expressions are equal). The term dV is a surface area element in itself, and we actually know its value, it is equal to $dudv$. We use surface area elements like dV whenever it is clear that there should be $dudv$ or $dx dy$ or any other standard "area form" there, and similarly we use dA to mean the same thing in the surface itself, which although we still don't know exactly what it is, we know that it must be $\sqrt{\mathcal{G}(x)}$ times dV after we pull it back to Ω . We will discuss the topic of surface area elements in more detail after learning about differential forms.

Remark 21.5. Notice the similarity of Definition 17.7 and Definition 21.2.

Remark 21.6. The surface integral of a scalar field is sometimes called a type one surface integral. That is because the result of the integral does not depend on the orientation of the surface.

Definition 21.3. We can generalize the definition of a surface integral of a scalar field to higher dimensions. Let $\Omega \subset \mathbb{R}^m$ be

Example 21.2 (Mass of a torus). Let

$$S = T^2 = \varphi([0, 2\pi] \times [0, 2\pi])$$

where $\varphi(u, v)$ is given by

$$\varphi(u, v) = \begin{pmatrix} (a + b \cos u) \cos v \\ b \sin u \\ (a + b \cos u) \sin v \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Let the mass density of the torus be

$$f(x, y, z) = \frac{1}{\sqrt{x^2 + y^2}} = \rho(x, y, z).$$

The total mass of the torus is then given by

$$\begin{aligned} \int_S f dA &= \int_{[0, 2\pi] \times [0, 2\pi]} f \circ \varphi(u, v) \|\varphi_u \times \varphi_v\| dudv \\ &= \int_0^{2\pi} \int_0^{2\pi} \frac{1}{a + b \cos u} b(a + b \cos u) dudv = \dots = 4b\pi^2. \end{aligned}$$

Definition 21.4 (Change in parameterization). Let $g(U) = S$ be a parameterization of an m -dimensional surface. A change of in parameterization on S is a map $\varphi \in C^1(V, \mathbb{R}^m)$ for $V \subset \mathbb{R}^m$ which is injective and $\varphi(V) = U$, such that $\text{rank}(D\varphi(x)) = m$ for all $x \in V$. We say that φ *preserves orientation*, if $\det(D\varphi(x)) > 0$ for all $x \in V$.

Remark 21.7. A change in parameterization defined a new m -parametric surface $\tilde{g} \in C^1(V, \mathbb{R}^n)$ with $\tilde{g}(x) = g(\varphi(x))$.

Proposition 21.1. Any surface integral $\int_S f dA$ is independent of the parameterization of S . More formally, if $\tilde{\Omega} \subset V \subset \mathbb{R}^k$ is a Jordan measurable set and $\Omega = \varphi(\tilde{\Omega}) \subset U \subset \mathbb{R}^k$ is a k -parametric surface, then for any two paratemetrizations g and $\tilde{g} := g \circ \varphi$ of $S = \tilde{g}(\tilde{\Omega}) = g(\Omega)$ we have $\text{Vol}_k(g(\Omega)) = \text{Vol}_k(\tilde{g}(\tilde{\Omega}))$.

Proof. By the chain rule we have that $D\tilde{g} = Dg \circ D\varphi$ so,

$$\begin{aligned}\tilde{\mathcal{G}}(x) &:= \det(D\tilde{g}^* D\tilde{g}) = \det(D\varphi^T Dg^T Dg D\varphi) = \det(Dg^T Dg) (\det(D\varphi))^2 \\ &= \mathcal{G}(\varphi(x)) (\det(D\varphi))^2.\end{aligned}$$

It is now left to use the change of variables theorem with $y = \varphi(x)$ and we get

$$\begin{aligned}\text{Vol}_k(g(\Omega)) &= \int_{\varphi(\tilde{\Omega})} \sqrt{\mathcal{G}(y)} |d^k y| = \int_{\tilde{\Omega}} \sqrt{\mathcal{G}(\varphi(x))} |\det(D\varphi)| |d^k x| \\ &= \int_{\tilde{\Omega}} \sqrt{\tilde{\mathcal{G}}(x)} |d^k x| = \text{Vol}_k(\tilde{g}(\tilde{\Omega})).\end{aligned}$$

This equality completes the proof. □

22 Differential forms

Remark 22.1. In this section we will introduce what I consider the most important object in this course, the differential form. Every integral we have seen so far can be described cleanly using differential forms.

We have seen in Theorem 21.1 that type one scalar/surface integrals are independent of the parameterization (in fact you can remember type one integrals as “integrals independent of orientation”). In this sense, they are different than the standard Riemann integrals we saw in analysis 2 which depend on the orientation of the interval:

$$\int_a^b f(x)dx = - \int_b^a f(x)dx.$$

We want to generalize these integrals to higher dimensions while preserving this property. We have already seen these integrals before.

22.1 Type two line integrals

Definition 22.1 (Type two line integral). Let $U \subset \mathbb{R}^n$ be open, let $f_1, \dots, f_n \in C^1(U, \mathbb{R})$, and let $\gamma \in C^1([a, b], \mathbb{R}^n)$ be a regular curve with $\gamma([a, b]) \subset U$. We then define the type two line integral of $\sum_{i=1}^n f_i(\mathbf{x})dx^i$ by

$$\int_{\gamma} \sum_{i=1}^n f_i(\mathbf{x})dx^i := \int_a^b \left(\sum_{i=1}^n f_i(\gamma(t)) \cdot \frac{d\gamma^i}{dt}(t) \right) dt.$$

Remark 22.2. The expression $\omega(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x})dx^i$ is called a differential 1-form.

Example 22.1. Let $w(x, y) = -ydx + xdy$, and let $\gamma(t) := (x(t), y(t)) = (R \cos t, R \sin t)$ for $t \in [0, 2\pi]$ be the parameterization of a circle of radius R . Then by definition we get

$$\begin{aligned} \int_{\gamma} \omega &= \int_{\gamma} -ydx + xdy = \int_0^{2\pi} \left(-y(t) \cdot \frac{dx}{dt}(t) + x(t) \cdot \frac{dy}{dt}(t) \right) dt \\ &= \int_0^{2\pi} ((-R \sin t)(-R \sin t) + (R \cos t)(R \cos t)) dt = 2\pi R^2. \end{aligned}$$

Example 22.2. Let $\omega(x, y) = xdx + e^y dy$, and let $\gamma(t) := (x(t), y(t))$ for $t \in [a, b]$ parameterize some arbitrary simple smooth closed curve. Then

$$\int_{\gamma} \omega = \int_{\gamma} xdx + e^y dy = \int_a^b \left(x(t) \cdot \frac{dx}{dt}(t) + e^{y(t)} \cdot \frac{dy}{dt}(t) \right) dt = \left(\frac{x^2(t)}{2} + e^{y(t)} \right) \Big|_a^b = 0.$$

This is not a coincidence. We will talk more about integrating differential forms over closed curves later.

Definition 22.2 (Sign). The sign of a change in parameterization $\varphi: [a, b] \rightarrow [\alpha, \beta]$ is defined to be the following function:

$$\text{sgn}(\varphi) := \begin{cases} 1, & \varphi \text{ preserves orientation,} \\ -1, & \varphi \text{ preserves does not preserve orientation.} \end{cases}$$

In this case, the orientation of an interval $I = [a, b]$ is defined by the sign of its boundary $\partial I = b - a$. We will talk more about orientation later.

Proposition 22.1. *The type two line integral is not dependent on the parameterization of the curve, only on its orientation. In particular, if $\varphi: [a, b] \rightarrow [\alpha, \beta]$ is a change in parameterization of a curve parameterized by $\gamma: [a, b] \rightarrow C$ so that $\tilde{\gamma} = \gamma \circ \varphi: [\alpha, \beta] \rightarrow C$, then $\int_{\tilde{\gamma}} \omega = \text{sgn}(\varphi) \int_{\gamma} \omega$.*

Proof. By the chain rule:

$$\begin{aligned} \int_{\tilde{\gamma}} \omega &= \int_{\alpha}^{\beta} \left(\sum_{i=1}^n f_i(\gamma(\varphi(\tau))) \cdot \frac{d\gamma^i}{dt}(\varphi(\tau)) \frac{d\varphi}{d\tau}(\tau) \right) d\tau \\ &= \int_{\varphi(\alpha)}^{\varphi(\beta)} \left(\sum_{i=1}^n f_i(\gamma(t)) \cdot \frac{d\gamma^i}{dt}(t) \right) dt = \text{sgn}(\varphi) \int_{\gamma} \omega. \end{aligned}$$

□

Remark 22.3. We denote a parameterization γ with the opposite orientation by $-\gamma$ and similarly the corresponding curve C by $-C$.

22.2 1-differential forms

In this subsection the goal is to understand the expression $\omega(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x}) dx^i$.

Let $x_0 \in U \subset \mathbb{R}^n$, and let $f \in C^1(U, \mathbb{R})$. Recall that $df(x_0): T_{x_0}\mathbb{R}^n \rightarrow \mathbb{R}$ given by $v \mapsto df(x_0)(v)$ is a linear map.

Under the standard coordinate system $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for each $1 \leq i \leq n$ we define the coordinate function $x^i: \mathbb{R}^n \rightarrow \mathbb{R}$ by $\mathbf{x} \mapsto x^i$. Thus, in standard coordinates the representing matrix of $dx^i(\mathbf{x}_0)$ is the matrix

$$Dx^i(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial x^i}{\partial x^1} & \frac{\partial x^i}{\partial x^2} & \cdots & \frac{\partial x^i}{\partial x^n} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}.$$

For any $\mathbf{v} \in T_{\mathbf{x}_0}\mathbb{R}^n$ we have $dx^i(\mathbf{x}_0)(\mathbf{v}) = v^i$. In particular we have that $dx^i(\mathbf{x}_0)(\mathbf{e}_j) = \delta_j^i$ where $\delta_j^i = 1$ if and only if $i = j$ and $\delta_j^i = 0$ otherwise. Therefore, $\{dx^1, \dots, dx^n\}$ is an orthonormal basis of the dual space $(T_{\mathbf{x}_0}\mathbb{R}^n)^*$ which is dual to the basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$.² Under the dual basis, since $df(\mathbf{x}_0) \in (T_{\mathbf{x}_0}\mathbb{R}^n)^*$ (formally $df(\mathbf{x}_0)$ is defined on $\mathbb{R}^n$ but since $T_{\mathbf{x}_0}\mathbb{R}^n \cong \mathbb{R}^n$ we can identify the same spaces as the same space) we have that

$$df(\mathbf{x}_0) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(\mathbf{x}_0) dx^i.$$

More generally, any linear combination $\sum_{i=1}^n \lambda_i dx^i$ is in $(T_{\mathbf{x}_0}\mathbb{R}^n)^*$ and maps $\mathbf{v} \in T_{\mathbf{x}_0}\mathbb{R}^n$ to the scalar $\sum_{i=1}^n \lambda_i v^i \in \mathbb{R}$.

Definition 22.3 (One-form). Let $U \subset \mathbb{R}^n$. A one-form (or, usually a 1-form) ω is a C^1 function $\omega: U \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that for all $\mathbf{x}_0 \in U$ we have $\omega(\mathbf{x}_0) \in (T_{\mathbf{x}_0}\mathbb{R}^n)^*$.

Remark 22.4. Since $\omega \in C^1$, and since $\omega(\mathbf{x}) \in (T_{\mathbf{x}_0}\mathbb{R}^n)^*$, we can always write $\omega(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x}) dx^i$ for some $f_1, \dots, f_n \in C^1(U, \mathbb{R})$.

Remark 22.5. We can also consider smooth 1-forms, which are the same as C^1 1-forms, except we require them to be C^∞ maps.

Remark 22.6. The space of all 1-forms in U is denoted by $\Omega^1(U)$.

²Recall that the dual space V^* of a vector space V over $\mathbb{F}$ is the space of linear functionals $V \rightarrow \mathbb{F}$. Given a basis $\{e_1, \dots, e_n\}$ of V , there exists a dual basis $\{\epsilon^1, \dots, \epsilon^n\}$ for V^* defined by $\epsilon^i(e_j) = \delta_j^i$.

Example 22.3. We can define a 1-form using constant functions $f_i \equiv \lambda_i$ for all $1 \leq i \leq n$, so that for any $\mathbf{v} \in T_{\mathbf{x}_0}$ we have

$$\omega(\mathbf{v}) := \omega(\mathbf{x}_0)(\mathbf{v}) = \left(\sum_{i=1}^n f_i(\mathbf{x}_0) dx^i \right) (\mathbf{v}) = \left(\sum_{i=1}^n \lambda_i dx^i \right) (\mathbf{v}) = \sum_{i=1}^n \lambda_i v^i.$$

Notice we can denote $\omega(\mathbf{v}) := \omega(\mathbf{x}_0)(\mathbf{v})$ since the functions $\{f_i\}_{i=1}^n$ are constant.

Remark 22.7. In general $\omega = \sum_{i=1}^n f_i dx^i$ is a 1-form, and is mapping from $U \times \mathbb{R}^n \rightarrow \mathbb{R}$, the expression $\omega(\mathbf{x}_0) = \sum_{i=1}^n f_i(\mathbf{x}_0) dx^i$ is the 1-form calculated at some point $\mathbf{x}_0 \in U$. It is a linear functional from $\mathbb{R}^n \cong T_{\mathbf{x}_0} \mathbb{R}^n$ to $\mathbb{R}$, therefore it is an element of the dual space $(T_{\mathbf{x}_0} \mathbb{R}^n)^*$. The expression $\omega(\mathbf{v})$ for some specific $v \in V$ is generally a function $\omega(\mathbf{x})(\mathbf{v}) = \left(\sum_{i=1}^n f_i(\mathbf{x}) dx^i \right) (\mathbf{v}) = \sum_{i=1}^n f_i(\mathbf{x}) v^i$ in $\mathbf{x}$ from U to $\mathbb{R}$. In Example 22.3 since the functions f_i are constant, we get that the function $\omega(\mathbf{v})$ is constant so we can consider it as a scalar (which is equal to its value $\omega(\mathbf{x}_0)(\mathbf{v})$).

Example 22.4. Let r^2 be the function $r^2(x, y, z) = x^2 + y^2 + z^2$. Then $df(x_0, y_0, z_0) \in (T_{(x_0, y_0, z_0)} \mathbb{R}^3)^*$. We get that $dr^2(x_0, y_0, z_0) = 2x_0 dx + 2y_0 dy + 2z_0 dz$. For any $(a, b, c) \in \mathbb{R}^3$ we get $dr^2(x_0, y_0, z_0)(a, b, c) = 2x_0 a + 2y_0 b + 2z_0 c \in \mathbb{R}$.

Example 22.5. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and let $\omega = df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i$. Then for $\mathbf{v} \in T_{\mathbf{x}_0} \mathbb{R}^n$ we get that

$$df(\mathbf{x}_0)(\mathbf{v}) = \left(\sum_{i=1}^n \frac{\partial f}{\partial x^i}(\mathbf{x}_0) dx^i \right) (\mathbf{v}) = \sum_{i=1}^n \frac{\partial f}{\partial x^i}(\mathbf{x}_0) v^i = \langle \nabla f(\mathbf{x}_0), \mathbf{v} \rangle,$$

which is exactly what we expected to get.

Remark 22.8. The expression df is also called the “exterior derivative” of f . We already know that df is a 1-form, and later we will see that f is a 0-form, so in this case d took an n -form and returned an $(n+1)$ -form.

Definition 22.4 (Type two line integral for 1-forms). Let $\gamma \in C^1([a, b], \mathbb{R}^n)$ be a simple smooth curve and let $\omega = \sum_{i=1}^n f_i dx^i \in \Omega^1(U)$ be a 1-form where $U \subset \mathbb{R}^n$ so that $\gamma([a, b]) \subset U$. The type two line integral for ω on γ is defined by

$$\int_{\gamma} \omega := \int_a^b \omega(\gamma(t))(\gamma'(t)) dt = \int_a^b \sum_{i=1}^n f_i(\gamma(t)) dx^i(\gamma'(t)) dt = \int_a^b \left(\sum_{i=1}^n f_i(\gamma(t)) \cdot \frac{d\gamma^i}{dt}(t) \right) dt.$$

Example 22.6. Let $\gamma = (x(t), y(t), z(t)) \in C^1([a, b], \mathbb{R}^3)$ be a simple smooth curve, and let $\omega = Pdx + Qdy + Rdz$ where $P, Q, R \in C^1(\mathbb{R}^3, \mathbb{R})$ be a 1-form. Then the integral of ω over γ is

$$\begin{aligned} \int_{\gamma} \omega &= \int_a^b \omega(\gamma(t))(\gamma'(t)) dt = \int_a^b (P(\gamma(t))x'(t) + Q(\gamma(t))y'(t) + R(\gamma(t))z'(t)) dt \\ &= \int_a^b \langle F, \gamma'(t) \rangle dt, \end{aligned}$$

where $F = (P, Q, R)$ is the vector field corresponding to ω .

Exercise 22.1. Let $\omega = \frac{-ydy+xdy}{x^2+y^2} \in \Omega^1(\mathbb{R}^2 \setminus \{0\})$, and let $\gamma(t) = (x(t), y(t)) = (R \cos t, R \sin t)$ where $t \in [0, 2\pi]$ and $R > 0$. Find $\int_{\gamma} \omega$.

Solution. By definition we have that

$$\begin{aligned}\int_{\gamma} \omega &= \int_0^{2\pi} \left(\frac{-y(t) \cdot x'(t) + x(t) \cdot y'(t)}{x^2(t) + y^2(t)} \right) dt \\ &= \int_0^{2\pi} \left(\frac{(-R \sin t)(-R \sin t) + (R \cos t)(R \cos t)}{R^2} \right) dt = 2\pi.\end{aligned}$$

Proposition 22.2 (Conservative vector field). *Let $U \subset \mathbb{R}^n$ be an open set, and let $\omega = df$ so that $f \in C^2(U, \mathbb{R})$. Then $\int_{\gamma} \omega = f(\gamma(b)) - f(\gamma(a))$. In particular if γ is closed ($\gamma(a) = \gamma(b)$) then $\int_{\gamma} \omega = 0$.*

Proof.

$$\begin{aligned}\int_{\gamma} \omega &= \int_a^b df(\gamma(t))(\gamma'(t))dt = \int_a^b \left(\sum_{i=1}^n \frac{\partial f}{\partial \gamma^i}(\gamma(t)) \cdot \frac{dx^i}{dt}(t) \right) dt \\ &= \int_a^b \frac{d}{dt}(f \circ \gamma)(t)dt = f(\gamma(b)) - f(\gamma(a)).\end{aligned}$$

□

Remark 22.9 (Exact form). In physics the function f is called the *potential function* and $\omega = df$ is called the *conservative field*. In mathematical terms $\omega = df$ is called an *exact form*. We will discuss these terms in more detail later.

Example 22.7. Let $\omega = (ye^{xy} + \cos x)dx + (xe^{xy} + \frac{2y}{1+y^2})dy \in \Omega^1(\mathbb{R}^2)$. Then for any closed smooth regular curve $\gamma: [a, b] \rightarrow \mathbb{R}^2$ we have $\int_{\gamma} \omega = 0$. That is because the function $f(x, y) = e^{xy} + \sin x + \ln(1 + y^2) + c$ is a “potential function” for ω , so by Theorem 22.2 we immediately have $\int_{\gamma} \omega = 0$.

Remark 22.10 (Method to find the potential function). We will focus on trying to find the potential function of ω from Example 22.7. If we know that a potential function f exists, then $f + c$ is a potential for any constant $c \in \mathbb{R}$. Therefore we may assume that $f(0, 0) = 0$. We can now define $f(x_0, y_0)$ simply by $f(x_0, y_0) := \int_{\gamma} \omega$ where γ is any path from $(0, 0)$ to (x_0, y_0) for example the concatenation $\gamma = a + b$ where $a(t) = (t, 0)$ for $t \in [0, x_0]$ and $b(s) = (x_0, s)$ for $s \in [0, y_0]$. Then we have

$$f(x_0, y_0) = \int_a \omega + \int_b \omega = \int_0^{x_0} \cos t dt + \int_0^{y_0} \left(x_0 e^{x_0 s} + \frac{2s}{1+s^2} \right) ds = \sin(x_0) + e^{x_0 y_0} + \ln(1+y_0^2) - 1.$$

This implies that $f(x, y) = \sin(x) + e^{xy} + \ln(1 + y^2) + c$ are all potential functions for ω .

Remark 22.11. Theorem 22.2 and Remark 22.10 show that a vector field has a potential function if and only if every curve integral on its corresponding 1-form only depends on the boundary of the curve, which is why some authors take this as a definition for a conservative vector field instead.

22.3 k -forms and the wedge product

The goal for the rest of this section is to generalize integration on curves $\int_{\gamma} \omega$ to integration on m -parametric surfaces. We will start by generalizing 1-forms to k -forms.

Definition 22.5 (Multilinear antisymmetric function). A function $L: (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ is called multilinear and antisymmetrical if it is

- (1) Linear in each component.
- (2) Changes sign when we interchange between any two variables:

$$L(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k) = -L(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_k).$$

Remark 22.12. The vector space of all multilinear antisymmetric functions $L: (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ is denoted by $\bigwedge^k(\mathbb{R}^n)^*$.

Definition 22.6 (k -form). Let $U \subset \mathbb{R}^n$ and let $k = 0, 1, 2, \dots$. A function $\omega \in C^1(U \times (\mathbb{R}^n)^k, \mathbb{R})$ is called a k -differential form if for all $\mathbf{x}_0 \in U$ the function $\omega(\mathbf{x}_0): (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ is multilinear and antisymmetric ($\omega(\mathbf{x}_0) \in \bigwedge^k(\mathbb{R}^n)^*$).

Remark 22.13. The space of all k -forms is denoted by $\Omega^k(U)$. In particular we have $\Omega^0(U) = C^1(U)$.

Remark 22.14 (Dimension of $\bigwedge^k(\mathbb{R}^n)^*$). We know that given a standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ on $\mathbb{R}^n$ each multilinear function $L: (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ is determined uniquely by its values on the basis vectors in each coordinate $L(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k})$ for $1 \leq i_j \leq n$ for $1 \leq j \leq k$ (n^k elements). If L is also antisymmetric, it is determined by the values of $L(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k})$ for $1 \leq i_j \leq n$ for $1 \leq j \leq k$ and $i_1 < i_2 < \dots < i_k$. Therefore the dimension of $\bigwedge^k(\mathbb{R}^n)^*$ is $\binom{n}{k}$.

Example 22.8. Consider the following cases:

- ($k = 1$) since $\bigwedge^1(\mathbb{R}^n)^* = (\mathbb{R}^n)^*$ we get $\dim(\bigwedge^1(\mathbb{R}^n)^*) = n$, and indeed $\{dx^i\}_{i=1}^n$ is the standard basis where $dx^i(\mathbf{v}) = v^i$ for $1 \leq i \leq n$;
- ($k = n$) we get $\dim(\bigwedge^n(\mathbb{R}^n)^*) = 1$, and since $\det \in \bigwedge^n(\mathbb{R}^n)^*$ we then must have that every function in $\bigwedge^n(\mathbb{R}^n)^*$ is a multiple of $\det$ by a constant so that $\omega(\mathbf{x}_0)(\mathbf{v}_1, \dots, \mathbf{v}_n) = c \det(\mathbf{v}_1, \dots, \mathbf{v}_n)$;
- ($k > n$) in this case $\dim(\bigwedge^k(\mathbb{R}^n)^*) = 0$.

Example 22.9. Let $\omega_{12} \in \bigwedge^2(\mathbb{R}^n)^*$ be the function

$$\omega_{12}(\mathbf{v}_1, \mathbf{v}_2) := \det \begin{pmatrix} v_1^1 & v_2^1 \\ v_1^2 & v_2^2 \end{pmatrix} = \text{Vol} \left(P \left(\begin{pmatrix} v_1^1 \\ v_1^2 \end{pmatrix}, \begin{pmatrix} v_2^1 \\ v_2^2 \end{pmatrix} \right) \right).$$

In other words, $\omega_{12}(\mathbf{v}_1, \mathbf{v}_2)$ is the signed (i.e. it can be negative depending on the orientation) of the parallelogram P generated by the projection of $\mathbf{v}_1$ and $\mathbf{v}_2$ onto the plane spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$. Similarly we can define $\omega_{ij} \in \bigwedge^2(\mathbb{R}^n)^*$ by

$$\omega_{ij}(\mathbf{v}_1, \mathbf{v}_2) := \det \begin{pmatrix} v_1^i & v_2^i \\ v_1^j & v_2^j \end{pmatrix} = \text{Vol} \left(P \left(\begin{pmatrix} v_1^i \\ v_1^j \end{pmatrix}, \begin{pmatrix} v_2^i \\ v_2^j \end{pmatrix} \right) \right)$$

for all $1 \leq i \leq j \leq n$. For example if $n = 3$ and $\mathbf{v}_1 = (1, 2, 3)$ and $\mathbf{v}_2 = (4, 5, 6)$ then we have

$$\begin{aligned} \omega_{12}(\mathbf{v}_1, \mathbf{v}_2) &= \det \begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix} = -3 \\ \omega_{13}(\mathbf{v}_1, \mathbf{v}_2) &= \det \begin{pmatrix} 1 & 4 \\ 3 & 6 \end{pmatrix} = -6 \\ \omega_{23}(\mathbf{v}_1, \mathbf{v}_2) &= \det \begin{pmatrix} 2 & 5 \\ 3 & 6 \end{pmatrix} = -3 \end{aligned}$$

and from Remark 22.14 it is easy to see that ω_{ij} form basis for $\bigwedge^2(\mathbb{R}^n)^*$.

Definition 22.7 (Wedge product). Let $dx^1, \dots, dx^n$ be the standard basis of $(T_{\mathbf{x}_0}\mathbb{R}^n)^*$. Then we define

$$dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_k}(\mathbf{v}_1, \dots, \mathbf{v}_k) := \det \begin{pmatrix} v_1^{i_1} & \dots & v_k^{i_1} \\ \vdots & \ddots & \vdots \\ v_1^{i_k} & \dots & v_k^{i_k} \end{pmatrix}.$$

In other words, the wedge product is the signed volume of the parallelogram generated by the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in T_{\mathbf{x}_0}\mathbb{R}^n$ projected onto the space spanned by $\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k}$. Moreover, if f is a 0-form we define $f \wedge dx^i = f dx^i$.

Remark 22.15. It is easy to see that for different $1 \leq i_1 < \dots < i_k \leq n$ the wedge products $dx^{i_1} \wedge \dots \wedge dx^{i_k}$ are linearly independent and thus from a cardinality argument they form a basis for $\bigwedge^k(\mathbb{R}^n)^*$.

Remark 22.16. The wedge product is indeed a product $\wedge: \bigwedge^k(\mathbb{R}^n)^* \times \bigwedge^l(\mathbb{R}^n)^* \rightarrow \bigwedge^{k+l}(\mathbb{R}^n)^*$ which is defined on the basis vectors by

$$(dx^{i_1} \wedge \dots \wedge dx^{i_k}) \wedge (dx^{j_1} \wedge \dots \wedge dx^{j_l}) := dx^{i_1} \wedge \dots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \dots \wedge dx^{j_l},$$

and is expanded linearly.

Proposition 22.3 (Properties of the wedge product). *The wedge product has the following properties:*

- (1) $\omega \wedge (\eta + \nu) = \omega \wedge \eta + \omega \wedge \nu$ (distributivity).
- (2) $\omega \wedge (\eta \wedge \nu) = (\omega \wedge \eta) \wedge \nu$ (associativity).
- (3) $\omega \wedge \nu = (-1)^{kl} \nu \wedge \omega$ (supercommutativity).

Remark 22.17. This means that the wedge product is distributive, associative, and property (3) is called supercommutativity, so it is also supercommutative.

Example 22.10. Let $\omega = a_1 dx + b_1 dy \in \bigwedge^1(\mathbb{R}^2)^*$ and $\nu = a_2 dx + b_2 dy \in \bigwedge^1(\mathbb{R}^2)^*$ (recall that these are like 1-forms computed at a point $\mathbf{x}_0 \in U$). Then, by distributivity of the wedge product:

$$\omega \wedge \nu = (a_1 dx + b_1 dy) \wedge (a_2 dx + b_2 dy) = a_1 a_2 dx \wedge dx + a_1 b_2 dx \wedge dy + b_1 a_2 dy \wedge dx + b_1 b_2 dy \wedge dy.$$

Since $dx \in \bigwedge^1(\mathbb{R}^1)$ by supercommutativity we get $dx \wedge dx = (-1)^1 dx \wedge dx$ so $dx \wedge dx = 0$ and similarly $dy \wedge dy = 0$ so we can write

$$\omega \wedge \nu = a_1 b_2 dx \wedge dy + b_1 a_2 dy \wedge dx.$$

Finally, we can use distributivity and supercommutativity again to get

$$\omega \wedge \nu = (a_1 b_2 - b_1 a_2) dx \wedge dy.$$

The coefficient looks like the determinant of a matrix, and the next exercise shows that this is not a coincidence.

Exercise 22.2. Let $\omega^i = a_1^i dx^1 + a_2^i dx^2 + \dots + a_n^i dx^n \in \bigwedge^1(\mathbb{R}^n)^*$ for all $1 \leq i \leq n$. Show that

$$\omega^1 \wedge \omega^2 \wedge \dots \wedge \omega^n = \det \begin{pmatrix} a_1^1 & \dots & a_1^n \\ \vdots & \ddots & \vdots \\ a_n^1 & \dots & a_n^n \end{pmatrix} dx^1 \wedge dx^2 \wedge \dots \wedge dx^n.$$

Explain what this means geometrically.

Corollary 22.4. Every k -form ω in $U \subset \mathbb{R}^n$ can be written as

$$\omega(\mathbf{x}) = \sum_{1 \leq i_1 < \dots < i_k \leq n} f_{i_1 \dots i_k}(\mathbf{x}) dx^{i_1} \wedge \dots \wedge dx^{i_k}$$

where $f_{i_1 \dots i_k}(\mathbf{x}) \in C^1(U, \mathbb{R})$.

Example 22.11. Let $\omega = P(x, y, z)dy \wedge dz + Q(x, y, z)dz \wedge dx + R(x, y, z)dx \wedge dy \in \Omega^2(\mathbb{R}^3)$. Then

$$\begin{aligned} \omega(\mathbf{p})(\mathbf{v}_1, \mathbf{v}_2) &= P \cdot \det \begin{pmatrix} v_1^2 & v_2^2 \\ v_1^3 & v_2^3 \end{pmatrix} + Q \cdot \det \begin{pmatrix} v_1^3 & v_2^3 \\ v_1^1 & v_2^1 \end{pmatrix} + R \cdot \det \begin{pmatrix} v_1^1 & v_2^1 \\ v_1^2 & v_2^2 \end{pmatrix} = \det \begin{pmatrix} P & v_1^1 & v_2^1 \\ Q & v_1^2 & v_2^2 \\ R & v_1^3 & v_2^3 \end{pmatrix} \\ &= \langle \mathbf{F}(\mathbf{p}), \mathbf{v}_1 \times \mathbf{v}_2 \rangle, \end{aligned}$$

where $\mathbf{p} \in \mathbb{R}^3$, $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathbf{p}}\mathbb{R}^3$ and $\mathbf{F}(\mathbf{p}) = (P(\mathbf{p}), Q(\mathbf{p}), R(\mathbf{p}))$.

Remark 22.18. Note that when working with 2-forms on $\mathbb{R}^3$ the most convenient way to express them is $\omega = Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy$ and not $\omega = Pdy \wedge dx + Qdx \wedge dz + Rdy \wedge dz$ for example, in order to maintain positive orientation.

22.4 Pullback of forms

Definition 22.8 (Pullback). Let $g \in C^1(U, \mathbb{R}^n)$ for $U \subset \mathbb{R}^m$ and let $\omega \in \Omega^k(g(U))$. We define $g^*\omega \in \Omega^k(U)$ by

$$g^*\omega(\mathbf{x}_0)(\mathbf{v}_1, \dots, \mathbf{v}_k) := \omega(g(\mathbf{x}_0))(dg(\mathbf{x}_0)(\mathbf{v}_1), \dots, dg(\mathbf{x}_0)(\mathbf{v}_k)).$$

Remark 22.19. The pullback operation is linear and is also compatible with multiplication by C^1 functions (this is actually a special case of distributivity over the wedge product). For example for $\omega = f_1 dx + f_2 dy$ we have $g^*(\omega) = g^*(f_1 dx + f_2 dy) = g^*(f_1)g^*(dx) + g^*(f_2)g^*(dy)$. The expression $g^*(f_1)$ is well defined since f_1 is a 0-form.

Remark 22.20 (Deriving pullback). The easiest way to derive the pullback is by understanding its connection to the change of variables theorem. We will talk about this more later, but even now it is possible to understand that in the same way that a parameterization g amplifies the volume of the parallelepiped generated by these k vectors by $Dg(\mathbf{x}_0)$, when we pull back here (change of variables is also pullback) we need to amplify the vectors by $Dg(\mathbf{x}_0)$, and in fact a more natural way to think about it is geometrically, where $\omega(g(\mathbf{x}_0))$ is a function that “eats” a parallelepiped (which is why it is multilinear and antisymmetric), and hence we amplify the parallelepiped by $Dg(\mathbf{x}_0)$ (except we use $dg(\mathbf{x}_0)$ because we are working without coordinates, which means without choosing a basis).

Example 22.12 (Pullback of a 0-form). Recall that $\Omega^0(g(U)) = C^1(g(U))$. For a function $f(x) \in \Omega^0(g(U))$ and $\mathbf{u} \in U \subset \mathbb{R}^m$ by definition

$$g^*f(\mathbf{u}) := f(g(\mathbf{u}))$$

so a pullback is simply composition.

Example 22.13 (Pullback of a 1-form). If $\omega \in \Omega^1(g(U))$, then it is a linear combination of the basis elements $\{dx^i\}_{i=1}^n$. Since the pullback is linear it suffices to understand what is g^*dx^i . By definition for $\mathbf{u} \in U$ and $\mathbf{v} \in T_{\mathbf{u}}\mathbb{R}^m \cong \mathbb{R}^m$ we have

$$g^*dx^i(\mathbf{u})(\mathbf{v}) := dx^i(g(\mathbf{u}))(dg(\mathbf{u})(\mathbf{v})) = [dg(\mathbf{u})(\mathbf{v})]^i = \sum_{j=1}^m \frac{\partial g^i}{\partial u^j}(\mathbf{u}) \cdot v^j = (dg^i)(\mathbf{u})(\mathbf{v}),$$

where $u_1, \dots, u^m$ are the variables of g , and dg^i is the 1-form that is the exterior derivative of g^i . Therefore we get that $g^*dx^i = dg^i$. For a general 1-form $\omega(\mathbf{x}) := \sum_{i=1}^n f_i(\mathbf{x})dx^i$ we get by linearity:

$$g^*\omega(\mathbf{u}) := \sum_{i=1}^n f_i(g(\mathbf{u}))dg^i = \sum_{i=1}^n f_i(g(\mathbf{u})) \sum_{j=1}^m \frac{\partial g^i}{\partial u^j}(\mathbf{u})du^j.$$

Remark 22.21. Pullback distributes over the wedge product. This means that $g^*(\omega \wedge \nu) = g^*\omega \wedge g^*\nu$. Notice that this also allows us to compute pullbacks of general k -forms.

Remark 22.22. As a corollary of the above computations we get a new chain rule for the pullback operation.

Corollary 22.5 (Chain rule). *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open sets. Next suppose that $f \in C^1(U, \mathbb{R}^m)$ so that $f(U) \subset V$ and $g \in C^1(V, \mathbb{R}^d)$. Then for $\omega \in \Omega^k(g(V))$ we get that $(g \circ f)^*(\omega) = f^*(g^*(\omega))$.*

Example 22.14. Let $\gamma: U \rightarrow \mathbb{R}^n$ where $U = (a, b)$ be a smooth regular curve in $\mathbb{R}^n$ and let $\omega = \sum_{i=1}^n f_i dx^i \in \Omega^1(\gamma(U))$. Then by Example 22.13 the pullback $\gamma^*\omega(t)$ is

$$\gamma^*\omega(t) = \sum_{i=1}^n f_i(\gamma(t)) d\gamma^i = \sum_{i=1}^n f_i(\gamma(t)) \frac{d\gamma^i}{dt}(t)dt \in \Omega^1(U).$$

Example 22.15. Let $g: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined by $g(u, v) = (u \cos v, u \sin v)$. Then

$$\begin{aligned} g^*(x dx + y dy) &= (u \cos v) d(u \cos v) + (u \sin v) d(u \sin v) \\ &= (u \cos v)(\cos v du - u \sin v dv) + (u \sin v)(\sin v du + u \cos v dv) \\ &= u(\sin^2 v + \cos^2 v) du + (-u^2 \cos v \sin v + u^2 \sin v \cos v) dv \\ &= u du, \end{aligned}$$

and similarly

$$\begin{aligned} g^*(-y dx + x dy) &= (-u \sin v) d(u \cos v) + (u \cos v) d(u \sin v) \\ &= (-u \sin v)(\cos v du - u \sin v dv) + (u \cos v)(\sin v du + u \cos v dv) \\ &= (-u \sin v \cos v + u \cos v \sin v) du + (u^2 \sin^2 v + u^2 \cos^2 v) dv \\ &= u^2 dv. \end{aligned}$$

We can also compute

$$\begin{aligned} g^*(dx \wedge dy) &= g^*(dx) \wedge g^*(dy) \\ &= d(u \cos v) \wedge d(u \sin v) \\ &= (\cos v du - u \sin v dv) \wedge (\sin v du + u \cos v dv), \end{aligned}$$

then by Exercise 22.2 we have that

$$\begin{aligned} g^*(dx \wedge dy) &= (\cos v du - u \sin v dv) \wedge (\sin v du + u \cos v dv) \\ &= \det \begin{pmatrix} \cos v & \sin v \\ -u \sin v & u \cos v \end{pmatrix} du \wedge dv \\ &= (u \cos^2 v + u \sin^2 v) du \wedge dv \\ &= u du \wedge dv. \end{aligned}$$

Remark 22.23. If we need to integrate $\omega = x dy + y dx$, we can notice that this is simply $d(xy)$ so $\int_\gamma \omega = \int_\gamma d(xy) = [x(t)y(t)]_a^b$.

Exercise 22.3. Let $g \in C^1(U, \mathbb{R}^n)$ for $U \subset \mathbb{R}^n$, and denote $\mathbf{x} = g(\mathbf{u})$. Show that

$$g^*(dx^1 \wedge \cdots \wedge dx^n) = \det \begin{pmatrix} \frac{\partial g^1}{\partial u^1} & \cdots & \frac{\partial g^1}{\partial u^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g^n}{\partial u^1} & \cdots & \frac{\partial g^n}{\partial u^n} \end{pmatrix} du^1 \wedge \cdots \wedge du^n,$$

where $u_1, \dots, u_n$ are the variables of g . Notice that this is an immediate corollary of a previous exercise. This is a very handy formula when using the change of variables theorem. Important bonus: generalize this and find the correct formula for when $U \subset \mathbb{R}^m$ and $m \neq n$. Note that the bonus can be solved (well not rigorously but you can find the formula) by intuition alone and previous things we saw. Think about the cases $n < m$ and $m < n$ separately.

22.5 Motivation for surface integrals and flux

Remark 22.24. The purpose of this section is to give intuition for surface integrals of m -forms by introducing the flux integral.

Remark 22.25. Recall that our goal is to generalize type two line integrals to k -parameteric surfaces.

Definition 22.9 (Unit normal vector). Let $S = g(\Omega)$ be a 2-parametric surface in $\mathbb{R}^3$ where $g(u, v) = (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in \Omega$. Then the normal vector at a point $g(u, v)$ on the surface is defined as

$$\mathbf{n}(u, v) := \frac{g'_u \times g'_v}{\|g'_u \times g'_v\|} \quad \text{or} \quad \mathbf{n}(u, v) := \frac{g'_v \times g'_u}{\|g'_v \times g'_u\|},$$

where g'_u and g'_v are the partial derivative of g with respect to u and v respectively.

Remark 22.26 (Intuition for the unit normal vector). The intuition behind this definition should be very clear. We take the vector product of the two vectors that span the tangent space $T_{(u,v)}S$, so the result is a vector that is orthogonal to $T_{(u,v)}S$ and then we normalize it by dividing by its norm. Notice that the normal vector is defined in two ways, that is because the normal vector can always have two orientations (usually we say that it points inward or points outward). For example on any point of the 2-sphere we can have a normal vector that points inside the sphere and one that points outside it, and we know how to write each one algebraically using the right hand rule.

Definition 22.10 (Flux). Let $S = g(\Omega)$ be a 2-parametric orientable surface in $\mathbb{R}^3$ where $g(u, v) = (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in \Omega$. We also assume that g is injective, C^1 , and $\text{rank}(dg) = 2$ in an open neighbourhood U of the compact Jordan measurable set $\Omega \subset \mathbb{R}^2$. Let $\mathbf{F} = (P, Q, R)$ be a continuous vector field on S . The flux of F through S is defined by:

$$\begin{aligned} \Phi &= \int_S \langle \mathbf{F}, \mathbf{n}(u, v) \rangle d\sigma \\ &= \int_\Omega \left\langle \mathbf{F}(g(u, v)), \frac{g'_u \times g'_v}{\|g'_u \times g'_v\|} \right\rangle \|g'_u \times g'_v\| du dv \\ &= \int_\Omega \langle \mathbf{F}(g(u, v)), g'_u \times g'_v \rangle |du dv| \\ &= \int_\Omega \left(P(g(u, v)) \cdot \begin{vmatrix} y'_u & y'_v \\ z'_u & z'_v \end{vmatrix} + Q(g(u, v)) \cdot \begin{vmatrix} z'_u & z'_v \\ x'_u & x'_v \end{vmatrix} + R(g(u, v)) \cdot \begin{vmatrix} x'_u & x'_v \\ y'_u & y'_v \end{vmatrix} \right) |du dv|, \end{aligned}$$

or symbolically, $\Phi = \int_S P dy \wedge dz + Q dz \wedge dx + R dx \wedge dy$.

Remark 22.27. Integrals of this kind are called surface integrals of differential 2-forms. Note that $d\sigma$ is a surface area element. We still don't have tools to define it, but we know it should be equal $\|g'_u \times g'_v\| du \wedge dv$ or $\|g'_u \times g'_v\| dudv$ or $\|g'_u \times g'_v\| dV$ (all options are equivalent).

Remark 22.28 (Flux form). When given a 2-form of the form $Pdx \wedge dy + Qdy \wedge dz + Rdz \wedge dx$, the corresponding vector field is not (P, Q, R) , instead we must first bring the 2-form to a flux form, $\tilde{P}dy \wedge dz + \tilde{Q}dz \wedge dx + \tilde{R}dx \wedge dy$ and then the vector field is $(\tilde{P}, \tilde{Q}, \tilde{R})$. The reason is that this is what we want intuitively, but you could also justify this more formally by noticing that this is the only proper way to compute the 2-form (we see this in the following section).

Remark 22.29. Since we have not defined surface integrals of m -forms yet, you may consider it as a given formula at this point. Physically, the flux integral measures how much “fluid” passes through the surface with respect to the orientation of the unit normal vector (this is why it is a type two integral). Intuitively, the vector field measures the velocity of the fluid at each point, so by taking the scalar product of the value of the vector field at the point with the normal vector at that point, we are actually projecting the vector field onto the normal unit vector, and check the “how much fluid” passed through in that direction.

Exercise 22.4. Let $\omega = xzdy \wedge dz + yzdz \wedge dx + dx \wedge dy$ be a 2-form in $\mathbb{R}^3$. Consider the closed upper hemisphere of the unit sphere $S \subset \mathbb{R}^3$, let the normal unit vector point outward, and parameterize S with

$$S = g(D) \quad \text{and} \quad g(x, y) = (x, y, \sqrt{1 - x^2 - y^2}) \quad \text{and} \quad D = \left\{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1 \right\}.$$

Compute the flux $\int_S \omega$.

Solution. We know that the flux formula is given by

$$\int_S \omega = \int_D \left\langle \mathbf{F}(g(x, y)), g'_x \times g'_y \right\rangle |dx dy|.$$

Notice that ω is given in flux form so we can simply write $\mathbf{F}(x, y, z) = (xz, yz, 1)$, and therefore:

$$\mathbf{F}(g(x, y)) = (x\sqrt{1 - x^2 - y^2}, y\sqrt{1 - x^2 - y^2}, 1).$$

We also have that

$$g'_x = \begin{pmatrix} 1 \\ 0 \\ \frac{-x}{\sqrt{1-x^2-y^2}} \end{pmatrix} \quad \text{and} \quad g'_y = \begin{pmatrix} 0 \\ 1 \\ \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix},$$

and so:

$$g'_x \times g'_y = \begin{vmatrix} i & j & k \\ 1 & 0 & \frac{-x}{\sqrt{1-x^2-y^2}} \\ 0 & 1 & \frac{-y}{\sqrt{1-x^2-y^2}} \end{vmatrix} = \begin{pmatrix} \frac{x}{\sqrt{1-x^2-y^2}} \\ \frac{y}{\sqrt{1-x^2-y^2}} \\ 1 \end{pmatrix}.$$

We see that at the point $(0, 0, 1)$ the normal we get is $(0, 0, 1)$ so indeed this order of the cross product is associated with the outward pointing normal. Therefore we can continue,

$$\begin{aligned} \left\langle \mathbf{F}(g(x, y)), g'_x \times g'_y \right\rangle &= \begin{pmatrix} x\sqrt{1-x^2-y^2} & y\sqrt{1-x^2-y^2} & 1 \end{pmatrix} \begin{pmatrix} \frac{x}{\sqrt{1-x^2-y^2}} \\ \frac{y}{\sqrt{1-x^2-y^2}} \\ 1 \end{pmatrix} \\ &= x^2 + y^2 + 1. \end{aligned}$$

Thus,

$$\int_S \omega = \int_D x^2 + y^2 + 1 |dx dy|.$$

Next we want to move to polar coordinates. The parameterization of the unit disc is given by $h(r, \theta) = (r \cos \theta, r \sin \theta)$ for $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Denote $A = [0, 1] \times [0, 2\pi]$. By the change of variables theorem,

$$\int_{h(A)} x^2 + y^2 + 1 |dx dy| = \int_A (r^2 + 1) \cdot |\det Dh(r, \theta)| |dr d\theta|.$$

Compute:

$$|\det Dh(r, \theta)| = \left| \begin{vmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{vmatrix} \right| = |r(\cos^2 \theta + \sin^2 \theta)| = r.$$

We can now complete the exercise since

$$\begin{aligned} \int_S \omega &= \int_A (r^2 + 1) \cdot r |dr d\theta| = \int_0^1 |dr| \int_0^{2\pi} (r^3 + r) |d\theta| = 2\pi \int_0^1 (r^3 + r) |dr| \\ &= 2\pi \left(\frac{1}{4} + \frac{1}{2} \right) = \frac{3\pi}{2}. \end{aligned}$$

22.6 Surface integrals of m -forms

Remark 22.30. We now have the tools to define type two line integrals on m -forms (also known as surface integrals). We will start with an underwhelming definition, and then proceed to the more interesting one.

Definition 22.11 (Surface integral of a simple m -form in $\mathbb{R}^m$ without pullback). Let $U \subset \mathbb{R}^m$ be a Jordan measurable set and let $\omega \in \Omega^m(U)$. Recall that $\dim(\bigwedge^k(\mathbb{R}^n)^*) = 1$ so we can write $\omega = f(\mathbf{u}) du^1 \wedge du^2 \wedge \cdots \wedge du^m$ (it is necessary for the indices to be of this specific increasing order!). We then define the surface integral of ω on U as

$$\int_U \omega := \int_U f dV.$$

Remark 22.31. This definition is essentially saying that when the m -form is in $U \subset \mathbb{R}^m$ and of the form $\omega = f(\mathbf{u}) du^1 \wedge du^2 \wedge \cdots \wedge du^m$ then the integral of ω over U is equal to the standard Riemann integral of f over U . We also know that $dV = du^1 \wedge \cdots \wedge du^m$ and $dV = du^1 du^2 \cdots du^m$ and this definition justifies this.

Definition 22.12 (Surface integral of an m -form on a manifold). Let $U \subset \mathbb{R}^m$ be an open set, let $\Omega \subset U$ be a compact Jordan measurable set, let $g \in C^1(U, \mathbb{R}^n)$ be an injective map such that $\text{rank}(dg(\mathbf{x})) = m$ for all $\mathbf{x} \in U$, and let $\omega \in \Omega^m(W)$ be an m -form on $g(U) \subset W$. Then the surface integral of ω on the m -parametric surface $S = g(\Omega)$ is defined by

$$\int_S \omega = \int_\Omega g^* \omega.$$

Remark 22.32. Note that this is a generalization of the change of variables theorem. Also, in particular, for 1-forms we get the type two integral on curves.

Example 22.16. Let $\gamma: [a, b] \rightarrow \mathbb{R}^n$ be a smooth regular curve in $\mathbb{R}^n$ and let $\omega = \sum_{i=1}^n f_i dx^i \in \Omega^1(\gamma([a, b]))$. Then if we denote $\mathbf{F} = (f_1, \dots, f_n)$ by Example 22.14 we get that

$$\int_\gamma \omega = \int_\gamma \sum_{i=1}^n f_i(\mathbf{x}) dx^i = \int_a^b \sum_{i=1}^n f_i(\gamma(t)) \frac{d\gamma^i}{dt}(t) dt = \int_a^b \langle \mathbf{F}(\gamma(t)), \gamma'(t) \rangle dt.$$

Remark 22.33. If we change the orientation of γ , for example by using the parameterization $t(\tau) = a + b - \tau$ and $\tilde{\gamma}(\tau) = \gamma(t(\tau)) = \gamma(a + b - \tau)$ for $\tau \in [a, b]$ we get

$$\begin{aligned} \int_{\tilde{\gamma}} \omega &= \int_a^b \sum_{i=1}^n f_i(\tilde{\gamma}(\tau)) \frac{d\tilde{\gamma}^i}{d\tau}(\tau) d\tau = \int_a^b \sum_{i=1}^n f_i(\gamma(t(\tau))) \frac{d\gamma^i}{dt}(t(\tau)) \cdot \frac{dt}{d\tau}(\tau) d\tau \\ &= \int_b^a \sum_{i=1}^n f_i(\gamma(t)) \frac{d\gamma^i}{dt}(t) dt = - \int_{\gamma} \omega, \end{aligned}$$

where the second to last transition is an application of the change of variables theorem.

Example 22.17. Let $\gamma: [a, b] \rightarrow \mathbb{R}^2$ be the graph $\gamma(x) = (x, f(x))$ of a function $f \in C^1([a, b] \subset \mathbb{R})$. Then since $\gamma'(x) = (1, f'(x)) \neq 0$, we get that $\gamma(x)$ is a smooth regular curve in $\mathbb{R}^2$. Therefore for $\omega = ydx \in \Omega^1(\mathbb{R}^2)$ we have $\mathbf{F}(x, y) = (y, 0)$ and thus $\mathbf{F}(\gamma(x)) = (f(x), 0)$ and so

$$\int_{\gamma} \omega = \int_{\gamma} ydx = \int_a^b f(x)dx.$$

Example 22.18. Let $\Omega := \{(r, \theta) \in \mathbb{R}^2 \mid 0 \leq r \leq 1 \text{ and } 0 \leq \theta \leq 2\pi\}$ and $g: \Omega \rightarrow \mathbb{R}^3$ be defined by $g(r, \theta) = (r \cos \theta, r \sin \theta, \sqrt{1-r^2})$. Then the parametric surface $S = g(\Omega)$ is the upper hemisphere of the unit sphere in $\mathbb{R}^3$, where (r, θ) are the standard polar coordinates on $\mathbb{R}^2$. For $\omega = z^2 dx \wedge dy$ we get

$$\begin{aligned} g^*\omega(r, \theta) &= g^*(z^2)(r, \theta) \cdot g^*(dx)(r, \theta) \wedge g^*(dy)(r, \theta) \\ &= (1-r^2) \cdot d(r \cos \theta) \wedge d(r \sin \theta) \\ &= (1-r^2) \cdot (\cos \theta dr - r \sin \theta d\theta) \wedge (\sin \theta dr + r \cos \theta d\theta) \\ &= (1-r^2) \cdot \det \begin{pmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{pmatrix} dr \wedge d\theta \\ &= (1-r^2) \cdot r dr \wedge d\theta. \end{aligned}$$

Therefore

$$\int_S \omega = \int_S z^2 dx \wedge dy = \int_{\Omega} (1-r^2) r dr \wedge d\theta = \int_{\Omega} (1-r^2) r |dr d\theta| = \int_0^{2\pi} \int_0^1 (1-r^2) r dr d\theta = \frac{\pi}{2}.$$

Remark 22.34. In the case of $n = m$ Definition 22.11 and Definition 22.12 are equivalent by the change of variables theorem and Exercise 22.3. Indeed for $U \subset \mathbb{R}^m$ and $g \in C^1(U, \mathbb{R}^m)$ which is injective and satisfies $\det(Dg) > 0$ on U , for any m -form $\omega = f(\mathbf{x}) dx^1 \wedge \cdots \wedge dx^m$ on $S = g(\Omega)$ where $\Omega \subset U$ is Jordan measurable, the equality $\int_S \omega = \int_{\Omega} g^*\omega$ follows by Exercise 22.3

$$\begin{aligned} g^*(f(\mathbf{x}) dx^1 \wedge \cdots \wedge dx^m)(\mathbf{u}) &= f(g(\mathbf{u})) \cdot \det(Dg(\mathbf{u})) du^1 \wedge \cdots \wedge du^m \\ &= f(g(\mathbf{u})) \cdot \left| \det(Dg(\mathbf{u})) \right| du^1 \wedge \cdots \wedge du^m, \end{aligned}$$

where the last step follows from $\det(Dg) > 0$.

Proposition 22.6. Let $\varphi: \tilde{\Omega} \rightarrow \Omega$ be a change in parameterization that preserves orientation. Then

$$\int_{\Omega} g^*\omega = \int_{\tilde{\Omega}} (g \circ \varphi)^*\omega.$$

Example 22.19. Let $\tilde{\Omega} := \{(\theta, \varphi) \in \mathbb{R}^2 \mid 0 \leq \theta \leq 2\pi \text{ and } 0 \leq \varphi \leq \frac{\pi}{2}\}$ and $\tilde{g}: \tilde{\Omega} \rightarrow \mathbb{R}^3$ be defined by $\tilde{g}(\theta, \varphi) = (\cos \theta \sin \varphi, \sin \theta \sin \varphi, \cos \varphi)$. The parametric surface $S = g(\tilde{\Omega})$ is still the upper hemisphere of the unit sphere in $\mathbb{R}^3$ and this time for $\omega = z^2 dx \wedge dy$ we get

$$\begin{aligned} \tilde{g}^* \omega(\theta, \varphi) &= \tilde{g}^*(z^2)(\theta, \varphi) \cdot \tilde{g}^*(dx)(\theta, \varphi) \wedge \tilde{g}^*(dy)(\theta, \varphi) \\ &= \cos^2 \varphi \cdot d(\cos \theta \sin \varphi) \wedge d(\sin \theta \sin \varphi) \\ &= \cos^2 \varphi \cdot (-\sin \varphi \sin \theta d\theta + \cos \theta \cos \varphi d\varphi) \wedge (\sin \varphi \cos \theta d\theta + \sin \theta \cos \varphi d\varphi) \\ &= \cos^2 \varphi \cdot \det \begin{pmatrix} -\sin \varphi \sin \theta & \sin \varphi \cos \theta \\ \cos \theta \cos \varphi & \sin \theta \cos \varphi \end{pmatrix} d\theta \wedge d\varphi \\ &= -\cos^3 \varphi \sin \varphi d\theta \wedge d\varphi. \end{aligned}$$

Therefore

$$\begin{aligned} \int_S \omega &= \int_S z^2 dx \wedge dy = \int_{\tilde{\Omega}} -\cos^3 \varphi \sin \varphi d\theta \wedge d\varphi = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} -\cos^3 \varphi \sin \varphi |d\theta d\varphi| \\ &= 2\pi \int_0^{\frac{\pi}{2}} \cos^3 \varphi (-\sin \varphi) d\varphi = 2\pi \left[\frac{\cos^4(\pi/2)}{4} - \frac{\cos^4(0)}{4} \right] = -\frac{\pi}{2}. \end{aligned}$$

Remark 22.35. In Example 22.18 and Example 22.19 we calculated the same surface integral of ω on S , and got both $\pi/2$ and $-\pi/2$. The sign is different because the change in parameterization $(\theta, \varphi) \mapsto (r, \theta)$ with $r = \sin \varphi$ does not preserve the orientation. In this case we can say that the change in parameterization is “orientation-reversing”. We will talk about orientation more in Section 23.4.

Exercise 22.5. Will the results in Example 22.18 and Example 22.19 still have different signs if the calculation was done on the lower hemisphere of the unit sphere in $\mathbb{R}^3$?

Solution. The original change in parameterization of $V := [0, 2\pi] \times [0, \pi/2]$ to $U := [0, 1] \times [0, 2\pi]$ is the map $\varphi \in C^1(V, \mathbb{R}^2)$ given by

$$\varphi(\theta, \varphi) = (\sin \varphi, \theta).$$

Therefore for $(\theta, \varphi) \in V$ we have

$$\det(D\varphi(\theta, \varphi)) = \det \begin{pmatrix} 0 & \cos \varphi \\ 1 & 0 \end{pmatrix} = -\cos \varphi.$$

In this case $0 \leq \varphi \leq \pi/2$ and thus $\det(D\varphi(\theta, \varphi)) < 0$ so φ is orientation-reversing, so now we have shown Remark 22.35 has justification. In the case of the lower hemisphere, in order to have correct parameterizations, we need to flip the sign of the z -coordinate in both g and $\tilde{g}$. Since $\omega = z^2 dx \wedge dy$ we can either manually verify that the values of the integrals do not change, or we can just notice that whether their values change or not, since the domains of the parameterizations are still U and V and since $\varphi: V \rightarrow U$ reverses orientation, the signs of the integrals will be different.

Exercise 22.6. Let $M \subset \mathbb{R}^3$ be the region between the coaxial cylinders $x^2 + y^2 = 4$ and $x^2 + y^2 = 1$ that is bounded by the planes $z = 0$ and $z = 1$. This type of shape is called a cylindrical shell. Let $\omega = z(x^2 + y^2)^2 dx \wedge dy \wedge dz$. Find $\int_M \omega$.

Solution. We will parameterize M by using cylindrical coordinates

$$g(r, \theta, z) = (r \cos(\theta), r \sin(\theta), z).$$

We now have $g(\Omega) = M$ for $\Omega := \{(r, \theta, z) \mid 1 \leq r \leq 2 \text{ and } 0 \leq \theta \leq 2\pi \text{ and } 0 \leq z \leq 1\}$. By Exercise 22.3 we have

$$\begin{aligned} g^*\omega(r, \theta, z) &= g^*(z(x^2 + y^2)^2)(r, \theta, z) \cdot \det \begin{pmatrix} \cos \theta & -r \sin \theta & 0 \\ \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} dr \wedge d\theta \wedge dz \\ &= z(r^2 \cos^2 \theta + r^2 \sin^2 \theta)^2 \cdot r dr \wedge d\theta \wedge dz \\ &= zr^5 dr \wedge d\theta \wedge dz. \end{aligned}$$

As we saw $\det(Dg(r, \theta, z)) = r > 0$ (when $(r, \theta, z) \in \Omega$). Therefore $\int_M \omega = + \int_\Omega g^*\omega$. Finally, we get that

$$\int_M z(x^2 + y^2)^2 dx \wedge dy \wedge dz = \int_\Omega zr^5 dr \wedge d\theta \wedge dz = \int_\Omega zr^5 |dr d\theta dz| = \dots = \frac{63\pi}{6}.$$

23 The exterior derivative and orientation

23.1 The exterior derivative

Remark 23.1. Recall that $\Omega^k(\mathbb{R}^n)$ is a vector space. Also, in this section and onwards, unless stated otherwise, we only consider smooth k -forms so $\Omega^0(U) = C^\infty(U)$ etc.

Theorem 23.1. Let A be an open set in $\mathbb{R}^n$. Then there exists a unique linear transformation

$$d: \Omega^k(A) \rightarrow \Omega^{k+1}(A),$$

defined for $k \geq 0$, such that

(1) If f is a 0-form, then df is the 1-form

$$df(\mathbf{x})(\mathbf{v}) = Df(\mathbf{x}) \cdot \mathbf{v}.$$

(2) If $\omega \in \Omega^k(A)$ and $\eta \in \Omega^l(A)$, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

(3) For every form ω ,

$$d(d\omega) = 0.$$

Remark 23.2. We call d the differential operator, and we call $d\omega$ the differential of ω .

Remark 23.3. Given a multi-index $I = (i_1, \dots, i_n)$ we introduce the following notation:

$$dx_I = dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_n}.$$

Definition 23.1 (Exterior derivative). Let $\omega = \sum_I f_I \wedge dx_I \in \Omega^k(M)$ be a k -form where M is a subset of $\mathbb{R}^n$ (usually a manifold), then the exterior derivative of ω is given by

$$d\omega = \sum_I df_I \wedge dx_I = \sum_I \sum_{j=1}^n \frac{\partial f_I}{\partial x_j} dx_j \wedge dx_{i_1} \wedge \dots \wedge dx_{i_k}.$$

Remark 23.4. We can see that our definition is sensible since by property (2) of Theorem 23.1, for a k -form $\omega = f_I \wedge dx_I \in \Omega^k(M)$ we have

$$d\omega = d(f_I \wedge dx_I) = df_I \wedge dx_I + (-1)^k f_I \wedge d(dx_I) = df_I \wedge dx_I.$$

This implies that $d\omega$ is determined by the value of d on the 0-form f_I . This is obvious, since any $\omega \in \Omega^k(M)$ is determined by the 0-forms that describe it, but it is not obvious that d can be described in such an elegant way, instead of a very complex expression of its 0-forms. In this sense, Definition 23.1 is not exactly a definition, rather it simply follows from the properties we want it to have.

Remark 23.5. Remember that the exterior derivative is a linear transformation between vector spaces.

Remark 23.6. As we mentioned in Theorem 23.1, the exterior derivative satisfies the rule $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$. This rule is called the “graded Leibniz rule”, because it reminds us of the product rule of standard differentiation, but it is called graded because it also involves degree. The word “grade” specifically comes from the fact that $\bigwedge^*(V^*)$ is a graded ring (we will not talk about graded rings in these notes).

Proposition 23.2. Let $g \in \Omega^0(M)$ for some manifold M . Then we have $d(g^n) = ng^{n-1}dg$.

Proof. We will prove this by induction. First, for $n = 2$ we see that

$$d(g^2) = d(g \wedge g) = dg \cdot g + (-1)^0 g \cdot dg = 2g dg.$$

Next assume that the identity holds for $n-1$, that is, $d(g^{n-1}) = (n-1)g^{n-2}dg$, and we will prove it holds for n . We notice that $g^n = g^{n-1} \wedge g$ (since wedge of 0-forms is simply multiplication), and use the Leibniz graded rule to get that

$$d(g^n) = d(g^{n-1} \wedge g) = d(g^{n-1})g + (-1)^0 g^{n-1}dg.$$

By the induction hypothesis we get that

$$d(g^n) = ((n-1)g^{n-2}dg)g + (-1)^0 g^{n-1}dg = ng^{n-1}dg,$$

which completes the proof. $\square$

Exercise 23.1. Let $g \in \Omega^0(U)$ for some connected open set $U \subset \mathbb{R}^2$, and let $\omega = g^{1234}(g'_x dx + g'_y dy) \in \Omega^1(U)$. Show that ω is an exact form. That is, there exists η so that $d\eta = \omega$.

Solution. Note that $dg = (g'_x dx + g'_y dy)$. By Theorem 23.2, we immediately get that for $\eta = \frac{1}{1235}g^{1235}$,

$$d\eta = \frac{1}{1235}d(g^{1235}) = g^{1234}dg = \omega,$$

which completes the proof.

Remark 23.7. For those of you who are already familiar with algebraic topology, and know about cohomology, it may be useful to keep in mind that the exterior derivative is a coboundary map in the cochain complex:

$$0 \xrightarrow{d} \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \xrightarrow{d} \Omega^3(M) \xrightarrow{d} \dots$$

With this cochain complex (the de Rham complex) we compute the famous de Rham cohomology. This is intuition for property (3) in Theorem 23.1. We have already seen intuition for property (1) in Remark 22.8, and property (2) makes sure property (3) is always satisfied when considering wedges of forms:

$$\begin{aligned} d^2(\omega \wedge \eta) &= d(d\omega \wedge \eta + (-1)^k \omega \wedge d\eta) \\ &= d(d\omega \wedge \eta) + (-1)^k d(\omega \wedge d\eta) \\ &= d(d\omega) \wedge \eta + (-1)^{k+1} d\omega \wedge d\eta + (-1)^k d\omega \wedge d\eta + (-1)^{2k} \omega \wedge dd(\eta) \\ &= (-1)^{k+1} d\omega \wedge d\eta + (-1)^k d\omega \wedge d\eta \\ &= 0. \end{aligned}$$

Notice that if both ω and η are 0-forms, then property (2) is simply the product differentiation rule.

Remark 23.8. For those who don't have much experience with topology, remember that in $\mathbb{R}^3$ we have the identity $\nabla \times \nabla f = 0$, when f is a smooth function. This is a special case of the identity $d^2 = 0$.

Remark 23.9. The exterior derivative commutes with the pullback operation.

Example 23.1. Let $\omega = ydx + xdy \in \Omega^1(\mathbb{R}^2)$. Then $d\omega = dy \wedge dx + dx \wedge dy = 0$.

Example 23.2. Let $\omega = -ydx + xdy \in \Omega^1(\mathbb{R}^2)$. Then $d\omega = -dy \wedge dx + dx \wedge dy = 2dx \wedge dy$.

Definition 23.2 (Exact form). Let M be an open set in $\mathbb{R}^n$. A 0-form f on A is said to be exact on A if it is constant on A . A k -form $\omega \in \Omega^k(A)$ with $k > 0$ is said to be exact if there exists a $k-1$ form $\eta \in \Omega^{k-1}(A)$ such that $\omega = d\eta$.

Definition 23.3 (Closed form). Let M be an open set in $\mathbb{R}^n$. A k -form $\omega \in \Omega^k(A)$ with $k \geq 0$ is said to be closed if $d\omega = 0$.

Remark 23.10. The form $\omega = d\left(\arctan \frac{y}{x}\right) = \frac{-ydx+xdy}{x^2+y^2} \in \Omega^1(\mathbb{R}^2 - \{0\})$ is closed. However, it can be shown that ω is not exact on $\mathbb{R}^2 - \{0\}$.

Example 23.3. In differential equations the equation

$$P(x, y)dx + Q(x, y)dy = 0,$$

is said to be exact if there is a function f such that $P = \frac{\partial f}{\partial x}$ and $Q = \frac{\partial f}{\partial y}$. But this simply means that the 1-form $Pdx + Qdy$ is the differential of the 0-form f , so it is exact.

Example 23.4. Recall that a vector field $\mathbf{F} = (P, Q, R)$ is conservative, if and only if it is the gradient of a scalar field f , that is, $\mathbf{F} = \nabla f$. And that is precisely the same as saying that the form $Pdx + Qdy + Rdz$ is the differential of the 0-form f .

Remark 23.11. Every exact form is closed. This follows from property (3) of the exterior derivative in Theorem 23.1.

Definition 23.4 (Star-convex). Let A be an open set in $\mathbb{R}^n$. We say that A is star-convex with respect to the point $p \in A$ if for each $x \in A$, the line segment $[p, x]$ is contained in A .

Remark 23.12 (Notes for the future). Let $\omega \in \Omega^k(A)$ be a closed form. Then if A is star-convex, ω is exact on A . This result is known as the Poincaré lemma. As you could expect from its name, we will not prove it right now, since we don't have the tools yet. The converse to Poincaré's lemma is false. If every closed $\omega \in \Omega^k(A)$ is exact on A , then we say that A is homologically trivial in dimension k .

Exercise 23.2. Let A be an open subset of $\mathbb{R}^n$, and let $\omega \in \Omega^{k-1}(A)$. Given $v_1, \dots, v_k \in \mathbb{R}^n$, define

$$g_j(\mathbf{x}) = \omega(\mathbf{x})(v_1, \dots, \widehat{v}_j, \dots, v_k),$$

where $\widehat{v}_j$ means that the component v_j is omitted. Then

$$d\omega(\mathbf{x})(v_1, \dots, v_k) = \sum_{j=1}^k (-1)^{j-1} Dg_j(\mathbf{x}) \cdot v_j.$$

Remark 23.13 (Topological perspective). When coming from the viewpoint of algebraic topology, it is very exciting to see these results. However, one may find it difficult to find a homology sequence, so that its dual is the de Rham cohomology sequence. In order to see a similar connection, reference the “de Rham theorem” which shows the connection between de Rham cohomology and the homology of chains, using the integral operation, and the Stokes theorem.

23.2 Green's theorem

Definition 23.5 (Natural orientation). Let $\Omega \subset \mathbb{R}^2$ be a Jordan measurable set, such that $\partial\Omega$ is a finite union of piecewise smooth, simple closed curves. The natural induced orientation (or standard induced orientation) of $\partial\Omega$, is the direction that requires Ω to be to the left when following along it.

Remark 23.14. This simply means that Ω is a Jordan measurable set, with boundary that could be either a circle, or if its a finite union of piecewise smooth, simple closed curves, it could be the figure eight space (a space that looks like the digit 8).

Remark 23.15. Notice that a circle in $\mathbb{R}^2$ does not have a natural induced orientation (as a boundary) by itself. If we consider the unit circle as the boundary of the unit disc, its natural orientation would be counterclockwise, but if we had considered it as the inner circle of the boundary of an annulus, its natural orientation would be clockwise. As you may expect, this definition we gave of natural orientation is not the most... “natural” one, and indeed, we will introduce a more general notion of induced orientation in Section 23.5.

Theorem 23.3 (Green’s theorem for a rectangle). *Let $R \subset \mathbb{R}^2$ be a rectangle, and let ω be a 1-form on R . Then*

$$\int_{\partial R} \omega = \int_R d\omega,$$

where ∂R is given the natural orientation.

Proof. Suppose that $R = [a, b] \times [c, d]$, and $\omega = Pdx + Qdy$. Then

$$d\omega = \left(\frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy \right) \wedge dx + \left(\frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy \right) \wedge dy = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy.$$

Now by a simple application of Fubini’s theorem we obtain that

$$\begin{aligned} \int_R d\omega &= \int_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy \\ &= \int_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dV \\ &= \int_c^d \left(\int_a^b \frac{\partial Q}{\partial x} dx \right) dy - \int_a^b \left(\int_c^d \frac{\partial P}{\partial y} dy \right) dx \\ &= \int_c^d (Q(b, y) - Q(a, y)) dy - \int_a^b (P(x, b) - P(x, c)) dx, \end{aligned}$$

where the last transition was the fundamental theorem of calculus. From here, we can simply use Definition 22.4, while considering that R is endowed with the natural orientation, and get that

$$\begin{aligned} \int_R d\omega &= \int_c^d (Q(b, y) - Q(a, y)) dy - \int_a^b (P(x, b) - P(x, c)) dx \\ &= \int_c^d Q(b, y) dy - \int_c^d Q(a, y) dy - \int_a^b P(x, b) dx + \int_a^b P(x, c) dx \\ &= \int_{\partial R} \omega. \end{aligned}$$

□

Remark 23.16. Green’s theorem is a special case of Stokes’ theorem, which we will get to soon.

Corollary 23.4. *Let $S \subset \mathbb{R}^2$ be parameterized by a rectangle, and let ω be a 1-form on S . Then*

$$\int_{\partial S} \omega = \int_S d\omega.$$

Proof. Let $g: R \rightarrow S \subset \mathbb{R}^2$ be a parameterization. Then, by Theorem 23.3 and Remark 23.9 we get that

$$\int_{\partial S} \omega = \int_{\partial R} g^* \omega = \int_R d(g^* \omega) = \int_R g^*(d\omega) = \int_S d\omega.$$

□

Remark 23.17. We can notice that Green's theorem also holds on the unit disc, since we can approximate it by $S_{\delta, \epsilon}$ sets parameterized by $g: [\delta, 1] \times [\epsilon, 2\pi]$ defined by $g(r, \theta) = r(\cos \theta, \sin \theta)$ with the natural orientation. Also, Green's theorem holds for any region S that can be decomposed as a finite union of parameterized rectangles overlapping along their edges, since the edges of the rectangles contained in S always cancel out (each edge is counted twice, once with positive and once with negative induced orientation, so long as all the boundaries are given natural orientation).

Example 23.5. Let $\Omega = \{(x, y) \mid x^2 + y^2 \leq 1\}$ and define the functions $P, Q: \partial\Omega \rightarrow \mathbb{R}$ by

$$\begin{aligned} P(x, y) &\equiv 0, \\ Q(x, y) &= x. \end{aligned}$$

Then by Green's theorem we get that

$$\int_{\partial\Omega} x dy = \int_{\partial\Omega} P dx + Q dy = \int_{\Omega} 1 dV = \text{Vol}(\Omega) = \pi$$

We can see that when calculating the integral directly

$$\int_{\partial\Omega} x dy = \int_0^{2\pi} (\cos t) \cos t dt = \int_0^{2\pi} \cos^2(t) dt = \pi$$

we get the same result.

Remark 23.18. More generally, for a planar region S , we can compute $\text{Vol}(S)$ by using line integration. We have

$$dx \wedge dy = d(xdy) = d(-ydx) = d\left(\frac{1}{2}(-ydx + xdy)\right),$$

and thus we get multiple ways to compute the area of S ,

$$\text{Vol}(S) = \int_S 1 dx \wedge dy = \int_{\partial S} x dy = \int_{\partial S} -y dx = \frac{1}{2} \int_{\partial S} -y dx + x dy.$$

Example 23.6. Let ω be the 1-form given by

$$\omega := -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy$$

which is defined and continuously differentiable on $\mathbb{R}^2 \setminus \{0\}$. We set $A := \{(x, 0) \mid x \geq 0\}$ and define a function $f: \mathbb{R}^2 \setminus A \rightarrow \mathbb{R}$ by

$$f(x, y) := \arg(x + iy).$$

Intuitively, the value $f(x, y)$ is the angle (x, y) makes with the positive side of the x -axis. This function is continuous and can also be written as

$$f(x, y) := \begin{cases} \arctan\left(\frac{y}{x}\right), & x > 0 \text{ and } y > 0 \\ \frac{\pi}{2} - \arctan\left(\frac{y}{x}\right), & y > 0 \\ \pi + \arctan\left(\frac{y}{x}\right), & x < 0 \\ \frac{3\pi}{2} - \arctan\left(\frac{y}{x}\right), & y < 0 \end{cases}.$$

We have (already mentioned) that

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy = \omega.$$

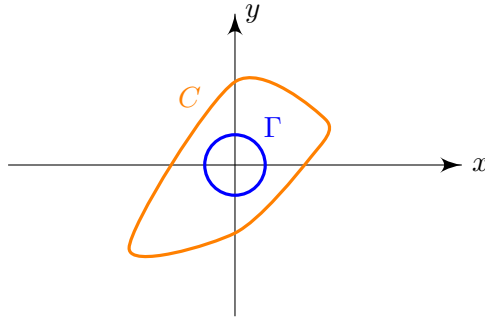
Indeed as mentioned, ω is exact on $\mathbb{R}^2 \setminus A$, so by Theorem 22.2, for any smooth simple curve contained in $\mathbb{R}^2 \setminus A$ we have the identity:

$$\int_C \omega = \int_C df = f(B) - f(A).$$

Now suppose we wanted to know if ω is exact on $\mathbb{R}^2 \setminus \{0\}$. We might want to apply the identity on the unit circle, in order to get a result that it is not 0, thus showing it cannot be exact on $\mathbb{R}^2 \setminus \{0\}$, however, since ω is exact on $\mathbb{R}^2 \setminus A$, this method only works for curves which are properly contained in $\mathbb{R}^2 \setminus A$. We may want to apply Green's theorem on the unit circle, in order to get a result that it is not 0, however, in order to use Green's theorem on the unit circle, ω must be smooth on the unit circle, which it is not because of the singularity at the origin. While these two methods fail, we can quickly compute by definition that $\int_{C_1(0)} \omega = 2\pi$, so indeed as expected, ω cannot be exact.

23.3 The winding number

Suppose C is any simple closed curve in the plane that encircles the origin, and let Γ be a circle entered at the origin lying inside the interior of C , as shown below:



We denote by Ω the set bounded by Γ and C . If we orient Ω and Γ counterclockwise, then if we recall Remark 23.15, we would get that $\partial\Omega = C + \Gamma^- = C - \Gamma$, where Γ^- is Γ with the opposite orientation. Once again, let

$$\omega := -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy.$$

Now, ω is smooth everywhere on Ω , so by exactness (in particular closedness) of ω , by Green's theorem we get that

$$0 = \int_{\Omega} d\omega = \int_{\partial\Omega} \omega = \int_C \omega - \int_{\Gamma} \omega.$$

This implies that

$$\int_C \omega = \int_{\Gamma} \omega = 2\pi,$$

where the last equality can be computed easily by definition, in a similar manner to what we saw in Example 23.6. By linearity of the integral over paths, we easily obtain that for a path C_n that winds around the origin n times, we have $\int_{C_n} \omega = 2\pi n$, and similarly if $\tilde{C}$ is a path that winds around the origin once in each direction, we would get $\int_{\tilde{C}} \omega = 2\pi - 2\pi = 0$.

Definition 23.6 (Winding number). Let C be a smooth closed curve on $\mathbb{R}^2 \setminus \{0\}$, the winding number (or index) of C is given by

$$\text{Ind}(C) := \frac{1}{2\pi} \int_C \omega = \frac{1}{2\pi} \int_C -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy.$$

Intuitively, it's the number of times C winds around the origin counterclockwise.

Remark 23.19. This result is very important in complex analysis.

23.4 Orientation

First recall the definition for a parameterized smooth surface.

Definition 23.7 (Parameterized smooth surface). Let $U \subset \mathbb{R}^2$ be a bounded open set, and let $g: U \rightarrow \mathbb{R}^n$ be an injective smooth map with $\text{rank } Dg(a) = 2$ for all $a \in U$. Then we call $S = g(U)$ a parameterized smooth surface of dimension n . In particular, it is a 2-dimensional manifold in $\mathbb{R}^n$.

Remember Remark 22.35? It is finally time to properly define orientation.

Definition 23.8 (Volume form). A volume form (or a top-dimensional form) is a differential form $\omega \in \Omega^n(M)$ of degree $n = \dim M$.

Remark 23.20. A top-dimensional form is called a volume form since it is the only type of form that allows us to compute the volume in the dimension of the manifold.

Definition 23.9 (Orientation via forms). An orientation of a smooth manifold M of dimension n is a smooth volume form $\omega \in \Omega^n(M)$ with $\omega(p) \neq 0$ for all $p \in M$ up to multiplication by a positive function $f: M \rightarrow \mathbb{R}^+$.

Remark 23.21 (Oriented manifold). A manifold is said to be orientable if it has some orientation. A manifold with a choice of orientation is called an oriented manifold. If M is a connected, orientable manifold, then it has exactly two possible orientations, because the space $\bigwedge^k(T_p^*M)$ is one dimensional and since the orientation is not zero it must be either in the positive side of the vector space or in its negative side, and the equivalence relation makes any two forms on the same side equivalent as orientations.

Definition 23.10 (Standard orientation). The standard volume form on $\mathbb{R}^n$ is $\omega = dx_1 \wedge \cdots \wedge dx_n$. When $n \leq 3$ we sometimes write $\omega = dx \wedge dy$ or $\omega = dx \wedge dy \wedge dz$ instead respectively.

Example 23.7. Consider the unit circle $S^1 := C_1(0) \subset \mathbb{R}^2$. A volume form on S^1 is a 1-form, for example $\omega = -ydx + xdy$. Since ω is nowhere-vanishing and smooth, it is an example of an orientation on S^1 .

Definition 23.11 (Positivity of an ordered basis). Let M be an n -dimensional manifold, let $B = \langle v_1, v_2, \dots, v_n \rangle \subset \mathbb{R}^n$ be an ordered basis for $T_p(M)$, and let ω be a nowhere vanishing volume form (an orientation) on M . We say that $B = \langle v_1, v_2, \dots, v_n \rangle$ is positive with respect to ω if $\omega(p)(v_1, v_2, \dots, v_n) > 0$, and otherwise we say that it is negative with respect to ω .

Remark 23.22. Note that a *choice of orientation* ω on a manifold M , is equivalent to a choice of an equivalence class of the nowhere vanishing forms in $\Omega^n(M)/\sim$ where $\omega \sim \tilde{\omega}$ if there exists a smooth function $f: M \rightarrow \mathbb{R}_{>0}$ such that $\omega = f\tilde{\omega}$. By a nowhere vanishing form we mean that ω is not identically 0 at any $p \in M$, and by equivalent, we mean that the positivity, or signs of ordered bases any every p would be the same with respect to any volume form in the same equivalency class.

Example 23.8. In $\mathbb{R}^2$, with respect to the standard orientation, the standard basis $\langle e_1, e_2 \rangle$ has positive orientation, $\langle e_2, e_1 \rangle$ has negative orientation, and $\langle e_1 + e_2, e_2 \rangle$ has positive orientation. We can see this quickly by recalling that $dx \wedge dy = \det$.

Remark 23.23. In Example 23.7, the orientation on the unit circle S^1 was $\omega = -ydx + xdy$. Define $g(\theta) = (\cos \theta, \sin \theta)$ so have that $T_{g(\theta)} = t(-\sin \theta, \cos \theta)$. Therefore $\langle (-\sin \theta, \cos \theta) \rangle$ is an ordered basis for $T_{g(\theta)}$. With respect to the orientation ω we notice that $\omega_{g(\theta)}(-\sin \theta, \cos \theta) = 1$, and thus the direction $(-\sin \theta, \cos \theta)$ is the positive direction of the tangent space. This orientation of the tangent space tells us intuitively, that the orientation ω on S^1 is the counterclockwise orientation.

Definition 23.12 (Framing). A framing of a smooth manifold M of dimension n is a continuous function that sends each point on the manifold $p \in M$ to an ordered basis $\langle v_1(p), v_2(p), \dots, v_n(p) \rangle$ of the tangent space $T_p(M)$.

Definition 23.13 (Parallelizable manifold). A manifold that admits a framing, is called a parallelizable manifold.

Example 23.9. The space S^1 from Example 23.7 has the framing $v_1(\theta) = \langle (-\sin \theta, \cos \theta) \rangle$, and therefore S^1 is a parallelizable manifold. However, $S^2 := \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ is not parallelizable since if $v = (v_1, v_2)$ were a framing, then the vector field v_1 is a contradiction to the hairy ball theorem.

Proposition 23.5 (Parallelizable implies orientable). *Every parallelizable manifold is orientable.*

Proof. Let $v = (v_1, \dots, v_n)$ be a framing on an n -dimensional manifold M . For every $p \in M$ define

$$\omega_p(v(p)) := 1,$$

and extend ω_p uniquely to an alternating multilinear volume form (this can be done since $\dim(\bigwedge^k(T_p M)^*) = 1$). Since v is smooth, ω is smooth, and since $\omega_p(v(p)) = 1$ for all $p \in M$, the volume form ω is nowhere vanishing. Therefore M is orientable which completes the proof. $\square$

Remark 23.24. By Theorem 23.5, all parallelizable manifolds are orientable, however the converse is not always true. For example, the 2-sphere given by $S^2 = \{p \in \mathbb{R}^3 \mid \|p\| = 1\}$ is orientable, but by a theorem (the hairy ball theorem), it does not admit a framing, so it is not parallelizable.

Definition 23.14 (Sign of an oriented manifold). Let M be an oriented manifold, let η be a reference orientation, and let ω be a volume form in its choice of orientation. Then we define the sign of M to be

$$\text{sgn}(M) := \begin{cases} 1 & \omega(p)(v_1, \dots, v_n) > 0, \\ -1 & \omega(p)(v_1, \dots, v_n) < 0, \end{cases}$$

for any arbitrary oriented basis $\langle v_1, \dots, v_n \rangle$ with positive orientation with respect to η .

Example 23.10. Not all smooth surfaces are orientable. A Möbius strip is not orientable. In topology we even say that a surface is orientable if it does not admit an embedding of a Möbius strip in it.

We will show that the Möbius strip is not orientable. Let $M = [0, 2\pi] \times (-1, 1) / \sim$ with $(x, y) \sim (x + 2\pi, -y)$ be the Möbius strip, and assume, by way of contradiction, that there exists a nowhere vanishing volume form $\omega(x, y) = f(x, y)dx \wedge dy \in \Omega^2(M)$. Consider the map $g: (x, y) \mapsto (x + 2\pi, -y)$. We have that

$$g^*(dx \wedge dy) = d(x + 2\pi) \wedge d(-y) = dx \wedge (-dy) = -dx \wedge dy,$$

so we can compute that

$$g^*(f(x, y)dx \wedge dy) = f(x + 2\pi, -y)g^*(dx \wedge dy) = -f(x + 2\pi, -y)dx \wedge dy,$$

so for a fixed point $(0, y)$ we have $g^*(f(0, y)dx \wedge dy) = -f(2\pi, -y)dx \wedge dy$. However, since $(0, y) \sim (2\pi, -y)$, we must have $g^*\omega(0, y) = \omega(2\pi, -y) = \omega(0, y)$, and thus $g^*(f(0, y)dx \wedge dy) = f(0, y)dx \wedge dy$. Therefore we obtained $f(0, y) = -f(2\pi, -y)$, but by the equivalence $f(0, y) = f(2\pi, -y)$, and therefore $f(0, y) = 0$, which is a contradiction, which completes the proof.

Remark 23.25 (Orientation and normal vectors). Let S be a parallelizable surface in $\mathbb{R}^n$, and let $v = (v_1, v_2)$ be some framing function on S . We know that $v_1(p) \times v_2(p)$ is a vector orthogonal to the tangent space $T_p(S)$. Since v_1 and v_2 are continuous functions, the function of the unit normal vector $n(p) = \frac{v_1(p) \times v_2(p)}{\|v_1(p) \times v_2(p)\|}$ is a continuous function as a composition of continuous functions, and by its construction, it is completely independent of the choice of the framing, only on its sign (whether it is positive or negative), and in fact if the orientation had a different sign, then its unit normal vector function would be $\tilde{n}(p) = -n(p)$. It follows that $\text{sgn}(S)$ has a one to one correspondence with the direction of the choice of a unit normal vector function. This also works in manifolds of higher dimensions when we generalize the vector product, and we could talk about it more when we will introduce the Hodge star operator. Note that even for manifolds that are not parallelizable like S^2 , there is a similar normal vector at every point.

Example 23.11 (Framings and orientations). Recall that a framing maps each point p on a surface S to an ordered basis $\langle v_1(p), v_2(p) \rangle$ of the tangent space of S at p . The standard framing is given by mapping each point to the standard basis of the tangent space at that point. Each framing $\langle v_1(p), v_2(p) \rangle$ corresponds to an orientation $\omega = v_1(p) \wedge v_2(p)$. For example for $S = \{(x, y) \mid x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$, the standard framing is the constant function $f: p \mapsto \langle dx, dy \rangle$. It corresponds to the standard volume form $\omega = dx \wedge dy$.

Definition 23.15 (Signed integration). Let Ω be a manifold with a choice of orientation, and let $f: \Omega \rightarrow \mathbb{R}$. We define:

$$\int_{\Omega} f dx_1 \wedge \cdots \wedge dx_n = \text{sgn}(\Omega) \int_{\Omega} f dV,$$

where $\text{sgn}(\Omega)$ is taken with respect to the standard orientation on Ω .

23.5 The interior product and induced orientation

Definition 23.16 (Interior product). Let M be a manifold, let $\omega \in \Omega^k(M)$, and let X be a vector field on M . Then the interior product (sometimes called a contraction, or interior derivative) of X is the map $\iota_X: \Omega^k(M) \rightarrow \Omega^{k-1}(M)$, defined by

$$(\iota_X \omega)(X_1, \dots, X_{p-1}) = \omega(X, X_1, \dots, X_{p-1}).$$

When ω is a 0-form, we define $\iota_X \omega = 0$ by convention.

Example 23.12. Let $\omega \in \Omega^1(M)$ be a 1-form, and let $X: M \rightarrow \mathbb{R}^{\dim M}$ be a vector field on M . Then we get

$$(\iota_X \omega) = \alpha(X) \in \Omega^0(M).$$

Note that $\alpha(X): M \rightarrow \mathbb{R}$ by $p \mapsto \alpha(X(p))$. Therefore, indeed $\alpha(X) \in \Omega^0(M) = C^\infty(M, \mathbb{R})$ (in the smooth case).

Remark 23.26. Let $\omega \in \Omega^p(M)$ and $\eta \in \Omega^q(M)$. Then

$$\iota_X(\omega \wedge \eta) = (\iota_X\omega) \wedge \eta + (-1)^p\omega \wedge (\iota_X\eta).$$

This means that the interior derivative ι_X , must like the exterior derivative, also satisfies the graded Leibniz rule (incidentally, this is why we call ι_X the interior *derivative*. It satisfies the rule of differentiation of a product).

Given an orientation (nowhere vanishing volume form) $\omega \in \Omega^n(M)$ on an n -dimensional manifold M , we want to find the induced orientation on submanifolds N of codimension 1 ($\dim N = n - 1$) of M (in particular we are interested in the induced orientation on ∂M). Therefore, we need to find the right nowhere vanishing $(n - 1)$ -form on ∂M . Now that we have the definition of the interior product in hand, this is equivalent to finding the right vector field $X: \partial M \rightarrow TM$, such that if B_p is a basis for $T_p\partial M$, then $\{X(p)\} \cup B_p$ is a basis for T_pM . In that case, we could define $\tilde{\omega} = \iota_X(\omega)$ to be the induced orientation on ∂M .

Definition 23.17 (Outward/Inward pointing vector). Let $p \in \partial M$. We then have the inclusion $T_p\partial M \subseteq T_pM$. If $X_p \in T_pM$, then we can write

$$X_p = \sum_{i=1}^n a_i e_i,$$

for $a_i \in \mathbb{R}$ and $\langle e_1, \dots, e_{n-1} \rangle$ an oriented basis for $T_p\partial M$, and e_n an orthogonal complement for $T_p\partial M$ in T_pM , so that M is given locally with coordinates $x_n \geq 0$. We say that X_p is outward pointing if $a_n < 0$, and inward pointing if $a_n > 0$.

Example 23.13. In the unit disc $\mathbb{D} \subset \mathbb{R}^2$ with the standard orientation $\omega = dx \wedge dy$,

Definition 23.18 (Generalized vector product). Let $v_1, \dots, v_{n-1}$ be vectors in $\mathbb{R}^n$. We define the generalized vector product $N = v_1 \times v_2 \times \dots \times v_{n-1}$ to be the unique vector in $\mathbb{R}^n$ so that for all $x \in \mathbb{R}^n$ we have

$$\langle N, x \rangle = \det \begin{pmatrix} | & | & & | \\ x & v_1 & \cdots & v_{n-1} \\ | & | & & | \end{pmatrix}.$$

Remark 23.27. It can be shown that if $\langle dx_1, \dots, dx_{n-1} \rangle$ is an oriented basis for $T_p\partial M$, then $\tilde{N}(p) = dx_1 \times \dots \times dx_{n-1}$ is an outward pointing normal vector to $T_p\partial M$. We can now define the vector field $N: \partial M \rightarrow TM$ by $N(p) = \frac{\tilde{N}(p)}{\|\tilde{N}(p)\|}$. This is the generalization to the unit normal vector function in Remark 23.25.

Definition 23.19 (Induced orientation). Let M be an oriented manifold with boundary, and let ω be a nowhere-vanishing volume form defining the orientation of M . For $p \in \partial M$ let $N(p)$ denote the outward-pointing unit normal vector at p . We say that an $(n - 1)$ -form η defines the *induced orientation* on ∂M if for any $e_1, \dots, e_{n-1} \in T_p\partial M$ so that $\langle e_1, \dots, e_{n-1} \rangle$ is an oriented basis for $T_p\partial M$ we have,

$$\eta(p)(e_1, \dots, e_{n-1}) = \omega(p)(N(p), e_1, \dots, e_{n-1}).$$

That is, $\eta = \iota_N\omega$.

Example 23.14. Let $M = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$, so that $\partial M = S^1$ is the unit circle. Give M the standard orientation via the volume form $\omega = dx \wedge dy$, so in particular $\text{sgn}(M) = 1$. Now we want to find the orientation of ∂M induced by M . First, we can compute and quickly find that $N(p) = (\cos \theta, \sin \theta)$ is the outward pointing vector at each $p \in \partial M$ where θ is

the angle between p and the x axis. The induced orientation is thus $\iota_N\omega$. We can now compute and get that for $v \in \mathbb{R}^2$,

$$(\iota_N\omega)(p)(v) = \omega(p)(N(p), v) = \det(N(p) \mid v) = \cos\theta v_y - \sin\theta v_x,$$

so the induced orientation is $\tilde{\omega} = -\sin\theta dx + \cos\theta dy$. In polar coordinates we get

$$\tilde{\omega} = -\sin\theta d(\cos\theta) + \cos\theta d(\sin\theta) = (-\sin\theta)^2 d\theta + (\cos\theta)^2 d\theta = d\theta.$$

Let $\gamma(t) = (1, t)$ be a parameterization for the circle in polar coordinates so that $\langle \gamma'(t) \rangle = \langle (0, 1) \rangle$ is an oriented basis for $T_p\partial M$ at any point $p \in \partial M$. We get that $d\theta(\gamma'(t)) = 1 > 0$, and therefore the basis $\langle \gamma'(t) \rangle$ has positive orientation. This implies that $\gamma'(t)$ has positive orientation, and in particular, the orientation is counterclockwise. This coincides with the natural induced orientation for the boundary we introduced in Definition 23.5. If our reference form for the sign of ∂M is $\eta = d\theta$, then clearly $\text{sgn}(\partial M) = 1$, and if $\eta = -d\theta$ then $\text{sgn}(\partial M) = -1$.

Exercise 23.3. Show that the induced orientation on $\partial\mathbb{R}_+^k$ is the usual one on $\mathbb{R}^{k-1}$ if and only if k is even.

Remark 23.28. When first stating Green's theorem we used the definition of "natural orientation". However, as we saw in Example 23.14, we can simply use the more general definition of induced orientation instead. Computationally, this does not change a thing, but mathematically, this implies there may be a way to generalize Green's theorem to higher dimensions.

Definition 23.20 (Flux). Let $S \subset \mathbb{R}^3$ be an oriented surface, let $\mathbf{F} = (F_1, F_2, F_3)$ be a continuous vector field, and let $\omega = F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dy$ be the flux form associated with $\mathbf{F}$. Suppose that $g: \Omega \rightarrow S$ is a smooth parameterization of S . Then the flux Φ of $\mathbf{F}$ across S is defined to be:

$$\int_S \omega = \int_\Omega g^*\omega = \int_\Omega \left\langle \mathbf{F}(g(u, v)), \frac{g'_u \times g'_v}{\|g'_u \times g'_v\|} \right\rangle \|g'_u \times g'_v\| du \wedge dv = \int_\Omega \langle \mathbf{F}(g(u, v)), g'_u \times g'_v \rangle dudv$$

Remark 23.29. Recall that we also described Φ by $\int_S \langle \mathbf{F}, \mathbf{n}(u, v) \rangle d\sigma$, where $d\sigma$ is the *surface area element*. We now understand that $d\sigma$ is the differential form that satisfies that its pullback is equal to the form $g^*(d\sigma) = \|g'_u \times g'_v\| du \wedge dv$.

Exercise 23.4. Let S be the orientable surface given by

$$S := \left\{ (x, y, z) \mid z \geq 0 \text{ and } z = 4 - x^2 - y^2 \right\}.$$

Compute the flux of the field $\mathbf{F}(x, y, z) = (xz, yz, x^2 + y^2)$ across S , where S is given the induced orientation as the boundary of the area bounded between S and the plane $z = 0$. Equivalently, it is given the orientation by declaring that an ordered basis $\langle v_1, v_2 \rangle$ of its tangent space, is positive if and only if $\langle n(p), v_1, v_2 \rangle$ is a positive basis for $\mathbb{R}^3$ where $n(p)$ is the outward pointing unit normal vector at $p \in S$.

Solution. In order to calculate this integral we first need to find a parameterization to this surface, for example $g: \Omega \rightarrow \mathbb{R}^3$ where

$$\Omega := \left\{ (u, v) \in \mathbb{R}^2 \mid u^2 + v^2 \leq 4 \right\} \quad g(u, v) := (u, v, 4 - u^2 - v^2)$$

Recall that

$$\Phi := \int_S \omega = \int_\Omega \langle \mathbf{F}(g(u, v)), g'_u \times g'_v \rangle dudv,$$

where the order $g'_u \times g'_v$ (and not $g'_v \times g'_u$) was chosen because it is the order that gives an outward pointing normal. By a simple computation:

$$g_u = \begin{pmatrix} 1 \\ 0 \\ -2u \end{pmatrix} \quad \text{and} \quad g_v = \begin{pmatrix} 0 \\ 1 \\ -2v \end{pmatrix},$$

and thus

$$g_u \times g_v = \hat{x} \begin{vmatrix} 0 & 1 \\ -2u & 2v \end{vmatrix} + \hat{y} \begin{vmatrix} -2u & -2v \\ 1 & 0 \end{vmatrix} + \hat{z} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = \cdots = \begin{pmatrix} 2u \\ 2v \\ 1 \end{pmatrix}.$$

We can now plug in the values:

$$\begin{aligned} \int_{\Omega} \mathbf{F}(g(u, v)) \cdot (g_u \times g_v) du \wedge dv &= \iint_{\Omega} [u(4 - u^2 - v^2), v(4 - u^2 - v^2), u^2 + v^2] \cdot [2u, 2v, 1]^T du dv \\ &= \iint_{\Omega} (u^2 + v^2) [2(4 - u^2 - v^2) + 1] du dv \\ &= \int_0^{2\pi} \int_0^2 (9r^2 - 2r^4) r dr d\theta. \end{aligned}$$

where the last transition is the change of variables theorem. The computation of the last integral is very basic, so we will omit it and conclude the exercise here.

23.6 Surface area element

The purpose of this short subsection is to understand what is the surface area element discussed in Remark 21.4, Remark 22.27, and Remark 23.29.

Remark 23.30 (Geometric interpretation of the interior product). Recall that the interior product takes a k -form ω , a vector field X and returns a $(k - 1)$ -form $\iota_X \omega$ which is simply ω with X inserted in its first argument.

Recall that when ω is a volume form ($\omega \in \Omega^n(M)$), it computes the n -dimensional volume of the parallelopiped spanned by n vectors in $\mathbb{R}^n$.

In this case, the geometrical interpretation of $\iota_X \omega$ at a point p is that it computes the n -dimensional volume of the parallelopiped spanned by $X(p)$ and $n - 1$ other vectors in $\mathbb{R}^n$.

We claim that that this value is equal to the $(n - 1)$ -dimensional volume of the parallelopiped spanned by $n - 1$ vectors in $\mathbb{R}^n$ projected onto the affine space that is orthogonal complement of $X(p)$ at p , or in other words projected onto the tangent space $T_p(M)$. This is the reason the interior product is sometimes called a “contraction”.

Example 23.15. The simplest example is to consider the area form $\omega = dx \wedge dy$, and the constant vector field $X(\mathbf{p}) \equiv (1, 0)$ for all $\mathbf{p} \in \mathbb{R}^2$. In this case we have

$$\iota_X \omega(\mathbf{p})(\mathbf{v}) = \det \begin{pmatrix} 1 & v_x \\ 0 & v_y \end{pmatrix} = v_y,$$

for all $\mathbf{v} = (v_x, v_y) \in \mathbb{R}^2$. This implies clearly that $\iota_X \omega = dy$, which coincides with the previously discussed geometric interpretation of the interior product.

Remark 23.31. Consider the manifold $M = S^2$ endowed with the standard orientation induced by $\mathbb{R}^3$, which is $\eta := \iota_N(dx \wedge dy \wedge dz)$ where N is the vector field that assigns each $p \in S^2$ an outward pointing normal vector, we get that η is the standard area form on $T_p(S^2)$. The form η is exactly what we need in order to integrate some any function over S^2 . This is true in general for any oriented manifold M . When M is a curve, the form η is called the *curve length element* associated with M . When M is a surface, the form η is called the *surface area element* associated with M .

Example 23.16. Let C be a 1-dimensional oriented curve in $\mathbb{R}^2$, and let $\eta = \iota_N(\omega) \in \Omega^1(C)$ be its curve length element, where $\omega = dx \wedge dy$ is the standard volume form on $\mathbb{R}^2$. Let $\gamma(t) = (x(t), y(t))$ be a local parameterization of C , and write η explicitly $\eta = N_x dy - N_y dx$, where N is the outward pointing normal ($\omega(N, \gamma') > 0$). Notice that we can write N explicitly by rotating the vector $\gamma'(t) = (x'(t), y'(t))$ and normalizing as follows:

$$N(t) = \frac{(y'(t), -x'(t))}{\sqrt{(x'(t))^2 + (y'(t))^2}}.$$

This means that we can compute the pullback of the functions $N_x(x, y)$ and $N_y(x, y)$ easily and get

$$\gamma^*\eta(t) = \frac{y'(t)}{\sqrt{(x'(t))^2 + (y'(t))^2}} \gamma^*dy - \frac{-x'(t)}{\sqrt{(x'(t))^2 + (y'(t))^2}} \gamma^*dx = \frac{x'(t)\gamma^*dx + y'(t)\gamma^*dy}{\sqrt{(x'(t))^2 + (y'(t))^2}}.$$

Note that $dx(t) = x'(t)dt$ and $dy(t) = y'(t)dt$, so we get that

$$\gamma^*\eta(t) = \frac{x'(t)\gamma^*dx + y'(t)\gamma^*dy}{\sqrt{(x'(t))^2 + (y'(t))^2}} = \frac{x'(t)x'(t)dt + y'(t)y'(t)dt}{\sqrt{(x'(t))^2 + (y'(t))^2}} = \sqrt{(x'(t))^2 + (y'(t))^2} dt$$

Recall that $\|\gamma'(t)\| = \sqrt{(x'(t))^2 + (y'(t))^2}$, and therefore the curve length element is

$$\gamma^*\eta = \|\gamma'(t)\| dt.$$

Compare this result with Remark 15.1.

Example 23.17. Let S be a 2-dimensional oriented surface, and let $\eta = \iota_N(\omega)$ be its surface area element. Let $g: U \rightarrow \mathbb{R}^3$ be a local parameterization of S , so that $g(u, v) = (x(u, v), y(u, v), z(u, v))$, and write η explicitly $\eta = N_x dy \wedge dz + N_y dz \wedge dx + N_z dx \wedge dy$. Using the notation ∂_u and ∂_v for the basis tangent vectors of $T_{(u,v)}(U)$, by definition of the pullback we get that

$$g^*\eta_{(u,v)}(\partial_u, \partial_v) = \eta_{g(u,v)}(dg_{(u,v)}(\partial_u), dg_{(u,v)}(\partial_v)) = \eta(g_u, g_v),$$

and so by definition of η we get $g^*\eta(\partial_u, \partial_v) = \omega(N, g_u, g_v)$, however since $N(u, v) = \frac{g_u \times g_v}{\|g_u \times g_v\|}$, we can compute that

$$\omega(N, g_u, g_v) = \frac{1}{\|g_u \times g_v\|} \det \begin{pmatrix} | & | & | \\ g_u \times g_v & g_u & g_v \\ | & | & | \end{pmatrix} = \frac{\|g_u \times g_v\|^2}{\|g_u \times g_v\|} = \|g_u \times g_v\|.$$

Since $g^*\eta(\partial_u, \partial_v) = \|g_u \times g_v\|$, we can expand $g^*\eta$ multilinearly and antisymmetrically, and get that $g^*\eta = \|g_u \times g_v\| du \wedge dv$. Compare this with Remark 21.4.

24 Generalized Stokes theorem

24.1 Partition of unity

So far we know how to integrate a k -form over a parameterized k -dimensional manifold by pulling back. In general, a manifold will be a union of parametrized pieces that overlap, and so summing the integrals will give a meaningless result. In this subsection, we will introduce a tool that will help us overcome this challenge by partitioning the global integration problem into local ones.

Definition 24.1 (Smooth manifold with boundary). A subset $M \subset \mathbb{R}^n$ is called a smooth k -manifold with boundary of dimension k if for every $p \in M$ there exists a neighbourhood $p \in U$, an open set $V \subset \mathbb{R}^k$ and an injection $H \in C^\infty(V, \mathbb{R}^n)$ such that:

- (1) $\text{rank } DH(v) = k$ for all $v \in V$.
- (2) $M \cap U = H(V \cap \mathcal{H}_k)$.
- (3) $H^{-1}: M \cap U \rightarrow V \cap \mathcal{H}_k$ is continuous with respect to the topology induced by $\mathbb{R}^n$.

In other words, H restricts to a homeomorphism $H: V \cap \mathcal{H}_k \rightarrow M \cap U$. The function H is called a *parametrization*, or a *chart* of M at p . Indeed, we also require charts to be compatible, that is, if there exist $H_i: V_i \rightarrow \mathbb{R}^n$ and $H_j: V_j \rightarrow \mathbb{R}^n$ which overlap, then we require the *transition map* $H_i^{-1} \circ H_j$ to be smooth. We say that $p \in \partial M$ if $p = H(u)$ for some $u \in \mathcal{H}_k$.

Remark 24.1. In Definition 24.1, $H(V)$ is sometimes called a *coordinate chart* on M , and a *coordinate ball* on M is the image of some ball under some parameterization H .

Theorem 24.1 (Partition of unity). Let $M \subset \mathbb{R}^n$ be a compact k -dimensional manifold with boundary. Then there are smooth real-valued functions $\varphi_1, \dots, \varphi_N$ on M so that:

- (1) $0 \leq \varphi_i \leq 1$ for all $1 \leq i \leq N$.
- (2) Each φ_i is zero outside some coordinate ball.
- (3) $\sum_{i=1}^N \varphi_i \equiv 1$.

Remark 24.2. We call $\{\varphi_i\}$ a *partition of unity* on M .

Definition 24.2 (Support). Let $f: X \rightarrow Y$ be a function between two manifolds. Then the support of f is defined to be the set

$$\text{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}}.$$

Definition 24.3 (Bump function). A function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that is smooth and has a compact support.

Remark 24.3. The set of all bump functions with domain $\mathbb{R}^n$ forms a vector space, denoted by $C_0^\infty(\mathbb{R}^n)$ or $C_c^\infty(\mathbb{R}^n)$.

Example 24.1. First define $h: \mathbb{R} \rightarrow \mathbb{R}$ be

$$h(x) = \begin{cases} e^{-\frac{1}{x}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Then we know from analysis 1 that h is smooth and all of its derivatives vanish at the origin. Next define:

$$j(x) = \frac{\int_0^x h(t)h(1-t)dt}{\int_0^1 h(t)h(1-t)dt},$$

and define $\psi: \mathbb{R}^k \rightarrow \mathbb{R}$ by $\psi(\mathbf{x}) = j(3 - 2\|\mathbf{x}\|)$. Then by a matter of technical calculation (or a particularly strong belief), we get that ψ is a smooth function so that $\psi(\mathbf{x}) = 1$ for $\|\mathbf{x}\| \leq 1$ and $\psi(\mathbf{x}) = 0$ whenever $\|\mathbf{x}\| \geq 3/2$, so ψ has compact support. Therefore ψ is an example of a bump function.

Proof of Theorem 24.1. Let ψ be the bump function from Example 24.1. For each point $\mathbf{p} \in M$, choose a coordinate chart whose domain is a ball of radius 2 in $\mathbb{R}^k$, and if $\mathbf{p} \in \partial M$ let that be the half-ball of radius 2 (this can be done since any open set in $\mathbb{R}^k$ contains an open ball, in particular with center $\mathbf{p}$, then by restricting the chart to that ball we get what we need). Since the images of those balls restricted to radius 1 cover the manifold, and since the manifold is compact, there exist a finite amount of $\mathbf{p}$, so that their respective balls of radius 1 contained in the balls of radius 2 denoted $V_1, \dots, V_n$, cover M . Let $g_i: B_2(0) \rightarrow V_i$ be their respective coordinate charts, and define $\theta_i = \psi \circ g_i^{-1}$, interpreting θ to be defined on all of M by letting it be 0 outside of V_i (note that θ_i is smooth since ψ is zero outside the ball of radius 3/2). Define

$$\rho_i = \frac{\theta_i}{\sum_{j=1}^N \theta_j}.$$

Note that for each $\mathbf{p} \in M$, we have $\mathbf{p} = g_j(\mathbf{u})$ for some j and $\mathbf{u} \in B_1(0)$, and hence $\theta_j(\mathbf{p}) = 1$ for some j , which implies that the sum is always greater than 1 and in particular it does not vanish. If we recall the original statement, we can now notice that the functions $\{\rho_i\}$ form a partition of unity on M . $\square$

Remark 24.4. Let $M \subset \mathbb{R}^n$ be a compact, oriented k -dimensional manifold (with piecewise-smooth boundary), let ω be a k -form on M , let $\{\rho_i\}$ be a partition of unity, and let g_i be the corresponding parameterizations, which we take to be orientation preserving (we can do this because if any g_i is not orientation preserving we just precompose it with a reflection). Then we define

$$\int_M \omega = \int_M \left(\sum_{i=1}^N \rho_i \omega \right) = \sum_{i=1}^N \int_{B_2(0)} g_i^*(\rho_i \omega) = \sum_{i=1}^N \int_{B_2(0)} \psi g_i^* \omega.$$

Note that we used all properties of a partition of unity in order for this to work.

Remark 24.5. When considering complex manifolds (manifolds with a complex structure), we can have a partition of unity on them only when we identify them as a smooth manifold. We cannot have a “holomorphic partition of unity” because by Liouville’s theorem all holomorphic functions with compact support must be identically zero.

24.2 Stokes’ theorem

Theorem 24.2 (Stokes’ theorem). *Let $k > 1$, and let M be a compact oriented k -dimensional manifold in $\mathbb{R}^n$. Give ∂M the induced orientation if ∂M is nonempty. Let ω be a $k-1$ form defined in an open set of $\mathbb{R}^n$ containing M . Then*

$$\int_M d\omega = \int_{\partial M} \omega$$

if ∂M is nonempty, and $\int_M d\omega = 0$ otherwise.

Remark 24.6. Note that the usual fundamental theorem of calculus, the fundamental theorem of calculus for line integrals, and Green’s theorem are all special cases of this theorem. The following equation shows the first example clearly:

$$\int_a^b df = \int_a^b f'(t)dt = f(b) - f(a).$$

Proof of Theorem 24.2. to be added □

Remark 24.7. Stokes' theorem is also valid when the boundary, rather than being a $(k-1)$ -manifold in itself, is piecewise smooth (when the manifold has “corners”). For example, we may apply it on shapes like a closed cube, or a cone.

Corollary 24.3. *Let M be a compact oriented k -dimensional manifold without boundary. Let ω be an exact k -form. Then $\int_M \omega = 0$.*

Proof. Immediate. □

Example 24.2. Define the smooth 2-form:

$$\omega := \frac{xdy \wedge dz + ydz \wedge dx + zdx \wedge dy}{(x^2 + y^2 + z^2)^{3/2}} \in \Omega^2(\mathbb{R}^3 - \{0\}).$$

On a sphere of radius a , centered at the origin, ω is simply $\frac{1}{a^2}$ times the standard area (volume) 2-form. Suppose g is the parameterization of spherical coordinates, then one can compute that

$$g^*\omega = \sin \theta d\theta \wedge d\varphi.$$

It is clear that $d(g^*\omega) = 0$, and since $\det Dg \neq 0$ whenever $\rho \neq 0$ and $\theta \neq 0, \pi$, it follows that $d\omega = 0$ (we could have also computed this directly). We can now compute that for any sphere S centered at the origin:

$$\int_S \omega = \int_0^{2\pi} |d\varphi| \int_0^\pi \sin \theta |d\theta| = 4\pi,$$

however, for the ball B associated with S we get $\int_B d\omega = \int_B 0 = 0$. This seems as if Stokes' theorem has failed, but the real reason behind this mismatch is that ω is not smooth (and it's not even defined) on all of B . We want to claim that

$$\int_{\partial\Omega} \omega = \begin{cases} 4\pi, & 0 \in \Omega \\ 0, & 0 \notin \Omega \end{cases}.$$

When $0 \notin \Omega$ we can apply Stokes' theorem and immediately get the desired result. When $0 \in \Omega$, we choose $\epsilon > 0$ sufficiently small so that $\overline{B}(0, \epsilon) \subset \Omega$, then denote $\Omega_\epsilon = \Omega - B(0, \epsilon)$, so that $\partial\Omega_\epsilon = \partial\Omega - S_\epsilon$. Then $0 \notin \Omega_\epsilon$ so

$$0 = \int_{\partial\Omega_\epsilon} \omega = \int_{\partial\Omega} \omega - \int_{S_\epsilon} \omega.$$

Thus,

$$\int_{\partial\Omega} \omega = \int_{S_\epsilon} \omega = 4\pi,$$

as we saw earlier.

25 Applications to physics

25.1 Classical Stokes' theorem

We have already seen that a vector field in $\mathbb{R}^3$ can be interpreted either as a 1-form (when we are computing work), or as a 2-form (when we are computing flux). We have seen that for a function (0-form) that df corresponds to the vector field ∇F , and now we want to give the traditional interpretations of the exterior derivative on 1-forms and 2-forms as well.

Definition 25.1 (Curl). Let $\omega = F_1 dx_1 + F_2 dx_2 + F_3 dx_3$ be a 1-form on $\mathbb{R}^3$, and let $\mathbf{F} = (F_1, F_2, F_3)$ be its corresponding vector field. By the definition of the exterior derivative:

$$d\omega = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) dx_3 \wedge dx_1 + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) dx_1 \wedge dx_2.$$

The curl of $\mathbf{F}$ is defined to be the following vector field:

$$\text{curl}(\mathbf{F}) := \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right).$$

Remark 25.1. In older texts, it is also common to use the notation $\text{rot}(\mathbf{F})$ for $\text{curl}(\mathbf{F})$, where rot stands for rotor. Both suggest that $\text{curl}(\mathbf{F})$ has something to do with rotation (curling).

Remark 25.2. Note that the curl can also be written as the vector product of ∇ and $\mathbf{F}$ as follows:

$$\text{curl}(\mathbf{F}) = \nabla \times \mathbf{F} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ F_1 & F_2 & F_3 \end{vmatrix}.$$

Remark 25.3. Note that for all C^2 functions f we have that $\text{curl}(\text{grad } f) = \text{curl}(\nabla f) = 0$, since we know by the identity $d^2 = 0$ we get that $d(df) = 0$. This is no more than a reformulation of Remark 23.8.

Theorem 25.1 (Classical Stokes' theorem). Let $S \subset \mathbb{R}^3$ be a compact, oriented surface with boundary. Let $\mathbf{F}$ be a smooth vector field defined on all of S . Then we have

$$\int_{\partial S} \underbrace{\mathbf{F} \cdot \mathbf{T} ds}_{\omega} = \int_S \underbrace{\text{curl}(\mathbf{F}) \cdot \mathbf{n} dS}_{d\omega},$$

where $\mathbf{T}$ is the tangent unit vector, ds is the curve length element, $\mathbf{n}$ is the outward pointing unit normal vector, and dS is the scalar surface area element.

Remark 25.4. Indeed, if $\omega = F_1 dx_1 + F_2 dx_2 + F_3 dx_3$ and $\gamma: [a, b] \rightarrow C \subset \partial S$ is a chart of ∂S , we get that

$$\gamma^*(\mathbf{F} \cdot \mathbf{T} ds) = \mathbf{F}(\gamma(t)) \cdot \frac{\gamma'(t)}{\|\gamma'(t)\|} \|\gamma'(t)\| dt = \mathbf{F}(\gamma(t)) \cdot \gamma'(t) dt = \gamma^* \omega,$$

therefore we get that $\omega = \mathbf{F} \cdot \mathbf{T} ds$ on the chart C , and since this holds for all charts of ∂S , which cover ∂S , we get that $\omega = \mathbf{F} \cdot \mathbf{T} ds$ on ∂S . Think about what similar argument using partition of unity could show that integrating over ω and $\mathbf{F} \cdot \mathbf{T} ds$ gives the same result without mentioning that they are globally equal.

Similarly for a chart (U, g) of S we get that

$$\begin{aligned} g^*(\operatorname{curl}(\mathbf{F}) \cdot \mathbf{n} dS) &= \operatorname{curl}(\mathbf{F}(g(u, v))) \cdot \frac{g'_u \times g'_v}{\|g'_u \times g'_v\|} \|g'_u \times g'_v\| du \wedge dv \\ &= \operatorname{curl}(\mathbf{F}(g(u, v))) \cdot g'_u \times g'_v du \wedge dv = g^* d\omega, \end{aligned}$$

where the last transition follows from a simple computation. Then we conclude in the same way that $d\omega = \operatorname{curl}(\mathbf{F}) \cdot \mathbf{n} dS$ on S , and therefore the classical Stokes' theorem is indeed a special case of the generalized Stokes' theorem.

25.2 Geometric intuition for curl

Visualize $\mathbf{F}$ as the velocity field of a fluid. Then the line integral $\int_C \mathbf{F} \cdot \mathbf{T} ds$ around a closed curve C could be interpreted as the *circulation* of $\mathbf{F}$ around C , which we may think of as the tendency of a piece of wire of the shape of C to spin when dropped in the fluid.

Example 25.1. Let $S = D_r$ be the 2-dimensional disc of radius r centered at a , with normal vector $\mathbf{n}$. Then Stokes' theorem gives us that

$$\int_{\partial D_r} \mathbf{F} \cdot \mathbf{T} ds = \int_{D_r} \operatorname{curl}(\mathbf{F}) \cdot \mathbf{n} dS,$$

and since the curl is a continuous function, the flux integral (note that the right hand side is a flux integral) satisfies the mean value theorem for integrals, which implies that for some $q \in D_r$ we have

$$\int_{D_r} \operatorname{curl}(\mathbf{F}) \cdot \mathbf{n} dS = \operatorname{Vol}(D_r) \operatorname{curl}(\mathbf{F})(q) \cdot \mathbf{n}(q),$$

so combined with Stokes' theorem

$$\operatorname{curl}(\mathbf{F})(q) \cdot \mathbf{n}(q) = \frac{1}{\pi r^2} \int_{\partial D_r} \mathbf{F} \cdot \mathbf{T} ds,$$

and when taking $r \rightarrow 0^+$ we get $q \rightarrow a$, and then

$$\operatorname{curl}(\mathbf{F})(a) \cdot \mathbf{n}(a) = \lim_{r \rightarrow 0^+} \frac{1}{\pi r^2} \int_{\partial D_r} \mathbf{F} \cdot \mathbf{T} ds,$$

so indeed the tendency to spin at a is greatest when $\mathbf{n}(a)$ is pointing in the direction of $\operatorname{curl}(\mathbf{F})(a)$. The most common metaphor is to imagine a paddlewheel with an axle pointing in the direction of $\mathbf{n}(a)$. In this case the paddlewheel will spin the fastest when $\mathbf{n}(a)$ points in the direction of $\operatorname{curl}(\mathbf{F})(a)$.

Remark 25.5. Some authors denote $\mathbf{T} ds = d\vec{r}$ and $\mathbf{n} dS = d\vec{S}$ or $\mathbf{n} dS = d\vec{n}$. In these cases Stokes' theorem becomes:

$$\int_{\partial S} \mathbf{F} \cdot d\vec{r} = \int_S \operatorname{curl}(\mathbf{F}) \cdot d\vec{n}.$$

In this form we can see more clearly that Stokes' theorem simply states that the line integral of a vector field over a loop is equal to the surface integral (the flux integral) of its curl over the enclosed surface.

25.3 Classical divergence theorem

Definition 25.2 (Divergence). Let $\omega = F_1 dx_2 \wedge dx_3 + F_2 dx_3 \wedge dx_1 + F_3 dx_1 \wedge dx_2$ be a 2-form on a surface $S \subset \mathbb{R}^3$ written as a flux form. Then we have

$$d\omega = \left(\frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3.$$

Correspondingly, for a vector field $\mathbf{F} = (F_1, F_2, F_3)$ we define its divergence to be the following scalar function:

$$\operatorname{div}(\mathbf{F}) := \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}.$$

Remark 25.6. Note that we can write $\operatorname{div}(\mathbf{F}) = \nabla \cdot \mathbf{F}$. The notation div is short for “divergence”. We shall soon see what it means geometrically. In this case $d^2 = 0$ can be restated as $\operatorname{div}(\operatorname{curl} \mathbf{F}) = 0$ for all C^2 vector fields $\mathbf{F}$. Using divergence, the generalized Stokes’ theorem takes the following form, sometimes called Gauss’ theorem.

Theorem 25.2 (Classical divergence theorem). Suppose $\mathbf{F}$ is a smooth vector field on a compact orientable 3-manifold with boundary $\Omega \subset \mathbb{R}^3$. Then

$$\int_{\partial\Omega} \underbrace{\mathbf{F} \cdot \mathbf{n} dS}_{\omega} = \int_{\Omega} \underbrace{\operatorname{div}(\mathbf{F}) dV}_{d\omega},$$

where $\mathbf{n}$ is the outward pointing unit normal.

Remark 25.7. The classical divergence theorem is sometimes called Gauss’ theorem, or Ostrogradsky’s theorem.

Example 25.2. We can use the divergence theorem to understand the relation between the volume of a sphere and its surface area. Let B_R be a ball of radius R around the origin. Let $\mathbf{F}$ be a vector field with $\mathbf{F} = \frac{1}{3}(x, y, z)$. We notice that

$$\operatorname{div}(\mathbf{F}) = \nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1.$$

The outward pointing unit normal vector to a point on the boundary of the ball $(x, y, z) \in \partial B_R$ is given by $\frac{1}{R}(x, y, z)$. Since we have that

$$\operatorname{Vol}(B_R) = \int_{B_R} 1 dV = \int_{B_R} \operatorname{div}(\mathbf{F}) dV$$

by the divergence theorem we conclude that

$$\operatorname{Vol}(B_R) = \int_{\partial B_R} \mathbf{F} \cdot \mathbf{n} dS = \int_{\partial B_R} \frac{1}{3R} \underbrace{(x^2 + y^2 + z^2)}_{R^2} dS = \frac{R}{3} \int_{\partial B_R} 1 dS = \frac{R}{3} \cdot A(\partial B_R),$$

where $A(\partial B_R)$ is the surface area of the ball. This is indeed the desired result because we know that

$$A(\partial B_R) = 4\pi R^2 \quad \text{and} \quad \operatorname{Vol}(B_R) = \frac{4\pi R^3}{3}.$$

Example 25.3. We want to calculate the flux of the vector field $\mathbf{F} = (x + z^2, \cos xz, z + x^3)$ on an arbitrary torus

$$T_{a,b} := \left\{ (x, y, z) \in \mathbb{R}^3 \mid \left(a^2 - \sqrt{x^2 + z^2} \right)^2 + y^2 = b^2 \right\}.$$

Denote by Ω the area enclosed by $T_{a,b}$, so that $\partial\Omega = T_{a,b}$. The divergence theorem states that

$$\int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n} dS = \int_{\Omega} \operatorname{div}(\mathbf{F}) dV,$$

so we can see that:

$$\operatorname{div}(\mathbf{F}) = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} = 1 + 0 + 1 = 2.$$

Therefore, the flux of $\mathbf{F}$ through $T_{a,b}$ is simply

$$\int_{\Omega} \operatorname{div}(\mathbf{F}) dV = \int_{\Omega} 2 dV = 2 \operatorname{Vol}(\Omega) = 4\pi^2 ab^2,$$

where the last transition follows from a computation we did in Example 20.1.

25.4 Geometric intuition for divergence

In a similar manner to what we did in Section 25.2, we can use the divergence theorem to conclude that for a point $a \in \Omega$ we have

$$\operatorname{div}(\mathbf{F})(a) = \lim_{r \rightarrow 0^+} \frac{1}{\frac{4}{3}\pi r^3} \int_{\partial \overline{B}_r(a)} \mathbf{F} \cdot \mathbf{n} dS.$$

In other words, $\operatorname{div}(\mathbf{F})(a)$ measures the flux (per unit volume) outward through an infinitesimally small sphere centered at a . If that flux is positive, we can visualize a as a *source* of the field, and if the flux is negative, we can visualize a as a *sink*.

25.5 Conservative fields

In this subsection we will supplement our understanding of conservative fields, and learn about locally conservative fields.

Corollary 25.3. *Let $U \subseteq \mathbb{R}^n$ be an open set and let $F \in C^1(U, \mathbb{R}^n)$ be a conservative vector field. Then for any point $p \in U$ and any i, j we have*

$$\frac{\partial F_i}{\partial x_j}(p) = \frac{\partial F_j}{\partial x_i}(p).$$

In particular when $n = 3$, if $\mathbf{F}$ is conservative then $\operatorname{curl}(\mathbf{F}) = \nabla \times \mathbf{F} = 0$ for any point in U .

Proof. We have already seen before that if $\mathbf{F}$ is conservative then $\mathbf{F} = \nabla f$ and then $\operatorname{curl}(\nabla f) = 0$. However, we have only really defined curl in 3 dimensions (although it's not too hard to see what the generalization should be), so let us add another explanation. We have seen in Section 25.2 that for any $a \in U$ we have

$$\operatorname{curl}(\mathbf{F})(a) \cdot \mathbf{n} = \lim_{r \rightarrow 0^+} \frac{1}{\pi r^2} \int_{\partial D_r} \mathbf{F} \cdot \mathbf{T} ds,$$

where $\mathbf{n}$ is the unit normal vector to D_r so that it is normal to the manifold, however here we can consider any manifold, so we may consider any unit vector $\mathbf{n}$ with its corresponding disc. On the right hand side, since the field is conservative the integral on any closed loop is zero so the right side is zero. Since this holds for any disc we get that $\operatorname{curl}(\mathbf{F})(a) \cdot \mathbf{n} = 0$ for all unit vectors $\mathbf{n} \in S^{n-1}$, and thus in particular for $\mathbf{n}$ in the direction of the curl we get

$$\|\operatorname{curl}(\mathbf{F})(a)\| = \operatorname{curl}(\mathbf{F})(a) \cdot \mathbf{n} = 0,$$

so $\operatorname{curl}(\mathbf{F})(a) = 0$. This holds for an arbitrary a so $\operatorname{curl}(\mathbf{F}) = 0$, which completes the proof. $\square$

Example 25.4. Recall Example 15.4. Let $\mathbf{F}: \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}^3$ be the gravitational force that a point mass m located at a point described by $\mathbf{r}$ feels as a result of a point mass M located at the origin. We know that

$$\mathbf{F} = G \frac{Mm}{\|\mathbf{r}\|^3} (-\mathbf{r}).$$

Under the assumption that m, M are constant we can write

$$\mathbf{F}(x, y, z) = -K \frac{(x, y, z)}{\|(x, y, z)\|^3}$$

where K is a constant. Without loss of generality we may assume $K = 1$. In Example 15.4 we calculated that the work done by this field along the metal spring is

$$\int_{\text{spring}} \mathbf{F} \cdot d\boldsymbol{\ell} = \int_0^L \mathbf{F}(f(t)) \cdot f'(t) dt = \dots = \frac{1}{\sqrt{a^2 + b^2 L^2}} - \frac{1}{a} = \frac{1}{\|r_1\|} - \frac{1}{\|r_0\|}$$

where r_0, r_1 are the hodgeting and end points of the path. We will now show that $\mathbf{F}$ is conservative. We can verify that we would get the same result if we had calculated the work of the field along the straight line segment connecting r_0 and r_1 , but we don't need to do this because we can guess that the potential function is

$$f(x, y, z) := \frac{1}{\sqrt{x^2 + y^2 + z^2}} = \frac{1}{r}.$$

We see that

$$\frac{\partial f}{\partial x} = \frac{-x}{r^3} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{-y}{r^3} \quad \text{and} \quad \frac{\partial f}{\partial z} = \frac{-z}{r^3}$$

and thus $\nabla f = \mathbf{F}$, so the field is conservative.

Remark 25.8. The same formula with different constants defines the electric field generated by a point charge located at the origin.

Definition 25.3 (Locally conservative field). Let $U \subseteq \mathbb{R}^n$ be an open set and let $\mathbf{F} \in C^1(U, \mathbb{R}^n)$ be a vector field. The field $\mathbf{F}$ is said to be locally conservative if for any point $p \in U$ there exists a neighbourhood of p where $\mathbf{F}$ is conservative.

Remark 25.9. The following proposition and theorem require introducing the notions of homotopy and a simply connected set. These are basic topological definitions which I would rather not introduce here.

Proposition 25.4. Let γ_0 and γ_1 be homotopic in an open set U , and let $\mathbf{F} \in C^1(U, \mathbb{R}^n)$ be a vector field so that $\frac{\partial F_i}{\partial x_j}(p) = \frac{\partial F_j}{\partial x_i}(p)$. Then

$$\int_{\gamma_0} \mathbf{F} \cdot \mathbf{T} ds = \int_{\gamma_1} \mathbf{F} \cdot \mathbf{T} ds.$$

Proposition 25.5. Let $U \subseteq \mathbb{R}^n$ be an open set and let $\mathbf{F} \in C^1(U, \mathbb{R}^n)$ be a vector field. then $\mathbf{F}$ is locally conservative if and only if for any point $p \in U$ and any i, j we have

$$\frac{\partial F_i}{\partial x_j}(p) = \frac{\partial F_j}{\partial x_i}(p).$$

Proof. Suppose that $\mathbf{F}$ is locally conservative in U . Then by Theorem 25.3 it satisfies

$$\frac{\partial F_i}{\partial x_j}(p) = \frac{\partial F_j}{\partial x_i}(p)$$

for all $p \in U$ and for all indices $1 \leq i, j \leq n$.

Next suppose that $\mathbf{F}$ satisfies

$$\frac{\partial F_i}{\partial x_j}(p) = \frac{\partial F_j}{\partial x_i}(p)$$

for all $p \in U$ and for all indices $1 \leq i, j \leq n$. We know by Theorem 25.4 that integrals of $\mathbf{F}$ on homotopic curves are equal, and since for an open ball B centered at any $p \in U$ any two curves are homotopic (open balls are simply connected), we obtain that the value of an integral of $\mathbf{F}$ over any closed curve is 0, which is equivalent to saying that $\mathbf{F}$ is conservative in B , so $\mathbf{F}$ is locally conservative. $\square$

Theorem 25.6. *Let U be simply connected region. Then any locally conservative field $\mathbf{F}$ in U is conservative in U . In particular, in 3 dimensions, if $\text{curl}(\mathbf{F}) = \nabla \times \mathbf{F} = 0$ at any point of a simply connected region U , then $\mathbf{F}$ is conservative.*

Proof. It suffices to prove that any locally conservative field is conservative when U is simply connected. Therefore, it suffices to show that for any closed curve $\gamma: [a, b] \rightarrow U$ we have that $\int_\gamma \mathbf{F} d\mathbf{r} = 0$. Let $\gamma: [a, b] \rightarrow U$ be a closed curve in U . Then since U is simply connected γ is homotopic to the constant loop at its basepoint. Therefore by Theorem 25.4 we get

$$\int_\gamma \mathbf{F} d\mathbf{r} = \int_{\{\gamma(a)\}} \mathbf{F} d\mathbf{r} = 0,$$

which completes the proof. $\square$

Example 25.5. Let $\omega \in \Omega^1(\mathbb{R}^2 \setminus \{0\})$ be the classic 1-form given by

$$\omega = -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy.$$

Its corresponding vector field is $\mathbf{F} = (-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2})$, defined on $U = \mathbb{R}^2 \setminus \{0\}$. As we saw in previous examples like Example 23.6, $\mathbf{F}$ is conservative on $\mathbb{R}^2 \setminus A$ but not on $\mathbb{R}^2 \setminus \{0\}$. We can see, however, that it is locally conservative, for example by computing that

$$\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x} = \frac{y^2 - x^2}{(x^2 + y^2)^2},$$

and applying Theorem 25.5. Now by Theorem 25.6 we can deduce that $\mathbb{R}^2 \setminus \{0\}$ is not simply connected.

26 The Hodge star operator

The purpose of this short section is to introduce the notion of the Hodge star operator. It is mostly based on the last part of the Youtube series on differential forms by Michael Penn.

26.1 A different view on differential operators

Remark 26.1. In Definition 25.1 and later in Definition 25.2 we defined the curl and divergence of vector fields, to be a vector field and a function that correspond to a 2-form and a 3-form respectively. Generally, the correspondences we are familiar with and accept is the correspondence between a 1-form $\omega = Pdx + Qdy + Rdz$ and a vector field $\mathbf{F} = (P, Q, R)$, and the correspondence between functions and 0-forms. This gives us a clue that there may be some operator that we can use to map the 2-forms and 3-forms to 1-forms and 0-forms, so that the correspondence with curl and divergence is natural.

Remark 26.2. Recall that $\bigwedge^m(\mathbb{R}^n)^*$ is the vector space of all m -forms on $\mathbb{R}^n$. Recall that we also have some abstract notion of an exterior algebra $\bigwedge^*(\mathbb{R}^n)^*$ which is the direct sum of all $\bigwedge^m(\mathbb{R}^n)^*$ for all natural m . We also have that $\dim \bigwedge^m(\mathbb{R}^n)^* = \binom{n}{m} = \binom{n}{n-m} = \dim \bigwedge^{n-m}(\mathbb{R}^n)^*$. This hints that there may be a connection between elements in $\bigwedge^m(\mathbb{R}^n)^*$ and elements in $\bigwedge^{n-m}(\mathbb{R}^n)^*$, and indeed there is.

Remark 26.3. Recall that for a multi-index $I = (i_1, \dots, i_m)$ we have that $dx_I = dx_{i_1} \wedge \dots \wedge dx_{i_m}$.

Definition 26.1 (Hodge star operator). Let $\star: \bigwedge^m(\mathbb{R}^n)^* \rightarrow \bigwedge^{n-m}(\mathbb{R}^n)^*$ be the operation so that for any $dx_I \in \bigwedge^m(\mathbb{R}^n)^*$ the element $\star dx_I \in \bigwedge^{n-m}(\mathbb{R}^n)^*$ satisfies $dx_I \wedge \star dx_I = dx_1 \wedge \dots \wedge dx_n$, which is the standard volume form on $\bigwedge^n(\mathbb{R}^n)^*$. Then for it to be defined on $\bigwedge^m(\mathbb{R}^n)^*$ we extend it linearly.

Example 26.1. When $n = 2$ and $m = 0$ we have that $\star 1 = dx \wedge dy$, and when $n = 2$ and $m = 1$ we have that $\star dx = dy$ and $\star dy = -dx$.

Example 26.2. Consider $\star: \bigwedge^2(\mathbb{R}^4)^* \rightarrow \bigwedge^2(\mathbb{R}^4)^*$. We then have that

$$\begin{aligned} \star(2dx_1 \wedge dx_3 + 11dx_2 \wedge dx_3) &= 2\star(dx_1 \wedge dx_3) + 11\star(dx_2 \wedge dx_3) \\ &= 2dx_4 \wedge dx_2 + 11\star(dx_2 \wedge dx_3) \\ &= 2dx_4 \wedge dx_2 + 11dx_1 \wedge dx_4. \end{aligned}$$

Remark 26.4. We identify vector fields $\mathbf{F} = (P, Q, R)$ with 1-forms $\omega = Pdx + Qdy + Rdz$ and vice versa. We have already seen that for a 0-form f we have $\text{grad}(f) = \nabla f$, which is identified nicely with the 1-form df . We will now use the Hodge star operator to find similar expressions which are identified with $\text{curl}(\mathbf{F})$ and $\text{div}(\mathbf{F})$.

Proposition 26.1. Let $\mathbf{F}$ be a vector field and $\omega_{\mathbf{F}}$ its corresponding 1-form. Then we have the identification $\text{curl}(\mathbf{F}) = \star d\omega_{\mathbf{F}}$.

Proof. By using the computation from Definition 25.1 we can compute that

$$\star d\omega_{\mathbf{F}} = \star \left(\left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) dx_3 \wedge dx_1 + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) dx_1 \wedge dx_2 \right)$$

Then by linearity of the Hodge star operator we get the expression:

$$\left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) \star(dx_2 \wedge dx_3) + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) \star(dx_3 \wedge dx_1) + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) \star(dx_1 \wedge dx_2),$$

and since $\star(dx_2 \wedge dx_3) = dx_1$, $\star(dx_3 \wedge dx_1) = dx_2$, and $\star(dx_1 \wedge dx_2) = dx_3$ we get

$$\star d\omega_{\mathbf{F}} = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) dx_1 + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) dx_2 + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) dx_3,$$

which indeed corresponds exactly to $\text{curl}(\mathbf{F})$. $\square$

Proposition 26.2. *Let $\mathbf{F}$ be a vector field and $\omega_{\mathbf{F}}$ its corresponding 1-form. Then we have the identification $\text{div}(\mathbf{F}) = \star d\star\omega_{\mathbf{F}}$.*

Proof. This identification is between 0-forms and functions, not 1-forms and vector fields. Indeed, like in the previous proposition, and by Definition 25.2 we get

$$\begin{aligned} \star d\star\omega_{\mathbf{F}} &= \star d\star(Pdx_1 + Qdx_2 + Rdx_3) \\ &= \star d(Pdx_2 \wedge dx_3 + Qdx_3 \wedge dx_1 + Rdx_1 \wedge dx_2) \\ &= \star \left(\frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3 \\ &= \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}, \end{aligned}$$

which corresponds precisely to $\text{div}(\mathbf{F})$. $\square$

Remark 26.5. Recall Remark 25.3 and Remark 25.6. It might not have been very clear why these follow from $d^2 = 0$, but now that we have Theorem 26.2 and Theorem 26.1, it should be very clear since

$$\begin{aligned} \text{curl}(\text{grad } f) &= \star d d f = \star(d^2 f) = 0, \\ \text{div}(\text{curl } \mathbf{F}) &= \star d \star d \omega_{\mathbf{F}} = \star(d^2 \omega_{\mathbf{F}}) = 0, \end{aligned}$$

where we must notice that $\star$ is an involution, meaning that it is the inverse of itself.

26.2 The general Hodge star operator

Definition 26.2 (Hodge star operator). Let $\langle \cdot, \cdot \rangle : (\bigwedge^k(\mathbb{R}^n))^* \rightarrow \mathbb{R}$ be a nondegenerate symmetric bilinear form (we may think of it as a scalar product, but it isn't always positive-definite) defined for all $0 \leq k \leq n$. Then for $\alpha \in \bigwedge^m(\mathbb{R}^n)^*$ we define $\star\alpha$ to be the unique $(n-m)$ -form such that for all $\beta \in \bigwedge^{n-m}(\mathbb{R}^n)^*$ we have $\beta \wedge (\star\alpha) = \langle \alpha, \beta \rangle dx_1 \wedge \cdots \wedge dx_n$.

Remark 26.6. We will call the bilinear form $\langle \cdot, \cdot \rangle$ the “scalar product”.

Example 26.3. Consider the scalar product $\langle dx_I, dx_J \rangle$ defined in the following way:

$$\langle dx_I, dx_J \rangle := \begin{cases} 1, & I = J, \\ 0, & I \neq J. \end{cases}$$

With respect to this bilinear form, Definition 26.2 coincides with Definition 26.1, and therefore Definition 26.2 is indeed a generalization of the Hodge star.

Remark 26.7. Note that in general for two 1-forms in $\bigwedge^1(\mathbb{R}^2)^*$ we have

$$\begin{aligned} \langle Adx + Bdy, Cdx + Ddy \rangle &= AC \langle dx, dx \rangle + AD \langle dx, dy \rangle + BC \langle dy, dx \rangle + BD \langle dy, dy \rangle \\ &= \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} \langle dx, dx \rangle & \langle dx, dy \rangle \\ \langle dy, dx \rangle & \langle dy, dy \rangle \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix}. \end{aligned}$$

This shows that the scalar product is determined completely and uniquely on $\bigwedge^1(\mathbb{R}^2)^*$ by the Gram matrix $G(dx, dy)$. In fact it is true in general that $\langle \cdot, \cdot \rangle$ is determined on $\bigwedge^1(\mathbb{R}^n)^*$ by the value of

$$G(dx_1, \dots, dx_n) = \begin{pmatrix} \langle dx_1, dx_1 \rangle & \cdots & \langle dx_1, dx_n \rangle \\ \vdots & \ddots & \vdots \\ \langle dx_n, dx_1 \rangle & \cdots & \langle dx_n, dx_n \rangle \end{pmatrix}.$$

We would now like to show that the scalar product is determined on $\bigwedge^k(\mathbb{R}^n)^*$ for all $0 \leq k \leq n$.

Remark 26.8. The scalar product on 0-forms acts as standard multiplication.

Definition 26.3 (Scalar product on higher forms). Let $\langle \cdot, \cdot \rangle$ be the scalar product on some exterior algebra of forms $\bigwedge^*(\mathbb{R}^n)^*$, and suppose we know the value of $G(dx_1, \dots, dx_n)$ (which determines its value on $\bigwedge^1(\mathbb{R}^n)^*$). For multi-indices $I = (i_1, \dots, i_m)$, $J = (j_1, \dots, j_m)$, we define:

$$\langle dx_I, dx_J \rangle = \sum_{\sigma \in S_m} \text{sgn}(\sigma) \prod_{l=1}^m \langle dx_{i_l}, dx_{j_{\sigma(l)}} \rangle,$$

where S_m is the group of symmetries of m elements.

Exercise 26.1. Let $\langle \cdot, \cdot \rangle : \bigwedge^1(\mathbb{R}^3)^* \times \bigwedge^3(\mathbb{R}^n)^* \rightarrow \mathbb{R}$ be given by the matrix

$$G(dx_1, dx_2, dx_3) = \begin{pmatrix} 1 & 3 & -2 \\ 3 & 2 & 0 \\ -2 & 0 & -1 \end{pmatrix}.$$

Find the value of $\langle dx_1 \wedge dx_2, dx_2 \wedge dx_3 \rangle$.

Solution. Note that $m = 2$, so $I = (1, 2)$, $J = (2, 3)$, and $S_2 = \{(1), (12)\}$. Also note that $\text{sgn}(1) = 1$ and $\text{sgn}(12) = -1$. By definition we have that

$$\langle dx_1 \wedge dx_2, dx_2 \wedge dx_3 \rangle = \langle dx_1, dx_2 \rangle \langle dx_2, dx_3 \rangle - \langle dx_1, dx_3 \rangle \langle dx_2, dx_2 \rangle.$$

Now by using the values given by the Gram matrix, we conclude that

$$\langle dx_1 \wedge dx_2, dx_2 \wedge dx_3 \rangle = 3 \times 0 - (-2) \times 2 = 4.$$

Remark 26.9. One of the interpretations of the Hodge star operator, is that for an m -hyperplane ω , it returns the orthogonal complementing $(n - m)$ -hyperplane η . Moreover, it returns the orthogonal complementing hyperplane in the exact orientation so that $\omega \wedge \eta$ is the form that gives the standard orientation on $(\mathbb{R}^n)^*$. For example let $n = 2$ and $m = 1$, and let $\star$ be induced by the scalar product from Example 26.3. If we consider dx and dy as the unit vectors and the orthonormal basis of the vector space $(\mathbb{R}^2)^*$, then for $dx + dy$ we would expect to get $-dx + dy$, and as we expect:

$$\star(dx + dy) = \star dx + \star dy = dy - dx,$$

and moreover, $(dx + dy) \wedge (-dx + dy) = 2dx \wedge dy$, which indeed is in the standard orientation choice of $(\mathbb{R}^2)^*$.

Example 26.4 (The Minkowski space). We will not talk about the Minkowski space in depth here, but we will talk about the Minkowski bilinear form, which as we know defines a Hodge star operator. It is simply given by a Gram matrix

$$G(dt, dx, dy, dz) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

In this case the Hodge star operator can describe Maxwell's equations from physics in a very natural and elegant way, but this is not a physics class, so we will not get into that.

Remark 26.10 (Final remark). As a final remark, unrelated to this section, I want to address the elephant in the room. We have discussed integrals over differential forms, and we mentioned that any integral we have seen in this course can be described as an integral over a differential form. However, we know that type one integrals do not change their signs when we change the orientation of the manifold we integrate over, in clear contradiction to Theorem 22.1. In the limited context of these notes, we will explain this by saying that the form we integrate over when we compute an integral of the first type completely depends on the orientation of the manifold, but this is not the full picture. It solves the problem since if a manifold has “negative” orientation we can give the form a negative sign, and if it has a “positive” orientation we could give it a positive sign, and then get the same result, which would explain why an integral of the first type “does not depend on orientation”.