Vector Calculus
Frankenstein's Note

Universidade Federal do ABC
Santo André
Version 0.9
05/2017

Written in LaTeX.

These notes were written based on and using excerpts from the book "Multivariable and Vector Calculus" by David Santos and includes excerpts from "Vector Calculus" by Michael Corral, from "Linear Algebra via Exterior Products" by Sergei Winitzki, "Linear Algebra" by David Santos and from "Introduction to Tensor Calculus" by Taha Sochi.

These books are also available under the terms of the GNU Free Documentation License.

History

These notes are based on the LaTeX source of the book “Multivariable and Vector Calculus”, which has undergone profound changes over time. In particular some examples and figures from “Vector Calculus” by Michael Corral have been added. The tensor part is based on “Linear algebra via exterior products” by Sergei Winitzki and on “Introduction to Tensor Calculus” by Taha Sochi.

What made possible the creation of these notes was the fact that these four books available are under the terms of the GNU Free Documentation License.

Second Version

This version was released 05/2017.

In this versions a lot of efforts were made to transform the notes into a more coherent text.

First Version

This version was released 02/2017.

The first version of the notes.

Contents

13 Appendix

Differential Vector Calculus

Chapter 1

Multidimensional Vectors

1.1 Vectors Space

In this section we introduce an algebraic structure for $\mathbb{R}^n$, the vector space in n-dimensions.

We assume that you are familiar with the geometric interpretation of members of $\mathbb{R}^2$ and $\mathbb{R}^3$ as the rectangular coordinates of points in a plane and three-dimensional space, respectively.

Although $\mathbb{R}^n$ cannot be visualized geometrically if $n \geqslant 4$, geometric ideas from $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$ often help us to interpret the properties of $\mathbb{R}^n$ for arbitrary n.

Definition 1.1.1 The n-dimensional space, $\mathbb{R}^n$, is defined as the set

$$\mathbb{R}^n = \left\{(x_1, x_2, \ldots, x_n) : x_k \in \mathbb{R}\right\}.$$

Elements $\mathbf{v} \in \mathbb{R}^n$ will be called **vectors** and will be written in boldface $\mathbf{v}$. In the blackboard the vectors generally are written with an arrow $\vec{v}$.

Definition 1.1.2 If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathbb{R}^n$ their **vector sum** $\mathbf{x}+\mathbf{y}$ is defined by the coordinatewise addition

$$\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n). \tag{1.1}$$

Note that the symbol "+" has two distinct meanings in (1.1): on the left, "+" stands for the newly defined addition of members of $\mathbb{R}^n$ and, on the right, for the usual addition of real numbers.

The vector with all components 0 is called the **zero vector** and is denoted by $\mathbf{0}$. It has the property that $\mathbf{v}+\mathbf{0} = \mathbf{v}$ for every vector $\mathbf{v}$; in other words, $\mathbf{0}$ is the identity element for vector addition.

Definition 1.1.3 A real number $\lambda \in \mathbb{R}$ will be called a **scalar**. If $\lambda \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$ we define **scalar multiplication** of a vector and a scalar by the coordinatewise multiplication

$$\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n). \tag{1.2}$$

The space $\mathbb{R}^n$ with the operations of sum and scalar multiplication defined above will be called n dimensional vector space.

The vector $(-1)\mathbf{x}$ is also denoted by $-\mathbf{x}$ and is called the **negative** or **opposite** of $\mathbf{x}$

We leave the proof of the following theorem to the reader

Theorem 1.1.1 If $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{y}$ are in $\mathbb{R}^n$ and λ, λ_1 and λ_2 are real numbers, then

❶ $\mathbf{x}+\mathbf{z} = \mathbf{z}+\mathbf{x}$ (vector addition is commutative).

❷ $(\mathbf{x}+\mathbf{z})+\mathbf{y} = \mathbf{x}+(\mathbf{z}+\mathbf{y})$ (vector addition is associative).

❸ There is a unique vector $\mathbf{0}$, called the zero vector, such that $\mathbf{x}+\mathbf{0} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

❹ For each $\mathbf{x}$ in $\mathbb{R}^n$ there is a unique vector $-\mathbf{x}$ such that $\mathbf{x}+(-\mathbf{x}) = \mathbf{0}$.

❺ $\lambda_1(\lambda_2\mathbf{x}) = (\lambda_1\lambda_2)\mathbf{x}$.

❻ $(\lambda_1+\lambda_2)\mathbf{x} = \lambda_1\mathbf{x}+\lambda_2\mathbf{x}$.

❼ $\lambda(\mathbf{x}+\mathbf{z}) = \lambda\mathbf{x}+\lambda\mathbf{z}$.

❽ $1\mathbf{x} = \mathbf{x}$.

Clearly, $\mathbf{0} = (0,0,\ldots,0)$ and, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, then

$$-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n).$$

We write $\mathbf{x}+(-\mathbf{z})$ as $\mathbf{x}-\mathbf{z}$. The vector $\mathbf{0}$ is called the **origin**.

In a more general context, a nonempty set V, together with two operations $+, \cdot$ is said to be a **vector space** if it has the properties listed in Theorem 1.1.1. The members of a vector space are called **vectors**.

When we wish to note that we are regarding a member of $\mathbb{R}^n$ as part of this algebraic structure, we will speak of it as a vector; otherwise, we will speak of it as a point.

Definition 1.1.4 The **canonical ordered basis** for $\mathbb{R}^n$ is the collection of vectors

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n,\}$$

with

$$\mathbf{e}_k = \underbrace{(0, \ldots, 1, \ldots, 0)}_{\text{a 1 in the } k \text{ slot and 0's everywhere else}}.$$

Observe that

$$\sum_{k=1}^{n} v_k \mathbf{e}_k = (v_1, v_2, \ldots, v_n). \tag{1.3}$$

This means that any vector can be written as sums of scalar multiples of the standard basis. We will discuss this fact more deeply in the next section.

Definition 1.1.5 Let $\mathbf{a}, \mathbf{b}$ be distinct points in $\mathbb{R}^n$ and let $\mathbf{x} = \mathbf{b} - \mathbf{a} \neq \mathbf{0}$,. The **parametric line** passing through $\mathbf{a}$ in the direction of $\mathbf{x}$ is the set

$$\{\mathbf{r} \in \mathbb{R}^n : \mathbf{r} = \mathbf{a} + t\mathbf{x}\}.$$

Example 1.1.2 Find the parametric equation of the line passing through the points $(1, 2, 3)$ and $(-2, -1, 0)$.

Solution: ► The line follows the direction

$$\left(1 - (-2), 2 - (-1), 3 - 0\right) = (3, 3, 3).$$

The desired equation is

$$(x, y, z) = (1, 2, 3) + t(3, 3, 3).$$

Equivalently

$$(x, y, z) = (-2, -1, 2) + t(3, 3, 3).$$

◄

Length, Distance, and Inner Product

Definition 1.1.6 Given vectors $\mathbf{x}, \mathbf{y}$ of $\mathbb{R}^n$, their **inner product** or **dot product** is defined as

$$\mathbf{x} \bullet \mathbf{y} = \sum_{k=1}^{n} x_k y_k.$$

Theorem 1.1.3 For $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, and α and β real numbers, we have>

❶ $(\alpha\mathbf{x} + \beta\mathbf{y}) \bullet \mathbf{z} = \alpha(\mathbf{x} \bullet \mathbf{z}) + \beta(\mathbf{y} \bullet \mathbf{z})$

❷ $\mathbf{x} \bullet \mathbf{y} = \mathbf{y} \bullet \mathbf{x}$

❸ $\mathbf{x} \bullet \mathbf{x} \geqslant 0$

❹ $\mathbf{x} \bullet \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$

The proof of this theorem is simple and will be left as exercise for the reader.

The **norm** or **length** of a vector $\mathbf{x}$, denoted as $\|\mathbf{x}\|$, is defined as

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \bullet \mathbf{x}}$$

Definition 1.1.7 Given vectors $\mathbf{x}, \mathbf{y}$ of $\mathbb{R}^n$, their **distance** is

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{(\mathbf{x} - \mathbf{y}) \bullet (\mathbf{x} - \mathbf{y})} = \sum_{i=1}^{n} (x_i - y_i)^2$$

If $n = 1$, the previous definition of length reduces to the familiar absolute value, for $n = 2$ and $n = 3$, the length and distance of Definition 1.1.7 reduce to the familiar definitions for the two and three dimensional space.

Definition 1.1.8 A vector $\mathbf{x}$ is called **unit vector**

$$\|\mathbf{x}\| = 1.$$

Definition 1.1.9 Let $\mathbf{x}$ be a non-zero vector, then the associated **versor** (or normalized vector) denoted $\hat{\mathbf{x}}$ is the unit vector

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

We now establish one of the most useful inequalities in analysis.

Theorem 1.1.4 — Cauchy-Bunyakovsky-Schwarz Inequality. Let $\mathbf{x}$ and $\mathbf{y}$ be any two vectors in $\mathbb{R}^n$. Then we have

$$|\mathbf{x} \bullet \mathbf{y}| \leqslant \|\mathbf{x}\| \|\mathbf{y}\|.$$

Proof: Since the norm of any vector is non-negative, we have

$$\begin{aligned}
\|\mathbf{x}+t\mathbf{y}\| \geqslant 0 &\iff (\mathbf{x}+t\mathbf{y})\bullet(\mathbf{x}+t\mathbf{y}) \geqslant 0 \\
&\iff \mathbf{x}\bullet\mathbf{x}+2t\mathbf{x}\bullet\mathbf{y}+t^2\mathbf{y}\bullet\mathbf{y} \geqslant 0 \\
&\iff \|\mathbf{x}\|^2+2t\mathbf{x}\bullet\mathbf{y}+t^2\|\mathbf{y}\|^2 \geqslant 0.
\end{aligned}$$

This last expression is a quadratic polynomial in t which is always non-negative. As such its discriminant must be non-positive, that is,

$$(2\mathbf{x}\bullet\mathbf{y})^2-4(\|\mathbf{x}\|^2)(\|\mathbf{y}\|^2) \leqslant 0 \iff |\mathbf{x}\bullet\mathbf{y}| \leqslant \|\mathbf{x}\|\|\mathbf{y}\|,$$

giving the theorem. ■

The Cauchy-Bunyakovsky-Schwarz inequality can be written as

$$\left|\sum_{k=1}^{n} x_k y_k\right| \leqslant \left(\sum_{k=1}^{n} x_k^2\right)^{1/2}\left(\sum_{k=1}^{n} y_k^2\right)^{1/2}, \tag{1.4}$$

for real numbers x_k, y_k.

Theorem 1.1.5 — Triangle Inequality. Let $\mathbf{x}$ and $\mathbf{y}$ be any two vectors in $\mathbb{R}^n$. Then we have

$$\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\|+\|\mathbf{y}\|.$$

Proof:

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\bullet(\mathbf{x}+\mathbf{y}) \\
&= \mathbf{x}\bullet\mathbf{x}+2\mathbf{x}\bullet\mathbf{y}+\mathbf{y}\bullet\mathbf{y} \\
&\leqslant \|\mathbf{x}\|^2+2\|\mathbf{x}\|\|\mathbf{y}\|+\|\mathbf{y}\|^2 \\
&= (\|\mathbf{x}\|+\|\mathbf{y}\|)^2,
\end{aligned}$$

from where the desired result follows. ■

Corollary 1.1.6 If $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are in $\mathbb{R}^n$, then

$$|\mathbf{x}-\mathbf{y}| \leqslant |\mathbf{x}-\mathbf{z}|+|\mathbf{z}-\mathbf{y}|.$$

Proof: Write

$$\mathbf{x}-\mathbf{y} = (\mathbf{x}-\mathbf{z})+(\mathbf{z}-\mathbf{y}),$$

and apply Theorem 1.1.5. ■

Definition 1.1.10 Let $\mathbf{x}$ and $\mathbf{y}$ be two non-zero vectors in $\mathbb{R}^n$. Then the angle $\widehat{(\mathbf{x},\mathbf{y})}$ between them is given by the relation

$$\cos\widehat{(\mathbf{x},\mathbf{y})} = \frac{\mathbf{x}\bullet\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

This expression agrees with the geometry in the case of the dot product for $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 1.1.11 Let $\mathbf{x}$ and $\mathbf{y}$ be two non-zero vectors in $\mathbb{R}^n$. This vectors are said orthogonal if the angle between them is 90 degrees. Equivalently, if: $\mathbf{x}\bullet\mathbf{y} = 0$

Let $P_0 = (p_1, p_2, \ldots, p_n)$, and $\mathbf{n} = (n_1, n_2, \ldots, n_n)$ be a nonzero vector.

Definition 1.1.12 The hyperplane defined by the point P_0 and the vector $\mathbf{n}$ is defined as the set of points $P : (x_1, , x_2, \ldots, x_n) \in \mathbb{R}^n$, such that the vector drawn from P_0 to P is perpendicular to $\mathbf{n}$.

$$\mathbf{n}\bullet(\mathbf{P} - \mathbf{P}_0) = 0.$$

Recalling that two vectors are perpendicular if and only if their dot product is zero, it follows that the desired hyperplane can be described as the set of all points P such that

$$\mathbf{n}\bullet(\mathbf{P} - \mathbf{P}_0) = 0.$$

Expanded this becomes

$$n_1(x_1 - p_1) + n_2(x_2 - p_2) + \cdots + n_n(x_n - p_n) = 0,$$

which is the point-normal form of the equation of a hyperplane. This is just a linear equation

$$n_1x_1 + n_2x_2 + \cdots n_nx_n + d = 0,$$

where

$$d = -(n_1p_1 + n_2p_2 + \cdots + n_np_n).$$

1.2 Basis and Change of Basis

1.2.1 Linear Independence and Spanning Sets

Definition 1.2.1 Let $\lambda_i \in \mathbb{R}, 1 \leqslant i \leqslant n$. Then the vectorial sum

$$\sum_{j=1}^{n} \lambda_j \mathbf{x}_j$$

is said to be a **linear combination** of the vectors $\mathbf{x}_i \in \mathbb{R}^n, 1 \leqslant i \leqslant n$.

Definition 1.2.2 The vectors $\mathbf{x}_i \in \mathbb{R}^n, 1 \leqslant i \leqslant n$, are **linearly dependent** or **tied** if

$$\exists (\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbb{R}^n \setminus \{\mathbf{0}\} \quad \text{such that} \quad \sum_{j=1}^{n} \lambda_j \mathbf{x}_j = \mathbf{0},$$

that is, if there is a non-trivial linear combination of them adding to the zero vector.

Definition 1.2.3 The vectors $\mathbf{x}_i \in \mathbb{R}^n, 1 \leqslant i \leqslant n$, are **linearly independent** or **free** if they are not linearly dependent. That is, if $\lambda_i \in \mathbb{R}, 1 \leqslant i \leqslant n$ then

$$\sum_{j=1}^{n} \lambda_j \mathbf{x}_j = \mathbf{0} \implies \lambda_1 = \lambda_2 = \cdots = \lambda_n = 0.$$

A family of vectors is linearly independent if and only if the only linear combination of them giving the zero-vector is the trivial linear combination.

Example 1.2.1

$$\{(1,2,3), (4,5,6), (7,8,9)\}$$

is a tied family of vectors in $\mathbb{R}^3$, since

$$(1)(1,2,3) + (-2)(4,5,6) + (1)(7,8,9) = (0,0,0).$$

Definition 1.2.4 A family of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots,\} \subseteq \mathbb{R}^n$ is said to **span** or **generate** $\mathbb{R}^n$ if every $\mathbf{x} \in \mathbb{R}^n$ can be written as a linear combination of the $\mathbf{x}_j$'s.

Example 1.2.2 Since

$$\sum_{k=1}^{n} v_k \mathbf{e}_k = (v_1, v_2, \ldots, v_n). \tag{1.5}$$

This means that the canonical basis generate $\mathbb{R}^n$.

Theorem 1.2.3 If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots,\} \subseteq \mathbb{R}^n$ spans $\mathbb{R}^n$, then any superset

$$\{\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots,\} \subseteq \mathbb{R}^n$$

also spans $\mathbb{R}^n$.

Proof: This follows at once from

$$\sum_{i=1}^{l} \lambda_i \mathbf{x}_i = 0\mathbf{y} + \sum_{i=1}^{l} \lambda_i \mathbf{x}_i.$$

■

Example 1.2.4 The family of vectors

$$\{\mathbf{i} = (1,0,0), \mathbf{j} = (0,1,0), \mathbf{k} = (0,0,1)\}$$

spans $\mathbb{R}^3$ since given $(a,b,c) \in \mathbb{R}^3$ we may write

$$(a,b,c) = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Example 1.2.5 Prove that the family of vectors

$$\{\mathbf{t}_1 = (1,0,0), \mathbf{t}_2 = (1,1,0), \mathbf{t}_3 = (1,1,1)\}$$

spans $\mathbb{R}^3$.

Solution: ► This follows from the identity

$$(a,b,c) = (a-b)(1,0,0) + (b-c)(1,1,0) + c(1,1,1) = (a-b)\mathbf{t}_1 + (b-c)\mathbf{t}_2 + c\mathbf{t}_3.$$

◄

1.2.2 Basis

Definition 1.2.5 A family $E = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots\} \subseteq \mathbb{R}^n$ is said to be a **basis** of $\mathbb{R}^n$ if

❶ are linearly independent,

❷ they span $\mathbb{R}^n$.

Example 1.2.6 The family

$$\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0),$$

where there is a 1 on the i-th slot and 0's on the other $n-1$ positions, is a basis for $\mathbb{R}^n$.

Theorem 1.2.7 All basis of $\mathbb{R}^n$ have the same number of vectors.

Definition 1.2.6 The **dimension** of $\mathbb{R}^n$ is the number of elements of any of its basis, n.

Theorem 1.2.8 Let $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ be a family of vectors in $\mathbb{R}^n$. Then the $\mathbf{x}$'s form a basis if and only if the $n\times n$ matrix A formed by taking the $\mathbf{x}$'s as the columns of A is invertible.

Proof: Since we have the right number of vectors, it is enough to prove that the $\mathbf{x}$'s are linearly independent. But if $X=(\lambda_1,\lambda_2,\ldots,\lambda_n)$, then

$$\lambda_1\mathbf{x}_1+\cdots+\lambda_n\mathbf{x}_n = AX.$$

If A is invertible, then $AX=\mathbf{0}_n \implies X=A^{-1}\mathbf{0}=\mathbf{0}$, meaning that $\lambda_1=\lambda_2=\cdots\lambda_n=0$, so the $\mathbf{x}$'s are linearly independent.

The reciprocal will be left as a exercise. ■

Definition 1.2.7

❶ A basis $E=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k\}$ of vectors in R^n is called **orthogonal** if

$$\mathbf{x}_i\bullet\mathbf{x}_j=0$$

for all $i\neq j$.

❷ An orthogonal basis of vectors is called **orthonormal** if all vectors in E are unit vectors, i.e, have norm equal to 1.

1.2.3 Coordinates

Theorem 1.2.9 Let $E=\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$ be a basis for a vector space $\mathbb{R}^n$. Then any $\mathbf{x}\in\mathbb{R}^n$ has a unique representation

$$\mathbf{x}=a_1\mathbf{e}_1+a_2\mathbf{e}_2+\cdots+a_n\mathbf{e}_n.$$

Proof: Let

$$\mathbf{x}=b_1\mathbf{e}_1+b_2\mathbf{e}_2+\cdots+y_n\mathbf{e}_n$$

be another representation of $\mathbf{x}$. Then

$$\mathbf{0} = (x_1 - b_1)\mathbf{e}_1 + (x_2 - b_2)\mathbf{e}_2 + \cdots + (x_n - y_n)\mathbf{e}_n.$$

Since $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ forms a basis for $\mathbb{R}^n$, they are a linearly independent family. Thus we must have

$$x_1 - y_1 = x_2 - y_2 = \cdots = x_n - y_n = 0_{\mathbb{R}},$$

that is

$$x_1 = b_1; x_2 = b_2; \cdots ; x_n = y_n,$$

proving uniqueness. ■

Definition 1.2.8 An **ordered basis** $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of a vector space $\mathbb{R}^n$ is a basis where the order of the $\mathbf{x}_k$ has been fixed. Given an ordered basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of a vector space $\mathbb{R}^n$, Theorem 1.2.9 ensures that there are unique $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ such that

$$\mathbf{x} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n.$$

The a_k's are called the **coordinates** of the vector $\mathbf{x}$.

We will denote the coordinates the vector $\mathbf{x}$ on the basis E by

$$[\mathbf{x}]_E$$

or simply $[\mathbf{x}]$.

Example 1.2.10 The standard ordered basis for $\mathbb{R}^3$ is $E = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. The vector $(1, 2, 3) \in \mathbb{R}^3$ for example, has coordinates $(1, 2, 3)_E$. If the order of the basis were changed to the ordered basis $F = \{\mathbf{i}, \mathbf{k}, \mathbf{j}\}$, then $(1, 2, 3) \in \mathbb{R}^3$ would have coordinates $(1, 3, 2)_F$.

Usually, when we give a coordinate representation for a vector $\mathbf{x} \in \mathbb{R}^n$, we assume that we are using the standard basis.

Example 1.2.11 Consider the vector $(1, 2, 3) \in \mathbb{R}^3$ (given in standard representation). Since

$$(1, 2, 3) = -1(1, 0, 0) - 1(1, 1, 0) + 3(1, 1, 1),$$

under the ordered basis $E = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$, $(1, 2, 3)$ has coordinates $(-1, -1, 3)_E$. We write

$$(1, 2, 3) = (-1, -1, 3)_E.$$

Example 1.2.12 The vectors of

$$E = \{(1,1),(1,2)\}$$

are non-parallel, and so form a basis for $\mathbb{R}^2$. So do the vectors

$$F = \{(2,1),(1,-1)\}.$$

Find the coordinates of $(3,4)_E$ in the base F.

Solution: ► We are seeking x,y such that

$$3(1,1)+4(1,2) = x\begin{bmatrix}2\\1\end{bmatrix}+y\begin{bmatrix}1\\-1\end{bmatrix} \implies \begin{bmatrix}1&1\\1&2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix} = \begin{bmatrix}2&1\\1&-1\end{bmatrix}(x,y)_F.$$

Thus

$$\begin{aligned}(x,y)_F &= \begin{bmatrix}2&1\\1&-1\end{bmatrix}^{-1}\begin{bmatrix}1&1\\1&2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{3}&\frac{1}{3}\\\frac{1}{3}&-\frac{2}{3}\end{bmatrix}\begin{bmatrix}1&1\\1&2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}\\ &= \begin{bmatrix}\frac{2}{3}&1\\-\frac{1}{3}&-1\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}\\ &= \begin{bmatrix}6\\-5\end{bmatrix}_F.\end{aligned}$$

Let us check by expressing both vectors in the standard basis of $\mathbb{R}^2$:

$$(3,4)_E = 3(1,1)+4(1,2) = (7,11),$$

$$(6,-5)_F = 6(2,1)-5(1,-1) = (7,11).$$

◄

In general let us consider basis E , F for the same vector space $\mathbb{R}^n$. We want to convert X_E to Y_F. We let A be the matrix formed with the column vectors of E in the given order an B be the matrix formed with the column vectors of F in the given order. Both A and B are invertible matrices since the E, F are basis, in view of Theorem 1.2.8. Then we must have

$$AX_E = BY_F \implies Y_F = B^{-1}AX_E.$$

Also,

$$X_E = A^{-1}BY_F.$$

This prompts the following definition.

Definition 1.2.9 Let $E = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $F = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ be two ordered basis for a vector space $\mathbb{R}^n$. Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$ be the matrix having the $\mathbf{x}$'s as its columns and let $B \in \mathbf{M}_{n\times n}(\mathbb{R})$ be the matrix having the $\mathbf{y}$'s as its columns. The matrix $P = B^{-1}A$ is called the **transition matrix** from E to F and the matrix $P^{-1} = A^{-1}B$ is called the **transition matrix** from F to E.

Example 1.2.13 Consider the basis of $\mathbb{R}^3$

$$E = \left\{(1,1,1),(1,1,0),(1,0,0)\right\},$$

$$F = \left\{(1,1,-1),(1,-1,0),(2,0,0)\right\}.$$

Find the transition matrix from E to F and also the transition matrix from F to E. Also find the coordinates of $(1,2,3)_E$ in terms of F.

Solution: ► Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

The transition matrix from E to F is

$$\begin{aligned} P &= B^{-1}A \\ &= \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & -1 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 & 0 \\ -2 & -1 & -0 \\ 2 & 1 & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

The transition matrix from F to E is

$$P^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ -2 & -1 & 0 \\ 2 & 1 & \frac{1}{2} \end{bmatrix}^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ 2 & -1 & 0 \\ 0 & 2 & 2 \end{bmatrix}.$$

Now,

$$Y_F = \begin{bmatrix} -1 & 0 & 0 \\ -2 & -1 & 0 \\ 2 & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_E = \begin{bmatrix} -1 \\ -4 \\ \frac{11}{2} \end{bmatrix}_F.$$

As a check, observe that in the standard basis for $\mathbb{R}^3$

$$\begin{bmatrix} 1, 2, 3 \end{bmatrix}_E = 1 \begin{bmatrix} 1, 1, 1 \end{bmatrix} + 2 \begin{bmatrix} 1, 1, 0 \end{bmatrix} + 3 \begin{bmatrix} 1, 0, 0 \end{bmatrix} = \begin{bmatrix} 6, 3, 1 \end{bmatrix},$$

$$\begin{bmatrix} -1, -4, \frac{11}{2} \end{bmatrix}_F = -1 \begin{bmatrix} 1, 1, -1 \end{bmatrix} - 4 \begin{bmatrix} 1, -1, 0 \end{bmatrix} + \frac{11}{2} \begin{bmatrix} 2, 0, 0 \end{bmatrix} = \begin{bmatrix} 6, 3, 1 \end{bmatrix}.$$

◀

1.3 Linear Transformations and Matrices

Definition 1.3.1 A **linear transformation** or **homomorphism** between $\mathbb{R}^n$ and $\mathbb{R}^m$

$$L : \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^m \\ \mathbf{x} & \mapsto & L(\mathbf{x}) \end{array},$$

is a function which is

- **A**dditive: $L(\mathbf{x}+\mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$,
- **H**omogeneous: $L(\lambda\mathbf{x}) = \lambda L(\mathbf{x})$, for $\lambda \in \mathbb{R}$.

It is clear that the above two conditions can be summarized conveniently into

$$L(\mathbf{x}+\lambda\mathbf{y}) = L(\mathbf{x}) + \lambda L(\mathbf{y}).$$

Assume that $\{\mathbf{x}_i\}_{i\in[1;n]}$ is an ordered basis for $\mathbb{R}^n$, and $E = \{\mathbf{y}_i\}_{i\in[1;m]}$ an ordered basis for $\mathbb{R}^m$.

Then

$$
\begin{aligned}
L(\mathbf{x}_1) &= a_{11}\mathbf{y}_1 + a_{21}\mathbf{y}_2 + \cdots + a_{m1}\mathbf{y}_m &= \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}_E \\
L(\mathbf{x}_2) &= a_{12}\mathbf{y}_1 + a_{22}\mathbf{y}_2 + \cdots + a_{m2}\mathbf{y}_m &= \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}_E \\
\vdots \quad & \vdots \qquad \qquad \vdots & \vdots \quad \vdots \\
L(\mathbf{x}_n) &= a_{1n}\mathbf{y}_1 + a_{2n}\mathbf{y}_2 + \cdots + a_{mn}\mathbf{y}_m &= \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}_E
\end{aligned}
.
$$

Definition 1.3.2 The $m \times n$ matrix

$$
M_L = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{12} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}
$$

formed by the column vectors above is called the **matrix representation of the linear map** L **with respect to the basis** $\{\mathbf{x}_i\}_{i\in[1;m]}, \{\mathbf{y}_i\}_{i\in[1;n]}$.

Example 1.3.1 Consider $L : \mathbb{R}^3 \to \mathbb{R}^3$,

$$
L(x,y,z) = (x-y-z, x+y+z, z).
$$

Clearly L is a linear transformation.

1. Find the matrix corresponding to L under the standard ordered basis.

2. Find the matrix corresponding to L under the ordered basis $(1,0,0), (1,1,0), (1,0,1)$, for both the domain and the image of L.

Solution: ▶

1. The matrix will be a 3×3 matrix. We have $L(1,0,0) = (1,1,0)$, $L(0,1,0) = (-1,1,0)$, and $L(0,0,1) = (-1,1,1)$, whence the desired matrix is

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

2. Call this basis E. We have

$$L(1,0,0) = (1,1,0) = 0(1,0,0) + 1(1,1,0) + 0(1,0,1) = (0,1,0)_E,$$

$$L(1,1,0) = (0,2,0) = -2(1,0,0) + 2(1,1,0) + 0(1,0,1) = (-2,2,0)_E,$$

and

$$L(1,0,1) = (0,2,1) = -3(1,0,0) + 2(1,1,0) + 1(1,0,1) = (-3,2,1)_E,$$

whence the desired matrix is

$$\begin{bmatrix} 0 & -2 & -3 \\ 1 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

◀

Definition 1.3.3 The column rank of A is the dimension of the space generated by the collums of A, while the row rank of A is the dimension of the space generated by the rows of A.

A fundamental result in linear algebra is that the column rank and the row rank are always equal. This number (i.e., the number of linearly independent rows or columns) is simply called the rank of A.

1.4 Three Dimensional Space

In this section we particularize some definitions to the important case of three dimensional space

Definition 1.4.1 The 3-dimensional space is defined and denoted by

$$\mathbb{R}^3 = \left\{ \mathbf{r} = (x,y,z) : x \in \mathbb{R},\ y \in \mathbb{R},\ z \in \mathbb{R} \right\}.$$

Having oriented the z axis upwards, we have a choice for the orientation of the the x and y-axis. We adopt a convention known as a **right-handed coordinate system**, as in figure 1.1. Let us explain. Put

$$\mathbf{i} = (1,0,0), \qquad \mathbf{j} = (0,1,0), \qquad \mathbf{k} = (0,0,1),$$

and observe that

$$\mathbf{r} = (x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

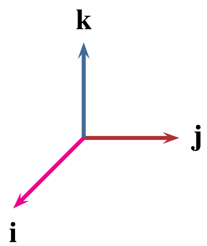


Figure 1.1 Right-handed system.

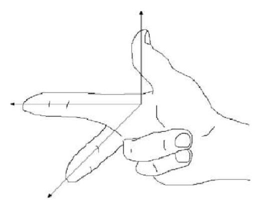


Figure 1.2 Right Hand.

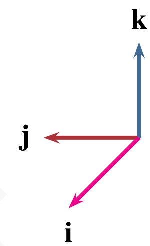


Figure 1.3 Left-handed system.

Definition 1.4.2 The dot product of two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^3$ is

$$\mathbf{x}\bullet\mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3.$$

The norm of a vector $\mathbf{x}$ in $\mathbb{R}^3$ is

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\bullet\mathbf{x}} = \sqrt{(x_1)^2 + (x_2)^2 + (x_3)^2}.$$

The Cauchy-Schwarz-Bunyakovsky Inequality takes the form

$$|\mathbf{x}\bullet\mathbf{y}| \leqslant \|\mathbf{x}\|\|\mathbf{y}\| \implies |x_1y_1 + x_2y_2 + x_3y_3| \leqslant (x_1^2 + x_2^2 + x_3^2)^{1/2}(y_1^2 + y_2^2 + y_3^2)^{1/2},$$

equality holding if an only if the vectors are parallel.

Consider now two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^3$. If $\mathbf{x} \parallel \mathbf{y}$, then the set

$$\{s\mathbf{x} + t\mathbf{y} : s \in \mathbb{R}, t \in \mathbb{R}\} = \{\lambda\mathbf{x} : \lambda \in \mathbb{R}\},$$

which is a line through the origin. Suppose now that $\mathbf{x}$ and $\mathbf{y}$ are not parallel. Then

$$\{s\mathbf{x} + t\mathbf{y} : s \in \mathbb{R}, t \in \mathbb{R}, \mathbf{x} \not\parallel \mathbf{y}\}$$

is a plane passing through the origin. We will say, abusing language, that two vectors are **coplanar** if there exists bi-point representatives of the vector that lie on the same plane. We will say, again abusing language, that a vector is **parallel to a specific plane** or that it **lies on a specific plane** if there exists a bi-point representative of the vector that lies on the particular plane. All the above gives the following result.

Theorem 1.4.1 Let $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^3$ be non-parallel vectors. Then every vector $\mathbf{v}$ of the form

$$\mathbf{v} = a\mathbf{x} + b\mathbf{y},$$

a, b arbitrary scalars, is coplanar with both $\mathbf{x}$ and $\mathbf{y}$. Conversely, any vector $\mathbf{t}$ coplanar with both $\mathbf{x}$ and $\mathbf{y}$ can be uniquely expressed in the form

$$\mathbf{t} = p\mathbf{x} + q\mathbf{y}.$$

See figure 1.4.

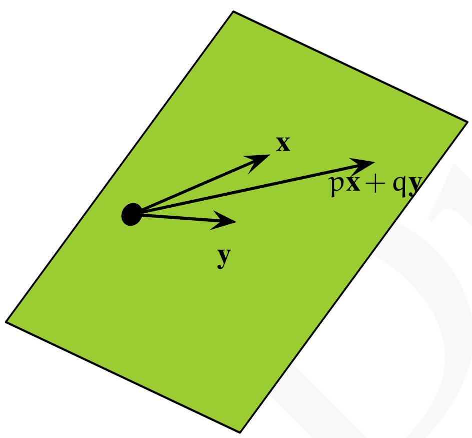


Figure 1.4 Theorem 1.4.1.

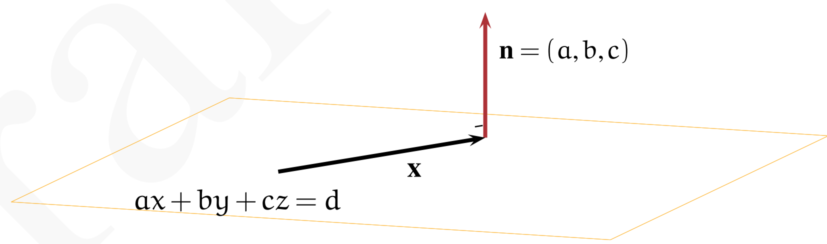


Figure 1.5 Theorem 1.4.3.

Theorem 1.4.2 Let **u** and **v** be linearly independent vectors. The **parametric equation** of a plane containing the point **a**, and parallel to the vectors **x** and **x** is given by

$$\mathbf{r} - \mathbf{x} = p\mathbf{u} + q\mathbf{v}.$$

Componentwise this takes the form

$$x - x_1 = pu_1 + qv_1,$$
$$y - x_2 = pu_2 + qv_2,$$
$$z - x_3 = pu_3 + qv_3.$$

Multiplying the first equation by $u_2v_3 - u_3v_2$, the second by $u_3v_1 - u_1v_3$, and the third by $u_1v_2 - u_2v_1$, we obtain,

$$(u_2v_3 - u_3v_2)(x - a_1) = (u_2v_3 - u_3v_2)(pu_1 + qv_1),$$
$$(u_3v_1 - u_1v_3)(y - a_2) = (u_3v_1 - u_1v_3)(pu_2 + qv_2),$$
$$(u_1v_2 - u_2v_1)(z - a_3) = (u_1v_2 - u_2v_1)(pu_3 + qv_3).$$

Adding gives,

$$(u_2v_3 - u_3v_2)(x - a_1) + (u_3v_1 - u_1v_3)(y - a_2) + (u_1v_2 - u_2v_1)(z - a_3) = 0.$$

Put

$$a = u_2v_3 - u_3v_2, \qquad b = u_3v_1 - u_1v_3, \qquad c = u_1v_2 - u_2v_1,$$

and

$$d = a_1(u_2v_3 - u_3v_2) + a_2(u_3v_1 - u_1v_3) + a_3(u_1v_2 - u_2v_1).$$

Since **u** is linearly independent from **v**, not all of a, b, c are zero. This gives the following theorem.

Theorem 1.4.3 The equation of the plane in space can be written in the form

$$ax + by + cz = d,$$

which is the **Cartesian equation** of the plane. Here $a^2 + b^2 + c^2 \neq 0$, that is, at least one of the coefficients is non-zero. Moreover, the vector $\mathbf{n} = (a, b, c)$ is normal to the plane with Cartesian equation $ax + by + cz = d$.

Proof: We have already proved the first statement. For the second statement, observe that if **x** and **x** are non-parallel vectors and $\mathbf{r} - \mathbf{x} = p\mathbf{x} + q\mathbf{x}$ is the equation of the plane containing the point

a and parallel to the vectors **x** and **x**, then if **n** is simultaneously perpendicular to **x** and **x** then $(\mathbf{r}-\mathbf{x})\bullet\mathbf{n}=0$ for $\mathbf{u}\bullet\mathbf{n}=0=\mathbf{v}\bullet\mathbf{n}$. Now, since at least one of a,b,c is non-zero, we may assume $a\neq 0$. The argument is similar if one of the other letters is non-zero and $a=0$. In this case we can see that

$$x=\frac{d}{a}-\frac{b}{a}y-\frac{c}{a}z.$$

Put $y=s$ and $z=t$. Then

$$\left(x-\frac{d}{a},y,z\right)=s\left(-\frac{b}{a},1,0\right)+t\left(-\frac{c}{a},0,1\right)$$

is a parametric equation for the plane. We have

$$a\left(-\frac{b}{a}\right)+b\,(1)+c\,(0)=0,\qquad a\left(-\frac{c}{a}\right)+b\,(0)+c\,(1)=0,$$

and so (a,b,c) is simultaneously perpendicular to $\left(-\frac{b}{a},1,0\right)$ and $\left(-\frac{c}{a},0,1\right)$, proving the second statement.■

Example 1.4.4 The equation of the plane passing through the point $(1,-1,2)$ and normal to the vector $(-3,2,4)$ is

$$-3(x-1)+2(y+1)+4(z-2)=0 \implies -3x+2y+4z=3.$$

Example 1.4.5 Find both the parametric equation and the Cartesian equation of the plane parallel to the vectors $(1,1,1)$ and $(1,1,0)$ and passing through the point $(0,-1,2)$.

Solution: ► The desired parametric equation is

$$(x,y+1,z-2)=s\,(1,1,1)+t\,(1,1,0).$$

This gives

$$s=z-2,\qquad t=y+1-s=y+1-z+2=y-z+3$$

and

$$x=s+t=z-2+y-z+3=y+1.$$

Hence the Cartesian equation is $x-y=1$. ◄

Definition 1.4.3 If **n** is perpendicular to plane Π_1 and $\mathbf{n}'$ is perpendicular to plane Π_2, the **angle between the two planes** is the angle between the two vectors **n** and $\mathbf{n}'$.

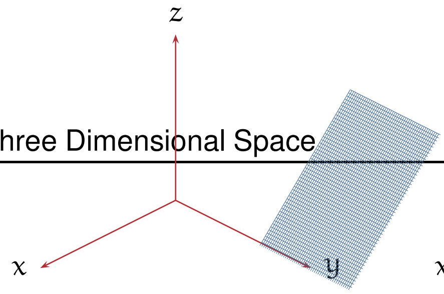


Figure 1.6 The plane $z = 1 - x$.

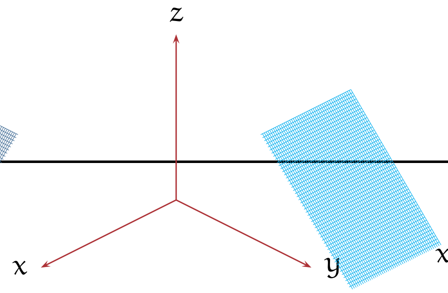


Figure 1.7 The plane $z = 1 - y$.

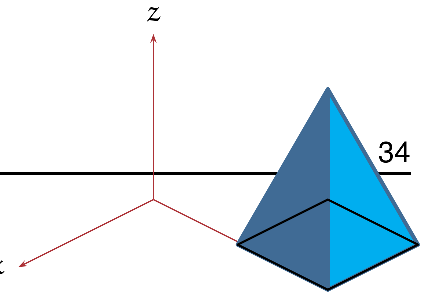


Figure 1.8 Solid bounded by the planes $z = 1 - x$ and $z = 1 - y$ in the first octant.

Example 1.4.6

1. Draw the intersection of the plane $z = 1 - x$ with the first octant.
2. Draw the intersection of the plane $z = 1 - y$ with the first octant.
3. Find the angle between the planes $z = 1 - x$ and $z = 1 - y$.
4. Draw the solid $\mathcal{S}$ which results from the intersection of the planes $z = 1 - x$ and $z = 1 - y$ with the first octant.
5. Find the volume of the solid $\mathcal{S}$.

Solution: ►

1. This appears in figure 1.6.
2. This appears in figure 1.7.
3. The vector $(1, 0, 1)$ is normal to the plane $x + z = 1$, and the vector $(0, 1, 1)$ is normal to the plane $y + z = 1$. If θ is the angle between these two vectors, then
$$\cos\theta = \frac{1 \cdot 0 + 0 \cdot 1 + 1 \cdot 1}{\sqrt{1^2 + 1^2} \cdot \sqrt{1^2 + 1^2}} \implies \cos\theta = \frac{1}{2} \implies \theta = \frac{\pi}{3}.$$
4. This appears in figure 1.8.
5. The resulting solid is a pyramid with square base of area $A = 1 \cdot 1 = 1$. Recall that the volume of a pyramid is given by the formula $V = \dfrac{Ah}{3}$, where A is area of the base of the pyramid and h is its height. Now, the height of this pyramid is clearly 1, and hence the volume required is $\dfrac{1}{3}$.

◄

1.4.1 Cross Product

The cross product of **two** vectors is defined **only** in three-dimensional space $\mathbb{R}^3$. We will define a generalization of the cross product for the n dimensional space in the chapter 13 The standard cross product is defined as a product satisfying the following properties.

Definition 1.4.4 Let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be vectors in $\mathbb{R}^3$, and let $\lambda \in \mathbb{R}$ be a scalar. The cross product $\times$ is a closed binary operation satisfying

❶ **A**nti-commutativity: $\mathbf{x} \times \mathbf{y} = -(\mathbf{y} \times \mathbf{x})$

❷ **B**ilinearity:

$$(\mathbf{x}+\mathbf{z}) \times \mathbf{y} = \mathbf{x} \times \mathbf{y} + \mathbf{z} \times \mathbf{y} \quad \text{and} \quad \mathbf{x} \times (\mathbf{z}+\mathbf{y}) = \mathbf{x} \times \mathbf{z} + \mathbf{x} \times \mathbf{y}$$

❸ **S**calar homogeneity: $(\lambda\mathbf{x}) \times \mathbf{y} = \mathbf{x} \times (\lambda\mathbf{y}) = \lambda(\mathbf{x} \times \mathbf{y})$

❹ $\mathbf{x} \times \mathbf{x} = \mathbf{0}$

❺ **R**ight-hand Rule:

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \;\; \mathbf{j} \times \mathbf{k} = \mathbf{i}, \;\; \mathbf{k} \times \mathbf{i} = \mathbf{j}.$$

It follows that the cross product is an operation that, given two non-parallel vectors on a plane, allows us to “get out” of that plane.

Example 1.4.7 Find

$$(1,0,-3) \times (0,1,2).$$

Solution: ► We have

$$\begin{aligned}(\mathbf{i}-3\mathbf{k}) \times (\mathbf{j}+2\mathbf{k}) &= \mathbf{i} \times \mathbf{j} + 2\mathbf{i} \times \mathbf{k} - 3\mathbf{k} \times \mathbf{j} - 6\mathbf{k} \times \mathbf{k} \\ &= \mathbf{k} - 2\mathbf{j} + 3\mathbf{i} + 6\mathbf{0} \\ &= 3\mathbf{i} - 2\mathbf{j} + \mathbf{k}\end{aligned}$$

Hence

$$(1,0,-3) \times (0,1,2) = (3,-2,1).$$

◄

The cross product of vectors in $\mathbb{R}^3$ is not associative, since

$$\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = \mathbf{i} \times \mathbf{k} = -\mathbf{j}$$

but

$$(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0} \times \mathbf{j} = \mathbf{0}.$$

Operating as in example 1.4.7 we obtain

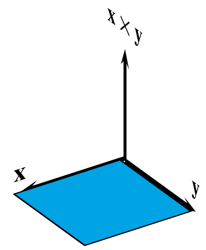


Figure 1.9 Theorem 1.4.11.

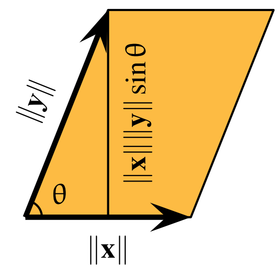


Figure 1.10 Area of a parallelogram

Theorem 1.4.8 Let $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ be vectors in $\mathbb{R}^3$. Then

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2)\mathbf{i} + (x_3y_1 - x_1y_3)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k}.$$

Proof: Since $\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$, we only worry about the mixed products, obtaining,

$$\begin{aligned}
\mathbf{x} \times \mathbf{y} &= (x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}) \times (y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k}) \\
&= x_1y_2\mathbf{i} \times \mathbf{j} + x_1y_3\mathbf{i} \times \mathbf{k} + x_2y_1\mathbf{j} \times \mathbf{i} + x_2y_3\mathbf{j} \times \mathbf{k} \\
&\qquad + x_3y_1\mathbf{k} \times \mathbf{i} + x_3y_2\mathbf{k} \times \mathbf{j} \\
&= (x_1y_2 - y_1x_2)\mathbf{i} \times \mathbf{j} + (x_2y_3 - x_3y_2)\mathbf{j} \times \mathbf{k} + (x_3y_1 - x_1y_3)\mathbf{k} \times \mathbf{i} \\
&= (x_1y_2 - y_1x_2)\mathbf{k} + (x_2y_3 - x_3y_2)\mathbf{i} + (x_3y_1 - x_1y_3)\mathbf{j},
\end{aligned}$$

proving the theorem. ■

The cross product can also be expressed as the formal/mnemonic determinant

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

Using cofactor expansion we have

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{i} + \begin{vmatrix} u_3 & u_1 \\ v_3 & v_1 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k}$$

Using the cross product, we may obtain a third vector simultaneously perpendicular to two other vectors in space.

Theorem 1.4.9 $\mathbf{x} \perp (\mathbf{x} \times \mathbf{y})$ and $\mathbf{y} \perp (\mathbf{x} \times \mathbf{y})$, that is, the cross product of two vectors is simultaneously perpendicular to both original vectors.

Proof: We will only check the first assertion, the second verification is analogous.

$$\begin{aligned} \mathbf{x} \bullet (\mathbf{x} \times \mathbf{y}) &= (x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}) \bullet ((x_2y_3 - x_3y_2)\mathbf{i} \\ &\qquad + (x_3y_1 - x_1y_3)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k}) \\ &= x_1x_2y_3 - x_1x_3y_2 + x_2x_3y_1 - x_2x_1y_3 + x_3x_1y_2 - x_3x_2y_1 \\ &= 0, \end{aligned}$$

completing the proof. ■ Although the cross product is not associative, we have, however, the following theorem.

Theorem 1.4.10

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \bullet \mathbf{c})\mathbf{b} - (\mathbf{a} \bullet \mathbf{b})\mathbf{c}.$$

Proof:

$$\begin{aligned} \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &= (a_1\mathbf{i} + x_2\mathbf{j} + a_3\mathbf{k}) \times ((b_2c_3 - b_3c_2)\mathbf{i} + \\ &\qquad + (b_3c_1 - b_1c_3)\mathbf{j} + (b_1c_2 - b_2c_1)\mathbf{k}) \\ &= a_1(b_3c_1 - b_1c_3)\mathbf{k} - a_1(b_1c_2 - b_2c_1)\mathbf{j} - a_2(b_2c_3 - b_3c_2)\mathbf{k} \\ &\qquad + a_2(b_1c_2 - b_2c_1)\mathbf{i} + a_3(b_2c_3 - b_3c_2)\mathbf{j} - a_3(b_3c_1 - b_1c_3)\mathbf{i} \\ &= (a_1c_1 + a_2c_2 + a_3c_3)(b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{i}) + \\ &\qquad (-a_1y_1 - a_2y_2 - a_3b_3)(c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{i}) \\ &= (\mathbf{a} \bullet \mathbf{c})\mathbf{b} - (\mathbf{a} \bullet \mathbf{b})\mathbf{c}, \end{aligned}$$

completing the proof. ■

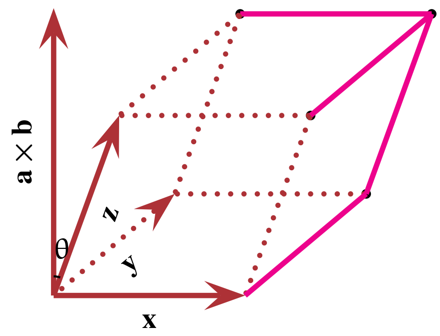


Figure 1.11 Theorem 9.7.1.

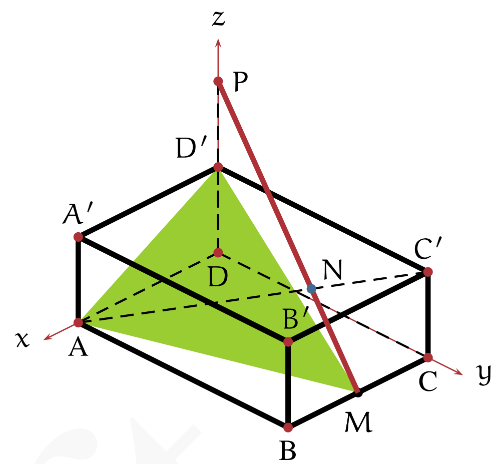


Figure 1.12 Example 1.4.15.

Theorem 1.4.11 Let $\widehat{(\mathbf{x},\mathbf{y})} \in [0;\pi]$ be the convex angle between two vectors **x** and **y**. Then

$$||\mathbf{x}\times\mathbf{y}|| = ||\mathbf{x}||||\mathbf{y}|| \sin\widehat{(\mathbf{x},\mathbf{y})}.$$

Proof: We have

$$\begin{aligned}
||\mathbf{x}\times\mathbf{y}||^2 &= (x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 + (x_1y_2 - x_2y_1)^2 \\
&= y^2y_3^2 - 2x_2y_3x_3y_2 + z^2y_2^2 + z^2y_1^2 - 2x_3y_1x_1y_3 + \\
&\qquad + x^2y_3^2 + x^2y_2^2 - 2x_1y_2x_2y_1 + y^2y_1^2 \\
&= (x^2 + y^2 + z^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2 \\
&= ||\mathbf{x}||^2||\mathbf{y}||^2 - (\mathbf{x}\bullet\mathbf{y})^2 \\
&= ||\mathbf{x}||^2||\mathbf{y}||^2 - ||\mathbf{x}||^2||\mathbf{y}||^2\cos^2\widehat{(\mathbf{x},\mathbf{y})} \\
&= ||\mathbf{x}||^2||\mathbf{y}||^2\sin^2\widehat{(\mathbf{x},\mathbf{y})},
\end{aligned}$$

whence the theorem follows. ■

Theorem 1.4.11 has the following geometric significance: $||\mathbf{x}\times\mathbf{y}||$ is the area of the parallelogram formed when the tails of the vectors are joined. See figure 1.10.

The following corollaries easily follow from Theorem 1.4.11.

Corollary 1.4.12 Two non-zero vectors $\mathbf{x}, \mathbf{y}$ satisfy $\mathbf{x} \times \mathbf{y} = \mathbf{0}$ if and only if they are parallel.

Corollary 1.4.13 — Lagrange's Identity.

$$\|\mathbf{x} \times \mathbf{y}\|^2 = \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - (\mathbf{x} \bullet \mathbf{y})^2.$$

The following result mixes the dot and the cross product.

Theorem 1.4.14 Let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, be linearly independent vectors in $\mathbb{R}^3$. The signed volume of the parallelepiped spanned by them is $(\mathbf{a} \times \mathbf{b}) \bullet \mathbf{z}$.

Proof: See figure 1.11. The area of the base of the parallelepiped is the area of the parallelogram determined by the vectors $\mathbf{x}$ and $\mathbf{y}$, which has area $\|\mathbf{a} \times \mathbf{b}\|$. The altitude of the parallelepiped is $\|\mathbf{z}\| \cos\theta$ where θ is the angle between $\mathbf{z}$ and $\mathbf{a} \times \mathbf{b}$. The volume of the parallelepiped is thus

$$\|\mathbf{a} \times \mathbf{b}\| \|\mathbf{z}\| \cos\theta = (\mathbf{a} \times \mathbf{b}) \bullet \mathbf{z},$$

proving the theorem. ■

Since we may have used any of the faces of the parallelepiped, it follows that

$$(\mathbf{a} \times \mathbf{b}) \bullet \mathbf{z} = (\mathbf{b} \times \mathbf{c}) \bullet \mathbf{x} = (\mathbf{c} \times \mathbf{a}) \bullet \mathbf{y}.$$

In particular, it is possible to "exchange" the cross and dot products:

$$\mathbf{x} \bullet (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \bullet \mathbf{z}$$

Example 1.4.15 Consider the rectangular parallelepiped $ABCDD'C'B'A'$ (figure 1.12) with vertices $A(2,0,0)$, $B(2,3,0)$, $C(0,3,0)$, $D(0,0,0)$, $D'(0,0,1)$, $C'(0,3,1)$, $B'(2,3,1)$, $A'(2,0,1)$. Let M be the midpoint of the line segment joining the vertices B and C.

1. Find the Cartesian equation of the plane containing the points A, D', and M.
2. Find the area of $\triangle AD'M$.
3. Find the parametric equation of the line $\overleftrightarrow{AC'}$.
4. Suppose that a line through M is drawn cutting the line segment $[AC']$ in N and the line $\overleftrightarrow{DD'}$ in P. Find the parametric equation of $\overleftrightarrow{MP}$.

Solution: ▶

1. Form the following vectors and find their cross product:

$$\mathbf{AD'} = (-2,0,1), \qquad \mathbf{AM} = (-1,3,0) \implies \mathbf{AD'} \times \mathbf{AM} = (-3,-1,-6).$$

The equation of the plane is thus

$$(x-2, y-0, z-0) \bullet (-3,-1,-6) = 0 \implies 3(x-2) + 1(y) + 6z = 0 \implies 3x + y + 6z = 6.$$

2. The area of the triangle is

$$\frac{\left\|\mathbf{AD'} \times \mathbf{AM}\right\|}{2} = \frac{1}{2}\sqrt{3^2+1^2+6^2} = \frac{\sqrt{46}}{2}.$$

3. We have $\mathbf{AC'} = (-2,3,1)$, and hence the line $\overleftrightarrow{\mathbf{A}C'}$ has parametric equation

$$(x,y,z) = (2,0,0) + t(-2,3,1) \implies x = 2-2t,\ y = 3t,\ z = t.$$

4. Since P is on the z-axis, $P = \left(0,0,z'\right)$ for some real number $z' > 0$. The parametric equation of the line $\overleftrightarrow{\mathbf{M}P}$ is thus

$$(x,y,z) = (1,3,0) + s\left(-1,-3,z'\right) \implies x = 1-s \qquad y = 3-3s, \qquad z = sz'.$$

Since N is on both $\overleftrightarrow{\mathbf{M}P}$ and $\overleftrightarrow{\mathbf{A}C'}$ we must have

$$2-2t = 1-s,\ 3t = 3-3s,\ t = sz'.$$

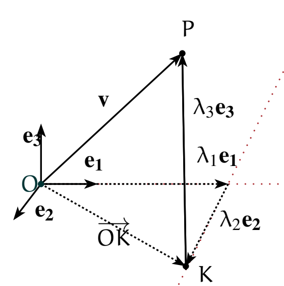
Solving the first two equations gives $s = \frac{1}{3}, t = \frac{2}{3}$. Putting this into the third equation we deduce $z' = 2$. Thus $P = (0,0,2)$ and the desired equation is

$$(x,y,z) = (1,3,0) + s(-1,-3,2) \implies x = 1-s, \quad y = 3-3s, \qquad z = 2s.$$

◀

1.4.2 Cylindrical and Spherical Coordinates

Let $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ be an ordered basis for $\mathbb{R}^3$. As we have already saw, for every $v \in \mathbb{R}^n$ there is a unique linear combination of the basis vectors that equals v:



$$v = x\mathbf{x}_1 + y\mathbf{x}_2 + z\mathbf{x}_3.$$

The coordinate vector of v relative to E is the sequence of coordinates

$$[v]_E = (x,y,z).$$

In this representation, the coordinates of a point (x, y, z) are determined by following straight paths starting from the origin: first parallel to $\mathbf{x}_1$, then parallel to the $\mathbf{x}_2$, then parallel to the $\mathbf{x}_3$, as in Figure 1.7.1.

In *curvilinear coordinate systems*, these paths can be curved. We will provide the definition of curvilinear coordinate systems in the section 3.10 and 8. In this section we provide some examples: the three types of curvilinear coordinates which we will consider in this section are polar coordinates in the plane cylindrical and spherical coordinates in the space.

Instead of referencing a point in terms of sides of a rectangular parallelepiped, as with Cartesian coordinates, we will think of the point as lying on a cylinder or sphere. Cylindrical coordinates are often used when there is symmetry around the z-axis; spherical coordinates are useful when there is symmetry about the origin.

Let $P = (x, y, z)$ be a point in Cartesian coordinates in $\mathbb{R}^3$, and let $P_0 = (x, y, 0)$ be the projection of P upon the xy-plane. Treating (x, y) as a point in $\mathbb{R}^2$, let (r, θ) be its polar coordinates (see Figure 1.7.2). Let ρ be the length of the line segment from the origin to P, and let ϕ be the angle between that line segment and the positive z-axis (see Figure 1.7.3). ϕ is called the *zenith angle*. Then the **cylindrical coordinates** (r, θ, z) and the **spherical coordinates** (ρ, θ, ϕ) of $P(x, y, z)$ are defined as follows:[1]

Cylindrical coordinates (r, θ, z):

$$x = r\cos\theta \qquad r = \sqrt{x^2 + y^2}$$

$$y = r\sin\theta \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

$$z = z \qquad z = z$$

where $0 \leqslant \theta \leqslant \pi$ if $y \geqslant 0$ and $\pi < \theta < 2\pi$ if $y < 0$

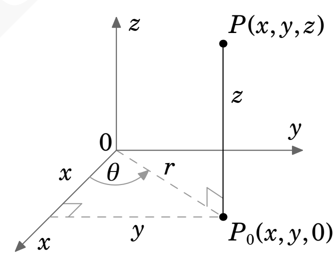


Figure 1.13
Cylindrical coordinates

[1] This "standard" definition of spherical coordinates used by mathematicians results in a left-handed system. For this reason, physicists usually switch the definitions of θ and ϕ to make (ρ, θ, ϕ) a right-handed system.

Spherical coordinates (ρ,θ,ϕ):

$$x = \rho\sin\phi\,\cos\theta \qquad \rho = \sqrt{x^2+y^2+z^2}$$

$$y = \rho\sin\phi\,\sin\theta \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

$$z = \rho\cos\phi \qquad \phi = \cos^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)$$

where $0 \leqslant \theta \leqslant \pi$ if $y \geqslant 0$ and $\pi < \theta < 2\pi$ if $y < 0$

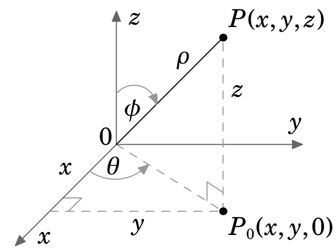


Figure 1.14
Spherical coordinates

Both θ and ϕ are measured in radians. Note that $r \geqslant 0$, $0 \leqslant \theta < 2\pi$, $\rho \geqslant 0$ and $0 \leqslant \phi \leqslant \pi$. Also, θ is undefined when $(x,y) = (0,0)$, and ϕ is undefined when $(x,y,z) = (0,0,0)$.

Example 1.4.16 Convert the point $(-2,-2,1)$ from Cartesian coordinates to (a) cylindrical and (b) spherical coordinates.

Solution: ► (a) $r = \sqrt{(-2)^2+(-2)^2} = 2\sqrt{2}$, $\theta = \tan^{-1}\left(\frac{-2}{-2}\right) = \tan^{-1}(1) = \frac{5\pi}{4}$, since $y = -2 < 0$.

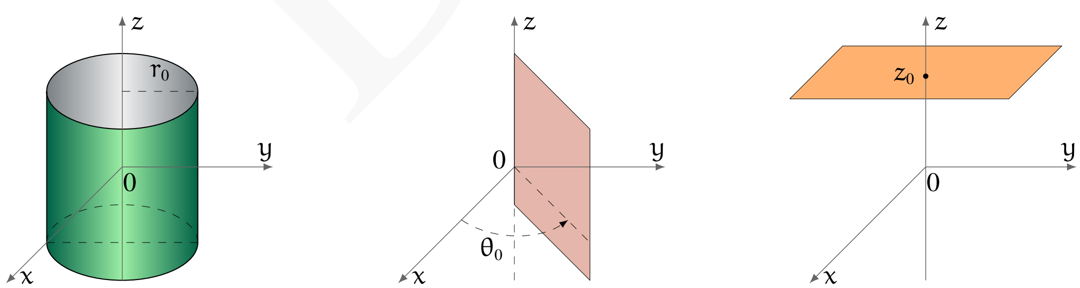
$\therefore\ (r,\theta,z) = \left(2\sqrt{2}, \frac{5\pi}{4}, 1\right)$

(b) $\rho = \sqrt{(-2)^2+(-2)^2+1^2} = \sqrt{9} = 3$, $\phi = \cos^{-1}\left(\frac{1}{3}\right) \approx 1.23$ radians.

$\therefore\ (\rho,\theta,\phi) = \left(3, \frac{5\pi}{4}, 1.23\right)$

◄

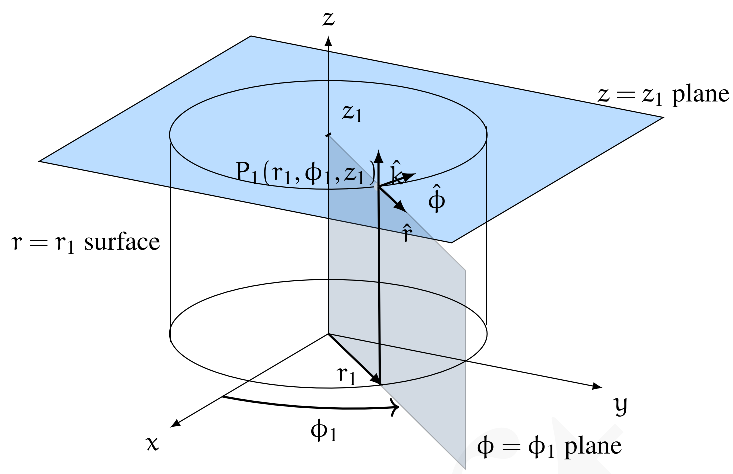
For cylindrical coordinates (r,θ,z), and constants r_0, θ_0 and z_0, we see from Figure 8.3 that the surface $r = r_0$ is a cylinder of radius r_0 centered along the z-axis, the surface $\theta = \theta_0$ is a half-plane emanating from the z-axis, and the surface $z = z_0$ is a plane parallel to the xy-plane.



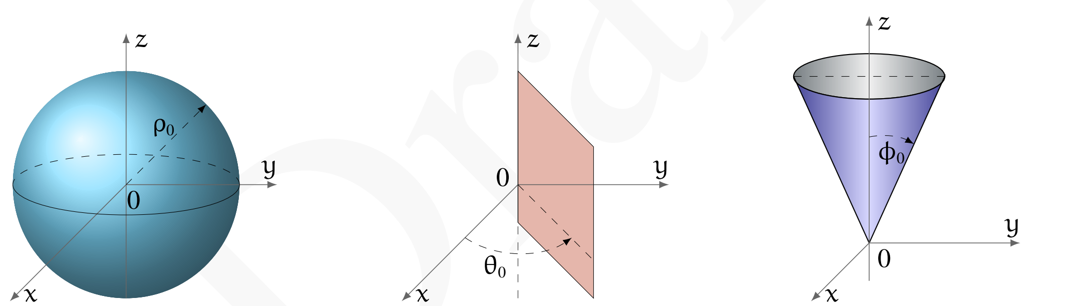
(a) $r = r_0$ **(b)** $\theta = \theta_0$ **(c)** $z = z_0$

Figure 1.15 Cylindrical coordinate surfaces

The unit vectors $\hat{r}, \hat{\theta}, \hat{k}$ at any point P are perpendicular to the surfaces $r =$ constant, $\theta =$ constant, $z =$ constant through P in the directions of increasing r, θ, z. Note that the direction of the unit vectors $\hat{r}, \hat{\theta}$ vary from point to point, unlike the corresponding Cartesian unit vectors.



For spherical coordinates (ρ, θ, ϕ), and constants ρ_0, θ_0 and ϕ_0, we see from Figure 1.16 that the surface $\rho = \rho_0$ is a sphere of radius ρ_0 centered at the origin, the surface $\theta = \theta_0$ is a half-plane emanating from the z-axis, and the surface $\phi = \phi_0$ is a circular cone whose vertex is at the origin.



(a) $\rho = \rho_0$ **(b)** $\theta = \theta_0$ **(c)** $\phi = \phi_0$

Figure 1.16 Spherical coordinate surfaces

Figures 8.3(a) and 1.16(a) show how these coordinate systems got their names.

Sometimes the equation of a surface in Cartesian coordinates can be transformed into a simpler equation in some other coordinate system, as in the following example.

Example 1.4.17 Write the equation of the cylinder $x^2 + y^2 = 4$ in cylindrical coordinates.

Solution: ► Since $r = \sqrt{x^2 + y^2}$, then the equation in cylindrical coordinates is $r = 2$. ◄

Using spherical coordinates to write the equation of a sphere does not necessarily make the equation simpler, if the sphere is not centered at the origin.

Example 1.4.18 Write the equation $(x-2)^2 + (y-1)^2 + z^2 = 9$ in spherical coordinates.

Solution: ► Multiplying the equation out gives

$$\begin{aligned} x^2+y^2+z^2-4x-2y+5=9\text{ , so we get} \\ \rho^2-4\rho\sin\phi\,\cos\theta-2\rho\sin\phi\,\sin\theta-4=0\text{ , or} \\ \rho^2-2\sin\phi\,(2\cos\theta-\sin\theta\,)\,\rho-4=0 \end{aligned}$$

after combining terms. Note that this actually makes it more difficult to figure out what the surface is, as opposed to the Cartesian equation where you could immediately identify the surface as a sphere of radius 3 centered at $(2,1,0)$. ◄

Example 1.4.19 Describe the surface given by $\theta = z$ in cylindrical coordinates.

Solution: ► This surface is called a *helicoid*. As the (vertical) z coordinate increases, so does the angle θ, while the radius r is unrestricted. So this sweeps out a (ruled!) surface shaped like a spiral staircase, where the spiral has an infinite radius. Figure 1.17 shows a section of this surface restricted to $0 \leqslant z \leqslant 4\pi$ and $0 \leqslant r \leqslant 2$. ◄

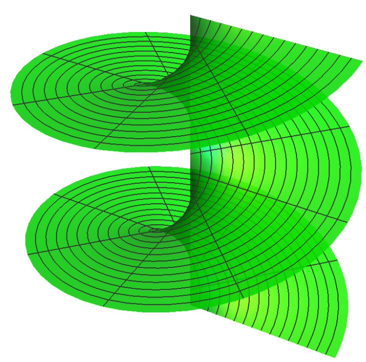


Figure 1.17 Helicoid $\theta = z$

Exercises

A

For Exercises 1-4, find the (a) cylindrical and (b) spherical coordinates of the point whose Cartesian coordinates are given.

1. $(2, 2\sqrt{3}, -1)$
2. $(-5, 5, 6)$
3. $(\sqrt{21}, -\sqrt{7}, 0)$
4. $(0, \sqrt{2}, 2)$

For Exercises 5-7, write the given equation in (a) cylindrical and (b) spherical coordinates.

5. $x^2 + y^2 + z^2 = 25$
6. $x^2 + y^2 = 2y$
7. $x^2 + y^2 + 9z^2 = 36$

B

8. Describe the intersection of the surfaces whose equations in spherical coordinates are $\theta = \dfrac{\pi}{2}$ and $\phi = \dfrac{\pi}{4}$.
9. Show that for $a \neq 0$, the equation $\rho = 2a \sin\phi \cos\theta$ in spherical coordinates describes a sphere centered at $(a, 0, 0)$ with radius $|a|$.

C

10. Let $P = (a, \theta, \phi)$ be a point in spherical coordinates, with $a > 0$ and $0 < \phi < \pi$. Then P lies on the sphere $\rho = a$. Since $0 < \phi < \pi$, the line segment from the origin to P can be extended to intersect the cylinder given by $r = a$ (in cylindrical coordinates). Find the cylindrical coordinates of that point of intersection.

11. Let P_1 and P_2 be points whose spherical coordinates are $(\rho_1, \theta_1, \phi_1)$ and $(\rho_2, \theta_2, \phi_2)$, respectively. Let $\mathbf{v}_1$ be the vector from the origin to P_1, and let $\mathbf{v}_2$ be the vector from the origin to P_2. For the angle γ between

$$\cos\gamma = \cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2 \cos(\theta_2 - \theta_1).$$

This formula is used in electrodynamics to prove the addition theorem for spherical harmonics, which provides a general expression for the electrostatic potential at a point due to a unit charge. See pp. 100-102 in [35].

12. Show that the distance d between the points P_1 and P_2 with cylindrical coordinates (r_1, θ_1, z_1) and (r_2, θ_2, z_2), respectively, is

$$d = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2 - \theta_1) + (z_2 - z_1)^2}.$$

13. Show that the distance d between the points P_1 and P_2 with spherical coordinates $(\rho_1, \theta_1, \phi_1)$ and $(\rho_2, \theta_2, \phi_2)$, respectively, is

$$d = \sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1 \rho_2 [\sin\phi_1 \sin\phi_2 \cos(\theta_2 - \theta_1) + \cos\phi_1 \cos\phi_2]}.$$

1.5 ⋆ Cross Product in the n-Dimensional Space

In this section we will answer the following question: Can one define a cross product in the n-dimensional space so that it will have properties similar to the usual 3 dimensional one?

Clearly the answer depends which properties we require.

The most direct generalizations of the cross product are to define either:

- a binary product $\times : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ which takes as input two vectors and gives as output a vector;
- a $n-1$-ary product $\times : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n-1 \text{ times}} \to \mathbb{R}^n$ which takes as input $n-1$ vectors, and gives as output one vector.

Under the correct assumptions it can be proved that a binary product exists only in the dimensions 3 and 7. A simple proof of this fact can be found in [50].

In this section we focus in the definition of the $n-1$-ary product.

Definition 1.5.1 Let $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ be vectors in $\mathbf{R}^n$,, and let $\lambda \in \mathbb{R}$ be a scalar. Then we define their generalized cross product $\mathbf{v}_n = \mathbf{v}_1 \times \cdots \times \mathbf{v}_{n-1}$ as the $(n-1)$-ary product satisfying

1. **A**nti-commutativity: $\mathbf{v}_1 \times \cdots \mathbf{v}_i \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1} = -\mathbf{v}_1 \times \cdots \mathbf{v}_{i+1} \times \mathbf{v}_i \times \cdots \times \mathbf{v}_{n-1}$, i.e, changing two consecutive vectors a minus sign appears.
2. **B**ilinearity: $\mathbf{v}_1 \times \cdots \mathbf{v}_i + \mathbf{x} \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1} = \mathbf{v}_1 \times \cdots \mathbf{v}_i \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1} + \mathbf{v}_1 \times \cdots \mathbf{x} \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1}$
3. **S**calar homogeneity: $\mathbf{v}_1 \times \cdots \lambda \mathbf{v}_i \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1} = \lambda \mathbf{v}_1 \times \cdots \mathbf{v}_i \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1}$
4. **R**ight-hand Rule: $\mathbf{e}_1 \times \cdots \times \mathbf{e}_{n-1} = \mathbf{e}_n$, $\mathbf{e}_2 \times \cdots \times \mathbf{e}_n = \mathbf{e}_1$, and so forth for cyclic permutations of indices.

We will also write

$$\bigtimes(\mathbf{v}_1, \dots, \mathbf{v}_{n-1}) := \mathbf{v}_1 \times \cdots \mathbf{v}_i \times \mathbf{v}_{i+1} \times \cdots \times \mathbf{v}_{n-1}$$

In coordinates, one can give a formula for this $(n-1)$-ary analogue of the cross product in $\mathbb{R}^n$ by:

Proposition 1.5.1 Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the canonical basis of $\mathbf{R}^n$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ be vectors in $\mathbf{R}^n$, with coordinates:

$$\begin{aligned} \mathbf{v}_1 &= (v_{11}, \ldots v_{1n}) \\ &\vdots \\ \mathbf{v}_i &= (v_{i1}, \ldots v_{in}) \\ &\vdots \\ \mathbf{v}_n &= (v_{n1}, \ldots v_{nn}) \end{aligned}$$

in the canonical basis. Then

$$\bigtimes(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) = \begin{vmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n-11} & \cdots & v_{n-1n} \\ \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{vmatrix}.$$

This formula is very similar to the determinant formula for the normal cross product in $\mathbb{R}^3$ except that the row of basis vectors is the last row in the determinant rather than the first.

The reason for this is to ensure that the ordered vectors

$$(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \bigtimes(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}))$$

have a positive orientation with respect to

$$(\mathbf{e}_1, \ldots, \mathbf{e}_n).$$

Proposition 1.5.2 The vector product have the following properties:

❶ The vector $\bigtimes(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$ is perpendicular to $\mathbf{v}_i$,

❷ the magnitude of $\bigtimes(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$ is the volume of the solid defined by the vectors $\mathbf{v}_1, \ldots \mathbf{v}_{i-1}$

❸ $$\mathbf{v}_n \bullet \mathbf{v}_1 \times \cdots \times \mathbf{v}_{n-1} = \begin{vmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n-11} & \cdots & v_{n-1n} \\ v_{n1} & \cdots & v_{n1} \end{vmatrix}.$$

1.6 Multivariable Functions

Let $A \subseteq \mathbb{R}^n$. For most of this course, our concern will be functions of the form

$$f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m.$$

If $m = 1$, we say that f is a **scalar field**. If $m \geqslant 2$, we say that f is a **vector field**.

We would like to develop a calculus analogous to the situation in $\mathbb{R}$. In particular, we would like to examine limits, continuity, differentiability, and integrability of multivariable functions. Needless to say, the introduction of more variables greatly complicates the analysis. For example, recall that the graph of a function $f : A \to \mathbb{R}^m$, $A \subseteq \mathbb{R}^n$. is the set

$$\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in A)\} \subseteq \mathbb{R}^{n+m}.$$

If $m+n > 3$, we have an object of more than three-dimensions! In the case $n = 2, m = 1$, we have a tri-dimensional surface. We will now briefly examine this case.

Definition 1.6.1 Let $A \subseteq \mathbb{R}^2$ and let $f : A \to \mathbb{R}$ be a function. Given $c \in \mathbb{R}$, the **level curve** at $z = c$ is the curve resulting from the intersection of the surface $z = f(x, y)$ and the plane $z = c$, if there is such a curve.

Example 1.6.1 The level curves of the surface $f(x, y) = x^2 + 3y^2$ (an elliptic paraboloid) are the concentric ellipses

$$x^2 + 3y^2 = c, \qquad c > 0.$$

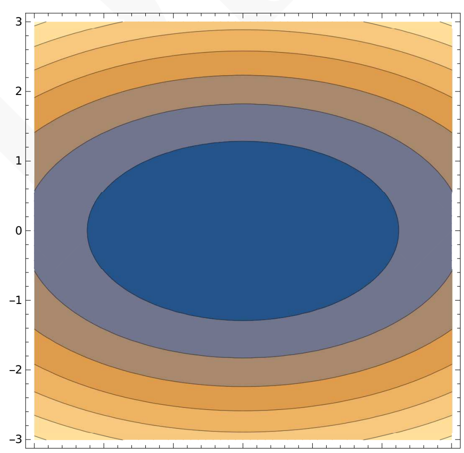


Figure 1.18 Level curves for $f(x, y) = x^2 + 3y^2$.

1.6.1 Graphical Representation of Vector Fields

In this section we present a graphical representation of vector fields. For this intent, we limit ourselves to low dimensional spaces.

A vector field $\mathbf{v} : \mathbb{R}^3 \to \mathbb{R}^3$ is an assignment of a vector $\mathbf{v} = \mathbf{v}(x,y,z)$ to each point (x,y,z) of a subset $U \subset \mathbb{R}^3$. Each vector $\mathbf{v}$ of the field can be regarded as a "bound vector" attached to the corresponding point (x,y,z). In components

$$\mathbf{v}(x,y,z) = v_1(x,y,z)\mathbf{i} + v_2(x,y,z)\mathbf{j} + v_3(x,y,z)\mathbf{k}.$$

Example 1.6.2 Sketch each of the following vector fields.

$\mathbf{F} = x\mathbf{i} + y\mathbf{j}$

$\mathbf{F} = -y\mathbf{i} + x\mathbf{j}$

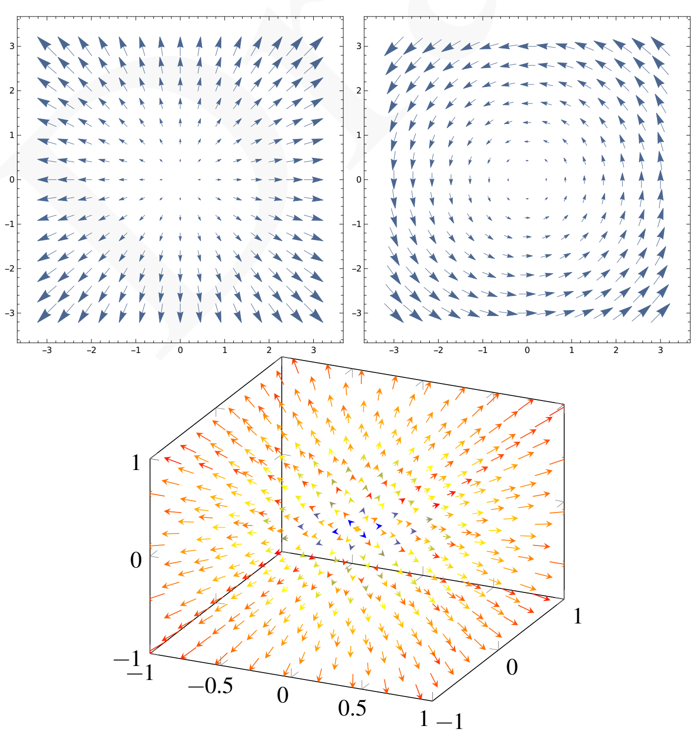
$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$

Solution: ►

a) The vector field is null at the origin; at other points, $\mathbf{F}$ is a vector pointing away from the origin;

b) This vector field is perpendicular to the first one at every point;

c) The vector field is null at the origin; at other points, $\mathbf{F}$ is a vector pointing away from the origin. This is the 3-dimensional analogous of the first one. ◄



Example 1.6.3 Suppose that an object of mass M is located at the origin of a three-dimensional coordinate system. We can think of this object as inducing a force field **g** in space. The effect of this gravitational field is to attract any object placed in the vicinity of the origin toward it with a force that is governed by Newton's Law of Gravitation.

$$\mathbf{F} = \frac{GmM}{r^2}$$

To find an expression for **g** , suppose that an object of mass m is located at a point with position vector $r = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

The gravitational field is the gravitational force exerted per unit mass on a small test mass (that won't distort the field) at a point in the field. Like force, it is a vector quantity: a point mass M at the origin produces the gravitational field

$$\mathbf{g} = \mathbf{g}(\mathbf{r}) = -\frac{GM}{r^3}\mathbf{r},$$

where **r** is the position relative to the origin and where $r = \|\mathbf{r}\|$. Its magnitude is

$$g = -\frac{GM}{r^2}$$

and, due to the minus sign, at each point **g** is directed opposite to **r**, i.e. towards the central mass.

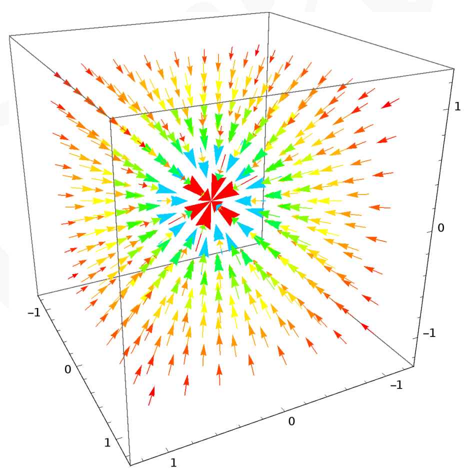


Figure 1.19 Gravitational Field

Exercises

Problem 1.1 Sketch the level curves for the following maps.

1. $(x,y) \mapsto x+y$
2. $(x,y) \mapsto xy$

3. $(x,y) \mapsto \min(|x|,|y|)$

4. $(x,y) \mapsto x^3 - x$

5. $(x,y) \mapsto x^2 + 4y^2$

6. $(x,y) \mapsto \sin(x^2+y^2)$

7. $(x,y) \mapsto \cos(x^2-y^2)$

Problem 1.2 Sketch the level surfaces for the following maps.

1. $(x,y,z) \mapsto x+y+z$

2. $(x,y,z) \mapsto xyz$

3. $(x,y,z) \mapsto \min(|x|,|y|,|z|)$

4. $(x,y,z) \mapsto x^2+y^2$

5. $(x,y,z) \mapsto x^2+4y^2$

6. $(x,y,z) \mapsto \sin(z-x^2-y^2)$

7. $(x,y,z) \mapsto x^2+y^2+z^2$

1.7 Levi-Civitta and Einstein Index Notation

We need an efficient abbreviated notation to handle the complexity of mathematical structure before us. We will use indices of a given "type" to denote all possible values of given index ranges. By index type we mean a collection of similar letter types, like those from the beginning or middle of the Latin alphabet, or Greek letters

$$a,b,c,\ldots$$
$$i,j,k,\ldots$$
$$\lambda,\beta,\gamma\ldots$$

each index of which is understood to have a given common range of successive integer values. Variations of these might be barred or primed letters or capital letters. For example, suppose we are looking at linear transformations between $\mathbb{R}^n$ and $\mathbb{R}^m$ where $m \neq n$. We would need two different index ranges to denote vector components in the two vector spaces of different dimensions, say $i,j,k,... = 1,2,\ldots,n$ and $\lambda,\beta,\gamma,\ldots = 1,2,\ldots,m$.

In order to introduce the so called Einstein summation convention, we agree to the following limitations on how indices may appear in formulas. A given index letter may occur only once in a given term in an expression (call this a "free index"), in which case the expression is understood to stand for the set of all such expressions for which the index assumes its allowed values, or it may occur twice but only as a superscript-subscript pair (one up, one down) which will stand for the sum over all allowed values (call this a "repeated index"). Here are some examples. If $i,j = 1,\ldots,n$ then

$$A^i \longleftrightarrow n \text{ expressions} : A^1, A^2, \ldots, A^n,$$
$$A^i{}_i \longleftrightarrow \sum_{i=1}^{n} A^i{}_i, \text{ a single expression with } n \text{ terms}$$

(this is called the trace of the matrix $\underline{A} = (A^i{}_j)$),

$A^{ji}{}_i \longleftrightarrow \sum_{i=1}^{n} A^{1i}{}_i, \ldots, \sum_{i=1}^{n} A^{ni}{}_i$, n expressions each of which has n terms in the sum,

$A_{ii} \longleftrightarrow$ no sum, just an expression for each i, if we want to refer to a specific diagonal component (entry) of a matrix, for example,

$A^i v_i + A^i w_i = A^i(v_i + w_i)$, 2 sums of n terms each (left) or one combined sum (right).

A repeated index is a "dummy index," like the dummy variable in a definite integral

$$\int_a^b f(x)\ dx = \int_a^b f(u)\ du.$$

We can change them at will: $A^i{}_i = A^j{}_j$.

In order to emphasize that we are using Einstein's convention, we will enclose any terms under consideration with $\ulcorner \cdot \lrcorner$.

Example 1.7.1 Using Einstein's Summation convention, the dot product of two vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ can be written as

$$\mathbf{x} \bullet \mathbf{y} = \sum_{i=1}^{n} x_i y_i = \ulcorner x_t y_t \lrcorner.$$

Example 1.7.2 Given that a_i, b_j, c_k, d_l are the components of vectors in $\mathbb{R}^3$, $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{d}$ respectively, what is the meaning of

$$\ulcorner a_i b_i c_k d_k \lrcorner?$$

Solution: ► We have

$$\ulcorner a_i b_i c_k d_k \lrcorner = \sum_{i=1}^{3} a_i b_i \ulcorner c_k d_k \lrcorner = \mathbf{x} \bullet \mathbf{y} \ulcorner c_k d_k \lrcorner = \mathbf{x} \bullet \mathbf{y} \sum_{k=1}^{3} c_k d_k = (\mathbf{x} \bullet \mathbf{y})(\mathbf{c} \bullet \mathbf{d}).$$

◄

Example 1.7.3 Using Einstein's Summation convention, the ij-th entry $(AB)_{ij}$ of the product of two matrices $A \in \mathbf{M}_{m \times n}(\mathbb{R})$ and $B \in \mathbf{M}_{n \times r}(\mathbb{R})$ can be written as

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} = \ulcorner A_{it} B_{tj} \lrcorner.$$

Example 1.7.4 Using Einstein's Summation convention, the trace $\operatorname{tr}(A)$ of a square matrix $A \in \mathbf{M}_{n \times n}(\mathbb{R})$ is $\operatorname{tr}(A) = \sum_{t=1}^{n} A_{tt} = \ulcorner A_{tt} \lrcorner$.

Example 1.7.5 Demonstrate, via Einstein's Summation convention, that if A, B are two $n \times n$ matrices, then

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Solution: ► We have

$$\operatorname{tr}(AB) = \operatorname{tr}\left((AB)_{ij}\right) = \operatorname{tr}\left(\ulcorner A_{ik}B_{kj}\lrcorner\right) = \ulcorner\ulcorner A_{tk}B_{kt}\lrcorner\lrcorner,$$

and

$$\operatorname{tr}(BA) = \operatorname{tr}\left((BA)_{ij}\right) = \operatorname{tr}\left(\ulcorner B_{ik}A_{kj}\lrcorner\right) = \ulcorner\ulcorner B_{tk}A_{kt}\lrcorner\lrcorner,$$

from where the assertion follows, since the indices are dummy variables and can be exchanged. ◄

Definition 1.7.1 — Kroenecker's Delta. The symbol δ_{ij} is defined as follows:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

Example 1.7.6 It is easy to see that $\ulcorner\delta_{ik}\delta_{kj}\lrcorner = \sum_{k=1}^{3}\delta_{ik}\delta_{kj} = \delta_{ij}$.

Example 1.7.7 We see that

$$\ulcorner\delta_{ij}a_i b_j\lrcorner = \sum_{i=1}^{3}\sum_{j=1}^{3}\delta_{ij}a_i b_j = \sum_{k=1} a_k b_k = \mathbf{x}\bullet\mathbf{y}.$$

Recall that a **permutation** of distinct objects is a reordering of them. The $3! = 6$ permutations of the index set $\{1,2,3\}$ can be classified into **even** or **odd**. We start with the identity permutation 123 and say it is even. Now, for any other permutation, we will say that it is even if it takes an even number of transpositions (switching only two elements in one move) to regain the identity permutation, and odd if it takes an odd number of transpositions to regain the identity permutation. Since

$$231 \to 132 \to 123, \quad 312 \to 132 \to 123,$$

the permutations 123 (identity), 231, and 312 are even. Since

$$132 \to 123, \quad 321 \to 123, \quad 213 \to 123,$$

the permutations 132, 321, and 213 are odd.

Definition 1.7.2 — Levi-Civitta's Alternating Tensor. The symbol ε_{jkl} is defined as follows:

$$\varepsilon_{jkl} = \begin{cases} 0 & \text{if } \{j,k,l\} \neq \{1,2,3\} \\ -1 & \text{if } \begin{pmatrix} 1 & 2 & 3 \\ j & k & l \end{pmatrix} \text{ is an odd permutation} \\ +1 & \text{if } \begin{pmatrix} 1 & 2 & 3 \\ j & k & l \end{pmatrix} \text{ is an even permutation} \end{cases}$$

In particular, if one subindex is repeated we have $\varepsilon_{rrs} = \varepsilon_{rsr} = \varepsilon_{srr} = 0$. Also,

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1, \qquad \varepsilon_{132} = \varepsilon_{321} = \varepsilon_{213} = -1.$$

Example 1.7.8 Using the Levi-Civitta alternating tensor and Einstein's summation convention, the cross product can also be expressed, if $\mathbf{i} = \mathbf{e_1}$, $\mathbf{j} = \mathbf{e_2}$, $\mathbf{k} = \mathbf{e_3}$, then

$$\mathbf{x} \times \mathbf{y} = \ulcorner \varepsilon_{jkl}(a_k b_l)\mathbf{e_j} \lrcorner.$$

Example 1.7.9 If $A = [a_{ij}]$ is a 3×3 matrix, then, using the Levi-Civitta alternating tensor,

$$\det A = \ulcorner \varepsilon_{ijk} a_{1i} a_{2j} a_{3k} \lrcorner.$$

Example 1.7.10 Let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be vectors in $\mathbb{R}^3$. Then

$$\mathbf{x} \bullet (\mathbf{y} \times \mathbf{z}) = \ulcorner x_i (\mathbf{y} \times \mathbf{z})_i \lrcorner = \ulcorner x_i \varepsilon_{ikl} (y_k z_l) \lrcorner.$$

Exercises

Problem 1.3 Let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be vectors in $\mathbb{R}^3$. Demonstrate that

$$\ulcorner x_i y_i z_j \lrcorner = (\mathbf{x} \bullet \mathbf{y}) \mathbf{z}.$$

CHAPTER 2

Limits and Continuity

2.1 Some Topology

Definition 2.1.1 Let $\mathbf{a} \in \mathbb{R}^n$ and let $\varepsilon > 0$. An **open ball** centered at $\mathbf{a}$ of radius ε is the set

$$B_\varepsilon(a) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| < \varepsilon\}.$$

An **open box** is a Cartesian product of open intervals

$$]a_1; b_1[\times]a_2; b_2[\times \cdots \times]a_{n-1}; b_{n-1}[\times]a_n; b_n[,$$

where the a_k, b_k are real numbers.

The set

$$B_\varepsilon(a) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| < \varepsilon\}.$$

are also called the ε**-neighborhood** of the point a.

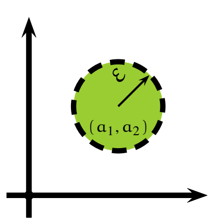


Figure 2.1 Open ball in $\mathbb{R}^2$.

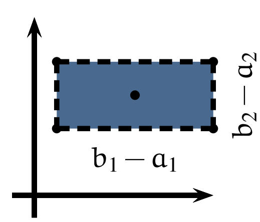


Figure 2.2 Open rectangle in $\mathbb{R}^2$.

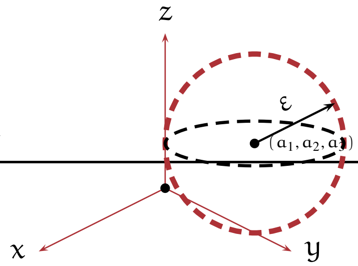


Figure 2.3 Open ball in $\mathbb{R}^3$.

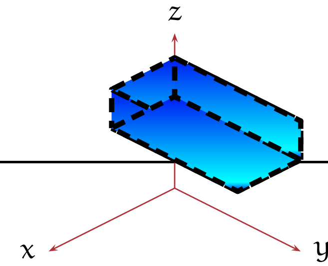


Figure 2.4 Open box in $\mathbb{R}^3$.

Example 2.1.1 An open ball in $\mathbb{R}$ is an open interval, an open ball in $\mathbb{R}^2$ is an open disk and an open ball in $\mathbb{R}^3$ is an open sphere. An open box in $\mathbb{R}$ is an open interval, an open box in $\mathbb{R}^2$ is a rectangle without its boundary and an open box in $\mathbb{R}^3$ is a box without its boundary.

Definition 2.1.2 A set $A \subseteq \mathbb{R}^n$ is said to be **open** if for every point belonging to it we can surround the point by a sufficiently small open ball so that this balls lies completely within the set. That is, $\forall \mathbf{a} \in A \ \ \exists \varepsilon > 0$ such that $B_\varepsilon(a) \subseteq A$.

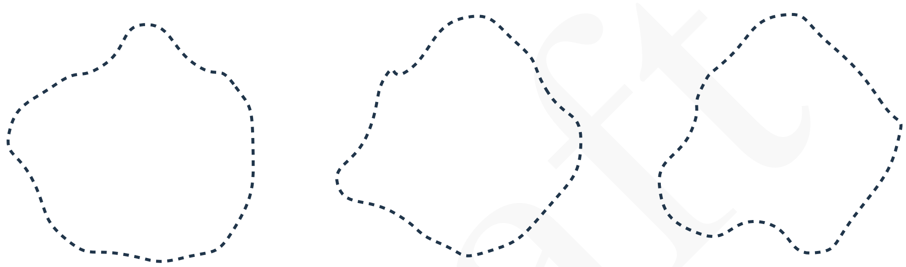


Figure 2.5 Open Sets

Example 2.1.2 The open interval $]-1;1[$ is open in $\mathbb{R}$. The interval $]-1;1]$ is not open, however, as no interval centred at 1 is totally contained in $]-1;1]$.

Example 2.1.3 The region $]-1;1[\times]0;+\infty[$ is open in $\mathbb{R}^2$.

Example 2.1.4 The ellipsoidal region $\left\{(x,y) \in \mathbb{R}^2 : x^2 + 4y^2 < 4\right\}$ is open in $\mathbb{R}^2$.

The reader will recognize that open boxes, open ellipsoids and their unions and finite intersections are open sets in $\mathbb{R}^n$.

Definition 2.1.3 A set $F \subseteq \mathbb{R}^n$ is said to be **closed** in $\mathbb{R}^n$ if its complement $\mathbb{R}^n \setminus F$ is open.

Example 2.1.5 The closed interval $[-1;1]$ is closed in $\mathbb{R}$, as its complement, $\mathbb{R} \setminus [-1;1] =]-\infty;-1[\cup]1;+\infty[$ is open in $\mathbb{R}$. The interval $]-1;1]$ is neither open nor closed in $\mathbb{R}$, however.

Example 2.1.6 The region $[-1;1] \times [0;+\infty[\times[0;2]$ is closed in $\mathbb{R}^3$.

Lemma 2.1.7 If $\mathbf{x}_1$ and $\mathbf{x}_2$ are in $S_r(\mathbf{x}_0)$ for some $r > 0$, then so is every point on the line segment from $\mathbf{x}_1$ to $\mathbf{x}_2$.

Proof: The line segment is given by

$$\mathbf{x} = t\mathbf{x}_2 + (1-t)\mathbf{x}_1, \quad 0 < t < 1.$$

Suppose that $r > 0$. If

$$|\mathbf{x}_1 - \mathbf{x}_0| < r, \quad |\mathbf{x}_2 - \mathbf{x}_0| < r,$$

and $0 < t < 1$, then

$$\begin{aligned}|\mathbf{x} - \mathbf{x}_0| &= |t\mathbf{x}_2 + (1-t)\mathbf{x}_1 - t\mathbf{x}_0 - (1-t)\mathbf{x}_0| \\ &= |t(\mathbf{x}_2 - \mathbf{x}_0) + (1-t)\mathbf{x}_1 - \mathbf{x}_0)| \\ &\leqslant t|\mathbf{x}_2 - \mathbf{x}_0| + (1-t)|\mathbf{x}_1 - \mathbf{x}_0| \\ &< tr + (1-t)r = r.\end{aligned}$$

■

Definition 2.1.4 A sequence of points $\{\mathbf{x}_r\}$ in $\mathbb{R}^n$ **converges to the limit $\overline{\mathbf{x}}$** if

$$\lim_{r\to\infty} |\mathbf{x}_r - \overline{\mathbf{x}}| = 0.$$

In this case we write

$$\lim_{r\to\infty} \mathbf{x}_r = \overline{\mathbf{x}}.$$

The next two theorems follow from this, the definition of distance in $\mathbb{R}^n$, and what we already know about convergence in $\mathbb{R}$.

Theorem 2.1.8 Let

$$\overline{\mathbf{x}} = (\overline{x}_1, \overline{x}_2, \dots, \overline{x}_n) \quad \text{and} \quad \mathbf{x}_r = (x_{1r}, x_{2r}, \dots, x_{nr}), \quad r \geqslant 1.$$

Then $\lim_{r\to\infty} \mathbf{x}_r = \overline{\mathbf{x}}$ if and only if

$$\lim_{r\to\infty} x_{ir} = \overline{x}_i, \quad 1 \leqslant i \leqslant n;$$

that is, a sequence $\{\mathbf{x}_r\}$ of points in $\mathbb{R}^n$ converges to a limit $\overline{\mathbf{x}}$ if and only if the sequences of components of $\{\mathbf{x}_r\}$ converge to the respective components of $\overline{\mathbf{x}}$.

Theorem 2.1.9 — Cauchy's Convergence Criterion. A sequence $\{\mathbf{x}_r\}$ in $\mathbb{R}^n$ converges if and only if for each $\varepsilon > 0$ there is an integer K such that

$$\|\mathbf{x}_r - \mathbf{x}_s\| < \varepsilon \quad \text{if} \quad r, s \geqslant K.$$

Definition 2.1.5 Let S be a subset of $\mathbb{R}$. Then

1. x_0 is a **limit point** of S if every deleted neighborhood of x_0 contains a point of S.
2. x_0 is a **boundary point** of S if every neighborhood of x_0 contains at least one point in S and one not in S. The set of boundary points of S is the **boundary** of S, denoted by ∂S. The **closure** of S, denoted by $\overline{S}$, is $\overline{S} = S \cup \partial S$.
3. x_0 is an **isolated point** of S if $x_0 \in S$ and there is a neighborhood of x_0 that contains no other point of S.
4. x_0 is **exterior** to S if x_0 is in the interior of S^c. The collection of such points is the **exterior** of S.

Example 2.1.10 Let $S = (-\infty, -1] \cup (1, 2) \cup \{3\}$. Then

1. The set of limit points of S is $(-\infty, -1] \cup [1, 2]$.
2. $\partial S = \{-1, 1, 2, 3\}$ and $\overline{S} = (-\infty, -1] \cup [1, 2] \cup \{3\}$.
3. 3 is the only isolated point of S.
4. The exterior of S is $(-1, 1) \cup (2, 3) \cup (3, \infty)$.

Example 2.1.11 For $n \geqslant 1$, let

$$I_n = \left[\frac{1}{2n+1}, \frac{1}{2n}\right] \quad \text{and} \quad S = \bigcup_{n=1}^{\infty} I_n.$$

Then

1. The set of limit points of S is $S \cup \{0\}$.
2. $\partial S = \{x \mid x = 0 \text{ or } x = 1/n \ (n \geqslant 2)\}$ and $\overline{S} = S \cup \{0\}$.
3. S has no isolated points.
4. The exterior of S is

$$(-\infty, 0) \cup \left[\bigcup_{n=1}^{\infty} \left(\frac{1}{2n+2}, \frac{1}{2n+1}\right)\right] \cup \left(\frac{1}{2}, \infty\right).$$

Example 2.1.12 Let S be the set of rational numbers. Since every interval contains a rational number, every real number is a limit point of S; thus, $\overline{S} = \mathbb{R}$. Since every interval also contains an irrational number, every real number is a boundary point of S; thus $\partial S = \mathbb{R}$. The interior and exterior of S are both empty, and S has no isolated points. S is neither open nor closed.

The next theorem says that S is closed if and only if $S = \overline{S}$ (Exercise 2.4).

Theorem 2.1.13 A set S is closed if and only if no point of S^c is a limit point of S.

Proof: Suppose that S is closed and $x_0 \in S^c$. Since S^c is open, there is a neighborhood of x_0 that is contained in S^c and therefore contains no points of S. Hence, x_0 cannot be a limit point of S. For the converse, if no point of S^c is a limit point of S then every point in S^c must have a neighborhood contained in S^c. Therefore, S^c is open and S is closed. ■

Theorem 2.1.13 is usually stated as follows.

Corollary 2.1.14 A set is closed if and only if it contains all its limit points.

A **polygonal curve** P is a curve specified by a sequence of points $(A_1, A_2, \ldots, A_n)$ called its vertices. The curve itself consists of the line segments connecting the consecutive vertices.

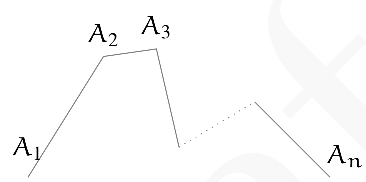


Figure 2.6 Polygonal curve

Definition 2.1.6 A **domain** is a path connected open set. A path connected set D means that any two points of this set can be connected by a polygonal curve lying within D.

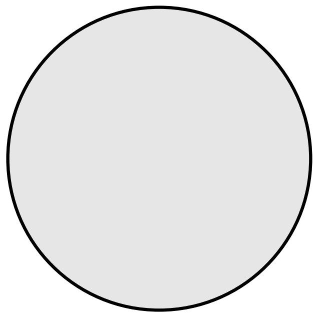
Definition 2.1.7 A **simply connected domain** is a path-connected domain where one can continuously shrink any simple closed curve into a point while remaining in the domain.

Equivalently a pathwise-connected domain $U \subseteq \mathbb{R}^3$ is called **simply connected** if for every simple closed curve $\Gamma \subseteq U$, there exists a surface $\Sigma \subseteq U$ whose boundary is exactly the curve Γ.

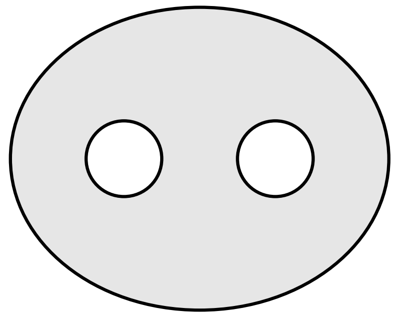
Exercises

Problem 2.1 Determine whether the following subsets of $\mathbb{R}^2$ are open, closed, or neither, in $\mathbb{R}^2$.

1. $A = \{(x,y) \in \mathbb{R}^2 : |x| < 1, |y| < 1\}$
2. $B = \{(x,y) \in \mathbb{R}^2 : |x| < 1, |y| \leqslant 1\}$
3. $C = \{(x,y) \in \mathbb{R}^2 : |x| \leqslant 1, |y| \leqslant 1\}$
4. $D = \{(x,y) \in \mathbb{R}^2 : x^2 \leqslant y \leqslant x\}$
5. $E = \{(x,y) \in \mathbb{R}^2 : xy > 1\}$
6. $F = \{(x,y) \in \mathbb{R}^2 : xy \leqslant 1\}$



(a) Simply connected domain



(b) Non-simply connected domain

Figure 2.7 Domains

7. $G = \{(x,y) \in \mathbb{R}^2 : |y| \leqslant 9, x < y^2\}$

Problem 2.2 — Putnam Exam 1969. Let $p(x,y)$ be a polynomial with real coefficients in the real variables x and y, defined over the entire plane $\mathbb{R}^2$. What are the possibilities for the image (range) of $p(x,y)$?

Problem 2.3 — Putnam 1998. Let $\mathcal{F}$ be a finite collection of open disks in $\mathbb{R}^2$ whose union contains a set $E \subseteq \mathbb{R}^2$. Shew that there is a pairwise disjoint subcollection $D_k, k \geqslant 1$ in $\mathcal{F}$ such that

$$E \subseteq \bigcup_{j=1}^{n} 3D_j.$$

Problem 2.4 A set S is closed if and only if no point of S^c is a limit point of S.

2.2 Limits

We will start with the notion of *limit*.

Definition 2.2.1 A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to have a limit $\mathbf{L} \in \mathbb{R}^m$ at $\mathbf{a} \in \mathbb{R}^n$ if $\forall \varepsilon > 0, \exists \delta > 0$ such that

$$0 < \|\mathbf{x} - \mathbf{a}\| < \delta \implies \|f(\mathbf{x}) - \mathbf{L}\| < \varepsilon.$$

In such a case we write,

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

The notions of infinite limits, limits at infinity, and continuity at a point, are analogously defined.

Theorem 2.2.1 A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ have limit

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

if and only if the coordinates functions $f_1, f_2, \ldots f_m$ have limits $L_1, L2, \ldots, L_m$ respectively, i.e., $f_i \to L_i$.

Proof:

We start with the following observation:

$$\left\|f(\mathbf{x})-L\right\|^2=\left|f_1(\mathbf{x})-L_1\right|^2+\left|f_2(\mathbf{x})-L2\right|^2+\cdots+\left|f_m(\mathbf{x})-L_m\right|^2.$$

So, if

$$\left|f_1(\mathbf{x})-L1\right|<\varepsilon$$
$$\left|f_2(\mathbf{x})-L2\right|<\varepsilon$$
$$\vdots$$
$$\left|f_m(\mathbf{x})-L_m\right|<\varepsilon$$

then $\left\|\mathbf{f}(t)-\mathbf{L}\right\|<\sqrt{n}\varepsilon$.

Now, if $\left\|\mathbf{f}(\mathbf{x})-\mathbf{L}\right\|<\varepsilon$ then

$$\left|f_1(\mathbf{x})-L1\right|<\varepsilon$$
$$\left|f_2(\mathbf{x})-L2\right|<\varepsilon$$
$$\vdots$$
$$\left|f_m(\mathbf{x})-L_m\right|<\varepsilon$$

■

Limits in more than one dimension are perhaps trickier to find, as one must approach the test point from infinitely many directions.

Example 2.2.2 Find $\lim\limits_{(x,y)\to(0,0)}\left(\dfrac{x^2y}{x^2+y^2},\dfrac{x^5y^3}{x^6+y^4}\right)$

Solution: ► First we will calculate $\lim\limits_{(x,y)\to(0,0)}\dfrac{x^2y}{x^2+y^2}$ We use the sandwich theorem. Observe that $0\leqslant x^2\leqslant x^2+y^2$, and so $0\leqslant\dfrac{x^2}{x^2+y^2}\leqslant 1$. Thus

$$\lim_{(x,y)\to(0,0)}0\leqslant\lim_{(x,y)\to(0,0)}\left|\frac{x^2y}{x^2+y^2}\right|\leqslant\lim_{(x,y)\to(0,0)}|y|,$$

and hence

$$\lim_{(x,y)\to(0,0)}\frac{x^2y}{x^2+y^2}=0.$$

Now we find $\lim\limits_{(x,y)\to(0,0)}\dfrac{x^5y^3}{x^6+y^4}$.

Either $|x| \leqslant |y|$ or $|x| \geqslant |y|$. Observe that if $|x| \leqslant |y|$, then

$$\left|\frac{x^5y^3}{x^6+y^4}\right| \leqslant \frac{y^8}{y^4} = y^4.$$

If $|y| \leqslant |x|$, then

$$\left|\frac{x^5y^3}{x^6+y^4}\right| \leqslant \frac{x^8}{x^6} = x^2.$$

Thus

$$\left|\frac{x^5y^3}{x^6+y^4}\right| \leqslant \max(y^4, x^2) \leqslant y^4 + x^2 \longrightarrow 0,$$

as $(x,y) \to (0,0)$.
Aliter: Let $X = x^3, Y = y^2$.

$$\left|\frac{x^5y^3}{x^6+y^4}\right| = \frac{X^{5/3}Y^{3/2}}{X^2+Y^2}.$$

Passing to polar coordinates $X = \rho\cos\theta, Y = \rho\sin\theta$, we obtain

$$\left|\frac{x^5y^3}{x^6+y^4}\right| = \frac{X^{5/3}Y^{3/2}}{X^2+Y^2} = \rho^{5/3+3/2-2}|\cos\theta|^{5/3}|\sin\theta|^{3/2} \leqslant \rho^{7/6} \to 0,$$

as $(x,y) \to (0,0)$.

◀

Example 2.2.3 Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{1+x+y}{x^2-y^2}$.

Solution: ► When $y = 0$,

$$\frac{1+x}{x^2} \to +\infty,$$

as $x \to 0$. When $x = 0$,

$$\frac{1+y}{-y^2} \to -\infty,$$

as $y \to 0$. The limit does not exist. ◀

Example 2.2.4 Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy^6}{x^6+y^8}$.

Solution: ► Putting $x = t^4, y = t^3$, we find

$$\frac{xy^6}{x^6+y^8} = \frac{1}{2t^2} \to +\infty,$$

as $t \to 0$. But when $y = 0$, the function is 0. Thus the limit does not exist. ◀

Example 2.2.5 Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{((x-1)^2+y^2)\log_e((x-1)^2+y^2)}{|x|+|y|}$.

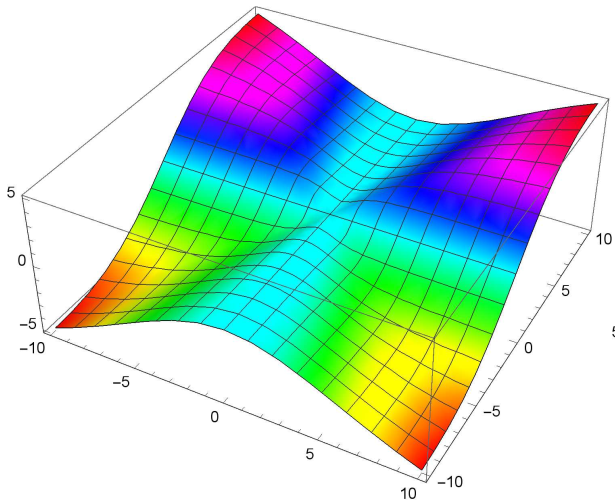


Figure 2.8 Example 2.2.5.

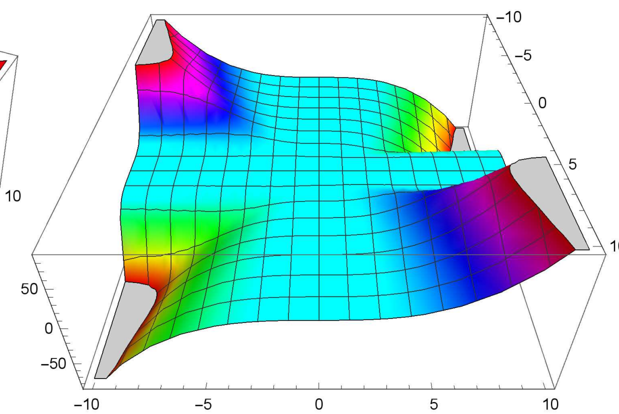


Figure 2.9 Example 2.2.6.

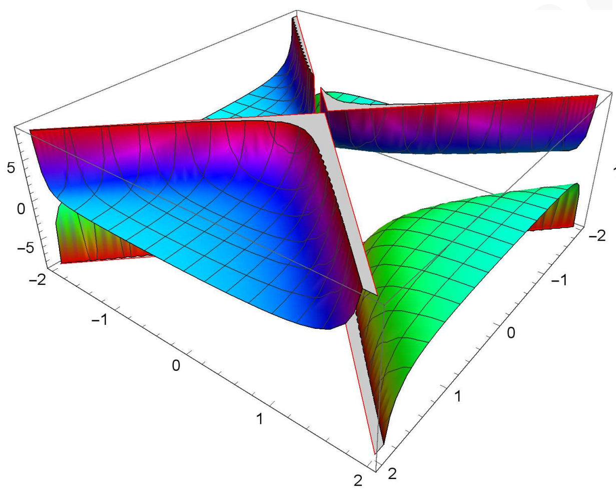


Figure 2.10 Example 2.2.7.

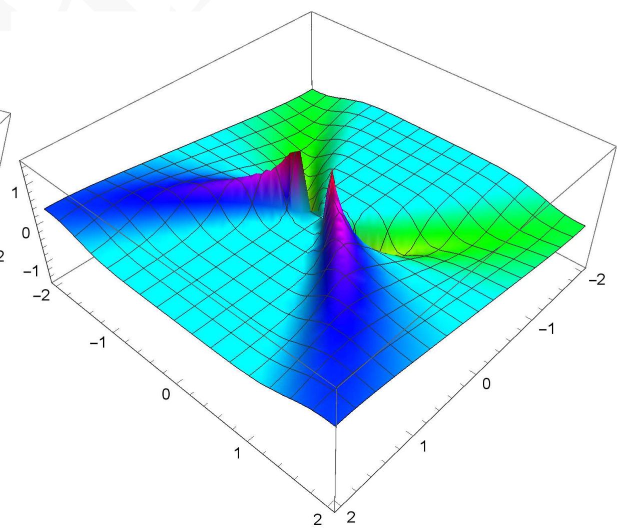


Figure 2.11 Example 2.2.4.

Solution: ► When $y=0$ we have

$$\frac{2(x-1)^2\ln(|1-x|)}{|x|}\sim-\frac{2x}{|x|},$$

and so the function does not have a limit at $(0,0)$. ◄

Example 2.2.6 Find $\lim_{(x,y)\to(0,0)}\frac{\sin(x^4)+\sin(y^4)}{\sqrt{x^4+y^4}}$.

Solution: ► $\sin(x^4)+\sin(y^4)\leqslant x^4+y^4$ and so

$$\left|\frac{\sin(x^4)+\sin(y^4)}{\sqrt{x^4+y^4}}\right|\leqslant\sqrt{x^4+y^4}\to 0,$$

as $(x,y)\to(0,0)$. ◄

Example 2.2.7 Find $\lim_{(x,y)\to(0,0)}\frac{\sin x-y}{x-\sin y}$.

Solution: ► When $y=0$ we obtain

$$\frac{\sin x}{x}\to 1,$$

as $x\to 0$. When $y=x$ the function is identically -1. Thus the limit does not exist. ◄

If $f:\mathbb{R}^2\to\mathbb{R}$, it may be that the limits

$$\lim_{y\to y_0}\left(\lim_{x\to x_0}f(x,y)\right),\qquad \lim_{x\to x_0}\left(\lim_{y\to y_0}f(x,y)\right),$$

both exist. These are called the **iterated limits of** f **as** $(x,y)\to(x_0,y_0)$. The following possibilities might occur.

1. If $\lim_{(x,y)\to(x_0,y_0)}f(x,y)$ exists, then each of the iterated limits $\lim_{y\to y_0}\left(\lim_{x\to x_0}f(x,y)\right)$ and $\lim_{x\to x_0}\left(\lim_{y\to y_0}f(x,y)\right)$ exists.

2. If the iterated limits exist and $\lim_{y\to y_0}\left(\lim_{x\to x_0}f(x,y)\right)\neq\lim_{x\to x_0}\left(\lim_{y\to y_0}f(x,y)\right)$ then $\lim_{(x,y)\to(x_0,y_0)}f(x,y)$ does not exist.

3. It may occur that $\lim_{y\to y_0}\left(\lim_{x\to x_0}f(x,y)\right)=\lim_{x\to x_0}\left(\lim_{y\to y_0}f(x,y)\right)$, but that $\lim_{(x,y)\to(x_0,y_0)}f(x,y)$ does not exist.

4. It may occur that $\lim_{(x,y)\to(x_0,y_0)}f(x,y)$ exists, but one of the iterated limits does not.

Exercises

Problem 2.5 Sketch the domain of definition of $(x,y) \mapsto \sqrt{4-x^2-y^2}$.

Problem 2.6 Sketch the domain of definition of $(x,y) \mapsto \log(x+y)$.

Problem 2.7 Sketch the domain of definition of $(x,y) \mapsto \dfrac{1}{x^2+y^2}$.

Problem 2.8 Find $\displaystyle\lim_{(x,y)\to(0,0)} (x^2+y^2)\sin\frac{1}{xy}$.

Problem 2.9 Find $\displaystyle\lim_{(x,y)\to(0,2)} \frac{\sin xy}{x}$.

Problem 2.10 For what c will the function

$$f(x,y) = \begin{cases} \sqrt{1-x^2-4y^2}, & \text{if } x^2+4y^2 \leqslant 1, \\ c, & \text{if } x^2+4y^2 > 1 \end{cases}$$

be continuous everywhere on the xy-plane?

Problem 2.11 Find

$$\lim_{(x,y)\to(0,0)} \sqrt{x^2+y^2}\sin\frac{1}{x^2+y^2}.$$

Problem 2.12 Find

$$\lim_{(x,y)\to(+\infty,+\infty)} \frac{\max(|x|,|y|)}{\sqrt{x^4+y^4}}.$$

Problem 2.13 Find

$$\lim_{(x,y)\to(0,0} \frac{2x^2\sin y^2+y^4e^{-|x|}}{\sqrt{x^2+y^2}}.$$

Problem 2.14 Demonstrate that

$$\lim_{(x,y,z)\to(0,0,0)} \frac{x^2y^2z^2}{x^2+y^2+z^2} = 0.$$

Problem 2.15 Prove that

$$\lim_{x\to 0}\left(\lim_{y\to 0}\frac{x-y}{x+y}\right) = 1 = -\lim_{y\to 0}\left(\lim_{x\to 0}\frac{x-y}{x+y}\right).$$

Does $\displaystyle\lim_{(x,y)\to(0,0)}\frac{x-y}{x+y}$ exist?.

Problem 2.16 Let

$$f(x,y) = \begin{cases} x\sin\dfrac{1}{x} + y\sin\dfrac{1}{y} & \text{if } x \neq 0,\ y \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Prove that $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ exists, but that the iterated limits $\displaystyle\lim_{x\to 0}\left(\lim_{y\to 0} f(x,y)\right)$ and $\displaystyle\lim_{y\to 0}\left(\lim_{x\to 0} f(x,y)\right)$ do not exist.

Problem 2.17 Prove that

$$\lim_{x\to 0}\left(\lim_{y\to 0}\frac{x^2y^2}{x^2y^2+(x-y)^2}\right) = 0,$$

and that

$$\lim_{y\to 0}\left(\lim_{x\to 0}\frac{x^2y^2}{x^2y^2+(x-y)^2}\right) = 0,$$

but still $\displaystyle\lim_{(x,y)\to(0,0)}\frac{x^2y^2}{x^2y^2+(x-y)^2}$ does not exist.

2.3 Continuity

Definition 2.3.1 Let $U \subset \mathbb{R}^m$ be a domain, and $f : U \to \mathbb{R}^d$ be a function. We say f is continuous at a if $\displaystyle\lim_{x\to a} f(x) = f(a)$.

Definition 2.3.2 If f is continuous at every point $a \in U$, then we say f is continuous on U (or sometimes simply f is continuous).

Again the standard results on continuity from one variable calculus hold. Sums, products, quotients (with a non-zero denominator) and composites of continuous functions will all yield continuous functions.

The notion of continuity gives us a generalization of Proposition **??** that is useful is computing the limits along arbitrary curves instead.

Proposition 2.3.1 Let $f:\mathbb{R}^d \to \mathbb{R}$ be a function, and $a \in \mathbb{R}^d$. Let $\gamma:[0,1]\to\mathbb{R}^d$ be a *any continuous function* with $\gamma(0)=a$, and $\gamma(t)\neq a$ for all $t>0$. If $\lim_{x\to a} f(x)=l$, then we must have $\lim_{t\to 0} f(\gamma(t))=l$.

Corollary 2.3.2 If there exists two continuous functions $\gamma_1,\gamma_2:[0,1]\to\mathbb{R}^d$ such that for $i\in 1,2$ we have $\gamma_i(0)=a$ and $\gamma_i(t)\neq a$ for all $t>0$. If $\lim_{t\to 0} f(\gamma_1(t)) \neq \lim_{t\to 0} f(\gamma_2(t))$ then $\lim_{x\to a} f(x)$ can not exist.

Theorem 2.3.3 The vector function $f:\mathbb{R}^d\to\mathbb{R}$ is **continuous at** t_0 if and only if the coordinates functions $f_1, f_2, \ldots f_n$ are continuous at t_0.

The proof of this Theorem is very similar to the proof of Theorem 2.2.1.

Exercises

Problem 2.18 Sketch the domain of definition of $(x,y)\mapsto\sqrt{4-x^2-y^2}$.

Problem 2.19 Sketch the domain of definition of $(x,y)\mapsto\log(x+y)$.

Problem 2.20 Sketch the domain of definition of $(x,y)\mapsto\dfrac{1}{x^2+y^2}$.

Problem 2.21 Find $\displaystyle\lim_{(x,y)\to(0,0)}(x^2+y^2)\sin\frac{1}{xy}$.

Problem 2.22 Find $\displaystyle\lim_{(x,y)\to(0,2)}\frac{\sin xy}{x}$.

Problem 2.23 For what c will the function

$$f(x,y)=\begin{cases}\sqrt{1-x^2-4y^2}, & \text{if } x^2+4y^2\leqslant 1,\\ c, & \text{if } x^2+4y^2>1\end{cases}$$

be continuous everywhere on the xy-plane?

Problem 2.24 Find

$$\lim_{(x,y)\to(0,0)}\sqrt{x^2+y^2}\sin\frac{1}{x^2+y^2}.$$

Problem 2.25 Find

$$\lim_{(x,y)\to(+\infty,+\infty)}\frac{\max(|x|,|y|)}{\sqrt{x^4+y^4}}.$$

Problem 2.26 Find

$$\lim_{(x,y)\to(0,0}\frac{2x^2\sin y^2+y^4e^{-|x|}}{\sqrt{x^2+y^2}}.$$

Problem 2.27 Demonstrate that

$$\lim_{(x,y,z)\to(0,0,0)}\frac{x^2y^2z^2}{x^2+y^2+z^2}=0.$$

Problem 2.28 Prove that

$$\lim_{x\to 0}\left(\lim_{y\to 0}\frac{x-y}{x+y}\right)=1=-\lim_{y\to 0}\left(\lim_{x\to 0}\frac{x-y}{x+y}\right).$$

Does $\displaystyle\lim_{(x,y)\to(0,0)}\frac{x-y}{x+y}$ exist?.

Problem 2.29 Let

$$f(x,y) = \begin{cases} x\sin\dfrac{1}{x} + y\sin\dfrac{1}{y} & \text{if } x \neq 0,\ y \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Prove that $\lim_{(x,y)\to(0,0)} f(x,y)$ exists, but that the iterated limits $\lim_{x\to 0}\left(\lim_{y\to 0} f(x,y)\right)$ and $\lim_{y\to 0}\left(\lim_{x\to 0} f(x,y)\right)$ do not exist.

Problem 2.30 Prove that

$$\lim_{x\to 0}\left(\lim_{y\to 0} \frac{x^2y^2}{x^2y^2+(x-y)^2}\right) = 0,$$

and that

$$\lim_{y\to 0}\left(\lim_{x\to 0} \frac{x^2y^2}{x^2y^2+(x-y)^2}\right) = 0,$$

but still $\lim_{(x,y)\to(0,0)} \dfrac{x^2y^2}{x^2y^2+(x-y)^2}$ does not exist.

2.4 ⋆ Compactness

The next definition generalizes the definition of the diameter of a circle or sphere.

Definition 2.4.1 If S is a nonempty subset of $\mathbb{R}^n$, then

$$d(S) = \sup\left\{|\mathbf{x}-\mathbf{Y}|\right\} \mathbf{x}, \mathbf{Y} \in S$$

is the **diameter** of S. If $d(S) < \infty$, S is **bounded**; if $d(S) = \infty$, S is **unbounded**.

Theorem 2.4.1 — Principle of Nested Sets. If $S_1, S_2, \ldots$ are closed nonempty subsets of $\mathbb{R}^n$ such that

$$S_1 \supset S_2 \supset \cdots \supset S_r \supset \cdots \tag{2.1}$$

and

$$\lim_{r\to\infty} d(S_r) = 0, \tag{2.2}$$

then the intersection

$$I = \bigcap_{r=1}^{\infty} S_r$$

contains exactly one point.

Proof: Let $\{\mathbf{x}_r\}$ be a sequence such that $\mathbf{x}_r \in S_r$ $(r \geqslant 1)$. Because of (2.1), $\mathbf{x}_r \in S_k$ if $r \geqslant k$, so

$$|\mathbf{x}_r - \mathbf{x}_s| < d(S_k) \quad \text{if} \quad r, s \geqslant k.$$

From (2.2) and Theorem 2.1.9, $\mathbf{x}_r$ converges to a limit $\overline{\mathbf{x}}$. Since $\overline{\mathbf{x}}$ is a limit point of every S_k and every S_k is closed, $\overline{\mathbf{x}}$ is in every S_k (Corollary 2.1.14). Therefore, $\overline{\mathbf{x}} \in I$, so $I \neq \emptyset$. Moreover, $\overline{\mathbf{x}}$ is the only point in I, since if $\mathbf{Y} \in I$, then

$$|\overline{\mathbf{x}} - \mathbf{Y}| \leqslant d(S_k), \quad k \geqslant 1,$$

and (2.2) implies that $\mathbf{Y} = \overline{\mathbf{x}}$. ■

We can now prove the Heine–Borel theorem for $\mathbb{R}^n$. This theorem concerns **compact** sets. As in $\mathbb{R}$, a compact set in $\mathbb{R}^n$ is a closed and bounded set.

Recall that a collection $\mathcal{H}$ of open sets is an open covering of a set S if

$$S \subset \cup\{H\} H \in \mathcal{H}.$$

Theorem 2.4.2 — Heine–Borel Theorem. If $\mathcal{H}$ is an open covering of a compact subset S, then S can be covered by finitely many sets from $\mathcal{H}$.

Proof: The proof is by contradiction. We first consider the case where $n = 2$, so that you can visualize the method. Suppose that there is a covering $\mathcal{H}$ for S from which it is impossible to select a finite subcovering. Since S is bounded, S is contained in a closed square

$$T = \{(x,y) | a_1 \leqslant x \leqslant a_1 + L, a_2 \leqslant x \leqslant a_2 + L\}$$

with sides of length L (Figure **??**).

Bisecting the sides of T as shown by the dashed lines in Figure **??** leads to four closed squares, $T^{(1)}, T^{(2)}$, $T^{(3)}$, and $T^{(4)}$, with sides of length $L/2$. Let

$$S^{(i)} = S \cap T^{(i)}, \quad 1 \leqslant i \leqslant 4.$$

Each $S^{(i)}$, being the intersection of closed sets, is closed, and

$$S = \bigcup_{i=1}^{4} S^{(i)}.$$

Moreover, $\mathcal{H}$ covers each $S^{(i)}$, but at least one $S^{(i)}$ cannot be covered by any finite subcollection of $\mathcal{H}$, since if all the $S^{(i)}$ could be, then so could S. Let S_1 be a set with this property, chosen from $S^{(1)}$, $S^{(2)}$, $S^{(3)}$, and $S^{(4)}$. We are now back to the situation we started from: a compact set S_1 covered by $\mathcal{H}$, but not by any finite subcollection of $\mathcal{H}$. However, S_1 is contained in a square T_1 with sides of length $L/2$ instead of L. Bisecting the sides of T_1 and repeating the argument, we obtain a subset S_2 of S_1 that has the same properties as S, except that it is contained in a square with sides of length $L/4$. Continuing in this way produces a sequence of nonempty closed sets $S_0 \,(= S)$, S_1, S_2, …, such that $S_k \supset S_{k+1}$ and $d(S_k) \leqslant L/2^{k-1/2}\,(k \geqslant 0)$. From Theorem 2.4.1, there is a point $\overline{\mathbf{x}}$ in $\bigcap_{k=1}^{\infty} S_k$. Since $\overline{\mathbf{x}} \in S$, there is an open set H in $\mathcal{H}$ that contains $\overline{\mathbf{x}}$, and this H must also contain some ε-neighborhood of $\overline{\mathbf{x}}$. Since every $\mathbf{x}$ in S_k satisfies the inequality

$$|\mathbf{x}-\overline{\mathbf{x}}| \leqslant 2^{-k+1/2}L,$$

it follows that $S_k \subset H$ for k sufficiently large. This contradicts our assumption on $\mathcal{H}$, which led us to believe that no S_k could be covered by a finite number of sets from $\mathcal{H}$. Consequently, this assumption must be false: $\mathcal{H}$ must have a finite subcollection that covers S. This completes the proof for $n = 2$.

The idea of the proof is the same for $n > 2$. The counterpart of the square T is the **hypercube** with sides of length L:

$$T = \left\{(x_1, x_2, \ldots, x_n)\right\} a_i \leqslant x_i \leqslant a_i + L, i = 1, 2, \ldots, n.$$

Halving the intervals of variation of the n coordinates x_1, x_2, …, x_n divides T into 2^n closed hypercubes with sides of length $L/2$:

$$T^{(i)} = \left\{(x_1, x_2, \ldots, x_n)\right\} b_i \leqslant x_i \leqslant b_i + L/2, 1 \leqslant i \leqslant n,$$

where $b_i = a_i$ or $b_i = a_i + L/2$. If no finite subcollection of $\mathcal{H}$ covers S, then at least one of these smaller hypercubes must contain a subset of S that is not covered by any finite subcollection of S. Now the proof proceeds as for $n = 2$. ■

Theorem 2.4.3 — Bolzano-Weierstrass. Every bounded infinite set of real numbers has at least one limit point.

Proof: We will show that a bounded nonempty set without a limit point can contain only a finite number of points. If S has no limit points, then S is closed (Theorem 2.1.13) and every point x of S has an open neighborhood N_x that contains no point of S other than x. The collection

$$\mathcal{H} = \{N_x\} x \in S$$

is an open covering for S. Since S is also bounded, implies that S can be covered by a finite collection of sets from $\mathcal{H}$, say N_{x_1}, …, N_{x_n}. Since these sets contain only x_1, …, x_n from S, it follows that $S = \{x_1, \ldots, x_n\}$. ■

Chapter 3

Differentiation of Vector Function

In this chapter we consider functions $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$. This functions are usually classified based on the dimensions n and m:

❶ if the dimensions n and m are equal to 1, such a function is called a **real function of a real variable**.

❷ if $m = 1$ and $n > 1$ the function is called a real-valued function of a vector variable or, more briefly, a **scalar field**.

❸ if $n = 1$ and $m > 1$ it is called a **vector-valued function of a real variable**.

❹ if $n > 1$ and $m > 1$ it is called a vector-valued function of a vector variable, or simply a **vector field**.

We suppose that the cases of real function of a real variable and of scalar fields have been studied before.

This chapter extends the concepts of limit, continuity, and derivative to vector-valued function and vector fields.

We start with the simplest one: vector-valued function.

3.1 Differentiation of Vector Function of a Real Variable

Definition 3.1.1 A **vector-valued function of a real variable** is a rule that associates a vector $\mathbf{f}(t)$ with a real number t, where t is in some subset D of $\mathbb{R}$ (called the **domain** of **f**). We write $\mathbf{f}: D \to \mathbb{R}^n$ to denote that **f** is a mapping of D into $\mathbb{R}^n$.

$$\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$$
$$\mathbf{f}(t) = \big(f_1(t), f_2(t), \ldots, f_n(t)\big)$$

with

$$f_1, f_2, \ldots, f_n : \mathbb{R} \to \mathbb{R}.$$

called the **component functions** of **f**.

In $\mathbb{R}^3$ vector-valued function of a real variable can be written in component form as

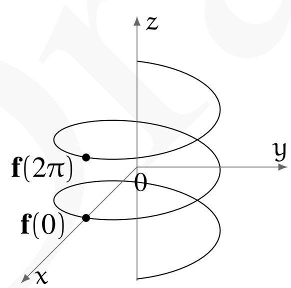
$$\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j} + f_3(t)\mathbf{k}$$

or in the form

$$\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$$

for some real-valued functions $f_1(t)$, $f_2(t)$, $f_3(t)$. The first form is often used when emphasizing that $\mathbf{f}(t)$ is a vector, and the second form is useful when considering just the terminal points of the vectors. By identifying vectors with their terminal points, a curve in space can be written as a vector-valued function.

Example 3.1.1 For example, $\mathbf{f}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k}$ is a vector-valued function in $\mathbb{R}^3$, defined for all real numbers t. At $t = 1$ the value of the function is the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$, which in Cartesian coordinates has the terminal point $(1, 1, 1)$.



Example 3.1.2 Define $\mathbf{f} : \mathbb{R} \to \mathbb{R}^3$ by $\mathbf{f}(t) = (\cos t, \sin t, t)$.
This is the equation of a *helix* (see Figure 1.8.1). As the value of t increases, the terminal points of $\mathbf{f}(t)$ trace out a curve spiraling upward. For each t, the x- and y-coordinates of $\mathbf{f}(t)$ are $x = \cos t$ and $y = \sin t$, so

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1.$$

Thus, the curve lies on the surface of the right circular cylinder $x^2 + y^2 = 1$.

It may help to think of vector-valued functions of a real variable in $\mathbb{R}^n$ as a generalization of the parametric functions in $\mathbb{R}^2$ which you learned about in single-variable calculus. Much of the theory of real-valued functions of a single real variable can be applied to vector-valued functions of a real variable.

Definition 3.1.2 Let $\mathbf{f}(t) = (f_1(t), f_2(t), \ldots, f_n(t))$ be a vector-valued function, and let a be a real number in its domain. The **derivative** of $\mathbf{f}(t)$ at a, denoted by $\mathbf{f}'(a)$ or $\frac{d\mathbf{f}}{dt}(a)$, is the limit

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

if that limit exists. Equivalently, $\mathbf{f}'(a) = (f_1'(a), f_2'(a), \ldots, f_n'(a))$, if the component derivatives exist. We say that $\mathbf{f}(t)$ is **differentiable** at a if $\mathbf{f}'(a)$ exists.

The derivative of a vector-valued function is a **tangent vector** to the curve in space which the function represents, and it lies on the *tangent line* to the curve (see Figure 3.1).

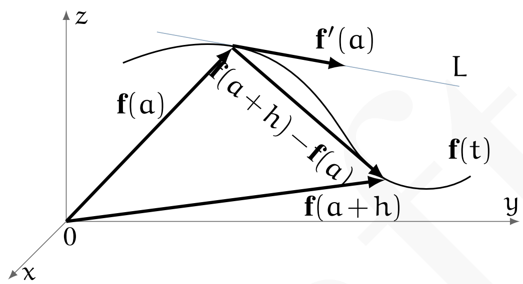


Figure 3.1 Tangent vector $\mathbf{f}'(a)$ and tangent line $L = \mathbf{f}(a) + s\mathbf{f}'(a)$

Example 3.1.3 Let $\mathbf{f}(t) = (\cos t, \sin t, t)$. Then $\mathbf{f}'(t) = (-\sin t, \cos t, 1)$ for all t. The tangent line L to the curve at $\mathbf{f}(2\pi) = (1, 0, 2\pi)$ is $L = \mathbf{f}(2\pi) + s\,\mathbf{f}'(2\pi) = (1, 0, 2\pi) + s(0, 1, 1)$, or in parametric form: $x = 1$, $y = s$, $z = 2\pi + s$ for $-\infty < s < \infty$.

Note that if $u(t)$ is a scalar function and $\mathbf{f}(t)$ is a vector-valued function, then their product, defined by $(u\mathbf{f})(t) = u(t)\mathbf{f}(t)$ for all t, is a vector-valued function (since the product of a scalar with a vector is a vector).

The basic properties of derivatives of vector-valued functions are summarized in the following theorem.

Theorem 3.1.4 Let $\mathbf{f}(t)$ and $\mathbf{g}(t)$ be differentiable vector-valued functions, let $u(t)$ be a differentiable scalar function, let k be a scalar, and let $\mathbf{c}$ be a constant vector. Then

❶ $\dfrac{d}{dt}\mathbf{c} = \mathbf{0}$

❷ $\dfrac{d}{dt}(k\mathbf{f}) = k\dfrac{d\mathbf{f}}{dt}$

❸ $\dfrac{d}{dt}(\mathbf{f}+\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} + \dfrac{d\mathbf{g}}{dt}$

❹ $\dfrac{d}{dt}(\mathbf{f}-\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} - \dfrac{d\mathbf{g}}{dt}$

❺ $\dfrac{d}{dt}(u\mathbf{f}) = \dfrac{du}{dt}\mathbf{f} + u\dfrac{d\mathbf{f}}{dt}$

❻ $\dfrac{d}{dt}(\mathbf{f}\bullet\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\bullet\mathbf{g} + \mathbf{f}\bullet\dfrac{d\mathbf{g}}{dt}$

❼ $\dfrac{d}{dt}(\mathbf{f}\times\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\times\mathbf{g} + \mathbf{f}\times\dfrac{d\mathbf{g}}{dt}$

Proof: The proofs of parts (1)-(5) follow easily by differentiating the component functions and using the rules for derivatives from single-variable calculus. We will prove part (6), and leave the proof of part (7) as an exercise for the reader.

(6) Write $\mathbf{f}(t) = \big(f_1(t), f_2(t), f_3(t)\big)$ and $\mathbf{g}(t) = \big(g_1(t), g_2(t), g_3(t)\big)$, where the component functions $f_1(t)$, $f_2(t)$, $f_3(t)$, $g_1(t)$, $g_2(t)$, $g_3(t)$ are all differentiable real-valued functions. Then

$$
\begin{aligned}
\frac{d}{dt}(\mathbf{f}(t)\bullet\mathbf{g}(t)) &= \frac{d}{dt}(f_1(t)\,g_1(t) + f_2(t)\,g_2(t) + f_3(t)\,g_3(t)) \\
&= \frac{d}{dt}(f_1(t)\,g_1(t)) + \frac{d}{dt}(f_2(t)\,g_2(t)) + \frac{d}{dt}(f_3(t)\,g_3(t)) \\
&= \frac{df_1}{dt}(t)\,g_1(t) + f_1(t)\,\frac{dg_1}{dt}(t) + \frac{df_2}{dt}(t)\,g_2(t) + f_2(t)\,\frac{dg_2}{dt}(t) + \frac{df_3}{dt}(t)\,g_3(t) + f_3(t)\,\frac{dg_3}{dt}(t) \\
&= \left(\frac{df_1}{dt}(t), \frac{df_2}{dt}(t), \frac{df_3}{dt}(t)\right)\bullet\big(g_1(t), g_2(t), g_3(t)\big) \\
&\quad + \big(f_1(t), f_2(t), f_3(t)\big)\bullet\left(\frac{dg_1}{dt}(t), \frac{dg_2}{dt}(t), \frac{dg_3}{dt}(t)\right) \qquad (3.1) \\
&= \frac{d\mathbf{f}}{dt}(t)\bullet\mathbf{g}(t) + \mathbf{f}(t)\bullet\frac{d\mathbf{g}}{dt}(t) \ \text{ for all } t. \ \blacksquare \qquad (3.2)
\end{aligned}
$$

■

Example 3.1.5 Suppose $\mathbf{f}(t)$ is differentiable. Find the derivative of $\left\|\mathbf{f}(t)\right\|$. **Solution:** ▶

Since $\left\|\mathbf{f}(t)\right\|$ is a real-valued function of t, then by the Chain Rule for real-valued functions, we know that $\frac{d}{dt}\left\|\mathbf{f}(t)\right\|^2 = 2\left\|\mathbf{f}(t)\right\| \frac{d}{dt}\left\|\mathbf{f}(t)\right\|$.

But $\left\|\mathbf{f}(t)\right\|^2 = \mathbf{f}(t)\cdot\mathbf{f}(t)$, so $\frac{d}{dt}\left\|\mathbf{f}(t)\right\|^2 = \frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t))$. Hence, we have

$$\begin{aligned} 2\left\|\mathbf{f}(t)\right\| \frac{d}{dt}\left\|\mathbf{f}(t)\right\| &= \frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t)) = \mathbf{f}'(t)\cdot\mathbf{f}(t) + \mathbf{f}(t)\cdot\mathbf{f}'(t) \text{ by Theorem 3.1.4(f), so} \\ &= 2\mathbf{f}'(t)\cdot\mathbf{f}(t)\text{ , so if } \left\|\mathbf{f}(t)\right\| \neq 0 \text{ then} \\ \frac{d}{dt}\left\|\mathbf{f}(t)\right\| &= \frac{\mathbf{f}'(t)\cdot\mathbf{f}(t)}{\left\|\mathbf{f}(t)\right\|}. \end{aligned}$$

◀

We know that $\left\|\mathbf{f}(t)\right\|$ is constant if and only if $\frac{d}{dt}\left\|\mathbf{f}(t)\right\| = 0$ for all t. Also, $\mathbf{f}(t) \perp \mathbf{f}'(t)$ if and only if $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$. Thus, the above example shows this important fact:

Proposition 3.1.6 If $\left\|\mathbf{f}(t)\right\| \neq 0$, then $\left\|\mathbf{f}(t)\right\|$ is constant if and only if $\mathbf{f}(t) \perp \mathbf{f}'(t)$ for all t.

This means that if a curve lies completely on a sphere (or circle) centered at the origin, then the tangent vector $\mathbf{f}'(t)$ is always perpendicular to the *position vector* $\mathbf{f}(t)$.

Example 3.1.7 The *spherical spiral* $\mathbf{f}(t) = \left(\frac{\cos t}{\sqrt{1+a^2t^2}}, \frac{\sin t}{\sqrt{1+a^2t^2}}, \frac{-at}{\sqrt{1+a^2t^2}}\right)$, for $a \neq 0$.

Figure 3.2 shows the graph of the curve when $a = 0.2$. In the exercises, the reader will be asked to show that this curve lies on the sphere $x^2+y^2+z^2 = 1$ and to verify directly that $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$ for all tJust as in single-variable calculus, higher-order derivatives of vector-valued functions are obtained by repeatedly differentiating the (first) derivative of the function:

$$\mathbf{f}''(t) = \frac{d}{dt}\mathbf{f}'(t)\,, \qquad \mathbf{f}'''(t) = \frac{d}{dt}\mathbf{f}''(t)\,, \quad \ldots \quad , \quad \frac{d^n\mathbf{f}}{dt^n} = \frac{d}{dt}\left(\frac{d^{n-1}\mathbf{f}}{dt^{n-1}}\right) \quad (\text{for } n = 2,3,4,\ldots)$$

We can use vector-valued functions to represent physical quantities, such as velocity, acceleration, force, momentum, etc. For example, let the real variable t represent time elapsed from some initial time ($t = 0$), and suppose that an object of constant mass m is subjected to some force so that it moves in space, with its position (x,y,z) at time t a function of t. That is, $x = x(t)$, $y = y(t)$, $z = z(t)$ for some real-valued functions $x(t)$, $y(t)$, $z(t)$. Call $\mathbf{r}(t) = (x(t),y(t),z(t))$ the **position vector** of the object. We can define various physical quantities associated with the object as

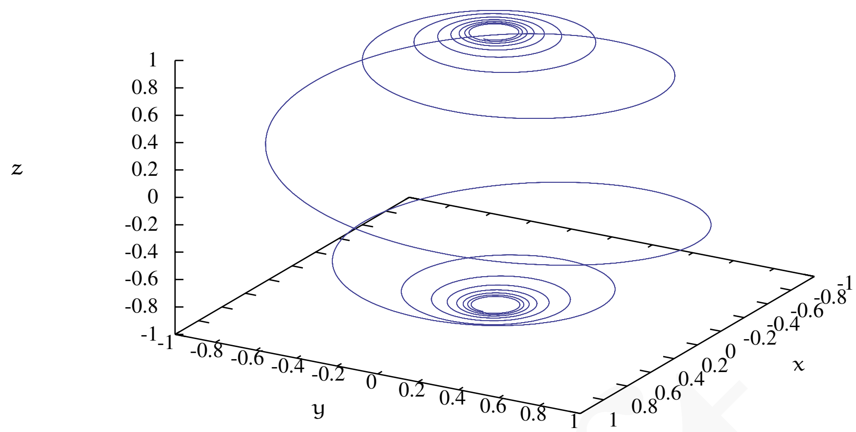


Figure 3.2 Spherical spiral with $a = 0.2$

follows:[1]

$$
\begin{aligned}
\textit{position}\colon\ \mathbf{r}(t) &= (x(t), y(t), z(t)) \\
\textit{velocity}\colon\ \mathbf{v}(t) &= \dot{\mathbf{r}}(t) = \mathbf{r}'(t) = \frac{d\mathbf{r}}{dt} \\
&= (x'(t), y'(t), z'(t)) \\
\textit{acceleration}\colon\ \mathbf{a}(t) &= \dot{\mathbf{v}}(t) = \mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} \\
&= \ddot{\mathbf{r}}(t) = \mathbf{r}''(t) = \frac{d^2\mathbf{r}}{dt^2} \\
&= (x''(t), y''(t), z''(t)) \\
\textit{momentum}\colon\ \mathbf{p}(t) &= m\mathbf{v}(t) \\
\textit{force}\colon\ \mathbf{F}(t) &= \dot{\mathbf{p}}(t) = \mathbf{p}'(t) = \frac{d\mathbf{p}}{dt} \quad \text{(Newton's Second Law of Motion)}
\end{aligned}
$$

The magnitude $\|\mathbf{v}(t)\|$ of the velocity vector is called the *speed* of the object. Note that since the mass m is a constant, the force equation becomes the familiar $\mathbf{F}(t) = m\mathbf{a}(t)$.

Example 3.1.8 Let $\mathbf{r}(t) = (5\cos t, 3\sin t, 4\sin t)$ be the position vector of an object at time $t \geqslant 0$. Find its (a) velocity and (b) acceleration vectors.

[1]We will often use the older dot notation for derivatives when physics is involved.

Solution: ►

(a) $\mathbf{v}(t) = \dot{\mathbf{r}}(t) = (-5\sin t, 3\cos t, 4\cos t)$

(b) $\mathbf{a}(t) = \dot{\mathbf{v}}(t) = (-5\cos t, -3\sin t, -4\sin t)$

Note that $\|\mathbf{r}(t)\| = \sqrt{25\cos^2 t + 25\sin^2 t} = 5$ for all t, so by Example 3.1.5 we know that $\mathbf{r}(t)\bullet\dot{\mathbf{r}}(t) = 0$ for all t (which we can verify from part (a)). In fact, $\|\mathbf{v}(t)\| = 5$ for all t also. And not only does $\mathbf{r}(t)$ lie on the sphere of radius 5 centered at the origin, but perhaps not so obvious is that it lies completely within a *circle* of radius 5 centered at the origin. Also, note that $\mathbf{a}(t) = -\mathbf{r}(t)$. It turns out (see Exercise 16) that whenever an object moves in a circle with constant speed, the acceleration vector will point in the opposite direction of the position vector (i.e. towards the center of the circle). ◄

3.1.1 Antiderivatives

Definition 3.1.3 An **antiderivative** of a vector-valued function $\mathbf{f}$ is a vector-valued function $\mathbf{F}$ such that

$$\mathbf{F}'(t) = \mathbf{f}(t).$$

The **indefinite integral** $\int \mathbf{f}(t)\ dt$ of a vector-valued function $\mathbf{f}$ is the general antiderivative of $\mathbf{f}$ and represents the collection of all antiderivatives of $\mathbf{f}$.

The same reasoning that allows us to differentiate a vector-valued function componentwise applies to integrating as well. Recall that the integral of a sum is the sum of the integrals and also that we can remove constant factors from integrals. So, given $\mathbf{f}(t) = x(t)\ vi + y(t)\mathbf{j} + z(t)\mathbf{k}$, it follows that we can integrate componentwise. Expressed more formally,

If $\mathbf{f}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, then

$$\int \mathbf{f}(t)\ dt = \left(\int x(t)\ dt\right)\mathbf{i} + \left(\int y(t)\ dt\right)\mathbf{j} + \left(\int z(t)\ dt\right)\mathbf{k}.$$

Proposition 3.1.9 Two antiderivarives of $\mathbf{f}(t)$ differs by a vector, i.e., if $\mathbf{F}(t)$ and $\mathbf{G}(t)$ are antiderivatives of f then exists $\mathbf{c} \in \mathbb{R}^n$ such that

$$\mathbf{F}(t) - \mathbf{G}(t) = \mathbf{c}$$

Exercises

Problem 3.1 For Exercises 1-4, calculate $\mathbf{f}'(t)$ and find the tangent line at $\mathbf{f}(0)$.

1. $\mathbf{f}(t) = (t+1, t^2+1, t^3+1)$

2. $\mathbf{f}(t) = (e^t+1, e^{2t}+1, e^{t^2}+1)$

3. $\mathbf{f}(t) = (\cos 2t, \sin 2t, t)$ **4.** $\mathbf{f}(t) = (\sin 2t, 2\sin^2 t, 2\cos t)$

For Exercises 5-6, find the velocity $\mathbf{v}(t)$ and acceleration $\mathbf{a}(t)$ of an object with the given position vector $\mathbf{r}(t)$.

5. $\mathbf{r}(t) = (t, t - \sin t, 1 - \cos t)$ **6.** $\mathbf{r}(t) = (3\cos t, 2\sin t, 1)$

Problem 3.2 1. Let

$$\mathbf{f}(t) = \left(\frac{\cos t}{\sqrt{1+a^2t^2}}, \frac{\sin t}{\sqrt{1+a^2t^2}}, \frac{-at}{\sqrt{1+a^2t^2}}\right),$$

with $a \neq 0$.

(a) Show that $\|\mathbf{f}(t)\| = 1$ for all t.

(b) Show directly that $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$ for all t.

2. If $\mathbf{f}'(t) = \mathbf{0}$ for all t in some interval (a, b), show that $\mathbf{f}(t)$ is a constant vector in (a, b).

3. For a constant vector $\mathbf{c} \neq \mathbf{0}$, the function $\mathbf{f}(t) = t\mathbf{c}$ represents a line parallel to $\mathbf{c}$.

(a) What kind of curve does $\mathbf{g}(t) = t^3\mathbf{c}$ represent? Explain.

(b) What kind of curve does $\mathbf{h}(t) = e^t\mathbf{c}$ represent? Explain.

(c) Compare $\mathbf{f}'(0)$ and $\mathbf{g}'(0)$. Given your answer to part (a), how do you explain the difference in the two derivatives?

4. Show that $\dfrac{d}{dt}\left(\mathbf{f} \times \dfrac{d\mathbf{f}}{dt}\right) = \mathbf{f} \times \dfrac{d^2\mathbf{f}}{dt^2}$.

5. Let a particle of (constant) mass m have position vector $\mathbf{r}(t)$, velocity $\mathbf{v}(t)$, acceleration $\mathbf{a}(t)$ and momentum $\mathbf{p}(t)$ at time t. The *angular momentum* $\mathbf{L}(t)$ of the particle with respect to the origin at time t is defined as $\mathbf{L}(t) = \mathbf{r}(t) \times \mathbf{p}(t)$. If $\mathbf{F}(t)$ is the force acting on the particle at time t, then define the *torque* $\mathbf{N}(t)$ acting on the particle with respect to the origin as $\mathbf{N}(t) = \mathbf{r}(t) \times \mathbf{F}(t)$. Show that $\mathbf{L}'(t) = \mathbf{N}(t)$.

6. Show that $\dfrac{d}{dt}(\mathbf{f} \cdot (\mathbf{g} \times \mathbf{h})) = \dfrac{d\mathbf{f}}{dt} \cdot (\mathbf{g} \times \mathbf{h}) + \mathbf{f} \cdot \left(\dfrac{d\mathbf{g}}{dt} \times \mathbf{h}\right) + \mathbf{f} \cdot \left(\mathbf{g} \times \dfrac{d\mathbf{h}}{dt}\right)$.

7. The Mean Value Theorem does not hold for vector-valued functions: Show that for $\mathbf{f}(t) = (\cos t, \sin t, t)$, there is no t in the interval $(0, 2\pi)$ such that

$$\mathbf{f}'(t) = \frac{\mathbf{f}(2\pi) - \mathbf{f}(0)}{2\pi - 0}.$$

Problem 3.3 1. The Bézier curve $\mathbf{b}_0^3(t)$ for four noncollinear points $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ in $\mathbb{R}^n$ is defined by the following algorithm (going from the left column to the right):

$$\begin{array}{lll}
\mathbf{b}_0^1(t) = (1-t)\mathbf{b}_0 + t\mathbf{b}_1 & \mathbf{b}_0^2(t) = (1-t)\mathbf{b}_0^1(t) + t\mathbf{b}_1^1(t) & \mathbf{b} \\
\mathbf{b}_1^1(t) = (1-t)\mathbf{b}_1 + t\mathbf{b}_2 & \mathbf{b}_1^2(t) = (1-t)\mathbf{b}_1^1(t) + t\mathbf{b}_2^1(t) & \\
\mathbf{b}_2^1(t) = (1-t)\mathbf{b}_2 + t\mathbf{b}_3 & &
\end{array}$$

(a) Show that $\mathbf{b}_0^3(t) = (1-t)^3\mathbf{b}_0 + 3t(1-t)^2\mathbf{b}_1 + 3t^2(1-t)\mathbf{b}_2 + t^3\mathbf{b}_3$.

(b) Write the explicit formula (as in Example **??**) for the Bézier curve for the points $\mathbf{b}_0 = (0,0,0)$, $\mathbf{b}_1 = (0,1,1)$, $\mathbf{b}_2 = (2,3,0)$, $\mathbf{b}_3 = (4,5,2)$.

2. Let $\mathbf{r}(t)$ be the position vector for a particle moving in $\mathbb{R}^n$. Show that

$$\frac{d}{dt}(\mathbf{r} \times (\mathbf{v} \times \mathbf{r})) = \|\mathbf{r}\|^2\mathbf{a} + (\mathbf{r} \cdot \mathbf{v})\mathbf{v} - (\|\mathbf{v}\|^2 + \mathbf{r} \cdot \mathbf{a})\mathbf{r}.$$

3. Let $\mathbf{r}(t)$ be the position vector in $\mathbb{R}^n$ for a particle that moves with constant speed $c > 0$ in a circle of radius $a > 0$ in the xy-plane. Show that $\mathbf{a}(t)$ points in the opposite direction as $\mathbf{r}(t)$ for all t. (*Hint: Use Example 3.1.5 to show that* $\mathbf{r}(t) \perp \mathbf{v}(t)$ *and* $\mathbf{a}(t) \perp \mathbf{v}(t)$, *and hence* $\mathbf{a}(t) \parallel \mathbf{r}(t)$.)

4. Prove Theorem 3.1.4(g).

3.2 Kepler Law

Why do planets have elliptical orbits? In this section we will solve the two body system problem, i.e., describe the trajectory of two body that interact under the force of gravity. In particular we will proof that the trajectory of a body is a ellipse with focus on the other body.

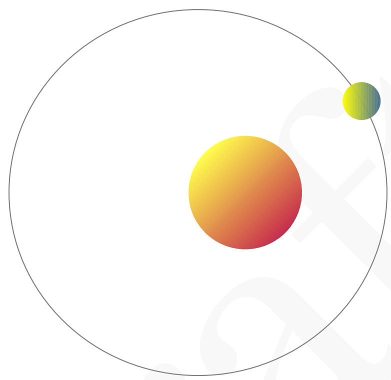


Figure 3.3 Two Body System

We will made two simplifying assumptions:

❶ The bodies are spherically symmetric and can be treated as point masses.

❷ There are no external or internal forces acting upon the bodies other than their mutual gravitation.

Two point mass objects with masses m_1 and m_2 and position vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ relative to some inertial reference frame experience gravitational forces:

$$m_1\ddot{\mathbf{x}}_1 = \frac{-Gm_1m_2}{r^2}\hat{\mathbf{r}}$$

$$m_2\ddot{\mathbf{x}}_2 = \frac{Gm_1m_2}{r^2}\hat{\mathbf{r}}$$

where $\mathbf{x}$ is the relative position vector of mass 1 with respect to mass 2, expressed as:

$$\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$$

and $\hat{\mathbf{r}}$ is the unit vector in that direction and r is the length of that vector.

Dividing by their respective masses and subtracting the second equation from the first yields the equation of motion for the acceleration of the first object with respect to the second:

$$\ddot{\mathbf{x}} = -\frac{\mu}{r^2}\hat{\mathbf{r}} \tag{3.3}$$

where μ is the parameter:

$$\mu = G(m_1 + m_2)$$

For movement under any central force, i.e. a force parallel to $\mathbf{r}$, the relative angular momentum

$$\mathbf{L} = \mathbf{r} \times \dot{\mathbf{r}}$$

stays constant:

$$\dot{\mathbf{L}} = \frac{d}{dt}(\mathbf{r} \times \dot{\mathbf{r}}) = \dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Since the cross product of the position vector and its velocity stays constant, they must lie in the same plane, orthogonal to $\mathbf{L}$. This implies the vector function is a plane curve.

With the versor $\hat{\mathbf{r}}$ we can write $\mathbf{r} = r\hat{\mathbf{r}}$ and with this notation equation 3.3 can be written

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^2}\hat{\mathbf{r}}.$$

It follows that

$$\mathbf{L} = \mathbf{r} \times \dot{\mathbf{r}} = r\hat{\mathbf{r}} \times \frac{d}{dt}(r\hat{\mathbf{r}}) = r\hat{\mathbf{r}} \times (r\dot{\hat{\mathbf{r}}} + \dot{r}\hat{\mathbf{r}}) = r^2(\hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}}) + r\dot{r}(\hat{\mathbf{r}} \times \hat{\mathbf{r}}) = r^2\hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}}$$

Now consider

$$\ddot{\mathbf{r}} \times \mathbf{L} = -\frac{\mu}{r^2}\hat{\mathbf{r}} \times (r^2\hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}}) = -\mu\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}}) = -\mu[(\hat{\mathbf{r}} \cdot \dot{\hat{\mathbf{r}}})\hat{\mathbf{r}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{r}})\dot{\hat{\mathbf{r}}}]$$

Since $\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = |\hat{\mathbf{r}}|^2 = 1$ we have that

$$\hat{\mathbf{r}} \cdot \dot{\hat{\mathbf{r}}} = \frac{1}{2}(\hat{\mathbf{r}} \cdot \dot{\hat{\mathbf{r}}} + \dot{\hat{\mathbf{r}}} \cdot \hat{\mathbf{r}}) = \frac{1}{2}\frac{d}{dt}(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}) = 0$$

Substituting these values into the previous equation, we have:

$$\ddot{\mathbf{r}} \times \mathbf{L} = \mu\dot{\hat{\mathbf{r}}}$$

Now, integrating both sides:

$$\dot{\mathbf{r}} \times \mathbf{L} = \mu \hat{\mathbf{r}} + \mathbf{c}$$

Where $\mathbf{c}$ is a constant vector. If we calculate the inner product of the previous equation this with $\mathbf{r}$ yields an interesting result:

$$\mathbf{r} \bullet (\dot{\mathbf{r}} \times \mathbf{L}) = \mathbf{r} \bullet (\mu \hat{\mathbf{r}} + \mathbf{c}) = \mu \mathbf{r} \bullet \hat{\mathbf{r}} + \mathbf{r} \bullet \mathbf{c} = \mu r (\hat{\mathbf{r}} \bullet \hat{\mathbf{r}}) + rc\cos(\theta) = r(\mu + c\cos(\theta))$$

Where θ is the angle between $\mathbf{r}$ and $\mathbf{c}$. Solving for r:

$$r = \frac{\mathbf{r} \bullet (\dot{\mathbf{r}} \times \mathbf{L})}{\mu + c\cos(\theta)} = \frac{(\mathbf{r} \times \dot{\mathbf{r}}) \bullet \mathbf{L}}{\mu + c\cos(\theta)} = \frac{|\mathbf{L}|^2}{\mu + c\cos(\theta)}$$

Finally, we note that

$$(r, \theta)$$

are effectively the polar coordinates of the vector function. Making the substitutions $p = \frac{|\mathbf{L}|^2}{\mu}$ and $e = \frac{c}{\mu}$, we arrive at the equation

$$r = \frac{p}{1 + e \cdot \cos\theta} \tag{3.4}$$

The Equation 3.4 is the equation in polar coordinates for a conic section with one focus at the origin.

3.3 Definition of the Derivative of Vector Function

Before we begin, let us introduce some necessary notation. Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}$ be a function. We write $\mathbf{f}(h) = \mathbf{o}(h)$ if $\mathbf{f}(h)$ goes faster to 0 than h, that is, if $\lim_{h\to 0} \frac{\mathbf{f}(h)}{h} = 0$. For example, $h^3 + 2h^2 = \mathbf{o}(h)$, since

$$\lim_{h\to 0} \frac{h^3 + 2h^2}{h} = \lim_{h\to 0} h^2 + 2h = 0.$$

We now define the derivative in the multidimensional space $\mathbb{R}^n$. Recall that in one variable, a function $\mathbf{g} : \mathbb{R} \to \mathbb{R}$ is said to be differentiable at $x = a$ if the limit

$$\lim_{x\to a} \frac{\mathbf{g}(x) - \mathbf{g}(a)}{x - a} = g'(a)$$

exists. The limit condition above is equivalent to saying that

$$\lim_{x\to a}\frac{\mathbf{g}(x)-\mathbf{g}(a)-g'(a)(x-a)}{x-a}=0,$$

or equivalently,

$$\lim_{h\to 0}\frac{\mathbf{g}(a+h)-\mathbf{g}(a)-g'(a)(h)}{h}=0.$$

We may write this as

$$\mathbf{g}(a+h)-\mathbf{g}(a)=g'(a)(h)+\mathbf{o}(h).$$

The above analysis provides an analogue definition for the higher-dimensional case. Observe that since we may not divide by vectors, the corresponding definition in higher dimensions involves quotients of norms.

Definition 3.3.1 Let $A\subseteq\mathbb{R}^n$. A function $\mathbf{f}:A\to\mathbb{R}^m$ is said to be **differentiable** at $\mathbf{a}\in A$ if there is a linear transformation, called the **derivative of f at a**, $D_{\mathbf{a}}(\mathbf{f}):\mathbb{R}^n\to\mathbb{R}^m$ such that

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{||\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{a})-D_{\mathbf{a}}(\mathbf{f})(\mathbf{x}-\mathbf{a})||}{||\mathbf{x}-\mathbf{a}||}=0.$$

Equivalently, $\mathbf{f}$ is differentiable at $\mathbf{a}$ if there is a linear transformation $D_{\mathbf{a}}(\mathbf{f})$ such that

$$\mathbf{f}(\mathbf{a}+\mathbf{h})-\mathbf{f}(\mathbf{a})=D_{\mathbf{a}}(\mathbf{f})(\mathbf{h})+\mathbf{o}\left(||\mathbf{h}||\right),$$

as $\mathbf{h}\to\mathbf{0}$.

The condition for differentiability at $\mathbf{a}$ is equivalent to

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{a})=D_{\mathbf{a}}(f)(\mathbf{x}-\mathbf{a})+\mathbf{o}\left(||\mathbf{x}-\mathbf{a}||\right),$$

as $\mathbf{x}\to\mathbf{a}$.

Theorem 3.3.1 If A is an open set in definition 3.3.1, $D_{\mathbf{a}}(f)$ is uniquely determined.

Proof: Let $L:\mathbb{R}^n\to\mathbb{R}^m$ be another linear transformation satisfying definition 3.3.1. We must prove that $\forall\mathbf{v}\in\mathbb{R}^n, L(\mathbf{v})=D_{\mathbf{a}}(f)(\mathbf{v})$. Since A is open, $\mathbf{a}+\mathbf{h}\in A$ for sufficiently small $||\mathbf{h}||$. By definition, as $\mathbf{h}\to\mathbf{0}$, we have

$$\mathbf{f}(\mathbf{a}+\mathbf{h})-\mathbf{f}(\mathbf{a})=D_{\mathbf{a}}(f)(\mathbf{h})+\mathbf{o}\left(||\mathbf{h}||\right).$$

and

$$\mathbf{f}(\mathbf{a}+\mathbf{h})-\mathbf{f}(\mathbf{a})=L(\mathbf{h})+\mathbf{o}\left(||\mathbf{h}||\right).$$

Now, observe that

$$D_{\mathbf{a}}(f)(\mathbf{v}) - L(\mathbf{v}) = D_{\mathbf{a}}(f)(\mathbf{h}) - \mathbf{f}(\mathbf{a}+\mathbf{h}) + \mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{h}).$$

By the triangle inequality,

$$\begin{aligned} \|D_{\mathbf{a}}(f)(\mathbf{v}) - L(\mathbf{v})\| &\leqslant \|D_{\mathbf{a}}(f)(\mathbf{h}) - \mathbf{f}(\mathbf{a}+\mathbf{h}) + \mathbf{f}(\mathbf{a})\| \\ &\quad + \|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{h})\| \\ &= \mathbf{o}\left(\|\mathbf{h}\|\right) + \mathbf{o}\left(\|\mathbf{h}\|\right) \\ &= \mathbf{o}\left(\|\mathbf{h}\|\right), \end{aligned}$$

as $\mathbf{h} \to \mathbf{0}$ This means that

$$\|L(\mathbf{v}) - D_{\mathbf{a}}(f)(\mathbf{v})\| \to 0,$$

i.e., $L(\mathbf{v}) = D_{\mathbf{a}}(f)(\mathbf{v})$, completing the proof. ■

If $A = \{\mathbf{a}\}$, a singleton, then $D_{\mathbf{a}}(f)$ is not uniquely determined. For $\|\mathbf{x}-\mathbf{a}\| < \delta$ holds only for $\mathbf{x} = \mathbf{a}$, and so $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$. Any linear transformation T will satisfy the definition, as $T(\mathbf{x}-\mathbf{a}) = T(\mathbf{0}) = \mathbf{0}$, and

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - T(\mathbf{x}-\mathbf{a})\| = \|\mathbf{0}\| = 0,$$

identically.

Example 3.3.2 If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $D_{\mathbf{a}}(L) = L$, for any $\mathbf{a} \in \mathbb{R}^n$.

Solution: ► Since $\mathbb{R}^n$ is an open set, we know that $D_{\mathbf{a}}(L)$ uniquely determined. Thus if L satisfies definition 3.3.1, then the claim is established. But by linearity

$$\|L(\mathbf{x}) - L(\mathbf{a}) - L(\mathbf{x}-\mathbf{a})\| = \|L(\mathbf{x}) - L(\mathbf{a}) - L(\mathbf{x}) + L(\mathbf{a})\| = \|0\| = 0,$$

whence the claim follows. ◄

Example 3.3.3 Let

$$f : \begin{array}{ccc} \mathbb{R}^3 \times \mathbb{R}^3 & \to & \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) & \mapsto & \mathbf{x}\bullet\mathbf{y} \end{array}$$

be the usual dot product in $\mathbb{R}^3$. Show that $\mathbf{f}$ is differentiable and that

$$D_{(\mathbf{x},\mathbf{y})}\mathbf{f}(\mathbf{h}, \mathbf{k}) = \mathbf{x}\bullet\mathbf{k} + \mathbf{h}\bullet\mathbf{y}.$$

Solution: ► We have

$$\begin{aligned}\mathbf{f}(\mathbf{x}+\mathbf{h},\mathbf{y}+\mathbf{k})-\mathbf{f}(\mathbf{x},\mathbf{y}) &= (\mathbf{x}+\mathbf{h})\bullet(\mathbf{y}+\mathbf{k})-\mathbf{x}\bullet\mathbf{y}\\ &= \mathbf{x}\bullet\mathbf{y}+\mathbf{x}\bullet\mathbf{k}+\mathbf{h}\bullet\mathbf{y}+\mathbf{h}\bullet\mathbf{k}-\mathbf{x}\bullet\mathbf{y}\\ &= \mathbf{x}\bullet\mathbf{k}+\mathbf{h}\bullet\mathbf{y}+\mathbf{h}\bullet\mathbf{k}.\end{aligned}$$

As $(\mathbf{h},\mathbf{k})\to(\mathbf{0},\mathbf{0})$, we have by the Cauchy-Buniakovskii-Schwarz inequality, $|\mathbf{h}\bullet\mathbf{k}|\leqslant\|\mathbf{h}\|\|\mathbf{k}\|=\mathbf{o}\left(\|\mathbf{h}\|\right)$, which proves the assertion. ◄

Just like in the one variable case, differentiability at a point, implies continuity at that point.

Theorem 3.3.4 Suppose $A\subseteq\mathbb{R}^n$ is open and $\mathbf{f}:A\to\mathbb{R}^n$ is differentiable on A. Then $\mathbf{f}$ is continuous on A.

Proof: Given $\mathbf{a}\in A$, we must shew that

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

Since $\mathbf{f}$ is differentiable at $\mathbf{a}$ we have

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{a})=D_{\mathbf{a}}(f)(\mathbf{x}-\mathbf{a})+\mathbf{o}\left(\|\mathbf{x}-\mathbf{a}\|\right),$$

and so

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{a})\to\mathbf{0},$$

as $\mathbf{x}\to\mathbf{a}$, proving the theorem. ■

Exercises

Problem 3.4 Let $L:\mathbb{R}^3\to\mathbb{R}^3$ be a linear transformation and

$$F:\begin{array}{ccc}\mathbb{R}^3 & \to & \mathbb{R}^3\\ \mathbf{x} & \mapsto & \mathbf{x}\times L(\mathbf{x})\end{array}.$$

Shew that F is differentiable and that

$$D_{\mathbf{x}}(F)(\mathbf{h})=\mathbf{x}\times L(\mathbf{h})+\mathbf{h}\times L(\mathbf{x}).$$

Problem 3.5 Let $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}, n\geqslant 1, \mathbf{f}(\mathbf{x})=\|\mathbf{x}\|$ be the usual norm in $\mathbb{R}^n$, with $\|\mathbf{x}\|^2=\mathbf{x}\bullet\mathbf{x}$. Prove that

$$D_{\mathbf{x}}(f)(\mathbf{v})=\frac{\mathbf{x}\bullet\mathbf{v}}{\|\mathbf{x}\|},$$

for $\mathbf{x}\neq\mathbf{0}$, but that $\mathbf{f}$ is not differentiable at $\mathbf{0}$.

3.4 Partial and Directional Derivatives

Definition 3.4.1 Let $A \subseteq \mathbb{R}^n$, $\mathbf{f} : A \to \mathbb{R}^m$, and put

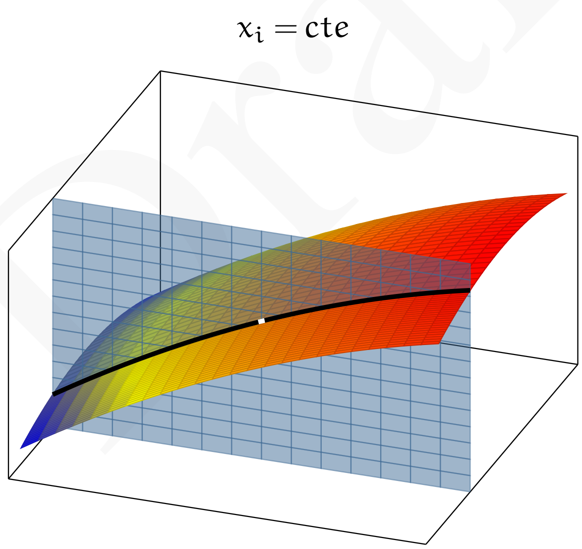
$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1,\ldots,x_n) \\ f_2(x_1,\ldots,x_n) \\ \vdots \\ f_m(x_1,\ldots,x_n) \end{bmatrix}.$$

Here $f_i : \mathbb{R}^n \to \mathbb{R}$. The **partial derivative** $\dfrac{\partial f_i}{\partial x_j}(\mathbf{x})$ is defined as

$$\partial_j f_i(x) := \frac{\partial f_i}{\partial x_j}(\mathbf{x}) := \lim_{h\to 0} \frac{f_i(x_1,,\ldots,x_j+h,\ldots,x_n) - f_i(x_1,\ldots,x_j,\ldots,x_n)}{h},$$

whenever this limit exists.

To find partial derivatives with respect to the j-th variable, we simply keep the other variables fixed and differentiate with respect to the j-th variable.



Example 3.4.1 If $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}$, and $\mathbf{f}(x,y,z) = x + y^2 + z^3 + 3xy^2z^3$ then

$$\frac{\partial f}{\partial x}(x,y,z) = 1 + 3y^2z^3,$$

$$\frac{\partial f}{\partial y}(x,y,z) = 2y + 6xyz^3,$$

and

$$\frac{\partial f}{\partial z}(x,y,z) = 3z^2 + 9xy^2z^2.$$

Let $\mathbf{f}(\mathbf{x})$ be a vector valued function. Then the derivative of $\mathbf{f}(\mathbf{x})$ in the direction $\mathbf{u}$ is defined as

$$D_{\mathbf{u}}\mathbf{f}(\mathbf{x}) := D\mathbf{f}(\mathbf{x})[\mathbf{u}] = \left[\frac{d}{d\alpha}\,\mathbf{f}(\mathbf{v} + \alpha\,\mathbf{u})\right]_{\alpha=0}$$

for all vectors $\mathbf{u}$.

Proposition 3.4.2

❶ If $\mathbf{f}(\mathbf{x}) = \mathbf{f}_1(\mathbf{x}) + \mathbf{f}_2(\mathbf{x})$ then $D_{\mathbf{u}}\mathbf{f}(\mathbf{x}) = D_{\mathbf{u}}\mathbf{f}_1(\mathbf{x}) + D_{\mathbf{u}}\mathbf{f}_2(\mathbf{x})$

❷ If $\mathbf{f}(\mathbf{x}) = \mathbf{f}_1(\mathbf{x}) \times \mathbf{f}_2(\mathbf{x})$ then $D_{\mathbf{u}}\mathbf{f}(\mathbf{x}) = \left(D_{\mathbf{u}}\mathbf{f}_1(\mathbf{x})\right) \times \mathbf{f}_2(\mathbf{x}) + \mathbf{f}_1(\mathbf{v}) \times \left(D_{\mathbf{u}}\mathbf{f}_2(\mathbf{x})\right)$

3.5 The Jacobi Matrix

We now establish a way which simplifies the process of finding the derivative of a function at a given point.

Since the derivative of a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, it can be represented by aid of matrices. The following theorem will allow us to determine the matrix representation for $D_{\mathbf{a}}(f)$ under the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Theorem 3.5.1 Let

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}.$$

Suppose $A \subseteq \mathbb{R}^n$ is an open set and $\mathbf{f} : A \to \mathbb{R}^m$ is differentiable. Then each partial derivative $\dfrac{\partial f_i}{\partial x_j}(\mathbf{x})$ exists, and the matrix representation of $D_{\mathbf{x}}(f)$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the **Jacobi matrix**

$$f'(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}) & \dfrac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \dfrac{\partial f_2}{\partial x_1}(\mathbf{x}) & \dfrac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \dfrac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathbf{x}) & \dfrac{\partial f_m}{\partial x_2}(\mathbf{x}) & \cdots & \dfrac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

Proof: Let $\mathbf{e}_j, 1 \leqslant j \leqslant n$, be the standard basis for $\mathbb{R}^n$. To obtain the Jacobi matrix, we must compute $D_{\mathbf{x}}(f)(\mathbf{e}_j)$, which will give us the j-th column of the Jacobi matrix. Let $f'(\mathbf{x}) = (J_{ij})$, and observe that

$$D_{\mathbf{x}}(f)(\mathbf{e}_j) = \begin{bmatrix} J_{1j} \\ J_{2j} \\ \vdots \\ J_{mj} \end{bmatrix}.$$

and put $\mathbf{y} = \mathbf{x} + \varepsilon \mathbf{e}_j$, $\varepsilon \in \mathbb{R}$. Notice that

$$\begin{aligned} & \frac{||\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x}) - D_{\mathbf{x}}(f)(\mathbf{y}-\mathbf{x})||}{||\mathbf{y}-\mathbf{x}||} \\ & = \frac{||\mathbf{f}(x_1, \ldots, x_j + h, \ldots, x_n) - \mathbf{f}(x_1, \ldots, x_j, \ldots, x_n) - \varepsilon D_{\mathbf{x}}(f)(\mathbf{e}_j)||}{|\varepsilon|}. \end{aligned}$$

Since the sinistral side $\to 0$ as $\varepsilon \to 0$, the so does the i-th component of the numerator, and so,

$$\frac{|f_i(x_1, \ldots, x_j + h, \ldots, x_n) - f_i(x_1, \ldots, x_j, \ldots, x_n) - \varepsilon J_{ij}|}{|\varepsilon|} \to 0.$$

This entails that

$$J_{ij} = \lim_{\varepsilon \to 0} \frac{f_i(x_1, \ldots, x_j + \varepsilon, \ldots, x_n) - f_i(x_1, \ldots, x_j, \ldots, x_n)}{\varepsilon} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}).$$

This finishes the proof. ■

Strictly speaking, the Jacobi matrix is not the derivative of a function at a point. It is a matrix representation of the derivative in the standard basis of $\mathbb{R}^n$. We will abuse language, however, and refer to f' when we mean the Jacobi matrix of $\mathbf{f}$.

Example 3.5.2 Let $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$\mathbf{f}(x, y) = (xy + yz, \log_e xy).$$

Compute the Jacobi matrix of $\mathbf{f}$.

Solution: ► The Jacobi matrix is the 2×3 matrix

$$f'(x, y) = \begin{bmatrix} \partial_x f_1(x, y) & \partial_y f_1(x, y) & \partial_z f_1(x, y) \\ \partial_x f_2(x, y) & \partial_y f_2(x, y) & \partial_z f_2(x, y) \end{bmatrix} = \begin{bmatrix} y & x+z & y \\ \frac{1}{x} & \frac{1}{y} & 0 \end{bmatrix}.$$

◄

Example 3.5.3 Let $\mathbf{f}(\rho, \theta, z) = (\rho \cos \theta, \rho \sin \theta, z)$ be the function which changes from cylindrical coordinates to Cartesian coordinates. We have

$$f'(\rho, \theta, z) = \begin{bmatrix} \cos\theta & -\rho\sin\theta & 0 \\ \sin\theta & \rho\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Example 3.5.4 Let $\mathbf{f}(\rho,\phi,\theta) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$ be the function which changes from spherical coordinates to Cartesian coordinates. We have

$$f'(\rho,\phi,\theta) = \begin{bmatrix} \cos\theta\sin\phi & \rho\cos\theta\cos\phi & -\rho\sin\phi\sin\theta \\ \sin\theta\sin\phi & \rho\sin\theta\cos\phi & \rho\cos\theta\sin\phi \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix}.$$

The concept of **repeated partial derivatives** is akin to the concept of repeated differentiation. Similarly with the concept of implicit partial differentiation. The following examples should be self-explanatory.

Example 3.5.5 Let $\mathbf{f}(u,v,w) = e^u v\cos w$. Determine $\dfrac{\partial^2}{\partial u \partial v}\mathbf{f}(u,v,w)$ at $(1,-1,\dfrac{\pi}{4})$.

Solution: ► We have

$$\frac{\partial^2}{\partial u \partial v}(e^u v\cos w) = \frac{\partial}{\partial u}(e^u\cos w) = e^u\cos w,$$

which is $\dfrac{e\sqrt{2}}{2}$ at the desired point. ◄

Example 3.5.6 The equation $z^{xy} + (xy)^z + xy^2z^3 = 3$ defines z as an implicit function of x and y. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ at $(1,1,1)$.

Solution: ► We have

$$\begin{aligned} \frac{\partial}{\partial x}z^{xy} &= \frac{\partial}{\partial x}e^{xy\log z} \\ &= \left(y\log z + \frac{xy}{z}\frac{\partial z}{\partial x}\right)z^{xy}, \\ \frac{\partial}{\partial x}(xy)^z &= \frac{\partial}{\partial x}e^{z\log xy} \\ &= \left(\frac{\partial z}{\partial x}\log xy + \frac{z}{x}\right)(xy)^z, \\ \frac{\partial}{\partial x}xy^2z^3 &= y^2z^3 + 3xy^2z^2\frac{\partial z}{\partial x}, \end{aligned}$$

Hence, at $(1,1,1)$ we have

$$\frac{\partial z}{\partial x} + 1 + 1 + 3\frac{\partial z}{\partial x} = 0 \implies \frac{\partial z}{\partial x} = -\frac{1}{2}.$$

Similarly,

$$\begin{aligned} \frac{\partial}{\partial y}z^{xy} &= \frac{\partial}{\partial y}e^{xy\log z} \\ &= \left(x\log z + \frac{xy}{z}\frac{\partial z}{\partial y}\right)z^{xy}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial y}(xy)^z &= \frac{\partial}{\partial y}e^{z\log xy} \\
&= \left(\frac{\partial z}{\partial y}\log xy + \frac{z}{y}\right)(xy)^z, \\
\frac{\partial}{\partial y}xy^2z^3 &= 2xyz^3 + 3xy^2z^2\frac{\partial z}{\partial y},
\end{aligned}$$

Hence, at $(1,1,1)$ we have

$$\frac{\partial z}{\partial y} + 1 + 2 + 3\frac{\partial z}{\partial y} = 0 \implies \frac{\partial z}{\partial y} = -\frac{3}{4}.$$

◀

Exercises

Problem 3.6 Let $\mathbf{f} : [0;+\infty[\times]0;+\infty[\to \mathbb{R}$, $\mathbf{f}(r,t) = t^n e^{-r^2/4t}$, where n is a constant. Determine n such that

$$\frac{\partial f}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right).$$

Problem 3.7 Let

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}, \qquad \mathbf{f}(x,y) = \min(x,y^2).$$

Find $\dfrac{\partial \mathbf{f}(x,y)}{\partial x}$ and $\dfrac{\partial \mathbf{f}(x,y)}{\partial y}$.

Problem 3.8 Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$\mathbf{f}(x,y) = \left(xy^2x^2y\right), \qquad \mathbf{g}(x,y,z) = (x-y+2zxy).$$

Compute $(f\circ g)'(1,0,1)$, if at all defined. If undefined, explain. Compute $(g\circ f)'(1,0)$, if at all defined. If undefined, explain.

Problem 3.9 Let $\mathbf{f}(x,y) = (xyx+y)$ and $\mathbf{g}(x,y) = \left(x - yx^2y^2x + y\right)$ Find $(g\circ f)'(0,1)$.

Problem 3.10 Prove that

$$\int_0^{+\infty}\frac{\arctan ax - \arctan x}{x}dx = \frac{\pi}{2}\log \pi.$$

Problem 3.11 Assuming that the equation $xy^2 + 3z = \cos z^2$ defines z implicitly as a function of x and y, find $\partial_x z$.

Problem 3.12 If $w = e^{uv}$ and $u = r+s$, $v = rs$, determine $\dfrac{\partial w}{\partial r}$.

Problem 3.13 Let z be an implicitly-defined function of x and y through the equation $(x+z)^2 + (y+z)^2 = 8$. Find $\dfrac{\partial z}{\partial x}$ at $(1,1,1)$.

3.6 Properties of Differentiable Transformations

Just like in the one-variable case, we have the following rules of differentiation.

Theorem 3.6.1 Let $A\subseteq\mathbb{R}^n, B\subseteq\mathbb{R}^m$ be open sets $f,\mathbf{g} : A\to\mathbb{R}^m, \alpha\in\mathbb{R}$, be differentiable on A, $h : B\to\mathbb{R}^l$ be differentiable on B, and $\mathbf{f}(A)\subseteq B$. Then we have

- Addition Rule: $D_{\mathbf{x}}((f+\alpha g)) = D_{\mathbf{x}}(f) + \alpha D_{\mathbf{x}}(g)$.
- Chain Rule: $D_{\mathbf{x}}((h\circ f)) = \left(D_{\mathbf{f}(x)}(h)\right)\circ\left(D_{\mathbf{x}}(f)\right)$.

Since composition of linear mappings expressed as matrices is matrix multiplication, the Chain Rule takes the alternative form when applied to the Jacobi matrix.

$$(h \circ f)' = (h' \circ f)(f'). \tag{3.5}$$

Example 3.6.2 Let

$$\mathbf{f}(u,v) = (ue^{v}u+vuv),$$
$$h(x,y) = \left(x^2+yy+z\right).$$

Find $(f \circ h)'(x,y)$.

Solution: ► We have

$$f'(u,v) = \begin{bmatrix} e^{v} & ue^{v} \\ 1 & 1 \\ v & u \end{bmatrix},$$

and

$$h'(x,y) = \begin{bmatrix} 2x & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Observe also that

$$f'(h(x,y)) = \begin{bmatrix} e^{y+z} & (x^2+y)e^{y+z} \\ 1 & 1 \\ y+z & x^2+y \end{bmatrix}.$$

Hence

$$\begin{aligned} (f \circ h)'(x,y) &= f'(h(x,y))h'(x,y) \\ &= \begin{bmatrix} e^{y+z} & (x^2+y)e^{y+z} \\ 1 & 1 \\ y+z & x^2+y \end{bmatrix} \begin{bmatrix} 2x & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2xe^{y+z} & (1+x^2+y)e^{y+z} & (x^2+y)e^{y+z} \\ 2x & 2 & 1 \\ 2xy+2xz & x^2+2y+z & x^2+y \end{bmatrix}. \end{aligned}$$

◄

Example 3.6.3 Let

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}, \quad \mathbf{f}(u,v) = u^2 + e^v,$$

$$u, v : \mathbb{R}^3 \to \mathbb{R} \quad u(x,y) = xz, \; v(x,y) = y + z.$$

Put $h(x,y) = f\left(u(x,y,z)v(x,y,z)\right)$. Find the partial derivatives of h.

Solution: ► Put $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2, \mathbf{g}(x,y) = \left(u(x,y)v(x,y)\right) = (xzy + z)$. Observe that $h = f \circ g$. Now,

$$g'(x,y) = \begin{bmatrix} z & 0 & x \\ 0 & 1 & 1 \end{bmatrix},$$

$$f'(u,v) = \begin{bmatrix} 2u & e^v \end{bmatrix},$$

$$f'(h(x,y)) = \begin{bmatrix} 2xz & e^{y+z} \end{bmatrix}.$$

Thus

$$\begin{aligned} \begin{bmatrix} \frac{\partial h}{\partial x}(x,y) & \frac{\partial h}{\partial y}(x,y) & \frac{\partial h}{\partial z}(x,y) \end{bmatrix} &= h'(x,y) \\ &= (f'(\mathbf{g}(x,y)))(g'(x,y)) \\ &= \begin{bmatrix} 2xz & e^{y+z} \end{bmatrix} \begin{bmatrix} z & 0 & x \\ 0 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2xz^2 & e^{y+z} & 2x^2z + e^{y+z} \end{bmatrix} \end{aligned}.$$

Equating components, we obtain

$$\frac{\partial h}{\partial x}(x,y) = 2xz^2,$$

$$\frac{\partial h}{\partial y}(x,y) = e^{y+z},$$

$$\frac{\partial h}{\partial z}(x,y) = 2x^2z + e^{y+z}.$$

◄

Theorem 3.6.4 Let $\mathbf{F} = (f_1, f_2, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$, and suppose that the partial derivatives

$$\frac{\partial f_i}{\partial x_j}, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n, \tag{3.6}$$

exist on a neighborhood of $\mathbf{x}_0$ and are continuous at $\mathbf{x}_0$. Then $\mathbf{F}$ is differentiable at $\mathbf{x}_0$.

We say that $\mathbf{F}$ is **continuously differentiable** on a set S if S is contained in an open set on which the partial derivatives in (3.6) are continuous. The next three lemmas give properties of continuously differentiable transformations that we will need later.

Lemma 3.6.5 Suppose that $\mathbf{F}:\mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable on a neighborhood N of $\mathbf{x}_0$. Then, for every $\epsilon > 0$, there is a $\delta > 0$ such that

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| < (\|\mathbf{F}'(\mathbf{x}_0)\| + \epsilon)|\mathbf{x}-\mathbf{y}| \quad \text{if} \quad \mathbf{A}, \mathbf{y} \in B_\delta(\mathbf{x}_0). \tag{3.7}$$

Proof: Consider the auxiliary function

$$\mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}'(\mathbf{x}_0)\mathbf{x}. \tag{3.8}$$

The components of $\mathbf{G}$ are

$$g_i(\mathbf{x}) = f_i(\mathbf{x}) - \sum_{j=1}^{n} \frac{\partial f_i(\mathbf{x}_0)\partial x_j}{x}{}_j,$$

so

$$\frac{\partial g_i(\mathbf{x})}{\partial x_j} = \frac{\partial f_i(\mathbf{x})}{\partial x_j} - \frac{\partial f_i(\mathbf{x}_0)}{\partial x_j}.$$

Thus, $\partial g_i/\partial x_j$ is continuous on N and zero at $\mathbf{x}_0$. Therefore, there is a $\delta > 0$ such that

$$\left|\frac{\partial g_i(\mathbf{x})}{\partial x_j}\right| < \frac{\epsilon}{\sqrt{mn}} \quad \text{for} \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n, \quad \text{if} \quad |\mathbf{x}-\mathbf{x}_0| < \delta. \tag{3.9}$$

Now suppose that $\mathbf{x}$, $\mathbf{y} \in B_\delta(\mathbf{x}_0)$. By Theorem **??**,

$$g_i(\mathbf{x}) - g_i(\mathbf{y}) = \sum_{j=1}^{n} \frac{\partial g_i(\mathbf{x}_i)}{\partial x_j}(x_j - y_j), \tag{3.10}$$

where $\mathbf{x}_i$ is on the line segment from $\mathbf{x}$ to $\mathbf{y}$, so $\mathbf{x}_i \in B_\delta(\mathbf{x}_0)$. From (3.9), (3.10), and Schwarz's inequality,

$$(g_i(\mathbf{x}) - g_i(\mathbf{y}))^2 \leqslant \left(\sum_{j=1}^{n}\left[\frac{\partial g_i(\mathbf{x}_i)}{\partial x_j}\right]^2\right)|\mathbf{x}-\mathbf{y}|^2 < \frac{\epsilon^2}{m}|\mathbf{x}-\mathbf{y}|^2.$$

Summing this from $i=1$ to $i=m$ and taking square roots yields

$$|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y})| < \epsilon|\mathbf{x}-\mathbf{y}| \quad \text{if} \quad \mathbf{x}, \mathbf{y} \in B_\delta(\mathbf{x}_0). \tag{3.11}$$

To complete the proof, we note that

$$\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) = \mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y}) + \mathbf{F}'(\mathbf{x}_0)(\mathbf{x}-\mathbf{y}), \tag{3.12}$$

so (3.11) and the triangle inequality imply (3.7). ■

Lemma 3.6.6 Suppose that $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable on a neighborhood of $\mathbf{x}_0$ and $\mathbf{F}'(\mathbf{x}_0)$ is nonsingular. Let

$$r = \frac{1}{\|(\mathbf{F}'(\mathbf{x}_0))^{-1}\|}. \tag{3.13}$$

Then, for every $\epsilon > 0$, there is a $\delta > 0$ such that

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \geqslant (r - \epsilon)|\mathbf{x} - \mathbf{y}| \quad \text{if} \quad \mathbf{x}, \mathbf{y} \in B_\delta(\mathbf{x}_0). \tag{3.14}$$

Proof: Let $\mathbf{x}$ and $\mathbf{y}$ be arbitrary points in $D_\mathbf{F}$ and let $\mathbf{G}$ be as in (3.8). From (3.12),

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \geqslant \big||\mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{y})| - |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y})|\big|, \tag{3.15}$$

Since

$$\mathbf{x} - \mathbf{y} = [\mathbf{F}'(\mathbf{x}_0)]^{-1}\mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{y}),$$

(3.13) implies that

$$|\mathbf{x} - \mathbf{y}| \leqslant \frac{1}{r}|\mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{y}|,$$

so

$$|\mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{y})| \geqslant r|\mathbf{x} - \mathbf{y}|. \tag{3.16}$$

Now choose $\delta > 0$ so that (3.11) holds. Then (3.15) and (3.16) imply (3.14). ■

Definition 3.6.1 A function $\mathbf{f}$ is said to be **continuously differentiable** if the derivative $\mathbf{f}'$ exists and is itself a continuous function.

Continuously differentiable functions are said to be of **class** C^1. A function is of **class** C^2 if the first and second derivative of the function both exist and are continuous. More generally, a function is said to be of **class** C^k if the first k derivatives exist and are continuous. If derivatives $\mathbf{f}^{(n)}$ exist for all positive integers n, the function is said **smooth** or equivalently, of **class** C^∞.

3.7 Gradients, Curls and Directional Derivatives

Definition 3.7.1 Let

$$f : \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ \mathbf{x} & \mapsto & f(\mathbf{x}) \end{array}$$

be a scalar field. The **gradient** of f is the vector defined and denoted by

$$\nabla f(\mathbf{x}) := Df(\mathbf{x}) := \left(\partial_1 f(\mathbf{x}), \partial_2 f(\mathbf{x}), \ldots, \partial_n f(\mathbf{x})\right).$$

The **gradient operator** is the operator

$$\nabla = (\partial_1, \partial_2, \ldots, \partial_n).$$

Theorem 3.7.1 Let $A \subseteq \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}$ be a scalar field, and assume that f is differentiable in A. Let $K \in \mathbb{R}$ be a constant. Then $\nabla f(\mathbf{x})$ is orthogonal to the surface implicitly defined by $f(\mathbf{x}) = K$.

Proof: Let

$$\mathbf{c} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R}^n \\ t & \mapsto & \mathbf{c}(t) \end{array}$$

be a curve lying on this surface. Choose t_0 so that $\mathbf{c}(t_0) = \mathbf{x}$. Then

$$(f \circ \mathbf{c})(t_0) = f(\mathbf{c}(t)) = K,$$

and using the chain rule

$$Df(c(t_0))Dc(t_0) = 0,$$

which translates to

$$(\nabla f(\mathbf{x})) \bullet (\mathbf{c}'(t_0)) = 0.$$

Since $\mathbf{c}'(t_0)$ is tangent to the surface and its dot product with $\nabla f(\mathbf{x})$ is 0, we conclude that $\nabla f(\mathbf{x})$ is normal to the surface. ■

R Now let $c(t)$ be a curve in $\mathbb{R}^n$ (not necessarily in the surface).
And let θ be the angle between $\nabla f(\mathbf{x})$ and $\mathbf{c}'(t_0)$. Since

$$|(\nabla f(\mathbf{x})) \bullet (\mathbf{c}'(t_0))| = ||\nabla f(\mathbf{x})|| ||\mathbf{c}'(t_0)|| \cos\theta,$$

$\nabla f(\mathbf{x})$ is the direction in which f is changing the fastest.

Example 3.7.2 Find a unit vector normal to the surface $x^3+y^3+z=4$ at the point $(1,1,2)$.

Solution: ▶ Here $f(x,y,z)=x^3+y^3+z-4$ has gradient

$$\nabla f(x,y,z)=\left(3x^2,3y^2,1\right)$$

which at $(1,1,2)$ is $(3,3,1)$. Normalizing this vector we obtain

$$\left(\frac{3}{\sqrt{19}},\frac{3}{\sqrt{19}},\frac{1}{\sqrt{19}}\right).$$

◀

Example 3.7.3 Find the direction of the greatest rate of increase of $f(x,y,z)=xye^z$ at the point $(2,1,2)$.

Solution: ▶ The direction is that of the gradient vector. Here

$$\nabla f(x,y,z)=(ye^z,xe^z,xye^z)$$

which at $(2,1,2)$ becomes $\left(e^2,2e^2,2e^2\right)$. Normalizing this vector we obtain

$$\frac{1}{\sqrt{5}}(1,2,2).$$

◀

Example 3.7.4 Sketch the gradient vector field for $f(x,y)=x^2+y^2$ as well as several contours for this function.

Solution: ▶ The contours for a function are the curves defined by,

$$f(x,y)=k$$

for various values of k. So, for our function the contours are defined by the equation,

$$x^2+y^2=k$$

and so they are circles centered at the origin with radius $\sqrt{k}$. The gradient vector field for this function is

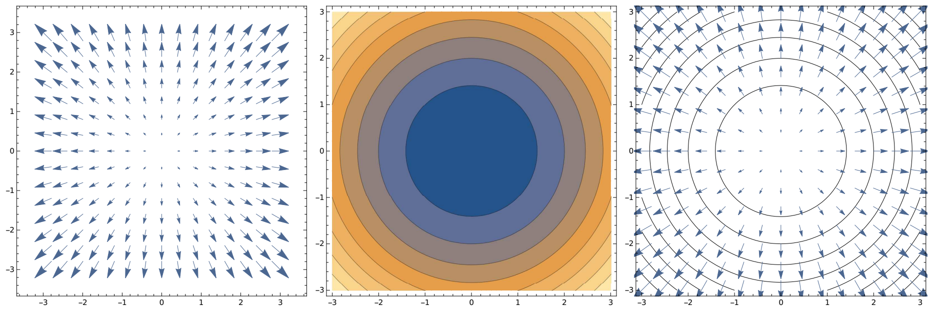
$$\nabla f(x,y)=2x\mathbf{i}+2y\mathbf{j}$$

Here is a sketch of several of the contours as well as the gradient vector field. ◀

Example 3.7.5 Let $f:\mathbb{R}^3\to\mathbb{R}$ be given by

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

Find the equation of the tangent plane to f at $(1,2,3)$.



Solution: ► A vector normal to the plane is $\nabla f(1,2,3)$. Now

$$\nabla f(x,y,z) = (1, 2y, -2z)$$

which is

$$(1,4,-6)$$

at $(1,2,3)$. The equation of the tangent plane is thus

$$1(x-1)+4(y-2)-6(z-3) = 0,$$

or

$$x+4y-6z = -9.$$

◄

Definition 3.7.2 Let

$$f : \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \mathbf{x} & \mapsto & \mathbf{f}(\mathbf{x}) \end{array}$$

be a vector field with

$$\mathbf{f}(\mathbf{x}) = \left(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})\right).$$

The **divergence** of **f** is defined and denoted by

$$\operatorname{div} \mathbf{f}(\mathbf{x}) = \nabla \bullet \mathbf{f}(\mathbf{x}) := \operatorname{Tr}\left(D\mathbf{f}(\mathbf{x})\right) := \partial_1 f_1(\mathbf{x}) + \partial_2 f_2(\mathbf{x}) + \cdots + \partial_n f_n(\mathbf{x}).$$

Example 3.7.6 If $\mathbf{f}(x,y,z) = (x^2, y^2, ye^{z^2})$ then

$$\operatorname{div} \mathbf{f}(\mathbf{x}) = 2x + 2y + 2yze^{z^2}.$$

Mean Value Theorem for Scalar Fields

The mean value theorem generalizes to scalar fields. The trick is to use parametrization to create a real function of one variable, and then apply the one-variable theorem.

Theorem 3.7.7 — Mean Value Theorem for Scalar Fields. Let U be an open connected subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}$ be a differentiable function. Fix points $\mathbf{x}, \mathbf{y} \in U$ such that the segment connecting $\mathbf{x}$ to $\mathbf{y}$ lies in U. Then

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla f(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{x})$$

where z is a point in the open segment connecting x to y

Proof: Let U be an open connected subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}$ be a differentiable function. Fix points $\mathbf{x}, \mathbf{y} \in U$ such that the segment connecting $\mathbf{x}$ to $\mathbf{y}$ lies in U , and define $g(t) := f\Big((1-t)\mathbf{x} + t\mathbf{y}\Big)$. Since f is a differentiable function in U the function g is continuous function in $[0,1]$ and differentiable in $(0,1)$. The mean value theorem gives:

$$g(1) - g(0) = g'(c)$$

for some $c \in (0,1)$. But since $g(0) = f(\mathbf{x})$ and $g(1) = f(\mathbf{y})$, computing $g'(c)$ explicitly we have:

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla f\Big((1-c)\mathbf{x} + c\mathbf{y}\Big) \cdot (\mathbf{y} - \mathbf{x})$$

or

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla f(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{x})$$

where $\mathbf{z}$ is a point in the open segment connecting $\mathbf{x}$ to $\mathbf{y}$ ■

By the Cauchy-Schwarz inequality, the equation gives the estimate:

$$\Bigg| f(\mathbf{y}) - f(\mathbf{x}) \Bigg| \leqslant \Bigg| \nabla f\Big((1-c)\mathbf{x} + c\mathbf{y}\Big) \Bigg| \, \Big| \mathbf{y} - \mathbf{x} \Big|.$$

Curl

Definition 3.7.3 If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field with components $\mathbf{F} = (F_1, F_2, F_3)$, we define the **curl of F**

$$\nabla \times \mathbf{F} \stackrel{\text{def}}{=} \begin{bmatrix} \partial_2 F_3 - \partial_3 F_2 \\ \partial_3 F_1 - \partial_1 F_3 \\ \partial_1 F_2 - \partial_2 F_1 \end{bmatrix}.$$

This is sometimes also denoted by $\operatorname{curl}(\mathbf{F})$.

R A mnemonic to remember this formula is to write

$$\nabla \times \mathbf{F} = \begin{bmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{bmatrix} \times \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix},$$

and compute the cross product treating both terms as 3-dimensional vectors.

Example 3.7.8 If $\mathbf{F}(x) = x/|x|^3$, then $\nabla \times \mathbf{F} = 0$.

R In the example above, $\mathbf{F}$ is proportional to a gravitational force exerted by a body at the origin. We know from experience that when a ball is pulled towards the earth by gravity alone, it doesn't start to rotate; which is consistent with our computation $\nabla \times \mathbf{F} = 0$.

Example 3.7.9 If $v(x,y,z) = (\sin z, 0, 0)$, then $\nabla \times v = (0, \cos z, 0)$.

R Think of v above as the velocity field of a fluid between two plates placed at $z = 0$ and $z = \pi$. A small ball placed closer to the bottom plate experiences a higher velocity near the top than it does at the bottom, and so should start rotating counter clockwise along the y-axis. This is consistent with our calculation of $\nabla \times v$.

The definition of the curl operator can be generalized to the n dimensional space.

Definition 3.7.4 Let $g_k : \mathbb{R}^n \to \mathbb{R}^n$, $1 \leqslant k \leqslant n-2$ be vector fields with $g_i = (g_{i1}, g_{i2}, \ldots, g_{in})$. Then the **curl** of $(g_1, g_2, \ldots, g_{n-2})$

$$\operatorname{curl}(g_1, g_2, \ldots, g_{n-2})(\mathbf{x}) = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ \partial_1 & \partial_2 & \ldots & \partial_n \\ g_{11}(\mathbf{x}) & g_{12}(\mathbf{x}) & \ldots & g_{1n}(\mathbf{x}) \\ g_{21}(\mathbf{x}) & g_{22}(\mathbf{x}) & \ldots & g_{2n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ g_{(n-2)1}(\mathbf{x}) & g_{(n-2)2}(\mathbf{x}) & \ldots & g_{(n-2)n}(\mathbf{x}) \end{bmatrix}.$$

Example 3.7.10 If $\mathbf{f}(x,y,z,w) = (e^{xyz}, 0, 0, w^2), g(x,y,z,w) = (0,0,z,0)$ then

$$\operatorname{curl}(f,g)(x,y,z,w) = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \\ \partial_1 & \partial_2 & \partial_3 & \partial_4 \\ e^{xyz} & 0 & 0 & w^2 \\ 0 & 0 & z & 0 \end{bmatrix} = (xz^2 e^{xyz})\mathbf{e}_4.$$

Definition 3.7.5 Let $A \subseteq \mathbb{R}^n$ be open and let $\mathbf{f} : A \to \mathbb{R}$ be a scalar field, and assume that $\mathbf{f}$ is differentiable in A. Let $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ be such that $\mathbf{x} + t\mathbf{v} \in A$ for sufficiently small $t \in \mathbb{R}$. Then the **directional derivative of f in the direction of v at the point x** is defined and denoted by

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) = \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{v}) - \mathbf{f}(\mathbf{x})}{t}.$$

Some authors require that the vector $\mathbf{v}$ in definition 3.7.5 be a unit vector.

Theorem 3.7.11 Let $A \subseteq \mathbb{R}^n$ be open and let $\mathbf{f} : A \to \mathbb{R}$ be a scalar field, and assume that $\mathbf{f}$ is differentiable in A. Let $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ be such that $\mathbf{x} + t\mathbf{v} \in A$ for sufficiently small $t \in \mathbb{R}$. Then the **directional derivative of f in the direction of v at the point x** is given by

$$\nabla f(\mathbf{x}) \bullet \mathbf{v}.$$

Example 3.7.12 Find the directional derivative of $\mathbf{f}(x, y, z) = x^3 + y^3 - z^2$ in the direction of $(1, 2, 3)$.

Solution: ► We have

$$\nabla f(x, y, z) = \left(3x^2, 3y^2, -2z\right)$$

and so

$$\nabla f(x, y, z) \bullet \mathbf{v} = 3x^2 + 6y^2 - 6z.$$

◄

The following is a collection of useful differentiation formulae in $\mathbb{R}^3$.

Theorem 3.7.13

❶ $\nabla \cdot \psi \mathbf{u} = \psi \nabla \cdot \mathbf{u} + \mathbf{u} \cdot \nabla \psi$

❷ $\nabla \times \psi \mathbf{u} = \psi \nabla \times \mathbf{u} + \nabla \psi \times \mathbf{u}$

❸ $\nabla \cdot \mathbf{u} \times \mathbf{v} = \mathbf{v} \cdot \nabla \times \mathbf{u} - \mathbf{u} \cdot \nabla \times \mathbf{v}$

❹ $\nabla \times (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \nabla \mathbf{u} - \mathbf{u} \cdot \nabla \mathbf{v} + \mathbf{u}(\nabla \cdot \mathbf{v}) - \mathbf{v}(\nabla \cdot \mathbf{u})$

❺ $\nabla(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot \nabla \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{u} + \mathbf{u} \times (\nabla \times \mathbf{v}) + \mathbf{v} \times (\nabla \times \mathbf{u})$

❻ $\nabla \times (\nabla \psi) = \mathrm{curl}\ \ (\mathrm{grad}\ \ \psi) = \mathbf{0}$

❼ $\nabla \cdot (\nabla \times \mathbf{u}) = \mathrm{div}\ \ (\mathrm{curl}\ \ \mathbf{u}) = 0$

❽ $\nabla \cdot (\nabla \psi_1 \times \nabla \psi_2) = 0$

❾ $\nabla \times (\nabla \times \mathbf{u}) = \mathrm{curl}\ \ (\mathrm{curl}\ \ \mathbf{u}) = \mathrm{grad}\ \ (\mathrm{div}\ \ \mathbf{u}) - \nabla^2 \mathbf{u}$

where

$$\Delta f = \nabla^2 f = \nabla \cdot \nabla f = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is the Laplacian operator and

$$\nabla^2 \mathbf{u} = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) =$$

$$(\frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2})\mathbf{i} + (\frac{\partial^2 u_y}{\partial x^2} + \frac{\partial^2 u_y}{\partial y^2} + \frac{\partial^2 u_y}{\partial z^2})\mathbf{j} + (\frac{\partial^2 u_z}{\partial x^2} + \frac{\partial^2 u_z}{\partial y^2} + \frac{\partial^2 u_z}{\partial z^2})\mathbf{k}$$

Finally, for the position vector $\mathbf{r}$ the following are valid

❶ $\nabla \cdot \mathbf{r} = 3$

❷ $\nabla \times \mathbf{r} = \mathbf{0}$

❸ $\mathbf{u} \cdot \nabla \mathbf{r} = \mathbf{u}$

where $\mathbf{u}$ is any vector.

Exercises

Problem 3.14 The temperature at a point in space is $T = xy + yz + zx$.

a) Find the direction in which the temperature changes most rapidly with distance from $(1,1,1)$. What is the maximum rate of change?

b) Find the derivative of T in the direction of the vector $3\mathbf{i} - 4\mathbf{k}$ at $(1,1,1)$.

Problem 3.15 For each of the following vector functions $\mathbf{F}$, determine whether $\nabla\phi = \mathbf{F}$ has a solution and determine it if it exists.

a) $\mathbf{F} = 2xyz^3\mathbf{i} - (x^2z^3 + 2y)\mathbf{j} + 3x^2yz^2\mathbf{k}$

b) $\mathbf{F} = 2xy\mathbf{i} + (x^2 + 2yz)\mathbf{j} + (y^2 + 1)\mathbf{k}$

Problem 3.16 Let $\mathbf{f}(x,y,z) = xe^{yz}$. Find

$$(\nabla f)(2,1,1).$$

Problem 3.17 Let $\mathbf{f}(x,y,z) = (xz, e^{xy}, z)$. Find

$$(\nabla \times f)(2,1,1).$$

Problem 3.18 Find the tangent plane to the surface $\frac{x^2}{2} - y^2 - z^2 = 0$ at the point $(2,-1,1)$.

Problem 3.19 Find the point on the surface

$$x^2 + y^2 - 5xy + xz - yz = -3$$

for which the tangent plane is $x - 7y = -6$.

Problem 3.20 Find a vector pointing in the direction in which $\mathbf{f}(x,y,z) = 3xy - 9xz^2 + y$ increases most rapidly at the point $(1,1,0)$.

Problem 3.21 Let $D_{\mathbf{u}}\mathbf{f}(x,y)$ denote the directional derivative of $\mathbf{f}$ at (x,y) in the direction of the unit vector $\mathbf{u}$. If $\nabla f(1,2) = 2\mathbf{i} - \mathbf{j}$, find $D_{\left(\frac{3}{5},\frac{4}{5}\right)}\mathbf{f}(1,2)$.

Problem 3.22 Use a linear approximation of the function $\mathbf{f}(x,y) = e^{x\cos 2y}$ at (0, 0) to estimate $\mathbf{f}(0.1, 0.2)$.

Problem 3.23 Prove that

$$\nabla \bullet (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \bullet (\nabla \times \mathbf{u}) - \mathbf{u} \bullet (\nabla \times \mathbf{v}).$$

Problem 3.24 Find the point on the surface

$$2x^2 + xy + y^2 + 4x + 8y - z + 14 = 0$$

for which the tangent plane is $4x + y - z = 0$.

Problem 3.25 Let $\phi : \mathbb{R}^3 \to \mathbb{R}$ be a scalar field, and let $\mathbf{U}, \mathbf{V} : \mathbb{R}^3 \to \mathbb{R}^3$ be vector fields. Prove that

1. $\nabla \bullet \phi\mathbf{V} = \phi\nabla \bullet \mathbf{V} + \mathbf{V} \bullet \nabla\phi$
2. $\nabla \times \phi\mathbf{V} = \phi\nabla \times \mathbf{V} + (\nabla\phi) \times \mathbf{V}$
3. $\nabla \times (\nabla\phi) = \mathbf{0}$
4. $\nabla \bullet (\nabla \times \mathbf{V}) = 0$
5. $\nabla(\mathbf{U} \bullet \mathbf{V}) = (\mathbf{U} \bullet \nabla)\mathbf{V} + (\mathbf{V} \bullet \nabla)\mathbf{U} + \mathbf{U} \times (\nabla \times \mathbf{V}) + +\mathbf{V} \times (\nabla \times \mathbf{U})$

Problem 3.26 Find the angles made by the gradient of $\mathbf{f}(x,y) = x^{\sqrt{3}} + y$ at the point $(1,1)$ with the coordinate axes.

3.8 The Geometrical Meaning of Divergence and Curl

In this section we provide some heuristics about the meaning of Divergence and Curl. This interpretations will be formally proved in the chapters 6 and 7.

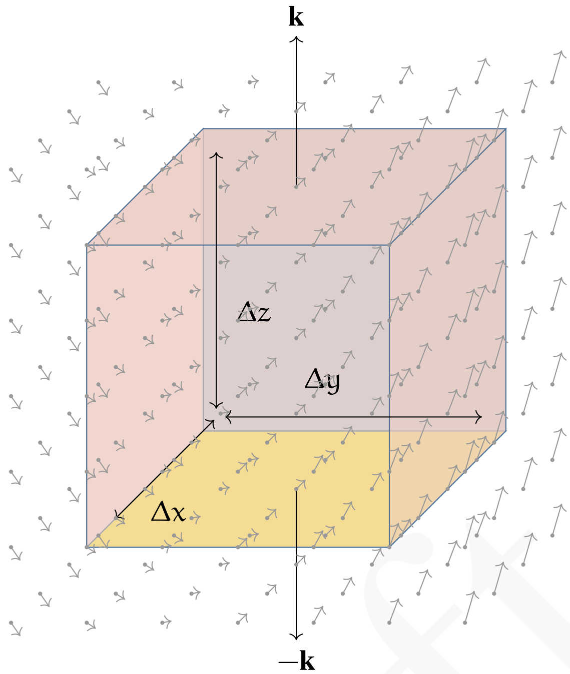


Figure 3.4 Computing the vertical contribution to the flux.

3.8.1 Divergence

Consider a small closed parallelepiped, with sides parallel to the coordinate planes, as shown in Figure 3.4. What is the flux of **F** out of the parallelepiped?

Consider first the vertical contribution, namely the flux up through the top face plus the flux through the bottom face. These two sides each have area $\Delta A = \Delta x\,\Delta y$, but the outward normal vectors point in opposite directions so we get

$$
\begin{aligned}
\sum_{\text{top+bottom}} \mathbf{F} \cdot \Delta\mathbf{A} &\approx \mathbf{F}(z+\Delta z) \cdot \mathbf{k}\,\Delta x\,\Delta y - \mathbf{F}(z) \cdot \mathbf{k}\,\Delta x\,\Delta y \\
&\approx \Big(F_z(z+\Delta z) - F_z(z)\Big)\,\Delta x\,\Delta y \\
&\approx \frac{F_z(z+\Delta z) - F_z(z)}{\Delta z}\,\Delta x\,\Delta y\,\Delta z \\
&\approx \frac{\partial F_z}{\partial z}\,\Delta x\,\Delta y\,\Delta z \qquad \text{by Mean Value Theorem}
\end{aligned}
$$

where we have multiplied and divided by Δz to obtain the volume $\Delta V = \Delta x\,\Delta y\,\Delta z$ in the third step, and used the definition of the derivative in the final step.

Repeating this argument for the remaining pairs of faces, it follows that the total flux out of the parallelepiped is

$$\text{total flux} = \sum_{\text{parallelepiped}} \mathbf{F} \cdot \Delta \mathbf{A} \approx \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) \Delta V$$

Since the total flux is proportional to the volume of the parallelepiped, it approaches zero as the volume of the parallelepiped shrinks down. The interesting quantity is therefore the ratio of the flux to volume; this ratio is called the divergence.

At any point P, we can define the divergence of a vector field $\mathbf{F}$, written $\nabla \cdot \mathbf{F}$, to be the flux of $\mathbf{F}$ per unit volume leaving a small parallelepiped around the point P.

Hence, the divergence of $\mathbf{F}$ at the point P is the flux per unit volume through a small parallelepiped around P, which is given in rectangular coordinates by

$$\nabla \cdot \mathbf{F} = \frac{\text{flux}}{\text{unit volume}} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

Analogous computations can be used to determine expressions for the divergence in other coordinate systems. This computations are presented in chapter 8.

3.8.2 Curl

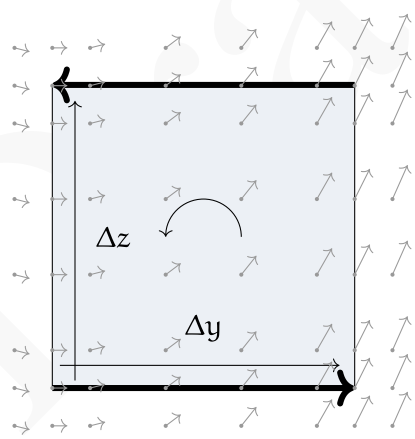


Figure 3.5 Computing the horizontal contribution to the circulation around a small rectangle.

Consider a small rectangle in the yz-plane, with sides parallel to the coordinate axes, as shown in Figure 1. What is the circulation of $\mathbf{F}$ around this rectangle?

Consider first the horizontal edges, on each of which $d\mathbf{r} = \Delta y\, \mathbf{j}$. However, when computing the circulation of $\mathbf{F}$ around this rectangle, we traverse these two edges in opposite directions. In particular, when traversing the rectangle in the counterclockwise direction, $\Delta y < 0$ on top and

$\Delta y > 0$ on the bottom.

$$\sum_{\text{top+bottom}} \mathbf{F}\bullet d\mathbf{r} \approx -\mathbf{F}(z+\Delta z)\bullet\mathbf{j}\,\Delta y + \mathbf{F}(z)\bullet\mathbf{j}\,\Delta y \tag{3.17}$$
$$\approx -\Big(F_y(z+\Delta z) - F_y(z)\Big)\,\Delta y$$
$$\approx -\frac{F_y(z+\Delta z)-F_y(z)}{\Delta z}\,\Delta y\,\Delta z$$
$$\approx -\frac{\partial F_y}{\partial z}\,\Delta y\,\Delta z \qquad \text{by Mean Value Theorem}$$

where we have multiplied and divided by Δz to obtain the surface element $\Delta A = \Delta y\,\Delta z$ in the third step, and used the definition of the derivative in the final step.

Just as with the divergence, in making this argument we are assuming that **F** doesn't change much in the x and y directions, while nonetheless caring about the change in the z direction.

Repeating this argument for the remaining two sides leads to

$$\sum_{\text{sides}} \mathbf{F}\bullet d\mathbf{r} \quad \approx \mathbf{F}(y+\Delta y)\bullet\mathbf{k}\,\Delta z - \mathbf{F}(y)\bullet\mathbf{k}\,\Delta z \tag{3.18}$$
$$\approx \Big(F_z(y+\Delta y) - F_z(y)\Big)\,\Delta z$$
$$\approx \frac{F_z(y+\Delta y)-F_z(y)}{\Delta y}\,\Delta y\,\Delta z$$
$$\approx \tfrac{\partial F_z}{\partial y}\,\Delta y\,\Delta z$$

where care must be taken with the signs, which are different from those in (3.17). Adding up both expressions, we obtain

$$\text{total } yz\text{-circulation} \approx \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\Delta x\,\Delta y \tag{3.19}$$

Since this is proportional to the area of the rectangle, it approaches zero as the area of the rectangle converges to zero. The interesting quantity is therefore the ratio of the circulation to area.

We are computing the **i**-component of the curl.

$$\text{curl}(\mathbf{F})\bullet\mathbf{i} := \frac{yz\text{-circulation}}{\text{unit area}} = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \tag{3.20}$$

The rectangular expression for the full curl now follows by cyclic symmetry, yielding

$$\text{curl}(\mathbf{F}) = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\mathbf{k} \tag{3.21}$$

which is more easily remembered in the form

$$\text{curl}(\mathbf{F}) = \nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \tag{3.22}$$

3.9 Maxwell's Equations

Maxwell's Equations is a set of four equations that describes the behaviors of electromagnetism. Together with the Lorentz Force Law, these equations describe completely (classical) electromagnetism, i. e., all other results are simply mathematical consequences of these equations.

To begin with, there are two fields that govern electromagnetism, known as the *electric* and *magnetic* field. These are denoted by $\mathbf{E}(r,t)$ and $\mathbf{B}(r,t)$ respectively.

To understand electromagnetism, we need to explain how the electric and magnetic fields are formed, and how these fields affect charged particles. The last is rather straightforward, and is described by the Lorentz force law.

Definition 3.9.1 — Lorentz force law. A point charge q experiences a force of

$$\mathbf{F} = q(\mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}).$$

The dynamics of the field itself is governed by Maxwell's Equations. To state the equations, first we need to introduce two more concepts.

Definition 3.9.2 — Charge and current density.

- $\rho(\mathbf{r},t)$ is the **charge density**, defined as the charge per unit volume.
- $\mathbf{j}(\mathbf{r},t)$ is the **current density**, defined as the electric current per unit area of cross section.

Then Maxwell's equations are

Definition 3.9.3 — Maxwell's equations.

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\varepsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \times \mathbf{B} - \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} &= \mu_0 \mathbf{j}, \end{aligned}$$

where ε_0 is the electric constant (i.e, the permittivity of free space) and μ_0 is the magnetic constant (i.e, the permeability of free space), which are constants.

3.10 Inverse Functions

A function $\mathbf{f}$ is said **one-to-one** if $\mathbf{f}(\mathbf{x}_1)$ and $\mathbf{f}(\mathbf{x}_2)$ are distinct whenever $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct points of $\mathrm{Dom}(\mathbf{f})$. In this case, we can define a function $\mathbf{g}$ on the image

$$\mathrm{Im}(\mathbf{f}) = \left\{\mathbf{u} | \mathbf{u} = \mathbf{f}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathrm{Dom}(\mathbf{f})\right\}$$

of $\mathbf{f}$ by defining $\mathbf{g}(\mathbf{u})$ to be the unique point in $\mathrm{Dom}(\mathbf{f})$ such that $\mathbf{f}(\mathbf{u}) = \mathbf{u}$. Then

$$\mathrm{Dom}(\mathbf{g}) = \mathrm{Im}(\mathbf{f}) \quad \text{and} \quad \mathrm{Im}(\mathbf{g}) = \mathrm{Dom}(\mathbf{f}).$$

Moreover, $\mathbf{g}$ is one-to-one,

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}, \quad \mathbf{x} \in \mathrm{Dom}(\mathbf{f}),$$

and

$$\mathbf{f}(\mathbf{g}(\mathbf{u})) = \mathbf{u}, \quad \mathbf{u} \in \mathrm{Dom}(\mathbf{g}).$$

We say that $\mathbf{g}$ is the **inverse** of $\mathbf{f}$, and write $\mathbf{g} = \mathbf{f}^{-1}$. The relation between $\mathbf{f}$ and $\mathbf{g}$ is symmetric; that is, $\mathbf{f}$ is also the inverse of $\mathbf{g}$, and we write $\mathbf{f} = \mathbf{g}^{-1}$.

A transformation $\mathbf{f}$ may fail to be one-to-one, but be one-to-one on a subset S of $\mathrm{Dom}(\mathbf{f})$. By this we mean that $\mathbf{f}(\mathbf{x}_1)$ and $\mathbf{f}(\mathbf{x}_2)$ are distinct whenever $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct points of S. In this case, $\mathbf{f}$ is not invertible, but if $\mathbf{f}|_S$ is defined on S by

$$\mathbf{f}|_S(\mathbf{x}) = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in S,$$

and left undefined for $\mathbf{x} \notin S$, then $\mathbf{f}|_S$ is invertible.

We say that $\mathbf{f}|_S$ is the **restriction of f to** S, and that $\mathbf{f}_S^{-1}$ is the **inverse of f restricted to** S. The domain of $\mathbf{f}_S^{-1}$ is $\mathbf{f}(S)$.

The question of invertibility of an arbitrary transformation $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is too general to have a useful answer. However, there is a useful and easily applicable sufficient condition which implies that one-to-one restrictions of continuously differentiable transformations have continuously differentiable inverses.

Definition 3.10.1 If the function $\mathbf{f}$ is one-to-one on a neighborhood of the point $\mathbf{x}_0$, we say that $\mathbf{f}$ is **locally invertible** at $\mathbf{x}_0$. If aa function is locally invertible for every $\mathbf{x}_0$ in a set S, then $\mathbf{f}$ is said **locally invertible on** S.

To motivate our study of this question, let us first consider the linear transformation

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The function $\mathbf{f}$ is invertible if and only if $\mathbf{A}$ is nonsingular, in which case $\mathrm{Im}(\mathbf{f}) = \mathbb{R}^n$ and

$$\mathbf{f}^{-1}(\mathbf{u}) = \mathbf{A}^{-1}\mathbf{u}.$$

Since $\mathbf{A}$ and $\mathbf{A}^{-1}$ are the differential matrices of $\mathbf{f}$ and $\mathbf{f}^{-1}$, respectively, we can say that a linear transformation is invertible if and only if its differential matrix $\mathbf{f}'$ is nonsingular, in which case the differential matrix of $\mathbf{f}^{-1}$ is given by

$$(\mathbf{f}^{-1})' = (\mathbf{f}')^{-1}.$$

Because of this, it is tempting to conjecture that if $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and $\mathbf{A}'(\mathbf{x})$ is nonsingular, or, equivalently, $D(\mathbf{f})(\mathbf{x}) \neq 0$, for $\mathbf{x}$ in a set S, then $\mathbf{f}$ is one-to-one on S. However, this is false. For example, if

$$\mathbf{f}(x,y) = [e^x \cos y, e^x \sin y],$$

then

$$D(\mathbf{f})(x,y) = \begin{vmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{vmatrix} = e^{2x} \neq 0, \tag{3.23}$$

but $\mathbf{f}$ is not one-to-one on $\mathbb{R}^2$. The best that can be said in general is that if $\mathbf{f}$ is continuously differentiable and $D(\mathbf{f})(\mathbf{x}) \neq 0$ in an open set S, then $\mathbf{f}$ is locally invertible on S, and the local inverses are continuously differentiable. This is part of the inverse function theorem, which we will prove presently.

Theorem 3.10.1 — Inverse Function Theorem. If $\mathbf{f} : U \to \mathbb{R}^n$ is differentiable at a and $D_{\mathbf{a}}(\mathbf{f})$ is invertible, then there exists a domains U', V' such that $a \in U' \subseteq U$, $\mathbf{f}(a) \in V'$ and $\mathbf{f} : U' \to V'$ is bijective. Further, the inverse function $\mathbf{g} : V' \to U'$ is differentiable.

The proof of the Inverse Function Theorem will be presented in the Section A.

We note that the condition about the invertibility of $D_{\mathbf{a}}(\mathbf{f})$ is necessary. If $\mathbf{f}$ has a differentiable inverse in a neighborhood of a, then $D_{\mathbf{a}}(\mathbf{f})$ must be invertible. To see this differentiate the identity

$$\mathbf{f}(\mathbf{g}(x)) = x$$

3.11 Implicit Functions

Let $U \subseteq \mathbb{R}^{n+1}$ be a domain and $f : U \to \mathbb{R}$ be a differentiable function. If $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we'll concatenate the two vectors and write $(\mathbf{x}, y) \in \mathbb{R}^{n+1}$.

Theorem 3.11.1 — Special Implicit Function Theorem. Suppose $c = f(\mathbf{a}, b)$ and $\partial_y f(\mathbf{a}, b) \neq 0$. Then, there exists a domain $U' \ni a$ and differentiable function $g : U' \to \mathbb{R}$ such that $g(\mathbf{a}) = b$ and $f(x, g(x)) = c$ for all $x \in U'$.

Further, there exists a domain $V' \ni b$ such that

$$\Big\{(x,y) \;\Big|\; x \in U', y \in V', f(x,y) = c\Big\} = \Big\{(x, g(x)) \;\Big|\; x \in U'\Big\}.$$

In other words, for all $x \in U'$ the equation $f(x,y) = c$ has a unique solution in V' and is given by $y = g(x)$.

R To see why $\partial_y f \neq 0$ is needed, let $f(x,y) = \alpha x + \beta y$ and consider the equation $f(x,y) = c$. To express y as a function of x we need $\beta \neq 0$ which in this case is equivalent to $\partial_y f \neq 0$.

R If $n = 1$, one expects $f(x,y) = c$ to some curve in $\mathbb{R}^2$. To write this curve in the form $y = g(x)$ using a differentiable function g, one needs the curve to never be vertical. Since ∇f is perpendicular to the curve, this translates to ∇f never being horizontal, or equivalently $\partial_y f \neq 0$ as assumed in the theorem.

R For simplicity we choose y to be the last coordinate above. It could have been any other, just as long as the corresponding partial was non-zero. Namely if $\partial_i f(a) \neq 0$, then one can locally solve the equation $f(x) = f(a)$ (uniquely) for the variable x_i and express it as a differentiable function of the remaining variables.

Example 3.11.2 $f(x,y) = x^2 + y^2$ with $c = 1$.

Proof: [of the Special Implicit Function Theorem] Let $\mathbf{f}(x,y) = (x, f(x,y))$, and observe $D(\mathbf{f})_{(a,b)} \neq 0$. By the inverse function theorem $\mathbf{f}$ has a unique local inverse $\mathbf{g}$. Note $\mathbf{g}$ must be of the form $\mathbf{g}(x,y) = (x, g(x,y))$. Also $\mathbf{f} \circ \mathbf{g} = \mathrm{Id}$ implies $(x,y) = \mathbf{f}(x, g(x,y)) = (x, f(x, g(x,y))$. Hence $y = g(x,c)$ uniquely solves $f(x,y) = c$ in a small neighbourhood of (a,b). ■

Instead of $y \in \mathbb{R}$ above, we could have been fancier and allowed $y \in \mathbb{R}^n$. In this case f needs to be an $\mathbb{R}^n$ valued function, and we need to replace $\partial_y f \neq 0$ with the assumption that the $n \times n$ minor in $D(f)$ (corresponding to the coordinate positions of y) is invertible. This is the general version of the implicit function theorem.

Theorem 3.11.3 — General Implicit Function Theorem. Let $U \subseteq \mathbb{R}^{m+n}$ be a domain. Suppose $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is C^1 on an open set containing (a, b) where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. Suppose $\mathbf{f}(a, b) = 0$ and that the $m \times m$ matrix $M = (D_{n+j}\mathbf{f}_i(a, b))$ is nonsingular. Then that there is an open set $A \subset \mathbb{R}^n$ containing a and an open set $B \subset \mathbb{R}^m$ containing b such that, for each $x \in A$, there is a unique $\mathbf{g}(x) \in B$ such that $\mathbf{f}(x, g(x)) = 0$. Furthermore, $\mathbf{g}$ is differentiable.

In other words: if the matrix M is invertible, then one can locally solve the equation $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$ (uniquely) for the variables $x_{i_1}, \ldots, x_{i_m}$ and express them as a differentiable function of the remaining n variables.

The proof of the General Implicit Function Theorem will be presented in the Section A.

Example 3.11.4 Consider the equations

$$(x-1)^2 + y^2 + z^2 = 5 \quad \text{and} \quad (x+1)^2 + y^2 + z^2 = 5$$

for which $x = 0$, $y = 0$, $z = 2$ is one solution. For all other solutions close enough to this point, determine which of variables x, y, z can be expressed as differentiable functions of the others.

Solution: ► Let $a = (0, 0, 1)$ and

$$F(x, y, z) = \begin{bmatrix} (x-1)^2 + y^2 + z^2 \\ (x+1)^2 + y^2 + z^2 \end{bmatrix}$$

Observe

$$DF_a = \begin{bmatrix} -2 & 0 & 4 \\ 2 & 0 & 4 \end{bmatrix},$$

and the 2×2 minor using the first and last column is invertible. By the implicit function theorem this means that in a small neighborhood of a, x and z can be (uniquely) expressed in terms of y. ◄

R In the above example, one can of course solve explicitly and obtain

$$x = 0 \quad \text{and} \quad z = \sqrt{4 - y^2},$$

but in general we won't be so lucky.

II

Integral Vector Calculus

CHAPTER 4

Multiple Integrals

In this chapter we develop the theory of integration for scalar functions.

In single-variable calculus, differentiation and integration are thought of as inverse operations. For instance, to integrate a function $f(x)$ it is necessary to find the antiderivative of f, that is, another function $F(x)$ whose derivative is $f(x)$. Is there a similar way of defining integration of real-valued functions of two or more variables? The answer is yes, as we will see shortly. Recall also that the definite integral of a nonnegative function $f(x) \geqslant 0$ represented the area "under" the curve $y = f(x)$. As we will now see, the *double integral* of a nonnegative real-valued function $f(x,y) \geqslant 0$ represents the *volume* "under" the surface $z = f(x,y)$.

4.1 Double Integrals

Let $R = [a,b] \times [c,d] \subseteq \mathbb{R}^2$ be a rectangle, and $f : R \to \mathbb{R}$ be continuous. Let $P = \{x_0, \ldots, x_M, y_0, \ldots, y_M\}$ where $a = x_0 < x_1 < \cdots < x_M = b$ and $c = y_0 < y_1 < \cdots < y_M = d$. The set P determines a partition of R into a grid of (non-overlapping) rectangles $R_{i,j} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ for $0 \leqslant i < M$ and $0 \leqslant j < N$. Given P, choose a collection of points $M = \left\{\xi_{i,j}\right\}$ so that $\xi_{i,j} \in R_{i,j}$ for all i,j.

Definition 4.1.1 The **Riemann sum** of f with respect to the partition P and points M is defined by

$$\mathcal{R}(f,P,M) \stackrel{\text{def}}{=} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(\xi_{i,j}) \operatorname{area}(R_{i,j}) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(\xi_{i,j})(x_{i+1} - x_i)(y_{j+1} - y_j)$$

Definition 4.1.2 The **mesh size** of a partition P is defined by

$$\|P\| = \max\left\{x_{i+1} - x_i \mid 0 \leqslant i < M\right\} \cup \left\{y_{j+1} - y_j \mid 0 \leqslant j \leqslant N\right\}.$$

Definition 4.1.3 The **Riemann integral** of f over the rectangle R is defined by

$$\int_R f(x,y)\,dx\,dy \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} \mathcal{R}(f,P,M),$$

provided the limit exists and is independent of the choice of the points M. A function is said to be **Riemann integrable** over R if the Riemann integral exists and is finite.

R A few other popular notation conventions used to denote the integral are

$$\iint_R f\,dA, \quad \iint_R f\,dx\,dy, \quad \iint_R f\,dx_1\,dx_2, \quad \text{and} \quad \iint_R f.$$

R The double integral represents the volume of the region under the graph of f. Alternately, if $f(x,y)$ is the density of a planar body at point (x,y), the double integral is the total mass.

Theorem 4.1.1 Any bounded continuous function is Riemann integrable on a bounded rectangle.

R Most bounded functions we will encounter will be Riemann integrable. Bounded functions with reasonable discontinuities (e.g. finitely many jumps) are usually Riemann integrable on bounded rectangle. An example of a "badly discontinuous" function that is not Riemann integrable is the function $f(x,y) = 1$ if $x,y \in \mathbb{Q}$ and 0 otherwise.

Now suppose $U \subseteq \mathbb{R}^2$ is an nice bounded[1] domain, and $f : U \to \mathbb{R}$ is a function. Find a bounded rectangle $R \supseteq U$, and as before let P be a partition of R into a grid of rectangles. Now we define the Riemann sum by only summing over all rectangles $R_{i,j}$ that are completely contained inside U. Explicitly, let

$$\chi_{i,j} = \begin{cases} 1 & R_{i,j} \subseteq U \\ 0 & \text{otherwise.} \end{cases}$$

[1] We will subsequently always assume U is "nice". Namely, U is open, connected and the boundary of U is a piecewise differentiable curve. More precisely, we need to assume that the "area" occupied by the boundary of U is 0. While you might suspect this should be true for all open sets, it isn't! There exist open sets of *finite area* whose boundary occupies an infinite area!

and define

$$\mathcal{R}(f,P,M,U) \stackrel{\text{def}}{=} \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \chi_{i,j} f(\xi_{i,j})(x_{i+1}-x_i)(y_{j+1}-y_j).$$

Definition 4.1.4 The **Riemann integral** of f over the domain $U$is defined by

$$\int_U f(x,y)\,dx\,dy \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} \mathcal{R}(f,P,M,U),$$

provided the limit exists and is independent of the choice of the points M. A function is said to be **Riemann integrable** over R if the Riemann integral exists and is finite.

Theorem 4.1.2 Any bounded continuous function is Riemann integrable on a bounded region.

R As before, most reasonable bounded functions we will encounter will be Riemann integrable.

To deal with unbounded functions over unbounded domains, we use a limiting process.

Definition 4.1.5 Let $U \subseteq \mathbb{R}^2$ be a domain (which is not necessarily bounded) and $f : U \to \mathbb{R}$ be a (not necessarily bounded) function. We say f is integrable if

$$\lim_{R\to\infty} \int_{U\cap B(0,R)} \chi_R |f|\, dA$$

exists and is finite. Here $\chi_R(x) = 1$ if $|f(x)| < R$ and 0 otherwise.

Proposition 4.1.3 If f is integrable on the domain U, then

$$\lim_{R\to\infty} \int_{U\cap B(0,R)} \chi_R f\, dA$$

exists and is finite.

R If f is integrable, then the above limit is independent of how you expand your domain. Namely, you can take the limit of the integral over $U \cap [-R,R]^2$ instead, and you will still get the same answer.

Definition 4.1.6 If f is integrable we define

$$\int_U f\,dx\,dy = \lim_{R\to\infty}\int_{U\cap B(0,R)} \chi_R f\,dA$$

4.2 Iterated integrals and Fubini's theorem

Let $f(x,y)$ be a continuous function such that $f(x,y) \geqslant 0$ for all (x,y) on the **rectangle** $R = \{(x,y) : a \leqslant x \leqslant b,\ c \leqslant y \leqslant d\}$ in $\mathbb{R}^2$. We will often write this as $R = [a,b]\times[c,d]$. For any number $x*$ in the interval $[a,b]$, slice the surface $z = f(x,y)$ with the plane $x = x*$ parallel to the yz-plane. Then the trace of the surface in that plane is the *curve* $f(x*,y)$, where $x*$ is fixed and only y varies. The area A under that curve (i.e. the area of the region between the curve and the xy-plane) as y varies over the interval $[c,d]$ then depends only on the value of $x*$. So using the variable x instead of $x*$, let $A(x)$ be that area (see Figure 4.1).

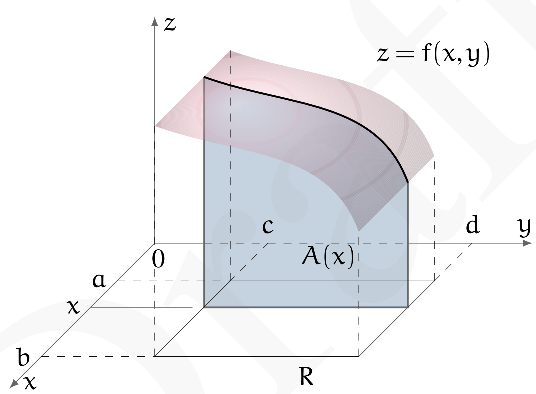


Figure 4.1 The area $A(x)$ varies with x

Then $A(x) = \displaystyle\int_c^d f(x,y)\,dy$ since we are treating x as fixed, and only y varies. This makes sense since for a fixed x the function $f(x,y)$ is a continuous function of y over the interval $[c,d]$, so we know that the area under the curve is the definite integral. The area $A(x)$ is a function of x, so by the "slice" or cross-section method from single-variable calculus we know that the volume V of the *solid* under the surface $z = f(x,y)$ but above the xy-plane over the rectangle R is the integral over $[a,b]$ of that cross-sectional area $A(x)$:

$$V = \int_a^b A(x)\,dx = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx \tag{4.1}$$

We will always refer to this volume as "the volume under the surface". The above expression uses what are called **iterated integrals**. First the function $f(x,y)$ is integrated as a function of y, treating the variable x as a constant (this is called *integrating with respect to* y). That is what occurs in the "inner" integral between the square brackets in equation (4.1). This is the first iterated integral. Once that integration is performed, the result is then an expression involving only x, which can then be *integrated with respect to* x. That is what occurs in the "outer" integral above (the second iterated integral). The final result is then a number (the volume). This process of going through two iterations of integrals is called *double integration*, and the last expression in equation (4.1) is called a **double integral**.

Notice that integrating $f(x,y)$ with respect to y is the inverse operation of taking the partial derivative of $f(x,y)$ with respect to y. Also, we could just as easily have taken the area of cross-sections under the surface which were parallel to the xz-plane, which would then depend only on the variable y, so that the volume V would be

$$V = \int_c^d \left[\int_a^b f(x,y)\,dx \right] dy\,. \tag{4.2}$$

It turns out that in general due to Fubini's Theorem the order of the iterated integrals does not matter. Also, we will usually discard the brackets and simply write

$$V = \int_c^d \int_a^b f(x,y)\,dx\,dy\,, \tag{4.3}$$

where it is understood that the fact that dx is written before dy means that the function $f(x,y)$ is first integrated with respect to x using the "inner" limits of integration a and b, and then the resulting function is integrated with respect to y using the "outer" limits of integration c and d. This order of integration can be changed if it is more convenient.

Let $U \subseteq \mathbb{R}^2$ be a domain.

Definition 4.2.1 For $x \in \mathbb{R}$, define

$$S_x U = \left\{ y \mid (x,y) \in U \right\} \quad \text{and} \quad T_y U = \left\{ x \mid (x,y) \in U \right\}$$

■ **Exemplo 4.1** If $U = [a,b] \times [c,d]$ then

$$S_x U = \begin{cases} [c,d] & x \in [a,b] \\ \emptyset & x \notin [a,b] \end{cases} \quad \text{and} \quad T_y U = \begin{cases} [a,b] & y \in [c,d] \\ \emptyset & y \notin [c,d]. \end{cases}$$

■

For domains we will consider, $S_x U$ and $T_y U$ will typically be an interval (or a finite union of intervals).

Definition 4.2.2 Given a function $f : U \to \mathbb{R}$, we define the two *iterated* integrals by

$$\int_{x\in\mathbb{R}} \Big(\int_{y\in S_x U} f(x,y)\,dy \Big)\,dx \quad \text{and} \quad \int_{y\in\mathbb{R}} \Big(\int_{x\in T_y U} f(x,y)\,dx \Big)\,dy,$$

with the convention that an integral over the empty set is 0. (We included the parenthesis above for clarity; and will drop them as we become more familiar with iterated integrals.)

Suppose $f(x,y)$ represents the density of a planar body at point (x,y). For any $x \in \mathbb{R}$,

$$\int_{y\in S_x U} f(x,y)\,dy$$

represents the mass of the body contained in the vertical line through the point $(x,0)$. It's only natural to expect that if we integrate this with respect to y, we will get the total mass, which is the double integral. By the same argument, we should get the same answer if we had sliced it horizontally first and then vertically. Consequently, we expect both iterated integrals to be equal to the double integral. This is true, under a finiteness assumption.

Theorem 4.2.1 — Fubini's theorem. Suppose $f : U \to \mathbb{R}$ is a function such that either

$$\int_{x\in\mathbb{R}} \Big(\int_{y\in S_x U} |f(x,y)|\,dy \Big)\,dx < \infty \quad \text{or} \quad \int_{y\in\mathbb{R}} \Big(\int_{x\in T_y U} |f(x,y)|\,dx \Big)\,dy < \infty, \tag{4.4}$$

then f is integrable over U and

$$\int_U f\,dA = \int_{x\in\mathbb{R}} \Big(\int_{y\in S_x U} f(x,y)\,dy \Big)\,dx = \int_{y\in\mathbb{R}} \Big(\int_{x\in T_y U} f(x,y)\,dx \Big)\,dy.$$

Without the assumption (4.4) the iterated integrals need not be equal, even though both may exist and be finite.

■ **Exemplo 4.2** Define

$$f(x,y) = -\partial_x \partial_y \tan^{-1}\Big(\frac{y}{x}\Big) = \frac{x^2-y^2}{(x^2+y^2)^2}.$$

Then

$$\int_{x=0}^{1} \int_{y=0}^{1} f(x,y)\,dy\,dx = \frac{\pi}{4} \quad \text{and} \quad \int_{y=0}^{1} \int_{x=0}^{1} f(x,y)\,dx\,dy = -\frac{\pi}{4}$$

■

■ **Exemplo 4.3** Let $f(x,y) = (x-y)/(x+y)^3$ if $x,y > 0$ and 0 otherwise, and $U = (0,1)^2$. The iterated integrals of f over U both exist, but are not equal. ■

■ **EXEMPLO 4.4** Define

$$f(x,y) = \begin{cases} 1 & y \in (x, x+1) \text{ and } x \geqslant 0 \\ -1 & y \in (x-1, x) \text{ and } x \geqslant 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then the iterated integrals of f both exist and are not equal. ■

Example 4.2.2 Find the volume V under the plane $z = 8x + 6y$ over the rectangle $R = [0,1] \times [0,2]$.

Solution: ► We see that $f(x,y) = 8x + 6y \geqslant 0$ for $0 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 2$, so:

$$\begin{aligned} V &= \int_0^2 \int_0^1 (8x + 6y)\,dx\,dy \\ &= \int_0^2 \left(4x^2 + 6xy \Big|_{x=0}^{x=1} \right) dy \\ &= \int_0^2 (4 + 6y)\,dy \\ &= 4y + 3y^2 \Big|_0^2 \\ &= 20 \end{aligned}$$

Suppose we had switched the order of integration. We can verify that we still get the same answer:

$$\begin{aligned} V &= \int_0^1 \int_0^2 (8x + 6y)\,dy\,dx \\ &= \int_0^1 \left(8xy + 3y^2 \Big|_{y=0}^{y=2} \right) dx \\ &= \int_0^1 (16x + 12)\,dx \\ &= 8x^2 + 12x \Big|_0^1 \\ &= 20 \end{aligned}$$

◄

Example 4.2.3 Find the volume V under the surface $z = e^{x+y}$ over the rectangle $R = [2,3] \times [1,2]$.

Solution: ► We know that $f(x,y) = e^{x+y} > 0$ for all (x,y), so

$$\begin{aligned} V &= \int_1^2 \int_2^3 e^{x+y}\,dx\,dy \\ &= \int_1^2 \left(e^{x+y} \Big|_{x=2}^{x=3} \right) dy \\ &= \int_1^2 (e^{y+3} - e^{y+2})\,dy \end{aligned}$$

$$= e^{y+3} - e^{y+2}\Big|_1^2$$
$$= e^5 - e^4 - (e^4 - e^3) = e^5 - 2e^4 + e^3$$

◀

Recall that for a general function $f(x)$, the integral $\int_a^b f(x)\,dx$ represents the difference of the area below the curve $y = f(x)$ but above the x-axis when $f(x) \geqslant 0$, and the area above the curve but below the x-axis when $f(x) \leqslant 0$. Similarly, the double integral of any continuous function $f(x,y)$ represents the difference of the volume below the surface $z = f(x,y)$ but above the xy-plane when $f(x,y) \geqslant 0$, and the volume above the surface but below the xy-plane when $f(x,y) \leqslant 0$. Thus, our method of double integration by means of iterated integrals can be used to evaluate the double integral of *any* continuous function over a rectangle, regardless of whether $f(x,y) \geqslant 0$ or not.

Example 4.2.4 Evaluate $\int_0^{2\pi}\int_0^{\pi} \sin(x+y)\,dx\,dy$.

Solution: ▶ Note that $f(x,y) = \sin(x+y)$ is both positive and negative over the rectangle $[0,\pi] \times [0,2\pi]$. We can still evaluate the double integral:

$$\int_0^{2\pi}\int_0^{\pi} \sin(x+y)\,dx\,dy = \int_0^{2\pi} \left(-\cos(x+y)\Big|_{x=0}^{x=\pi}\right) dy$$
$$= \int_0^{2\pi} (-\cos(y+\pi) + \cos y)\,dy$$
$$= -\sin(y+\pi) + \sin y\Big|_0^{2\pi} = -\sin 3\pi + \sin 2\pi - (-\sin\pi + \sin 0)$$
$$= 0$$

◀

Exercises

A

For Exercises 1-4, find the volume under the surface $z = f(x,y)$ over the rectangle R.

1. $f(x,y) = 4xy$, $R = [0,1] \times [0,1]$

2. $f(x,y) = e^{x+y}$, $R = [0,1] \times [-1,1]$

3. $f(x,y) = x^3 + y^2$, $R = [0,1] \times [0,1]$

4. $f(x,y) = x^4 + xy + y^3$, $R = [1,2] \times [0,2]$

For Exercises 5-12, evaluate the given double integral.

5. $\int_0^1\int_1^2 (1-y)x^2\,dx\,dy$

6. $\int_0^1\int_0^2 x(x+y)\,dx\,dy$

7. $\displaystyle\int_0^2 \int_0^1 (x+2)\,dx\,dy$

8. $\displaystyle\int_{-1}^2 \int_{-1}^1 x(xy+\sin x)\,dx\,dy$

9. $\displaystyle\int_0^{\pi/2} \int_0^1 xy\cos(x^2y)\,dx\,dy$

10. $\displaystyle\int_0^{\pi} \int_0^{\pi/2} \sin x\cos(y-\pi)\,dx\,dy$

11. $\displaystyle\int_0^2 \int_1^4 xy\,dx\,dy$

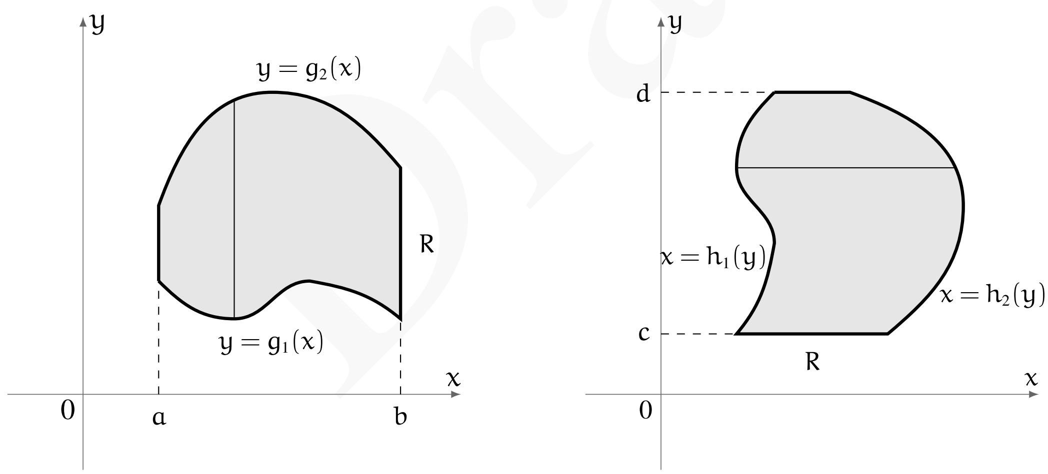
12. $\displaystyle\int_{-1}^1 \int_{-1}^2 1\,dx\,dy$

13. Let M be a constant. Show that $\displaystyle\int_c^d \int_a^b M\,dx\,dy = M(d-c)(b-a)$.

4.3 Double Integrals Over a General Region

In the previous section we got an idea of what a double integral over a rectangle represents. We can now define the double integral of a real-valued function $f(x,y)$ over more general regions in $\mathbb{R}^2$.

Suppose that we have a region R in the xy-plane that is bounded on the left by the vertical line $x = a$, bounded on the right by the vertical line $x = b$ (where $a < b$), bounded below by a curve $y = g_1(x)$, and bounded above by a curve $y = g_2(x)$, as in Figure 4.2(a). We will assume that $g_1(x)$ and $g_2(x)$ do not intersect on the open interval (a,b) (they could intersect at the endpoints $x = a$ and $x = b$, though).



(a) Vertical slice: $\displaystyle\int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx$

(b) Horizontal slice: $\displaystyle\int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy$

Figure 4.2 Double integral over a nonrectangular region R

Then using the slice method from the previous section, the double integral of a real-valued function $f(x,y)$ over the region R, denoted by $\iint\limits_R f(x,y)\,dA$, is given by

$$\iint\limits_R f(x,y)\,dA \;=\; \int_a^b \left[\int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\right] dx \tag{4.5}$$

This means that we take vertical slices in the region R between the curves $y = g_1(x)$ and $y = g_2(x)$. The symbol dA is sometimes called an *area element* or *infinitesimal*, with the A signifying area. Note that $f(x,y)$ is first integrated with respect to y, with functions of x as the limits of integration. This makes sense since the result of the first iterated integral will have to be a function of x alone, which then allows us to take the second iterated integral with respect to x.

Similarly, if we have a region R in the xy-plane that is bounded on the left by a curve $x = h_1(y)$, bounded on the right by a curve $x = h_2(y)$, bounded below by the horizontal line $y = c$, and bounded above by the horizontal line $y = d$ (where $c < d$), as in Figure 4.2(b) (assuming that $h_1(y)$ and $h_2(y)$ do not intersect on the open interval (c,d)), then taking horizontal slices gives

$$\iint\limits_R f(x,y)\,dA \;=\; \int_c^d \left[\int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\right] dy \tag{4.6}$$

Notice that these definitions include the case when the region R is a rectangle. Also, if $f(x,y) \geqslant 0$ for all (x,y) in the region R, then $\iint\limits_R f(x,y)\,dA$ is the volume under the surface $z = f(x,y)$ over the region R.

Example 4.3.1 Find the volume V under the plane $z = 8x + 6y$ over the region $R = \{(x,y) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 2x^2\}$.

Solution: ► The region R is shown in Figure 3.2.2. Using vertical slices we get:

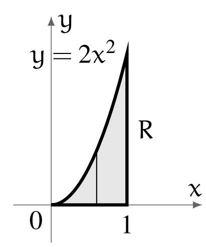


Figure 4.3

$$\begin{aligned}
V &= \iint\limits_R (8x+6y)\,dA \\
&= \int_0^1 \left[\int_0^{2x^2} (8x+6y)\,dy\right] dx \\
&= \int_0^1 \left(8xy + 3y^2 \Big|_{y=0}^{y=2x^2}\right) dx \\
&= \int_0^1 (16x^3 + 12x^4)\,dx \\
&= 4x^4 + \tfrac{12}{5}x^5 \Big|_0^1 \;=\; 4 + \tfrac{12}{5} \;=\; \tfrac{32}{5} \;=\; 6.4
\end{aligned}$$

We get the same answer using horizontal slices (see Figure 3.2.3):

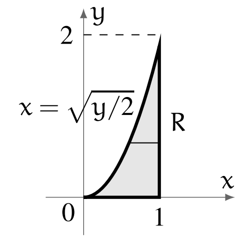


Figure 4.4

$$
\begin{aligned}
V &= \iint\limits_R (8x+6y)\,dA \\
&= \int_0^2 \left[\int_{\sqrt{y/2}}^1 (8x+6y)\,dx \right] dy \\
&= \int_0^2 \left(4x^2 + 6xy \Big|_{x=\sqrt{y/2}}^{x=1} \right) dy \\
&= \int_0^2 (4+6y-(2y+\tfrac{6}{\sqrt{2}}y\sqrt{y}))\,dy \;=\; \int_0^2 (4+4y-3\sqrt{2}y^{3/2})\,dy \\
&= 4y+2y^2 - \tfrac{6\sqrt{2}}{5}y^{5/2}\Big|_0^2 \;=\; 8+8-\tfrac{6\sqrt{2}\sqrt{32}}{5} \;=\; 16-\tfrac{48}{5} \;=\; \tfrac{32}{5} \;=\; 6.4
\end{aligned}
$$

◄

Example 4.3.2 Find the volume V of the solid bounded by the three coordinate planes and the plane $2x+y+4z=4$.

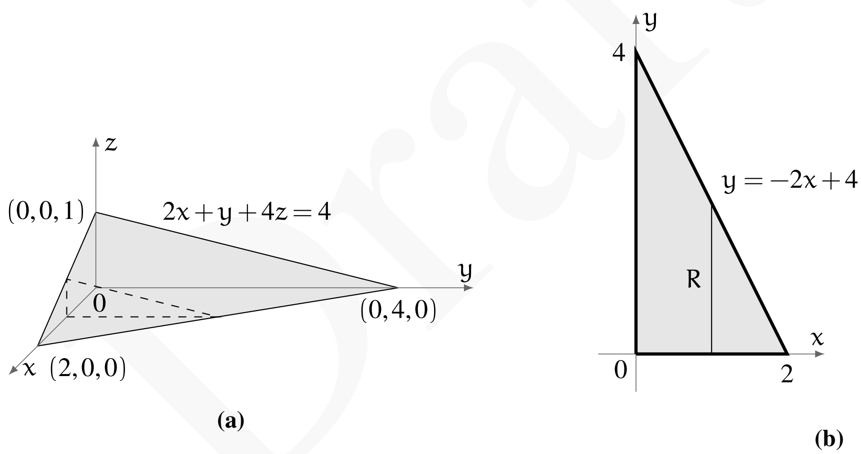


Figure 4.5

Solution: ► The solid is shown in Figure 4.5(a) with a typical vertical slice. The volume V is given by $\iint\limits_R f(x,y)\,dA$, where $f(x,y) = z = \frac{1}{4}(4-2x-y)$ and the region R, shown in Figure 4.5(b), is $R = \{(x,y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant -2x+4\}$. Using vertical slices in R gives

$$
\begin{aligned}
V &= \iint\limits_R \tfrac{1}{4}(4-2x-y)\,dA \\
&= \int_0^2 \left[\int_0^{-2x+4} \tfrac{1}{4}(4-2x-y)\,dy \right] dx
\end{aligned}
$$

$$= \int_0^2 \left(-\tfrac{1}{8}(4-2x-y)^2 \Big|_{y=0}^{y=-2x+4} \right) dx$$
$$= \int_0^2 \tfrac{1}{8}(4-2x)^2 \, dx$$
$$= -\tfrac{1}{48}(4-2x)^3 \Big|_0^2 = \tfrac{64}{48} = \tfrac{4}{3}$$

◀

For a general region R, which may not be one of the types of regions we have considered so far, the double integral $\iint\limits_R f(x,y)\,dA$ is defined as follows. Assume that $f(x,y)$ is a nonnegative real-valued function and that R is a bounded region in $\mathbb{R}^2$, so it can be enclosed in some rectangle $[a,b] \times [c,d]$. Then divide that rectangle into a grid of subrectangles. Only consider the subrectangles that are enclosed completely within the region R, as shown by the shaded subrectangles in Figure 4.6(a). In any such subrectangle $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$, pick a point (x_{i*}, y_{j*}). Then the volume under the surface $z = f(x,y)$ over that subrectangle is approximately $f(x_{i*}, y_{j*})\,\Delta x_i\,\Delta y_j$, where $\Delta x_i = x_{i+1} - x_i$, $\Delta y_j = y_{j+1} - y_j$, and $f(x_{i*}, y_{j*})$ is the height and $\Delta x_i\,\Delta y_j$ is the base area of a parallelepiped, as shown in Figure 4.6(b). Then the total volume under the surface is approximately the sum of the volumes of all such parallelepipeds, namely

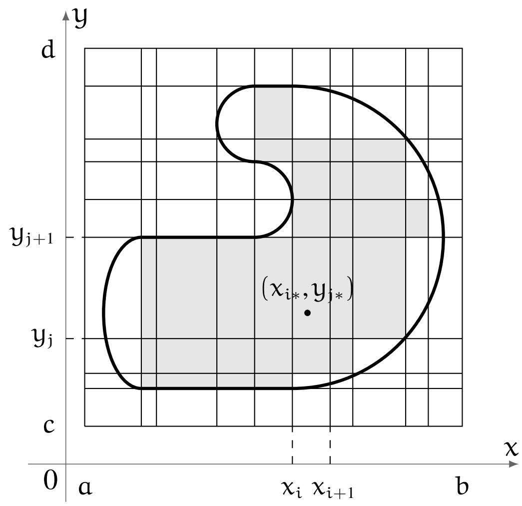
$$\sum_j \sum_i f(x_{i*}, y_{j*})\,\Delta x_i\,\Delta y_j \,, \tag{4.7}$$

where the summation occurs over the indices of the subrectangles inside R. If we take smaller and smaller subrectangles, so that the length of the largest diagonal of the subrectangles goes to 0, then the subrectangles begin to fill more and more of the region R, and so the above sum approaches the actual volume under the surface $z = f(x,y)$ over the region R. We then *define* $\iint\limits_R f(x,y)\,dA$ as the limit of that double summation (the limit is taken over all subdivisions of the rectangle $[a,b] \times [c,d]$ as the largest diagonal of the subrectangles goes to 0).

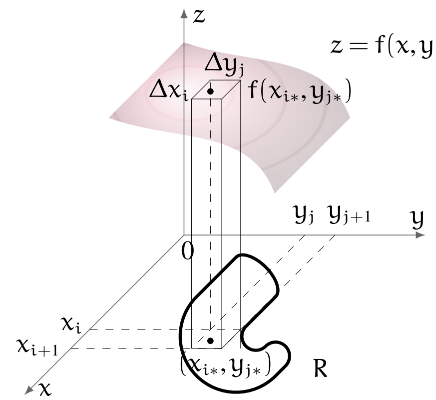
A similar definition can be made for a function $f(x,y)$ that is not necessarily always nonnegative: just replace each mention of volume by the negative volume in the description above when $f(x,y) < 0$. In the case of a region of the type shown in Figure 4.2, using the definition of the Riemann integral from single-variable calculus, our definition of $\iint\limits_R f(x,y)\,dA$ reduces to a sequence of two iterated integrals.

Finally, the region R does not have to be bounded. We can evaluate *improper* double integrals (i.e. over an unbounded region, or over a region which contains points where the function $f(x,y)$ is not defined) as a sequence of iterated improper single-variable integrals.

Example 4.3.3 Evaluate $\displaystyle\int_1^\infty \int_0^{1/x^2} 2y \, dy \, dx$.



(a) Subrectangles inside the region R



(b) Parallelepiped over a subrectangle, with volume $f(x_{i*}, y_{j*})\,\Delta x_i\,\Delta y_j$

Figure 4.6 Double integral over a general region R

Solution: ▶

$$\int_1^\infty \int_0^{1/x^2} 2y\,dy\,dx = \int_1^\infty \left(y^2 \Big|_{y=0}^{y=1/x^2} \right) dx$$
$$= \int_1^\infty x^{-4}\,dx = -\tfrac{1}{3}x^{-3}\Big|_1^\infty = 0 - (-\tfrac{1}{3}) = \tfrac{1}{3}$$

◀

Exercises

A

For Exercises 1-6, evaluate the given double integral.

1. $\displaystyle\int_0^1 \int_{\sqrt{x}}^1 24x^2 y\,dy\,dx$

2. $\displaystyle\int_0^\pi \int_0^y \sin x\,dx\,dy$

3. $\displaystyle\int_1^2 \int_0^{\ln x} 4x\,dy\,dx$

4. $\displaystyle\int_0^2 \int_0^{2y} e^{y^2}\,dx\,dy$

5. $\displaystyle\int_0^{\pi/2} \int_0^y \cos x \sin y\,dx\,dy$

6. $\displaystyle\int_0^\infty \int_0^\infty xye^{-(x^2+y^2)}\,dx\,dy$

7. $\displaystyle\int_0^2 \int_0^y 1\,dx\,dy$

8. $\displaystyle\int_0^1 \int_0^{x^2} 2\,dy\,dx$

9. Find the volume V of the solid bounded by the three coordinate planes and the plane $x+y+z=1$.

10. Find the volume V of the solid bounded by the three coordinate planes and the plane $3x+2y+5z=6$.

B

11. Explain why the double integral $\iint\limits_R 1\,dA$ gives the area of the region R. For simplicity, you can assume that R is a region of the type shown in Figure 4.2(a).

C

12. Prove that the volume of a tetrahedron with mutually perpendicular adjacent sides of lengths a, b, and c, as in Figure 3.2.6, is $\frac{abc}{6}$. (*Hint: Mimic Example 4.3.2, and recall from Section 1.5 how three noncollinear points determine a plane.*)

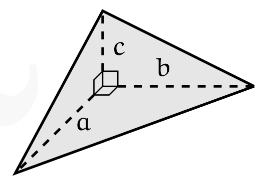


Figure 4.7

13. Show how Exercise 12 can be used to solve Exercise 10.

4.4 Triple Integrals

Our definition of a double integral of a real-valued function $f(x,y)$ over a region R in $\mathbb{R}^2$ can be extended to define a *triple integral* of a real-valued function $f(x,y,z)$ over a *solid* S in $\mathbb{R}^3$. We simply proceed as before: the solid S can be enclosed in some rectangular parallelepiped, which is then divided into subparallelepipeds. In each subparallelepiped inside S, with sides of lengths Δx, Δy and Δz, pick a point (x_*, y_*, z_*). Then define the triple integral of $f(x,y,z)$ over S, denoted by $\iiint\limits_S f(x,y,z)\,dV$, by

$$\iiint\limits_S f(x,y,z)\,dV \;=\; \lim \sum\sum\sum f(x_*,y_*,z_*)\,\Delta x\,\Delta y\,\Delta z\,, \tag{4.8}$$

where the limit is over all divisions of the rectangular parallelepiped enclosing S into subparallelepipeds whose largest diagonal is going to 0, and the triple summation is over all the subparallelepipeds inside S. It can be shown that this limit does not depend on the choice of the rectangular parallelepiped enclosing S. The symbol dV is often called the *volume element*.

Physically, what does the triple integral represent? We saw that a double integral could be thought of as the volume under a two-dimensional surface. It turns out that the triple integral simply generalizes this idea: it can be thought of as representing the *hypervolume* under a three-dimensional *hypersurface* $w = f(x,y,z)$ whose graph lies in $\mathbb{R}^4$. In general, the word "volume" is often used as a general term to signify the same concept for any n-dimensional object (e.g. length in $\mathbb{R}^1$, area in $\mathbb{R}^2$). It may be hard to get a grasp on the concept of the "volume" of a four-dimensional object, but at least we now know how to calculate that volume!

In the case where S is a rectangular parallelepiped $[x_1,x_2] \times [y_1,y_2] \times [z_1,z_2]$, that is, $S = \{(x,y,z) : x_1 \leqslant x \leqslant x_2,\ y_1 \leqslant y \leqslant y_2,\ z_1 \leqslant z \leqslant z_2\}$, the triple integral is a sequence of three iterated integrals, namely

$$\iiint\limits_S f(x,y,z)\,dV \;=\; \int_{z_1}^{z_2}\int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x,y,z)\,dx\,dy\,dz\,, \tag{4.9}$$

where the order of integration does not matter. This is the simplest case.

A more complicated case is where S is a solid which is bounded below by a surface $z = g_1(x,y)$, bounded above by a surface $z = g_2(x,y)$, y is bounded between two curves $h_1(x)$ and $h_2(x)$, and x varies between a and b. Then

$$\iiint\limits_S f(x,y,z)\,dV \;=\; \int_a^b\int_{h_1(x)}^{h_2(x)}\int_{g_1(x,y)}^{g_2(x,y)} f(x,y,z)\,dz\,dy\,dx\,. \tag{4.10}$$

Notice in this case that the first iterated integral will result in a function of x and y (since its limits of integration are functions of x and y), which then leaves you with a double integral of a type that we learned how to evaluate in Section 3.2. There are, of course, many variations on this case (for example, changing the roles of the variables x, y, z), so as you can probably tell, triple integrals can be quite tricky. At this point, just learning how to evaluate a triple integral, regardless of what it represents, is the most important thing. We will see some other ways in which triple integrals are used later in the text.

Example 4.4.1 Evaluate $\displaystyle\int_0^3\int_0^2\int_0^1 (xy+z)\,dx\,dy\,dz$.

Solution: ►

$$\begin{aligned}
\int_0^3\int_0^2\int_0^1 (xy+z)\,dx\,dy\,dz \;&=\; \int_0^3\int_0^2 \left(\tfrac{1}{2}x^2y + xz\,\Big|_{x=0}^{x=1}\right)\,dy\,dz\\
&=\; \int_0^3\int_0^2 \left(\tfrac{1}{2}y + z\right)\,dy\,dz\\
&=\; \int_0^3 \left(\tfrac{1}{4}y^2 + yz\,\Big|_{y=0}^{y=2}\right)\,dz\\
&=\; \int_0^3 (1+2z)\,dz\\
&=\; z + z^2\,\Big|_0^3 \;=\; 12
\end{aligned}$$

◄

Example 4.4.2 Evaluate $\int_0^1 \int_0^{1-x} \int_0^{2-x-y} (x+y+z)\,dz\,dy\,dx$.

Solution: ▶

$$\begin{aligned}
\int_0^1 \int_0^{1-x} \int_0^{2-x-y} (x+y+z)\,dz\,dy\,dx &= \int_0^1 \int_0^{1-x} \left((x+y)z + \tfrac{1}{2}z^2 \Big|_{z=0}^{z=2-x-y} \right) dy\,dx \\
&= \int_0^1 \int_0^{1-x} \left((x+y)(2-x-y) + \tfrac{1}{2}(2-x-y)^2 \right) dy\,dx \\
&= \int_0^1 \int_0^{1-x} \left(2 - \tfrac{1}{2}x^2 - xy - \tfrac{1}{2}y^2 \right) dy\,dx \\
&= \int_0^1 \left(2y - \tfrac{1}{2}x^2 y - xy - \tfrac{1}{2}xy^2 - \tfrac{1}{6}y^3 \Big|_{y=0}^{y=1-x} \right) dx \\
&= \int_0^1 \left(\tfrac{11}{6} - 2x + \tfrac{1}{6}x^3 \right) dx \\
&= \tfrac{11}{6}x - x^2 + \tfrac{1}{24}x^4 \Big|_0^1 = \tfrac{7}{8}
\end{aligned}$$

◀ Note that the volume V of a solid in $\mathbb{R}^3$ is given by

$$V = \iiint\limits_S 1\,dV\,. \tag{4.11}$$

Since the function being integrated is the constant 1, then the above triple integral reduces to a double integral of the types that we considered in the previous section if the solid is bounded above by some surface $z = f(x,y)$ and bounded below by the xy-plane $z = 0$. There are many other possibilities. For example, the solid could be bounded below and above by surfaces $z = g_1(x,y)$ and $z = g_2(x,y)$, respectively, with y bounded between two curves $h_1(x)$ and $h_2(x)$, and x varies between a and b. Then

$$V = \iiint\limits_S 1\,dV = \int_a^b \int_{h_1(x)}^{h_2(x)} \int_{g_1(x,y)}^{g_2(x,y)} 1\,dz\,dy\,dx = \int_a^b \int_{h_1(x)}^{h_2(x)} \left(g_2(x,y) - g_1(x,y) \right) dy\,dx$$

just like in equation (4.10). See Exercise 10 for an example.

Exercises

A

For Exercises 1-8, evaluate the given triple integral.

1. $\int_0^3 \int_0^2 \int_0^1 xyz\,dx\,dy\,dz$

2. $\int_0^1 \int_0^x \int_0^y xyz\,dz\,dy\,dx$

3. $\displaystyle\int_0^{\pi}\int_0^{x}\int_0^{xy} x^2 \sin z \, dz\, dy\, dx$

4. $\displaystyle\int_0^{1}\int_0^{z}\int_0^{y} z e^{y^2} \, dx\, dy\, dz$

5. $\displaystyle\int_1^{e}\int_0^{y}\int_0^{1/y} x^2 z \, dx\, dz\, dy$

6. $\displaystyle\int_1^{2}\int_0^{y^2}\int_0^{z^2} yz \, dx\, dz\, dy$

7. $\displaystyle\int_1^{2}\int_2^{4}\int_0^{3} 1 \, dx\, dy\, dz$

8. $\displaystyle\int_0^{1}\int_0^{1-x}\int_0^{1-x-y} 1 \, dz\, dy\, dx$

9. Let M be a constant. Show that $\displaystyle\int_{z_1}^{z_2}\int_{y_1}^{y_2}\int_{x_1}^{x_2} M \, dx\, dy\, dz = M(z_2-z_1)(y_2-y_1)(x_2-x_1)$.

B

10. Find the volume V of the solid S bounded by the three coordinate planes, bounded above by the plane $x+y+z=2$, and bounded below by the plane $z=x+y$.

C

11. Show that $\displaystyle\int_a^{b}\int_a^{z}\int_a^{y} f(x)\, dx\, dy\, dz = \int_a^{b} \tfrac{(b-x)^2}{2} f(x)\, dx$. (*Hint: Think of how changing the order of integration in the triple integral changes the limits of integration.*)

4.5 Change of Variables in Multiple Integrals

Given the difficulty of evaluating multiple integrals, the reader may be wondering if it is possible to simplify those integrals using a suitable substitution for the variables. The answer is yes, though it is a bit more complicated than the substitution method which you learned in single-variable calculus.

Recall that if you are given, for example, the definite integral

$$\int_1^2 x^3 \sqrt{x^2-1}\, dx \,,$$

then you would make the substitution

$$\begin{aligned} u &= x^2 - 1 \;\Rightarrow\; x^2 = u+1 \\ du &= 2x\, dx \end{aligned}$$

which changes the limits of integration

$$\begin{aligned} x = 1 &\;\Rightarrow\; u = 0 \\ x = 2 &\;\Rightarrow\; u = 3 \end{aligned}$$

so that we get

$$\begin{aligned}
\int_1^2 x^3\sqrt{x^2-1}\,dx &= \int_1^2 \tfrac{1}{2}x^2\cdot 2x\sqrt{x^2-1}\,dx\\
&= \int_0^3 \tfrac{1}{2}(u+1)\sqrt{u}\,du\\
&= \tfrac{1}{2}\int_0^3 \left(u^{3/2}+u^{1/2}\right)\,du \quad \text{, which can be easily integrated to give}\\
&= \tfrac{14\sqrt{3}}{5}\,.
\end{aligned}$$

Let us take a different look at what happened when we did that substitution, which will give some motivation for how substitution works in multiple integrals. First, we let $u = x^2-1$. On the interval of integration $[1,2]$, the function $x \mapsto x^2-1$ is strictly increasing (and maps $[1,2]$ onto $[0,3]$) and hence has an inverse function (defined on the interval $[0,3]$). That is, on $[0,3]$ we can define x as a function of u, namely

$$x = g(u) = \sqrt{u+1}\,.$$

Then substituting that expression for x into the function $f(x) = x^3\sqrt{x^2-1}$ gives

$$f(x) = f(g(u)) = (u+1)^{3/2}\sqrt{u}\,,$$

and we see that

$$\begin{aligned}
\frac{dx}{du} = g\,'(u) \;\Rightarrow\; dx &= g\,'(u)\,du\\
dx &= \tfrac{1}{2}(u+1)^{-1/2}\,du\,,
\end{aligned}$$

so since

$$\begin{aligned}
g(0) = 1 &\;\Rightarrow\; 0 = g^{-1}(1)\\
g(3) = 2 &\;\Rightarrow\; 3 = g^{-1}(2)
\end{aligned}$$

then performing the substitution as we did earlier gives

$$\begin{aligned}
\int_1^2 f(x)\,dx &= \int_1^2 x^3\sqrt{x^2-1}\,dx\\
&= \int_0^3 \tfrac{1}{2}(u+1)\sqrt{u}\,du \quad \text{, which can be written as}\\
&= \int_0^3 (u+1)^{3/2}\sqrt{u}\cdot\tfrac{1}{2}(u+1)^{-1/2}\,du \quad \text{, which means}\\
\int_1^2 f(x)\,dx &= \int_{g^{-1}(1)}^{g^{-1}(2)} f(g(u))\,g\,'(u)\,du\,.
\end{aligned}$$

In general, if $x = g(u)$ is a one-to-one, differentiable function from an interval $[c, d]$ (which you can think of as being on the "u-axis") onto an interval $[a, b]$ (on the x-axis), which means that $g'(u) \neq 0$ on the interval (c, d), so that $a = g(c)$ and $b = g(d)$, then $c = g^{-1}(a)$ and $d = g^{-1}(b)$, and

$$\int_a^b f(x)\, dx \;=\; \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))\, g'(u)\, du\,. \tag{4.12}$$

This is called the *change of variable* formula for integrals of single-variable functions, and it is what you were implicitly using when doing integration by substitution. This formula turns out to be a special case of a more general formula which can be used to evaluate multiple integrals. We will state the formulas for double and triple integrals involving real-valued functions of two and three variables, respectively. We will assume that all the functions involved are continuously differentiable and that the regions and solids involved all have "reasonable" boundaries. The proof of the following theorem is beyond the scope of the text.

Theorem 4.5.1 Change of Variables Formula for Multiple Integrals
Let $x = x(u,v)$ and $y = y(u,v)$ define a one-to-one mapping of a region R' in the uv-plane onto a region R in the xy-plane such that the determinant

$$J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \tag{4.13}$$

is never 0 in R'. Then

$$\iint\limits_R f(x,y)\, dA(x,y) = \iint\limits_{R'} f(x(u,v),y(u,v)) \left|J(u,v)\right| \, dA(u,v)\,. \tag{4.14}$$

We use the notation $dA(x,y)$ and $dA(u,v)$ to denote the area element in the (x,y) and (u,v) coordinates, respectively.

Similarly, if $x = x(u,v,w)$, $y = y(u,v,w)$ and $z = z(u,v,w)$ define a one-to-one mapping of a solid S' in uvw-space onto a solid S in xyz-space such that the determinant

$$J(u,v,w) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} \tag{4.15}$$

is never 0 in S', then

$$\iiint\limits_S f(x,y,z)\, dV(x,y,z) = \iiint\limits_{S'} f(x(u,v,w),y(u,v,w),z(u,v,w)) \left|J(u,v,w)\right| \, dV(u,v,w)\,. \tag{4.16}$$

The determinant $J(u,v)$ in formula (4.13) is called the **Jacobian** of x and y with respect to u and v, and is sometimes written as

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)}\,. \tag{4.17}$$

Similarly, the Jacobian $J(u,v,w)$ of three variables is sometimes written as

$$J(u,v,w) = \frac{\partial(x,y,z)}{\partial(u,v,w)}\,. \tag{4.18}$$

Notice that formula (4.14) is saying that $dA(x,y) = \left|J(u,v)\right| \, dA(u,v)$, which you can think of as a two-variable version of the relation $dx = g'(u)\, du$ in the single-variable case.

The following example shows how the change of variables formula is used.

Example 4.5.2 Evaluate $\displaystyle\iint_R e^{\frac{x-y}{x+y}}\,dA$, where $R = \{(x,y) : x \geqslant 0, y \geqslant 0, x+y \leqslant 1\}$.

Solution: ▶ First, note that evaluating this double integral *without* using substitution is probably impossible, at least in a closed form. By looking at the numerator and denominator of the exponent of e, we will try the substitution $u = x-y$ and $v = x+y$. To use the change of variables formula (4.14), we need to write both x and y in terms of u and v. So solving for x and y gives $x = \frac{1}{2}(u+v)$ and $y = \frac{1}{2}(v-u)$. In Figure 4.8 below, we see how the mapping $x = x(u,v) = \frac{1}{2}(u+v)$, $y = y(u,v) = \frac{1}{2}(v-u)$ maps the region R' onto R in a one-to-one manner.

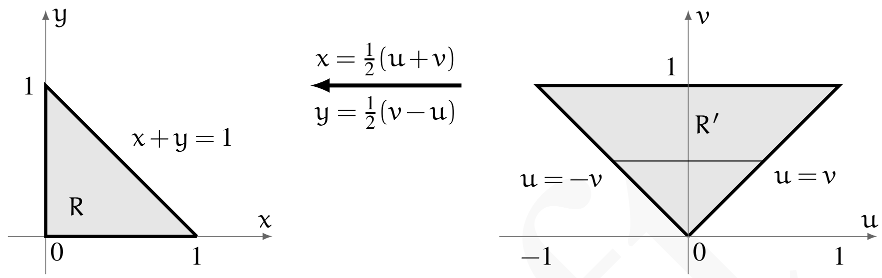


Figure 4.8 The regions R and R'

Now we see that

$$J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{2} \Rightarrow |J(u,v)| = \left|\frac{1}{2}\right| = \frac{1}{2},$$

so using horizontal slices in R', we have

$$\begin{aligned}
\iint_R e^{\frac{x-y}{x+y}}\,dA &= \iint_{R'} f(x(u,v),y(u,v))\,|J(u,v)|\,dA \\
&= \int_0^1 \int_{-v}^{v} e^{\frac{u}{v}}\,\tfrac{1}{2}\,du\,dv \\
&= \int_0^1 \left(\tfrac{v}{2} e^{\frac{u}{v}} \Big|_{u=-v}^{u=v} \right) dv \\
&= \int_0^1 \tfrac{v}{2}(e - e^{-1})\,dv \\
&= \frac{v^2}{4}(e-e^{-1})\Big|_0^1 = \frac{1}{4}\left(e - \frac{1}{e}\right) = \frac{e^2-1}{4e}
\end{aligned}$$

◀ The change of variables formula can be used to evaluate double integrals in polar coordinates. Letting

$$x = x(r,\theta) = r\cos\theta \quad \text{and} \quad y = y(r,\theta) = r\sin\theta\,,$$

we have

$$J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r \Rightarrow \left|J(u,v)\right| = |r| = r\,,$$

so we have the following formula:

Double Integral in Polar Coordinates

$$\iint\limits_R f(x,y)\,dx\,dy = \iint\limits_{R'} f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta\,, \tag{4.19}$$

where the mapping $x = r\cos\theta$, $y = r\sin\theta$ maps the region R' in the $r\theta$-plane onto the region R in the xy-plane in a one-to-one manner.

Example 4.5.3 Find the volume V inside the paraboloid $z = x^2 + y^2$ for $0 \leqslant z \leqslant 1$.

Solution: Using vertical slices, we see that

$$V = \iint\limits_R (1-z)\,dA = \iint\limits_R (1-(x^2+y^2))\,dA\,,$$

where $R = \{(x,y) : x^2+y^2 \leqslant 1\}$ is the unit disk in $\mathbb{R}^2$ (see Figure 3.5.2). In polar coordinates (r,θ) we know that $x^2 + y^2 = r^2$ and that the unit disk R is the set $R' = \{(r,\theta) : 0 \leqslant r \leqslant 1, 0 \leqslant \theta \leqslant 2\pi\}$. Thus,

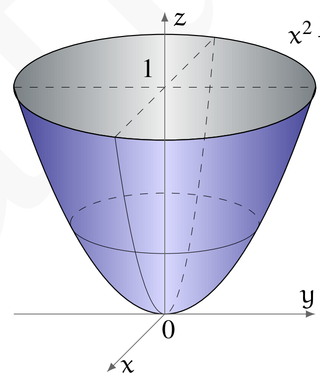


Figure 4.9 $z = x^2 + y^2$

$$\begin{aligned} V &= \int_0^{2\pi}\int_0^1 (1-r^2)\,r\,dr\,d\theta \\ &= \int_0^{2\pi}\int_0^1 (r-r^3)\,dr\,d\theta \\ &= \int_0^{2\pi} \left(\frac{r^2}{2} - \frac{r^4}{4} \Big|_{r=0}^{r=1} \right) d\theta \\ &= \int_0^{2\pi} \tfrac{1}{4}\,d\theta \\ &= \frac{\pi}{2} \end{aligned}$$

Example 4.5.4 Find the volume V inside the cone $z = \sqrt{x^2+y^2}$ for $0 \leqslant z \leqslant 1$.

Solution: Using vertical slices, we see that

$$V = \iint\limits_R (1-z)\,dA = \iint\limits_R \left(1-\sqrt{x^2+y^2}\right)\,dA\,,$$

where $R = \{(x,y) : x^2+y^2 \leqslant 1\}$ is the unit disk in $\mathbb{R}^2$ (see Figure 3.5.3). In polar coordinates (r,θ) we know that $\sqrt{x^2+y^2} = r$ and that the unit disk R is the set $R' = \{(r,\theta) : 0 \leqslant r \leqslant 1, 0 \leqslant \theta \leqslant 2\pi\}$. Thus,

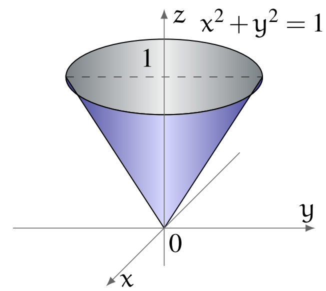


Figure 4.10 $z = \sqrt{x^2+y^2}$

$$\begin{aligned}
V &= \int_0^{2\pi}\int_0^1 (1-r)\,r\,dr\,d\theta \\
&= \int_0^{2\pi}\int_0^1 (r-r^2)\,dr\,d\theta \\
&= \int_0^{2\pi} \left(\frac{r^2}{2}-\frac{r^3}{3}\Bigg|_{r=0}^{r=1}\right)d\theta \\
&= \int_0^{2\pi} \tfrac{1}{6}\,d\theta \\
&= \frac{\pi}{3}
\end{aligned}$$

In a similar fashion, it can be shown (see Exercises 5-6) that triple integrals in cylindrical and spherical coordinates take the following forms:

Triple Integral in Cylindrical Coordinates

$$\iiint\limits_S f(x,y,z)\,dx\,dy\,dz = \iiint\limits_{S'} f(r\cos\theta, r\sin\theta, z)\,r\,dr\,d\theta\,dz\,, \tag{4.20}$$

where the mapping $x = r\cos\theta$, $y = r\sin\theta$, $z = z$ maps the solid S' in $r\theta z$-space onto the solid S in xyz-space in a one-to-one manner.

Triple Integral in Spherical Coordinates

$$\iiint\limits_S f(x,y,z)\,dx\,dy\,dz = \iiint\limits_{S'} f(\rho\sin\phi\,\cos\theta, \rho\sin\phi\,\sin\theta, \rho\cos\phi)\,\rho^2\,\sin\phi\,d\rho\,d\phi\,d\theta\,, \tag{4.21}$$

where the mapping $x = \rho\sin\phi\,\cos\theta$, $y = \rho\sin\phi\,\sin\theta$, $z = \rho\cos\phi$ maps the solid S' in $\rho\phi\theta$-space onto the solid S in xyz-space in a one-to-one manner.

Example 4.5.5 For $a > 0$, find the volume V inside the sphere $S = x^2+y^2+z^2 = a^2$.

Solution: We see that S is the set $\rho = a$ in spherical coordinates, so

$$\begin{aligned}
V &= \iiint\limits_S 1\,dV = \int_0^{2\pi}\int_0^{\pi}\int_0^{a} 1\,\rho^2\sin\phi\,d\rho\,d\phi\,d\theta \\
&= \int_0^{2\pi}\int_0^{\pi}\left(\frac{\rho^3}{3}\Big|_{\rho=0}^{\rho=a}\right)\sin\phi\,d\phi\,d\theta = \int_0^{2\pi}\int_0^{\pi}\frac{a^3}{3}\sin\phi\,d\phi\,d\theta \\
&= \int_0^{2\pi}\left(-\frac{a^3}{3}\cos\phi\Big|_{\phi=0}^{\phi=\pi}\right)d\theta = \int_0^{2\pi}\frac{2a^3}{3}\,d\theta = \frac{4\pi a^3}{3}\,.
\end{aligned}$$

Exercises

A

1. Find the volume V inside the paraboloid $z = x^2 + y^2$ for $0 \leqslant z \leqslant 4$.

2. Find the volume V inside the cone $z = \sqrt{x^2 + y^2}$ for $0 \leqslant z \leqslant 3$.

B

3. Find the volume V of the solid inside both $x^2 + y^2 + z^2 = 4$ and $x^2 + y^2 = 1$.

4. Find the volume V inside both the sphere $x^2 + y^2 + z^2 = 1$ and the cone $z = \sqrt{x^2 + y^2}$.

5. Prove formula (4.20).

6. Prove formula (4.21).

7. Evaluate $\iint\limits_R \sin\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)dA$, where R is the triangle with vertices $(0,0)$, $(2,0)$ and $(1,1)$. (*Hint: Use the change of variables* $u = (x+y)/2$, $v = (x-y)/2$.)

8. Find the volume of the solid bounded by $z = x^2 + y^2$ and $z^2 = 4(x^2 + y^2)$.

9. Find the volume inside the elliptic cylinder $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ for $0 \leqslant z \leqslant 2$.

C

10. Show that the volume inside the ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ is $\frac{4\pi abc}{3}$. (*Hint: Use the change of variables* $x = au$, $y = bv$, $z = cw$, *then consider Example 4.5.5.*)

11. Show that the *Beta function*, defined by

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt\,, \quad \text{for } x > 0,\ y > 0,$$

satisfies the relation $B(y,x) = B(x,y)$ for $x > 0$, $y > 0$.

12. Using the substitution $t = u/(u+1)$, show that the Beta function can be written as

$$B(x,y) = \int_0^{\infty}\frac{u^{x-1}}{(u+1)^{x+y}}\,du\,, \quad \text{for } x > 0,\ y > 0.$$

4.6 Application: Center of Mass

Recall from single-variable calculus that for a region $R = \{(x,y) : a \leqslant x \leqslant b, 0 \leqslant y \leqslant f(x)\}$ in $\mathbb{R}^2$ that represents a thin, flat plate (see Figure 3.6.1), where $f(x)$ is a continuous function on $[a,b]$, the *center of mass* of R has coordinates $(\bar{x},\bar{y})$ given by

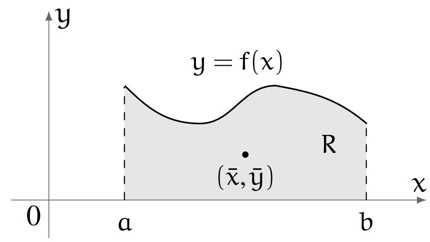


Figure 4.11 Center of mass of R

$$\bar{x} = \frac{M_y}{M} \quad \text{and} \quad \bar{y} = \frac{M_x}{M} ,$$

where

$$M_x = \int_a^b \frac{(f(x))^2}{2}\, dx , \quad M_y = \int_a^b x f(x)\, dx , \quad M = \int_a^b f(x)\, dx , \tag{4.22}$$

assuming that R has *uniform density*, i.e the *mass* of R is uniformly distributed over the region. In this case the area M of the region is considered the mass of R (the density is constant, and taken as 1 for simplicity).

In the general case where the density of a region (or *lamina*) R is a continuous function $\delta = \delta(x,y)$ of the coordinates (x,y) of points inside R (where R can be *any* region in $\mathbb{R}^2$) the coordinates $(\bar{x},\bar{y})$ of the center of mass of R are given by

$$\bar{x} = \frac{M_y}{M} \quad \text{and} \quad \bar{y} = \frac{M_x}{M} , \tag{4.23}$$

where

$$M_y = \iint\limits_R x\delta(x,y)\, dA , \quad M_x = \iint\limits_R y\delta(x,y)\, dA , \quad M = \iint\limits_R \delta(x,y)\, dA , \tag{4.24}$$

The quantities M_x and M_y are called the *moments* (or *first moments*) of the region R about the x-axis and y-axis, respectively. The quantity M is the mass of the region R. To see this, think of taking a small rectangle inside R with dimensions Δx and Δy close to 0. The mass of that rectangle is approximately $\delta(x_*,y_*)\Delta x\,\Delta y$, for some point (x_*,y_*) in that rectangle. Then the mass of R is the limit of the sums of the masses of all such rectangles inside R as the diagonals of the rectangles approach 0, which is the double integral $\iint\limits_R \delta(x,y)\, dA$.

Note that the formulas in (4.22) represent a special case when $\delta(x,y) = 1$ throughout R in the formulas in (4.24).

Example 4.6.1 Find the center of mass of the region $R = \{(x,y) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 2x^2\}$, if the density function at (x,y) is $\delta(x,y) = x+y$.

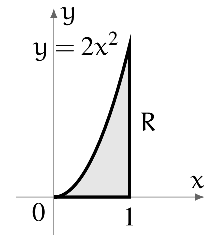


Figure 4.12

Solution: ► The region R is shown in Figure 3.6.2. We have

$$\begin{aligned} M &= \iint_R \delta(x,y)\,dA \\ &= \int_0^1 \int_0^{2x^2} (x+y)\,dy\,dx \\ &= \int_0^1 \left(xy + \frac{y^2}{2} \Bigg|_{y=0}^{y=2x^2} \right) dx \\ &= \int_0^1 (2x^3+2x^4)\,dx \\ &= \frac{x^4}{2} + \frac{2x^5}{5} \Bigg|_0^1 = \frac{9}{10} \end{aligned}$$

and

$$\begin{aligned} M_x &= \iint_R y\delta(x,y)\,dA \\ &= \int_0^1 \int_0^{2x^2} y(x+y)\,dy\,dx \\ &= \int_0^1 \left(\frac{xy^2}{2} + \frac{y^3}{3} \Bigg|_{y=0}^{y=2x^2} \right) dx \\ &= \int_0^1 (2x^5 + \frac{8x^6}{3})\,dx \\ &= \frac{x^6}{3} + \frac{8x^7}{21} \Bigg|_0^1 = \frac{5}{7} \end{aligned}$$

$$\begin{aligned} M_y &= \iint_R x\delta(x,y)\,dA \\ &= \int_0^1 \int_0^{2x^2} x(x+y)\,dy\,dx \\ &= \int_0^1 \left(x^2y + \frac{xy^2}{2} \Bigg|_{y=0}^{y=2x^2} \right) dx \\ &= \int_0^1 (2x^4+2x^5)\,dx \\ &= \frac{2x^5}{5} + \frac{x^6}{3} \Bigg|_0^1 = \frac{11}{15}, \end{aligned}$$

so the center of mass $(\bar{x},\bar{y})$ is given by

$$\bar{x} = \frac{M_y}{M} = \frac{11/15}{9/10} = \frac{22}{27}, \quad \bar{y} = \frac{M_x}{M} = \frac{5/7}{9/10} = \frac{50}{63}.$$

Note how this center of mass is a little further towards the upper corner of the region R than when the density is uniform (it is easy to use the formulas in (4.22) to show that $(\bar{x},\bar{y}) = \left(\frac{3}{4},\frac{3}{5}\right)$ in that case). This makes sense since the density function $\delta(x,y) = x+y$ increases as (x,y) approaches that upper corner, where there is quite a bit of area. ◄

In the special case where the density function $\delta(x,y)$ is a constant function on the region R, the center of mass $(\bar{x},\bar{y})$ is called the *centroid* of R.

The formulas for the center of mass of a region in $\mathbb{R}^2$ can be generalized to a solid S in $\mathbb{R}^3$. Let S be a solid with a continuous mass density function $\delta(x,y,z)$ at any point (x,y,z) in S. Then the center of mass of S has coordinates $(\bar{x},\bar{y},\bar{z})$, where

$$\bar{x} = \frac{M_{yz}}{M}\,, \quad \bar{y} = \frac{M_{xz}}{M}\,, \quad \bar{z} = \frac{M_{xy}}{M}\,, \tag{4.25}$$

where

$$M_{yz} = \iiint\limits_S x\delta(x,y,z)\,dV\,, \quad M_{xz} = \iiint\limits_S y\delta(x,y,z)\,dV\,, \quad M_{xy} = \iiint\limits_S z\delta(x,y,z)\,dV\,, \tag{4.26}$$

$$M = \iiint\limits_S \delta(x,y,z)\,dV\,. \tag{4.27}$$

In this case, M_{yz}, M_{xz} and M_{xy} are called the *moments* (or *first moments*) of S around the yz-plane, xz-plane and xy-plane, respectively. Also, M is the mass of S.

Example 4.6.2 Find the center of mass of the solid $S = \{(x,y,z) : z \geqslant 0,\ x^2+y^2+z^2 \leqslant a^2\}$, if the density function at (x,y,z) is $\delta(x,y,z) = 1$.

Solution: ► The solid S is just the upper hemisphere inside the sphere of radius a centered at the origin (see Figure 3.6.3). So since the density function is a constant and S is symmetric about the z-axis, then it is clear that $\bar{x} = 0$ and $\bar{y} = 0$, so we need only find $\bar{z}$. We have

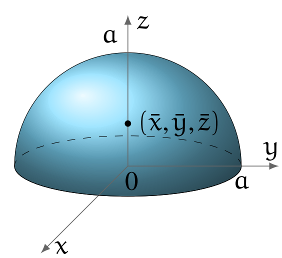


Figure 4.13

$$M = \iiint\limits_S \delta(x,y,z)\,dV = \iiint\limits_S 1\,dV = \mathrm{Volume}(S).$$

But since the volume of S is half the volume of the sphere of radius a, which we know by Example 4.5.5 is $\frac{4\pi a^3}{3}$, then $M = \frac{2\pi a^3}{3}$. And

$$\begin{aligned}
M_{xy} &= \iiint\limits_S z\delta(x,y,z)\,dV \\
&= \iiint\limits_S z\,dV \text{ , which in spherical coordinates is} \\
&= \int_0^{2\pi}\int_0^{\pi/2}\int_0^a (\rho\cos\phi)\,\rho^2\sin\phi\,d\rho\,d\phi\,d\theta \\
&= \int_0^{2\pi}\int_0^{\pi/2} \sin\phi\cos\phi\left(\int_0^a \rho^3\,d\rho\right)d\phi\,d\theta \\
&= \int_0^{2\pi}\int_0^{\pi/2} \frac{a^4}{4}\sin\phi\cos\phi\,d\phi\,d\theta
\end{aligned}$$

$$\begin{aligned} M_{xy} &= \int_0^{2\pi}\int_0^{\pi/2} \tfrac{a^4}{8}\sin 2\phi \, d\phi \, d\theta \quad (\text{since } \sin 2\phi = 2\sin\phi\cos\phi) \\ &= \int_0^{2\pi}\left(-\tfrac{a^4}{16}\cos 2\phi \,\Big|_{\phi=0}^{\phi=\pi/2} \right) d\theta \\ &= \int_0^{2\pi} \tfrac{a^4}{8}\, d\theta \\ &= \frac{\pi a^4}{4}, \end{aligned}$$

$$\bar{z} = \frac{M_{xy}}{M} = \frac{\frac{\pi a^4}{4}}{\frac{2\pi a^3}{3}} = \frac{3a}{8}.$$

Thus, the center of mass of S is $(\bar{x},\bar{y},\bar{z}) = \left(0,0,\frac{3a}{8}\right)$. ◀

Exercises

A

For Exercises 1-5, find the center of mass of the region R with the given density function $\delta(x,y)$.

1. $R = \{(x,y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant 4\}$, $\delta(x,y) = 2y$

2. $R = \{(x,y) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant x^2\}$, $\delta(x,y) = x+y$

3. $R = \{(x,y) : y \geqslant 0,\ x^2+y^2 \leqslant a^2\}$, $\delta(x,y) = 1$

4. $R = \{(x,y) : y \geqslant 0,\ x \geqslant 0,\ 1 \leqslant x^2+y^2 \leqslant 4\}$, $\delta(x,y) = \sqrt{x^2+y^2}$

5. $R = \{(x,y) : y \geqslant 0,\ x^2+y^2 \leqslant 1\}$, $\delta(x,y) = y$

B

For Exercises 6-10, find the center of mass of the solid S with the given density function $\delta(x,y,z)$.

6. $S = \{(x,y,z) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1\}$, $\delta(x,y,z) = xyz$

7. $S = \{(x,y,z) : z \geqslant 0,\ x^2+y^2+z^2 \leqslant a^2\}$, $\delta(x,y,z) = x^2+y^2+z^2$

8. $S = \{(x,y,z) : x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0,\ x^2+y^2+z^2 \leqslant a^2\}$, $\delta(x,y,z) = 1$

9. $S = \{(x,y,z) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1\}$, $\delta(x,y,z) = x^2+y^2+z^2$

10. $S = \{(x,y,z) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1-x-y\}$, $\delta(x,y,z) = 1$

4.7 Application: Probability and Expected Value

In this section we will briefly discuss some applications of multiple integrals in the field of probability theory. In particular we will see ways in which multiple integrals can be used to calculate *probabilities* and *expected values*.

Probability

Suppose that you have a standard six-sided (fair) die, and you let a variable X represent the value rolled. Then the *probability* of rolling a 3, written as $P(X=3)$, is $\frac{1}{6}$, since there are six sides on the die and each one is equally likely to be rolled, and hence in particular the 3 has a one out of six chance of being rolled. Likewise the probability of rolling *at most* a 3, written as $P(X \leqslant 3)$, is $\frac{3}{6} = \frac{1}{2}$, since of the six numbers on the die, there are three equally likely numbers (1, 2, and 3) that are less than or equal to 3. Note that $P(X \leqslant 3) = P(X=1) + P(X=2) + P(X=3)$. We call X a *discrete random variable* on the *sample space* (or *probability space*) Ω consisting of all possible outcomes. In our case, $\Omega = \{1,2,3,4,5,6\}$. An *event* A is a subset of the sample space. For example, in the case of the die, the event $X \leqslant 3$ is the set $\{1,2,3\}$.

Now let X be a variable representing a random real number in the interval $(0,1)$. Note that the set of all real numbers between 0 and 1 is *not* a discrete (or *countable*) set of values, i.e. it can not be put into a one-to-one correspondence with the set of positive integers.[2] In this case, for any real number x in $(0,1)$, it makes no sense to consider $P(X=x)$ since it *must* be 0 (why?). Instead, we consider the probability $P(X \leqslant x)$, which is given by $P(X \leqslant x) = x$. The reasoning is this: the interval $(0,1)$ has length 1, and for x in $(0,1)$ the interval $(0,x)$ has length x. So since X represents a *random* number in $(0,1)$, and hence is *uniformly distributed* over $(0,1)$, then

$$P(X \leqslant x) = \frac{\text{length of } (0,x)}{\text{length of } (0,1)} = \frac{x}{1} = x\,.$$

We call X a *continuous random variable* on the *sample space* $\Omega = (0,1)$. An *event* A is a subset of the sample space. For example, in our case the event $X \leqslant x$ is the set $(0,x)$.

In the case of a discrete random variable, we saw how the probability of an event was the *sum* of the probabilities of the individual outcomes comprising that event (e.g. $P(X \leqslant 3) = P(X = 1) + P(X=2) + P(X=3)$ in the die example). For a continuous random variable, the probability of an event will instead be the *integral* of a function, which we will now describe.

Let X be a continuous real-valued random variable on a sample space Ω in $\mathbb{R}$. For simplicity,

[2] For a proof see p. 9-10 in **kam**

let $\Omega = (a, b)$. Define the *distribution function* F of X as

$$F(x) = P(X \leqslant x), \quad \text{for } -\infty < x < \infty \tag{4.28}$$

$$= \begin{cases} 1, & \text{for } x \geqslant b \\ P(X \leqslant x), & \text{for } a < x < b \\ 0, & \text{for } x \leqslant a\,. \end{cases} \tag{4.29}$$

Suppose that there is a nonnegative, continuous real-valued function f on $\mathbb{R}$ such that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy\,, \quad \text{for } -\infty < x < \infty\,, \tag{4.30}$$

and

$$\int_{-\infty}^{\infty} f(x)\,dx = 1\,. \tag{4.31}$$

Then we call f the *probability density function* (or *p.d.f.* for short) for X. We thus have

$$P(X \leqslant x) = \int_{a}^{x} f(y)\,dy\,, \quad \text{for } a < x < b\,. \tag{4.32}$$

Also, by the Fundamental Theorem of Calculus, we have

$$F'(x) = f(x)\,, \quad \text{for } -\infty < x < \infty. \tag{4.33}$$

Example 4.7.1 Let X represent a randomly selected real number in the interval $(0, 1)$. We say that X has the *uniform distribution* on $(0, 1)$, with distribution function

$$F(x) = P(X \leqslant x) = \begin{cases} 1, & \text{for } x \geqslant 1 \\ x, & \text{for } 0 < x < 1 \\ 0, & \text{for } x \leqslant 0\,, \end{cases} \tag{4.34}$$

and probability density function

$$f(x) = F'(x) = \begin{cases} 1, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere.} \end{cases} \tag{4.35}$$

In general, if X represents a randomly selected real number in an interval (a, b), then X has the uniform distribution function

$$F(x) = P(X \leqslant x) = \begin{cases} 1, & \text{for } x \geqslant b \\ \frac{x}{b-a}, & \text{for } a < x < b \\ 0, & \text{for } x \leqslant a\,, \end{cases} \tag{4.36}$$

and probability density function

$$f(x) = F'(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a < x < b \\ 0, & \text{elsewhere.} \end{cases} \tag{4.37}$$

Example 4.7.2 A famous distribution function is given by the *standard normal distribution*, whose probability density function f is

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,, \quad \text{for } -\infty < x < \infty. \tag{4.38}$$

This is often called a "bell curve", and is used widely in statistics. Since we are claiming that f is a p.d.f., we *should* have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx \;=\; 1 \tag{4.39}$$

by formula (4.31), which is equivalent to

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx \;=\; \sqrt{2\pi}\,. \tag{4.40}$$

We can use a double integral in polar coordinates to verify this integral. First,

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy \;&=\; \int_{-\infty}^{\infty} e^{-y^2/2}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)\,dy\\
&=\; \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right)\\
&=\; \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)^2
\end{aligned}$$

since the same function is being integrated twice in the middle equation, just with different variables. But using polar coordinates, we see that

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy \;&=\; \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\,r\,dr\,d\theta\\
&=\; \int_0^{2\pi}\left(-e^{-r^2/2}\,\Bigg|_{r=0}^{r=\infty}\right)\,d\theta\\
&=\; \int_0^{2\pi}(0-(-e^0))\,d\theta = \int_0^{2\pi} 1\,d\theta \;=\; 2\pi\,,
\end{aligned}$$

and so

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right)^2 \;=\; 2\pi \;\;, \text{ and hence}$$

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx \;=\; \sqrt{2\pi}\,.$$

In addition to individual random variables, we can consider *jointly distributed* random variables. For this, we will let X, Y and Z be three real-valued continuous random variables defined on the same sample space Ω in $\mathbb{R}$ (the discussion for two random variables is similar). Then the *joint distribution function* F of X, Y and Z is given by

$$F(x,y,z) \;=\; P(X \leqslant x,\; Y \leqslant y,\; Z \leqslant z)\,, \quad \text{for } -\infty < x,y,z < \infty. \tag{4.41}$$

If there is a nonnegative, continuous real-valued function f on $\mathbb{R}^3$ such that

$$F(x,y,z) \;=\; \int_{-\infty}^{z}\int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v,w)\,du\,dv\,dw\,, \quad \text{for } -\infty < x,y,z < \infty \tag{4.42}$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\,dx\,dy\,dz \;=\; 1\,, \tag{4.43}$$

then we call f the *joint probability density function* (or *joint p.d.f.* for short) for X, Y and Z. In general, for $a_1 < b_1$, $a_2 < b_2$, $a_3 < b_3$, we have

$$P(a_1 < X \leqslant b_1,\; a_2 < Y \leqslant b_2,\; a_3 < Z \leqslant b_3) \;=\; \int_{a_3}^{b_3}\int_{a_2}^{b_2}\int_{a_1}^{b_1} f(x,y,z)\,dx\,dy\,dz\,, \tag{4.44}$$

with the $\leqslant$ and $<$ symbols interchangeable in any combination. A triple integral, then, can be thought of as representing a probability (for a function f which is a p.d.f.).

Example 4.7.3 Let a, b, and c be real numbers selected randomly from the interval $(0,1)$. What is the probability that the equation $ax^2 + bx + c = 0$ has at least one real solution x?

Solution: ▶ We know by the quadratic formula that there is at least one real solution if $b^2 - 4ac \geqslant 0$. So we need to calculate $P(b^2 - 4ac \geqslant 0)$. We will use three jointly distributed random variables to do this. First, since $0 < a,b,c < 1$, we have

$$b^2 - 4ac \;\geqslant\; 0 \;\Leftrightarrow\; 0 \;<\; 4ac \;\leqslant\; b^2 \;<\; 1 \;\Leftrightarrow\; 0 \;<\; 2\sqrt{a}\sqrt{c} \;\leqslant\; b \;<\; 1\,,$$

where the last relation holds for all $0 < a,c < 1$ such that

$$0 \;<\; 4ac \;<\; 1 \;\Leftrightarrow\; 0 \;<\; c \;<\; \frac{1}{4a}\,.$$

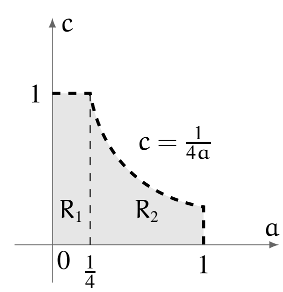


Figure 4.14 Region $R = R_1 \cup R_2$

Considering a, b and c as real variables, the region R in the ac-plane where the above relation holds is given by $R = \{(a,c) : 0 < a < 1,\; 0 < c < 1,\; 0 < c < \frac{1}{4a}\}$, which we can see is a union of two regions R_1 and R_2, as in Figure 3.7.1 above.

Now let X, Y and Z be continuous random variables, each representing a randomly selected real number from the interval $(0,1)$ (think of X, Y and Z representing a, b and c, respectively). Then, similar to how we showed that $f(x) = 1$ is the p.d.f. of the uniform distribution on $(0,1)$, it can be

shown that $f(x,y,z) = 1$ for x, y, z in $(0,1)$
(0 elsewhere) is the joint p.d.f. of X, Y and Z. Now,

$$P(b^2 - 4ac \geqslant 0) \;=\; P((a,c) \in R,\; 2\sqrt{a}\sqrt{c} \leqslant b < 1)\;,$$

so this probability is the triple integral of $f(a,b,c) = 1$ as b varies from $2\sqrt{a}\sqrt{c}$ to 1 and as (a,c) varies over the region R. Since R can be divided into two regions R_1 and R_2, then the required triple integral can be split into a sum of two triple integrals, using vertical slices in R:

$$\begin{aligned}
P(b^2 - 4ac \geqslant 0) \;&=\; \underbrace{\int_0^{1/4}\int_0^1}_{R_1}\int_{2\sqrt{a}\sqrt{c}}^1 1\,db\,dc\,da \;+\; \underbrace{\int_{1/4}^1\int_0^{1/4a}}_{R_2}\int_{2\sqrt{a}\sqrt{c}}^1 1\,db\,dc\,da\\
&=\; \int_0^{1/4}\int_0^1 (1 - 2\sqrt{a}\sqrt{c})\,dc\,da \;+\; \int_{1/4}^1\int_0^{1/4a} (1 - 2\sqrt{a}\sqrt{c})\,dc\,da\\
&=\; \int_0^{1/4}\left(c - \tfrac{4}{3}\sqrt{a}\,c^{3/2}\Big|_{c=0}^{c=1}\right)da \;+\; \int_{1/4}^1\left(c - \tfrac{4}{3}\sqrt{a}\,c^{3/2}\Big|_{c=0}^{c=1/4a}\right)da\\
&=\; \int_0^{1/4}\left(1 - \tfrac{4}{3}\sqrt{a}\right)da \;+\; \int_{1/4}^1 \tfrac{1}{12a}\,da\\
&=\; a - \frac{8}{9}a^{3/2}\Big|_0^{1/4} \;+\; \frac{1}{12}\ln a\Big|_{1/4}^1\\
&=\; \left(\frac{1}{4} - \frac{1}{9}\right) \;+\; \left(0 - \frac{1}{12}\ln\frac{1}{4}\right) \;=\; \frac{5}{36} + \frac{1}{12}\ln 4\\
P(b^2 - 4ac \geqslant 0) \;&=\; \frac{5 + 3\ln 4}{36} \;\approx\; 0.2544
\end{aligned}$$

In other words, the equation $ax^2 + bx + c = 0$ has about a 25% chance of being solved! ◀

Expected Value

The *expected value* EX of a random variable X can be thought of as the "average" value of X as it varies over its sample space. If X is a discrete random variable, then

$$EX \;=\; \sum_x x\,P(X = x)\;, \tag{4.45}$$

with the sum being taken over all elements x of the sample space. For example, if X represents the number rolled on a six-sided die, then

$$EX \;=\; \sum_{x=1}^{6} x\,P(X = x) \;=\; \sum_{x=1}^{6} x\,\frac{1}{6} \;=\; 3.5 \tag{4.46}$$

is the expected value of X, which is the average of the integers $1 - 6$.

If X is a real-valued continuous random variable with p.d.f. f, then

$$EX = \int_{-\infty}^{\infty} x\,f(x)\,dx\,. \tag{4.47}$$

For example, if X has the uniform distribution on the interval $(0,1)$, then its p.d.f. is

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere,} \end{cases} \tag{4.48}$$

and so

$$EX = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^1 x\,dx = \frac{1}{2}\,. \tag{4.49}$$

For a pair of jointly distributed, real-valued continuous random variables X and Y with joint p.d.f. $f(x,y)$, the expected values of X and Y are given by

$$EX = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,f(x,y)\,dx\,dy \quad \text{and} \quad EY = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,f(x,y)\,dx\,dy\,, \tag{4.50}$$

respectively.

Example 4.7.4 If you were to pick $n > 2$ random real numbers from the interval $(0,1)$, what are the expected values for the smallest and largest of those numbers?

Solution: ▶ Let $U_1,\ldots,U_n$ be n continuous random variables, each representing a randomly selected real number from $(0,1)$, i.e. each has the uniform distribution on $(0,1)$. Define random variables X and Y by

$$X = \min(U_1,\ldots,U_n) \quad \text{and} \quad Y = \max(U_1,\ldots,U_n)\,.$$

Then it can be shown[3] that the joint p.d.f. of X and Y is

$$f(x,y) = \begin{cases} n(n-1)(y-x)^{n-2}, & \text{for } 0 \leqslant x \leqslant y \leqslant 1 \\ 0, & \text{elsewhere.} \end{cases} \tag{4.51}$$

Thus, the expected value of X is

$$\begin{aligned} EX &= \int_0^1 \int_x^1 n(n-1)x(y-x)^{n-2}\,dy\,dx \\ &= \int_0^1 \left(nx(y-x)^{n-1} \Big|_{y=x}^{y=1} \right) dx \\ &= \int_0^1 nx(1-x)^{n-1}\,dx \ \text{, so integration by parts yields} \\ &= -x(1-x)^n - \frac{1}{n+1}(1-x)^{n+1} \Big|_0^1 \\ EX &= \frac{1}{n+1}\,, \end{aligned}$$

[3] See Ch. 6 in [33].

and similarly (see Exercise 3) it can be shown that

$$\mathrm{E}Y = \int_0^1 \int_0^y n(n-1)y(y-x)^{n-2}\,dx\,dy = \frac{n}{n+1}.$$

So, for example, if you were to repeatedly take samples of $n = 3$ random real numbers from $(0,1)$, and each time store the minimum and maximum values in the sample, then the average of the minimums would approach $\frac{1}{4}$ and the average of the maximums would approach $\frac{3}{4}$ as the number of samples grows. It would be relatively simple (see Exercise 4) to write a computer program to test this. ◀

Exercises

B

1. Evaluate the integral $\int_{-\infty}^{\infty} e^{-x^2}\,dx$ using anything you have learned so far.

2. For $\sigma > 0$ and $\mu > 0$, evaluate $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\,dx$.

3. Show that $\mathrm{E}Y = \frac{n}{n+1}$ in Example 4.7.4

C

4. Write a computer program (in the language of your choice) that verifies the results in Example 4.7.4 for the case $n = 3$ by taking large numbers of samples.

5. Repeat Exercise 4 for the case when $n = 4$.

6. For continuous random variables X, Y with joint p.d.f. $f(x,y)$, define the *second moments* $\mathrm{E}(X^2)$ and $\mathrm{E}(Y^2)$ by

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy \quad\text{and}\quad \mathrm{E}(Y^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)\,dx\,dy\,,$$

and the *variances* $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ by

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2 \quad\text{and}\quad \mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}Y)^2\,.$$

Find $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ for X and Y as in Example 4.7.4.

7. Continuing Exercise 6, the *correlation* ρ between X and Y is defined as

$$\rho = \frac{\mathrm{E}(XY) - (\mathrm{E}X)(\mathrm{E}Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}\,,$$

where $\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x,y)\,dx\,dy$. Find ρ for X and Y as in Example 4.7.4.
(Note: The quantity $\mathrm{E}(XY) - (\mathrm{E}X)(\mathrm{E}Y)$ is called the *covariance* of X and Y.)

8. In Example 4.7.3 would the answer change if the interval $(0,100)$ is used instead of $(0,1)$? Explain.

Chapter 5

Curves and Surfaces

5.1 Parametric Curves

There are many ways we can described a curve. We can, say, describe it by a equation that the points on the curve satisfy. For example, a circle can be described by $x^2 + y^2 = 1$. However, this is not a good way to do so, as it is rather difficult to work with. It is also often difficult to find a closed form like this for a curve.

Instead, we can imagine the curve to be specified by a particle moving along the path. So it is represented by a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$, and the curve itself is the image of the function. This is known as a **parametrization** of a curve. In addition to simplified notation, this also has the benefit of giving the curve an **orientation**.

Definition 5.1.1 We say $\Gamma \subseteq \mathbb{R}^n$ is a differentiable **curve** if exists a differentiable function $\gamma : I = [a,b] \to \mathbb{R}^n$ such that $\Gamma = \gamma([a,b])$.

The function γ is said a parametrization of the curve γ. And the function $\gamma : I = [a,b] \to \mathbb{R}^n$ is said a parametric curve.

Sometimes $\Gamma = \gamma[I] \subseteq \mathbb{R}^n$ is called the **image** of the parametric curve. We note that a curve $\mathbb{R}^n$ can be the image of several distinct parametric curves.

R Usually we will denote the image of the curve and its parametrization by the same letter and we will talk about the curve γ with parametrization $\gamma(t)$.

Definition 5.1.2 A parametrization $\gamma(t) : I \to \mathbb{R}^n$ is **regular** if $\gamma'(t) \neq 0$ for all $t \in I$.

The parametrization provide the curve with an orientation. Since $\gamma = \gamma([a,b])$, we can think the curve as the trace of a motion that starts at $\gamma(a)$ and ends on $\gamma(b)$.

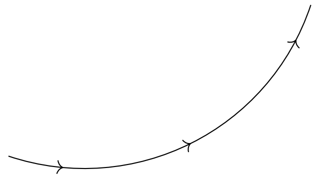


Figure 5.1 Orientation of a Curve

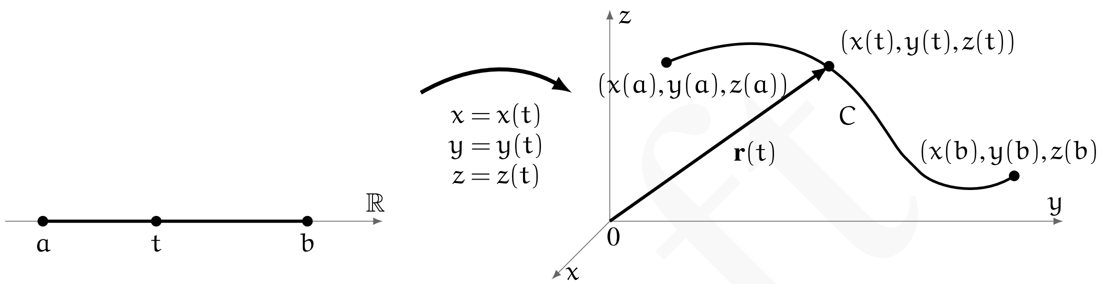


Figure 5.2 Parametrization of a curve C in $\mathbb{R}^3$

Example 5.1.1 The curve $x^2+y^2 = 1$ can be parametrized by $\gamma(t) = (\cos t, \sin t)$ for $t \in [0,2\pi]$

Given a parametric curve **fl** : $I = [a,b] \to \mathbb{R}^n$

- The curve is said to be **simple** if γ is injective, i.e. if for all x, y in (a,b), we have $\gamma(x) = \gamma(y)$ implies $x = y$.
- If $\gamma(x) = \gamma(y)$ for some $x \neq y$ in (a,b), then $\gamma(x)$ is called a **multiple point** of the curve.
- A curve γ is said to be **closed** if $\gamma(a) = \gamma(b)$.

Theorem 5.1.2 — Jordan Curve Theorem. Let γ be a simple closed curve in the plane $\mathbb{R}^2$. Then its complement, $\mathbb{R}^2 \setminus \gamma$, consists of exactly two connected components. One of these components is bounded (the interior) and the other is unbounded (the exterior), and the curve γ is the boundary of each component.

The Jordan Curve Theorem asserts that every simple closed curve in the plane curve divides the plane into an "interior" region bounded by the curve and an "exterior" region. While the statement of this theorem is intuitively obvious, it's demonstration is intricate.

Example 5.1.3 Find a parametric representation for the curve resulting by the intersection of the plane $3x+y+z = 1$ and the cylinder $x^2+2y^2 = 1$ in $\mathbb{R}^3$.

Solution: ► The projection of the intersection of the plane $3x+y+z=1$ and the cylinder is the ellipse $x^2+2y^2=1$, on the xy-plane. This ellipse can be parametrized as

$$x=\cos t,\ y=\frac{\sqrt{2}}{2}\sin t, \qquad 0\leqslant t\leqslant 2\pi.$$

From the equation of the plane,

$$z=1-3x-y=1-3\cos t-\frac{\sqrt{2}}{2}\sin t.$$

Thus we may take the parametrization

$$\mathbf{r}(t)=\big(x(t),y(t),z(t)\big)=\left(\cos t,\frac{\sqrt{2}}{2}\sin t,1-3\cos t-\frac{\sqrt{2}}{2}\sin t\right).$$

◄

Proposition 5.1.4 Let $\mathbf{f}:\mathbb{R}^{n+1}\to\mathbb{R}^n$ is differentiable, $c\in\mathbb{R}^n$ and $\gamma=\left\{x\in\mathbb{R}^{n+1}\ \middle|\ f(x)=c\right\}$ be the level set of f. If at every point in γ, the matrix $\mathbf{D}\mathbf{f}$ has rank n then γ is a curve.

Proof: Let $a\in\gamma$. Since $\operatorname{rank}(D(\mathbf{f})_a)=d$, there must be d linearly independent columns in the matrix $D(\mathbf{f})_a$. For simplicity assume these are the first d ones. The implicit function theorem applies and guarantees that the equation $f(x)=c$ can be solved for $x_1,\ldots,x_n$, and each x_i can be expressed as a differentiable function of x_{n+1} (close to a). That is, there exist open sets $U'\subseteq\mathbb{R}^n$, $V'\subseteq\mathbb{R}$ and a differentiable function g such that $a\in U'\times V'$ and $\gamma\bigcap(U'\times V')=\left\{(g(x_{n+1}),x_{n+1})\ \middle|\ x_{n+1}\in V'\right\}$. ■

R A curve can have many parametrizations. For example, $\delta(t)=(\cos t,\sin(-t))$ also parametrizes the unit circle, but runs clockwise instead of counter clockwise. Choosing a parametrization requires choosing the direction of traversal through the curve.

We can change parametrization of $\mathbf{r}$ by taking an invertible smooth function $u\mapsto\tilde{u}$, and have a new parametrization $\mathbf{r}(\tilde{u})=\mathbf{r}(\tilde{u}(u))$. Then by the chain rule,

$$\frac{d\mathbf{r}}{du}=\frac{d\mathbf{r}}{d\tilde{u}}\cdot\frac{d\tilde{u}}{du}$$
$$\frac{d\mathbf{r}}{d\tilde{u}}=\frac{d\mathbf{r}}{du}\Big/\frac{d\tilde{u}}{du}$$

Proposition 5.1.5 Let γ be a regular curve and γ be a parametrization, $a = \gamma(t_0) \in \gamma$. Then the tangent line through a is $\left\{\gamma(t_0) + t\gamma'(t_0) \,\middle|\, t \in \mathbb{R}\right\}$.

If we think of $\gamma(t)$ as the position of a particle at time t, then the above says that the tangent space is spanned by the *velocity* of the particle.

That is, the velocity of the particle is always tangent to the curve it traces out. However, the acceleration of the particle (defined to be γ'') *need not* be tangent to the curve! In fact if the magnitude of the velocity $|\gamma'|$ is constant, then the acceleration will be *perpendicular* to the curve!

5.2 Surfaces

We have seen that a space curve C can be parametrized by a vector function $\mathbf{r} = \mathbf{r}(u)$ where u ranges over some interval I of the u-axis. In an analogous manner we can parametrize a surface S in space by a vector function $\mathbf{r} = \mathbf{r}(u,v)$ where (u,v) ranges over some region Ω of the uv-plane.

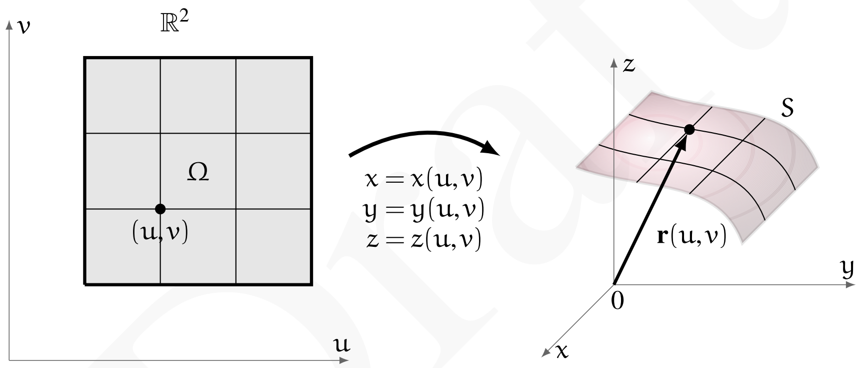


Figure 5.3 Parametrization of a surface S in $\mathbb{R}^3$

Definition 5.2.1 A **parametrized surface** is given by a one-to-one transformation $\mathbf{r} : \Omega \to \mathbb{R}^n$, where Ω is a domain in the plane $\mathbb{R}^2$. The transformation is then given by

$$\mathbf{r}(u,v) = (x_1(u,v), \ldots, x_n(u,v)).$$

Example 5.2.1 (The graph of a function) The graph of a function

$$y = f(x), x \in [a,b]$$

can be parametrized by setting

$$r(u) = u\mathbf{i} + f(u)\mathbf{j}, u \in [a,b].$$

In the same vein the graph of a function

$$z = f(x,y), (x,y) \in \Omega$$

can be parametrized by setting

$$\mathbf{r}(u,v) = u\mathbf{i} + v\mathbf{j} + f(u,v)\mathbf{k}, (u,v) \in \Omega.$$

As (u,v) ranges over Ω, the tip of $\mathbf{r}(u,v)$ traces out the graph of f.

Example 5.2.2 — Plane. If two vectors $\mathbf{a}$ and $\mathbf{b}$ are not parallel, then the set of all linear combinations $ua + vb$ generate a plane p_0 that passes through the origin. We can parametrize this plane by setting

$$\mathbf{r}(u,v) = u\mathbf{a} + v\mathbf{b}, (u,v) \in \mathbb{R} \times \mathbb{R}.$$

The plane p that is parallel to p_0 and passes through the tip of c can be parametrized by setting

$$\mathbf{r}(u,v) = u\mathbf{a} + v\mathbf{b} + \mathbf{c}, (u,v) \in \mathbb{R} \times \mathbb{R}.$$

Note that the plane contains the lines

$$l_1 : r(u,0) = u\mathbf{a} + \mathbf{c} \text{ and } l_2 : r(0,v) = v\mathbf{b} + \mathbf{c}.$$

Example 5.2.3 — Sphere. The sphere of radius a centered at the origin can be parametrized by

$$\mathbf{r}(u,v) = a\cos u \cos v\mathbf{i} + a\sin u \cos v\mathbf{j} + a\sin v\mathbf{k}$$

with (u,v) ranging over the rectangle $R : 0 \leqslant u \leqslant 2\pi, -\frac{\pi}{2} \leqslant v \leqslant \frac{\pi}{2}$.

Derive this parametrization. The points of latitude v form a circle of radius $a\cos v$ on the horizontal plane $z = a\sin v$. This circle can be parametrized by

$$R(u) = a\cos v(\cos u\mathbf{i} + \sin u\mathbf{j}) + a\sin v\mathbf{k}, u \in [0,2\pi].$$

This expands to give

$$R(u,v) = a\cos u \cos v\mathbf{i} + a\sin u \cos v\mathbf{j} + a\sin v\mathbf{k}, u \in [0,2\pi].$$

Letting v range from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, we obtain the entire sphere. The xyz-equation for this same sphere is $x^2 + y^2 + z^2 = a^2$. It is easy to verify that the parametrization satisfies this equation:

$$\begin{aligned} x^2 + y^2 + z^2 &= a^2\cos^2 u\cos^2 v + a^2\sin^2 u\cos^2 v + a^2\sin^2 v \\ &= a^2\left(\cos^2 u + \sin^2 u\right)\cos^2 v + a^2\sin^2 v \\ &= a^2\left(\cos^2 v + \sin^2 v\right) = a^2. \end{aligned}$$

Example 5.2.4 — Cone. Considers a cone with apex semiangle α and slant height s. The points of slant height v form a circle of radius $v\sin\alpha$ on the horizontal plane $z = v\cos a$. This circle can be parametrized by

$$\mathbf{C}(u) = v\sin\alpha(\cos u\mathbf{i} + \sin u\mathbf{j}) + v\cos\alpha\mathbf{k}$$

$$= v\cos u\sin\alpha\mathbf{i} + v\sin u\sin\alpha\mathbf{j} + v\cos\alpha\mathbf{k}, u \in [0, 2\pi].$$

Since we can obtain the entire cone by letting v range from 0 to s, the cone is parametrized by

$$\mathbf{r}(u,v) = v\cos u\sin\alpha\mathbf{i} + v\sin u\sin\alpha\mathbf{j} + v\cos\alpha\mathbf{k},$$

with $0 \leqslant u \leqslant 2\pi, 0 \leqslant v \leqslant s$.

Example 5.2.5 — Spiral Ramp. A rod of length l initially resting on the x-axis and attached at one end to the z-axis sweeps out a surface by rotating about the z-axis at constant rate ω while climbing at a constant rate b.

To parametrize this surface we mark the point of the rod at a distance u from the z-axis ($0 \leqslant u \leqslant l$) and ask for the position of this point at time v. At time v the rod will have climbed a distance bv and rotated through an angle ωv. Thus the point will be found at the tip of the vector

$$u(\cos\omega v\mathbf{i} + \sin\omega v\mathbf{j}) + bv\mathbf{k} = u\cos\omega v\mathbf{i} + u\sin\omega v\mathbf{j} + bv\mathbf{k}.$$

The entire surface can be parametrized by

$$\mathbf{r}(u,v) = u\cos\omega v\mathbf{i} + u\sin\omega v\mathbf{j} + bv\mathbf{k} \text{ with } 0 \leqslant u \leqslant l, 0 \leqslant v.$$

Definition 5.2.2 A **regular parametrized surface** is a smooth mapping $\varphi : U \to \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^2$, of maximal rank. This is equivalent to saying that the rank of φ is 2

Let (u,v) be coordinates in $\mathbb{R}^2$, $(x_1, \ldots, x_n)$ be coordinates in $\mathbb{R}^n$. Then

$$\varphi(u,v) = (x_1(u,v), \ldots, x_n(u,v)),$$

where ${}_ix(u,v)$ admit partial derivatives and the Jacobian matrix has rank two.

5.3 Classical Examples of Surfaces

In this section we consider various surfaces that we shall periodically encounter in subsequent sections.

Let us start with the plane. Recall that if a, b, c are real numbers, not all zero, then the Cartesian equation of a plane with normal vector (a,b,c) and passing through the point (x_0, y_0, z_0) is

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

If we know that the vectors **u** and **v** are on the plane (parallel to the plane) then with the parameters $p, q$the equation of the plane is

$$x - x_0 = pu_1 + qv_1,$$

$$y - y_0 = pu_2 + qv_2,$$

$$z - z_0 = pu_3 + qv_3.$$

Definition 5.3.1 A surface S consisting of all lines parallel to a given line Δ and passing through a given curve γ is called a **cylinder**. The line Δ is called the **directrix** of the cylinder.

To recognise whether a given surface is a cylinder we look at its Cartesian equation. If it is of the form $f(A, B) = 0$, where A, B are secant planes, then the curve is a cylinder. Under these conditions, the lines generating S will be parallel to the line of equation $A = 0, B = 0$. In practice, if one of the variables x, y, or z is missing, then the surface is a cylinder, whose directrix will be the axis of the missing coordinate.

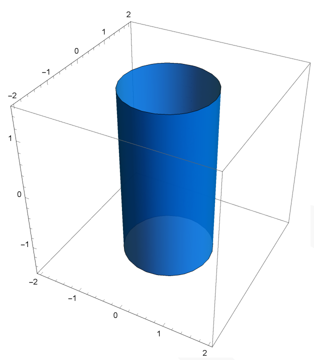


Figure 5.4 Circular cylinder $x^2 + y^2 = 1$.

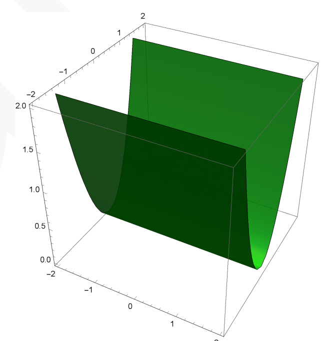


Figure 5.5 The parabolic cylinder $z = y^2$.

Example 5.3.1 Figure 5.4 shews the cylinder with Cartesian equation $x^2 + y^2 = 1$. One starts with the circle $x^2 + y^2 = 1$ on the xy-plane and moves it up and down the z-axis. A parametrization for this cylinder is the following:

$$x = \cos v, \qquad y = \sin v, \qquad z = u, \qquad u \in \mathbb{R}, v \in [0; 2\pi].$$

Example 5.3.2 Figure 5.5 shews the parabolic cylinder with Cartesian equation $z = y^2$. One starts with the parabola $z = y^2$ on the yz-plane and moves it up and down the x-axis. A parametrization for this parabolic cylinder is the following:

$$x = u, \qquad y = v, \qquad z = v^2, \qquad u \in \mathbb{R}, v \in \mathbb{R}.$$

Example 5.3.3 Figure 5.6 shews the hyperbolic cylinder with Cartesian equation $x^2 - y^2 = 1$. One starts with the hyperbola $x^2 - y^2$ on the xy-plane and moves it up and down the z-axis. A parametrization for this parabolic cylinder is the following:

$$x = \pm \cosh v, \qquad y = \sinh v, \qquad z = u, \qquad u \in \mathbb{R}, v \in \mathbb{R}.$$

We need a choice of sign for each of the portions. We have used the fact that $\cosh^2 v - \sinh^2 v = 1$.

Definition 5.3.2 Given a point $\Omega \in \mathbb{R}^3$ (called the **apex**) and a curve γ (called the generating curve), the surface S obtained by drawing rays from Ω and passing through γ is called a **cone**.

In practice, if the Cartesian equation of a surface can be put into the form $f(\frac{A}{C}, \frac{B}{C}) = 0$, where A, B, C, are planes secant at exactly one point, then the surface is a cone, and its apex is given by $A = 0, B = 0, C = 0$.

Example 5.3.4 The surface in $\mathbb{R}^3$ implicitly given by

$$z^2 = x^2 + y^2$$

is a cone, as its equation can be put in the form $\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 - 1 = 0$. Considering the planes $x = 0, y = 0, z = 0$, the apex is located at $(0,0,0)$. The graph is shewn in figure 5.8.

Definition 5.3.3 A surface S obtained by making a curve γ turn around a line Δ is called a **surface of revolution.** We then say that Δ is the axis of revolution. The intersection of S with a half-plane bounded by Δ is called a **meridian**.

If the Cartesian equation of S can be put in the form $f(A, S) = 0$, where A is a plane and S is a sphere, then the surface is of revolution. The axis of S is the line passing through the centre of S and perpendicular to the plane A.

Example 5.3.5 Find the equation of the surface of revolution generated by revolving the hyperbola

$$x^2 - 4z^2 = 1$$

about the z-axis.

Solution: ► Let (x, y, z) be a point on S. If this point were on the xz plane, it would be on the hyperbola, and its distance to the axis of rotation would be $|x| = \sqrt{1 + 4z^2}$. Anywhere else, the distance of (x, y, z) to the axis of rotation is the same as the distance of (x, y, z) to $(0, 0, z)$, that is $\sqrt{x^2 + y^2}$. We must have

$$\sqrt{x^2 + y^2} = \sqrt{1 + 4z^2},$$

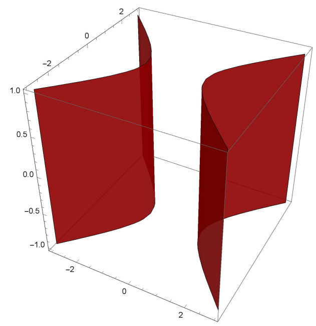


Figure 5.6 The hyperbolic cylinder $x^2 - y^2 = 1$.

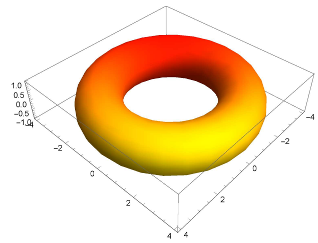


Figure 5.7 The torus.

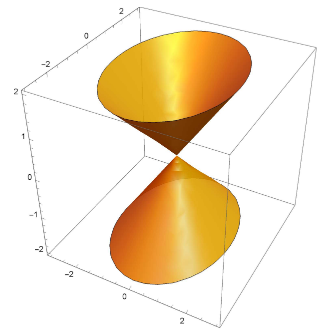


Figure 5.8 Cone $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = \dfrac{z^2}{c^2}$.

which is to say

$$x^2 + y^2 - 4z^2 = 1.$$

This surface is called a **hyperboloid of one sheet**. See figure 5.12. Observe that when $z = 0$, $x^2 + y^2 = 1$ is a circle on the xy plane. When $x = 0$, $y^2 - 4z^2 = 1$ is a hyperbola on the yz plane. When $y = 0$, $x^2 - 4z^2 = 1$ is a hyperbola on the xz plane.

A parametrization for this hyperboloid is

$$x = \sqrt{1+4u^2}\cos v, \qquad y = \sqrt{1+4u^2}\sin v, \qquad z = u, \qquad u \in \mathbb{R}, v \in [0;2\pi].$$

◀

Example 5.3.6 The circle $(y-a)^2 + z^2 = r^2$, on the yz plane (a, r are positive real numbers) is revolved around the z-axis, forming a torus T. Find the equation of this torus.

Solution: ▶ Let (x, y, z) be a point on T. If this point were on the yz plane, it would be on the circle, and the of the distance to the axis of rotation would be $y = a + \operatorname{sgn}(y-a)\sqrt{r^2 - z^2}$, where $\operatorname{sgn}(t)$ (with $\operatorname{sgn}(t) = -1$ if $t < 0$, $\operatorname{sgn}(t) = 1$ if $t > 0$, and $\operatorname{sgn}(0) = 0$) is the sign of t. Anywhere else, the distance from (x, y, z) to the z-axis is the distance of this point to the point $(x, y, z) : \sqrt{x^2 + y^2}$. We must have

$$x^2 + y^2 = (a + \operatorname{sgn}(y-a)\sqrt{r^2 - z^2})^2 = a^2 + 2a\operatorname{sgn}(y-a)\sqrt{r^2 - z^2} + r^2 - z^2.$$

Rearranging

$$x^2 + y^2 + z^2 - a^2 - r^2 = 2a\operatorname{sgn}(y-a)\sqrt{r^2 - z^2},$$

or

$$(x^2 + y^2 + z^2 - (a^2 + r^2))^2 = 4a^2r^2 - 4a^2z^2$$

since $(\operatorname{sgn}(y-a))^2 = 1$, (it could not be 0, why?). Rearranging again,

$$(x^2 + y^2 + z^2)^2 - 2(a^2 + r^2)(x^2 + y^2) + 2(a^2 - r^2)z^2 + (a^2 - r^2)^2 = 0.$$

The equation of the torus thus, is of fourth degree, and its graph appears in figure 7.4.

A parametrization for the torus generated by revolving the circle $(y-a)^2+z^2=r^2$ around the z-axis is

$$x = a\cos\theta + r\cos\theta\cos\alpha, \qquad y = a\sin\theta + r\sin\theta\cos\alpha, \qquad z = r\sin\alpha,$$

with $(\theta,\alpha) \in [-\pi;\pi]^2$.

◀

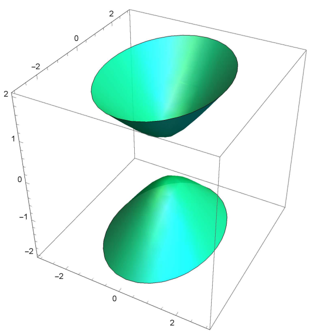


Figure 5.9 Paraboloid $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$

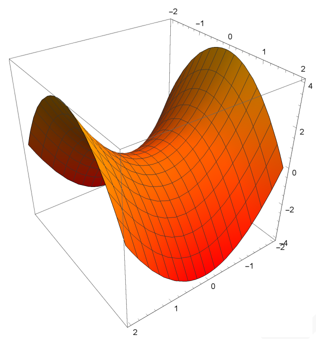


Figure 5.10 Hyperbolic paraboloid $z = \frac{x^2}{a^2}\frac{y^2}{b^2}$

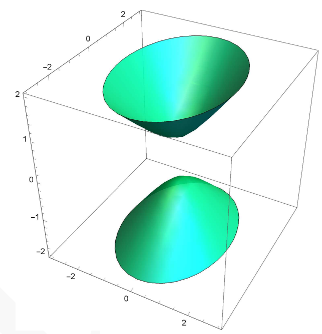


Figure 5.11 Two-sheet hyperboloid $\frac{z^2}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2} + 1$.

Example 5.3.7 The surface $z = x^2+y^2$ is called an **elliptic paraboloid**. The equation clearly requires that $z \geqslant 0$. For fixed $z = c$, $c > 0$, $x^2+y^2 = c$ is a circle. When $y = 0$, $z = x^2$ is a parabola on the xz plane. When $x = 0$, $z = y^2$ is a parabola on the yz plane. See figure 5.9. The following is a parametrization of this paraboloid:

$$x = \sqrt{u}\cos v, \qquad y = \sqrt{u}\sin v, \qquad z = u, \qquad u \in [0;+\infty[, v \in [0;2\pi].$$

Example 5.3.8 The surface $z = x^2-y^2$ is called a **hyperbolic paraboloid** or **saddle**. If $z = 0$, $x^2-y^2 = 0$ is a pair of lines in the xy plane. When $y = 0$, $z = x^2$ is a parabola on the xz plane. When $x = 0$, $z = -y^2$ is a parabola on the yz plane. See figure 5.10. The following is a parametrization of this hyperbolic paraboloid:

$$x = u, \qquad y = v, \qquad z = u^2-v^2, \qquad u \in \mathbb{R}, v \in \mathbb{R}.$$

Example 5.3.9 The surface $z^2 = x^2+y^2+1$ is called an **hyperboloid of two sheets**. For $z^2-1 < 0$, $x^2+y^2 < 0$ is impossible, and hence there is no graph when $-1 < z < 1$. When $y = 0$, $z^2-x^2 = 1$ is a hyperbola on the xz plane. When $x = 0$, $z^2-y^2 = 1$ is a hyperbola on the yz plane. When $z = c$ is a constant $c > 1$, then the $x^2+y^2 = c^2-1$ are circles. See figure 5.11. The following is a parametrization for the top sheet of this hyperboloid of two sheets

$$x = u\cos v, \qquad y = u\sin v, \qquad z = u^2+1, \qquad u \in \mathbb{R}, v \in [0;2\pi]$$

and the following parametrizes the bottom sheet,

$$x = u\cos v, \qquad y = u\sin v, \qquad z = -u^2 - 1, \qquad u \in \mathbb{R}, v \in [0; 2\pi],$$

Example 5.3.10 The surface $z^2 = x^2 + y^2 - 1$ is called an **hyperboloid of one sheet**. For $x^2 + y^2 < 1$, $z^2 < 0$ is impossible, and hence there is no graph when $x^2 + y^2 < 1$. When $y = 0$, $z^2 - x^2 = -1$ is a hyperbola on the xz plane. When $x = 0$, $z^2 - y^2 = -1$ is a hyperbola on the yz plane. When $z = c$ is a constant, then the $x^2 + y^2 = c^2 + 1$ are circles See figure 5.12. The following is a parametrization for this hyperboloid of one sheet

$$x = \sqrt{u^2+1}\cos v, \qquad y = \sqrt{u^2+1}\sin v, \qquad z = u, \qquad u \in \mathbb{R}, v \in [0; 2\pi],$$

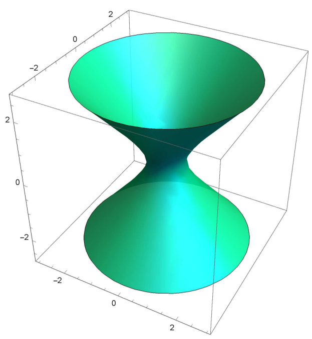


Figure 5.12 One-sheet hyperboloid $\dfrac{z^2}{c^2} = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - 1$.

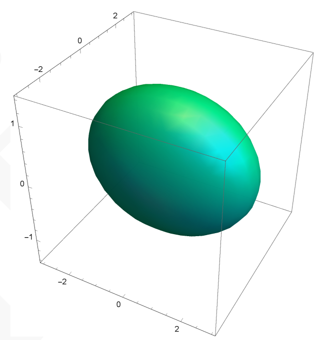


Figure 5.13 Ellipsoid $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$.

Example 5.3.11 Let a, b, c be strictly positive real numbers. The surface $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ is called an **ellipsoid**. For $z = 0$, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} 1$ is an ellipse on the xy plane.When $y = 0$, $\dfrac{x^2}{a^2} + \dfrac{z^2}{c^2} = 1$ is an ellipse on the xz plane. When $x = 0$, $\dfrac{z^2}{c^2} + \dfrac{y^2}{b^2} = 1$ is an ellipse on the yz plane. See figure 5.13. We may parametrize the ellipsoid using spherical coordinates:

$$x = a\cos\theta\sin\phi, \qquad y = b\sin\theta\sin\phi, \qquad z = c\cos\phi, \qquad \theta \in [0; 2\pi], \phi \in [0; \pi].$$

Exercises

Problem 5.1 Find the equation of the surface of revolution S generated by revolving the ellipse $4x^2 + z^2 = 1$ about the z-axis.

Problem 5.2 Find the equation of the surface of revolution generated by revolving the line $3x + 4y = 1$ about the y-axis .

Problem 5.3 Describe the surface parametrized by $\varphi(u, v) \mapsto (v\cos u, v\sin u, au)$, $(u, v) \in (0, 2\pi) \times (0, 1)$, $a > 0$.

Problem 5.4 Describe the surface parametrized by $\varphi(u,v) = (au\cos v, bu\sin v, u^2)$, $(u,v) \in (1,+\infty)\times(0,2\pi)$, $a,b>0$.

Problem 5.5 Consider the spherical cap defined by

$$S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1,\ z \geqslant 1/\sqrt{2}\}.$$

Parametrise S using Cartesian, Spherical, and Cylindrical coordinates.

Problem 5.6 Demonstrate that the surface in $\mathbb{R}^3$

$$S : e^{x^2+y^2+z^2} - (x+z)e^{-2xz} = 0,$$

implicitly defined, is a cylinder.

Problem 5.7 Shew that the surface in $\mathbb{R}^3$ implicitly defined by

$$x^4+y^4+z^4-4xyz(x+y+z) = 1$$

is a surface of revolution, and find its axis of revolution.

Problem 5.8 Shew that the surface S in $\mathbb{R}^3$ given implicitly by the equation

$$\frac{1}{x-y}+\frac{1}{y-z}+\frac{1}{z-x} = 1$$

is a cylinder and find the direction of its directrix.

Problem 5.9 Shew that the surface S in $\mathbb{R}^3$ implicitly defined as

$$xy+yz+zx+x+y+z+1 = 0$$

is of revolution and find its axis.

Problem 5.10 Demonstrate that the surface in $\mathbb{R}^3$ given implicitly by

$$z^2 - xy = 2z-1$$

is a cone

Problem 5.11 — Putnam Exam 1970. Determine, with proof, the radius of the largest circle which can lie on the ellipsoid

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2} = 1, \qquad a>b>c>0.$$

Problem 5.12 The hyperboloid of one sheet in figure 5.14 has the property that if it is cut by planes at $z=\pm 2$, its projection on the xy plane produces the ellipse $x^2+\frac{y^2}{4}=1$, and if it is cut by a plane at $z=0$, its projection on the xy plane produces the ellipse $4x^2+y^2=1$. Find its equation.

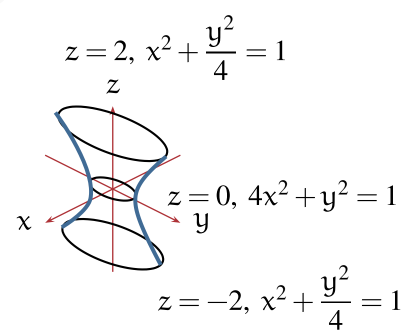


Figure 5.14 Problem 5.12.

5.4 ⋆ Manifolds

Definition 5.4.1 We say $M \subseteq \mathbb{R}^n$ is a d-dimensional (differentiable) **manifold** if for every $a \in M$ there exists domains $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^n$ and a differentiable function $\mathbf{f} : V \to U$ such that $\operatorname{rank}(D(\mathbf{f})) = d$ at every point in V and $U \bigcap M = \mathbf{f}(V)$.

R For $d=1$ this is just a curve, and for $d=2$ this is a surface.

R If $d=1$ and γ is a connected, then there exists an interval U and an injective differentiable function $\gamma: U \to \mathbb{R}^n$ such that $D\gamma \neq 0$ on U and $\gamma(U) = \gamma$. If $d > 1$ this is no longer true: even though near every point the surface is a differentiable image of a rectangle, the entire surface need not be one.

As before d-dimensional manifolds can be obtained as level sets of functions $f: \mathbb{R}^{n+d} \to \mathbb{R}^n$ provided we have $\operatorname{rank}(D(\mathbf{f})) = d$ on the entire level set.

Proposition 5.4.1 Let $f: \mathbb{R}^{n+d} \to \mathbb{R}^n$ is differentiable, $c \in \mathbb{R}^n$ and $\gamma = \left\{x \in \mathbb{R}^{n+1} \mid f(x) = c\right\}$ be the level set of f. If at every point in γ, the matrix $D(\mathbf{f})$ has rank d then γ is a d-dimensional manifold.

The results from the previous section about tangent spaces of implicitly defined manifolds generalize naturally in this context.

Definition 5.4.2 Let $U \subseteq \mathbb{R}^n$, $f: U \to R$ be a differentiable function, and $M = \left\{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in U\right\}$ be the graph of f. (Note M is a d-dimensional manifold in $\mathbb{R}^{n+1}$.) Let $(a, f(a)) \in M$.

- The **tangent "plane"** at the point $(a, f(a))$ is defined by
$$\left\{(x,y) \in \mathbb{R}^{n+1} \mid y = f(a) + Df_a(x-a)\right\}$$
- The **tangent space** at the point $(a, f(a))$ (denoted by $TM_{(a,f(a))}$) is the subspace defined by
$$TM_{(a,f(a))} = \left\{(x,y) \in \mathbb{R}^{n+1} \mid y = Df_a x\right\}.$$

R When $d=2$ the tangent plane is really a plane. For $d=1$ it is a line (the tangent line), and for other values it is a d-dimensional hyper-plane.

Proposition 5.4.2 Suppose $\mathbf{f}: \mathbb{R}^{n+d} \to \mathbb{R}^n$ is differentiable, and the level set $\gamma = \left\{x \mid f(x) = c\right\}$ is a d-dimensional manifold. Suppose further that $D(\mathbf{f})_a$ has rank n for all $a \in \gamma$. Then the tangent space at a is precisely the kernel of $D(\mathbf{f})_a$, and the vectors $\nabla f_1, \dots \nabla f_n$ are n linearly independent vectors that are normal to the tangent space.

5.5 Constrained optimization.

Consider an implicitly defined surface $S = \{g = c\}$, for some $g: \mathbb{R}^3 \to \mathbb{R}$. Our aim is to maximise or minimise a function f *on this surface*.

Definition 5.5.1 We say a function f attains a local maximum at a on the surface S, if there exists $\epsilon > 0$ such that $|x - a| < \epsilon$ and $x \in S$ imply $f(a) \geqslant f(x)$.

R This is sometimes called constrained local maximum, or local maximum subject to the constraint $g = c$.

Proposition 5.5.1 If f attains a local maximum at a on the surface S, then $\exists \lambda \in \mathbb{R}$ such that $\nabla f(a) = \lambda \nabla g(a)$.

Proof:[Intuition] If $\nabla f(a) \neq 0$, then $S' \stackrel{\text{def}}{=} \{f = f(a)\}$ is a surface. If f attains a constrained maximum at a then S' must be tangent to S at the point a. This forces $\nabla f(a)$ and $\nabla g(a)$ to be parallel.
■

Proposition 5.5.2 — Multiple constraints. Let $f, g_1, \ldots, g_n : \mathbb{R}^d \to \mathbb{R}$ be $: \mathbb{R}^d \to \mathbb{R}$ be differentiable. If f attains a local maximum at a subject to the constraints $g_1 = c_1$, $g_2 = c_2$, $\ldots g_n = c_n$ then $\exists \lambda_1, \ldots \lambda_n \in \mathbb{R}$ such that $\nabla f(a) = \sum_1^n \lambda_i \nabla g_i(a)$.

To explicitly find constrained local maxima in $\mathbb{R}^n$ with n constraints we do the following:

- Simultaneously solve the system of equations

$$\begin{gathered}
\nabla f(x) = \lambda_1 \nabla g_1(x) + \cdots \lambda_n \nabla g_n(x) \\
g_1(x) = c_1, \\
\ldots \\
g_n(x) = c_n.
\end{gathered}$$

- The unknowns are the d-coordinates of x, and the Lagrange multipliers $\lambda_1, \ldots, \lambda_n$. This is $n + d$ variables.
- The first equation above is a vector equation where both sides have d coordinates. The remaining are scalar equations. So the above system is a system of $n + d$ equations with $n + d$ variables.
- The typical situation will yield a finite number of solutions.

- There is a test involving the *bordered Hessian* for whether these points are constrained local minima / maxima or neither. These are quite complicated, and are usually more trouble than they are worth, so one usually uses some ad-hoc method to decide whether the solution you found is a local maximum or not.

Example 5.5.3 Find necessary conditions for $f(x,y)=y$ to attain a local maxima/minima of subject to the constraint $y=g(x)$.

Of course, from one variable calculus, we know that the local maxima / minima must occur at points where $g'=0$. Let's revisit it using the constrained optimization technique above.
Proof:[Solution] Note our constraint is of the form $y-g(x)=0$. So at a local maximum we must have

$$\begin{bmatrix}0\\1\end{bmatrix}=\nabla f=\lambda\nabla(y-g(x))=\begin{bmatrix}-g'(x)\\1\end{bmatrix}\quad\text{and}\quad y=g(x).$$

This forces $\lambda=1$ and hence $g'(x)=0$, as expected. ■

Example 5.5.4 Maximise xy subject to the constraint $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$.

Proof:[Solution] At a local maximum,

$$\begin{bmatrix}y\\x\end{bmatrix}=\nabla(xy)=\lambda\nabla\Big(\frac{x^2}{a^2}+\frac{y^2}{b^2}\Big)=\lambda\begin{bmatrix}2x/a^2\\2y/b^2.\end{bmatrix}$$

which forces $y^2=x^2b^2/a^2$. Substituting this in the constraint gives $x=\pm a/\sqrt{2}$ and $y=\pm b/\sqrt{2}$. This gives four possibilities for xy to attain a maximum. Directly checking shows that the points $(a/\sqrt{2},b/\sqrt{2})$ and $(-a/\sqrt{2},-b/\sqrt{2})$ both correspond to a local maximum, and the maximum value is $ab/2$. ■

Proposition 5.5.5 — Cauchy-Schwartz. If $x,y\in\mathbb{R}^n$ then $|x\cdot y|\leqslant|x||y|$.

Proof: Maximise $x\cdot y$ subject to the constraint $|x|=a$ and $|y|=b$. ■

Proposition 5.5.6 — Inequality of the means. If $x_i\geqslant 0$, then

$$\frac{1}{n}\sum_1^n x_i\geqslant\Big(\prod_1^n x_i\Big)^{1/n}.$$

Proposition 5.5.7 — Young's inequality. If $p, q > 1$ and $1/p + 1/q = 1$ then

$$|xy| \leqslant \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$

CHAPTER 6

Line Integrals

So far we have always insisted all curves and parametrizations are differentiable or C^1. We now relax this requirement and subsequently only assume that all curves (and parametrizations) are **piecewise differentiable**, or **piecewise** C^1.

Definition 6.0.1 A function $f : [a, b] \to \mathbb{R}^n$ is called **piecewise** C^1 if there exists a finite set $F \subseteq [a, b]$ such that f is C^1 on $[a, b] - F$, and further both left and right limits of f and f' exist at all points in F.

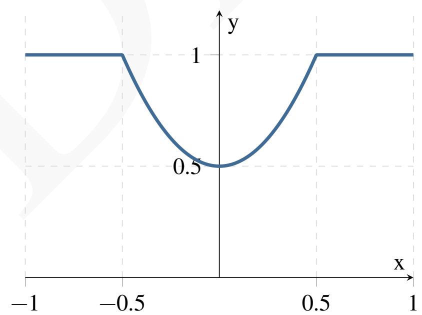


Figure 6.1 Piecewise C^1 function

Definition 6.0.2 A (connected) curve γ is **piecewise** C^1 if it has a parametrization which is continuous *and* piecewise C^1.

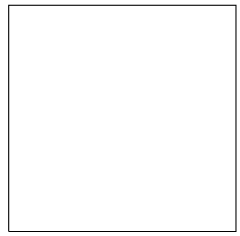


Figure 6.2 The boundary of a square is a piecewise C^1 curve, but not a differentiable curve.

R A piecewise C^1 function need not be continuous. But curves are always assumed to be at least continuous; so for notational convenience, we define a piecewise C^1 curve to be one which has a parametrization which is both continuous and piecewise C^1.

6.1 Line Integrals of Vector Fields

We start with some motivation. With this objective we remember the definition of the work:

Definition 6.1.1 If a constant force $\mathbf{f}$ acting on a body produces an displacement $\Delta\mathbf{x}$, then the **work** done by the force is $\mathbf{f}\bullet\Delta\mathbf{x}$.

We want to generalize this definition to the case in which the force is not constant. For this purpose let $\gamma \subseteq \mathbb{R}^3$ be a curve, with a given direction of traversal, and $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector function.

Here $\mathbf{f}$ represents the force that acts on a body and pushes it along the curve γ. The work done by the force can be approximated by

$$W \approx \sum_{i=0}^{N-1} \mathbf{f}(x_i)\bullet(\mathbf{x}_{i+1} - \mathbf{x}_i) = \sum_{i=0}^{N-1} \mathbf{f}(x_i)\bullet\Delta\mathbf{x}_i$$

where $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ are N points on γ, *chosen along the direction of traversal*. The limit as the largest distance between neighbors approaches 0 is the work done:

$$W = \lim_{\|P\|\to 0} \sum_{i=0}^{N-1} \mathbf{f}(x_i)\bullet\Delta\mathbf{x}_i$$

This motivates the following definition:

Definition 6.1.2 Let $\gamma \subseteq \mathbb{R}^n$ be a curve with a given direction of traversal, and $\mathbf{f}: \gamma \to \mathbb{R}^n$ be a (vector) function. The **line integral** of $\mathbf{f}$ over γ is defined to be

$$\int_\gamma \mathbf{f} \bullet d\ell = \lim_{\|P\| \to 0} \sum_{i=0}^{N-1} \mathbf{f}(\mathbf{x}_i^*) \bullet (\mathbf{x}_{i+1} - \mathbf{x}_i)$$
$$= \lim_{\|P\| \to 0} \sum_{i=0}^{N-1} \mathbf{f}(\mathbf{x}_i^*) \bullet \Delta \mathbf{x}_i.$$

Here $P = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}\}$, the points x_i are chosen along the direction of traversal, and $\|P\| = \max|\mathbf{x}_{i+1} - \mathbf{x}_i|$.

R If $\mathbf{f} = (f_1, \ldots, f_n)$, where $f_i : \gamma \to \mathbb{R}$ are functions, then one often writes the line integral in the **differential form** notation as

$$\int_\gamma \mathbf{f} \bullet d\ell = \int_\gamma f_1 \ dx_1 + \cdots + f_n \ dx_n$$

The following result provides a explicit way of calculating line integrals using a parametrization of the curve.

Theorem 6.1.1 If $\gamma : [a, b] \to \mathbb{R}^n$ is a parametrization of γ (in the direction of traversal), then

$$\int_\gamma \mathbf{f} \bullet d\ell = \int_a^b \mathbf{f} \circ \gamma(t) \bullet \gamma'(t) \ dt \tag{6.1}$$

Proof:

Let $a = t_0 < t_1 < \cdots < t_n = b$ be a partition of a, b and let $\mathbf{x}_i = \gamma(t_i)$.
The line integral of $\mathbf{f}$ over γ is defined to be

$$\int_\gamma \mathbf{f} \bullet d\ell = \lim_{\|P\| \to 0} \sum_{i=0}^{N-1} \mathbf{f}(\mathbf{x}_i) \bullet \Delta \mathbf{x}_i$$
$$= \lim_{\|P\| \to 0} \sum_{i=0}^{N-1} \sum_{j=1}^{n} f_j(\mathbf{x}_i) \cdot (\Delta \mathbf{x}_i)_j$$

By the Mean Value Theorem, we have $(\Delta \mathbf{x}_i)_j = \left(x'^*_i\right)_j \Delta t_i$

$$\sum_{j=1}^{n} \sum_{i=0}^{N-1} f_j(\mathbf{x}_i) \cdot (\Delta x_i)_j = \sum_{j=1}^{n} \sum_{i=0}^{N-1} f_j(\mathbf{x}_i) \cdot \left(x'^*_i\right)_j \Delta t_i$$

$$= \sum_{j=1}^{n} \int f_j(\gamma(x)) \cdot \gamma_j'(t)\ dt = \int_a^b \mathbf{f} \circ \gamma(t) \bullet \gamma'(t)\ dt$$

■

In the differential form notation (when $d = 2$) say

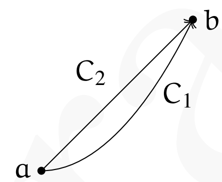
$$\mathbf{f} = (f, g) \quad \text{and} \quad \gamma(t) = \big(x(t), y(t)\big),$$

where $f, g : \gamma \to \mathbb{R}$ are functions. Then Proposition 6.1.1 says

$$\int_\gamma \mathbf{f} \bullet d\ell = \int_\gamma f\ dx + g\ dy = \int_\gamma \big[f(x(t), y(t))\, x'(t) + g(x(t), y(t))\, y'(t)\big]\ dt$$

R Sometimes (6.1) is used as the definition of the line integral. In this case, one needs to verify that this definition is **independent** of the parametrization. Since this is a good exercise, we'll do it anyway a little later.

Example 6.1.2 Take $\mathbf{F}(\mathbf{r}) = (xe^y, z^2, xy)$ and we want to find the line integral from $\mathbf{a} = (0,0,0)$ to $\mathbf{b} = (1,1,1)$.



We first integrate along the curve $C_1 : \mathbf{r}(u) = (u, u^2, u^3)$. Then $\mathbf{r}'(u) = (1, 2u, 3u^2)$, and $\mathbf{F}(\mathbf{r}(u)) = (ue^{u^2}, u^6, u^3)$. So

$$\begin{aligned}
\int_{C_1} \mathbf{F} \bullet d\mathbf{r} &= \int_0^1 \mathbf{F} \bullet \mathbf{r}'(u)\ du \\
&= \int_0^1 ue^{u^2} + 2u^7 + 3u^5\ du \\
&= \frac{e}{2} - \frac{1}{2} + \frac{1}{4} + \frac{1}{2} \\
&= \frac{e}{2} + \frac{1}{4}
\end{aligned}$$

Now we try to integrate along another curve $C_2 : \mathbf{r}(t) = (t, t, t)$. So $\mathbf{r}'(t) = (1, 1, 1)$.

$$\begin{aligned}
\int_{C_2} \mathbf{F} \bullet d\ell &= \int \mathbf{F} \bullet \mathbf{r}'(t) dt \\
&= \int_0^1 te^t + 2t^2\ dt \\
&= \frac{5}{3}.
\end{aligned}$$

We see that the line integral depends on the curve C in general, not just $\mathbf{a}, \mathbf{b}$.

Example 6.1.3 Suppose a body of mass M is placed at the origin. The force experienced by a body of mass m at the point $x \in \mathbb{R}^3$ is given by $\mathbf{f}(x) = \dfrac{-GMx}{|x|^3}$, where G is the **gravitational constant**. Compute the work done when the body is moved from a to b along a straight line.

Solution: ▶ Let γ be the straight line joining a and b. Clearly $\gamma : [0,1] \to \gamma$ defined by $\gamma(t) = a + t(b - a)$ is a parametrization of γ. Now

$$W = \int_\gamma \mathbf{f} \bullet d\ell = -GMm \int_0^1 \frac{\gamma(t)}{|\gamma(t)|^3} \bullet \gamma'(t)\ dt = \frac{GMm}{|b|} - \frac{GMm}{|a|}. \quad \blacksquare$$

◀

R If the line joining through a and b passes through the origin, then some care has to be taken when doing the above computation. We will see later that gravity is a **conservative force**, and that the above line integral only depends on the endpoints and not the actual path taken.

6.2 Parametrization Invariance and Others Properties of Line Integrals

Since line integrals can be defined in terms of ordinary integrals, they share many of the properties of ordinary integrals.

Definition 6.2.1 The curve γ is said to be the union of two curves γ_1 and γ_2 if γ is defined on an interval $[a, b]$, and the curves γ_1 and γ_2 are the restriction $\gamma|_{[a,d]}$ and $\gamma|_{[d,b]}$.

Proposition 6.2.1

- linearity property with respect to the integrand,

$$\int_\gamma (\alpha \mathbf{f} + \beta \mathbf{G}) \bullet d\ell = \alpha \int_\gamma \mathbf{f} \bullet d\ell + \beta \int_\gamma \mathbf{G} \bullet d\ell$$

- additive property with respect to the path of integration: where the union of the two curves γ_1 and γ_2 is the curve γ.

$$\int_\gamma \mathbf{f} \bullet d\ell = \int_{\gamma_1} \mathbf{f} \bullet d\ell + \int_{\gamma_2} \mathbf{f} \bullet d\ell$$

The proofs of these properties follows immediately from the definition of the line integral.

Definition 6.2.2 Let $h : I \to I_1$ be aC^1 real-valued function that is a one-to-one map of an interval $I = [a, b]$ onto another interval $I = [a_1, b_1]$. Let $\gamma : I_1 \to \mathbb{R}^3$ be a piecewise C^1 path. Then we call the composition

$$\gamma_2 = \gamma_1 \circ h : I \to \mathbb{R}^3$$

a **reparametrization** of γ.

It is implicit in the definition that h must carry endpoints to endpoints; that is, either $h(a) = a_1$ and $h(b) = b_1$, or $h(a) = b_1$ and $h(b) = a_1$. We distinguish these two types of reparametrizations.

- In the first case, the reparametrization is said to be **orientation-preserving**, and a particle tracing the path $\gamma_1 \circ$ moves in the same direction as a particle tracing γ_1.
- In the second case, the reparametrization is described as **orientation-reversing**, and a particle tracing the path $\gamma_1 \circ$ moves in the opposite direction to that of a particle tracing γ_1

Proposition 6.2.2 — Parametrization invariance. If $\gamma_1 : [a_1, b_1] \to \gamma$ and $\gamma_2 : [a_2, b_2] \to \gamma$ are two parametrizations of γ that traverse it in the same direction, then

$$\int_{a_1}^{b_1} \mathbf{f} \circ \gamma_1(t) \boldsymbol{\cdot} \gamma_1'(t) \;\, dt = \int_{a_2}^{b_2} \mathbf{f} \circ \gamma_2(t) \boldsymbol{\cdot} \gamma_2'(t) \;\, dt.$$

Proof: Let $\varphi : [a_1, b_1] \to [a_2, b_2]$ be defined by $\varphi = \gamma_2^{-1} \circ \gamma_1$. Since γ_1 and γ_2 traverse the curve in the same direction, φ must be increasing. One can also show (using the inverse function theorem) that φ is continuous and piecewise C^1. Now

$$\int_{a_2}^{b_2} \mathbf{f} \circ \gamma_2(t) \boldsymbol{\cdot} \gamma_2'(t) \;\, dt = \int_{a_2}^{b_2} \mathbf{f}(\gamma_1(\varphi(t))) \boldsymbol{\cdot} \gamma_1'(\varphi(t)) \varphi'(t) \;\, dt.$$

Making the substitution $s = \varphi(t)$ finishes the proof. ■

6.3 Line Integral of Scalar Fields

Definition 6.3.1 If $\gamma \subseteq \mathbb{R}^n$ is a piecewise C^1 curve, then

$$\text{length}(\gamma) = \int_\gamma f \, |d\ell| = \lim_{\|P\| \to 0} \sum_{i=0}^{N} |x_{i+1} - x_i|,$$

where as before $P = \{x_0, \ldots, x_{N-1}\}$.

More generally:

Definition 6.3.2 If $f : \gamma \to \mathbb{R}$ is any scalar function, we define[a]

$$\int_\gamma f\,|d\ell| \overset{\text{def}}{=} \lim_{\|P\|\to 0} \sum_{i=0}^{N} f(x_i^*)\,|x_{i+1} - x_i|,$$

[a] Unfortunately $\int_\gamma f|d\ell|$ is also called the line integral. To avoid confusion, we will call this the **line integral with respect to arc-length** instead.

The integral $\int_\gamma f\,|d\ell|$ is also denoted by

$$\int_\gamma f \;\, ds = \int_\gamma f\,|d\ell|$$

Theorem 6.3.1 Let $\gamma \subseteq \mathbb{R}^n$ be a piecewise C^1 curve, $\gamma : [a,b] \to \mathbb{R}$ be any parametrization (in the given direction of traversal), $f : \gamma \to \mathbb{R}$ be a scalar function. Then

$$\int_\gamma f\,|d\ell| = \int_a^b f(\gamma(t))\left|\gamma'(t)\right| \; dt,$$

and consequently

$$\operatorname{length}(\gamma) = \int_\gamma 1\,|d\ell| = \int_a^b \left|\gamma'(t)\right| \; dt.$$

Example 6.3.2 Compute the circumference of a circle of radius r.

Example 6.3.3 The trace of

$$\mathbf{r}(t) = \mathbf{i}\cos t + \mathbf{j}\sin t + \mathbf{k}t$$

is known as a *cylindrical helix*. To find the length of the helix as t traverses the interval $[0; 2\pi]$, first observe that

$$\|d\ell\| = \left\|(\sin t)^2 + (-\cos t)^2 + 1\right\| dt = \sqrt{2}dt,$$

and thus the length is

$$\int_0^{2\pi} \sqrt{2}dt = 2\pi\sqrt{2}.$$

6.3.1 Area above a Curve

If γ is a curve in the xy-plane and $\mathbf{f}(x,y)$ is a nonnegative continuous function defined on the curve γ, then the integral

$$\int_\gamma f(x,y)|d\ell|$$

can be interpreted as the area A of the curtain that obtained by the union of all vertical line segment that extends upward from the point (x, y) to a height of $f(x, y)$, i.e, the area bounded by the curve γ and the graph of f

This fact come from the approximation by rectangles:

$$\text{area} = \lim_{\|P\| \to 0} \sum_{i=0}^{N} f(x, y)|x_{i+1} - x_i|,$$

Example 6.3.4 Use a line integral to show that the lateral surface area A of a right circular cylinder of radius r and height h is $2\pi r h$.

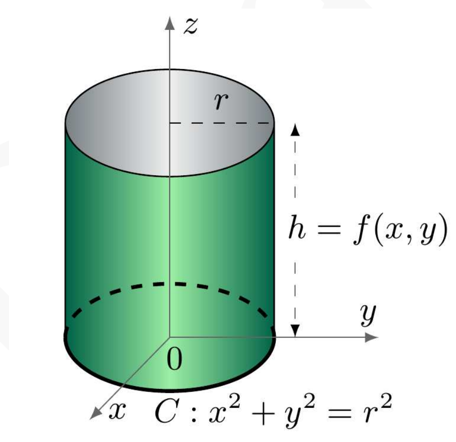


Figure 6.3 Right circular cylinder of radius r and height h

Solution: ► We will use the right circular cylinder with base circle C given by $x^2 + y^2 = r^2$ and with height h in the positive z direction (see Figure 4.1.3). Parametrize C as follows:

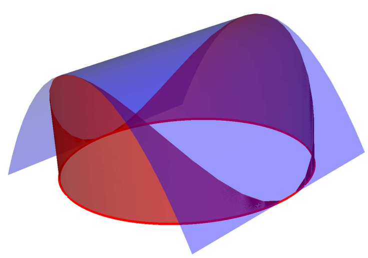
$$x = x(t) = r\cos t\,, \quad y = y(t) = r\sin t\,, \quad 0 \leqslant t \leqslant 2\pi$$

Let $f(x,y) = h$ for all (x,y). Then

$$\begin{aligned} A &= \int_C f(x,y)\,ds = \int_a^b f(x(t),y(t))\sqrt{x'(t)^2 + y'(t)^2}\,dt \\ &= \int_0^{2\pi} h\sqrt{(-r\sin t)^2 + (r\cos t)^2}\,dt \\ &= h\int_0^{2\pi} r\sqrt{\sin^2 t + \cos^2 t}\,dt \\ &= rh\int_0^{2\pi} 1\,dt = 2\pi rh \end{aligned}$$

◀

Example 6.3.5 Find the area of the surface extending upward from the circle $x^2 + y^2 = 1$ in the xy-plane to the parabolic cylinder $z = 1 - y^2$



Solution: ► The circle circle C given by $x^2 + y^2 = 1$ can be parametrized as as follows:

$$x = x(t) = \cos t, \quad y = y(t) = \sin t, \quad 0 \leqslant t \leqslant 2\pi$$

Let $f(x,y) = 1 - y^2$ for all (x,y). Above the circle he have $f(\theta) = 1 - \sin^2 t$ Then

$$\begin{aligned} A &= \int_C f(x,y)\,ds = \int_a^b f(x(t),y(t))\sqrt{x'(t)^2 + y'(t)^2}\,dt \\ &= \int_0^{2\pi} (1 - \sin^2 t)\sqrt{(-\sin t)^2 + (\cos t)^2}\,dt \\ &= \int_0^{2\pi} 1 - \sin^2 t\,dt = \pi \end{aligned}$$

◀

6.4 The First Fundamental Theorem

Definition 6.4.1 Suppose $U \subseteq \mathbb{R}^n$ is a domain. A vector field $\mathbf{F}$ is a **gradient field** in U if exists an C^1 function $\varphi : U \to \mathbb{R}$ such that

$$\mathbf{F} = \nabla\varphi$$

Definition 6.4.2 Suppose $U \subseteq \mathbb{R}^n$ is a domain. A vector field $\mathbf{f}: U \to \mathbb{R}^n$ is a **path-independent** vector field if the integral of $\mathbf{f}$ over a piecewise C^1 curve is dependent only on end points, for all piecewise C^1 curve in U.

Theorem 6.4.1 — First Fundamental theorem for line integrals. Suppose $U \subseteq \mathbb{R}^n$ is a domain, $\varphi : U \to \mathbb{R}$ is C^1 and $\gamma \subseteq \mathbb{R}^n$ is any differentiable curve that starts at a, ends at b and is completely contained in U. Then

$$\int_\gamma \nabla\varphi \bullet d\ell = \varphi(b) - \varphi(a).$$

Proof: Let $\gamma : [0,1] \to \gamma$ be a parametrization of γ. Note

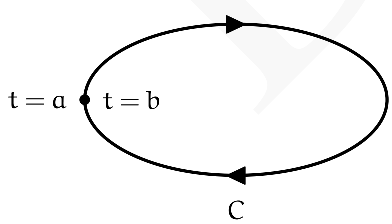
$$\int_\gamma \nabla\varphi \bullet d\ell = \int_0^1 \nabla\varphi(\gamma(t)) \bullet \gamma'(t)\ dt = \int_0^1 \frac{d}{dt}\varphi(\gamma(t))\ dt = \varphi(b) - \varphi(a). \quad \blacksquare$$

■

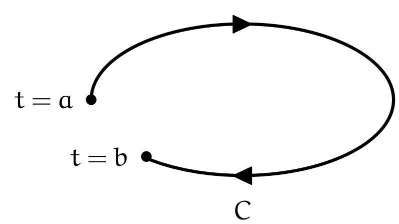
The above theorem can be restated as: a gradient vector field is a path-independent vector field.

Definition 6.4.3 A **closed curve** is a curve that starts and ends at the same point, i.e. for C : $x = x(t)$, $y = y(t)$, $a \leqslant t \leqslant b$, we have $(x(a), y(a)) = (x(b), y(b))$. A **simple closed curve** is a closed curve which does not intersect itself.

Note that any closed curve can be regarded as a union of simple closed curves (think of the loops in a figure eight). We use the special notation



(a) Closed



(b) Not closed

Figure 6.4 Closed vs nonclosed curves

If γ is a closed curve, then line integrals over γ are denoted by

$$\oint_\gamma \mathbf{f} \bullet d\ell.$$

Corollary 6.4.2 If $\gamma \subseteq \mathbb{R}^n$ is a closed curve, and $\varphi : \gamma \to \mathbb{R}$ is C^1, then

$$\oint_\gamma \nabla\varphi \cdot d\ell = 0.$$

Definition 6.4.4 Let $U \subseteq \mathbb{R}^n$, and $\mathbf{f} : U \to \mathbb{R}^n$ be a vector function. We say $\mathbf{f}$ is a **conservative force** (or **conservative vector field**) if

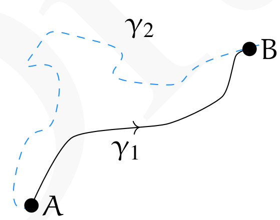
$$\oint \mathbf{f} \cdot d\ell = 0,$$

for all closed curves γ which are completely contained inside U.

Clearly if $\mathbf{f} = -\nabla\phi$ for some C^1 function $V : U \to \mathbb{R}$, then $\mathbf{f}$ is conservative. The converse is also true provided U is **simply connected**, which we'll return to later. For conservative vector field:

$$\begin{aligned}\int_\gamma \mathbf{F} \cdot d\ell &= \int_\gamma \nabla\phi \cdot d\ell \\ &= [\phi]_a^b \\ &= \phi(b) - \phi(a)\end{aligned}$$

We note that the result is *independent of the path* γ joining a to b.



Example 6.4.3 If φ fails to be C^1 even at one point, the above can fail quite badly. Let $\varphi(x,y) = \tan^{-1}(y/x)$, extended to $\mathbb{R}^2 - \left\{(x,y) \mid x \leqslant 0\right\}$ in the usual way. Then

$$\nabla\varphi = \frac{1}{x^2+y^2}\begin{bmatrix} -y \\ x \end{bmatrix}$$

which is defined on $\mathbb{R}^2 - (0,0)$. In particular, if $\gamma = \left\{(x,y) \mid x^2+y^2 = 1\right\}$, then $\nabla\varphi$ is defined on all of γ. However, you can easily compute

$$\oint_\gamma \nabla\varphi \cdot d\ell = 2\pi \neq 0.$$

The reason this doesn't contradict the previous corollary is that Corollary 6.4.2 requires φ itself to be defined on all of γ, and not just $\nabla\varphi$! This example leads into something called the **winding number** which we will return to later.

6.5 Test for a Gradient Field

If a vector field $\mathbf{F}$ is a gradient field, and the potential φ has continuous second derivatives, then the second-order mixed partial derivatives must be equal:

$$\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x}) \text{ for all } i,j$$

So if $\mathbf{F} = (F_1, \ldots, F_n)$ is a gradient field and the components of $\mathbf{F}$ have continuous partial derivatives, then we must have

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \frac{\partial f_j}{\partial x_i}(\mathbf{x}) \text{ for all } i,j$$

If these partial derivatives do not agree, then the vector field cannot be a gradient field.

This gives us an easy way to determine that a vector field is *not* a gradient field.

Example 6.5.1 The vector field $(-y, x, -yx)$ is not a gradient field because $M_y = -1$ is not equal to $N_x = 1$.

When $\mathbf{F}$ is defined on simple connected domain and has continuous partial derivatives, the check works the other way as well.

If $\mathbf{F} = (F_1, \ldots, F_n)$ is field and the components of $\mathbf{F}$ have continuous partial derivatives, satisfying

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \frac{\partial f_j}{\partial x_i}(\mathbf{x}) \text{ for all } i,j$$

then $\mathbf{F}$ is a gradient field (i.e., there is a potential function f such that $\mathbf{F} = \nabla f$). This gives us a very nice way of checking if a vector field is a gradient field.

Example 6.5.2 The vector field $\mathbf{F} = (x, z, y)$ is a gradient field because $\mathbf{F}$ is defined on all of $\mathbb{R}^3$, each component has continuous partial derivatives, and $M_y = 0 = N_x$, $M_z = 0 = P_x$, and $N_z = 1 = P_y$. Notice that $f = x^2/2 + yz$ gives $\nabla f = \langle x, z, y \rangle = \mathbf{F}$.

6.5.1 Irrotational Vector Fields

In this section we restrict our attention to three dimensional space .

Definition 6.5.1 Let $\mathbf{f} : U \to \mathbb{R}^3$ be a C^1 vector field defined in the open set U. Then the vector $\mathbf{f}$ is called **irrotational** if and only if its curl is $\mathbf{0}$ everywhere in U, i.e., if

$$\nabla \times \mathbf{f} \equiv \mathbf{0}.$$

For any C^2 scalar field φ on U, we have

$$\nabla \times (\nabla \varphi) \equiv \mathbf{0}.$$

so every C^1 conservative vector field on U is also an irrotational vector field on U.

Provided that U is simply connected, the converse of this is also true:

Theorem 6.5.3 Let $U \subset \mathbb{R}^3$ be a simply connected domain and let f be a C^1 vector field in U. Then are equivalents

- f is a irrotational vector field;
- f is a conservative vector field on U

The proof of this theorem is presented in the Section 7.7.1.

The above statement is *not* true in general if U is not simply connected as we have already seen in the example 6.4.3.

6.6 Conservative Fields

Definition 6.6.1 A force field F, defined everywhere in space (or within a simply-connected domain), is called a **conservative force** or conservative vector field if the curl of $\mathbf{f}$ is zero: $\nabla \times \mathbf{f} = \mathbf{0}$.

As a particle moves through a force field along a path γ, the work done by the force is given by the line integral

$$W = \int_\gamma \mathbf{f} \cdot \mathbf{dr}$$

By the parametrization invariance this value is independent of how the particle travels along the path. And for a conservative force field, it is also independent of the path itself, and depends only on the starting and ending points. We observe also that if the field is conservative, the work done can be more easily evaluated by realizing that a conservative vector field can be written as the gradient of some scalar potential function:

$$\mathbf{f} = \nabla \phi$$

The work done is then simply the difference in the value of this potential in the starting and end points of the path. If these points are given by x = a and x = b, respectively:

$$W = \phi(b) - \phi(a)$$

The function ϕ is called the potential energy.

Therefore, if the starting and ending points are the same, the work is zero for a conservative field:

$$\oint_\gamma \mathbf{f}\cdot \mathrm{d}\mathbf{r} = 0$$

6.6.1 Work and potential energy

Definition 6.6.2 — Work and potential energy. If $\mathbf{F}(\mathbf{r})$ is a force, then $\int_C \mathbf{F}\cdot \mathrm{d}\boldsymbol{\ell}$ is the *work done* by the force along the curve C. It is the limit of a sum of terms $\mathbf{F}(\mathbf{r})\cdot\delta\mathbf{r}$, ie. the force along the direction of $\delta\mathbf{r}$.

Consider a point particle moving under $\mathbf{F}(\mathbf{r})$ according to Newton's second law: $\mathbf{F}(\mathbf{r}) = m\ddot{\mathbf{r}}$. Since the kinetic energy is defined as

$$T(t) = \frac{1}{2}m\dot{\mathbf{r}}^2,$$

the rate of change of energy is

$$\frac{d}{dt}T(t) = m\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}} = \mathbf{F}\cdot\dot{\mathbf{r}}.$$

Suppose the path of particle is a curve C from $\mathbf{a} = \mathbf{r}(\alpha)$ to $\mathbf{b} = \mathbf{r}(\beta)$, Then

$$T(\beta) - T(\alpha) = \int_\alpha^\beta \frac{dT}{dt}\,dt = \int_\alpha^\beta \mathbf{F}\cdot\dot{\mathbf{r}}\,dt = \int_C \mathbf{F}\cdot d\boldsymbol{\ell}.$$

So the work done on the particle is the change in kinetic energy.

Definition 6.6.3 — Potential energy. Given a conservative force $\mathbf{F} = -\nabla V$, $V(\mathbf{x})$ is the *potential energy*. Then

$$\int_C \mathbf{F}\cdot d\boldsymbol{\ell} = V(\mathbf{a}) - V(\mathbf{b}).$$

Therefore, for a conservative force, we have $\mathbf{F} = \nabla V$, where $V(\mathbf{r})$ is the potential energy.

So the work done (gain in kinetic energy) is the loss in potential energy. So the total energy $T + V$ is conserved, ie. constant during motion.

We see that energy is conserved for conservative forces. In fact, the converse is true — the energy is conserved only for conservative forces.

6.7 The Second Fundamental Theorem

The gradient theorem states that if the vector field $\mathbf{f}$ is the gradient of some scalar-valued function, then $\mathbf{f}$ is a path-independent vector field. This theorem has a powerful converse:

Theorem 6.7.1 Suppose $U \subseteq \mathbb{R}^n$ is a domain of $\mathbb{R}^n$. If $\mathbf{f}$ is a path-independent vector field in U, then $\mathbf{f}$ is the gradient of some scalar-valued function.

It is straightforward to show that a vector field is path-independent if and only if the integral of the vector field over every closed loop in its domain is zero. Thus the converse can alternatively be stated as follows: If the integral of $\mathbf{f}$ over every closed loop in the domain of $\mathbf{f}$ is zero, then $\mathbf{f}$ is the gradient of some scalar-valued function.

Proof:

Suppose U is an open, path-connected subset of $\mathbb{R}^n$, and $\mathbf{F} : U \to \mathbb{R}^n$ is a continuous and path-independent vector field. Fix some point $\mathbf{a}$ of U, and define $f : U \to \mathbb{R}$ by

$$f(\mathbf{x}) := \int_{\gamma[\mathbf{a},\mathbf{x}]} \mathbf{F}(\mathbf{u}) \bullet d\mathbf{u}$$

Here $\gamma[\mathbf{a},\mathbf{x}]$ is any differentiable curve in U originating at $\mathbf{a}$ and terminating at $\mathbf{x}$. We know that f is well-defined because $\mathbf{f}$ is path-independent.

Let $\mathbf{v}$ be any nonzero vector in $\mathbb{R}^n$. By the definition of the directional derivative,

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}) = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} \tag{6.2}$$

$$= \lim_{t\to 0} \frac{\displaystyle\int_{\gamma[\mathbf{a},\mathbf{x}+t\mathbf{v}]} \mathbf{F}(\mathbf{u}) \bullet d\mathbf{u} - \int_{\gamma[\mathbf{a},\mathbf{x}]} \mathbf{F}(\mathbf{u}) \bullet d\mathbf{u}}{t} \tag{6.3}$$

$$= \lim_{t\to 0} \frac{1}{t} \int_{\gamma[\mathbf{x},\mathbf{x}+t\mathbf{v}]} \mathbf{F}(\mathbf{u}) \bullet d\mathbf{u} \tag{6.4}$$

To calculate the integral within the final limit, we must parametrize $\gamma[\mathbf{x},\mathbf{x}+t\mathbf{v}]$. Since $\mathbf{f}$ is path-independent, U is open, and t is approaching zero, we may assume that this path is a straight line, and parametrize it as $\mathbf{u}(s) = \mathbf{x} + s\mathbf{v}$ for $0 < s < t$. Now, since $\mathbf{u}'(s) = \mathbf{v}$, the limit becomes

$$\lim_{t\to 0} \frac{1}{t} \int_0^t \mathbf{F}(\mathbf{u}(s)) \bullet \mathbf{u}'(s)\ ds = \frac{d}{dt} \int_0^t \mathbf{F}(\mathbf{x}+s\mathbf{v}) \bullet \mathbf{v}\ ds \bigg|_{t=0} = \mathbf{F}(\mathbf{x}) \bullet \mathbf{v}$$

Thus we have a formula for $\partial_{\mathbf{v}} f$, where $\mathbf{v}$ is arbitrary.. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right) = \mathbf{F}(\mathbf{x})$$

Thus we have found a scalar-valued function f whose gradient is the path-independent vector field $\mathbf{f}$, as desired.

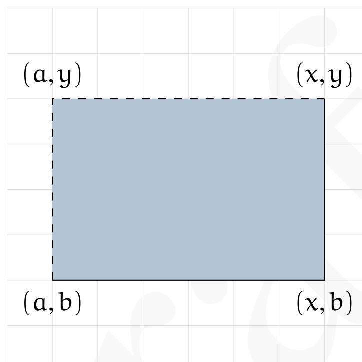
■

6.8 Constructing Potentials Functions

If $\mathbf{f}$ is a conservative field on an open connected set U, the line integral of $\mathbf{f}$ is independent of the path in U. Therefore we can find a potential simply by integrating $\mathbf{f}$ from some fixed point $\mathbf{a}$ to an arbitrary point $\mathbf{x}$ in U, using any piecewise smooth path lying in U. The scalar field so obtained depends on the choice of the initial point a. If we start from another initial point, say b, we obtain a new potential. But, because of the additive property of line integrals, and can differ only by a constant, this constant being the integral of $\mathbf{f}$ from $\mathbf{a}$ to $\mathbf{b}$.

Construction of a potential on an open rectangle.

If $\mathbf{f}$ is a conservative vector field on an open rectangle in $\mathbb{R}^n$, a potential f can be constructed by integrating from a fixed point to an arbitrary point along a set of line segments parallel to the coordinate axes.



We will simplify the deduction, assuming that $n = 2$. In this case we can integrate first from (a, b) to (x, b) along a horizontal segment, then from (x, b) to (x,y) along a vertical segment. Along the horizontal segment we use the parametric representation

$$\gamma(t) = t\mathbf{i} + b\mathbf{j}, a <, t <, x,$$

and along the vertical segment we use the parametrization

$$\gamma_2(t) = x\mathbf{i} + t\mathbf{j}, b < t < y.$$

If $F(x, y) = F_1(x, y)\mathbf{i} + F_2(x, y)\mathbf{j}$, the resulting formula for a potential $f(x, y)$ is

$$f(x, y) = \int_a^b F_1(t, b)\ dt + \int_b^y F_2(x, t)\ dt.$$

We could also integrate first from (a, b) to (a, y) along a vertical segment and then from (a, y) to (x, y) along a horizontal segment as indicated by the dotted lines in Figure. This gives us another formula for f(x, y),

$$f(x, y) = \int_b^y F_2(a, t)\ dt + \int_a^x F_2(t, y)\ dt.$$

Both formulas give the same value for $f(x,y)$ because the line integral of a gradient is independent of the path.

Construction of a potential using anti-derivatives

But there's another way to find a potential of a conservative vector field: you use the fact that $\frac{\partial V}{\partial x} = F_x$ to conclude that $V(x,y)$ must be of the form $\int_a^x F_x(u,y)\,du + G(y)$, and similarly $\frac{\partial V}{\partial y} = F_y$ implies that $V(x,y)$ must be of the form $\int_b^y F_y(x,v)\,du + H(x)$. So you find functions $G(y)$ and $H(x)$ such that $\int_a^x F_x(u,y)\,du + G(y) = \int_b^y F_y(x,v)\,du + H(x)$

Example 6.8.1 Show that

$$\mathbf{F} = (e^x \cos y + yz)\mathbf{i} + (xz - e^x \sin y)\mathbf{j} + (xy + z)\mathbf{k}$$

is conservative over its natural domain and find a potential function for it.

Solution: ►

The natural domain of F is all of space, which is connected and simply connected. Let's define the following:

$$M = e^x \cos y + yz$$
$$N = xz - e^x \sin y$$
$$P = xy + z$$

and calculate

$$\frac{\partial P}{\partial x} = y = \frac{\partial M}{\partial z}$$
$$\frac{\partial P}{\partial y} = x = \frac{\partial N}{\partial z}$$
$$\frac{\partial N}{\partial x} = -e^x \sin y = \frac{\partial M}{\partial y}$$

Because the partial derivatives are continuous, F is conservative. Now that we know there exists a function f where the gradient is equal to F, let's find f.

$$\frac{\partial f}{\partial x} = e^x \cos y + yz$$
$$\frac{\partial f}{\partial y} = xz - e^x \sin y$$
$$\frac{\partial f}{\partial z} = xy + z$$

If we integrate the first of the three equations with respect to x, we find that

$$f(x,y,z) = \int (e^x \cos y + yz)dx = e^x \cos y + xyz + g(y,z)$$

where g(y,z) is a constant dependant on y and z variables. We then calculate the partial derivative with respect to y from this equation and match it with the equation of above.

$$\frac{\partial}{\partial y}(f(x,y,z)) = -e^x \sin y + xz + \frac{\partial g}{\partial y} = xz - e^x \sin y$$

This means that the partial derivative of g with respect to y is 0, thus eliminating y from g entirely and leaving at as a function of z alone.

$$f(x,y,z) = e^x \cos y + xyz + h(z)$$

We then repeat the process with the partial derivative with respect to z.

$$\frac{\partial}{\partial z}(f(x,y,z)) = xy + \frac{dh}{dz} = xy + z$$

which means that

$$\frac{dh}{dz} = z$$

so we can find h(z) by integrating:

$$h(z) = \frac{z^2}{2} + C$$

Therefore,

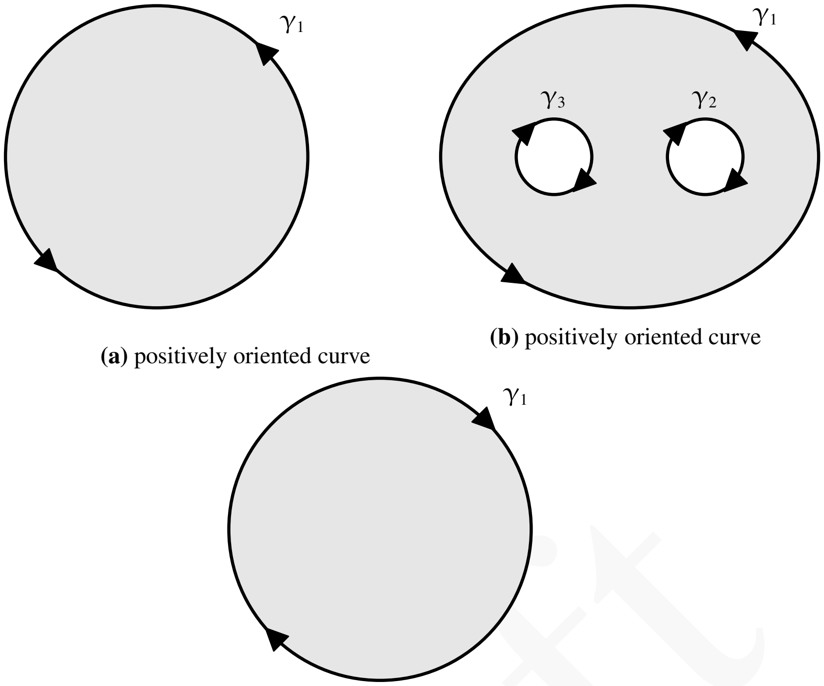
$$f(x,y,z) = e^x \cos y + xyz + \frac{z^2}{2} + C$$

We still have infinitely many potential functions for F, one at each value of C. ◀

6.9 Green's Theorem in the Plane

Definition 6.9.1 A **positively oriented curve** is a planar simple closed curve such that when travelling on it one always has the curve interior to the left. If in the previous definition one interchanges left and right, one obtains a **negatively oriented curve**.

We will now see a way of evaluating the line integral of a *smooth* vector field around a simple closed curve. A vector field $\mathbf{f}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$ is **smooth** if its component functions $P(x,y)$ and $Q(x,y)$ are smooth. We will use *Green's Theorem* (sometimes called *Green's Theorem in the plane*) to relate the *line* integral around a closed curve with a *double* integral over the region inside the curve:



(a) positively oriented curve

(b) positively oriented curve

(c) negatively oriented curve

Figure 6.5 Orientations of Curves

Theorem 6.9.1 — Green's Theorem - Simple Version. Let Ω be a region in $\mathbb{R}^2$ whose boundary is a simple closed curve γ which is piecewise smooth. Let $\mathbf{f}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$ be a smooth vector field defined on both Ω and γ. Then

$$\oint_{\gamma} \mathbf{f}\cdot d\mathbf{r} = \iint\limits_{\Omega} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA\,, \tag{6.5}$$

where γ is traversed so that Ω is always on the left side of γ.

Proof: We will prove the theorem in the case for a *simple* region Ω, that is, where the boundary curve γ can be written as $C = \gamma_1 \cup \gamma_2$ in two distinct ways:

$$\gamma_1 = \text{the curve } y = y_1(x) \text{ from the point } X_1 \text{ to the point } X_2 \tag{6.6}$$

$$\gamma_2 = \text{the curve } y = y_2(x) \text{ from the point } X_2 \text{ to the point } X_1, \tag{6.7}$$

where X_1 and X_2 are the points on C farthest to the left and right, respectively; and

$$\gamma_1 = \text{the curve } x = x_1(y) \text{ from the point } Y_2 \text{ to the point } Y_1 \tag{6.8}$$

$$\gamma_2 = \text{the curve } x = x_2(y) \text{ from the point } Y_1 \text{ to the point } Y_2, \tag{6.9}$$

where Y_1 and Y_2 are the lowest and highest points, respectively, on γ. See Figure

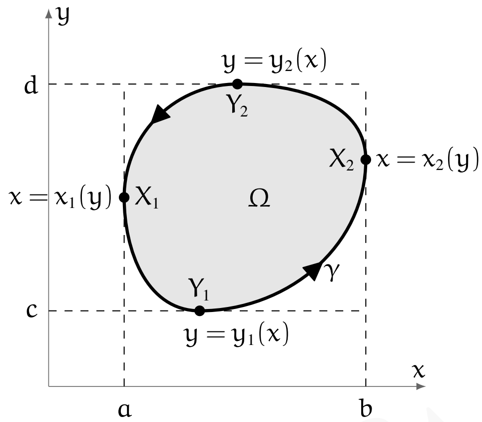


Figure 6.6

Integrate $P(x,y)$ around γ using the representation $\gamma = \gamma_1 \cup \gamma_2$ Since $y = y_1(x)$ along γ_1 (as x goes from a to b) and $y = y_2(x)$ along γ_2 (as x goes from b to a), as we see from Figure, then we have

$$\begin{aligned}
\oint_\gamma P(x,y)\,dx &= \int_{\gamma_1} P(x,y)\,dx + \int_{\gamma_2} P(x,y)\,dx \\
&= \int_a^b P(x,y_1(x))\,dx + \int_b^a P(x,y_2(x))\,dx \\
&= \int_a^b P(x,y_1(x))\,dx - \int_a^b P(x,y_2(x))\,dx \\
&= -\int_a^b \big(P(x,y_2(x)) - P(x,y_1(x))\big)\,dx \\
&= -\int_a^b \left(P(x,y)\Big|_{y=y_1(x)}^{y=y_2(x)} \right) dx \\
&= -\int_a^b \int_{y_1(x)}^{y_2(x)} \frac{\partial P(x,y)}{\partial y}\,dy\,dx \quad \text{(by the Fundamental Theorem of Calculus)} \\
&= -\iint_\Omega \frac{\partial P}{\partial y}\,dA\,.
\end{aligned}$$

Likewise, integrate $Q(x,y)$ around γ using the representation $\gamma = \gamma_1 \cup \gamma_2$. Since $x = x_1(y)$ along γ_1 (as y goes from d to c) and $x = x_2(y)$ along γ_2 (as y goes from c to d), as we see from Figure , then we have

$$
\begin{aligned}
\oint_\gamma Q(x,y)\,dy &= \int_{\gamma_1} Q(x,y)\,dy + \int_{\gamma_2} Q(x,y)\,dy \\
&= \int_d^c Q(x_1(y),y)\,dy + \int_c^d Q(x_2(y),y)\,dy \\
&= -\int_c^d Q(x_1(y),y)\,dy + \int_c^d Q(x_2(y),y)\,dy \\
&= \int_c^d \big(Q(x_2(y),y) - Q(x_1(y),y)\big)\,dy \\
&= \int_c^d \left(Q(x,y)\Big|_{x=x_1(y)}^{x=x_2(y)} \right) dy \\
&= \int_c^d \int_{x_1(y)}^{x_2(y)} \frac{\partial Q(x,y)}{\partial x}\,dx\,dy \quad \text{(by the Fundamental Theorem of Calculus)} \\
&= \iint_\Omega \frac{\partial Q}{\partial x}\,dA \text{ , and so}
\end{aligned}
$$

$$
\begin{aligned}
\oint_\gamma \mathbf{f}\bullet d\mathbf{r} &= \oint_\gamma P(x,y)\,dx + \oint_\gamma Q(x,y)\,dy \\
&= -\iint_\Omega \frac{\partial P}{\partial y}\,dA + \iint_\Omega \frac{\partial Q}{\partial x}\,dA \\
&= \iint_\Omega \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \,.
\end{aligned}
$$

■

The Green Theorem can be generalized:

Theorem 6.9.2 — Green's Theorem. Let $\Omega \subseteq \mathbb{R}^2$ be a bounded domain whose exterior boundary is a piecewise C^1 curve γ. If Ω has holes, let $\gamma_1, \ldots, \gamma_N$ be the interior boundaries. If $\mathbf{f} : \bar{\Omega} \to \mathbb{R}^2$ is C^1, then

$$\int_\Omega [\partial_1 F_2 - \partial_2 F_1]\,dA = \oint_\gamma \mathbf{f}\bullet d\ell + \sum_{i=1}^N \oint_{\gamma_i} \mathbf{f}\bullet d\ell,$$

where all line integrals above are computed by traversing the exterior boundary **counter clockwise**, and every interior boundary **clockwise**.

R A common convention is to denote the **boundary** of Ω by $\partial\Omega$ and write

$$\partial\Omega = \gamma \cup \left[\bigcup_{i=1}^{N} \gamma_i\right].$$

Then Theorem 6.9.2 becomes

$$\int_{\Omega} [\partial_1 F_2 - \partial_2 F_1]\, dA = \oint_{\partial\Omega} \mathbf{f} \bullet d\ell,$$

where again the exterior boundary is oriented **counter clockwise** and the interior boundaries are all oriented **clockwise**.

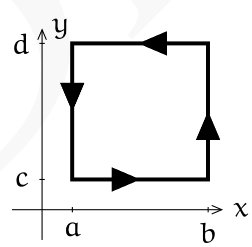
R In the differential form notation, Green's theorem is stated as

$$\int_{\Omega} \left[\partial_x Q - \partial_y P\right] dA = \int_{\partial\Omega} P\ dx + Q\ dy,$$

$P, Q : \bar{\Omega} \to \mathbb{R}$ are C^1 functions. (We use the same assumptions as before on the domain Ω, and orientations of the line integrals on the boundary.)

R Note, Green's theorem requires that Ω is bounded and **f** (or P and Q) is C^1 on *all* of Ω. If this fails at even one point, Green's theorem need not apply anymore!

Proof: The full proof is a little cumbersome. But the main idea can be seen by first proving it when Ω is a square. Indeed, suppose first $\Omega = (0,1)^2$.

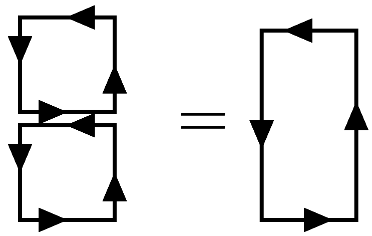


Then the fundamental theorem of calculus gives

$$\int_{\Omega} [\partial_1 F_2 - \partial_2 F_1]\, dA = \int_{y=0}^{1} \left[F_2(1,y) - F_2(0,y)\right]\ dy - \int_{x=0}^{1} \left[F_1(x,1) - F_1(x,0)\right]\ dx$$

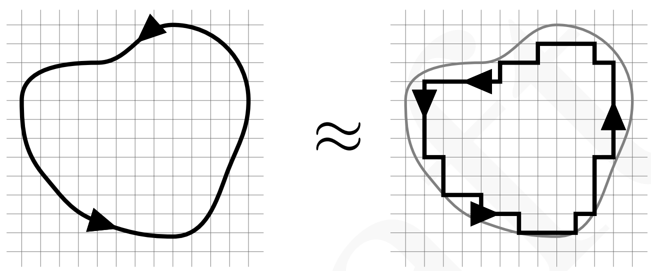
The first integral is the line integral of **f** on the two vertical sides of the square, and the second one is line integral of **f** on the two horizontal sides of the square. This proves Theorem 6.9.2 in the case when Ω is a square.

For line integrals, when adding two rectangles with a common edge the common edges are traversed in opposite directions so the sum is just the line integral over the outside boundary.

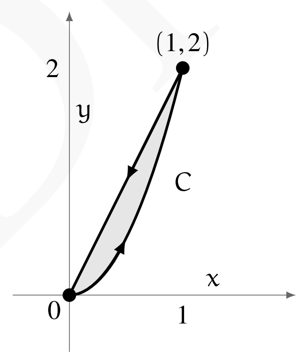


Similarly when adding a lot of rectangles: everything cancels except the outside boundary. This extends Green's Theorem on a rectangle to Green's Theorem on a sum of rectangles. Since any region can be approximated as closely as we want by a sum of rectangles, Green's Theorem must hold on arbitrary regions.

■



Example 6.9.3 Evaluate $\oint_C (x^2+y^2)\,dx + 2xy\ \ dy$, where C is the boundary traversed counterclockwise of the region $R = \{(x,y) : 0 \leqslant x \leqslant 1,\ 2x^2 \leqslant y \leqslant 2x\}$.



Solution: ► R is the shaded region in Figure above. By Green's Theorem, for $P(x,y) = x^2+y^2$ and $Q(x,y) = 2xy$, we have

$$\begin{aligned}\oint_C (x^2+y^2)\,dx + 2xy\ \ dy &= \iint_\Omega \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA \\ &= \iint_\Omega (2y-2y)\,dA = \iint_\Omega 0\,dA = 0\,.\end{aligned}$$

There is another way to see that the answer is zero. The vector field $\mathbf{f}(x,y) = (x^2+y^2)\,\mathbf{i} + 2xy\,\mathbf{j}$ has a potential function $F(x,y) = \frac{1}{3}x^3 + xy^2$, and so $\oint_C \mathbf{f}\cdot d\mathbf{r} = 0$. ◄

Example 6.9.4 Let $\mathbf{f}(x,y) = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$, where

$$P(x,y) = \frac{-y}{x^2+y^2} \quad \text{and} \quad Q(x,y) = \frac{x}{x^2+y^2},$$

and let $R = \{(x,y) : 0 < x^2+y^2 \leqslant 1\}$. For the boundary curve $C : x^2+y^2 = 1$, traversed counterclockwise, it was shown in Exercise 9(b) in Section 4.2 that $\oint_C \mathbf{f}\cdot d\mathbf{r} = 2\pi$. But

$$\frac{\partial Q}{\partial x} = \frac{y^2-x^2}{(x^2+y^2)^2} = \frac{\partial P}{\partial y} \Rightarrow \iint_\Omega \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA = \iint_\Omega 0\, dA = 0\,.$$

This would seem to contradict Green's Theorem. However, note that R is not the *entire* region enclosed by C, since the point $(0,0)$ is not contained in R. That is, R has a "hole" at the origin, so Green's Theorem does not apply.

6.10 Application of Green's Theorem: Area

Green's theorem can be used to compute area by line integral. Let C be a positively oriented, piecewise smooth, simple closed curve in a plane, and let U be the region bounded by C The area of domain U is given by $A = \iint_U dA$.

Then if we choose P and M such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$, the area is given by

$$A = \oint_C (P\ dx + Q\ dy).$$

Possible formulas for the area of U include:

$$A = \oint_C x\ dy = -\oint_C y\ dx = \frac{1}{2}\oint_C (-y\ dx + x\ dy).$$

Corollary 6.10.1 Let $\Omega \subseteq \mathbb{R}^2$ be bounded set with a C^1 boundary $\partial\Omega$, then

$$\text{area}\,(\Omega) = \frac{1}{2}\int_{\partial\Omega} [-y\ dx + x\ dy] = \int_{\partial\Omega} -y\ dx = \int_{\partial\Omega} x\ dy$$

Example 6.10.2 Use Green's Theorem to calculate the area of the disk D of radius r.

Solution: ► The boundary of D is the circle of radius r:

$$C(t) = (r\cos t, r\sin t), \quad 0 \leqslant t \leqslant 2\pi.$$

Then

$$C'(t) = (-r\sin t, r\cos t),$$

and, by Corollary 6.10.1,

$$\begin{aligned}
\text{area of } D &= \iint dA \\
&= \frac{1}{2}\int_C x\,dy - y\,dx \\
&= \frac{1}{2}\int_0^{2\pi} [(r\cos t)(r\cos t) - (r\sin t)(-r\sin t)]dt \\
&= \frac{1}{2}\int_0^{2\pi} r^2(\sin^2 t + \cos^2 t)dt = \frac{r^2}{2}\int_0^{2\pi} dt = \pi r^2.
\end{aligned}$$

◄

Example 6.10.3 Use the Green's theorem for computing the area of the region bounded by the x -axis and the arch of the cycloid:

$$x = t - \sin(t), \quad y = 1 - \cos(t), \quad 0 \leqslant t \leqslant 2\pi$$

Solution: ►

$$\text{Area}(D) = \iint_D dA = \oint_C -y\,dx.$$

Along the x-axis, you have $y = 0$, so you only need to compute the integral over the arch of the cycloid. Note that your parametrization of the arch is a clockwise parametrization, so in the following calculation, the answer will be the minus of the area:

$$\int_0^{2\pi} (\cos(t) - 1)(1 - \cos(t))dt = -\int_0^{2\pi} 1 - 2\cos(t) + \cos^2(t)dt = -3\pi.$$

◄

Corollary 6.10.4 — Surveyor's Formula. Let $P \subseteq \mathbb{R}^2$ be a (not necessarily convex) polygon whose vertices, ordered counter clockwise, are $(x_1, y_1), \ldots, (x_N, y_N)$. Then

$$\text{area}\,(P) = \frac{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + \cdots + (x_Ny_1 - x_1y_N)}{2}.$$

Proof: Let P be the set of points belonging to the polygon. We have that

$$A = \iint_P dx\ dy.$$

Using the Corollary 6.10.1 we have

$$\iint_P dx dy = \int_{\partial P} \frac{x\,dy}{2} - \frac{y\,dx}{2}.$$

We can write $\partial P = \bigcup_{i=1}^{n} L(i)$, where $L(i)$ is the line segment from (x_i, y_i) to (x_{i+1}, y_{i+1}). With this notation, we may write

$$\int_{\partial P} \frac{x\,dy}{2} - \frac{y\,dx}{2} = \sum_{i=1}^{n} \int_{A(i)} \frac{x\,dy}{2} - \frac{y\,dx}{2} = \frac{1}{2}\sum_{i=1}^{n} \int_{A(i)} x\,dy - y\,dx.$$

Parameterizing the line segment, we can write the integrals as

$$\frac{1}{2}\sum_{i=1}^{n} \int_0^1 (x_i + (x_{i+1} - x_i)t)(y_{i+1} - y_i) - (y_i + (y_{i+1} - y_i)t)(x_{i+1} - x_i)\,dt.$$

Integrating we get

$$\frac{1}{2}\sum_{i=1}^{n} \frac{1}{2}[(x_i + x_{i+1})(y_{i+1} - y_i) - (y_i + y_{i+1})(x_{i+1} - x_i)].$$

simplifying yields the result

$$\text{area}\,(P) = \frac{1}{2}\sum_{i=1}^{n} (x_i y_{i+1} - x_{i+1} y_i).$$

■

6.11 Vector forms of Green's Theorem

Theorem 6.11.1 — Stokes' Theorem in the Plane. Let $\mathbf{F} = L\mathbf{i} + M\mathbf{j}$. Then

$$\oint_\gamma \mathbf{F}\cdot d\ell = \int_\Omega \nabla\times\mathbf{F}\cdot d\mathbf{S}$$

Proof:

$$\nabla\times\mathbf{F} = \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\right)\hat{k}$$

Over the region R we can write $dx\,dy = dS$ and $d\mathbf{S} = \hat{k}dS$. Thus using Green's Theorem:

$$\begin{aligned}\oint_\gamma \mathbf{F}\cdot d\ell &= \int_\Omega \hat{k}\cdot\nabla\times\mathbf{F}\,dS\\ &= \int_\Omega \nabla\times\mathbf{F}\cdot d\mathbf{S}\end{aligned}$$

■

Theorem 6.11.2 — Divergence Theorem in the Plane. . Let $\mathbf{F} = M\mathbf{i} - L\mathbf{j}$ Then

$$\int_R \nabla \bullet \mathbf{F}\, dx\, dy = \oint_\gamma \mathbf{F} \cdot \hat{n}\, ds$$

Proof:

$$\nabla \bullet \mathbf{F} = \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}$$

and so Green's theorem can be rewritten as

$$\int_\Omega \nabla \bullet \mathbf{F}\, dx\, dy = \oint_\gamma F_1\, dy - F_2\, dx$$

Now it can be shown that

$$\hat{n}\, ds = (dy\mathbf{i} - dx\mathbf{j})$$

here s is arclength along C, and $\hat{n}$ is the unit normal to C. Therefore we can rewrite Green's theorem as

$$\int_R \nabla \bullet \mathbf{F}\, dx\, dy = \oint_\gamma \mathbf{F} \cdot \hat{n}\, ds$$

■

Theorem 6.11.3 — Green's identities in the Plane. Let $\phi(x,y)$ and $\psi(x,y)$ be two scalar functions $\mathcal{C}^2$, defined in the open set $\Omega \subset \mathbb{R}^2$.

$$\oint_\gamma \phi \frac{\partial \psi}{\partial n}\, ds = \int_\Omega \phi \nabla^2 \psi + (\partial \phi) \cdot (\partial \psi)]\, dx\, dy$$

and

$$\oint_\gamma \left[\phi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \phi}{\partial n}\right] ds = \int_\tau (\phi \nabla^2 \psi - \psi \nabla^2 \phi)\, dx\, dy$$

Proof: If we use the divergence theorem:

$$\int_S \nabla \bullet \mathbf{F}\, dx\, dy = \oint_\gamma \mathbf{F} \cdot \hat{n}\, ds$$

then we can calculate down the corresponding Green identities. These are

$$\oint_\gamma \phi \frac{\partial \psi}{\partial n}\, ds = \int_\Omega \phi \nabla^2 \psi + (\partial \phi) \cdot (\partial \psi)]\, dx\, dy$$

and

$$\oint_{\gamma}\left[\phi\frac{\partial\psi}{\partial n}-\psi\frac{\partial\phi}{\partial n}\right]\,ds=\int_{\tau}(\phi\nabla^2\psi-\psi\nabla^2\phi)\,dx\,dy$$

■

7

CHAPTER

Surface Integrals

In this chapter we restrict our study to the case of surfaces in three-dimensional space. Similar results for manifolds in the n-dimensional space are presented in the chapter 13.

7.1 The Fundamental Vector Product

Definition 7.1.1 A **parametrized surface** is given by a one-to-one transformation $\mathbf{r}:\Omega\to\mathbb{R}^n$, where Ω is a domain in the plane $\mathbb{R}^2$. This amounts to being given three scalar functions, $x=x(u,v)$, $y=y(u,v)$ and $z=z(u,v)$ of two variables, u and v, say. The transformation is then given by

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

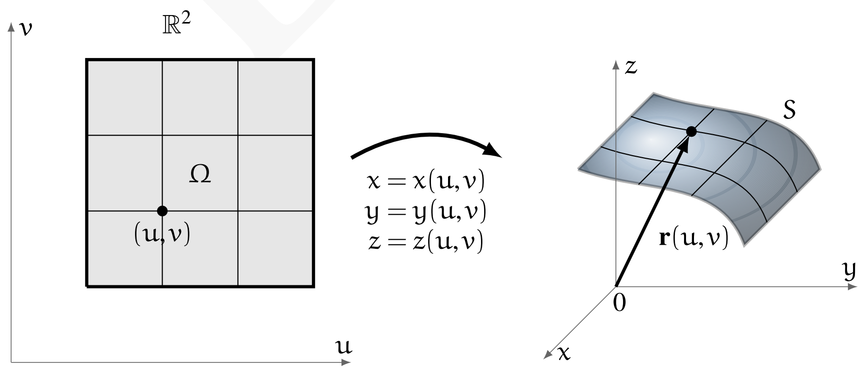


Figure 7.1 Parametrization of a surface S in $\mathbb{R}^3$

Definition 7.1.2

- A parametrization is said **regular** at the point (u_0, v_0) in Ω if

$$\partial_u \mathbf{r}(u_0, v_0) \times \partial_v \mathbf{r}(u_0, v_0) \neq \mathbf{0}.$$

- The parametrization is regular if its regular for all points in Ω.
- A surface that admits a regular parametrization is said **regular parametrized surface**.

Henceforth, we will assume that all surfaces are regular parametrized surface.

Now we consider two curves in S. The first one C_1 is given by the vector function

$$r_1(u) = \mathbf{r}(u, v_0), u \in (a, b)$$

obtained keeping the variable v fixed at v_0. The second curve C_2 is given by the vector function

$$r_2(u) = \mathbf{r}(u_0, v), v \in (c, d)$$

this time we are keeping the variable u fixed at u_0).

Both curves pass through the point $\mathbf{r}(u_0, v_0)$:

- The curve C_1 has tangent vector $r_1'(u_0) = \partial_u \mathbf{r}(u_0, v_0)$
- The curve C_2 has tangent vector $r_2'(v_0) = \partial_v \mathbf{r}(u_0, v_0)$.

The cross product $\mathbf{n}(u_0, v_0) = \partial_u \mathbf{r}(u_0, v_0) \times \partial_v \mathbf{r}'(u_0, v_0)$, which we have assumed to be different from zero, is thus perpendicular to both curves at the point $\mathbf{r}(u_0, v_0)$ and can be taken as a normal vector to the surface at that point.

We record the result as follows:

Definition 7.1.3 If S is a regular surface given by a differentiable function $\mathbf{r} = \mathbf{r}(u, v)$, then the cross product

$$\mathbf{n}(u, v) = \partial_u \mathbf{r} \times \partial_v \mathbf{r}$$

is called the **fundamental vector product** of the surface.

Example 7.1.1 For the plane $\mathbf{r}(u, v) = u\mathbf{a} + v\mathbf{b} + \mathbf{c}$ we have

$\partial_u \mathbf{r}(u, v) = \mathbf{a}$, $\partial_v \mathbf{r}(u, v) = \mathbf{b}$ and therefore $\hat{\mathbf{n}}(u, v) = \mathbf{a} \times \mathbf{b}$. The vector $\mathbf{a} \times \mathbf{b}$ is normal to the plane.

Example 7.1.2 We parametrized the sphere $x^2 + y^2 + z^2 = a^2$ by setting

$$\mathbf{r}(u, v) = a \cos u \sin v \mathbf{i} + a \sin u \sin v \mathbf{j} + a \cos v \mathbf{k},$$

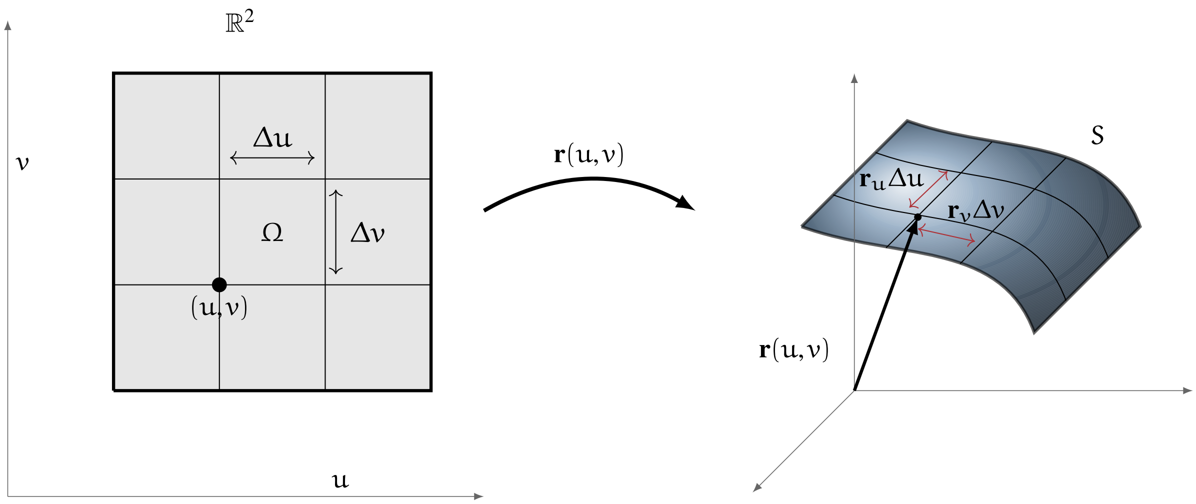


Figure 7.2 Parametrization of a surface S in $\mathbb{R}^3$

with $0 \leqslant u \leqslant 2\pi, 0 \leqslant v \leqslant \pi$. In this case

$$\partial_u \mathbf{r}(u, v) = -a \sin u \sin v \mathbf{i} + a \cos u \sin v \mathbf{j}$$

and

$$\partial_v \mathbf{r}(u, v) = a \cos u \cos v \mathbf{i} + a \sin u \cos v \mathbf{j} - a \sin v \mathbf{k}.$$

Thus

$$\begin{aligned} \mathbf{n}(u, v) &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -a \sin u \cos v & a \cos u \cos v & 0 \\ a \cos u \cos v & a \sin u \cos v & a \sin v \end{vmatrix} \\ &= -a \sin v (a \cos u \sin v \mathbf{i} + a \sin u \sin v \mathbf{j} + a \cos v \mathbf{k},) \\ &= a \sin v \mathbf{r}(u, v). \end{aligned}$$

As was to be expected, the fundamental vector product of a sphere is parallel to the radius vector $\mathbf{r}(u, v)$.

Example 7.1.3

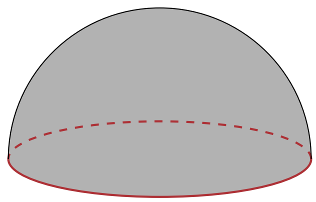
1. Take the sphere $f(\mathbf{r}) = x^2 + y^2 + z^2 = c$ for $c > 0$. Then $\nabla f = 2(x, y, z) = 2\mathbf{r}$, which is clearly normal to the sphere.
2. Take $f(\mathbf{r}) = x^2 + y^2 - z^2 = c$, which is a hyperboloid. Then $\nabla f = 2(x, y, -z)$.

   In the special case where $c = 0$, we have a double cone, with a singular apex $\mathbf{0}$. Here $\nabla f = \mathbf{0}$, and we cannot find a meaningful direction of normal.

Definition 7.1.4 — Boundary. A surface S can have a **boundary** ∂S. We are interested in the case where the boundary consist of a piecewise smooth curve or in a union of piecewise smooth curves.

A surface is **bounded** if it can be contained in a solid sphere of radius R, and is called *unbounded* otherwise. A bounded surface with no boundary is called **closed**.

Example 7.1.4 The boundary of a hemisphere is a circle (drawn in red).



Example 7.1.5 The sphere and the torus are examples of closed surfaces. Both are bounded and without boundaries.

7.2 The Area of a Parametrized Surface

We will now learn how to perform integration over a *surface* in $\mathbb{R}^3$.

Similar to how we used a parametrization of a curve to define the line integral along the curve, we will use a parametrization of a surface to define a *surface integral*. We will use *two* variables, u and v, to parametrize a surface S in $\mathbb{R}^3$: $x = x(u,v)$, $y = y(u,v)$, $z = z(u,v)$, for (u,v) in some region Ω in $\mathbb{R}^2$ (see Figure 7.3).

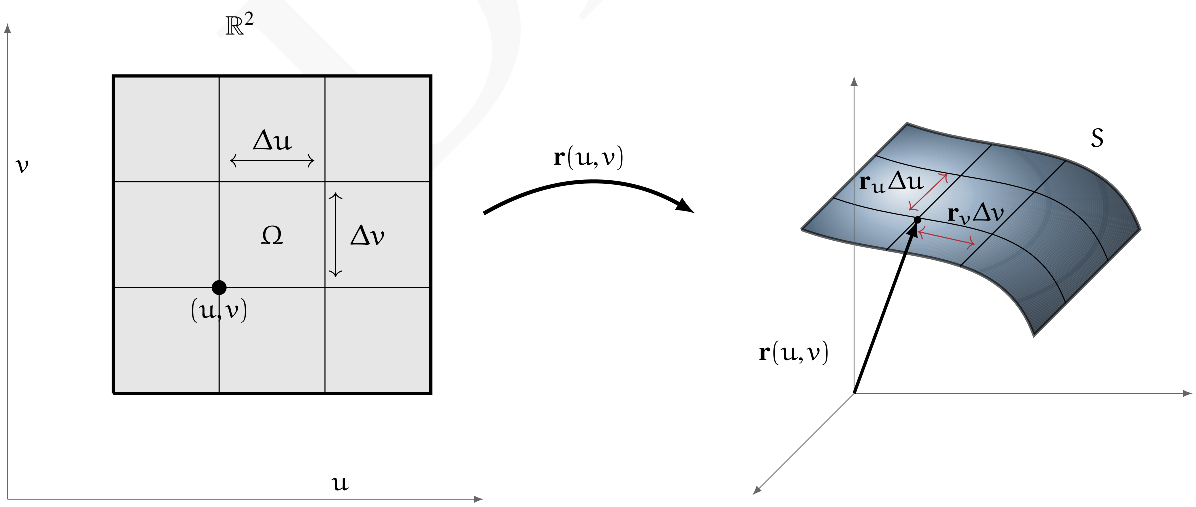


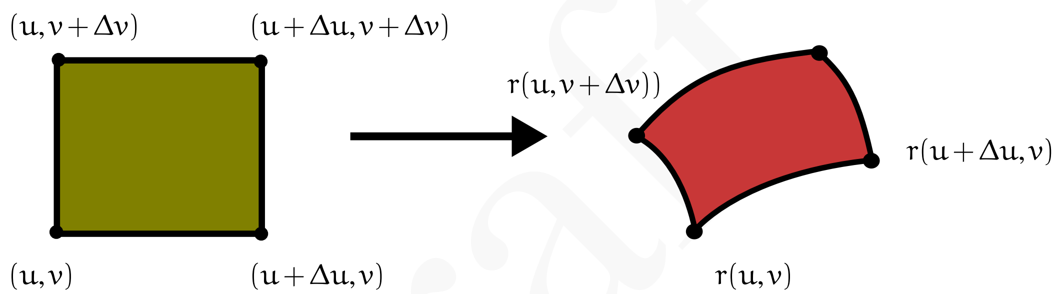
Figure 7.3 Parametrization of a surface S in $\mathbb{R}^3$

In this case, the position vector of a point on the surface S is given by the vector-valued function

$$\mathbf{r}(u,v) \;=\; x(u,v)\mathbf{i} \;+\; y(u,v)\mathbf{j} \;+\; z(u,v)\mathbf{k} \;\text{ for } (u,v) \text{ in } \Omega.$$

The parametrization of S can be thought of as "transforming" a region in $\mathbb{R}^2$ (in the uv-plane) into a 2-dimensional surface in $\mathbb{R}^3$. This parametrization of the surface is sometimes called a *patch*, based on the idea of "patching" the region Ω onto S in the grid-like manner shown in Figure 7.3.

In fact, those gridlines in Ω lead us to how we will define a surface integral over S. Along the vertical gridlines in Ω, the variable u is constant. So those lines get mapped to curves on S, and the variable u is constant along the position vector $\mathbf{r}(u,v)$. Thus, the tangent vector to those curves at a point (u,v) is $\dfrac{\partial \mathbf{r}}{\partial v}$. Similarly, the horizontal gridlines in Ω get mapped to curves on S whose tangent vectors are $\dfrac{\partial \mathbf{r}}{\partial u}$.



Now take a point (u,v) in Ω as, say, the lower left corner of one of the rectangular grid sections in Ω, as shown in Figure 7.3. Suppose that this rectangle has a small width and height of Δu and Δv, respectively. The corner points of that rectangle are (u,v), $(u+\Delta u,v)$, $(u+\Delta u,v+\Delta v)$ and $(u,v+\Delta v)$. So the area of that rectangle is $A = \Delta u\,\Delta v$. Then that rectangle gets mapped by the parametrization onto some section of the surface S which, for Δu and Δv small enough, will have a surface area (call it dS) that is very close to the area of the parallelogram which has adjacent sides $\mathbf{r}(u+\Delta u,v) - \mathbf{r}(u,v)$ (corresponding to the line segment from (u,v) to $(u+\Delta u,v)$ in Ω) and $\mathbf{r}(u,v+\Delta v) - \mathbf{r}(u,v)$ (corresponding to the line segment from (u,v) to $(u,v+\Delta v)$ in Ω). But by combining our usual notion of a partial derivative with that of the derivative of a vector-valued function applied to a function of two variables, we have

$$\begin{aligned}
\frac{\partial \mathbf{r}}{\partial u} &\approx \frac{\mathbf{r}(u+\Delta u,v) - \mathbf{r}(u,v)}{\Delta u}\,, \;\text{ and} \\
\frac{\partial \mathbf{r}}{\partial v} &\approx \frac{\mathbf{r}(u,v+\Delta v) - \mathbf{r}(u,v)}{\Delta v}\,,
\end{aligned}$$

and so the surface area element dS is approximately

$$\left\| (\mathbf{r}(u+\Delta u,v) - \mathbf{r}(u,v)) \times (\mathbf{r}(u,v+\Delta v) - \mathbf{r}(u,v)) \right\| \approx \left\| (\Delta u \frac{\partial \mathbf{r}}{\partial u}) \times (\Delta v \frac{\partial \mathbf{r}}{\partial v}) \right\| = \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| \Delta u\,\Delta v$$

Thus, the total surface area S of S is approximately the sum of all the quantities $\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| \Delta u \, \Delta v$, summed over the rectangles in Ω. Taking the limit of that sum as the diagonal of the largest rectangle goes to 0 gives

$$S = \iint_{\Omega} \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du \, dv \,. \tag{7.1}$$

We will write the double integral on the right using the special notation

$$\iint_{S} dS = \iint_{\Omega} \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du \, dv \,. \tag{7.2}$$

This is a special case of a *surface integral* over the surface S, where the surface area element dS can be thought of as $1\,dS$. Replacing 1 by a general real-valued function $f(x,y,z)$ defined in $\mathbb{R}^3$, we have the following:

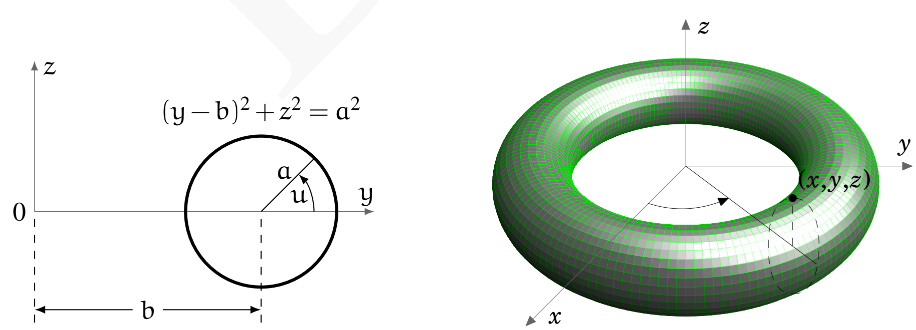
Definition 7.2.1 Let S be a surface in $\mathbb{R}^3$ parametrized by

$$x = x(u,v), \quad y = y(u,v), \quad z = z(u,v),$$

for (u,v) in some region Ω in $\mathbb{R}^2$. Let $\mathbf{r}(u,v) = x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k}$ be the position vector for any point on S. The surface area S of S is defined as

$$S = \iint_{S} 1\,dS = \iint_{\Omega} \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du \, dv \tag{7.3}$$

Example 7.2.1 A *torus* T is a surface obtained by revolving a circle of radius a in the yz-plane around the z-axis, where the circle's center is at a distance b from the z-axis ($0 < a < b$), as in Figure 7.4. Find the surface area of T.



(a) Circle in the yz-plane

(b) Torus T

Figure 7.4

Solution: ►

For any point on the circle, the line segment from the center of the circle to that point makes an angle u with the y-axis in the positive y direction (see Figure 7.4(a)). And as the circle revolves around the z-axis, the line segment from the origin to the center of that circle sweeps out an angle v with the positive x-axis (see Figure 7.4(b)). Thus, the torus can be parametrized as:

$$x = (b + a\cos u)\cos v\,, \quad y = (b + a\cos u)\sin v\,, \quad z = a\sin u\,, \quad 0 \leqslant u \leqslant 2\pi\,, \quad 0 \leqslant v \leqslant 2\pi$$

So for the position vector

$$\begin{aligned} \mathbf{r}(u,v) &= x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k} \\ &= (b + a\cos u)\cos v\,\mathbf{i} + (b + a\cos u)\sin v\,\mathbf{j} + a\sin u\,\mathbf{k} \end{aligned}$$

we see that

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial u} &= -a\sin u\cos v\,\mathbf{i} - a\sin u\sin v\,\mathbf{j} + a\cos u\,\mathbf{k} \\ \frac{\partial \mathbf{r}}{\partial v} &= -(b + a\cos u)\sin v\,\mathbf{i} + (b + a\cos u)\cos v\,\mathbf{j} + 0\mathbf{k}\,, \end{aligned}$$

and so computing the cross product gives

$$\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = -a(b + a\cos u)\cos v\cos u\,\mathbf{i} - a(b + a\cos u)\sin v\cos u\,\mathbf{j} - a(b + a\cos u)\sin u\,\mathbf{k}\,,$$

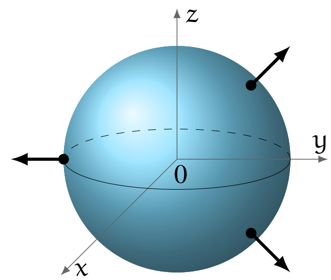
which has magnitude

$$\left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| = a(b + a\cos u)\,.$$

Thus, the surface area of T is

$$\begin{aligned} S &= \iint\limits_S 1\,dS \\ &= \int_0^{2\pi}\int_0^{2\pi} \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| du\,dv \\ &= \int_0^{2\pi}\int_0^{2\pi} a(b + a\cos u)\,du\,dv \\ &= \int_0^{2\pi} \left(abu + a^2\sin u \Big|_{u=0}^{u=2\pi} \right) dv \\ &= \int_0^{2\pi} 2\pi ab\,dv \\ &= 4\pi^2 ab \end{aligned}$$

◄



Example 7.2.2 [The surface area of a sphere] The function

$$\mathbf{r}(u,v) = a\cos u \sin v\mathbf{i} + a\sin u \sin v\mathbf{j} + a\cos v\mathbf{k},$$

with (u,v) ranging over the set $0 \leqslant u \leqslant 2\pi, 0 \leqslant v \leqslant \frac{\pi}{2}$ parametrizes a sphere of radius a. For this parametrization

$$\mathbf{n}(u,v) = a\sin v\,\mathbf{r}(u,v) \text{ and } \left\|\mathbf{n}(u,v)\right\| = a^2|\sin v| = a^2 \sin v.$$

So,

$$\begin{aligned}\text{area of the sphere} &= \iint_{\Omega} a^2 \sin v \; du \; dv \\ &= \int_0^{2\pi}\left(\int_0^{\pi} a^2 \sin v \; dv\right) du = 2\pi a^2 \int_0^{\pi} \sin v \; dv = 4\pi a^2,\end{aligned}$$

which is known to be correct.

Example 7.2.3 — The area of a region of the plane. If S is a plane region Ω, then S can be parametrized by setting

$$\mathbf{r}(u,v) = ui + vj, (u,v) \in \Omega.$$

Here $\mathbf{n}(u,v) = \partial_u \mathbf{r}(u,v) \times \partial_v \mathbf{r}(u,v) = i \times j = k$ and $\left\|\mathbf{n}(u,v)\right\| = 1$. In this case we reobtain the familiar formula

$$A = \iint_{\Omega} du \; dv.$$

Example 7.2.4 — The area of a surface of revolution. Let S be the surface generated by revolving the graph of a function

$$y = f(x), x \in [a,b]$$

about the x-axis. We will assume that f is positive and continuously differentiable.

We can parametrize S by setting

$$\mathbf{r}(u,v) = vi + f(v)\cos u\, j + f(v) \sin u\, k$$

with (u,v) ranging over the set $\Omega: 0 \leqslant u \leqslant 2\pi, a \leqslant v \leqslant b$. In this case

$$\mathbf{n}(u,v) = \partial_u \mathbf{r}(u,v) \times \partial_v \mathbf{r}(u,v) = \begin{vmatrix} i & j & k \\ 0 & -f(v)\sin u & f(v)\cos u \\ 1 & f'(v)\cos u & f'(v)\sin u \end{vmatrix}$$

$$= -f(v)f'(v)i + f(v)\cos u\, j + f(v)\sin u\, k.$$

Therefore $\|\mathbf{n}(u,v)\| = f(v)\sqrt{[f'(v)]^2+1}$ and

$$\text{area}(S) = \iint_\Omega f(v)\sqrt{[f'(v)]^2+1}\ du\ dv$$

$$\int_0^{2\pi}\left(\int_a^b f(v)\sqrt{[f'(v)]^2+1}\ dv\right) du = \int_a^b 2\pi f(v)\sqrt{[f'(v)]^2+1}\ dv.$$

Example 7.2.5 — Spiral ramp. One turn of the spiral ramp of Example 5 is the surface

$$S : \mathbf{r}(u,v) = u\cos\omega v\,\mathbf{i} + u\sin\omega v\,\mathbf{j} + bv\,\mathbf{k}$$

with (u,v) ranging over the set $\Omega : 0 \leqslant u \leqslant l, 0 \leqslant v \leqslant 2\pi/\omega$. In this case

$$\partial_u \mathbf{r}(u,v) = \cos\omega v\,\mathbf{i} + \sin\omega v\,\mathbf{j},\ \partial_v \mathbf{r}'(u,v) = -\omega u\sin\omega v\,\mathbf{i} + \omega u\cos\omega v\,\mathbf{j} + b\mathbf{k}.$$

Therefore

$$\mathbf{n}(u,v) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos\omega v & \sin\omega v & 0 \\ -\omega u\sin\omega v & \omega u\cos\omega v & b \end{vmatrix} = b\sin\omega v\,\mathbf{i} - b\cos\omega v\,\mathbf{j} + \omega u\mathbf{k}$$

and

$$\|\mathbf{n}(u,v)\| = \sqrt{b^2+\omega^2u^2}.$$

Thus

$$\text{area of } S = \iint_\Omega \sqrt{b^2+\omega^2u^2}\ du\ dv$$

$$= \int_0^{2\pi/\omega}\left(\int_0^l \sqrt{b^2+\omega^2u^2}du\right) dv = \frac{2\pi}{\omega}\int_0^l \sqrt{b^2+\omega^2u^2}du.$$

The integral can be evaluated by setting $u = (b/\omega)\tan x$.

7.2.1 The Area of a Graph of a Function

Let S be the surface of a function $f(x,y)$:

$$z = f(x,y), (x,y) \in \Omega.$$

We are to show that if f is continuously differentiable, then

$$\operatorname{area}(S) = \iint_{\Omega} \sqrt{\left[f_x'(x,y)\right]^2 + \left[f_y'(x,y)\right]^2 + 1}\; dx\, dy.$$

We can parametrize S by setting

$$\mathbf{r}(u,v) = u\mathbf{i} + v\mathbf{j} + f(u,v)\mathbf{k}, (u,v) \in \Omega.$$

We may just as well use x and y and write

$$\mathbf{r}(x,y) = x\mathbf{i} + y\mathbf{j} + f(x,y)\mathbf{k}, (x,y) \in \Omega.$$

Clearly

$$\mathbf{r}_x(x,y) = \mathbf{i} + f_x(x,y)\mathbf{k} \text{ and } \mathbf{r}_y(x,y) = \mathbf{j} + f_y(x,y)\mathbf{k}.$$

Thus

$$\mathbf{n}(x,y) = \begin{vmatrix} i & j & k \\ 1 & 0 & f_x(x,y) \\ 0 & 1 & f_y(x,y) \end{vmatrix} = -f_x(x,y)\,\mathbf{i} - f_y(x,y)\,\mathbf{j} + \mathbf{k}.$$

Therefore $\left\|\mathbf{n}(x,y)\right\| = \sqrt{\left[f_x'(x,y)\right]^2 + \left[f_y'(x,y)\right]^2 + 1}$ and the formula is verified.

Example 7.2.6 Find the surface area of that part of the parabolic cylinder $z = y^2$ that lies over the triangle with vertices $(0,0), (0,1), (1,1)$ in the xy-plane.

Solution: ►

Here $f(x,y) = y^2$ so that

$$f_x(x,y) = 0, f_y(x,y) = 2y.$$

The base triangle can be expressed by writing

$$\Omega : 0 \leqslant y \leqslant 1, 0 \leqslant x \leqslant y.$$

The surface has area

$$\begin{aligned} \text{area} &= \iint_{\Omega} \sqrt{\left[f_x'(x,y)\right]^2 + \left[f_y'(x,y)\right]^2 + 1}\; dx\, dy \\ &= \int_0^1 \int_0^y \sqrt{4y^2+1}\; dx\, dy \\ &= \int_0^1 y\sqrt{4y^2+1}\; dy = \frac{5\sqrt{5}-1}{12}. \end{aligned}$$

◄

Example 7.2.7 Find the surface area of that part of the hyperbolic paraboloid $z = xy$ that lies inside the cylinder $x^2 + y^2 = a^2$.

Solution: ► Let $f(x,y) = xy$ so that

$$f_x(x,y) = y, f_y(x,y) = x.$$

The formula gives

$$A = \iint_\Omega \sqrt{x^2+y^2+1}\ dx\ dy.$$

In polar coordinates the base region takes the form

$$0 \leqslant r \leqslant a, 0 \leqslant \theta \leqslant 2\pi.$$

Thus we have

$$\begin{aligned} A &= \iint_\Omega \sqrt{r^2+1}\, r dr d\theta = \int_0^{2\pi} \int_0^a \sqrt{r^2+1}\, r dr d\theta \\ &= \frac{2}{3}\pi[(a^2+1)^{3/2} - 1]. \end{aligned}$$

There is an elegant version of this last area formula that is geometrically vivid. We know that the vector

$$\mathbf{r}_x(x,y) \times \mathbf{r}_y(x,y) = -f_x(x,y)\mathbf{i} - f_y(x,y)\mathbf{j} + \mathbf{k}$$

is normal to the surface at the point $(x, y, f(x,y))$. The unit vector in that direction, the vector

$$\mathbf{n}(x,y) = \frac{-f_x(x,y)\mathbf{i} - f_y(x,y)\mathbf{j} + \mathbf{k}}{\sqrt{\left[f_x(x,y)\right]^2 + \left[f_y(x,y)\right]^2 + 1}},$$

is called the **upper unit normal** (It is the unit normal with a nonnegative k-component.)

Now let $\gamma(x,y)$ be the angle between $\mathbf{n}(x,y)$ and $\mathbf{k}$. Since $\mathbf{n}(x,y)$ and $\mathbf{k}$ are both unit vectors,

$$\cos[\gamma(x,y)] = \mathbf{n}(x,y) \bullet \mathbf{k} = \frac{1}{\sqrt{\left[f'_x(x,y)\right]^2 + \left[f'_y(x,y)\right]^2 + 1}}.$$

Taking reciprocals we have

$$\sec[\gamma(x,y)] = \sqrt{\left[f'_x(x,y)\right]^2 + \left[f'_y(x,y)\right]^2 + 1}.$$

The area formula can therefore be written

$$A = \iint_\Omega \sec[\gamma(x,y)]\ dx\ dy.$$

◄

7.3 Surface Integrals of Scalar Functions

We start with a motivation the Mass of a Material Surface.

7.3.1 The Mass of a Material Surface

Imagine a thin distribution of matter spread out over a surface S. We call this a *material surface.* If the mass density (the mass per unit area) is constant throughout, then the total mass of the material surface is the density λ times the area of S :

$$M = \lambda \text{ area of } S.$$

If, however, the mass density varies continuously from point to point, $\lambda = \lambda(x, y, z)$, then the total mass must be calculated by integration.

To develop the appropriate integral we suppose that

$$S : \mathbf{r} = \mathbf{r}(u, v) = x(u, v)i + y(u, v)j + z(u, v)k, (u, v) \in \Omega$$

is a continuously differentiable surface, a surface that meets the conditions for area formula. Our first step is to break up Ω into n little basic regions $\Omega_1, ..., \Omega_N$. This decomposes the surface into n little pieces $S_1, ..., S_n$. The area of S_i is given by the integral

$$\iint_{\Omega_i} \left\|\mathbf{n}(u, v)\right\| \, du \, dv.$$

By the mean-value theorem for double integrals, there exists a point (u_i^*, v_i^*) in Ω_i for which

$$\iint_{\Omega_i} \left\|\mathbf{n}(u, v)\right\| \, du \, dv = \left\|\mathbf{n}(u_i^*, v_i^*)\right\| (\text{area of } \Omega_i).$$

It follows that

$$\text{area of } S_i = \left\|\mathbf{n}(u_i^*, v_i^*)\right\| (\text{area of } \Omega_i).$$

Since the point (u_i^*, v_i^*) is in Ω_i, the tip of $\mathbf{r}(u_i^*, v_i^*)$ is on S_i. The mass density at this point is $\lambda[\mathbf{r}(u_i^*, v_i^*)]$. If S_i is small (which we can guarantee by choosing Ω_i, small), then the mass density on S_i is approximately the same throughout. Thus we can estimate M_i, the mass contribution of S_i, by writing

$$M_i \approx \lambda[\mathbf{r}(u_i^*, v_i^*)](\text{ area } (S)_i) = \lambda[\mathbf{r}(u_i^*, v_i^*)] \left\|\mathbf{n}(u_i^*, v_i^*)\right\| (\text{area of } \Omega_i)$$

Adding up these estimates we have an estimate for the total mass of the surface:

$$M \approx \sum_{i=1}^{N} \lambda[\mathbf{r}(u_i^*, v_i^*)] \left\|\mathbf{n}(u_i^*, v_i^*)\right\| (\text{area of } \Omega_i)$$

$$= \sum_{i=1}^{N} \lambda[x(u_i^*, v_i^*), y(u_i^*, v_i^*), z(u_i^*, v_i^*)] \left\| \mathbf{n}(u_i^*, v_i^*) \right\| (\text{area of } \Omega_i)$$

This last expression is a Riemann sum for

$$\iint_{\Omega} \lambda[x(u,v), y(u,v), z(u,v)] \left\| \mathbf{n}(u,v) \right\| \, du \, dv.$$

7.3.2 Surface Integrals

The above double integral can be calculated not only for a mass density function λ but for any scalar field f continuous over S. We call this integral the **surface integral** of f over S.

Definition 7.3.1 Let $S \subseteq \mathbb{R}^3$ be a surface, and $f : S \to \mathbb{R}$ be a continuous function. Define

$$\iint_S f \, dS = \lim_{\|P\| \to 0} \sum_{i=1}^{n} f(\xi_i) \, \text{area}(R_i),$$

where P is a partition of S into the regions $R_1, \ldots, R_n$, and $\|P\|$ is the diameter of the largest region R_i and $(\xi_i \in R_i$.

R Other common notation for the surface integral is

$$\int_S f \, dS = \iint_S f \, dS = \iint_{\Omega} f \, dS = \iint_{\Omega} f \, dA$$

As with line integrals, we obtained a formula in terms of a parametrization.

Proposition 7.3.1 If $\mathbf{r}(u,v) : U \to S$ is a regular parametrization of the surface S, then

$$\iint_S f \, dS = \int_U f(\mathbf{r}(u,v) \, |\partial_u \mathbf{r} \times \partial_v \mathbf{r}| \, dA.$$

R If the surface cannot be parametrized by a unique function, the integral can be computed by breaking up S into finitely many pieces which *can* be parametrized.

The formula above will yield an answer that is independent of the chosen parametrization and how you break up the surface (if necessary).

Example 7.3.2 Evaluate

$$\iint_S z dS$$

where S is the upper half of a sphere of radius 2.

Solution: ► As we already computed $\mathbf{n} =$ ◄

Example 7.3.3 Integrate the function $g(x,y,z) = yz$ over the surface of the wedge in the first octant bounded by the coordinate planes and the planes $x = 2$ and $y + z = 1$.

Solution: ► If a surface consists of many different pieces, then a surface integral over such a surface is the sum of the integrals over each of the surfaces.

The portions are S_1: $x = 0$ for $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1-y$; S_2: $x = 2$ for $0 \leqslant y \leqslant 1$, $\leqslant z \leqslant 1-y$; S_3: $y = 0$ for $0 \leqslant x \leqslant 2$, $0 \leqslant z \leqslant 1$; S_4: $z = 0$ for $0 \leqslant x \leqslant 2$, $0 \leqslant y \leqslant 1$; and S_5: $z = 1-y$ for $0 \leqslant x \leqslant 2$, $0 \leqslant y \leqslant 1$. Hence, to find $\iint_S g dS$, we must evaluate all 5 integrals. We compute $dS_1 = \sqrt{1+0+0}dzdy$, $dS_2 = \sqrt{1+0+0}dzdy$, $dS_3 = \sqrt{0+1+0}dzdx$, $dS_4 = \sqrt{0+0+1}dydx$, $dS_5 = \sqrt{0+(-1)^2+1}dydx$, and so

$$\begin{array}{lllllll} & \iint_{S_1} gdS & +\iint_{S_2} gdS & +\iint_{S_3} gdS & +\iint_{S_4} gdS & +\iint_{S_5} gdS \\ = & \int_0^1 \int_0^{1-y} yzdzdy & +\int_0^1 \int_0^{1-y} yzdzdy & +\int_0^2 \int_0^1 (0)zdzdx & +\int_0^2 \int_0^1 y(0)dydx & +\int_0^2 \int_0^1 y(1-y)\sqrt{2}dyd \\ = & \int_0^1 \int_0^{1-y} yzdzdy & +\int_0^1 \int_0^{1-y} yzdzdy & +0 & +0 & +\int_0^2 \int_0^1 y(1-y)\sqrt{2}dyd \\ = & 1/24 & +1/24 & +0 & +0 & +\sqrt{2}/3 \end{array}$$

◄

Example 7.3.4 The temperature at each point in space on the surface of a sphere of radius 3 is given by $T(x,y,z) = \sin(xy+z)$. Calculate the average temperature.

Solution: ►

The average temperature on the sphere is given by the surface integral

$$AV = \frac{1}{S}\iint_S fdS$$

A parametrization of the surface is

$$\mathbf{r}(\theta,\phi) = \langle 3\cos\theta\sin\phi, 3\sin\theta\sin\phi, 3\cos\phi\rangle$$

for $0 \leqslant \theta \leqslant 2\pi$ and $0 \leqslant \phi \leqslant \pi$. We have

$$T(\theta,\phi) = \sin((3\cos\theta\sin\phi)(3\sin\theta\sin\phi) + 3\cos\phi),$$

and the surface area differential is $dS = |\mathbf{r}_\theta \times \mathbf{r}_\phi| = 9\sin\phi$.

The surface area is

$$\sigma = \int_0^{2\pi}\int_0^{\pi} 9\sin\phi d\phi d\theta$$

and the average temperature on the surface is

$$AV = \frac{1}{\sigma}\int_0^{2\pi}\int_0^{\pi} \sin((3\cos\theta\sin\phi)(3\sin\theta\sin\phi) + 3\cos\phi)9\sin\phi d\phi d\theta.$$

◄

Example 7.3.5 Consider the surface which is the upper hemisphere of radius 3 with density $\delta(x,y,z) = z^2$. Calculate its surface, the mass and the center of mass

Solution: ►

A parametrization of the surface is

$$\mathbf{r}(\theta,\phi) = \langle 3\cos\theta\sin\phi, 3\sin\theta\sin\phi, 3\cos\phi\rangle$$

for $0 \leqslant \theta \leqslant 2\pi$ and $0 \leqslant \phi \leqslant \pi/2$. The surface area differential is

$$dS = |\mathbf{r}_\theta \times \mathbf{r}_\phi| d\theta d\phi = 9\sin\phi\, d\theta d\phi.$$

The surface area is

$$S = \int_0^{2\pi}\int_0^{\pi/2} 9\sin\phi\, d\phi d\theta.$$

If the density is $\delta(x,y,z) = z^2$, then we have

$$\bar{y} = \frac{\iint_S y\delta dS}{\iint_S \delta dS} = \frac{\int_0^{2\pi}\int_0^{\pi/2} (3\sin\theta\sin\phi)(3\cos\phi)^2(9\sin\phi)\, d\phi d\theta}{\int_0^{2\pi}\int_0^{\pi/2} (3\cos\phi)^2(9\sin\phi)\, d\phi d\theta}$$

◄

7.4 Surface Integrals of Vector Functions

Like curves, we can parametrize a surface in two different orientations. The orientation of a curve is given by the unit tangent vector n; the orientation of a surface is given by the unit normal vector n. Unless we are dealing with an unusual surface, a surface has two sides. We can pick the normal vector to point out one side of the surface, or we can pick the normal vector to point out the other side of the surface. Our choice of normal vector specifies the orientation of the surface. We call the side of the surface with the normal vector the positive side of the surface.

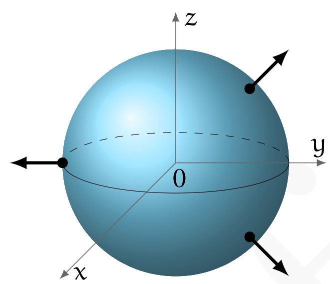
Definition 7.4.1 We say $(S, \hat{\mathbf{n}})$ is an **oriented surface** if $S \subseteq \mathbb{R}^3$ is a C^1 surface, $\hat{\mathbf{n}} : S \to \mathbb{R}^3$ is a continuous function such that for every $x \in S$, the vector $\hat{\mathbf{n}}(x)$ is normal to the surface S at the point x, and $\left|\hat{\mathbf{n}}(x)\right| = 1$.

Example 7.4.1 Let $S = \left\{x \in \mathbb{R}^3 \mid |x| = 1\right\}$, and choose $\hat{\mathbf{n}}(x) = x/|x|$.

R At any point $x \in S$ there are exactly two possible choices of $\hat{\mathbf{n}}(x)$. An oriented surface simply provides a consistent choice of one of these **in a continuous way** on the *entire surface*. Surprisingly this isn't always possible! If S is the surface of a Möbius strip, for instance, cannot be oriented.

Example 7.4.2 If S is the graph of a function, we orient S by chosing $\hat{\mathbf{n}}$ to always be the unit normal vector with a positive z coordinate.

Example 7.4.3 If S is a closed surface, then we will typically orient S by letting $\hat{\mathbf{n}}$ to be the **outward pointing** normal vector.



Recall that normal vectors to a plane can point in two opposite directions. By an **outward unit normal vector** to a surface S, we will mean the unit vector that is normal to S and points to the "outer" part of the surface.

If S is some oriented surface with unit normal $\hat{\mathbf{n}}$, then the amount of fluid flowing through S per unit time is exactly

$$\iint_S \mathbf{f} \cdot \hat{\mathbf{n}} \, dS.$$

Note, both $\mathbf{f}$ and $\hat{\mathbf{n}}$ above are **vector functions**, and $\mathbf{f} \cdot \hat{\mathbf{n}} : S \to \mathbb{R}$ is a scalar function. The surface integral of this was defined in the previous section.

Example 7.4.4 If S is the surface of a Möbius strip, for instance, cannot be oriented.

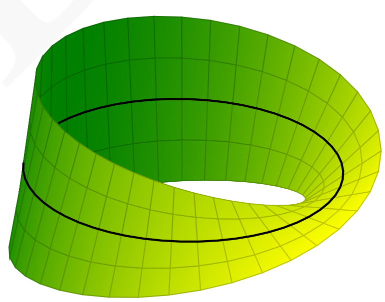


Figure 7.5 The Moebius Strip is an example of a surface that is not orientable

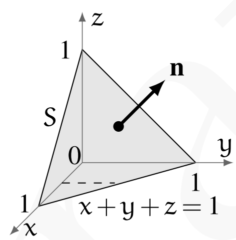
Definition 7.4.2 Let $(S, \hat{\mathbf{n}})$ be an oriented surface, and $\mathbf{f} : S \to \mathbb{R}^3$ be a C^1 vector field. The **surface integral** of $\mathbf{f}$ over S is defined to be

$$\iint_S \mathbf{f} \cdot \hat{\mathbf{n}} \, dS.$$

R Other common notation for the surface integral is

$$\iint_S \mathbf{f} \cdot \hat{\mathbf{n}} \, dS = \iint_S \mathbf{f} \cdot d\mathbf{S} = \iint_S \mathbf{f} \cdot dA$$

Example 7.4.5 Evaluate the surface integral $\iint_S \mathbf{f} \cdot dS$, where $\mathbf{f}(x,y,z) = yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$ and S is the part of the plane $x+y+z=1$ with $x \geqslant 0$, $y \geqslant 0$, and $z \geqslant 0$, with the outward unit normal $\mathbf{n}$ pointing in the positive z direction.



Solution: ▶

Since the vector $\mathbf{v} = (1,1,1)$ is normal to the plane $x+y+z=1$ (why?), then dividing $\mathbf{v}$ by its length yields the outward unit normal vector $\mathbf{n} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$. We now need to parametrize S. As we can see from Figure projecting S onto the xy-plane yields a triangular region $R = \{(x,y) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1-x\}$. Thus, using (u,v) instead of (x,y), we see that

$$x = u,\ y = v,\ z = 1-(u+v), \quad \text{for } 0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1-u$$

is a parametrization of S over Ω (since $z = 1-(x+y)$ on S). So on S,

$$\begin{aligned}
\mathbf{f} \cdot \mathbf{n} &= (yz, xz, xy) \cdot \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \frac{1}{\sqrt{3}}(yz + xz + xy) \\
&= \frac{1}{\sqrt{3}}((x+y)z + xy) = \frac{1}{\sqrt{3}}((u+v)(1-(u+v)) + uv) \\
&= \frac{1}{\sqrt{3}}((u+v) - (u+v)^2 + uv)
\end{aligned}$$

for (u,v) in Ω, and for $\mathbf{r}(u,v) = x(u,v)\mathbf{i} + y(u,v)\mathbf{j} + z(u,v)\mathbf{k} = u\mathbf{i} + v\mathbf{j} + (1-(u+v))\mathbf{k}$ we have

$$\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \;=\; (1,0,-1)\times(0,1,-1) \;=\; (1,1,1) \quad \Rightarrow \quad \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| = \sqrt{3}\,.$$

Thus, integrating over Ω using vertical slices (e.g. as indicated by the dashed line in Figure 4.4.5) gives

$$\begin{aligned}
\iint\limits_S \mathbf{f} \bullet d\mathbf{S} &= \iint\limits_S \mathbf{f} \bullet \mathbf{n}\, dS \\
&= \iint\limits_\Omega (\mathbf{f}(x(u,v), y(u,v), z(u,v)) \bullet \mathbf{n}) \left\| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right\| dv\, du \\
&= \int_0^1 \int_0^{1-u} \frac{1}{\sqrt{3}}((u+v) - (u+v)^2 + uv)\sqrt{3}\, dv\, du \\
&= \int_0^1 \left(\frac{(u+v)^2}{2} - \frac{(u+v)^3}{3} + \frac{uv^2}{2} \Bigg|_{v=0}^{v=1-u} \right) du \\
&= \int_0^1 \left(\frac{1}{6} + \frac{u}{2} - \frac{3u^2}{2} + \frac{5u^3}{6} \right) du \\
&= \frac{u}{6} + \frac{u^2}{4} - \frac{u^3}{2} + \frac{5u^4}{24} \Bigg|_0^1 = \frac{1}{8}\,.
\end{aligned}$$

◀

Proposition 7.4.6 Let $\mathbf{r} : U \to S$ be a parametrization of the oriented surface $(S, \hat{\mathbf{n}})$. Then either

$$\hat{\mathbf{n}} \circ \mathbf{r} = \frac{\partial_u \mathbf{r} \times \partial_v \mathbf{r}}{|\partial_u \mathbf{r} \times \partial_v \mathbf{r}|} \tag{7.4}$$

on all of S, or

$$\hat{\mathbf{n}} \circ \mathbf{r} = -\frac{\partial_u \mathbf{r} \times \partial_v \mathbf{r}}{|\partial_u \mathbf{r} \times \partial_v \mathbf{r}|} \tag{7.5}$$

on all of S. Consequently, in the case (7.4) holds, we have

$$\iint_S \mathbf{F} \bullet \hat{\mathbf{n}}\, dS = \int_U (\mathbf{F} \circ \mathbf{r}) \bullet (\partial_u \mathbf{r} \times \partial_v \mathbf{r})\, dA. \tag{7.6}$$

Proof: The vector $\partial_u \mathbf{r} \times \partial_v \mathbf{r}$ is **normal** to S and hence parallel to $\hat{\mathbf{n}}$. Thus

$$\hat{\mathbf{n}} \bullet \frac{\partial_u \mathbf{r} \times \partial_v \mathbf{r}}{|\partial_u \mathbf{r} \times \partial_v \mathbf{r}|}$$

must be a function that only takes on the values ± 1. Since s is also continuous, it must either be identically 1 or identically -1, finishing the proof. ■

Example 7.4.7 Gauss's law sates that the total charge enclosed by a surface S is given by

$$Q = \epsilon_0 \iint_S E \bullet d\mathbf{S},$$

where ϵ_0 the permittivity of free space, and E is the electric field. By convention, the normal vector is chosen to be pointing outward.

If $E(x) = e_3$, compute the charge enclosed by the top half of the hemisphere bounded by $|x| = 1$ and $x_3 = 0$.

7.5 Kelvin-Stokes Theorem

Given a surface $S \subset \mathbb{R}^3$ with boundary ∂S you are free to chose the orientation of S, i.e., the direction of the normal, but you have to orient S and ∂S coherently. This means that if you are an "observer" walking along the boundary of the surface with the normal as your upright direction; you are moving in the positive direction if onto the surface the boundary the interior of S is on *to the left* of ∂S.

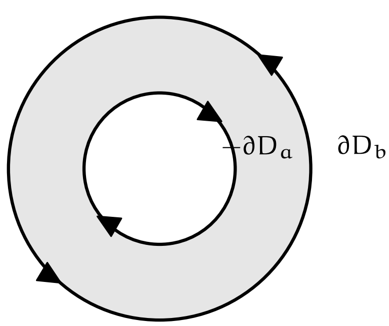
Example 7.5.1 Consider the annulus

$$A := \{(x, y, 0) \mid a^2 \leqslant x^2 + y^2 \leqslant b^2\}$$

in the (x, y)-plane, and from the two possible normal unit vectors $(0, 0, \pm 1)$ choose $\hat{n} := (0, 0, 1)$. If you are an "observer" walking along the boundary of the surface with the normal as $\hat{n}$ means that the outer boundary circle of A should be oriented counterclockwise. Staring at the figure you can convince yourself that the inner boundary circle has to be oriented clockwise to make the interior of A lie to the left of ∂A. One might write

$$\partial A = \partial D_b - \partial D_a \ ,$$

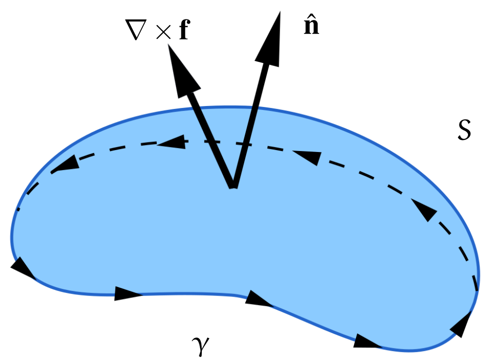
where D_r is the disk of radius r centered at the origin, and its boundary circle ∂D_r is oriented counterclockwise.



Theorem 7.5.2 — Kelvin–Stokes Theorem. Let $U \subseteq \mathbb{R}^3$ be a domain, $(S, \hat{\mathbf{n}}) \subseteq U$ be a bounded, oriented, piecewise C^1, surface whose boundary is the (piecewise C^1) curve γ. If $\mathbf{f}: U \to \mathbb{R}^3$ is a C^1 vector field, then

$$\int_S \nabla \times \mathbf{f} \bullet \hat{\mathbf{n}} \, dS = \oint_\gamma \mathbf{f} \bullet d\ell.$$

Here γ is traversed in the counter clockwise direction when viewed by an observer standing with his feet on the surface and head in the direction of the normal vector.



Proof: Let $\mathbf{f} = f_1\mathbf{i} + f_2\mathbf{j} + f_3\mathbf{k}$. Consider

$$\nabla \times (f_1\mathbf{i}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ f_1 & 0 & 0 \end{vmatrix} = \mathbf{j}\frac{\partial f_1}{\partial z} - \mathbf{k}\frac{\partial f_1}{\partial y}$$

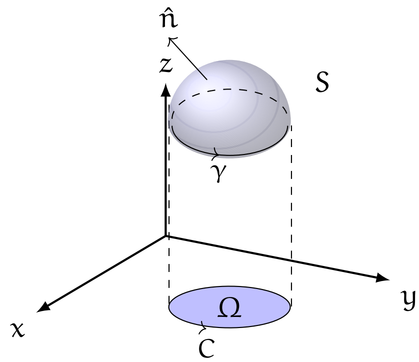
Then we have

$$\begin{aligned} \int_S [\nabla \times (f_1\mathbf{i})] \cdot dS &= \int_S (\hat{n} \cdot \nabla \times (f_1\mathbf{i}) \, dS \\ &= \int_S \frac{\partial f_1}{\partial z}(\mathbf{j} \cdot \hat{n}) - \frac{\partial f_1}{\partial y}(\mathbf{k} \cdot \hat{n}) \, dS \end{aligned}$$

We prove the theorem in the case S is a graph of a function, i.e., S is parametrized as

$$r = x\mathbf{i} + y\mathbf{j} + g(x, y)\mathbf{k}$$

where $g(x, y) : \Omega \to \mathbb{R}$. In this case the boundary γ of S is given by the image of the curve C boundary of Ω:



Let the equation of S be $z = g(x, y)$. Then we have

$$\hat{n} = \frac{-\partial g/\partial x\mathbf{i} - \partial g/\partial y\mathbf{j} + \mathbf{k}}{((\partial g/\partial x)^2 + (\partial g/\partial y)^2 + 1)^{1/2}}$$

Therefore on Ω:

$$\mathbf{j} \cdot \hat{n} = -\frac{\partial g}{\partial y}(\mathbf{k} \cdot \hat{n}) = -\frac{\partial z}{\partial y}(\mathbf{k} \cdot \hat{n})$$

Thus

$$\int_S [\nabla \times (f_1\mathbf{i})] \cdot dS = \int_S \left(-\left.\frac{\partial f_1}{\partial y}\right|_{z,x} - \left.\frac{\partial f_1}{\partial z}\right|_{y,x} \left.\frac{\partial z}{\partial y}\right|_x \right) (\mathbf{k} \cdot \hat{n})\, dS$$

Using the chain rule for partial derivatives

$$= -\int_S \left.\frac{\partial}{\partial y}\right|_x f_1(x, y, z)(\mathbf{k} \cdot \hat{n})\, dS$$

Then:

$$= -\int_\Omega \frac{\partial}{\partial y} f_1(x, y, g)\, dx\, dy$$
$$= \oint_C f_1(x, y, f(x, y))$$

with the last line following by using Green's theorem. However on γ we have $z = g$ and

$$\oint_C f_1(x, y, g)\, dx = \oint_\gamma f_1(x, y, z)\, dx$$

We have therefore established that

$$\int_S (\nabla \times f_1\mathbf{i}) \cdot d\mathbf{f} = \oint_\gamma f_1\, dx$$

In a similar way we can show that

$$\int_S (\nabla \times A_2\mathbf{j}) \cdot d\mathbf{f} = \oint_\gamma A_2\, dy$$

and

$$\int_S (\nabla \times A_3\mathbf{k}) \cdot d\mathbf{f} = \oint_\gamma A_3\, dz$$

and so the theorem is proved by adding all three results together. ■

Example 7.5.3 Verify Stokes' Theorem for $\mathbf{f}(x,y,z) = z\,\mathbf{i} + x\,\mathbf{j} + y\,\mathbf{k}$ when S is the paraboloid $z = x^2 + y^2$ such that $z \leqslant 1$.

Solution: ► The positive unit normal vector to the surface $z = z(x,y) = x^2 + y^2$ is

$$\mathbf{n} = \frac{-\frac{\partial z}{\partial x}\mathbf{i} - \frac{\partial z}{\partial y}\mathbf{j} + \mathbf{k}}{\sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}} = \frac{-2x\,\mathbf{i} - 2y\,\mathbf{j} + \mathbf{k}}{\sqrt{1 + 4x^2 + 4y^2}},$$

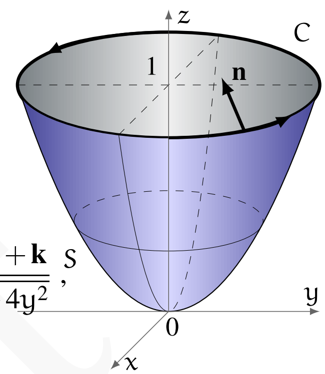


Figure 7.6 $z = x^2 + y^2$

and $\nabla \times \mathbf{f} = (1-0)\,\mathbf{i} + (1-0)\,\mathbf{j} + (1-0)\,\mathbf{k} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, so

$$(\nabla \times \mathbf{f}) \boldsymbol{\cdot} \mathbf{n} = (-2x - 2y + 1)/\sqrt{1 + 4x^2 + 4y^2}\,.$$

Since S can be parametrized as $\mathbf{r}(x,y) = x\,\mathbf{i} + y\,\mathbf{j} + (x^2 + y^2)\,\mathbf{k}$ for (x,y) in the region $D = \{(x,y) : x^2 + y^2 \leqslant 1\}$, then

$$\begin{aligned}
\iint_S (\nabla \times \mathbf{f}) \boldsymbol{\cdot} \mathbf{n}\, dS &= \iint_D (\nabla \times \mathbf{f}) \boldsymbol{\cdot} \mathbf{n} \left\| \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right\| dA \\
&= \iint_D \frac{-2x - 2y + 1}{\sqrt{1 + 4x^2 + 4y^2}} \sqrt{1 + 4x^2 + 4y^2}\, dA \\
&= \iint_D (-2x - 2y + 1)\, dA \text{ , so switching to polar coordinates gives} \\
&= \int_0^{2\pi} \int_0^1 (-2r\cos\theta - 2r\sin\theta + 1) r\, dr\, d\theta \\
&= \int_0^{2\pi} \int_0^1 (-2r^2\cos\theta - 2r^2\sin\theta + r)\, dr\, d\theta \\
&= \int_0^{2\pi} \left(-\frac{2r^3}{3}\cos\theta - \frac{2r^3}{3}\sin\theta + \frac{r^2}{2} \Big|_{r=0}^{r=1} \right) d\theta \\
&= \int_0^{2\pi} \left(-\frac{2}{3}\cos\theta - \frac{2}{3}\sin\theta + \frac{1}{2} \right) d\theta \\
&= -\frac{2}{3}\sin\theta + \frac{2}{3}\cos\theta + \frac{1}{2}\theta \Big|_0^{2\pi} = \pi\,.
\end{aligned}$$

The boundary curve C is the unit circle $x^2+y^2=1$ laying in the plane $z=1$ (see Figure), which can be parametrized as $x=\cos t$, $y=\sin t$, $z=1$ for $0 \leqslant t \leqslant 2\pi$. So

$$\begin{aligned}
\oint_C \mathbf{f}\bullet d\mathbf{r} &= \int_0^{2\pi} ((1)(-\sin t)+(\cos t)(\cos t)+(\sin t)(0))\,dt \\
&= \int_0^{2\pi} \left(-\sin t + \frac{1+\cos 2t}{2}\right) dt \quad \left(\text{here we used } \cos^2 t = \frac{1+\cos 2t}{2}\right) \\
&= \cos t + \frac{t}{2} + \frac{\sin 2t}{4}\bigg|_0^{2\pi} = \pi\,.
\end{aligned}$$

So we see that $\oint_C \mathbf{f}\bullet d\mathbf{r} = \iint_S (\nabla \times \mathbf{f})\bullet\mathbf{n}\,dS$, as predicted by Stokes' Theorem. ◄

The line integral in the preceding example was far simpler to calculate than the surface integral, but this will not always be the case.

Example 7.5.4 Let S be the section of a sphere of radius a with $0 \leqslant \theta \leqslant \alpha$. In spherical coordinates,

$$dS = a^2 \sin\theta \mathbf{e}_r \, d\theta \, d\varphi.$$

Let $\mathbf{F} = (0, xz, 0)$. Then $\nabla \times \mathbf{F} = (-x, 0, z)$. Then

$$\int_S \nabla \times \mathbf{F}\bullet d\mathbf{S} = \pi a^3 \cos\alpha \sin^2\alpha.$$

Our boundary ∂C is

$$\mathbf{r}(\varphi) = a(\sin\alpha\cos\varphi, \sin\alpha\sin\varphi, \cos\alpha).$$

The right hand side of Stokes' is

$$\begin{aligned}
\int_C \mathbf{F}\bullet d\ell &= \int_0^{2\pi} \underbrace{a\sin\alpha\cos\varphi}_{x}\,\underbrace{a\cos\alpha}_{z}\,\underbrace{a\sin\alpha\cos\varphi\, d\varphi}_{dy} \\
&= a^3 \sin^2\alpha\cos\alpha \int_0^{2\pi} \cos^2\varphi \, d\varphi \\
&= \pi a^3 \sin^2\alpha\cos\alpha.
\end{aligned}$$

So they agree.

R The rule determining the direction of traversal of γ is often called the **right hand rule**. Namely, if you put your right hand on the surface with thumb aligned with $\hat{\mathbf{n}}$, then γ is traversed in the pointed to by your index finger.

R If the surface S has holes in it, then (as we did with Greens theorem) we orient each of the holes clockwise, and the exterior boundary counter clockwise following the right hand rule. Now Kelvin–Stokes theorem becomes

$$\int_S \nabla \times \mathbf{f}\bullet\hat{\mathbf{n}}\,dS = \int_{\partial S} \mathbf{f}\bullet d\ell,$$

where the line integral over ∂S is defined to be the sum of the line integrals over each component of the boundary.

R If S is contained in the x, y plane and is oriented by choosing $\hat{\mathbf{n}} = e_3$, then Kelvin–Stokes theorem reduces to Greens theorem.

Kelvin–Stokes theorem allows us to quickly see how the curl of a vector field measures the infinitesimal circulation.

Proposition 7.5.5 Suppose a small, rigid paddle wheel of radius a is placed in a fluid with center at x_0 and rotation axis parallel to $\hat{\mathbf{n}}$. Let $v : \mathbb{R}^3 \to \mathbb{R}^3$ be the vector field describing the velocity of the ambient fluid. If ω the angular speed of rotation of the paddle wheel about the axis $\hat{\mathbf{n}}$, then

$$\lim_{a \to 0} \omega = \frac{\nabla \times v(x_0) \bullet \hat{\mathbf{n}}}{2}.$$

Proof: Let S be the surface of a disk with center x_0, radius a, and face perpendicular to $\hat{\mathbf{n}}$, and $\gamma = \partial S$. (Here S represents the face of the paddle wheel, and γ the boundary.) The angular speed ω will be such that

$$\oint_\gamma (v - a\omega\hat{\tau}) \bullet d\ell = 0,$$

where $\hat{\tau}$ is a unit vector tangent to γ, pointing in the direction of traversal. Consequently

$$\omega = \frac{1}{2\pi a^2} \oint_\gamma v \bullet d\ell = \frac{1}{2\pi a^2} \iint_S \nabla \times v \bullet \hat{\mathbf{n}} \, dS \xrightarrow{a \to 0} \frac{\nabla \times v(x_0) \bullet \hat{\mathbf{n}}}{2}. \quad \blacksquare$$

■

Example 7.5.6 Let S be the elliptic paraboloid $z = \dfrac{x^2}{4} + \dfrac{y^2}{9}$ for $z \leqslant 1$, and let C be its boundary curve. Calculate $\oint_C \mathbf{f} \bullet d\mathbf{r}$ for $\mathbf{f}(x, y, z) = (9xz + 2y)\mathbf{i} + (2x + y^2)\mathbf{j} + (-2y^2 + 2z)\mathbf{k}$, where C is traversed counterclockwise.

Solution: ► The surface is similar to the one in Example 7.5.3, except now the boundary curve C is the ellipse $\dfrac{x^2}{4} + \dfrac{y^2}{9} = 1$ laying in the plane $z = 1$. In this case, using Stokes' Theorem is easier than computing the line integral directly. As in Example 7.5.3, at each point $(x, y, z(x, y))$ on the surface $z = z(x, y) = \dfrac{x^2}{4} + \dfrac{y^2}{9}$ the vector

$$\mathbf{n} = \frac{-\dfrac{\partial z}{\partial x}\mathbf{i} - \dfrac{\partial z}{\partial y}\mathbf{j} + \mathbf{k}}{\sqrt{1 + \left(\dfrac{\partial z}{\partial x}\right)^2 + \left(\dfrac{\partial z}{\partial y}\right)^2}} = \frac{-\dfrac{x}{2}\mathbf{i} - \dfrac{2y}{9}\mathbf{j} + \mathbf{k}}{\sqrt{1 + \dfrac{x^2}{4} + \dfrac{4y^2}{9}}},$$

is a positive unit normal vector to S. And calculating the curl of $\mathbf{f}$ gives

$$\nabla \times \mathbf{f} = (-4y-0)\mathbf{i} + (9x-0)\mathbf{j} + (2-2)\mathbf{k} = -4y\,\mathbf{i} + 9x\,\mathbf{j} + 0\,\mathbf{k}\,,$$

so

$$(\nabla \times \mathbf{f})\cdot\mathbf{n} = \frac{(-4y)(-\frac{x}{2}) + (9x)(-\frac{2y}{9}) + (0)(1)}{\sqrt{1+\frac{x^2}{4}+\frac{4y^2}{9}}} = \frac{2xy - 2xy + 0}{\sqrt{1+\frac{x^2}{4}+\frac{4y^2}{9}}} = 0\,,$$

and so by Stokes' Theorem

$$\oint_C \mathbf{f}\cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{f})\cdot\mathbf{n}\,dS = \iint_S 0\,dS = 0\,.$$

◀

7.6 Divergence Theorem

Theorem 7.6.1 — Divergence Theorem. Let $U \subseteq \mathbb{R}^3$ be a bounded domain whose boundary is a (piecewise) C^1 surface denoted by ∂U. If $\mathbf{f} : U \to \mathbb{R}^3$ is a C^1 vector field, then

$$\iiint_U (\nabla\cdot\mathbf{f})\,dV = \oiint_{\partial U} \mathbf{f}\cdot\hat{\mathbf{n}}\,dS,$$

where $\hat{\mathbf{n}}$ is the outward pointing unit normal vector.

R Similar to our convention with line integrals, we denote surface integrals over **closed surfaces** with the symbol $\oiint$.

R Let $B_R = B(x_0, R)$ and observe

$$\lim_{R\to 0} \frac{1}{\mathrm{volume}(\partial B_R)} \int_{\partial B_R} \mathbf{f}\cdot\hat{\mathbf{n}}\,dS = \lim_{R\to 0} \frac{1}{\mathrm{volume}(\partial B_R)} \int_{B_R} \nabla\cdot\mathbf{f}\,dV = \nabla\cdot\mathbf{f}(x_0),$$

which justifies our intuition that $\nabla\cdot\mathbf{f}$ measures the outward flux of a vector field.

R If $V \subseteq \mathbb{R}^2$, $U = V \times [a, b]$ is a cylinder, and $\mathbf{f} : \mathbb{R}^3 \to R^3$ is a vector field that doesn't depend on x_3, then the divergence theorem reduces to Greens theorem.

Proof:[Proof of the Divergence Theorem] Suppose first that the domain U is the unit cube $(0,1)^3 \subseteq \mathbb{R}^3$. In this case

$$\iiint_U \nabla \bullet \mathbf{f}\, dV = \iiint_U (\partial_1 v_1 + \partial_2 v_2 + \partial_3 v_3)\, dV.$$

Taking the first term on the right, the fundamental theorem of calculus gives

$$\begin{aligned}\iiint_U \partial_1 v_1\, dV &= \int_{x_3=0}^{1} \int_{x_2=0}^{1} (v_1(1,x_2,x_3) - v_1(0,x_2,x_3))\ dx_2\ dx_3 \\ &= \int_L v \bullet \hat{\mathbf{n}}\, dS + \int_R v \bullet \hat{\mathbf{n}}\, dS,\end{aligned}$$

where L and B are the left and right faces of the cube respectively. The $\partial_2 v_2$ and $\partial_3 v_3$ terms give the surface integrals over the other four faces. This proves the divergence theorem in the case that the domain is the unit cube.

■

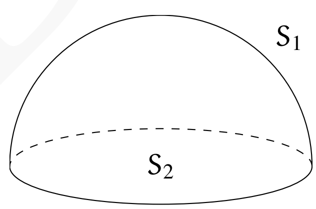
Example 7.6.2 Evaluate $\displaystyle\iint_S \mathbf{f} \bullet dS$, where $\mathbf{f}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and S is the unit sphere $x^2+y^2+z^2 = 1$.

Solution: ► We see that div $\mathbf{f} = 1+1+1 = 3$, so

$$\begin{aligned}\iint_S \mathbf{f} \bullet dS &= \iiint_S \operatorname{div} \mathbf{f}\, dV = \iiint_S 3\, dV \\ &= 3 \iiint_S 1\, dV = 3 \operatorname{vol}(S) = 3 \cdot \frac{4\pi(1)^3}{3} = 4\pi\,.\end{aligned}$$

◄

Example 7.6.3 Consider a hemisphere.



V is a solid hemisphere

$$x^2+y^2+z^2 \leqslant a^2, \quad z \geqslant 0,$$

and $\partial V = S_1 + S_2$, the hemisphere and the disc at the bottom.

Take $\mathbf{F} = (0,0,z+a)$ and $\nabla \bullet \mathbf{F} = 1$. Then

$$\int_V \nabla \bullet \mathbf{F}\, dV = \frac{2}{3}\pi a^3,$$

the volume of the hemisphere.

On S_1,

$$d\mathbf{S} = \mathbf{n}\, dS = \frac{1}{a}(x, y, z)\, dS.$$

Then

$$\mathbf{F}\bullet d\mathbf{S} = \frac{1}{a} z(z+a)\, dS = \cos\theta\, a(\cos\theta + 1)\underbrace{a^2 \sin\theta\, d\theta\, d\varphi}_{dS}.$$

Then

$$\begin{aligned}
\int_{S_1} \mathbf{F}\bullet d\mathbf{S} &= a^3 \int_0^{2\pi} d\varphi \int_0^{\pi/2} \sin\theta(\cos^2\theta + \cos\theta)\, d\theta \\
&= 2\pi a^3 \left[\frac{-1}{3}\cos^3\theta - \frac{1}{2}\cos^2\theta\right]_0^{\pi/2} \\
&= \frac{5}{3}\pi a^3.
\end{aligned}$$

On S_2, $d\mathbf{S} = \mathbf{n}\, dS = -(0, 0, 1)\, dS$. Then $\mathbf{F}\bullet d\mathbf{S} = -a\, dS$. So

$$\int_{S_2} \mathbf{F}\bullet d\mathbf{S} = -\pi a^3.$$

So

$$\int_{S_1} \mathbf{F}\bullet d\mathbf{S} + \int_{S_2} \mathbf{F}\bullet d\mathbf{S} = \left(\frac{5}{3} - 1\right)\pi a^3 = \frac{2}{3}\pi a^3,$$

in accordance with Gauss' theorem.

7.6.1 Gauss's Law For Inverse-Square Fields

Proposition 7.6.4 — Gauss's gravitational law. Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the gravitational field of a mass distribution (i.e. $g(x)$ is the force experienced by a point mass located at x). If S is any closed C^1 surface, then

$$\oint_S g\bullet\hat{\mathbf{n}}\, dS = -4\pi G M,$$

where M is the mass enclosed by the region S. Here G is the gravitational constant, and $\hat{\mathbf{n}}$ is the outward pointing unit normal vector.

Proof: The core of the proof is the following calculation. Given a fixed $y \in \mathbb{R}^3$, define the vector field $\mathbf{f}$ by

$$\mathbf{f}(x) = \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|^3}.$$

The vector field $-Gm\mathbf{f}(x)$ represents the gravitational field of a mass located at $\mathbf{y}$ Then

$$\oint_S \mathbf{f}\bullet\hat{\mathbf{n}}\, dS = \begin{cases} 4\pi & \text{if } \mathbf{y} \text{ is in the region enclosed by } S, \\ 0 & \text{otherwise.} \end{cases} \tag{7.7}$$

For simplicity, we subsequently assume $\mathbf{y} = \mathbf{0}$.

To prove (7.7), observe

$$\nabla \bullet \mathbf{f} = 0,$$

when $\mathbf{x} \neq \mathbf{0}$. Let U be the region enclosed by S. If $\mathbf{0} \notin U$, then the divergence theorem will apply to in the region U and we have

$$\oint_S g \bullet \hat{\mathbf{n}} \, dS = \int_U \nabla \bullet g \, dV = 0.$$

On the other hand, if $\mathbf{0} \in U$, the divergence theorem will not directly apply, since $\mathbf{f} \notin C^1(U)$. To circumvent this, let $\epsilon > 0$ and $U' = U - B(0, \epsilon)$, and S' be the boundary of U'. Since $0 \notin U'$, $\mathbf{f}$ is C^1 on all of U' and the divergence theorem gives

$$0 = \int_{U'} \nabla \bullet \mathbf{f} \, dV = \int_{\partial U'} \mathbf{f} \bullet \hat{\mathbf{n}} \, dS,$$

and hence

$$\oint_S \mathbf{f} \bullet \hat{\mathbf{n}} \, dS = -\oint_{\partial B(0,\epsilon)} \mathbf{f} \bullet \hat{\mathbf{n}} \, dS = \oint_{\partial B(0,\epsilon)} \frac{1}{\epsilon^2} \, dS = -4\pi,$$

as claimed. (Above the normal vector on $\partial B(0, \epsilon)$ points outward with respect to the domain U', and *inward* with respect to the ball $B(0, \epsilon)$.)

Now, in the general case, suppose the mass distribution has density ρ. Then the gravitational field $g(x)$ will be the super-position of the gravitational fields at x due to a point mass of size $\rho(y) \, dV$ placed at y. Namely, this means

$$g(x) = -G \int_{\mathbb{R}^3} \frac{\rho(y)(x - y)}{|x - y|^3} \, dV(y).$$

Now using Fubini's theorem,

$$\begin{aligned}\iint_S g(x) \bullet \hat{\mathbf{n}}(x) \, dS(x) &= -G \int_{y \in \mathbb{R}^3} \rho(y) \int_{x \in S} \frac{x - y}{|x - y|^3} \bullet \hat{\mathbf{n}}(x) \, dS(x) \, dV(y) \\ &= -4\pi G \int_{y \in U} \rho(y) \, dV(y) = -4\pi G M,\end{aligned}$$

where the second last equality followed from (7.7). ■

Example 7.6.5 A system of electric charges has a *charge density* $\rho(x, y, z)$ and produces an electrostatic field $\mathbf{E}(x, y, z)$ at points (x, y, z) in space. *Gauss' Law* states that

$$\iint\limits_S \mathbf{E} \bullet d\mathbf{S} \;=\; 4\pi \iiint\limits_S \rho \, dV$$

for any closed surface S which encloses the charges, with S being the solid region enclosed by S. Show that $\nabla \bullet \mathbf{E} = 4\pi\rho$. This is one of *Maxwell's Equations*.[1]

[1] In Gaussian (or CGS) units.

Solution: ► By the Divergence Theorem, we have

$$\begin{aligned}
\iiint_S \nabla\bullet\mathbf{E}\,dV &= \iint_S \mathbf{E}\bullet d\mathbf{S} \\
&= 4\pi \iiint_S \rho\,dV \quad \text{by Gauss' Law, so combining the integrals gives} \\
\iiint_S (\nabla\bullet\mathbf{E} - 4\pi\rho)\,dV &= 0 \quad \text{, so} \\
\nabla\bullet\mathbf{E} - 4\pi\rho &= 0 \quad \text{since } S \text{ and hence } S \text{ was arbitrary, so} \\
\nabla\bullet\mathbf{E} &= 4\pi\rho\,.
\end{aligned}$$

◄

7.7 Applications of Surface Integrals

7.7.1 Conservative and Potential Forces

We've seen before that any potential force must be conservative. We demonstrate the converse here.

Theorem 7.7.1 Let $U \subseteq \mathbb{R}^3$ be a simply connected domain, and $\mathbf{f}: U \to \mathbb{R}^3$ be a C^1 vector field. Then $\mathbf{f}$ is a conservative force, if and only if $\mathbf{f}$ is a potential force, if and only if $\nabla\times\mathbf{f} = 0$.

Proof: Clearly, if $\mathbf{f}$ is a potential force, equality of mixed partials shows $\nabla\times\mathbf{f} = 0$. Suppose now $\nabla\times\mathbf{f} = 0$. By Kelvin–Stokes theorem

$$\oint_\gamma \mathbf{f}\bullet d\ell = \int_S \nabla\times\mathbf{f}\bullet\hat{\mathbf{n}}\,dS = 0,$$

and so $\mathbf{f}$ is conservative. Thus to finish the proof of the theorem, we only need to show that a conservative force is a potential force. We do this next.

Suppose $\mathbf{f}$ is a conservative force. Fix $x_0 \in U$ and define

$$V(x) = -\int_\gamma \mathbf{f}\bullet d\ell,$$

where γ is *any* path joining x_0 and x that is completely contained in U. Since $\mathbf{f}$ is conservative, we seen before that the line integral above *will not* depend on the path itself but only on the endpoints.

Now let $h > 0$, and let γ be a path that joins x_0 to a, and is a straight line between a and $a + he_1$. Then

$$-\partial_1 V(a) = \lim_{h\to 0}\frac{1}{h}\int_{a_1}^{a_1+h} F_1(a + te_1)\;dt = F_1(a).$$

The other partials can be computed similarly to obtain $\mathbf{f} = -\nabla V$ concluding the proof. ■

7.7.2 Conservation laws

Definition 7.7.1 — Conservation equation. Suppose we are interested in a quantity Q. Let $\rho(\mathbf{r},t)$ be the amount of stuff per unit volume and $\mathbf{j}(\mathbf{r},t)$ be the flow rate of the quantity (eg if Q is charge, j is the current density).

The conservation equation is

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\mathbf{j}=0.$$

This is stronger than the claim that the total amount of Q in the universe is fixed. It says that Q cannot just disappear here and appear elsewhere. It must continuously flow out.

In particular, let V be a fixed time-independent volume with boundary $S=\partial V$. Then

$$Q(t)=\int_V \rho(\mathbf{r},t)\,dV$$

Then the rate of change of amount of Q in V is

$$\frac{dQ}{dt}=\int_V \frac{\partial\rho}{\partial t}\,dV=-\int_V \nabla\cdot\mathbf{j}\,dV=-\int_S \mathbf{j}\cdot d\mathbf{S}.$$

by divergence theorem. So this states that the rate of change of the quantity Q in V is the flux of the stuff flowing out of the surface. ie Q cannot just disappear but must smoothly flow out.

In particular, if V is the whole universe (ie $\mathbb{R}^3$), and $\mathbf{j}\to 0$ sufficiently rapidly as $|\mathbf{r}|\to\infty$, then we calculate the total amount of Q in the universe by taking V to be a solid sphere of radius Ω, and take the limit as $R\to\infty$. Then the surface integral $\to 0$, and the equation states that

$$\frac{dQ}{dt}=0,$$

Example 7.7.2 If $\rho(\mathbf{r},t)$ is the charge density (ie. $\rho\delta V$ is the amount of charge in a small volume δV), then $Q(t)$ is the total charge in V. $\mathbf{j}(\mathbf{r},t)$ is the electric current density. So $\mathbf{j}\cdot d\mathbf{S}$ is the charge flowing through $\delta\mathbf{S}$ per unit time.

Example 7.7.3 Let $\mathbf{j}=\rho\mathbf{u}$ with $\mathbf{u}$ being the velocity field. Then $(\rho\mathbf{u}\,\delta t)\cdot\delta\mathbf{S}$ is equal to the mass of fluid crossing $\delta\mathbf{S}$ in time δt. So

$$\frac{dQ}{dt}=-\int_S \mathbf{j}\cdot d\mathbf{S}$$

does indeed imply the conservation of mass. The conservation equation in this case is

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{u})=0$$

For the case where ρ is constant and uniform (ie. independent of $\mathbf{r}$ and t), we get that $\nabla\cdot\mathbf{u}=0$. We say that the fluid is *incompressible*.

7.7.3 Maxell Equation

7.8 Helmholtz Decomposition

The Helmholtz theorem, also known as the **Fundamental Theorem of Vector Calculus**, states that a vector field $\mathbf{F}$ which vanishes at the boundaries can be written as the sum of two terms, one of which is irrotational and the other, solenoidal.

Theorem 7.8.1 — Helmholtz Decomposition for $\mathbb{R}^3$. If $\mathbf{F}$ is a C^2 vector function on $\mathbb{R}^3$ and $\mathbf{F}$ vanishes faster than $1/r$ as $r \to \infty$. Then $\mathbf{F}$ can be decomposed into a curl-free component and a divergence-free component:

$$\mathbf{F} = -\nabla\mathbf{\Phi} + \nabla \times \mathbf{A},$$

Proof: We will demonstrate first the case when $\mathbf{F}$ satisfies

$$\mathbf{F} = -\nabla^2 \mathbf{Z} \tag{7.8}$$

for some vector field $\mathbf{Z}$

Now, consider the following identity for an arbitrary vector field $\mathbf{Z}(\mathbf{r})$:

$$-\nabla^2 \mathbf{Z} = -\nabla(\nabla \cdot \mathbf{Z}) + \nabla \times \nabla \times \mathbf{Z} \tag{7.9}$$

then it follows that

$$\mathbf{F} = -\nabla U + \nabla \times \mathbf{W} \tag{7.10}$$

with

$$U = \nabla.\mathbf{Z} \tag{7.11}$$

and

$$\mathbf{W} = \nabla \times \mathbf{Z} \tag{7.12}$$

Eq.(7.10) is Helmholtz's theorem, as ∇U is irrotational and $\nabla \times \mathbf{W}$ is solenoidal.

Now we will generalize for all vector field: if $\mathbf{V}$ vanishes at infinity fast enough, for, then, the equation

$$\nabla^2 \mathbf{Z} = -\mathbf{V}\,, \tag{7.13}$$

which is Poisson's equation, has always the solution

$$\mathbf{Z}(\mathbf{r}) = \frac{1}{4\pi} \int d^3\mathbf{r}' \frac{\mathbf{V}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{7.14}$$

It is now a simple matter to prove, from Eq.(7.10), that $\mathbf{V}$ is determined from its div and curl. Taking, in fact, the divergence of Eq.(7.10), we have:

$$\mathrm{div}(\mathbf{V}) = -\nabla^2 U \tag{7.15}$$

which is, again, Poisson's equation, and, so, determines U as

$$U(\mathbf{r}) = \frac{1}{4\pi}\int d^3\mathbf{r}'\frac{\nabla'.\mathbf{V}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{7.16}$$

Take now the curl of Eq.(7.10). We have

$$\begin{aligned} \nabla\times\mathbf{V} &= \nabla\times\nabla\times\mathbf{W} \\ &= \nabla(\nabla.\mathbf{W}) - \nabla^2\mathbf{W} \end{aligned} \tag{7.17}$$

Now, $\nabla.\mathbf{W} = 0$, as $\mathbf{W} = \nabla\times\mathbf{Z}$, so another Poisson equation determines $\mathbf{W}$. Using U and $\mathbf{W}$ so determined in Eq.(7.10) proves the decomposition ■

Theorem 7.8.2 — Helmholtz Decomposition for Bounded Domains. If $\mathbf{F}$ is a C^2 vector function on a bounded domain $V\subset\mathbb{R}^3$ and let S be the surface that encloses the domain Vthen Then F can be decomposed into a curl-free component and a divergence-free component:

$$\mathbf{F} = -\nabla\boldsymbol{\Phi} + \nabla\times\mathbf{A},$$

where

$$\boldsymbol{\Phi}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{\nabla'\boldsymbol{\cdot}\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV' - \frac{1}{4\pi}\oint_S \hat{\mathbf{n}}'\boldsymbol{\cdot}\frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dS'$$

$$\mathbf{A}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{\nabla'\times\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV' - \frac{1}{4\pi}\oint_S \hat{\mathbf{n}}'\times\frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dS'$$

and ∇' is the gradient with respect to $\mathbf{r}'$ not $\mathbf{r}$.

7.9 Green's Identities

Theorem 7.9.1 Let ϕ and ψ be two scalar fields with continuous second derivatives. Then

- $\displaystyle\int_S \left[\phi\frac{\partial\psi}{\partial n}\right] dS = \int_U [\phi\nabla^2\psi + (\nabla\phi)\cdot(\nabla\psi)]\, dV$ *Green's first identity*
- $\displaystyle\int_S \left[\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right] dS = \int_U (\phi\nabla^2\psi - \psi\nabla^2\phi)\, dV$ *Green's second identity.*

Proof:

Consider the quantity

$$\mathbf{F} = \phi\nabla\psi$$

It follows that

$$\operatorname{div}\mathbf{F} = \phi\nabla^2\psi + (\nabla\phi)\cdot(\nabla\psi)$$
$$\hat{n}\cdot\mathbf{F} = \phi\partial\psi/\partial n$$

Applying the divergence theorem we obtain

$$\int_S \left[\phi\frac{\partial\psi}{\partial n}\right] dS = \int_U [\phi\nabla^2\psi + (\nabla\phi)\cdot(\nabla\psi)]\, dV$$

which is known as *Green's first identity*. Interchanging ϕ and ψ we have

$$\int_S \left[\psi\frac{\partial\phi}{\partial n}\right] dS = \int_U [\psi\nabla^2\phi + (\nabla\psi)\cdot(\nabla\phi)]\, dV$$

Subtracting (2) from (1) we obtain

$$\int_S \left[\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right] dS = \int_U (\phi\nabla^2\psi - \psi\nabla^2\phi)\, dV$$

which is known as *Green's second identity.*

■

Tensor Calculus

CHAPTER 8

Curvilinear Coordinates

8.1 Curvilinear Coordinates

The location of a point P in space can be represented in many different ways. Three systems commonly used in applications are the *rectangular cartesian system of Coordinates* (x, y, z), the *cylindrical polar system of Coordinates* (r, ϕ, z) and the *spherical system of Coordinates* (r, φ, ϕ). The last two are the best examples of *orthogonal curvilinear* systems of coordinates (u_1, u_2, u_3) .

Definition 8.1.1 A function $\mathbf{u} : U \to V$ is called a (differentiable) **coordinate change** if

- $\mathbf{u}$ is bijective
- $\mathbf{u}$ is differentiable
- $D\mathbf{u}$ is invertible at every point.

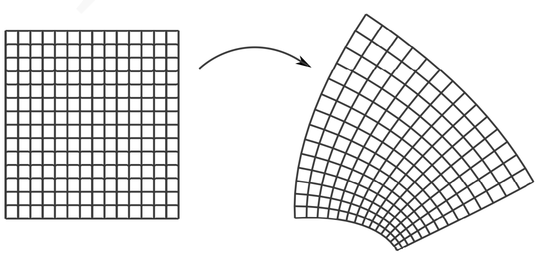
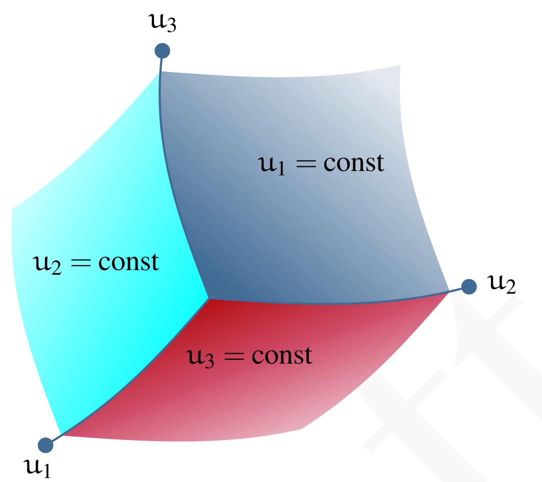


Figure 8.1 Coordinate System

In the tridimensional case, suppose that (x, y, z) are expressible as single-valued functions $\mathbf{u}$ of the variables (u_1, u_2, u_3). Suppose also that (u_1, u_2, u_3) can be expressed as single-valued functions of (x, y, z).

Through each point $P : (a, b, c)$ of the space we have three surfaces: $u_1 = c_1$, $u_2 = c_2$ and $u_3 = c_3$, where the constants c_i are given by $c_i = u_i(a, b, c)$

If say u_2 and u_3 are held fixed and u_1 is made to vary, a *path* results. Such path is called a u_1 *curve*. u_2 and u_3 curves can be constructed in analogous manner.



The system (u_1, u_2, u_3) is said to be a **curvilinear coordinate system**.

Example 8.1.1 The parabolic cylindrical coordinates are defined in terms of the Cartesian coordinates by:

$$\begin{aligned} x &= \sigma\tau \\ y &= \frac{1}{2}\left(\tau^2 - \sigma^2\right) \\ z &= z \end{aligned}$$

The constant surfaces are the plane

$$z = z_1$$

and the parabolic cylinders

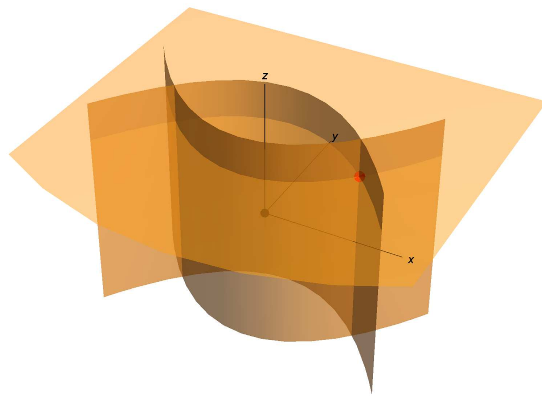
$$2y = \frac{x^2}{\sigma^2} - \sigma^2$$

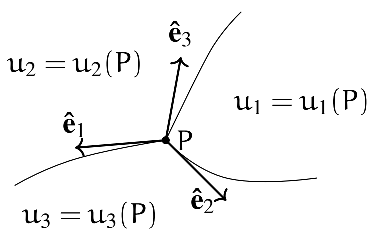
and

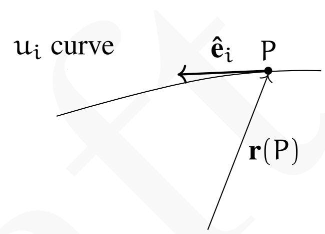
$$2y = -\frac{x^2}{\tau^2} + \tau^2$$

Coordinates I

The surfaces $u_2 = u_2(P)$ and $u_3 = u_3(P)$ intersect in a curve, along which only u_1 varies.







Let $\hat{\mathbf{e}}_1$ be the unit vector tangential to the curve at P. Let $\hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$ be unit vectors tangential to curves along which only u_2, u_3 vary.

Clearly

$$\hat{\mathbf{e}}_i = \frac{\partial \mathbf{r}}{\partial u_1} \Big/ \left\| \frac{\partial \mathbf{r}}{\partial u_i} \right\|.$$

And if we define $h_i = |\partial \mathbf{r} / \partial u_i|$ then

$$\frac{\partial \mathbf{r}}{\partial u_i} = \hat{\mathbf{e}}_i \cdot h_i$$

The quantities h_i are often known as the **length scales** for the coordinate system.

Coordinates II

Let $(\hat{\mathbf{e}}^i, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3)$ be unit vectors at P in the directions normal to $u_1 = u_1(P)$, $u_2 = u_2(P), u_3 = u_3(P)$ respectively, such that u_1, u_2, u_3 increase in the directions $\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{a}_3$. Clearly we must have

$$\hat{\mathbf{e}}^i = \nabla(u_i) / |\nabla u_i|$$

Definition 8.1.2 If $(\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3)$ are mutually orthogonal, the coordinate system is said to be an **orthogonal** curvilinear coordinate system.

In an orthogonal system we have

$$\hat{\mathbf{e}}_i = \hat{\mathbf{e}}^i = \frac{\partial \mathbf{r}/\partial u_i}{|\partial \mathbf{r}/\partial u_i|} = \nabla u_i / |\nabla u_i| \quad \text{for } i = 1,2,3$$

So we associate to a general curvilinear coordinate system two sets of basis vectors for every point:

$$\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$$

is the **covariant basis**, and

$$\{\hat{\mathbf{e}}^1, \hat{\mathbf{e}}^2, \hat{\mathbf{e}}^3\}$$

is the **contravariant basis**.

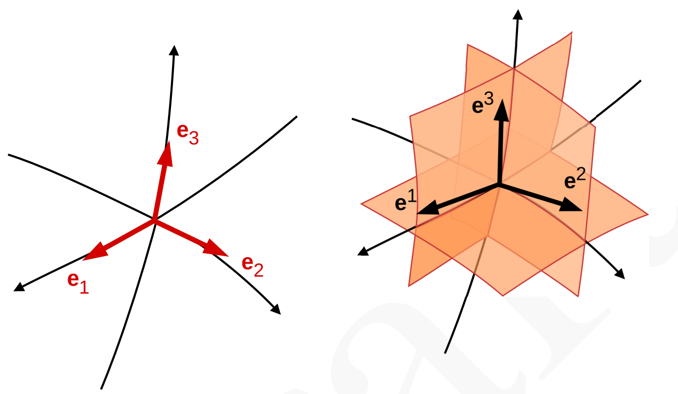


Figure 8.2 Covariant and Contravariant Basis

Note the following important equality:

$$\hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}_j = \delta^i_j.$$

Example 8.1.2 Cylindrical coordinates (r, θ, z):

$$x = r\cos\theta \qquad r = \sqrt{x^2 + y^2}$$

$$y = r\sin\theta \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

$$z = z \qquad z = z$$

where $0 \leqslant \theta \leqslant \pi$ if $y \geqslant 0$ and $\pi < \theta < 2\pi$ if $y < 0$

For cylindrical coordinates (r, θ, z), and constants r_0, θ_0 and z_0, we see from Figure 8.3 that the surface $r = r_0$ is a cylinder of radius r_0 centered along the z-axis, the surface $\theta = \theta_0$ is a half-plane emanating from the z-axis, and the surface $z = z_0$ is a plane parallel to the xy-plane.

The unit vectors $\hat{r}, \hat{\theta}, \hat{k}$ at any point P are perpendicular to the surfaces $r =$ constant, $\theta =$ constant, $z =$ constant through P in the directions of increasing r, θ, z. Note that the direction of the unit vectors $\hat{r}, \hat{\theta}$ vary from point to point, unlike the corresponding Cartesian unit vectors.

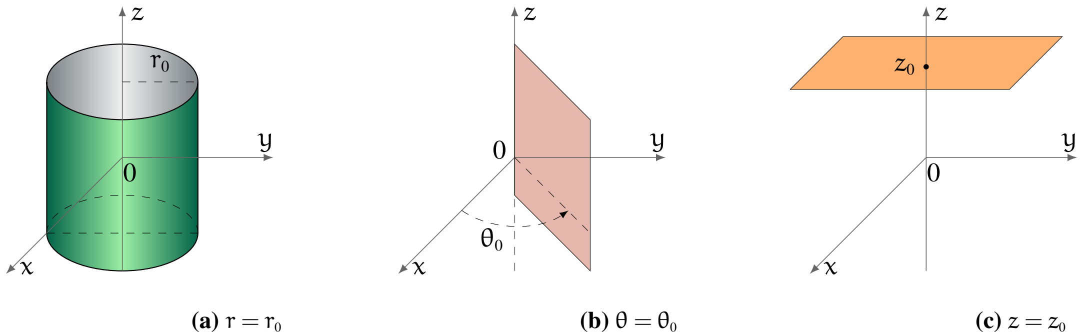


Figure 8.3 Cylindrical coordinate surfaces

8.2 Line and Volume Elements in Orthogonal Coordinate Systems

Definition 8.2.1 — Line Element. Since $\mathbf{r} = \mathbf{r}(u_1, u_2, u_3)$, the *line element* $d\mathbf{r}$ is given by

$$\begin{aligned} d\mathbf{r} &= \frac{\partial \mathbf{r}}{\partial u_1} du_1 + \frac{\partial \mathbf{r}}{\partial u_2} du_2 + \frac{\partial \mathbf{r}}{\partial u_3} du_3 \\ &= h_1 du_1 \hat{\mathbf{e}}_1 + h_2 du_2 \hat{\mathbf{e}}_2 + h_3 du_3 \hat{\mathbf{e}}_3 \end{aligned}$$

If the system is orthogonal, then it follows that

$$(ds)^2 = (d\mathbf{r}) \cdot (d\mathbf{r}) = h_1^2 (du_1)^2 + h_2^2 (du_2)^2 + h_3^2 (du_3)^2$$

In what follows we will assume we have an orthogonal system so that

$$\hat{\mathbf{e}}_i = \hat{\mathbf{e}}^i = \frac{\partial \mathbf{r}/\partial u_i}{|\partial \mathbf{r}/\partial u_i|} = \nabla u_i / |\nabla u_i| \quad \text{for } i = 1, 2, 3$$

In particular, line elements along curves of intersection of u_i surfaces have lengths $h_1 du_1$, $h_2 du_2$, $h_3 du_3$ respectively.

Definition 8.2.2 — Volume Element. In $\mathbb{R}^3$, the volume element is given by

$$dV = dx\ dy\ dz.$$

In a coordinate systems $x = x(u_1, u_2, u_3), y = y(u_1, u_2, u_3), z = z(u_1, u_2, u_3)$, the volume element is:

$$dV = \left| \frac{\partial(x, y, z)}{\partial(u_1, u_2, u_3)} \right| du_1\, du_2\, du_3.$$

Proposition 8.2.1 In an orthogonal system we have

$$\begin{aligned} dV &= (h_1 du_1)(h_2 du_2)(h_3 du_3) \\ &= h_1 h_2 h_3\, du_1 du_2 du_3 \end{aligned}$$

In this section we find the expression of the line and volume elements in some classics orthogonal coordinate systems.

(i) Cartesian Coordinates (x, y, z)

$$\begin{aligned} dV &= dx dy dz \\ d\mathbf{r} &= dx\hat{\imath} + dy\hat{\jmath} + dz\hat{k} \\ (ds)^2 &= (d\mathbf{r})\cdot(d\mathbf{r}) = (dx)^2 + (dy)^2 + (dz)^2 \end{aligned}$$

(ii) Cylindrical polar coordinates (r, θ, z) The coordinates are related to Cartesian by

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

We have that $(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$, but we can write

$$\begin{aligned} dx &= \left(\frac{\partial x}{\partial r}\right) dr + \left(\frac{\partial x}{\partial \theta}\right) d\theta + \left(\frac{\partial x}{\partial z}\right) dz \\ &= (\cos\theta)\, dr - (r\sin\theta)\, d\theta \end{aligned}$$

and

$$\begin{aligned} dy &= \left(\frac{\partial y}{\partial r}\right) dr + \left(\frac{\partial y}{\partial \theta}\right) d\theta + \left(\frac{\partial y}{\partial z}\right) dz \\ &= (\sin\theta)\, dr + (r\cos\theta) d\theta \end{aligned}$$

Therefore we have

$$\begin{aligned} (ds)^2 &= (dx)^2 + (dy)^2 + (dz)^2 \\ &= \cdots = (dr)^2 + r^2(d\theta)^2 + (dz)^2 \end{aligned}$$

Thus we see that for this coordinate system, the length scales are

$$h_1 = 1,\ h_2 = r,\ h_3 = 1$$

and the element of volume is

$$dV = r\, dr d\theta dz$$

(iii) Spherical Polar coordinates (r, ϕ, θ) In this case the relationship between the coordinates is

$$x = r\sin\phi\cos\theta;\ y = r\sin\phi\sin\theta;\ z = r\cos\phi$$

Again, we have that $(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$ and we know that

$$\begin{aligned} dx &= \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta + \frac{\partial x}{\partial \phi}d\phi \\ &= (\sin\phi\cos\theta)dr + (-r\sin\phi\sin\theta)d\theta + r\cos\phi\cos\theta d\phi \end{aligned}$$

and

$$\begin{aligned} dy &= \frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial \theta}d\theta + \frac{\partial y}{\partial \phi}d\phi \\ &= \sin\phi\sin\theta dr + r\sin\phi\cos\theta d\theta + r\cos\phi\sin\theta d\phi \end{aligned}$$

together with

$$\begin{aligned} dz &= \frac{\partial z}{\partial r}dr + \frac{\partial z}{\partial \theta}d\theta + \frac{\partial z}{\partial \phi}d\phi \\ &= (\cos\phi)dr - (r\sin\phi)d\phi \end{aligned}$$

Therefore in this case, we have (after some work)

$$\begin{aligned} (ds)^2 &= (dx)^2 + (dy)^2 + (dz)^2 \\ &= \cdots = (dr)^2 + r^2(d\phi)^2 + r^2\sin^2\phi(d\theta)^2 \end{aligned}$$

Thus the length scales are

$$h_1 = 1,\ h_2 = r,\ h_3 = r\sin\phi$$

and the volume element is

$$dV = r^2\sin\phi\, dr d\phi d\theta$$

Example 8.2.2 *Find the volume and surface area of a sphere of radius a, and also find the surface area of a cap of the sphere that subtends on angle α at the centre of the sphere.*

$$dV = r^2\sin\phi\, dr d\phi d\theta$$

and an element of surface of a sphere of radius a is (by removing $h_1 du_1 = dr$):

$$dS = a^2\sin\phi\, d\phi\, d\theta$$

$\therefore$ total volume is

$$\begin{aligned} \int_V dV &= \int_{\theta=0}^{2\pi}\int_{\phi=0}^{\pi}\int_{r=0}^{a} r^2\sin\phi\, dr d\phi d\theta \\ &= 2\pi[-\cos\phi]_0^{\pi}\int_0^a r^2\, dr \\ &= 4\pi a^3/3 \end{aligned}$$

Surface area is

$$\begin{aligned}\int_S dS &= \int_{\theta=0}^{2\pi}\int_{\phi=0}^{\pi} a^2 \sin\phi \, d\phi \, d\theta \\ &= 2\pi a^2[-\cos\phi]_0^{\pi} \\ &= 4\pi a^2\end{aligned}$$

Surface area of cap is

$$\begin{aligned}\int_{\theta=0}^{2\pi}\int_{\phi=0}^{\alpha} a^2 \sin\phi \, d\phi \, d\theta &= 2\pi a^2[-\cos\phi]_0^{\alpha} \\ &= 2\pi a^2(1-\cos\alpha)\end{aligned}$$

8.3 Gradient in Orthogonal Curvilinear Coordinates

Let

$$\nabla\Phi = \lambda_1\hat{\mathbf{e}}_1 + \lambda_2\hat{\mathbf{e}}_2 + \lambda_3\hat{\mathbf{e}}_3$$

in a general coordinate system, where $\lambda_1, \lambda_2, \lambda_3$ are to be found. Recall that the element of length is given by

$$d\mathbf{r} = h_1 du_1\hat{\mathbf{e}}_1 + h_2 du_2\hat{\mathbf{e}}_2 + h_3 du_3\hat{\mathbf{e}}_3$$

Now

$$\begin{aligned}d\Phi &= \frac{\partial\Phi}{\partial u_1}du_1 + \frac{\partial\Phi}{\partial u_2}du_2 + \frac{\partial\Phi}{\partial u_3}du_3 \\ &= \frac{\partial\Phi}{\partial x}dx + \frac{\partial\Phi}{\partial y}dy + \frac{\partial\Phi}{\partial z}dz \\ &= (\nabla\Phi)\cdot d\mathbf{r}\end{aligned}$$

But, using our expressions for $\nabla\Phi$ and $d\mathbf{r}$ above:

$$(\nabla\Phi)\cdot d\mathbf{r} = \lambda_1 h_1 du_1 + \lambda_2 h_2 du_2 + \lambda_3 h_3 du_3$$

and so we see that

$$h_i\lambda_i = \frac{\partial\Phi}{\partial u_i} \; (i = 1,2,3)$$

Thus we have the result that

Proposition 8.3.1 — Gradient in Orthogonal Curvilinear Coordinates.

$$\nabla\Phi = \frac{\hat{\mathbf{e}}_1}{h_1}\frac{\partial\Phi}{\partial u_1} + \frac{\hat{\mathbf{e}}_2}{h_2}\frac{\partial\Phi}{\partial u_2} + \frac{\hat{\mathbf{e}}_3}{h_3}\frac{\partial\Phi}{\partial u_3}$$

This proposition allows us to write down ∇ easily for other coordinate systems.

(i) Cylindrical polars (r, θ, z) Recall that $h_1 = 1$, $h_2 = r$, $h_3 = 1$. Thus

$$\nabla = \hat{r}\frac{\partial}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial}{\partial \theta} + \hat{z}\frac{\partial}{\partial z}$$

(ii) Spherical Polars (r, ϕ, θ) We have $h_1 = 1$, $h_2 = r$, $h_3 = r\sin\phi$, and so

$$\nabla = \hat{r}\frac{\partial}{\partial r} + \frac{\hat{\phi}}{r}\frac{\partial}{\partial \phi} + \frac{\hat{\theta}}{r\sin\phi}\frac{\partial}{\partial \theta}$$

8.3.1 Expressions for Unit Vectors

From the expression for ∇ we have just derived, it is easy to see that

$$\hat{\mathbf{e}}_i = h_i \nabla u_i$$

Alternatively, since the unit vectors are orthogonal, if we know two unit vectors we can find the third from the relation

$$\hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_2 \times \hat{\mathbf{e}}_3 = h_2 h_3 (\nabla u_2 \times \nabla u_3)$$

and similarly for the other components, by permuting in a cyclic fashion.

8.4 Divergence in Orthogonal Curvilinear Coordinates

Suppose we have a vector field

$$\mathbf{A} = A_1\hat{\mathbf{e}}_1 + A_2\hat{\mathbf{e}}_2 + A_3\hat{\mathbf{e}}_3$$

Then consider

$$\begin{aligned}\nabla \cdot (A_1\hat{\mathbf{e}}_1) &= \nabla \cdot [A_1 h_2 h_3 (\nabla u_2 \times \nabla u_3)] \\ &= A_1 h_2 h_3 \nabla \cdot (\nabla u_2 \times \nabla u_3) + \nabla(A_1 h_2 h_3) \cdot \frac{\hat{\mathbf{e}}_1}{h_2 h_3}\end{aligned}$$

using the results established just above. Also we know that

$$\nabla \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{C} \cdot \operatorname{curl} \mathbf{B} - \mathbf{B} \cdot \operatorname{curl} \mathbf{C}$$

and so it follows that

$$\nabla \cdot (\nabla u_2 \times \nabla u_3) = (\nabla u_3) \cdot \operatorname{curl}(\nabla u_2) - (\nabla u_2) \cdot \operatorname{curl}(\nabla u_3) = 0$$

since the curl of a gradient is always zero. Thus we are left with

$$\nabla \cdot (A_1\hat{\mathbf{e}}_1) = \nabla(A_1 h_2 h_3) \cdot \frac{\hat{\mathbf{e}}_1}{h_2 h_3} = \frac{1}{h_1 h_2 h_3}\frac{\partial}{\partial u_1}(A_1 h_2 h_3)$$

We can proceed in a similar fashion for the other components, and establish that

Proposition 8.4.1 — Divergence in Orthogonal Curvilinear Coordinates.

$$\nabla\cdot\mathbf{A}=\frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(h_2h_3A_1)+\frac{\partial}{\partial u_2}(h_3h_1A_2)+\frac{\partial}{\partial u_3}(h_1h_2A_3)\right]$$

Using the above proposition is now easy to write down the divergence in other coordinate systems.

(i) Cylindrical polars (r,θ,z)

Since $h_1=1$, $h_2=r$, $h_3=1$ using the above formula we have :

$$\begin{aligned}\nabla\cdot\mathbf{A}&=\frac{1}{r}\left[\frac{\partial}{\partial r}(rA_1)+\frac{\partial}{\partial\theta}(A_2)+\frac{\partial}{\partial z}(rA_3)\right]\\&=\frac{\partial A_1}{\partial r}+\frac{A_1}{r}+\frac{1}{r}\frac{\partial A_2}{\partial\theta}+\frac{\partial A_3}{\partial z}\end{aligned}$$

(ii) Spherical polars (r,ϕ,θ)

We have $h_1=1$, $h_2=r$, $h_3=r\sin\phi$. So

$$\nabla\cdot\mathbf{A}=\frac{1}{r^2\sin\phi}\left[\frac{\partial}{\partial r}(r^2\sin\phi A_1)+\frac{\partial}{\partial\phi}(r\sin\phi A_2)+\frac{\partial}{\partial\theta}(rA_3)\right]$$

8.5 Curl in Orthogonal Curvilinear Coordinates

We will calculate the curl of the first component of $\mathbf{A}$:

$$\begin{aligned}\nabla\times(A_1\hat{\mathbf{e}}_1)&=\nabla\times(A_1h_1\nabla u_1)\\&=A_1h_2\nabla\times(\nabla u_1)+\nabla(A_1h_1)\times\nabla u_1\\&=0+\nabla(A_1h_1)\times\nabla u_1\\&=\left[\frac{\hat{\mathbf{e}}_1}{h_1}\frac{\partial}{\partial u_1}(A_1h_1)+\frac{\hat{\mathbf{e}}_2}{h_2}\frac{\partial}{\partial u_2}(A_1h_1)+\frac{\hat{\mathbf{e}}_3}{h_3}\frac{\partial}{\partial u_3}(A_1h_1)\right]\times\frac{\hat{\mathbf{e}}_1}{h_1}\\&=\frac{\hat{\mathbf{e}}_2}{h_1h_3}\frac{\partial}{\partial u_3}(h_1A_1)-\frac{\hat{\mathbf{e}}_3}{h_1h_2}\frac{\partial}{\partial u_2}(h_1A_1)\end{aligned}$$

(since $\hat{\mathbf{e}}_1\times\hat{\mathbf{e}}_1=0$, $\hat{\mathbf{e}}_2\times\hat{\mathbf{e}}_1=-\hat{\mathbf{e}}_3$, $\hat{\mathbf{e}}_3\times\hat{\mathbf{e}}_1=\hat{\mathbf{e}}_2$).

We can obviously find $\operatorname{curl}(A_2\hat{\mathbf{e}}_2)$ and $\operatorname{curl}(A_3\hat{\mathbf{e}}_3)$ in a similar way. These can be shown to be

$$\begin{aligned}\nabla\times(A_2\hat{\mathbf{e}}_2)&=\frac{\hat{\mathbf{e}}_3}{h_2h_1}\frac{\partial}{\partial u_1}(h_2A_2)-\frac{\hat{\mathbf{e}}_1}{h_2h_3}\frac{\partial}{\partial u_3}(h_2A_2)\\\nabla\times(A_3\hat{\mathbf{e}}_3)&=\frac{\hat{\mathbf{e}}_1}{h_3h_2}\frac{\partial}{\partial u_2}(h_3A_3)-\frac{\hat{\mathbf{e}}_2}{h_3h_1}\frac{\partial}{\partial u_1}(h_3A_3)\end{aligned}$$

Adding these three contributions together, we find we can write this in the form of a determinant as

Proposition 8.5.1 — Curl in Orthogonal Curvilinear Coordinates.

$$\operatorname{curl}\mathbf{A} = \frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\hat{\mathbf{e}}_1 & h_2\hat{\mathbf{e}}_2 & h_3\hat{\mathbf{e}}_3 \\ \partial_{u_1} & \partial_{u_2} & \partial_{u_3} \\ h_1A_1 & h_2A_2 & h_3A_3 \end{vmatrix}$$

It's then straightforward to write down the expressions of the curl in various orthogonal coordinate systems.

(i) Cylindrical polars

$$\operatorname{curl}\mathbf{A} = \frac{1}{r}\begin{vmatrix} \hat{r} & r\hat{\theta} & \hat{z} \\ \partial_r & \partial_\theta & \partial_z \\ A_1 & rA_2 & A_3 \end{vmatrix}$$

(ii) Spherical polars

$$\operatorname{curl}\mathbf{A} = \frac{1}{r^2\sin\phi}\begin{vmatrix} \hat{r} & r\hat{\phi} & r\sin\phi\hat{\theta} \\ \partial_r & \partial_\phi & \partial_\theta \\ A_1 & rA_2 & r\sin\phi A_3 \end{vmatrix}$$

8.6 The Laplacian in Orthogonal Curvilinear Coordinates

From the formulae already established for the gradient and the divergent, we can see that

Proposition 8.6.1 — The Laplacian in Orthogonal Curvilinear Coordinates.

$$\begin{aligned}\nabla^2\Phi &= \nabla\cdot(\nabla\Phi) \\ &= \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(h_2h_3\frac{1}{h_1}\frac{\partial\Phi}{\partial u_1}) + \frac{\partial}{\partial u_2}(h_3h_1\frac{1}{h_2}\frac{\partial\Phi}{\partial u_2}) + \frac{\partial}{\partial u_3}(h_1h_2\frac{1}{h_3}\frac{\partial\Phi}{\partial u_3})\right]\end{aligned}$$

(i) Cylindrical polars (r, θ, z)

$$\begin{aligned}\nabla^2\Phi &= \frac{1}{r}\left[\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(\frac{1}{r}\frac{\partial\Phi}{\partial\theta}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial\Phi}{\partial z}\right)\right] \\ &= \frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2} + \frac{\partial^2\Phi}{\partial z^2}\end{aligned}$$

(ii) Spherical polars (r,ϕ,θ)

$$\nabla^2\Phi = \frac{1}{r^2\sin\phi}\left[\frac{\partial}{\partial r}\left(r^2\sin\phi\frac{\partial\Phi}{\partial r}\right)+\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial\Phi}{\partial\phi}\right)+\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\phi}\frac{\partial\Phi}{\partial\theta}\right)\right]$$
$$= \frac{\partial^2\Phi}{\partial r^2}+\frac{2}{r}\frac{\partial\Phi}{\partial r}+\frac{\cot\phi}{r^2}\frac{\partial\Phi}{\partial\phi}+\frac{1}{r^2}\frac{\partial^2\Phi}{\partial\phi^2}+\frac{1}{r^2\sin^2\phi}\frac{\partial^2\Phi}{\partial\theta^2}$$

Example 8.6.2 In Example **??** we showed that $\nabla\|\mathbf{r}\|^2 = 2\mathbf{r}$ and $\Delta\|\mathbf{r}\|^2 = 6$, where $\mathbf{r}(x,y,z) = x\,\mathbf{i}+y\,\mathbf{j}+z\,\mathbf{k}$ in Cartesian coordinates. Verify that we get the same answers if we switch to spherical coordinates.

Solution: Since $\|\mathbf{r}\|^2 = x^2+y^2+z^2 = \rho^2$ in spherical coordinates, let $F(\rho,\theta,\phi) = \rho^2$ (so that $F(\rho,\theta,\phi) = \|\mathbf{r}\|^2$). The gradient of F in spherical coordinates is

$$\begin{aligned}
\nabla F &= \frac{\partial F}{\partial\rho}\mathbf{e}_\rho + \frac{1}{\rho\sin\phi}\frac{\partial F}{\partial\theta}\mathbf{e}_\theta + \frac{1}{\rho}\frac{\partial F}{\partial\phi}\mathbf{e}_\phi \\
&= 2\rho\,\mathbf{e}_\rho + \frac{1}{\rho\sin\phi}(0)\,\mathbf{e}_\theta + \frac{1}{\rho}(0)\,\mathbf{e}_\phi \\
&= 2\rho\,\mathbf{e}_\rho = 2\rho\frac{\mathbf{r}}{\|\mathbf{r}\|}\text{ , as we showed earlier, so} \\
&= 2\rho\frac{\mathbf{r}}{\rho} = 2\mathbf{r}\text{ , as expected. And the Laplacian is} \\
\Delta F &= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial F}{\partial\rho}\right) + \frac{1}{\rho^2\sin^2\phi}\frac{\partial^2 F}{\partial\theta^2} + \frac{1}{\rho^2\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial F}{\partial\phi}\right) \\
&= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}(\rho^2 2\rho) + \frac{1}{\rho^2\sin\phi}(0) + \frac{1}{\rho^2\sin\phi}\frac{\partial}{\partial\phi}(\sin\phi\,(0)) \\
&= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}(2\rho^3) + 0 + 0 \\
&= \frac{1}{\rho^2}(6\rho^2) = 6\text{ , as expected.}
\end{aligned}$$

8.7 Examples of Orthogonal Coordinates

Spherical Polar Coordinates

$(r,\phi,\theta)\in[0,\infty)\times[0,\pi]\times[0,2\pi)$

$$x = r\sin\phi\cos\theta \tag{8.1}$$
$$y = r\sin\phi\sin\theta \tag{8.2}$$
$$z = r\cos\phi \tag{8.3}$$

$\nabla\Phi$	$=$	$\frac{\hat{\mathbf{e}}_1}{h_1}\frac{\partial\Phi}{\partial u_1}+\frac{\hat{\mathbf{e}}_2}{h_2}\frac{\partial\Phi}{\partial u_2}+\frac{\hat{\mathbf{e}}_3}{h_3}\frac{\partial\Phi}{\partial u_3}$
$\nabla\cdot\mathbf{A}$	$=$	$\frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(h_2h_3A_1)+\frac{\partial}{\partial u_2}(h_3h_1A_2)+\frac{\partial}{\partial u_3}(h_1h_2A_3)\right]$
$\operatorname{curl}\mathbf{A}$	$=$	$\frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\hat{\mathbf{e}}_1 & h_2\hat{\mathbf{e}}_2 & h_3\hat{\mathbf{e}}_3 \\ \partial_{u_1} & \partial_{u_2} & \partial_{u_3} \\ h_1A_1 & h_2A_2 & h_3A_3 \end{vmatrix}$
$\nabla^2\Phi$	$=$	$\frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(h_2h_3\frac{1}{h_1}\frac{\partial\Phi}{\partial u_1})+\frac{\partial}{\partial u_2}(h_3h_1\frac{1}{h_2}\frac{\partial\Phi}{\partial u_2})+\frac{\partial}{\partial u_3}(h_1h_2\frac{1}{h_3}\frac{\partial\Phi}{\partial u_3})\right]$

Table 8.1 Vector operators in orthogonal curvilinear coordinates u_1, u_2, u_3.

The scale factors for the Spherical Polar Coordinates are:

$$h_1 = 1 \tag{8.4}$$

$$h_2 = r \tag{8.5}$$

$$h_3 = r\sin\phi \tag{8.6}$$

Cylindrical Polar Coordinates

$(r,\theta,z) \in [0,\infty)\times[0,2\pi)\times(-\infty,\infty)$

$$x = r\cos\theta \tag{8.7}$$

$$y = r\sin\theta \tag{8.8}$$

$$z = z \tag{8.9}$$

The scale factors for the Cylindrical Polar Coordinates are:

$$h_1 = h_3 = 1 \tag{8.10}$$

$$h_2 = r \tag{8.11}$$

Parabolic Cylindrical Coordinates

$(u,v,z) \in (-\infty,\infty)\times[0,\infty)\times(-\infty,\infty)$

$$x = \frac{1}{2}(u^2-v^2) \tag{8.12}$$

$$y = uv \tag{8.13}$$

$$z = z \tag{8.14}$$

The scale factors for the Parabolic Cylindrical Coordinates are:

$$h_1 = h_2 = \sqrt{u^2 + v^2} \tag{8.15}$$
$$h_3 = 1 \tag{8.16}$$

Paraboloidal Coordinates

$(u, v, \theta) \in [0, \infty) \times [0, \infty) \times [0, 2\pi)$

$$x = uv \cos\theta \tag{8.17}$$
$$y = uv \sin\theta \tag{8.18}$$
$$z = \frac{1}{2}(u^2 - v^2) \tag{8.19}$$

The scale factors for the Paraboloidal Coordinates are:

$$h_1 = h_2 = \sqrt{u^2 + v^2} \tag{8.20}$$
$$h_3 = uv \tag{8.21}$$

Elliptic Cylindrical Coordinates

$(u, v, z) \in [0, \infty) \times [0, 2\pi) \times (-\infty, \infty)$

$$x = a \cosh u \cos v \tag{8.22}$$
$$y = a \sinh u \sin v \tag{8.23}$$
$$z = z \tag{8.24}$$

The scale factors for the Elliptic Cylindrical Coordinates are:

$$h_1 = h_2 = a\sqrt{\sinh^2 u + \sin^2 v} \tag{8.25}$$
$$h_3 = 1 \tag{8.26}$$

Prolate Spheroidal Coordinates

$(\xi, \eta, \theta) \in [0, \infty) \times [0, \pi] \times [0, 2\pi)$

$$x = a \sinh\xi \sin\eta \cos\theta \tag{8.27}$$
$$y = a \sinh\xi \sin\eta \sin\theta \tag{8.28}$$
$$z = a \cosh\xi \cos\eta \tag{8.29}$$

The scale factors for the Prolate Spheroidal Coordinates are:

$$h_1 = h_2 = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{8.30}$$
$$h_3 = a \sinh\xi \sin\eta \tag{8.31}$$

Oblate Spheroidal Coordinates

$(\xi,\eta,\theta) \in [0,\infty) \times \left[-\frac{\pi}{2},\frac{\pi}{2}\right] \times [0,2\pi)$

$$x = a\cosh\xi\cos\eta\cos\theta \tag{8.32}$$
$$y = a\cosh\xi\cos\eta\sin\theta \tag{8.33}$$
$$z = a\sinh\xi\sin\eta \tag{8.34}$$

The scale factors for the Oblate Spheroidal Coordinates are:

$$h_1 = h_2 = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{8.35}$$
$$h_3 = a\cosh\xi\cos\eta \tag{8.36}$$

Ellipsoidal Coordinates

$$(\lambda,\mu,\nu) \tag{8.37}$$
$$\lambda < c^2 < b^2 < a^2, \tag{8.38}$$
$$c^2 < \mu < b^2 < a^2, \tag{8.39}$$
$$c^2 < b^2 < \nu < a^2, \tag{8.40}$$

$\frac{x^2}{a^2-q_i} + \frac{y^2}{b^2-q_i} + \frac{z^2}{c^2-q_i} = 1$ where $(q_1,q_2,q_3) = (\lambda,\mu,\nu)$

The scale factors for the Ellipsoidal Coordinates are: $h_i = \frac{1}{2}\sqrt{\frac{(q_j-q_i)(q_k-q_i)}{(a^2-q_i)(b^2-q_i)(c^2-q_i)}}$

Bipolar Coordinates

$(u,v,z) \in [0,2\pi) \times (-\infty,\infty) \times (-\infty,\infty)$

$$x = \frac{a\sinh v}{\cosh v - \cos u} \tag{8.41}$$
$$y = \frac{a\sin u}{\cosh v - \cos u} \tag{8.42}$$
$$z = z \tag{8.43}$$

The scale factors for the Bipolar Coordinates are:

$$h_1 = h_2 = \frac{a}{\cosh v - \cos u} \tag{8.44}$$
$$h_3 = 1 \tag{8.45}$$

Toroidal Coordinates

$(u,v,\theta) \in (-\pi,\pi] \times [0,\infty) \times [0,2\pi)$

$$x = \frac{a\sinh v\cos\theta}{\cosh v - \cos u} \tag{8.46}$$

$$y = \frac{a\sinh v\sin\theta}{\cosh v - \cos u} \tag{8.47}$$

$$z = \frac{a\sin u}{\cosh v - \cos u} \tag{8.48}$$

The scale factors for the Toroidal Coordinates are:

$$h_1 = h_2 = \frac{a}{\cosh v - \cos u} \tag{8.49}$$

$$h_3 = \frac{a\sinh v}{\cosh v - \cos u} \tag{8.50}$$

Conical Coordinates

$$(\lambda,\mu,\nu) \tag{8.51}$$

$$\nu^2 < b^2 < \mu^2 < a^2 \tag{8.52}$$

$$\lambda \in [0,\infty) \tag{8.53}$$

$$x = \frac{\lambda\mu\nu}{ab} \tag{8.54}$$

$$y = \frac{\lambda}{a}\sqrt{\frac{(\mu^2-a^2)(\nu^2-a^2)}{a^2-b^2}} \tag{8.55}$$

$$z = \frac{\lambda}{b}\sqrt{\frac{(\mu^2-b^2)(\nu^2-b^2)}{a^2-b^2}} \tag{8.56}$$

The scale factors for the Conical Coordinates are:

$$h_1 = 1 \tag{8.57}$$

$$h_2^2 = \frac{\lambda^2(\mu^2-\nu^2)}{(\mu^2-a^2)(b^2-\mu^2)} \tag{8.58}$$

$$h_3^2 = \frac{\lambda^2(\mu^2-\nu^2)}{(\nu^2-a^2)(\nu^2-b^2)} \tag{8.59}$$

8.8 Alternative Definitions for Grad, Div, Curl

Let V be a region enclosed by a surface S and let P be a general point of V. We established earlier that

$$\int_V \nabla\theta\,dV = \int_S \hat{n}\theta\,dS$$

It follows that

$$\int_V \hat{\imath}\cdot\nabla\theta\,dV = \int_S (\hat{\imath}\cdot\hat{n})\theta\,dS$$

Now the left hand side of the above equation can be written as $V\overline{\{\hat{\imath}\cdot\nabla\theta\}}$, where the bar denotes the mean value of this quantity over V. Since we are assuming that θ has continuous derivates throughout V, we can write

$$\overline{\{\hat{\imath}\cdot\nabla\theta\}} = \{\hat{\imath}\cdot\nabla\theta\}_Q$$

for some point Q of V. Thus we have that

$$\{\hat{\imath}\cdot\nabla\theta\}_Q = \frac{1}{V}\int_S (\hat{\imath}\cdot\hat{n})\theta\,dS$$

Now let $V\to 0$ about P. Then $P\to Q$ and we have that at any point P of V:

$$\hat{\imath}\cdot\nabla\theta = \lim_{V\to 0}\frac{1}{V}\int_S (\hat{\imath}\cdot\hat{n})\theta\,dS$$

Similar results can be established for $\hat{\jmath}\cdot\nabla\theta$ and $\hat{k}\cdot\nabla\theta$. Taken together, these imply that

$$\nabla\theta = \lim_{V\to 0}\frac{1}{V}\int_S \hat{n}\theta\,dS$$

This can be regarded an an alternative way of defining $\nabla\theta$, rather than defining it as $(\partial\theta/\partial x)\hat{\imath} + (\partial\theta/\partial u)\hat{\jmath} + (\partial\theta/\partial z)\hat{k}$. We can similarly establish that

$$\operatorname{div}\mathbf{A} = \lim_{V\to 0}\frac{1}{V}\int_S (\hat{n}\cdot\mathbf{A})\,dS$$
$$\operatorname{curl}\mathbf{A} = \lim_{V\to 0}\frac{1}{V}\int_S (\hat{n}\times\mathbf{A})\,dS$$

which are alternative definitions of the divergence and curl, and are clearly independent of the choice of coordinates, which is one of the advantages of this approach. In particular we can see that the divergence is a measure of the flux of a quantity.

Equivalence of definitions

Let's show that the definition of divergence given here is consistent with the curvilinear formula given earlier. Consider δV to be the volume of a curvilinear volume element located at the point P, with edges of length $h_1\delta u_1$, $h_2\delta u_2$, $h_3\delta u_3$, and unit vectors aligned as shown in the picture:

The volume of the element $\delta V \approx h_1h_2h_3\,\delta u_1\delta u_2\delta u_3$. We start with our definition

$$\operatorname{div}\mathbf{A} = \lim_{V\to 0}\frac{1}{V}\int_S (\hat{n}\cdot\mathbf{A})\,dS$$

and aim to compute explicitly the right-hand-side. This involves calculating the contributions to $\int_S$ arising from the six faces of the volume element. If we start with the contribution from the face $PP'S'S$, this is

$$-(A_1h_2h_3)_P\,\delta u_2\delta u_3 + \text{ higher order terms}$$

The contribution from the face $QQ'R'R$ is

$$(A_1h_2h_3)_Q\delta u_3\delta u_3 + \text{h.o.t} = \left[(A_1h_2h_3) + \frac{\partial}{\partial u_1}(A_1h_2h_3)\delta u_1\right]_P \delta u_2\delta u_3 + \text{h.o.t}$$

using a Taylor series expansion. Adding together the contributions from these two faces, we get

$$\left[\frac{\partial}{\partial u_1}(A_1h_2h_3)\right]_P \delta u_1\delta u_2\delta u_3 + \text{ h.o.t.}$$

Similarly the sum of the contributions from the faces $PSRQ$, $P'S'R'Q'$ is

$$\left[\frac{\partial}{\partial u_3}(A_3h_1h_2)\right]_P \delta u_1\delta u_2\delta u_3 + \text{ h.o.t.}$$

while the combined contributions from $PQQ'P'$, $SRR'S'$ is

$$\left[\frac{\partial}{\partial u_2}(A_2h_3h_1)\right]_P \delta u_1\delta u_2\delta u_3 + \text{ h.o.t.}$$

If we then let $\delta V \to 0$, we have that

$$\lim_{\delta V\to 0}\frac{1}{\delta V}\int_S \hat{n}\cdot\mathbf{A}\,dS = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}(A_1h_2h_3) + \frac{\partial}{\partial u_2}(A_2h_3h_1) + \frac{\partial}{\partial u_3}(A_3h_1h_2)\right]$$

and so we can see that the integral expression for div $\mathbf{A}$ is consistent with the formula in curvilinear coordinates derived earlier.

Exercises

A

For Exercises 1-6, find the Laplacian of the function $f(x,y,z)$ in Cartesian coordinates.

1. $f(x,y,z) = x + y + z$

2. $f(x,y,z) = x^5$

3. $f(x,y,z) = (x^2 + y^2 + z^2)^{3/2}$

4. $f(x,y,z) = e^{x+y+z}$

5. $f(x,y,z) = x^3 + y^3 + z^3$

6. $f(x,y,z) = e^{-x^2-y^2-z^2}$

7. Find the Laplacian of the function in Exercise 3 in spherical coordinates.

8. Find the Laplacian of the function in Exercise 6 in spherical coordinates.

9. Let $f(x,y,z) = \dfrac{z}{x^2+y^2}$ in Cartesian coordinates. Find ∇f in cylindrical coordinates.

10. For $\mathbf{f}(r,\theta,z) = r\,\mathbf{e}_r + z\sin\theta\,\mathbf{e}_\theta + rz\,\mathbf{e}_z$ in cylindrical coordinates, find div $\mathbf{f}$ and curl $\mathbf{f}$.

11. For $\mathbf{f}(\rho,\theta,\phi) = \mathbf{e}_\rho + \rho\cos\theta\,\mathbf{e}_\theta + \rho\,\mathbf{e}_\phi$ in spherical coordinates, find div $\mathbf{f}$ and curl $\mathbf{f}$.

B

For Exercises 12-23, prove the given formula ($r = \|\mathbf{r}\|$ is the length of the position vector field $\mathbf{r}(x,y,z) = x\,\mathbf{i} + y\,\mathbf{j} + z\,\mathbf{k}$).

12. $\nabla(1/r) = -\mathbf{r}/r^3$ **13.** $\Delta(1/r) = 0$ **14.** $\nabla\cdot(\mathbf{r}/r^3) = 0$ **15.** $\nabla(\ln r) = \mathbf{r}/r^2$

16. $\operatorname{div}(\mathbf{F}+\mathbf{G}) = \operatorname{div}\mathbf{F} + \operatorname{div}\mathbf{G}$

17. $\operatorname{curl}(\mathbf{F}+\mathbf{G}) = \operatorname{curl}\mathbf{F} + \operatorname{curl}\mathbf{G}$

18. $\operatorname{div}(f\mathbf{F}) = f\operatorname{div}\mathbf{F} + \mathbf{F}\cdot\nabla f$

19. $\operatorname{div}(\mathbf{F}\times\mathbf{G}) = \mathbf{G}\cdot\operatorname{curl}\mathbf{F} - \mathbf{F}\cdot\operatorname{curl}\mathbf{G}$

20. $\operatorname{div}(\nabla f\times\nabla g) = 0$

21. $\operatorname{curl}(f\mathbf{F}) = f\operatorname{curl}\mathbf{F} + (\nabla f)\times\mathbf{F}$

22. $\operatorname{curl}(\operatorname{curl}\mathbf{F}) = \nabla(\operatorname{div}\mathbf{F}) - \Delta\mathbf{F}$

23. $\Delta(fg) = f\Delta g + g\Delta f + 2(\nabla f\cdot\nabla g)$

C

24. Derive the gradient formula in cylindrical coordinates: $\nabla F = \frac{\partial F}{\partial r}\mathbf{e}_r + \frac{1}{r}\frac{\partial F}{\partial \theta}\mathbf{e}_\theta + \frac{\partial F}{\partial z}\mathbf{e}_z$

25. Use $\mathbf{f} = u\nabla v$ in the Divergence Theorem to prove:

(a) *Green's first identity:* $\iiint\limits_S (u\Delta v + (\nabla u)\cdot(\nabla v))\,dV = \iint\limits_\Sigma (u\nabla v)\cdot d\sigma$

(b) *Green's second identity:* $\iiint\limits_S (u\Delta v - v\Delta u)\,dV = \iint\limits_\Sigma (u\nabla v - v\nabla u)\cdot d\sigma$

26. Suppose that $\Delta u = 0$ (i.e. u is *harmonic*) over $\mathbb{R}^3$. Define the *normal derivative* $\frac{\partial u}{\partial n}$ of u over a closed surface Σ with outward unit normal vector $\mathbf{n}$ by $\frac{\partial u}{\partial n} = D_{\mathbf{n}}u = \mathbf{n}\cdot\nabla u$. Show that $\iint\limits_\Sigma \frac{\partial u}{\partial n}\,d\sigma = 0$. (*Hint: Use Green's second identity.*)

CHAPTER 9

Tensors

In this chapter we define a tensor as a multilinear map.

9.1 Linear Functional

Definition 9.1.1 A function $f:\mathbb{R}^n \to \mathbb{R}$ is a **linear functional** if satisfies the

$$\text{linearity condition:} \quad f(a\mathbf{u}+b\mathbf{v}) = af(\mathbf{u})+bf(\mathbf{v}),$$

or in words: “the value on a linear combination is the the linear combination of the values."

A linear functional is also called **linear function**, **1-form**, or **covector**.

This easily extends to linear combinations with any number of terms; for example

$$f(\mathbf{v}) = f\left(\sum_{i=1}^{N} v_i \mathbf{e}_i\right) = \sum_{i=1}^{N} v_i f(\mathbf{e}_i)$$

where the coefficients $f_i \equiv f(\mathbf{e}_i)$ are the “components" of a covector with respect to the basis $\{\mathbf{e}_i\}$, or in our shorthand notation

$$\begin{aligned} f(\mathbf{v}) &= f(v_i \mathbf{e}_i) && \text{(express in terms of basis)} \\ &= v_i f(\mathbf{e}_i) && \text{(linearity)} \\ &= v_i f_i\,. && \text{(definition of components)} \end{aligned}$$

A covector f is entirely determined by its values f_i on the basis vectors, namely its components with respect to that basis.

Our linearity condition is usually presented separately as a pair of separate conditions on the two operations which define a vector space:

- sum rule: the value of the function on a sum of vectors is the sum of the values, $f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u})+f(\mathbf{v})$,
- scalar multiple rule: the value of the function on a scalar multiple of a vector is the scalar times the value on the vector, $f(c\mathbf{u}) = cf(\mathbf{u})$.

Example 9.1.1 In the usual notation on $\mathbb{R}^3$, with Cartesian coordinates $(x^1,x^2,x^3) = (x,y,z)$, linear functions are of the form $f(x,y,z) = ax+by+cz$,

Example 9.1.2 If we fixed a vector $\mathbf{n}$ we have a function $\mathbf{n}^* : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathbf{n}^*(\mathbf{v}) := \mathbf{n}\cdot\mathbf{v}$$

is a linear function.

9.2 Dual Spaces

Definition 9.2.1 We define the **dual space** of $\mathbb{R}^n$, denoted as $(\mathbb{R}^n)^*$, as the set of all real-valued linear functions on $\mathbb{R}^n$;

$$(\mathbb{R}^n)^* = \{f : f : \mathbb{R}^n \to \mathbb{R} \text{ is a linear function } \}$$

The dual space $(\mathbb{R}^n)^*$ is itself an n-dimensional vector space, with linear combinations of covectors defined in the usual way that one can takes linear combinations of any functions, i.e., in terms of values

$$\text{covector addition:} \quad (af+bg)(\mathbf{v}) \equiv af(\mathbf{v})+bg(\mathbf{v}), \qquad f,g \text{ covectors, } v \text{ a vector}.$$

Theorem 9.2.1 Suppose that vectors in $\mathbb{R}^n$ represented as column vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

For each row vector

$$[a] = [a^1 \dots a^n]$$

there is a linear functional f defined by

$$f(\mathbf{x}) = [a_1 \dots a_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

$$f(\mathbf{x}) = a_1 x_1 + \cdots + a_n x_n,$$

and each linear functional in $\mathbb{R}^n$ can be expressed in this form

R As consequence of the previous theorem we can see vectors as column and covectors as row matrix. And the action of covectors in vectors as the matrix product of the row vector and the column vector.

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, x_i \in \mathbb{R} \right\} \tag{9.1}$$

$$(\mathbb{R}^n)^* = \{[a_1 \dots a_n], a_i \in \mathbb{R}\} \tag{9.2}$$

Remark 9.2.2 closure of the dual space

Show that the dual space is closed under this linear combination operation. In other words, show that if f, g are linear functions, satisfying our linearity condition, then $a\,f + b\,g$ also satisfies the linearity condition for linear functions:

$$(a\,f + b\,g)(c_1\mathbf{u} + c_2\mathbf{v}) = c_1(a\,f + b\,g)(\mathbf{u}) + c_2(a\,f + b\,g)(\mathbf{v}).$$

9.2.1 Duas Basis

Let us produce a basis for $(\mathbb{R}^n)^*$, called the dual basis $\{\mathbf{e}^i\}$ or "the basis dual to $\{\mathbf{e}_i\}$," by defining n covectors which satisfy the following "duality relations"

$$\mathbf{e}^i(\mathbf{e}_j) = \delta^i{}_j \equiv \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{if } i \neq j\,, \end{cases}$$

where the symbol $\delta^i{}_j$ is called the "Kronecker delta," nothing more than a symbol for the components of the $n \times n$ identity matrix $I = (\delta^i{}_j)$. We then extend them to any other vector by linearity. Then by linearity

$$\begin{aligned} \mathbf{e}^i(\mathbf{v}) &= \mathbf{e}^i(v_j\mathbf{e}_j) && \text{(expand in basis)} \\ &= v_j\mathbf{e}^i(\mathbf{e}_j) && \text{(linearity)} \\ &= v_j\delta^i{}_j && \text{(duality)} \\ &= v_i && \text{(Kronecker delta definition)} \end{aligned}$$

where the last equality follows since for each i, only the term with $j = i$ in the sum over j contributes to the sum. Alternatively matrix multiplication of a vector on the left by the identity matrix $\delta^i{}_j v_j = v_i$ does not change the vector. Thus the calculation shows that the i-th dual basis covector $\mathbf{e}^i$ picks out the i-th component v_i of a vector v.

Theorem 9.2.3 The n covectors $\{\mathbf{e}^i\}$ form a basis of $(\mathbb{R}^n)^*$.

Proof:

1. **spanning condition:**
   Using linearity and the definition $f_i = f(\mathbf{e}_i)$, this calculation shows that every linear function f can be written as a linear combination of these covectors

$$\begin{aligned} f(\mathbf{v}) &= f(v_i\mathbf{e}_i) && \text{(expand in basis)} \\ &= v_i f(\mathbf{e}_i) && \text{(linearity)} \\ &= v_i f_i && \text{(definition of components)} \\ &= v_i \delta^j{}_i f_j && \text{(Kronecker delta definition)} \\ &= v_i \mathbf{e}^j(\mathbf{e}_i) f_j && \text{(dual basis definition)} \\ &= (f_j\mathbf{e}^j)(v^i\mathbf{e}_i) && \text{(linearity)} \\ &= (f_j\mathbf{e}^j)(\mathbf{v})\,. && \text{(expansion in basis, in reverse)} \end{aligned}$$

   Thus f and $f_i\mathbf{e}^i$ have the same value on every $\mathbf{v} \in \mathbb{R}^n$ so they are the same function: $f = f_i\mathbf{e}^i$, where $f_i = f(\mathbf{e}_i)$ are the "components" of f with respect to the basis $\{\mathbf{e}^i\}$ of $(\mathbb{R}^n)^*$ also said to be the “components" of f with respect to the basis $\{\mathbf{e}_i\}$ of $\mathbb{R}^n$ already introduced above. The index i on f_i labels the components of f, while the index i on $\mathbf{e}^i$ labels the dual basis covectors.

2. **linear independence**:
Suppose $f_i\mathbf{e}^i = 0$ is the zero covector. Then evaluating each side of this equation on $\mathbf{e}_j$ and using linearity

$$\begin{aligned}
0 &= 0(\mathbf{e}_j) && \text{(zero scalar = value of zero linear function)}\\
&= (f_i\mathbf{e}^i)(\mathbf{e}_j) && \text{(expand zero vector in basis)}\\
&= f_i\mathbf{e}^i(\mathbf{e}_j) && \text{(definition of linear combination function value)}\\
&= f_i\delta^i{}_j && \text{(duality)}\\
&= f_j && \text{(Knonecker delta definition)}
\end{aligned}$$

forces all the coefficients of $\mathbf{e}^i$ to vanish, i.e., no nontrivial linear combination of these covectors exists which equals the zero covector so these covectors are linearly independent. Thus $(\mathbb{R}^n)^*$ is also an n-dimensional vector space.

■

9.3 Bilinear Forms

A bilinear form is a function that is linear in each argument separately:

1. $B(\mathbf{u}+\mathbf{v},\mathbf{w}) = B(\mathbf{u},\mathbf{w})+B(\mathbf{v},\mathbf{w})$ and $B(\lambda\mathbf{u},\mathbf{v}) = \lambda B(\mathbf{u},\mathbf{v})$
2. $B(\mathbf{u},\mathbf{v}+\mathbf{w}) = B(\mathbf{u},\mathbf{v})+B(\mathbf{u},\mathbf{w})$ and $B(\mathbf{u},\lambda\mathbf{v}) = \lambda B(\mathbf{u},\mathbf{v})$

Let $f(\mathbf{v},\mathbf{w})$ be a bilinear form and let $e_1,\ldots,\mathbf{e}_n$ be a basis in this space. The numbers f_{ij} determined by formula

$$f_{ij} = f(\mathbf{e}_i,\mathbf{e}_j) \tag{9.3}$$

are called the **coordinates** or the **components** of the form f in the basis $e_1,\ldots,\mathbf{e}_n$. The numbers 9.3 are written in form of a matrix

$$F = \begin{bmatrix} f_{11} & \ldots & f_{1n}\\ \vdots & \ddots & \vdots\\ f_{n1} & \ldots & f_{nn}\end{bmatrix}, \tag{9.4}$$

which is called the matrix of the bilinear form f in the basis $e_1, \dots, \mathbf{e}_n$. For the element f_{ij} in the matrix 9.4 the first index i specifies the row number, the second index j specifies the column number. The matrix of a symmetric bilinear form g is also symmetric: $g_{ij} = g_{ji}$. Further, saying the matrix of a quadratic form $g(v)$, we shall assume the matrix of an associated symmetric bilinear form $g(\mathbf{v}, \mathbf{w})$.

Let $v^1, \dots, v^n$ and $w^1, \dots, w^n$ be coordinates of two vectors $\mathbf{v}$ and $\mathbf{w}$ in the basis $e_1, \dots, \mathbf{e}_n$. Then the values $f(\mathbf{v}, \mathbf{w})$ and $g(v)$ of a bilinear form and of a quadratic form respectively are calculated by the following formulas:

$$f(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} v^i w^j, g(\mathbf{v}) = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} v^i v^j. \tag{9.5}$$

In the case when g_{ij} is a diagonal matrix, the formula for $g(v)$ contains only the squares of coordinates of a vector $\mathbf{v}$:

$$g(v) = g_{11} (v^1)^2 + \cdots + g_{nn} (v^n)^2. \tag{9.6}$$

This supports the term quadratic form. Bringing a quadratic form to the form 9.6 by means of choosing proper basis $e_1, \dots, \mathbf{e}_n$ in a linear space V is one of the problems which are solved in the theory of quadratic form.

9.4 Tensor

Let $V = \mathbb{R}^n$ and let $V^* = \mathbb{R}^{n*}$ denote its dual space. We let

$$V^k = \underbrace{V \times \cdots \times V}_{\text{k times}}.$$

Definition 9.4.1 A k-**multilinear map** on V is a function $T : V^k \to \mathbb{R}$ which is linear in each variable.

$$T(\mathbf{v}_1, \dots, \lambda \mathbf{v} + \mathbf{w}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_k) = \lambda T(\mathbf{v}_1, \dots, \mathbf{v}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_k) + T(\mathbf{v}_1, \dots, \mathbf{w}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_k)$$

In other words, given $(k-1)$ vectors $v_1, v_2, \dots, v_{i-1}, v_{i+1}, \dots, v_k$, the map $T_i : V \to \mathbb{R}$ defined by $T_i(\mathbf{v}) = T(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_k)$ is linear.

Definition 9.4.2

- A **tensor of type** (r,s) on V is a multilinear map $T\colon V^r \times (V^*)^s \to \mathbb{R}$.
- A **covariant** k-tensor on V is a multilinear map $T\colon V^k \to \mathbb{R}$
- A **contravariant** k-tensor on V is a multilinear map $T\colon (V^*)^k \to \mathbb{R}$.

In other words, a covariant k-tensor is a tensor of type $(k,0)$ and a contravariant k-tensor is a tensor of type $(0,k)$.

Example 9.4.1

- Vectors can be seem as functions $V^* \to \mathbb{R}$, so vectors are contravariant tensor.
- Linear functionals are covariant tensors.
- Inner product are functions from $V \times V \to \mathbb{R}$ so covariant tensor.
- The determinant of a matrix is an multilinear function of the columns (or rows) of a square matrix, so is a covariant tensor.

The above terminology seems backwards, Michael Spivak explains:

> ”Nowadays such situations are always distinguished by calling the things which go in the same direction “covariant” and the things which go in the opposite direction “contravariant.” Classical terminology used these same words, and it just happens to have reversed this... And no one had the gall or authority to reverse terminology sanctified by years of usage. So it’s very easy to remember which kind of tensor is covariant, and which is contravariant — it’s just the opposite of what it logically ought to be.”

Definition 9.4.3 We denote the **space of tensors** of type (r,s) by $T^r_s(V)$. We will write also:

$$T^r_s(V) = \underbrace{V^* \otimes \cdots \otimes V^*}_{r} \otimes \underbrace{V \otimes \cdots \otimes V}_{s} = V^{*\otimes r} \otimes V^{\otimes s}.$$

The operation $\otimes$ is denoted tensor product and will be explained better latter.

So, in particular,

$$T^k(V) := T^k_0(V) = \{\text{covariant } k\text{-tensors}\}$$
$$T_k(V) := T^0_k(V) = \{\text{contravariant } k\text{-tensors}\}.$$

Two important special cases are:

$$\mathsf{T}^1(V) = \{\text{covariant 1-tensors}\} = V^*$$
$$\mathsf{T}_1(V) = \{\text{contravariant 1-tensors}\} = V^{**} \cong V.$$

This last line means that we can regard vectors $\mathbf{v} \in V$ as contravariant 1-tensors. That is, every vector $\mathbf{v} \in V$ can be regarded as a linear functional $V^* \to \mathbb{R}$ via

$$v(\omega) := \omega(\mathbf{v}),$$

where $\omega \in V^*$.

The **rank of an** (r,s)**-tensor** is defined to be $r+s$.

In particular, vectors (contravariant 1-tensors) and dual vectors (covariant 1-tensors) have rank 1.

Definition 9.4.4 If $\mathsf{S} \in \mathsf{T}^{r_1}_{s_1}(V)$ is an (r_1,s_1)-tensor, and $\mathsf{T} \in \mathsf{T}^{r_2}_{s_2}(V)$ is an (r_2,s_2)-tensor, we can define their **tensor product** $\mathsf{S} \otimes \mathsf{T} \in \mathsf{T}^{r_1+r_2}_{s_1+s_2}(V)$ by

$$(\mathsf{S}\otimes\mathsf{T})(v_1,\dots,v_{r_1+r_2},\omega_1,\dots,\omega_{s_1+s_2}) = \mathsf{S}(v_1,\dots,v_{r_1},\omega_1,\dots,\omega_{s_1})\cdot\mathsf{T}(v_{r_1+1},\dots,v_{r_1+r_2},\omega_{s_1+1},\dots,\omega_{s_1+s_2}).$$

Example 9.4.2 Let $u, \mathbf{v} \in V$. Again, since $V \cong \mathsf{T}_1(V)$, we can regard $u, \mathbf{v} \in \mathsf{T}_1(V)$ as $(0,1)$-tensors. Their tensor product $\mathbf{u} \otimes \mathbf{v} \in \mathsf{T}_2(V)$ is a $(0,2)$-tensor defined by

$$(\mathbf{u}\otimes\mathbf{v})(\omega,\eta) = u(\omega)\cdot v(\eta)$$

Example 9.4.3 Let $V = \mathbb{R}^3$. Write $\mathbf{u} = (1,2,3) \in V$ in the standard basis, and $\eta = (4,5,6)^\top \in V^*$ in the dual basis. For the inputs, let's also write $\omega = (x,y,z)^\top \in V^*$ and $\mathbf{v} = (p,q,r) \in V$. Then

$$\begin{aligned}
(\mathbf{u}\otimes\eta)(\omega,\mathbf{v}) &= u(\omega)\cdot\eta(\mathbf{v}) \\
&= \begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}x,y,z\end{bmatrix}^\top \cdot \begin{bmatrix}4,5,6\end{bmatrix}^\top \begin{bmatrix}p\\q\\r\end{bmatrix} \\
&= (x+2y+3z)(4p+5q+6r) \\
&= 4px+5qx+6rx \\
&\quad 8py+10qy+12py \\
&\quad 12pz+15qz+18rz
\end{aligned}$$

$$= \begin{bmatrix} x, y, z \end{bmatrix}^\top \begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix}$$
$$= \omega \begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix} v.$$

Exemplo 9.1 If S has components $\alpha_i{}^j{}_k$, and T has components β^{rs} then $\mathsf{S} \otimes \mathsf{T}$ has components $\alpha_i{}^j{}_k \beta^{rs}$, because

$$\mathsf{S} \otimes \mathsf{T}(u_i, u^j, u_k, u^r, u^s) = \mathsf{S}(u_i, u^j, u_k)\mathsf{T}(u^r, u^s).$$

■

Tensors satisfy algebraic laws such as:

(i) $\mathsf{R} \otimes (\mathsf{S} + \mathsf{T}) = \mathsf{R} \otimes \mathsf{S} + \mathsf{R} \otimes \mathsf{T}$,

(ii) $(\lambda \mathsf{R}) \otimes \mathsf{S} = \lambda(\mathsf{R} \otimes \mathsf{S}) = \mathsf{R} \otimes (\lambda \mathsf{S})$,

(iii) $(\mathsf{R} \otimes \mathsf{S}) \otimes \mathsf{T} = \mathsf{R} \otimes (\mathsf{S} \otimes \mathsf{T})$.

But

$$\mathsf{S}) \otimes \mathsf{T} \neq \mathsf{T} \otimes \mathsf{S})$$

in general. To prove those we look at components wrt a basis, and note that

$$\alpha^i{}_{jk}(\beta^r{}_s + \gamma^r{}_s) = \alpha^i{}_{jk}\beta^r{}_s + \alpha^i{}_{jk}\gamma^r{}_s,$$

for example, but

$$\alpha^i \beta^j \neq \beta^j \alpha^i$$

in general.

Some authors take the definition of an (r, s)-tensor to mean a multilinear map $V^s \times (V^*)^r \to \mathbb{R}$ (note that the r and s are reversed).

9.4.1 Basis of Tensor

Theorem 9.4.4 Let $\mathsf{T}^r_s(V)$ be the space of tensors of type (r, s). Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for V, and $\{\mathbf{e}^1, \ldots, \mathbf{e}^n\}$ be the dual basis for V^*

Then

$$\{\mathbf{e}^{j_1} \otimes \ldots, \otimes \mathbf{e}^{j_r} \otimes \mathbf{e}_{j_{r+1}} \otimes \ldots \otimes \mathbf{e}_{j_n} \quad 1 \leqslant j_i \leqslant n\}$$

is a base for $\mathsf{T}^r_s(V)$.

So any tensor $T \in T^r_s(V)$ can be written as combination of this basis. Let $T \in T^r_s(V)$ be a (r,s) tensor and let $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$ be a basis for V, and $\{\mathbf{e}^1,\dots,\mathbf{e}^n\}$ be the dual basis for V^* then we can define a collection of scalars $Aj_1\cdots j_n{}^{r+s}$ by

$$T(\mathbf{e}_{j_1},\dots,\mathbf{e}_{j_r},\mathbf{e}^{j_{r+1}}\dots\mathbf{e}^{j_n}) = Aj_1\cdots j_r{}^{j_{r+1}\cdots j_n}$$

Then the scalars $Aj_1\cdots j_r{}^{j_{r+1}\cdots j_n} \mid 1 \leqslant j_i \leqslant n\}$ completely determine the multilinear function T

Theorem 9.4.5 Given $T \in T^r_s(V)$ a (r,s) tensor. Then we can define a collection of scalars $Aj_1\cdots j_n{}^{r+s}$ by

$$Aj_1\cdots j_r{}^{j_{r+1}\cdots j_n} = T(\mathbf{e}_{j_1},\dots,\mathbf{e}_{j_r},\mathbf{e}^{j_{r+1}}\dots\mathbf{e}^{j_n})$$

The tensor T can be expressed as:

$$T = \sum_{j_1=1}^{n}\cdots\sum_{j_n=1}^{n} Aj_1\cdots j_r{}^{j_{r+1}\cdots j_n}\mathbf{e}^{j_1}\otimes\mathbf{e}^{j_r}\otimes\mathbf{e}_{j_{r+1}}\cdots\otimes\mathbf{e}_{j_{r+s}}$$

As consequence of the previous theorem we have the following expression for the value of a tensor:

Theorem 9.4.6 Given $T\, inT^r_s(V)$ be a (r,s) tensor. And

$$\mathbf{v}_i = \sum_{j_i=1}^{n} v_{ij_i}\mathbf{e}_{j_i}$$

for $1 < i < r$, and

$$\mathbf{v}^i = \sum_{j_i=1}^{n} v^{ij_i}\mathbf{e}^{j_i}$$

for $r+1 < i < r+s$, then

$$T(\mathbf{v}_1,\dots,\mathbf{v}^n) = \sum_{j_1=1}^{n}\cdots\sum_{j_n=1}^{n} Aj_1\cdots j_r{}^{j_{r+1}\cdots j_n} v_{1j_1}\cdots v^{nj_n}$$

Example 9.4.7 Let's take a trilinear function

$$f\colon \mathbb{R}^2\times\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}.$$

A basis for $\mathbb{R}^2$ is $\{\mathbf{e}_1,\mathbf{e}_2\} = \{(1,0),(0,1)\}$. Let

$$f(\mathbf{e}_i,\mathbf{e}_j,\mathbf{e}_k) = Aijk,$$

where $i,j,k \in \{1,2\}$. In other words, the constant $Aijk$ is a function value at one of the eight possible triples of basis vectors (since there are two choices for each of the three V_i), namely:

$$\{\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_1\},\{\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_2\},\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_1\},\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_2\},\{\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_1\},\{\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_2\},\{\mathbf{e}_2,\mathbf{e}_2,\mathbf{e}_1\},\{\mathbf{e}_2,\mathbf{e}_2,\mathbf{e}_2\}.$$

Each vector $\mathbf{v}_i \in V_i = \mathbb{R}^2$ can be expressed as a linear combination of the basis vectors

$$\mathbf{v}_i = \sum_{j=1}^{2} v_{ij}\mathbf{e}_{ij} = v_{i1} \times \mathbf{e}_1 + v_{i2} \times \mathbf{e}_2 = v_{i1} \times (1,0) + v_{i2} \times (0,1).$$

The function value at an arbitrary collection of three vectors $\mathbf{v}_i \in \mathbb{R}^2$ can be expressed as

$$f(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3) = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} \mathsf{Aijk} v_{1i}v_{2j}v_{3k}.$$

Or, in expanded form as

$$f((a,b),(c,d),(e,f)) = ace \times f(\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_1) + acf \times f(\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_2) \tag{9.7}$$

$$+ ade \times f(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_1) + adf \times f(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_2) + bce \times f(\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_1) + bcf \times f(\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_2) \tag{9.8}$$

$$+ bde \times f(\mathbf{e}_2,\mathbf{e}_2,\mathbf{e}_1) + bdf \times f(\mathbf{e}_2,\mathbf{e}_2,\mathbf{e}_2). \tag{9.9}$$

9.4.2 Contraction

The simplest case of contraction is the pairing of V with its dual vector space V^*.

$$C : V^* \otimes V \to \mathbb{R} \tag{9.10}$$

$$C(f \otimes \mathbf{v}) = f(\mathbf{v}) \tag{9.11}$$

where f is in V^* and v is in V.

The above operation can be generalized to a tensor of type (r,s) (with $r > 1, s > 1$)

$$C_{ks} : T^r_s(V) \to T^{r-1}_{s-1}(V) \tag{9.12}$$

$$C_{ks}(v_1,\ldots,v_k,\ldots v_r, \tag{9.13}$$

9.5 Change of Coordinates

9.5.1 Vectors and Covectors

Suppose that V is a vector space and $E = \{v_1,\cdots,v_n\}$ and $F = \{w_1,\cdots,w_n\}$ are two ordered basis for V. E and F give rise to the dual basis $E^* = \{v^1,\cdots,v^n\}$ and $F^* = \{w^1,\cdots,w^n\}$ for V^* respectively.

If $[T]_F^E = [\lambda_i^j]$ is the matrix representation of coordinate transformation from E to F, i.e.

$$\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} \lambda_1^1 & \lambda_1^2 & \dots & \lambda_1^n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_n^1 & \lambda_n^2 & \dots & \lambda_n^n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

What is the matrix of coordinate transformation from E^* to F^*?

We can write $w^j \in F^*$ as a linear combination of basis elements in E^*:

$$w^j = \mu_1^j v^1 + \dots + \mu_n^j v^n$$

We get a matrix representation $[S]_{F^*}^{E^*} = [\mu_i^j]$ as the following:

$$\begin{bmatrix} w^1 & \dots & w^n \end{bmatrix} = \begin{bmatrix} v^1 & \dots & v^n \end{bmatrix} \begin{bmatrix} \mu_1^1 & \mu_1^2 & \dots & \mu_1^n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n^1 & \mu_n^2 & \dots & \mu_n^n \end{bmatrix}$$

We know that $w_i = \lambda_i^1 v_1 + \dots + \lambda_i^n v_n$. Evaluating this functional at $w_i \in V$ we get:

$$w^j(w_i) = \mu_1^j v^1(w_i) + \dots + \mu_n^j v^n(w_i) = \delta_i^j$$
$$w^j(w_i) = \mu_1^j v^1(\lambda_i^1 v_1 + \dots + \lambda_i^n v_n) + \dots + \mu_n^j v^n(\lambda_i^1 v_1 + \dots + \lambda_i^n v_n) = \delta_i^j$$
$$w^j(w_i) = \mu_1^j \lambda_i^1 + \dots + \mu_n^j \lambda_i^n = \sum_{k=1}^{n} \mu_k^j \lambda_i^k = \delta_i^j$$

But $\sum_{k=1}^{n} \mu_k^j \lambda_i^k$ is the (i,j) entry of the matrix product TS. Therefore $TS = I_n$ and $S = T^{-1}$.

If we want to write down the transformation from E^* to F^* as column vectors instead of row vector and name the new matrix that represents this transformation as U, we observe that $U = S^t$ and therefore $U = (T^{-1})^t$.

Therefore if T represents the transformation from E to F by the equation $\mathbf{w} = T\mathbf{v}$, then $\mathbf{w}^* = U\mathbf{v}^*$.

9.5.2 Bilinear Forms

Let $e_1, \dots, \mathbf{e}_n$ and $\tilde{e_1}, \dots, \tilde{\mathbf{e}_n}$ be two basis in a linear vector space V. Let's denote by S the transition matrix for passing from the first basis to the second one. Denote $T = S^{-1}$. From 9.3 we easily derive the formula relating the components of a bilinear form $f(\mathbf{v}, \mathbf{w})$ these two basis. For this purpose it is sufficient to substitute the expression for a change of basis into the formula 9.3 and use the bilinearity of the form $f(\mathbf{v},\mathbf{w})$:

$$f_{ij} = f(e_i, e_j) = \sum_{k=1}^{n}\sum_{q=1}^{n} T_i^k T_j^q f(\tilde{e_k}, \tilde{\mathbf{e}_q}) = \sum_{k=1}^{n}\sum_{q=1}^{n} T_i^k T_j^q \tilde{f_{kq}}.$$

The reverse formula expressing $\tilde{f}_{kq}$ through f_{ij} is derived similarly:

$$f_{ij} = \sum_{k=1}^{n}\sum_{q=1}^{n} T_i^k T_j^q \tilde{f}_{kq}, \tilde{f}_{kq} = \sum_{i=1}^{n}\sum_{j=1}^{n} S_k^i S_q^j f_{ij}. \tag{9.14}$$

In matrix form these relationships are written as follows:

$$F = T^T \tilde{F} T, \tilde{F} = S^T F S. \tag{9.15}$$

Here S^T and T^T are two matrices obtained from S and T by transposition.

9.6 Symmetry properties of tensors

Symmetry properties involve the behavior of a tensor under the interchange of two or more arguments. Of course to even consider the value of a tensor after the permutation of some of its arguments, the arguments must be of the same type, i.e., covectors have to go in covector arguments and vectors in vectors arguments and no other combinations are allowed.

The simplest case to consider are tensors with only 2 arguments of the same type. For vector arguments we have $(0,2)$-tensors. For such a tensor T introduce the following terminology:

$T(Y,X) = T(X,Y)$, T is **symmetric** in X and Y,

$T(Y,X) = -T(X,Y)$, T is **antisymmetric** or " **alternating**" in X and Y.

Letting $(X,Y) = (\mathbf{e}_i, \mathbf{e}_j)$ and using the definition of components, we get a corresponding condition on the components

$T_{ji} = T_{ij}$, T is symmetric in the index pair (i,j),

$T_{ji} = -T_{ij}$, T is antisymmetric (alternating) in the index pair (i,j).

For an antisymmetric tensor, the last condition immediately implies that no index can be repeated without the corresponding component being zero

$$T_{ji} = -T_{ij} \to T_{ii} = 0.$$

Any $(0,2)$-tensor can be decomposed into symmetric and antisymmetric parts by defining

$$[\mathrm{SYM}(T)](X,Y) = \frac{1}{2}[T(X,Y)+T(Y,X)]\,, \qquad \text{(“the symmetric part of } T\text{"),}$$

$$[\mathrm{ALT}(T)](X,Y) = \frac{1}{2}[T(X,Y)-T(Y,X)]\,, \qquad \text{(“the antisymmetric part of } T\text{"),}$$

$$T = \mathrm{SYM}(T)+\mathrm{ALT}(T)\,.$$

The last equality holds since evaluating it on the pair (X,Y) immediately leads to an identity. [Check.]

Again letting $(X,Y) = (\mathbf{e}_i,\mathbf{e}_j)$ leads to corresponding component formulas

$$[\mathrm{SYM}(T)]_{ij} = \frac{1}{2}(T_{ij}+T_{ji}) \equiv T_{(ij)}\,, \qquad (n(n+1)/2 \text{ independent components}),$$

$$[\mathrm{ALT}(T)]_{ij} = \frac{1}{2}(T_{ij}-T_{ji}) \equiv T_{[ij]}\,, \qquad (n(n-1)/2 \text{ independent components}),$$

$$T_{ij} = T_{(ij)}+T_{[ij]}\,, \qquad (n^2 = n(n+1)/2+n(n-1)/2 \text{ independent components}).$$

Round brackets around a pair of indices denote the symmetrization operation, while square brackets denote antisymmetrization. This is a very convenient shorthand. All of this can be repeated for $\binom{2}{0}$-tensors and just reflects what we already know about the symmetric and antisymmetric parts of matrices.

9.7 Forms

9.7.1 Motivation

Oriented area and Volume

We define the **oriented area** function $A(\mathbf{a},\mathbf{b})$ by

$$A(\mathbf{a},\mathbf{b}) = \pm|\mathbf{a}|\cdot|\mathbf{b}|\cdot\sin\alpha,$$

where the sign is chosen positive when the angle α is measured from the vector **a** to the vector **b** in the counterclockwise direction, and negative otherwise.

Statement:

The oriented area $A(\mathbf{a},\mathbf{b})$ of a parallelogram spanned by the vectors **a** and **b** in the two-dimensional Euclidean space is an antisymmetric and bilinear function of the vectors **a** and **b**:

$$A(\mathbf{a},\mathbf{b}) = -A(\mathbf{b},\mathbf{a}),$$

$$A(\lambda\mathbf{a},\mathbf{b}) = \lambda A(\mathbf{a},\mathbf{b}),$$

$$A(\mathbf{a},\mathbf{b}+\mathbf{c}) = A(\mathbf{a},\mathbf{b})+A(\mathbf{a},\mathbf{c}). \qquad \text{(the sum law)}$$

The ordinary (unoriented) area is then obtained as the absolute value of the oriented area, $Ar(\mathbf{a},\mathbf{b}) = \big|A(\mathbf{a},\mathbf{b})\big|$. It turns out that the oriented area, due to its strict linearity properties, is a much more convenient and powerful construction than the unoriented area.

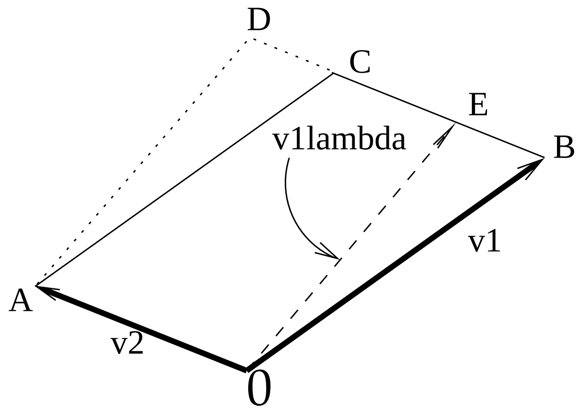


Figure 9.1 The area of the parallelogram $0ACB$ spanned by $\mathbf{a}$ and $\mathbf{b}$ is equal to the area of the parallelogram $0ADE$ spanned by $\mathbf{a}$ and $\mathbf{b} + \alpha\mathbf{a}$ due to the equality of areas ACD and $0BE$.

Theorem 9.7.1 Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, be linearly independent vectors in $\mathbb{R}^3$. The signed volume of the parallelepiped spanned by them is $(\mathbf{a} \times \mathbf{b}) \bullet \mathbf{c}$.

Statement:

The oriented volume $V(\mathbf{a}, \mathbf{b}, \mathbf{c})$ of a parallelepiped spanned by the vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ in the three-dimensional Euclidean space is an antisymmetric and trilinear function of the vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$:

$$
\begin{aligned}
V(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= -V(\mathbf{b}, \mathbf{a}, \mathbf{c}), \\
V(\lambda\mathbf{a}, \mathbf{b}, \mathbf{c}) &= \lambda V(\mathbf{a}, \mathbf{b}, \mathbf{c}), \\
V(\mathbf{a}, \mathbf{b} + \mathbf{d}, \mathbf{c}) &= V(\mathbf{a}, \mathbf{b}) + V(\mathbf{a}, \mathbf{d}, \mathbf{c}). \qquad \text{(the sum law)}
\end{aligned}
$$

9.7.2 Exterior product

In three dimensions, an oriented area is represented by the cross product $\mathbf{a} \times \mathbf{b}$, which is indeed an antisymmetric and bilinear product. So we expect that the oriented area in higher dimensions can be represented by some kind of new antisymmetric product of $\mathbf{a}$ and $\mathbf{b}$; let us denote this product (to be defined below) by $\mathbf{a} \wedge \mathbf{b}$, pronounced "a wedge b." The value of $\mathbf{a} \wedge \mathbf{b}$ will be a vector in a *new* vector space. We will also construct this new space explicitly.

Definition of exterior product

We will construct an antisymmetric product using the tensor product space.

Definition 9.7.1 Given a vector space V, we define a new vector space $V \wedge V$ called the **exterior product** (or antisymmetric tensor product, or alternating product, or **wedge product**) of two copies of V. The space $V \wedge V$ is the subspace in $V \otimes V$ consisting of all **antisymmetric** tensors, i.e. tensors of the form

$$\mathbf{v}_1 \otimes \mathbf{v}_2 - \mathbf{v}_2 \otimes \mathbf{v}_1, \quad \mathbf{v}_{1,2} \in V,$$

and all linear combinations of such tensors. The exterior product of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is the expression shown above; it is obviously an antisymmetric and bilinear function of $\mathbf{v}_1$ and $\mathbf{v}_2$.

For example, here is one particular element from $V \wedge V$, which we write in two different ways using the properties of the tensor product:

$$\begin{aligned}(\mathbf{u}+\mathbf{v}) \otimes (\mathbf{v}+\mathbf{w}) - (\mathbf{v}+\mathbf{w}) \otimes (\mathbf{u}+\mathbf{v}) &= \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} \\ &+ \mathbf{u} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v} \in V \wedge V. \end{aligned} \tag{9.16}$$

Remark: A tensor $\mathbf{v}_1 \otimes \mathbf{v}_2 \in V \otimes V$ is not equal to the tensor $\mathbf{v}_2 \otimes \mathbf{v}_1$ if $\mathbf{v}_1 \neq \mathbf{v}_2$.

It is quite cumbersome to perform calculations in the tensor product notation as we did in Eq. (9.16). So let us write the exterior product as $\mathbf{u} \wedge \mathbf{v}$ instead of $\mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u}$. It is then straightforward to see that the “wedge” symbol $\wedge$ indeed works like an anti-commutative multiplication, as we intended. The rules of computation are summarized in the following statement.

Statement 1:

One may save time and write $\mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} \equiv \mathbf{u} \wedge \mathbf{v} \in V \wedge V$, and the result of any calculation will be correct, as long as one follows the rules:

$$\mathbf{u} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{u}, \tag{9.17}$$

$$(\lambda \mathbf{u}) \wedge \mathbf{v} = \lambda (\mathbf{u} \wedge \mathbf{v}), \tag{9.18}$$

$$(\mathbf{u}+\mathbf{v}) \wedge \mathbf{x} = \mathbf{u} \wedge \mathbf{x} + \mathbf{v} \wedge \mathbf{x}. \tag{9.19}$$

It follows also that $\mathbf{u} \wedge (\lambda \mathbf{v}) = \lambda (\mathbf{u} \wedge \mathbf{v})$ and that $\mathbf{v} \wedge \mathbf{v} = 0$. (These identities hold for any vectors $\mathbf{u}, \mathbf{v} \in V$ and any scalars $\lambda \in \mathbb{K}$.)

Proof: These properties are direct consequences of the properties of the tensor product when applied to antisymmetric tensors. For example, the calculation (9.16) now requires a simple expansion of brackets,

$$(\mathbf{u}+\mathbf{v}) \wedge (\mathbf{v}+\mathbf{w}) = \mathbf{u} \wedge \mathbf{v} + \mathbf{u} \wedge \mathbf{w} + \mathbf{v} \wedge \mathbf{w}.$$

Here we removed the term $\mathbf{v} \wedge \mathbf{v}$ which vanishes due to the antisymmetry of $\wedge$. Details left as exercise. ■

Elements of the space $V \wedge V$, such as $\mathbf{a} \wedge \mathbf{b} + \mathbf{c} \wedge \mathbf{d}$, are sometimes called **bivectors**.[1] We will also want to define the exterior product of more than two vectors. To define the exterior product of *three* vectors, we consider the subspace of $V \otimes V \otimes V$ that consists of antisymmetric tensors of the form

$$\begin{aligned}\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} - \mathbf{b} \otimes \mathbf{a} \otimes \mathbf{c} + \mathbf{c} \otimes \mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{b} \otimes \mathbf{a} \\ + \mathbf{b} \otimes \mathbf{c} \otimes \mathbf{a} - \mathbf{a} \otimes \mathbf{c} \otimes \mathbf{b}\end{aligned} \tag{9.20}$$

and linear combinations of such tensors. These tensors are called **totally antisymmetric** because they can be viewed as (tensor-valued) functions of the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ that change sign under exchange of any two vectors. The expression in Eq. (9.20) will be denoted for brevity by $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$, similarly to the exterior product of two vectors, $\mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}$, which is denoted for brevity by $\mathbf{a} \wedge \mathbf{b}$. Here is a general definition.

Definition 2:

The **exterior product** of k copies of V (also called the k**-th exterior power** of V) is denoted by $\wedge^k V$ and is defined as the subspace of totally antisymmetric tensors within $V \otimes ... \otimes V$. In the concise notation, this is the space spanned by expressions of the form

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge ... \wedge \mathbf{v}_k, \quad \mathbf{v}_j \in V,$$

assuming that the properties of the wedge product (linearity and antisymmetry) hold as given by Statement 1. For instance,

$$\mathbf{u} \wedge \mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k = (-1)^k \mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k \wedge \mathbf{u} \tag{9.21}$$

("pulling a vector through k other vectors changes sign k times"). ■

The previously defined space of bivectors is in this notation $V \wedge V \equiv \wedge^2 V$. A natural extension of this notation is $\wedge^0 V = \mathbb{K}$ and $\wedge^1 V = V$. I will also use the following "wedge product" notation,

$$\bigwedge_{k=1}^{n} \mathbf{v}_k \equiv \mathbf{v}_1 \wedge \mathbf{v}_2 \wedge ... \wedge \mathbf{v}_n.$$

Tensors from the space $\wedge^n V$ are also called n**-vectors** or **antisymmetric tensors** of rank n.

Question:

How to compute expressions containing multiple products such as $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$?

[1] It is important to note that a bivector is not necessarily expressible as a single-term product of two vectors; see the Exercise at the end of Sec. **??**.

Answer: Apply the rules shown in Statement 1. For example, one can permute adjacent vectors and change sign,

$$\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}=-\mathbf{b}\wedge\mathbf{a}\wedge\mathbf{c}=\mathbf{b}\wedge\mathbf{c}\wedge\mathbf{a},$$

one can expand brackets,

$$\mathbf{a}\wedge(\mathbf{x}+4\mathbf{y})\wedge\mathbf{b}=\mathbf{a}\wedge\mathbf{x}\wedge\mathbf{b}+4\mathbf{a}\wedge\mathbf{y}\wedge\mathbf{b},$$

and so on. If the vectors $\mathbf{a},\mathbf{b},\mathbf{c}$ are given as linear combinations of some basis vectors $\{\mathbf{e}_j\}$, we can thus reduce $\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}$ to a linear combination of exterior products of basis vectors, such as $\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_3$, $\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_4$, etc.

Example 1:

Suppose we work in $\mathbb{R}^3$ and have vectors $\mathbf{a}=\left(0,\frac{1}{2},-\frac{1}{2}\right)$, $\mathbf{b}=(2,-2,0)$, $\mathbf{c}=(-2,5,-3)$. Let us compute various exterior products. Calculations are easier if we introduce the basis $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ explicitly:

$$\mathbf{a}=\frac{1}{2}(\mathbf{e}_2-\mathbf{e}_3),\quad \mathbf{b}=2(\mathbf{e}_1-\mathbf{e}_2),\quad \mathbf{c}=-2\mathbf{e}_1+5\mathbf{e}_2-3\mathbf{e}_3.$$

We compute the 2-vector $\mathbf{a}\wedge\mathbf{b}$ by using the properties of the exterior product, such as $\mathbf{x}\wedge\mathbf{x}=0$ and $\mathbf{x}\wedge\mathbf{y}=-\mathbf{y}\wedge\mathbf{x}$, and simply expanding the brackets as usual in algebra:

$$\begin{aligned}\mathbf{a}\wedge\mathbf{b}&=\frac{1}{2}(\mathbf{e}_2-\mathbf{e}_3)\wedge 2(\mathbf{e}_1-\mathbf{e}_2)\\&=(\mathbf{e}_2-\mathbf{e}_3)\wedge(\mathbf{e}_1-\mathbf{e}_2)\\&=\mathbf{e}_2\wedge\mathbf{e}_1-\mathbf{e}_3\wedge\mathbf{e}_1-\mathbf{e}_2\wedge\mathbf{e}_2+\mathbf{e}_3\wedge\mathbf{e}_2\\&=-\mathbf{e}_1\wedge\mathbf{e}_2+\mathbf{e}_1\wedge\mathbf{e}_3-\mathbf{e}_2\wedge\mathbf{e}_3.\end{aligned}$$

The last expression is the result; note that now there is nothing more to compute or to simplify. The expressions such as $\mathbf{e}_1\wedge\mathbf{e}_2$ are the basic expressions out of which the space $\mathbb{R}^3\wedge\mathbb{R}^3$ is built.

Let us also compute the 3-vector $\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}$,

$$\begin{aligned}\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}&=(\mathbf{a}\wedge\mathbf{b})\wedge\mathbf{c}\\&=(-\mathbf{e}_1\wedge\mathbf{e}_2+\mathbf{e}_1\wedge\mathbf{e}_3-\mathbf{e}_2\wedge\mathbf{e}_3)\wedge(-2\mathbf{e}_1+5\mathbf{e}_2-3\mathbf{e}_3).\end{aligned}$$

When we expand the brackets here, terms such as $\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_1$ will vanish because

$$\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_1=-\mathbf{e}_2\wedge\mathbf{e}_1\wedge\mathbf{e}_1=0,$$

so only terms containing all different vectors need to be kept, and we find

$$\begin{aligned}\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}&=3\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_3+5\mathbf{e}_1\wedge\mathbf{e}_3\wedge\mathbf{e}_2+2\mathbf{e}_2\wedge\mathbf{e}_3\wedge\mathbf{e}_1\\&=(3-5+2)\,\mathbf{e}_1\wedge\mathbf{e}_2\wedge\mathbf{e}_3=0.\end{aligned}$$

We note that all the terms are proportional to the 3-vector $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$, so only the coefficient in front of $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$ was needed; then, by coincidence, that coefficient turned out to be zero. So the result is the zero 3-vector. ■

Remark: Origin of the name "exterior."

The construction of the exterior product is a modern formulation of the ideas dating back to H. Grassmann (1844). A 2-vector $\mathbf{a} \wedge \mathbf{b}$ is interpreted geometrically as the oriented area of the parallelogram spanned by the vectors $\mathbf{a}$ and $\mathbf{b}$. Similarly, a 3-vector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ represents the oriented 3-volume of a parallelepiped spanned by $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Due to the antisymmetry of the exterior product, we have $(\mathbf{a} \wedge \mathbf{b}) \wedge (\mathbf{a} \wedge \mathbf{c}) = 0$, $(\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}) \wedge (\mathbf{b} \wedge \mathbf{d}) = 0$, etc. We can interpret this geometrically by saying that the "product" of two volumes is zero if these volumes have a vector in common. This motivated Grassmann to call his antisymmetric product "exterior." In his reasoning, the product of two "extensive quantities" (such as lines, areas, or volumes) is nonzero only when each of the two quantities is geometrically "to the exterior" (outside) of the other.

Exercise 2:

Show that in a *two*-dimensional space V, any 3-vector such as $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ can be simplified to the zero 3-vector. Prove the same for n-vectors in N-dimensional spaces when $n > N$. ■

One can also consider the exterior powers of the *dual* space V^*. Tensors from $\wedge^n V^*$ are usually (for historical reasons) called n-**forms** (rather than "n-covectors").

Definition 3:

The action of a k-form $\mathbf{f}_1^* \wedge ... \wedge \mathbf{f}_k^*$ on a k-vector $\mathbf{v}_1 \wedge ... \wedge \mathbf{v}_k$ is defined by

$$\sum_{\sigma} (-1)^{|\sigma|} \mathbf{f}_1^*(\mathbf{v}_{\sigma(1)}) ... \mathbf{f}_k^*(\mathbf{v}_{\sigma(k)}),$$

where the summation is performed over all permutations σ of the ordered set $(1, ..., k)$.

Example 2:

With $k = 3$ we have

$$\begin{aligned}
&(\mathbf{p}^* \wedge \mathbf{q}^* \wedge \mathbf{r}^*)(\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}) \\
&= \mathbf{p}^*(\mathbf{a})\mathbf{q}^*(\mathbf{b})\mathbf{r}^*(\mathbf{c}) - \mathbf{p}^*(\mathbf{b})\mathbf{q}^*(\mathbf{a})\mathbf{r}^*(\mathbf{c}) \\
&+ \mathbf{p}^*(\mathbf{b})\mathbf{q}^*(\mathbf{c})\mathbf{r}^*(\mathbf{a}) - \mathbf{p}^*(\mathbf{c})\mathbf{q}^*(\mathbf{b})\mathbf{r}^*(\mathbf{a}) \\
&+ \mathbf{p}^*(\mathbf{c})\mathbf{q}^*(\mathbf{a})\mathbf{r}^*(\mathbf{b}) - \mathbf{p}^*(\mathbf{c})\mathbf{q}^*(\mathbf{b})\mathbf{r}^*(\mathbf{a}).
\end{aligned}$$

Exercise 3:

a) Show that $\mathbf{a} \wedge \mathbf{b} \wedge \omega = \omega \wedge \mathbf{a} \wedge \mathbf{b}$ where ω is any antisymmetric tensor (e.g. $\omega = \mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$).

b) Show that

$$\omega_1 \wedge \mathbf{a} \wedge \omega_2 \wedge \mathbf{b} \wedge \omega_3 = -\omega_1 \wedge \mathbf{b} \wedge \omega_2 \wedge \mathbf{a} \wedge \omega_3,$$

where ω_1, ω_2, ω_3 are arbitrary antisymmetric tensors and $\mathbf{a}, \mathbf{b}$ are vectors.

c) Due to antisymmetry, $\mathbf{a} \wedge \mathbf{a} = 0$ for any vector $\mathbf{a} \in V$. Is it also true that $\omega \wedge \omega = 0$ for any bivector $\omega \in \wedge^2 V$?

9.7.3 Forms

We will now consider integration in several variables. In order to smooth our discussion, we need to consider the concept of differential forms.

Definition 9.7.2 Consider n variables

$$x_1, x_2, \ldots, x_n$$

in n-dimensional space (used as the names of the axes), and let

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \in \mathbb{R}^n, \quad 1 \leqslant j \leqslant k,$$

be $k \leqslant n$ vectors in $\mathbb{R}^n$. Moreover, let $\{j_1, j_2, \ldots, j_k\} \subseteq \{1, 2, \ldots, n\}$ be a collection of k sub-indices. An **elementary k-differential form ($k > 1$) acting on the vectors $\mathbf{a}_j$**, $1 \leqslant j \leqslant k$ is defined and denoted by

$$dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_k}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k) = \det \begin{bmatrix} a_{j_1 1} & a_{j_1 2} & \cdots & a_{j_1 k} \\ a_{j_2 1} & a_{j_2 2} & \cdots & a_{j_2 k} \\ \vdots & \vdots & \cdots & \vdots \\ a_{j_k 1} & a_{j_k 2} & \cdots & a_{j_k k} \end{bmatrix}.$$

In other words, $dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_k}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k)$ is the $x_{j_1} x_{j_2} \ldots x_{j_k}$ component of the signed k-volume of a k-parallelotope in $\mathbb{R}^n$ spanned by $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$.

By virtue of being a determinant, the wedge product $\wedge$ of differential forms has the following properties

❶ **anti-commutativity:** $da \wedge db = -db \wedge da$.

❷ **linearity:** $d(a + b) = da + db$.

❸ **scalar homogeneity:** if $\lambda \in \mathbb{R}$, then $d\lambda a = \lambda da$.

❹ **associativity:** $(da \wedge db) \wedge dc = da \wedge (db \wedge dc)$.[2]

Anti-commutativity yields
$$da \wedge da = 0.$$

Example 9.7.2 Consider
$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \in \mathbb{R}^3.$$
Then
$$dx(\mathbf{a}) = \det(1) = 1,$$
$$dy(\mathbf{a}) = \det(0) = 0,$$
$$dz(\mathbf{a}) = \det(-1) = -1,$$
are the (signed) 1-volumes (that is, the length) of the projections of **a** onto the coordinate axes.

Example 9.7.3 In $\mathbb{R}^3$ we have $dx \wedge dy \wedge dx = 0$, since we have a repeated variable.

Example 9.7.4 In $\mathbb{R}^3$ we have
$$dx \wedge dz + 5dz \wedge dx + 4dx \wedge dy - dy \wedge dx + 12dx \wedge dx = -4dx \wedge dz + 5dx \wedge dy.$$

In order to avoid redundancy we will make the convention that if a sum of two or more terms have the same differential form up to permutation of the variables, we will simplify the summands and express the other differential forms in terms of the one differential form whose indices appear in increasing order.

9.7.4 Hodge star operator

Given an orthonormal basis $(\mathbf{e}_1, \cdots, e_n)$ we define
$$\star : \bigwedge^{k} V \to \bigwedge^{n-k} V$$
$$\star(e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}) = e_{i_{k+1}} \wedge e_{i_{k+2}} \wedge \cdots \wedge e_{i_n},$$
where
$$(i_1, i_2, \cdots, i_n)$$
is an even permutation of $\{1, 2, ..., n\}$. [3]

[2] Notice that associativity does not hold for the wedge product of *vectors*.

[3] The parity of a permutation σ of $\{1, 2, \ldots, n\}$ can be defined as the parity of the number of inversions, i.e., of pairs of elements x, y of $\{1, 2, ..., n\}$ such that $x < y$ and $\sigma(x) > \sigma(y)$.

■ **Exemplo 9.2** In $\mathbb{R}^n$:

$$\star(\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \cdots \wedge \mathbf{e}_k) = e_{k+1} \wedge e_{k+2} \wedge \cdots \wedge e_n.$$

■

Chapter 10

Tensors in Coordinates

"The introduction of numbers as coordinates is an act of violence."
Hermann Weyl.

10.1 Index notation for tensors

So far we have used a coordinate-free formalism to define and describe tensors. However, in many calculations a basis in V is fixed, and one needs to compute the components of tensors in that basis. In this cases the **index notation** makes such calculations easier.

Suppose a basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ in V is fixed; then the dual basis $\left\{\mathbf{e}^k\right\}$ is also fixed. Any vector $\mathbf{v} \in V$ is decomposed as $\mathbf{v} = \sum_k v^k \mathbf{e}_k$ and any covector as $\mathbf{f}^* = \sum_k f_k \mathbf{e}^k$.

Any tensor from $V \otimes V$ is decomposed as

$$A = \sum_{j,k} A^{jk} \mathbf{e}_j \otimes \mathbf{e}_k \in V \otimes V$$

and so on. The action of a covector on a vector is $\mathbf{f}^* (\mathbf{v}) = \sum_k f_k v_k$, and the action of an operator on a vector is $\sum_{j,k} \mathsf{A}_{jk} v_k \mathbf{e}_k$. However, it is cumbersome to keep writing these sums. In the index notation, one writes *only* the components v_k or A_{jk} of vectors and tensors.

Definition 10.1.1 Given $T \in T^r_s(V)$:

$$T = \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} T^{j_{r+1} \cdots j_n}_{j_1 \cdots j_r} \mathbf{e}^{j_1} \otimes \mathbf{e}^{j_r} \otimes \mathbf{e}_{j_{r+1}} \cdots \otimes \mathbf{e}_{j_{r+s}}$$

The index notation of this tensor is

$$T^{j_{r+1} \cdots j_n}_{j_1 \cdots j_r}$$

10.1.1 Definition of index notation

The rules for expressing tensors in the index notations are as follows:

- Basis vectors $\mathbf{e}_k$ and basis tensors (e.g. $\mathbf{e}_k \otimes \mathbf{e}^*_l$) are never written explicitly. (It is assumed that the basis is fixed and known.)
- Instead of a vector $\mathbf{v} \in V$, one writes its array of components v^k with the *superscript* index. Covectors $\mathbf{f}^* \in V^*$ are written f_k with the *subscript* index. The index k runs over integers from 1 to N. Components of vectors and tensors may be thought of as numbers.
- Tensors are written as multidimensional arrays of components with superscript or subscript indices as necessary, for example $A_{jk} \in V^* \otimes V^*$ or $B^{lm}_k \in V \otimes V \otimes V^*$. Thus e.g. the Kronecker delta symbol is written as δ^j_k when it represents the identity operator $\hat{1}_V$.
- Tensors with subscript indices, like A_{ij}, are called covariant, while tensors with superscript indices, like A^k, are called contravariant. Tensors with both types of indices, like A^{lmn}_{lk}, are called mixed type.
- Subscript indices, rather than subscripted tensors, are also dubbed “covariant” and superscript indices are dubbed “contravariant”.
- For tensor invariance, a pair of dummy indices should in general be complementary in their variance type, i.e. one covariant and the other contravariant.
- As indicated earlier, tensor order is equal to the number of its indices while tensor rank is equal to the number of its free indices; hence vectors (terms, expressions and equalities) are represented by a single free index and rank-2 tensors are represented by two free indices. The dimension of a tensor is determined by the range taken by its indices.
- The choice of indices must be consistent; each index corresponds to a particular copy of V or V^*. Thus it is wrong to write $v_j = u_k$ or $v_i + u^i = 0$. Correct equations are $v_j = u_j$ and $v^i + u^i = 0$. This disallows meaningless expressions such as $\mathbf{v}^* + \mathbf{u}$ (one cannot add vectors from different spaces).

- Sums over indices such as $\sum_{k=1}^{n} a_k b_k$ are not written explicitly, the $\sum$ symbol is omitted, and the **Einstein summation convention** is used instead: Summation over all values of an index is *always implied* when that index letter appears once as a subscript and once as a superscript. In this case the letter is called a **dummy** (or **mute**) **index**. Thus one writes $f_k v^k$ instead of $\sum_k f_k v_k$ and $A^j_k v^k$ instead of $\sum_k A_{jk} v_k$.

- Summation is allowed *only* over one subscript and one superscript but never over two subscripts or two superscripts and never over three or more coincident indices. This corresponds to requiring that we are only allowed to compute the canonical pairing of V and V^* but no other pairing. The expression $v^k v^k$ is not allowed because there is no canonical pairing of V and V, so, for instance, the sum $\sum_{k=1}^{n} v^k v^k$ depends on the choice of the basis. For the same reason (dependence on the basis), expressions such as $u^i v^i w^i$ or $A_{ii} B^{ii}$ are not allowed. Correct expressions are $u_i v^i w_k$ and $A_{ik} B^{ik}$.

- One needs to pay close attention to the choice and the position of the letters such as $j, k, l,\ldots$ used as indices. Indices that are not repeated are **free** indices. The rank of a tensor expression is equal to the number of free subscript and superscript indices. Thus $A^j_k v^k$ is a rank 1 tensor (i.e. a vector) because the expression $A^j_k v^k$ has a single free index, j, and a summation over k is implied.

- The tensor product symbol $\otimes$ is never written. For example, if $\mathbf{v} \otimes \mathbf{f}^* = \sum_{jk} v_j f^*_k \mathbf{e}_j \otimes \mathbf{e}^k$, one writes $v^k f_j$ to represent the tensor $\mathbf{v} \otimes \mathbf{f}^*$. The index letters in the expression $v^k f_j$ are intentionally chosen to be *different* (in this case, k and j) so that no summation would be implied. In other words, a tensor product is written simply as a product of components, and the index letters are chosen appropriately. Then one can interpret $v^k f_j$ as simply the product of *numbers*. In particular, it makes no difference whether one writes $f_j v^k$ or $v^k f_j$. The *position of the indices* (rather than the ordering of vectors) shows in every case how the tensor product is formed. Note that it is not possible to distinguish $V \otimes V^*$ from $V^* \otimes V$ in the index notation.

Example 10.1.1 It follows from the definition of δ^i_j that $\delta^i_j v^j = v^i$. This is the index representation of the identity transformation $\hat{1}\mathbf{v} = \mathbf{v}$.

Example 10.1.2 Suppose $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are vectors from V whose components are w^i, x^i, y^i, z^i. What are the components of the tensor $\mathbf{w} \otimes \mathbf{x} + 2\mathbf{y} \otimes \mathbf{z} \in V \otimes V$?

Solution: ► $w^i x^k + 2y^i z^k$. (We need to choose another letter for the second free index, k, which corresponds to the second copy of V in $V \otimes V$.) ◄

Example 10.1.3 The operator $\hat{A} \equiv \hat{1}_V + \lambda \mathbf{v} \otimes \mathbf{u}^* \in V \otimes V^*$ acts on a vector $\mathbf{x} \in V$. Calculate the resulting vector $\mathbf{y} \equiv \hat{A}\mathbf{x}$.

In the index-free notation, the calculation is

$$\mathbf{y} = \hat{A}\mathbf{x} = \left(\hat{1}_V + \lambda \mathbf{v} \otimes \mathbf{u}^*\right)\mathbf{x} = \mathbf{x} + \lambda \mathbf{u}^*(\mathbf{x})\,\mathbf{v}.$$

In the index notation, the calculation looks like this:

$$y^k = \left(\delta_j^k + \lambda v^k u_j\right) x^j = x^k + \lambda v^k u_j x^j.$$

In this formula, j is a dummy index and k is a free index. We could have also written $\lambda x^j v^k u_j$ instead of $\lambda v^k u_j x^j$ since the ordering of components makes no difference in the index notation.

Example 10.1.4 In a physics book you find the following formula,

$$H^\alpha_{\mu\nu} = \frac{1}{2}\left(h_{\beta\mu\nu} + h_{\beta\nu\mu} - h_{\mu\nu\beta}\right) g^{\alpha\beta}.$$

To what spaces do the tensors H, g, h belong (assuming these quantities represent tensors)? Rewrite this formula in the coordinate-free notation.

Solution: ► $H \in V \otimes V^* \otimes V^*$, $h \in V^* \otimes V^* \otimes V^*$, $g \in V \otimes V$. Assuming the simplest case,

$$h = \mathbf{h}_1^* \otimes \mathbf{h}_2^* \otimes \mathbf{h}_3^*,\; g = \mathbf{g}_1 \otimes \mathbf{g}_2,$$

the coordinate-free formula is

$$H = \frac{1}{2}\mathbf{g}_1 \otimes \left(\mathbf{h}_1^*(\mathbf{g}_2)\,\mathbf{h}_2^* \otimes \mathbf{h}_3^* + \mathbf{h}_1^*(\mathbf{g}_2)\,\mathbf{h}_3^* \otimes \mathbf{h}_2^* - \mathbf{h}_3^*(\mathbf{g}_2)\,\mathbf{h}_1^* \otimes \mathbf{h}_2^*\right).$$

◄

10.1.2 Advantages and disadvantages of index notation

Index notation is conceptually easier than the index-free notation because one can imagine manipulating "merely" some tables of numbers, rather than "abstract vectors." In other words, we are working with *less abstract objects*. The price is that we obscure the geometric interpretation of what we are doing, and proofs of general theorems become more difficult to understand.

The main advantage of the index notation is that it makes computations with complicated tensors quicker.

Some *disadvantages* of the index notation are:

- If the basis is changed, all components need to be recomputed. In textbooks that use the index notation, quite some time is spent studying the transformation laws of tensor components under a change of basis. If different basis are used simultaneously, confusion may result.

- The geometrical meaning of many calculations appears hidden behind a mass of indices. It is sometimes unclear whether a long expression with indices can be simplified and how to proceed with calculations.

Despite these disadvantages, the index notation enables one to perform practical calculations with high-rank tensor spaces, such as those required in field theory and in general relativity. For this reason, and also for historical reasons (Einstein used the index notation when developing the theory of relativity), most physics textbooks use the index notation. In some cases, calculations can be performed equally quickly using index and index-free notations. In other cases, especially when deriving general properties of tensors, the index-free notation is superior.

10.2 Tensor Revisited: Change of Coordinate

Vectors, covectors, linear operators, and bilinear forms are examples of tensors. They are multilinear maps that are represented numerically when some basis in the space is chosen.

This numeric representation is specific to each of them: vectors and covectors are represented by one-dimensional arrays, linear operators and quadratic forms are represented by two-dimensional arrays. Apart from the number of indices, their position does matter. The coordinates of a vector are numerated by one upper index, which is called the contravariant index. The coordinates of a covector are numerated by one lower index, which is called the covariant index. In a matrix of bilinear form we use two lower indices; therefore bilinear forms are called **twice-covariant tensors**. Linear operators are tensors of **mixed type**; their components are numerated by one upper and one lower index. The number of indices and their positions determine the transformation rules, i˙ ethe way the components of each particular tensor behave under a change of basis. In the general case, any tensor is represented by a multidimensional array with a definite number of upper indices and a definite number of lower indices. Let's denote these numbers by r and s. Then we have **a tensor of the type** (r, s), or sometimes the term **valency** is used. A tensor of type (r, s), or of valency (r, s) is called **an r-times contravariant** and **an s-times covariant** tensor. This is terminology; now let's proceed to the exact definition. It is based on the following general transformation formulas:

$$X^{i_1\dots i_r}_{j_1\dots j_s} = \sum^{n}_{\substack{h_1,\dots,h_r\\k_1,\dots,k_s}} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{X}^{h_1\dots h_r}_{k_1\dots k_s}, \tag{10.1}$$

$$\tilde{X}^{i_1\dots i_r}_{j_1\dots j_s} = \sum^{n}_{\substack{h_1,\dots,h_r\\k_1,\dots,k_s}} \dots \sum^{n} T^{i_1}_{h_1} \dots T^{i_r}_{h_r} S^{k_1}_{j_1} \dots S^{k_s}_{j_s} X^{h_1\dots h_r}_{k_1\dots k_s}. \tag{10.2}$$

Definition 10.2.1 — Tensor Definition in Coordinate. A $(r+s)$-dimensional array $X^{i_1\dots i_r}_{j_1\dots j_s}$ of real numbers and such that the components of this array obey the transformation rules

$$X^{i_1\dots i_r}_{j_1\dots j_s} = \sum_{\substack{h_1,\dots,h_r\\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{X}^{h_1\dots h_r}_{k_1\dots k_s}, \tag{10.3}$$

$$\tilde{X}^{i_1\dots i_r}_{j_1\dots j_s} = \sum_{\substack{h_1,\dots,h_r\\ k_1,\dots,k_s}}^{n} \dots \sum^{n} T^{i_1}_{h_1} \dots T^{i_r}_{h_r} S^{k_1}_{j_1} \dots S^{k_s}_{j_s} X^{h_1\dots h_r}_{k_1\dots k_s}. \tag{10.4}$$

under a change of basis is called **tensor** of type (r,s), or of valency (r,s).

Formula 10.4 is derived from 10.3, so it is sufficient to remember only one of them. Let it be the formula 10.3. Though huge, formula 10.3 is easy to remember.

Indices $i_1,\dots,i_r$ and $j_1,\dots,j_s$ are free indices. In right hand side of the equality 10.3 they are distributed in S-s and T-s, each having only one entry and each keeping its position, i˙ eupper indices $i_1,\dots,i_r$ remain upper and lower indices $j_1,\dots,j_s$ remain lower in right hand side of the equality 10.3.

Other indices $h_1,\dots,h_r$ and $k_1,\dots,k_s$ are summation indices; they enter the right hand side of 10.3 pairwise: once as an upper index and once as a lower index, once in S-s or T-s and once in components of array $\tilde{X}^{h_1\dots h_r}_{k_1\dots k_s}$.

When expressing $X^{i_1\dots i_r}_{j_1\dots j_s}$ through $\tilde{X}^{h_1\dots h_r}_{k_1\dots k_s}$ each upper index is served by direct transition matrix S and produces one summation in 10.3:

$$X^{\dots\, i_\alpha\, \dots}_{\dots\dots\dots} = \sum \dots \sum_{h_\alpha=1}^{n} \dots \sum \dots S^{i_\alpha}_{h_\alpha} \dots \tilde{X}^{\dots\, h_\alpha\, \dots}_{\dots\dots\dots}. \tag{10.5}$$

In a similar way, each lower index is served by inverse transition matrix T and also produces one summation in formula 10.3:

$$X^{\dots\dots\dots}_{\dots\, j_\alpha\, \dots} = \sum \dots \sum_{k_\alpha=1}^{n} \dots \sum \dots T^{k_\alpha}{}_{j_\alpha} \dots \tilde{X}^{\dots\dots\dots}_{\dots\, k_\alpha\, \dots}. \tag{10.6}$$

Formulas 10.5 and 10.6 are the same as 10.3 and used to highlight how 10.3 is written. So tensors are defined. Further we shall consider more examples showing that many well-known objects undergo the definition 12.1.

Example 10.2.1 Verify that formulas for change of basis of vectors, covectors, linear transformation and bilinear forms are special cases of formula 10.3. What are the valencies of vectors, covectors, linear operators, and bilinear forms when they are considered as tensors.

Example 10.2.2 The δ^j_i is a tensor.

Solution: ►

$$\delta'^j_i = A^j_k (A^{-1})^l_i \delta^k_l = A^j_k (A^{-1})^k_i = \delta^j_i$$

◀

Example 10.2.3 The ϵ_{ijk} is a pseudo-tensor.

Example 10.2.4 Let a_{ij} be the matrix of some bilinear form a. Let's denote by b^{ij} components of inverse matrix for a_{ij}. Prove that matrix b^{ij} under a change of basis transforms like matrix of twice-contravariant tensor. Hence it determines tensor b of valency $(2,0)$. Tensor b is called **a dual bilinear form** for a.

10.2.1 Rank

The order of a tensor is identified by the number of its indices (e.g. A^i_{jk} is a tensor of order 3) which normally identifies the tensor rank as well. However, when contraction (see S 10.3.4) takes place once or more, the order of the tensor is not affected but its rank is reduced by two for each contraction operation.[1]

- "Zero tensor" is a tensor whose all components are zero.
- "Unit tensor" or "unity tensor", which is usually defined for rank-2 tensors, is a tensor whose all elements are zero except the ones with identical values of all indices which are assigned the value 1.
- While tensors of rank-0 are generally represented in a common form of light face non-indexed symbols, tensors of rank $\geqslant 1$ are represented in several forms and notations, the main ones are the index-free notation, which may also be called direct or symbolic or Gibbs notation, and the indicial notation which is also called index or component or tensor notation. The first is a geometrically oriented notation with no reference to a particular reference frame and hence it is intrinsically invariant to the choice of coordinate systems, whereas the second takes an algebraic form based on components identified by indices and hence the notation is suggestive of an underlying coordinate system, although being a tensor makes it form-invariant under certain coordinate transformations and therefore it possesses certain invariant properties. The index-free notation is usually identified by using bold face symbols, like **a** and **B**, while the indicial notation is identified by using light face indexed symbols such as a^i and B_{ij}.

10.2.2 Examples of Tensors of Different Ranks

- Examples of rank-0 tensors (scalars) are energy, mass, temperature, volume and density. These are totally identified by a single number regardless of any coordinate system and hence they are invariant under coordinate transformations.

[1] In the literature of tensor calculus, rank and order of tensors are generally used interchangeably; however some authors differentiate between the two as they assign order to the total number of indices, including repetitive indices, while they keep rank to the number of free indices. We think the latter is better and hence we follow this convention in the present text.

- Examples of rank-1 tensors (vectors) are displacement, force, electric field, velocity and acceleration. These need for their complete identification a number, representing their magnitude, and a direction representing their geometric orientation within their space. Alternatively, they can be uniquely identified by a set of numbers, equal to the number of dimensions of the underlying space, in reference to a particular coordinate system and hence this identification is system-dependent although they still have system-invariant properties such as length.

- Examples of rank-2 tensors are Kronecker delta (see S 10.5.1), stress, strain, rate of strain and inertia tensors. These require for their full identification a set of numbers each of which is associated with two directions.

- Examples of rank-3 tensors are the Levi-Civita tensor (see S 10.5.2) and the tensor of piezoelectric moduli.

- Examples of rank-4 tensors are the elasticity or stiffness tensor, the compliance tensor and the fourth-order moment of inertia tensor.

- Tensors of high ranks are relatively rare in science.

10.3 Tensor Operations in Coordinates

There are many operations that can be performed on tensors to produce other tensors in general. Some examples of these operations are addition/subtraction, multiplication by a scalar (rank-0 tensor), multiplication of tensors (each of rank > 0), contraction and permutation. Some of these operations, such as addition and multiplication, involve more than one tensor while others are performed on a single tensor, such as contraction and permutation.

In tensor algebra, division is allowed only for scalars, hence if the components of an indexed tensor should appear in a denominator, the tensor should be redefined to avoid this, e.g. $B_i = \frac{1}{A_i}$.

10.3.1 Addition and Subtraction

Tensors of the same rank and type can be added algebraically to produce a tensor of the same rank and type, e.g.

$$a = b + c \tag{10.7}$$

$$A_i = B_i - C_i \tag{10.8}$$

$$A^i_j = B^i_j + C^i_j \tag{10.9}$$

Definition 10.3.1 Given two tensors $Y^{i_1 \dots i_r}_{j_1 \dots j_s}$ and $Z^{i_1 \dots i_r}_{j_1 \dots j_s}$ of the same type then we define their sum as

$$X^{i_1 \dots i_r}_{j_1 \dots j_s} + Y^{i_1 \dots i_r}_{j_1 \dots j_s} = Z^{i_1 \dots i_r}_{j_1 \dots j_s}.$$

Theorem 10.3.1 Given two tensors $Y^{i_1 \dots i_r}_{j_1 \dots j_s}$ and $Z^{i_1 \dots i_r}_{j_1 \dots j_s}$ of type (r,s) then their sum

$$Z^{i_1 \dots i_r}_{j_1 \dots j_s} = X^{i_1 \dots i_r}_{j_1 \dots j_s} + Y^{i_1 \dots i_r}_{j_1 \dots j_s}.$$

is also a tensor of type (r,s).

Proof:

$$X^{i_1 \dots i_r}_{j_1 \dots j_s} = \sum_{\substack{h_1,\dots,h_r \\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{X}^{h_1 \dots h_r}_{k_1 \dots k_s},$$

$$Y^{i_1 \dots i_r}_{j_1 \dots j_s} = \sum_{\substack{h_1,\dots,h_r \\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{Y}^{h_1 \dots h_r}_{k_1 \dots k_s},$$

Then

$$Z^{i_1 \dots i_r}_{j_1 \dots j_s} = \sum_{\substack{h_1,\dots,h_r \\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{X}^{h_1 \dots h_r}_{k_1 \dots k_s}$$

$$+ \sum_{\substack{h_1,\dots,h_r \\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{Y}^{h_1 \dots h_r}_{k_1 \dots k_s}$$

$$Z^{i_1 \dots i_r}_{j_1 \dots j_s} = \sum_{\substack{h_1,\dots,h_r \\ k_1,\dots,k_s}}^{n} \dots \sum^{n} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \left(\tilde{X}^{h_1 \dots h_r}_{k_1 \dots k_s} + \tilde{Y}^{h_1 \dots h_r}_{k_1 \dots k_s} \right)$$

■

Addition of tensors is associative and commutative:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{10.10}$$

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{10.11}$$

10.3.2 Multiplication by Scalar

A tensor can be multiplied by a scalar, which generally should not be zero, to produce a tensor of the same variance type and rank, e.g.

$$A^{j}_{ik} = aB^{j}_{ik} \tag{10.12}$$

where a is a non-zero scalar.

Definition 10.3.2 Given $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ a tensor of type (r,s) and α a scalar we define the multiplication of $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ by α as:

$$Y^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \alpha X^{i_1 \ldots i_r}_{j_1 \ldots j_s}.$$

Theorem 10.3.2 Given $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ a tensor of type (r,s) and α a scalar then

$$Y^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \alpha X^{i_1 \ldots i_r}_{j_1 \ldots j_s}.$$

is also a tensor of type (r,s)

The proof of this Theorem is very similar to the proof of the Theorem 10.3.1 and the proof is left as an exercise to the reader.

As indicated above, multiplying a tensor by a scalar means multiplying each component of the tensor by that scalar.

Multiplication by a scalar is commutative, and associative when more than two factors are involved.

10.3.3 Tensor Product

This may also be called outer or exterior or direct or dyadic multiplication, although some of these names may be reserved for operations on vectors.

The tensor product is defined by a more tricky formula. Suppose we have tensor X of type (r,s) and tensor Y of type (p,q), then we can write:

$$Z^{i_1 \ldots i_{r+p}}_{j_1 \ldots j_{s+q}} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s} Y^{i_{r+1} \ldots i_{r+p}}_{j_{s+1} \ldots j_{s+q}}.$$

Formula 10.3.3 produces new tensor **Z** of the type $(r+p, s+q)$. It is called **the tensor product** of X and Y and denoted $\mathbf{Z} = X \otimes Y$.

Example 10.3.3

$$A_i B_j = C_{ij} \tag{10.13}$$

$$A^{ij} B_{kl} = C^{ij}_{kl} \tag{10.14}$$

Direct multiplication of tensors is not commutative.

Example 10.3.4 — Outer Product of Vectors. The outer product of two vectors is equivalent to a matrix multiplication $\mathbf{u}\mathbf{v}^T$, provided that $\mathbf{u}$ is represented as a column vector and $\mathbf{v}$ as a column vector. And so $\mathbf{v}^T$ is a row vector.

$$\mathbf{u}\otimes\mathbf{v}=\mathbf{u}\mathbf{v}^T=\begin{bmatrix}u_1\\u_2\\u_3\\u_4\end{bmatrix}\begin{bmatrix}v_1 & v_2 & v_3\end{bmatrix}=\begin{bmatrix}u_1v_1 & u_1v_2 & u_1v_3\\u_2v_1 & u_2v_2 & u_2v_3\\u_3v_1 & u_3v_2 & u_3v_3\\u_4v_1 & u_4v_2 & u_4v_3\end{bmatrix}. \tag{10.15}$$

In index notation:

$$(\mathbf{u}\mathbf{v}^T)_{ij}=u_iv_j$$

The outer product operation is distributive with respect to the algebraic sum of tensors:

$$\mathbf{A}(\mathbf{B}\pm\mathbf{C})=\mathbf{AB}\pm\mathbf{AC} \quad \& \quad (\mathbf{B}\pm\mathbf{C})\mathbf{A}=\mathbf{BA}\pm\mathbf{CA} \tag{10.16}$$

Multiplication of a tensor by a scalar (refer to S 10.3.2) may be regarded as a special case of direct multiplication.

The rank-2 tensor constructed as a result of the direct multiplication of two vectors is commonly called dyad.

Tensors may be expressed as an outer product of vectors where the rank of the resultant product is equal to the number of the vectors involved (e.g. 2 for dyads and 3 for triads).

Not every tensor can be synthesized as a product of lower rank tensors.

10.3.4 Contraction

Contraction of a tensor of rank >1 is to make two free indices identical, by unifying their symbols, and perform summation over these repeated indices, e.g.

$$A_i^j \xrightarrow{\text{contraction}} A_i^i \tag{10.17}$$

$$A_{il}^{jk} \xrightarrow{\text{contraction on } jl} A_{im}^{mk} \tag{10.18}$$

Contraction results in a reduction of the rank by 2 since it implies the annihilation of two free indices. Therefore, the contraction of a rank-2 tensor is a scalar, the contraction of a rank-3 tensor is a vector, the contraction of a rank-4 tensor is a rank-2 tensor, and so on.

For non-Cartesian coordinate systems, the pair of contracted indices should be different in their variance type, i.e. one upper and one lower. Hence, contraction of a mixed tensor of type (m,n) will, in general, produce a tensor of type $(m-1,n-1)$.

A tensor of type (p,q) can have $p\times q$ possible contractions, i.e. one contraction for each pair of lower and upper indices.

Example 10.3.5 — Trace. In matrix algebra, taking the trace (summing the diagonal elements) can also be considered as contraction of the matrix, which under certain conditions can represent a rank-2 tensor, and hence it yields the trace which is a scalar.

10.3.5 Inner Product

On taking the outer product of two tensors of rank $\geqslant 1$ followed by a contraction on two indices of the product, an inner product of the two tensors is formed. Hence if one of the original tensors is of rank-m and the other is of rank-n, the inner product will be of rank-$(m+n-2)$.

The inner product operation is usually symbolized by a single dot between the two tensors, e.g. $\mathbf{A} \cdot \mathbf{B}$, to indicate contraction following outer multiplication.

In general, the inner product is not commutative. When one or both of the tensors involved in the inner product are of rank > 1 the order of the multiplicands does matter.

The inner product operation is distributive with respect to the algebraic sum of tensors:

$$\mathbf{A} \cdot (\mathbf{B} \pm \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} \pm \mathbf{A} \cdot \mathbf{C} \quad \& \quad (\mathbf{B} \pm \mathbf{C}) \cdot \mathbf{A} = \mathbf{B} \cdot \mathbf{A} \pm \mathbf{C} \cdot \mathbf{A} \tag{10.19}$$

Example 10.3.6 — Dot Product. A common example of contraction is the dot product operation on vectors which can be regarded as a direct multiplication (refer to S 10.3.3) of the two vectors, which results in a rank-2 tensor, followed by a contraction.

Example 10.3.7 — Matrix acting on vectors. Another common example (from linear algebra) of inner product is the multiplication of a matrix (representing a rank-2 tensor) by a vector (rank-1 tensor) to produce a vector, e.g.

$$[\mathbf{A}\mathbf{b}]_{ij}^{k} = A_{ij} b^{k} \quad \xrightarrow{\text{contraction on } jk} \quad [\mathbf{A} \cdot \mathbf{b}]_{i} = A_{ij} b^{j} \tag{10.20}$$

The multiplication of two $n \times n$ matrices is another example of inner product (see Eq. **??**).

For tensors whose outer product produces a tensor of rank > 2, various contraction operations between different sets of indices can occur and hence more than one inner product, which are different in general, can be defined. Moreover, when the outer product produces a tensor of rank > 3 more than one contraction can take place simultaneously.

10.3.6 Permutation

A tensor may be obtained by exchanging the indices of another tensor, e.g. transposition of rank-2 tensors.

Obviously, tensor permutation applies only to tensors of rank $\geqslant 2$.

The collection of tensors obtained by permuting the indices of a basic tensor may be called **isomers**.

10.4 Tensor Test: Quotient Rule

Sometimes a tensor-like object may be suspected for being a tensor; in such cases a test based on the "quotient rule" can be used to clarify the situation. According to this rule, if the inner product of a suspected tensor with a known tensor is a tensor then the suspect is a tensor. In more formal terms, if it is not known if **A** is a tensor but it is known that **B** and **C** are tensors; moreover it is known that the following relation holds true in all rotated (properly-transformed) coordinate frames:

$$A_{pq\ldots k\ldots m}B_{ij\ldots k\ldots n} = C_{pq\ldots mij\ldots n} \tag{10.21}$$

then **A** is a tensor. Here, **A**, **B** and **C** are respectively of ranks m, n and $(m+n-2)$, due to the contraction on k which can be any index of **A** and **B** independently.

Testing for being a tensor can also be done by applying first principles through direct substitution in the transformation equations. However, using the quotient rule is generally more convenient and requires less work.

The quotient rule may be considered as a replacement for the division operation which is not defined for tensors.

10.5 Kronecker and Levi-Civita Tensors

These tensors are of particular importance in tensor calculus due to their distinctive properties and unique transformation attributes. They are numerical tensors with fixed components in all coordinate systems. The first is called Kronecker delta or unit tensor, while the second is called Levi-Civita

The δ and ϵ tensors are conserved under coordinate transformations and hence they are the same for all systems of coordinate.[2]

10.5.1 Kronecker δ

This is a rank-2 symmetric tensor in all dimensions, i.e.

$$\delta_{ij} = \delta_{ji} \qquad (i,j = 1,2,\ldots,n) \tag{10.22}$$

Similar identities apply to the contravariant and mixed types of this tensor.

It is invariant in all coordinate systems, and hence it is an isotropic tensor.[3]

It is defined as:

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{10.23}$$

[2] For the permutation tensor, the statement applies to proper coordinate transformations.

[3] In fact it is more general than isotropic as it is invariant even under improper coordinate transformations.

and hence it can be considered as the identity matrix, e.g. for 3D

$$[\delta_{ij}] = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10.24}$$

Covariant, contravariant and mixed type of this tensor are the same, that is

$$\delta^i_j = \delta_i^j = \delta^{ij} = \delta_{ij} \tag{10.25}$$

10.5.2 Permutation ϵ

This is an isotropic tensor. It has a rank equal to the number of dimensions; hence, a rank-n permutation tensor has n^n components.

It is totally anti-symmetric in each pair of its indices, i.e. it changes sign on swapping any two of its indices, that is

$$\epsilon_{i_1 \ldots i_k \ldots i_l \ldots i_n} = -\epsilon_{i_1 \ldots i_l \ldots i_k \ldots i_n} \tag{10.26}$$

The reason is that any exchange of two indices requires an even/odd number of single-step shifts to the right of the first index plus an odd/even number of single-step shifts to the left of the second index, so the total number of shifts is odd and hence it is an odd permutation of the original arrangement.

It is a pseudo tensor since it acquires a minus sign under improper orthogonal transformation of coordinates (inversion of axes with possible superposition of rotation).

Definition of rank-2 ϵ (ϵ_{ij}):

$$\epsilon_{12} = 1, \quad \epsilon_{21} = -1 \quad \& \quad \epsilon_{11} = \epsilon_{22} = 0 \tag{10.27}$$

Definition of rank-3 ϵ (ϵ_{ijk}):

$$\epsilon_{ijk} = \begin{cases} 1 & (i, j, k \text{ is even permutation of 1,2,3}) \\ -1 & (i, j, k \text{ is odd permutation of 1,2,3}) \\ 0 & (\text{repeated index}) \end{cases} \tag{10.28}$$

The definition of rank-n ϵ ($\epsilon_{i_1 i_2 \ldots i_n}$) is similar to the definition of rank-3 ϵ considering index repetition and even or odd permutations of its indices $(i_1, i_2, \cdots, i_n)$ corresponding to $(1, 2, \cdots, n)$, that is

$$\epsilon_{i_1 i_2 \ldots i_n} = \begin{cases} 1 & [(i_1, i_2, \ldots, i_n) \text{ is even permutation of } (1, 2, \ldots, n)] \\ -1 & [(i_1, i_2, \ldots, i_n) \text{ is odd permutation of } (1, 2, \ldots, n)] \\ 0 & [\text{repeated index}] \end{cases} \tag{10.29}$$

ϵ may be considered a contravariant relative tensor of weight $+1$ or a covariant relative tensor of weight -1. Hence, in 2, 3 and n dimensional spaces respectively we have:

$$\epsilon_{ij} = \epsilon^{ij} \tag{10.30}$$
$$\epsilon_{ijk} = \epsilon^{ijk} \tag{10.31}$$
$$\epsilon_{i_1 i_2 \ldots i_n} = \epsilon^{i_1 i_2 \ldots i_n} \tag{10.32}$$

10.5.3 Useful Identities Involving δ or/and ϵ

Identities Involving δ

When an index of the Kronecker delta is involved in a contraction operation by repeating an index in another tensor in its own term, the effect of this is to replace the shared index in the other tensor by the other index of the Kronecker delta, that is

$$\delta_{ij} A_j = A_i \tag{10.33}$$

In such cases the Kronecker delta is described as the substitution or index replacement operator. Hence,

$$\delta_{ij}\delta_{jk} = \delta_{ik} \tag{10.34}$$

Similarly,

$$\delta_{ij}\delta_{jk}\delta_{ki} = \delta_{ik}\delta_{ki} = \delta_{ii} = n \tag{10.35}$$

where n is the space dimension.

Because the coordinates are independent of each other:

$$\frac{\partial x_i}{\partial x_j} = \partial_j x_i = x_{i,j} = \delta_{ij} \tag{10.36}$$

Hence, in an n dimensional space we have

$$\partial_i x_i = \delta_{ii} = n \tag{10.37}$$

For orthonormal Cartesian systems:

$$\frac{\partial x^i}{\partial x^j} = \frac{\partial x^j}{\partial x^i} = \delta_{ij} = \delta^{ij} \tag{10.38}$$

For a set of orthonormal basis vectors in orthonormal Cartesian systems:

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \tag{10.39}$$

The double inner product of two dyads formed by orthonormal basis vectors of an orthonormal Cartesian system is given by:

$$\mathbf{e}_i \mathbf{e}_j : \mathbf{e}_k \mathbf{e}_l = \delta_{ik}\delta_{jl} \tag{10.40}$$

Identities Involving ϵ

For rank-3 ϵ:

$$\epsilon_{ijk} = \epsilon_{kij} = \epsilon_{jki} = -\epsilon_{ikj} = -\epsilon_{jik} = -\epsilon_{kji} \quad \text{(sense of cyclic order)} \tag{10.41}$$

These equations demonstrate the fact that rank-3 ϵ is totally anti-symmetric in all of its indices since a shift of any two indices reverses the sign. This also reflects the fact that the above tensor system has only one independent component.

For rank-2 ϵ:

$$\epsilon_{ij} = (j - i) \tag{10.42}$$

For rank-3 ϵ:

$$\epsilon_{ijk} = \frac{1}{2}(j-i)(k-i)(k-j) \tag{10.43}$$

For rank-4 ϵ:

$$\epsilon_{ijkl} = \frac{1}{12}(j-i)(k-i)(l-i)(k-j)(l-j)(l-k) \tag{10.44}$$

For rank-n ϵ:

$$\epsilon_{a_1 a_2 \cdots a_n} = \prod_{i=1}^{n-1}\left[\frac{1}{i!}\prod_{j=i+1}^{n}\left(a_j - a_i\right)\right] = \frac{1}{S(n-1)}\prod_{1\leqslant i<j\leqslant n}\left(a_j - a_i\right) \tag{10.45}$$

where $S(n-1)$ is the super-factorial function of $(n-1)$ which is defined as

$$S(k) = \prod_{i=1}^{k} i! = 1! \cdot 2! \cdot \ldots \cdot k! \tag{10.46}$$

A simpler formula for rank-n ϵ can be obtained from the previous one by ignoring the magnitude of the multiplication factors and taking only their signs, that is

$$\epsilon_{a_1 a_2 \cdots a_n} = \prod_{1\leqslant i<j\leqslant n} \sigma\left(a_j - a_i\right) = \sigma\left(\prod_{1\leqslant i<j\leqslant n}\left(a_j - a_i\right)\right) \tag{10.47}$$

where

$$\sigma(k) = \begin{cases} +1 & (k > 0) \\ -1 & (k < 0) \\ 0 & (k = 0) \end{cases} \tag{10.48}$$

For rank-n ϵ:

$$\epsilon_{i_1 i_2 \cdots i_n}\,\epsilon_{i_1 i_2 \cdots i_n} = n! \tag{10.49}$$

because this is the sum of the squares of $\epsilon_{i_1 i_2 \cdots i_n}$ over all the permutations of n different indices which is equal to $n!$ where the value of ϵ of each one of these permutations is either $+1$ or -1 and hence in both cases their square is 1.

For a symmetric tensor A_{jk}:

$$\epsilon_{ijk}\mathsf{A}_{jk} = 0 \tag{10.50}$$

because an exchange of the two indices of A_{jk} does not affect its value due to the symmetry whereas a similar exchange in these indices in ϵ_{ijk} results in a sign change; hence each term in the sum has its own negative and therefore the total sum will vanish.

$$\epsilon_{ijk}\mathsf{A}_i\mathsf{A}_j = \epsilon_{ijk}\mathsf{A}_i\mathsf{A}_k = \epsilon_{ijk}\mathsf{A}_j\mathsf{A}_k = 0 \tag{10.51}$$

because, due to the commutativity of multiplication, an exchange of the indices in A's will not affect the value but a similar exchange in the corresponding indices of ϵ_{ijk} will cause a change in sign; hence each term in the sum has its own negative and therefore the total sum will be zero.

For a set of orthonormal basis vectors in a 3D space with a right-handed orthonormal Cartesian coordinate system:

$$\mathbf{e}_i \times \mathbf{e}_j = \epsilon_{ijk}\mathbf{e}_k \tag{10.52}$$

$$\mathbf{e}_i \cdot \left(\mathbf{e}_j \times \mathbf{e}_k\right) = \epsilon_{ijk} \tag{10.53}$$

Identities Involving δ and ϵ

$$\epsilon_{ijk}\delta_{1i}\delta_{2j}\delta_{3k} = \epsilon_{123} = 1 \tag{10.54}$$

For rank-2 ϵ:

$$\epsilon_{ij}\epsilon_{kl} = \begin{vmatrix} \delta_{ik} & \delta_{il} \\ \delta_{jk} & \delta_{jl} \end{vmatrix} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} \tag{10.55}$$

$$\epsilon_{il}\epsilon_{kl} = \delta_{ik} \tag{10.56}$$

$$\epsilon_{ij}\epsilon_{ij} = 2 \tag{10.57}$$

For rank-3 ϵ:

$$\epsilon_{ijk}\epsilon_{lmn} = \begin{vmatrix} \delta_{il} & \delta_{im} & \delta_{in} \\ \delta_{jl} & \delta_{jm} & \delta_{jn} \\ \delta_{kl} & \delta_{km} & \delta_{kn} \end{vmatrix} = \delta_{il}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{kl} + \delta_{in}\delta_{jl}\delta_{km} - \delta_{il}\delta_{jn}\delta_{km} - \delta_{im}\delta_{jl}\delta_{kn} - \delta_{in}\delta_{jm}\delta_{kl} \tag{10.58}$$

$$\epsilon_{ijk}\epsilon_{lmk} = \begin{vmatrix} \delta_{il} & \delta_{im} \\ \delta_{jl} & \delta_{jm} \end{vmatrix} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \tag{10.59}$$

The last identity is very useful in manipulating and simplifying tensor expressions and proving vector and tensor identities.

$$\epsilon_{ijk}\epsilon_{ljk} = 2\delta_{il} \tag{10.60}$$

$$\epsilon_{ijk}\epsilon_{ijk} = 2\delta_{ii} = 6 \tag{10.61}$$

since the rank and dimension of ϵ are the same, which is 3 in this case.

For rank-n ϵ:

$$\epsilon_{i_1 i_2 \cdots i_n}\epsilon_{j_1 j_2 \cdots j_n} = \begin{vmatrix} \delta_{i_1 j_1} & \delta_{i_1 j_2} & \cdots & \delta_{i_1 j_n} \\ \delta_{i_2 j_1} & \delta_{i_2 j_2} & \cdots & \delta_{i_2 j_n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i_n j_1} & \delta_{i_n j_2} & \cdots & \delta_{i_n j_n} \end{vmatrix} \tag{10.62}$$

According to Eqs. 10.28 and 10.33:

$$\epsilon_{ijk}\delta_{ij} = \epsilon_{ijk}\delta_{ik} = \epsilon_{ijk}\delta_{jk} = 0 \tag{10.63}$$

10.5.4 ⋆ Generalized Kronecker delta

The generalized Kronecker delta is defined by:

$$\delta^{i_1 \ldots i_n}_{j_1 \ldots j_n} = \begin{cases} 1 & \left[(j_1 \ldots j_n) \text{ is even permutation of } (i_1 \ldots i_n)\right] \\ -1 & \left[(j_1 \ldots j_n) \text{ is odd permutation of } (i_1 \ldots i_n)\right] \\ 0 & [\text{repeated } j\text{'s}] \end{cases} \tag{10.64}$$

It can also be defined by the following $n \times n$ determinant:

$$\delta^{i_1 \ldots i_n}_{j_1 \ldots j_n} = \begin{vmatrix} \delta^{i_1}_{j_1} & \delta^{i_1}_{j_2} & \cdots & \delta^{i_1}_{j_n} \\ \delta^{i_2}_{j_1} & \delta^{i_2}_{j_2} & \cdots & \delta^{i_2}_{j_n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta^{i_n}_{j_1} & \delta^{i_n}_{j_2} & \cdots & \delta^{i_n}_{j_n} \end{vmatrix} \tag{10.65}$$

where the δ^i_j entries in the determinant are the normal Kronecker delta as defined by Eq. 10.23.

Accordingly, the relation between the rank-n ϵ and the generalized Kronecker delta in an n dimensional space is given by:

$$\epsilon_{i_1 i_2 \ldots i_n} = \delta^{12 \ldots n}_{i_1 i_2 \ldots i_n} \quad \& \quad \epsilon^{i_1 i_2 \ldots i_n} = \delta^{i_1 i_2 \ldots i_n}_{12 \ldots n} \tag{10.66}$$

Hence, the permutation tensor ϵ may be considered as a special case of the generalized Kronecker delta. Consequently the permutation symbol can be written as an $n \times n$ determinant consisting of the normal Kronecker deltas.

If we define

$$\delta^{ij}_{lm} = \delta^{ijk}_{lmk} \tag{10.67}$$

then Eq. 10.59 will take the following form:

$$\delta^{ij}_{lm} = \delta^{i}_{l}\delta^{j}_{m} - \delta^{i}_{m}\delta^{j}_{l} \tag{10.68}$$

Other identities involving δ and ϵ can also be formulated in terms of the generalized Kronecker delta.

On comparing Eq. 10.62 with Eq. 10.65 we conclude

$$\delta^{i_1 \dots i_n}_{j_1 \dots j_n} = \epsilon^{i_1 \dots i_n}\,\epsilon_{j_1 \dots j_n} \tag{10.69}$$

10.6 Types of Tensors Fields

In the following subsections we introduce a number of tensor types and categories and highlight their main characteristics and differences. These types and categories are not mutually exclusive and hence they overlap in general; moreover they may not be exhaustive in their classes as some tensors may not instantiate any one of a complementary set of types such as being symmetric or anti-symmetric.

10.6.1 Isotropic and Anisotropic Tensors

Isotropic tensors are characterized by the property that the values of their components are invariant under coordinate transformation by proper rotation of axes. In contrast, the values of the components of anisotropic tensors are dependent on the orientation of the coordinate axes. Notable examples of isotropic tensors are scalars (rank-0), the vector $\mathbf{0}$ (rank-1), Kronecker delta δ_{ij} (rank-2) and Levi-Civita tensor ϵ_{ijk} (rank-3). Many tensors describing physical properties of materials, such as stress and magnetic susceptibility, are anisotropic.

Direct and inner products of isotropic tensors are isotropic tensors.

The zero tensor of any rank is isotropic; therefore if the components of a tensor vanish in a particular coordinate system they will vanish in all properly and improperly rotated coordinate systems.[4] Consequently, if the components of two tensors are identical in a particular coordinate system they are identical in all transformed coordinate systems.

As indicated, all rank-0 tensors (scalars) are isotropic. Also, the zero vector, $\mathbf{0}$, of any dimension is isotropic; in fact it is the only rank-1 isotropic tensor.

Theorem 10.6.1 Any isotropic second order tensor T_{ij} we can be written as

$$T_{ij} = \lambda\delta_{ij}$$

for some scalar λ.

[4] For improper rotation, this is more general than being isotropic.

Proof: First we will prove that T is diagonal. Let R be the reflection in the hyperplane perpendicular to the j-th vector in the standard ordered basis.

$$R_{kl} = \begin{cases} -1 & \text{if } k = l = j \\ \delta_{kl} & \text{otherwise} \end{cases}$$

therefore

$$R = R^T \wedge R^2 = I \Rightarrow R^T R = RR^T = I$$

Therefore:

$$T_{ij} = \sum_{p,q} R_{ip} R_{jq} T_{pq} = R_{ii} R_{jj} T_{ij} i \neq j \Rightarrow T_{ij} = -T_{ij} \Rightarrow T_{ij} = 0$$

Now we will prove that $T_{jj} = T_{11}$. Let P be the permutation matrix that interchanges the 1st and j-th rows when acrting by left multiplication.

$$P_{kl} = \begin{cases} \delta_{jl} & \text{if } k = 1 \\ \delta_{1l} & \text{if } k = j \\ \delta_{kl} & \text{otherwise} \end{cases}$$

$$(P^T P)_{kl} = \sum_m P^T_{km} P_{ml} = \sum_m P_{mk} P_{ml} = \sum_{m \neq 1,j} P_{mk} P_{ml} + \sum_{m=1,j} P_{mk} P_{ml} = \sum_{m \neq 1,j} \delta_{mk}\delta_{ml} + \delta_{jk}\delta_{jl} + \delta_{1k}\delta_{1l} = \sum_m \delta$$

Therefore:

$$T_{jj} = \sum_{p,q} P_{jp} P_{jq} T_{pq} = \sum_q P^2_{jq} T_{qq} = \sum_q \delta^2_{1q} T_{qq} = \sum_q \delta_{1q} T_{qq} = T_{11}$$

■

10.6.2 Symmetric and Anti-symmetric Tensors

These types of tensor apply to high ranks only (rank $\geqslant 2$). Moreover, these types are not exhaustive, even for tensors of ranks $\geqslant 2$, as there are high-rank tensors which are neither symmetric nor anti-symmetric.

A rank-2 tensor A_{ij} is symmetric *iff* for all i and j

$$A_{ji} = A_{ij} \tag{10.70}$$

and anti-symmetric or skew-symmetric *iff*

$$A_{ji} = -A_{ij} \tag{10.71}$$

Similar conditions apply to contravariant type tensors (refer also to the following).

A rank-n tensor $A_{i_1 \dots i_n}$ is symmetric in its two indices i_j and i_l *iff*

$$A_{i_1 \dots i_l \dots i_j \dots i_n} = A_{i_1 \dots i_j \dots i_l \dots i_n} \tag{10.72}$$

and anti-symmetric or skew-symmetric in its two indices i_j and i_l *iff*

$$A_{i_1 \dots i_l \dots i_j \dots i_n} = -A_{i_1 \dots i_j \dots i_l \dots i_n} \tag{10.73}$$

Any rank-2 tensor A_{ij} can be synthesized from (or decomposed into) a symmetric part $A_{(ij)}$ (marked with round brackets enclosing the indices) and an anti-symmetric part $A_{[ij]}$ (marked with square brackets) where

$$A_{ij} = A_{(ij)} + A_{[ij]}, \quad A_{(ij)} = \frac{1}{2}\left(A_{ij} + A_{ji}\right) \quad \& \quad A_{[ij]} = \frac{1}{2}\left(A_{ij} - A_{ji}\right) \tag{10.74}$$

A rank-3 tensor A_{ijk} can be symmetrized by

$$A_{(ijk)} = \frac{1}{3!}\left(A_{ijk} + A_{kij} + A_{jki} + A_{ikj} + A_{jik} + A_{kji}\right) \tag{10.75}$$

and anti-symmetrized by

$$A_{[ijk]} = \frac{1}{3!}\left(A_{ijk} + A_{kij} + A_{jki} - A_{ikj} - A_{jik} - A_{kji}\right) \tag{10.76}$$

A rank-n tensor $A_{i_1 \dots i_n}$ can be symmetrized by

$$A_{(i_1 \dots i_n)} = \frac{1}{n!}\left(\text{sum of all even \& odd permutations of indices } i\text{'s}\right) \tag{10.77}$$

and anti-symmetrized by

$$A_{[i_1 \dots i_n]} = \frac{1}{n!}\left(\text{sum of all even permutations minus sum of all odd permutations}\right) \tag{10.78}$$

For a symmetric tensor A_{ij} and an anti-symmetric tensor B^{ij} (or the other way around) we have

$$A_{ij} B^{ij} = 0 \tag{10.79}$$

The indices whose exchange defines the symmetry and anti-symmetry relations should be of the same variance type, i.e. both upper or both lower.

The symmetry and anti-symmetry characteristic of a tensor is invariant under coordinate transformation.

A tensor of high rank (> 2) may be symmetrized or anti-symmetrized with respect to only some of its indices instead of all of its indices, e.g.

$$A_{(ij)k} = \frac{1}{2}\left(A_{ijk} + A_{jik}\right) \quad \& \quad A_{[ij]k} = \frac{1}{2}\left(A_{ijk} - A_{jik}\right) \tag{10.80}$$

A tensor is totally symmetric *iff*

$$\mathsf{A}_{i_1 \dots i_n} = \mathsf{A}_{(i_1 \dots i_n)} \tag{10.81}$$

and totally anti-symmetric *iff*

$$\mathsf{A}_{i_1 \dots i_n} = \mathsf{A}_{[i_1 \dots i_n]} \tag{10.82}$$

For a totally skew-symmetric tensor (i.e. anti-symmetric in all of its indices), nonzero entries can occur only when all the indices are different.

CHAPTER 11

Tensor Calculus

11.1 Tensor Fields

In many applications, especially in differential geometry and physics, it is natural to consider a tensor with components that are functions of the point in a space. This was the setting of Ricci's original work. In modern mathematical terminology such an object is called a tensor field and often referred to simply as a tensor.

Definition 11.1.1 A **tensor field** of type (r, s) is a map $T : V \to T_r^s(V)$.

The space of all tensor fields of type (r, s) is denoted $\mathcal{T}_r^s(V)$. In this way, given $T \in \mathcal{T}_r^s(V)$, if we apply this to a point $p \in V$, we obtain $T(p) \in T_r^s(V)$

It's usual to write the point p as an index:

$$T_p : (v_1, \ldots, \omega_n) \mapsto T_p(v_1, \ldots, \omega_n) \in \mathbb{R}$$

Example 11.1.1

- If $f \in \mathcal{T}_0^0(V)$ then f is a scalar function.
- If $T \in \mathcal{T}_1^0(V)$ then T is a vector field.
- If $T \in \mathcal{T}_0^1(V)$ then f is called differential form of rank 1.

Differential

Now we will construct the one of the most important tensor field: the differential.

Given a differentiable scalar function f the directional derivative

$$D_v f(p) := \frac{d}{dt} f(p + tv)\Big|_{t=0}$$

is a linear function of v.

$$(D_{v+w}f)(p) = (D_v f)(p) + (D_w f)(p) \tag{11.1}$$

$$(D_{cv}f)(p) = c(D_v f)(p) \tag{11.2}$$

In other words $D_v f(p) \in \mathcal{T}_0^1(V)$

Definition 11.1.2 Let $f : V \to \mathbb{R}$ be a differentiable function. The *differential* of f, denoted by df, is the differential form defined by

$$df_p v = D_v f(p).$$

Clearly, $df \in \mathcal{T}_0^1(V)$

Let $\{u^1, u^2, \ldots, u^n\}$ be a coordinate system. Since the coordinates $\{u^1, u^2, \ldots, u^n\}$ are themselves functions, we define the associated differential-forms $\{du^1, du^2, \ldots, du^n\}$.

Proposition 11.1.2 Let $\{u^1, u^2, \ldots, u^n\}$ be a coordinate system and $\frac{\partial \mathbf{r}}{\partial u_i}(p)$ the corresponding basis of V. Then the differential-forms $\{du^1, du^2, \ldots, du^n\}$ are the corresponding dual basis:

$$du_p^i \left(\frac{\partial \mathbf{r}}{\partial u_j}(p) \right) = \delta_i^j$$

Since $\frac{\partial u^i}{\partial u^j} = \delta_j^i$, it follows that

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial u^i}\, du^i.$$

We also have the following product rule

$$d(fg) = (df)g + f(dg)$$

As consequence of Theorem 11.1.3 and Proposition 11.1.2 we have:

Theorem 11.1.3 Given $T \in \mathcal{T}_s^r(V)$ be a (r, s) tensor. Then T can be expressed in coordinates as:

$$T = \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} A_{j_{r+1} \cdots j_n}^{j_1 \cdots j_r}\, du^{j_1} \otimes du^{j_r} \otimes \frac{\partial r}{\partial u_{j_{r+1}}}(p) \cdots \otimes \frac{\partial r}{\partial u_{j_{r+s}}}(p)$$

11.1.1 Change of Coordinates

Let $\{u^1, u^2, \ldots, u^n\}$ and $\{\bar{u}^1, \bar{u}^2, \ldots, \bar{u}^n\}$ two coordinates system and $\{\frac{\partial \mathbf{r}}{\partial u_i}(p)\}$ and $\{\frac{\partial \mathbf{r}}{\partial \bar{u}_i}(p)\}$ the basis of V with $\{du^j\}$ and $\{d\bar{u}^j\}$ are the corresponding dual basis:

By the chain rule we have that the vectors change of basis as:

$$\frac{\partial \mathbf{r}}{\partial \bar{u}_j}(p) = \frac{\partial u_i}{\partial \bar{u}_j}(p)\frac{\partial \mathbf{r}}{\partial u_i}(p)$$

So the matrix of change of basis is:

$$A_i^j = \frac{\partial u_i}{\partial \bar{u}_j}$$

And the covectors changes by the inverse:

$$\left(A^{-1}\right)_i^j = \frac{\partial \bar{u}_j}{\partial u_i}$$

Theorem 11.1.4 — Change of Basis For Tensor Fields. Let $\{u^1, u^2, \ldots, u^n\}$ and $\{\bar{u}^1, \bar{u}^2, \ldots, \bar{u}^n\}$ two coordinates system and T a tensor

$$\hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n) = \frac{\partial \bar{u}^{i'_1}}{\partial u^{i_1}} \cdots \frac{\partial \bar{u}^{i'_p}}{\partial u^{i_p}} \frac{\partial u^{j_1}}{\partial \bar{u}^{j'_1}} \cdots \frac{\partial u^{j_q}}{\partial \bar{u}^{j'_q}} T^{i_1 \ldots i_p}_{j_1 \ldots j_q}(u^1, \ldots, u^n).$$

Example 11.1.5 — Contravariance. The tangent vector to a curve is a contravariant vector.

Solution: ► Let the curve be given by the parameterization $x^i = x^i(t)$. Then the tangent vector to the curve is

$$T^i = \frac{dx^i}{dt}$$

Under a change of coordinates, the curve is given by

$$x'^i = x'^i(t) = x'^i(x^1(t), \cdots, x^n(t))$$

and the tangent vector in the new coordinate system is given by:

$$T'^i = \frac{dx'^i}{dt}$$

By the chain rule,

$$\frac{dx'^i}{dt} = \frac{\partial x'^i}{\partial x^j}\frac{dx^j}{dt}$$

Therefore,

$$T'^i = T^j \frac{\partial x'^i}{\partial x^j}$$

which shows that the tangent vector transforms contravariantly and thus it is a contravariant vector.
◄

■ **EXEMPLO 11.1 — Covariance.** The gradient of a scalar field is a covariant vector field. ■

Solution: ► Let $\phi(\mathbf{x})$ be a scalar field. Then let

$$\mathbf{G} = \nabla\phi = \left(\frac{\partial\phi}{\partial x^1}, \frac{\partial\phi}{\partial x^2}, \frac{\partial\phi}{\partial x^3}, \cdots, \frac{\partial\phi}{\partial x^n}\right)$$

thus

$$G_i = \frac{\partial\phi}{\partial x^i}$$

In the primed coordinate system, the gradient is

$$G'_i = \frac{\partial\phi'}{\partial x'^i}$$

where $\phi' = \phi'(\mathbf{x}') = \phi(\mathbf{x}(\mathbf{x}'))$ By the chain rule,

$$\frac{\partial\phi'}{\partial x'^i} = \frac{\partial\phi}{\partial x^j}\frac{\partial x^j}{\partial x'^i}$$

Thus

$$G'_i = G_j\frac{\partial x^j}{\partial x'^i}$$

which shows that the gradient is a covariant vector.

◄

■ **EXEMPLO 11.2** A covariant tensor has components $xy, z^2, 3yz - x$ in rectangular coordinates. Write its components in spherical coordinates. ■

Solution: ► Let A_i denote its coordinates in rectangular coordinates $(x^1, x^2, x^3) = (x, y, z)$.

$$A_1 = xy \qquad A_2 = z^2, \qquad A_3 = 3y - x$$

Let $\bar{A}_k$ denote its coordinates in spherical coordinates $(\bar{x}^1, \bar{x}^2, \bar{x}^3) = (r, \phi, \theta)$:

Then

$$\bar{A}_k = \frac{\partial x^j}{\partial \bar{x}^k} A_j$$

The relation between the two coordinates systems are given by:

$$x = r\sin\phi\cos\theta;\; y = r\sin\phi\sin\theta;\; z = r\cos\phi$$

And so:

$$\bar{A}_1 = \frac{\partial x^1}{\partial \bar{x}^1}A_1 + \frac{\partial x^2}{\partial \bar{x}^1}A_2 + \frac{\partial x^3}{\partial \bar{x}^1}A_3 \tag{11.3}$$

$$= \sin\phi\cos\theta(xy) + \sin\phi\sin\theta(z^2) + \cos\phi(3y - x) \tag{11.4}$$

$$= \sin\phi\cos\theta(r\sin\phi\cos\theta)(r\sin\phi\sin\theta) + \sin\phi\sin\theta(r\cos\phi)^2 \tag{11.5}$$

$$+ \cos\phi(3r\sin\phi\sin\theta - r\sin\phi\cos\theta) \tag{11.6}$$

$$
\begin{aligned}
\bar{A}_2 &= \frac{\partial x^1}{\partial \bar{x}^2}A_1 + \frac{\partial x^2}{\partial \bar{x}^2}A_2 + \frac{\partial x^3}{\partial \bar{x}^2}A_3 && (11.7)\\
&= r\cos\phi\cos\theta(xy) + r\cos\phi\sin\theta(z^2) + -r\sin\phi(3y - x) && (11.8)\\
&= r\cos\phi\cos\theta(r\sin\phi\cos\theta)(r\sin\phi\sin\theta) + r\cos\phi\sin\theta(r\cos\phi)^2 && (11.9)\\
&+ r\sin\phi(3r\sin\phi\sin\theta - r\sin\phi\cos\theta) && (11.10)
\end{aligned}
$$

$$
\begin{aligned}
\bar{A}_3 &= \frac{\partial x^1}{\partial \bar{x}^3}A_1 + \frac{\partial x^2}{\partial \bar{x}^3}A_2 + \frac{\partial x^3}{\partial \bar{x}^3}A_3 && (11.11)\\
&= -r\sin\phi\sin\theta(xy) + r\sin\phi\cos\theta(z^2) + 0) && (11.12)\\
&= -r\sin\phi\sin\theta(r\sin\phi\cos\theta)(r\sin\phi\sin\theta) + r\sin\phi\cos\theta(r\cos\phi)^2 && (11.13)\\
& && (11.14)
\end{aligned}
$$

◀

11.2 Derivatives

In this section we consider two different types of derivatives of tensor fields: differentiation with respect to spacial variables $x^1, \ldots, x^n$ and differentiation with respect to parameters other than the spatial ones.

The second type of derivatives are simpler to define. Suppose we have tensor field T of type (r, s) and depending on the additional parameter t (for instance, this could be a time variable). Then, upon choosing some Cartesian coordinate system, we can write

$$\frac{\partial X^{i_1\ldots i_r}_{j_1\ldots j_s}}{\partial t} = \lim_{h\to 0} \frac{X^{i_1\ldots i_r}_{j_1\ldots j_s}(t+h, x^1, \ldots, x^n) - X^{i_1\ldots i_r}_{j_1\ldots j_s}(t, x^1, \ldots, x^n)}{h}. \tag{11.15}$$

The left hand side of 11.15 is a tensor since the fraction in right hand side is constructed by means of two tensorial operations: difference and scalar multiplication. Taking the limit $h \to 0$ preserves the tensorial nature of this fraction since the matrices of change of coordinates are time-independent.

So the differentiation with respect to external parameters is a tensorial operation producing new tensors from existing ones.

Now let's consider the spacial derivative of tensor field T, e.g, the derivative with respect to x^1. In this case we want to write the derivative as

$$\frac{\partial T^{i_1\ldots i_r}_{j_1\ldots j_s}}{\partial x^1} = \lim_{h\to 0} \frac{T^{i_1\ldots i_r}_{j_1\ldots j_s}(x^1+h, \ldots, x^n) - T^{i_1\ldots i_r}_{j_1\ldots j_s}(x^1, \ldots, x^n)}{h}, \tag{11.16}$$

but in numerator of the fraction in the right hand side of 11.16 we get the difference of two tensors bound to different points of space: the point $x^1, \ldots, x^n$ and the point $x^1 + h, \ldots, x^n$.

In general we can't sum the coordinates of tensors defined in different points since these tensors are written with respect to distinct basis of vector and covectors, as both basis varies with the point. In Cartesian coordinate system we don't have this dependence. And both tensors are written in the same basis and everything is well defined.

We now claim:

Theorem 11.2.1 For any tensor field $\mathbf{T}$ of type (r,s) partial derivatives with respect to spacial variables $u_1, \ldots, u_n$

$$\underbrace{\frac{\partial}{\partial u^a} \cdots \frac{\partial}{\partial x_c}}_{m} T^{i_1 \ldots i_r}_{j_1 \ldots j_s},$$

in any Cartesian coordinate system represent another tensor field of the type $(r, s+m)$.

Proof: Since T is a Tensor

$$T^{i_1 \ldots i_p}_{j_1 \ldots j_q}(u^1, \ldots, u^n) = \frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n).$$

and so:

$$\frac{\partial}{\partial u^a} T^{i_1 \ldots i_p}_{j_1 \ldots j_q}(u^1, \ldots, u^n) = \frac{\partial}{\partial u^a} \left(\frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n) \right) \tag{11.17}$$

$$= \frac{\partial}{\partial u^a} \left(\frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \right) \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n) + \tag{11.18}$$

$$\frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \frac{\partial}{\partial u^a} \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n) \tag{11.19}$$

We are assuming that the matrices

$$\frac{\partial u^{i_s}}{\partial \bar{u}^{i'_s}} \qquad \frac{\partial \bar{u}^{j'_l}}{\partial u^{j_l}}$$

are constant matrices.

And so

$$\frac{\partial}{\partial u^a} \frac{\partial u^{i_s}}{\partial \bar{u}^{i'_s}} = 0 \qquad \frac{\partial}{\partial u^a} \frac{\partial \bar{u}^{j'_l}}{\partial u^{j_l}} = 0$$

Hence

$$\frac{\partial}{\partial u^a} \left(\frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \right) \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1, \ldots, \bar{u}^n) = 0$$

And

$$\frac{\partial}{\partial u^a} T^{i_1 \ldots i_p}_{j_1 \ldots j_q}(u^1,\ldots,u^n) = \frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \frac{\partial}{\partial u^a} \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1,\ldots,\bar{u}^n) \tag{11.20}$$

$$= \frac{\partial u^{i_1}}{\partial \bar{u}^{i'_1}} \cdots \frac{\partial u^{i_p}}{\partial \bar{u}^{i'_p}} \frac{\partial \bar{u}^{j'_1}}{\partial u^{j_1}} \cdots \frac{\partial \bar{u}^{j'_q}}{\partial u^{j_q}} \frac{\partial \bar{u}'_a}{\partial u^a} \left[\frac{\partial}{\partial \bar{u}'_a} \hat{T}^{i'_1 \ldots i'_p}_{j'_1 \ldots j'_q}(\bar{u}^1,\ldots,\bar{u}^n) \right] \tag{11.21}$$

■

R We note that in general the partial derivative is not a tensor. Given a vector field

$$\mathbf{v} = v^j \frac{\partial \mathbf{r}}{\partial u^j} \quad ,$$

then

$$\frac{\partial \mathbf{v}}{\partial u^i} = \frac{\partial v^j}{\partial u^i} \frac{\partial \mathbf{r}}{\partial u^j} + v^j \frac{\partial^2 \mathbf{r}}{\partial u^i \partial u^j}.$$

The term $\dfrac{\partial^2 \mathbf{r}}{\partial u^i \partial u^j}$ in general is not null if the coordinate system is not the Cartesian.

■ **Exemplo 11.3** Calculate

$$\partial_{x^m} \partial_{\lambda^n} (A^{ij} \lambda^i x^j + B^{ij} x^i \lambda^j)$$

■

Solution: ►

$$\partial_{x^m} \partial_{\lambda^n} (A^{ij} \lambda^i x^j + B^{ij} x^i \lambda^j) = A^{ij} \delta^{in} \delta^{jm} + B^{ij} \delta^{im} \delta^{jn} \tag{11.22}$$

$$= A^{nm} + B^{mn} \tag{11.23}$$

◄

■ **Exemplo 11.4** Prove that if F_{ik} is an antisymmetric tensor then

$$T_{ijk} = \partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij}$$

is a tensor .

■

Solution: ►

The tensor F_{ik} changes as:

$$F_{jk} = \frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \bar{F}_{ab}$$

Then

$$\partial_i F_{jk} = \partial_i \left(\frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \bar{F}_{ab} \right) \tag{11.24}$$

$$= \partial_i \left(\frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \right) \bar{F}_{ab} + \frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \partial_i \bar{F}_{ab} \tag{11.25}$$

$$= \partial_i \left(\frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \right) \bar{F}_{ab} + \frac{\partial x^j}{\partial x'^a} \frac{\partial x^k}{\partial x'^b} \frac{\partial x^i}{\partial x'^a} \partial_a \bar{F}_{ab} \tag{11.26}$$

The tensor

$$T_{ijk} = \partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij}$$

is totally antisymmetric under any index pair exchange. Now perform a coordinate change, T_{ijk} will transform as

$$T_{abc} = \frac{\partial x^i}{\partial x'^a} \frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} T_{ijk} + I_{abc}$$

where this I_{abc} is given by:

$$I_{abc} = \frac{\partial x^i}{\partial x'^a} \partial_i \left(\frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \right) F_{jk} + \cdots$$

such I_{abc} will clearly be also totally antisymmetric under exchange of any pair of the indices a, b, c. Notice now that we can rewrite:

$$I_{abc} = \frac{\partial}{\partial x'^a} \left(\frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \right) F_{jk} + \cdots = \frac{\partial^2 x^j}{\partial x'^a x'^b} \frac{\partial x^k}{\partial x'^c} F_{jk} + \frac{\partial x^j}{\partial x'^b} \frac{\partial^2 x^j}{\partial x'^a x'^c} F_{jk} + \cdots$$

and they all vanish because the object is antisymmetric in the indices a, b, c while the mixed partial derivatives are symmetric (remember that an object both symmetric and antisymmetric is zero), hence T_{ijk} is a tensor. ◄

Problem 11.1 Give a more detailed explanation of why the time derivative of a tensor of type (r, s) is tensor of type (r, s).

11.3 Integrals and the Tensor Divergence Theorem

It is also straightforward to do integrals. Since we can sum tensors and take limits, the definition of a tensor-valued integral is straightforward.

For example, $\int_V T_{ij\cdots k}(\mathbf{x})\, dV$ is a tensor of the same rank as $T_{ij\cdots k}$ (think of the integral as the limit of a sum).

It is easy to generalize the divergence theorem from vectors to tensors.

Theorem 11.3.1 — Divergence Theorem for Tensors. Let $T_{ijk\dots}$ be a continuously differentiable tensor defined on a domain V with a piecewise-differentiable boundary (i.e. for almost all points, we have a well-defined normal vector n^l), then we have

$$\int_S T_{ij\dots k\ell} n^\ell \, dS = \int_V \frac{\partial}{\partial x_\ell}(T_{ij\dots k\ell}) \, dV,$$

with $\mathbf{n}$ being an outward pointing normal.

The regular divergence theorem is the case where T has one index and is a vector field.

Proof: The tensor form of the divergence theorem can be obtained applying the usual divergence theorem to the vector field $\mathbf{v}$ defined by $v_\ell = a^i b^j \cdots c^k T_{ij\dots k\ell}$, where $\mathbf{a}, \mathbf{b}, \cdots, \mathbf{c}$ are fixed constant vectors.

Then

$$\nabla \cdot \mathbf{v} = \frac{\partial v_\ell}{\partial x^\ell} = a^i b^j \cdots c^k \frac{\partial}{\partial x^\ell} T^{ij\cdots k\ell},$$

and

$$\mathbf{n} \cdot \mathbf{v} = n^\ell v_\ell = a^i b^j \cdots c^k T_{ij\dots k\ell} n^\ell.$$

Since $\mathbf{a}, \mathbf{b}, \cdots, \mathbf{c}$ are arbitrary, therefore they can be eliminated, and the tensor divergence theorem follows. ■

11.4 Metric Tensor

This is a rank-2 tensor which may also be called the fundamental tensor.

The main purpose of the metric tensor is to generalize the concept of distance to general curvilinear coordinate frames and maintain the invariance of distance in different coordinate systems.

In orthonormal Cartesian coordinate systems the distance element squared, $(ds)^2$, between two infinitesimally neighboring points in space, one with coordinates x^i and the other with coordinates $x^i + dx^i$, is given by

$$(ds)^2 = dx^i dx^i = \delta_{ij} dx^i dx^j \tag{11.27}$$

This definition of distance is the key to introducing a rank-2 tensor, g_{ij}, called the metric tensor which, for a general coordinate system, is defined by

$$(ds)^2 = g_{ij} dx^i dx^j \tag{11.28}$$

The metric tensor has also a contravariant form, i.e. g^{ij}.

The components of the metric tensor are given by:

$$g_{ij} = \mathbf{E}_i \cdot \mathbf{E}_j \qquad \& \qquad g^{ij} = \mathbf{E}^i \cdot \mathbf{E}^j \tag{11.29}$$

where the indexed **E** are the covariant and contravariant basis vectors:

$$\mathbf{E}_i = \frac{\partial \mathbf{r}}{\partial u^i} \qquad \& \qquad \mathbf{E}^i = \nabla u^i \tag{11.30}$$

where **r** is the position vector in Cartesian coordinates and u^i is a generalized curvilinear coordinate.

The mixed type metric tensor is given by:

$$g^i_j = \mathbf{E}^i \cdot \mathbf{E}_j = \delta^i_j \qquad \& \qquad g_i^j = \mathbf{E}_i \cdot \mathbf{E}^j = \delta_i^j \tag{11.31}$$

and hence it is the same as the unity tensor.

For a coordinate system in which the metric tensor can be cast in a diagonal form where the diagonal elements are ± 1 the metric is called flat.

For Cartesian coordinate systems, which are orthonormal flat-space systems, we have

$$g^{ij} = \delta^{ij} = g_{ij} = \delta_{ij} \tag{11.32}$$

The metric tensor is symmetric, that is

$$g_{ij} = g_{ji} \qquad \& \qquad g^{ij} = g^{ji} \tag{11.33}$$

The contravariant metric tensor is used for raising indices of covariant tensors and the covariant metric tensor is used for lowering indices of contravariant tensors, e.g.

$$A^i = g^{ij}\mathsf{A}_j \qquad \mathsf{A}^i = g_{ij}A^j \tag{11.34}$$

where the metric tensor acts, like a Kronecker delta, as an index replacement operator. Hence, any tensor can be cast into a covariant or a contravariant form, as well as a mixed form. However, the order of the indices should be respected in this process, e.g.

$$\mathsf{A}^i_j = g_{jk}A^{ik} \neq \mathsf{A}_j{}^i = g_{jk}A^{ki} \tag{11.35}$$

Some authors insert dots (e.g. $\mathsf{A}_{j}^{\cdot i}$) to remove any ambiguity about the order of the indices.

The covariant and contravariant metric tensors are inverses of each other, that is

$$\left[g_{ij}\right] = \left[g^{ij}\right]^{-1} \qquad \& \qquad \left[g^{ij}\right] = \left[g_{ij}\right]^{-1} \tag{11.36}$$

Hence

$$g^{ik}g_{kj} = \delta^i_j \qquad \& \qquad g_{ik}g^{kj} = \delta_i^j \tag{11.37}$$

It is common to reserve the "metric tensor" to the covariant form and call the contravariant form, which is its inverse, the "associate" or "conjugate" or "reciprocal" metric tensor.

As a tensor, the metric has a significance regardless of any coordinate system although it requires a coordinate system to be represented in a specific form.

For orthogonal coordinate systems the metric tensor is diagonal, i.e. $g_{ij} = g^{ij} = 0$ for $i \neq j$.

For flat-space orthonormal Cartesian coordinate systems in a 3D space, the metric tensor is given by:

$$[g_{ij}] = [\delta_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \left[\delta^{ij}\right] = \left[g^{ij}\right] \tag{11.38}$$

For cylindrical coordinate systems with coordinates (ρ, ϕ, z), the metric tensor is given by:

$$[g_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \& \quad \left[g^{ij}\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\rho^2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11.39}$$

For spherical coordinate systems with coordinates (r, θ, ϕ), the metric tensor is given by:

$$[g_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{bmatrix} \quad \& \quad \left[g^{ij}\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{1}{r^2 \sin^2\theta} \end{bmatrix} \tag{11.40}$$

11.5 Covariant Differentiation

Let $\{x^1, \dots, x^n\}$ be a coordinate system. And

$$\left\{ \left.\frac{\partial \mathbf{r}}{\partial x^i}\right|_p : i \in \{1, \dots, n\} \right\}$$

the associated basis

The metric tensor $g_{ij} = \left\langle \frac{\partial \mathbf{r}}{\partial x^i}; \frac{\partial \mathbf{r}}{\partial x^j} \right\rangle$.

Given a vector field

$$\mathbf{v} = v^j \frac{\partial \mathbf{r}}{\partial x^j} \quad ,$$

then

$$\frac{\partial \mathbf{v}}{\partial x^i} = \frac{\partial v^j}{\partial x^i} \frac{\partial \mathbf{r}}{\partial x^j} + v^j \frac{\partial^2 \mathbf{r}}{\partial x^i \partial x^j}.$$

The last term but can be expressed as a linear combination of the tangent space base vectors using the Christoffel symbols

$$\frac{\partial^2 \mathbf{r}}{\partial x^i \partial x^j} = \Gamma^k{}_{ij} \frac{\partial \mathbf{r}}{\partial x^k}.$$

Definition 11.5.1 The covariant derivative $\nabla_{\mathbf{e}_i}\mathbf{v}$, also written $\nabla_i\mathbf{v}$, is defined as:

$$\nabla_{\mathbf{e}_i}\mathbf{v} := \frac{\partial \mathbf{v}}{\partial x^i} = \left(\frac{\partial v^k}{\partial x^i} + v^j \Gamma^k{}_{ij}\right)\frac{\partial \mathbf{r}}{\partial x^k}.$$

The Christoffel symbols can be calculated using the inner product:

$$\left\langle \frac{\partial^2 \mathbf{r}}{\partial x^i \partial x^j}, \frac{\partial \mathbf{r}}{\partial x^l}\right\rangle = \Gamma^k{}_{ij}\left\langle \frac{\partial \mathbf{r}}{\partial x^k}, \frac{\partial \mathbf{r}}{\partial x^l}\right\rangle = \Gamma^k{}_{ij}\, g_{kl}.$$

On the other hand,

$$\frac{\partial g_{ab}}{\partial x^c} = \left\langle \frac{\partial^2 \mathbf{r}}{\partial x^c \partial x^a}, \frac{\partial \mathbf{r}}{\partial x^b}\right\rangle + \left\langle \frac{\partial \mathbf{r}}{\partial x^a}, \frac{\partial^2 \mathbf{r}}{\partial x^c \partial x^b}\right\rangle$$

using the symmetry of the scalar product and swapping the order of partial differentiations we have

$$\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} = 2\left\langle \frac{\partial^2 \mathbf{r}}{\partial x^i \partial x^j}, \frac{\partial \mathbf{r}}{\partial x^k}\right\rangle$$

and so we have expressed the Christoffel symbols for the Levi-Civita connection in terms of the metric:

$$g_{kl}\Gamma^k{}_{ij} = \frac{1}{2}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right).$$

Definition 11.5.2 Christoffel symbol of the second kind is defined by:

$$\Gamma^k_{ij} = \frac{g^{kl}}{2}\left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l}\right) \tag{11.41}$$

where the indexed g is the metric tensor in its contravariant and covariant forms with implied summation over l. It is noteworthy that Christoffel symbols are not tensors.

The Christoffel symbols of the second kind are symmetric in their two lower indices:

$$\Gamma^k_{ij} = \Gamma^k_{ji} \tag{11.42}$$

■ **Exemplo 11.5** For Cartesian coordinate systems, the Christoffel symbols are zero for all the values of indices. ■

EXEMPLO 11.6 For cylindrical coordinate systems (ρ,ϕ,z), the Christoffel symbols are zero for all the values of indices except:

$$\begin{aligned}\Gamma^{k}_{22} &= -\rho \\ \Gamma^{2}_{12} &= \Gamma^{2}_{21} = \frac{1}{\rho}\end{aligned} \tag{11.43}$$

where $(1,2,3)$ stand for (ρ,ϕ,z). ■

EXEMPLO 11.7 For spherical coordinate systems (r,θ,ϕ), the Christoffel symbols can be computed from

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\varphi^2$$

We can easily then see that the metric tensor and the inverse metric tensor are:

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}$$

$$g^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & r^{-2}\sin^{-2}\theta \end{pmatrix}$$

Using the formula:

$$\Gamma^{m}_{ij} = \frac{1}{2} g^{ml}(\partial_j g_{il} + \partial_i g_{lj} - \partial_l g_{ji})$$

Where upper indices indicate the inverse matrix. And so:

$$\Gamma^{1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -r & 0 \\ 0 & 0 & -r\sin^2\theta \end{pmatrix}$$

$$\Gamma^{2} = \begin{pmatrix} 0 & \frac{1}{r} & 0 \\ \frac{1}{r} & 0 & 0 \\ 0 & 0 & -\sin\theta\cos\theta \end{pmatrix}$$

$$\Gamma^3 = \begin{pmatrix} 0 & 0 & \frac{1}{r} \\ 0 & 0 & \cot\theta \\ \frac{1}{r} & \cot\theta & 0 \end{pmatrix}$$

■

Theorem 11.5.1 Under a change of variable from $(y^1, \ldots, y^n)$ to $(x^1, \ldots, x^n)$, the Christoffel symbol transform as

$$\bar{\Gamma}^k{}_{ij} = \frac{\partial x^p}{\partial y^i}\frac{\partial x^q}{\partial y^j}\Gamma^r{}_{pq}\frac{\partial y^k}{\partial x^r} + \frac{\partial y^k}{\partial x^m}\frac{\partial^2 x^m}{\partial y^i \partial y^j}$$

where the overline denotes the Christoffel symbols in the y coordinate system.

Definition 11.5.3 — Derivatives of Tensors in Coordinates.

- For a differentiable scalar f the covariant derivative is the same as the normal partial derivative, that is:

$$f_{;i} = f_{,i} = \partial_i f \tag{11.44}$$

This is justified by the fact that the covariant derivative is different from the normal partial derivative because the basis vectors in general coordinate systems are dependent on their spatial position, and since a scalar is independent of the basis vectors the covariant and partial derivatives are identical.

- For a differentiable vector $\mathbf{A}$ the covariant derivative is:

$$\begin{aligned} A_{j;i} &= \partial_i A_j - \Gamma^k_{ji} A_k && \text{(covariant)} \\ A^j_{;i} &= \partial_i A^j + \Gamma^j_{ki} A^k && \text{(contravariant)} \end{aligned} \tag{11.45}$$

- For a differentiable rank-2 tensor $\mathbf{A}$ the covariant derivative is:

$$\begin{aligned} A_{jk;i} &= \partial_i A_{jk} - \Gamma^l_{ji} A_{lk} - \Gamma^l_{ki} A_{jl} && \text{(covariant)} \\ A^{jk}_{;i} &= \partial_i A^{jk} + \Gamma^j_{li} A^{lk} + \Gamma^k_{li} A^{jl} && \text{(contravariant)} \\ A^k_{j;i} &= \partial_i A^k_j + \Gamma^k_{li} A^k_{jl - \Gamma^l_{ji} A_l} && \text{(mixed)} \end{aligned} \tag{11.46}$$

- For a differentiable rank-n tensor $\mathbf{A}$ the covariant derivative is:

$$\begin{aligned} A^{ij\ldots k}_{lm\ldots p;q} = {} & \partial_q A^{ij\ldots k}_{lm\ldots p} + \Gamma^i_{aq} A^{aj\ldots k}_{lm\ldots p} \Gamma^j_{aq} A^{ia\ldots k}_{lm\ldots p} + \cdots + \Gamma^k_{aq} A^{ij\ldots a}_{lm\ldots p} \\ & - \Gamma^a_{lq} A^{ij\ldots k}_{am\ldots p} - \Gamma^a_{mq} A^{ij\ldots k}_{la\ldots p} - \cdots - \Gamma^a_{pq} A^{ij\ldots k}_{lm\ldots a} \end{aligned} \tag{11.47}$$

Since the Christoffel symbols are identically zero in Cartesian coordinate systems, the covariant derivative is the same as the normal partial derivative for all tensor ranks.

The covariant derivative of the metric tensor is zero in all coordinate systems.

Several rules of normal differentiation similarly apply to covariant differentiation. For example, covariant differentiation is a linear operation with respect to algebraic sums of tensor terms:

$$\partial_{;i} (a\mathbf{A} \pm b\mathbf{B}) = a\partial_{;i}\mathbf{A} \pm b\partial_{;i}\mathbf{B} \tag{11.48}$$

where a and b are scalar constants and $\mathbf{A}$ and $\mathbf{B}$ are differentiable tensor fields. The product rule of normal differentiation also applies to covariant differentiation of tensor multiplication:

$$\partial_{;i} (\mathbf{AB}) = \left(\partial_{;i}\mathbf{A}\right)\mathbf{B} + \mathbf{A}\partial_{;i}\mathbf{B} \tag{11.49}$$

This rule is also valid for the inner product of tensors because the inner product is an outer product operation followed by a contraction of indices, and covariant differentiation and contraction of indices commute.

The covariant derivative operator can bypass the raising/lowering index operator:

$$A^i = g_{ij} A^j \qquad \Longrightarrow \qquad \partial_{;m} A^i = g_{ij} \partial_{;m} A^j \tag{11.50}$$

and hence the metric behaves like a constant with respect to the covariant operator.

A principal difference between normal partial differentiation and covariant differentiation is that for successive differential operations the partial derivative operators do commute with each other (assuming certain continuity conditions) but the covariant operators do not commute, that is

$$\partial_i \partial_j = \partial_j \partial_i \qquad \text{but} \qquad \partial_{;i} \partial_{;j} \neq \partial_{;j} \partial_{;i} \tag{11.51}$$

Higher order covariant derivatives are similarly defined as derivatives of derivatives; however the order of differentiation should be respected (refer to the previous point).

11.6 Geodesics and The Euler-Lagrange Equations

Given the metric tensor g in some domain $U \subset \mathbb{R}^n$, the length of a continuously differentiable curve $\gamma : [a, b] \to \mathbb{R}^n$ is defined by

$$L(\gamma) = \int_a^b \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}\, dt.$$

In coordinates if $\gamma(t) = (x^1, \dots x^n)$ then:

$$L(\gamma) = \int_a^b \sqrt{-g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}}\, dt$$

The distance $d(p, q)$ between two points p and q is defined as the infimum of the length taken over all continuous, piecewise continuously differentiable curves $\gamma : [a, b] \to \mathbb{R}^n$ such that $\gamma(a) = p$ and $\gamma(b) = q$. The **geodesics** are then defined as the locally distance-minimizing paths.

So the geodesics are the curve $y(x)$ such that the functional

$$L(\gamma) = \int_a^b \sqrt{g_{\gamma(x)}(\dot{\gamma}(x), \dot{\gamma}(x))}\, dx.$$

is minimized over all smooth (or piecewise smooth) functions $y(x)$ such that $x(a) = p$ and $x(b) = q$.

This problem can be simplified, if we introduce the energy functional

$$E(\gamma) = \frac{1}{2} \int_a^b g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))\, dt.$$

For a piecewise C^1 curve, the Cauchy–Schwarz inequality gives

$$L(\gamma)^2 \leqslant 2(b-a)E(\gamma)$$

with equality if and only if

$$g(\gamma',\gamma')$$

is constant.

Hence the minimizers of $E(\gamma)$ also minimize $L(\gamma)$.

The previous problem is an example of calculus of variations is concerned with the extrema of functionals. The fundamental problem of the calculus of variations is to find a function $x(t)$ such that the functional

$$I(x) = \int_a^b f(t, x(t), y'(t))\, dt$$

is minimized over all smooth (or piecewise smooth) functions $x(t)$ satisfying certain boundary conditions—for example, $x(a) = A$ and $x(b) = B$.

If $\hat{x}(t)$ is the smooth function at which the desired minimum of $I(x)$ occurs, and if $I(\hat{x}(t) + \varepsilon\eta(t))$ is defined for some arbitrary smooth function $eta(x)$ with $\eta(a) = 0$ and $\eta(b) = 0$, for small enough ε, then

$$I(\hat{x} + \varepsilon\eta) = \int_a^b f(t, \hat{x} + \varepsilon\eta, \hat{x}' + \varepsilon\eta')\, dt$$

is now a function of ε, which must have a minimum at $\varepsilon = 0$. In that case, if $I(\varepsilon)$ is smooth enough, we must have

$$\frac{dI}{d\varepsilon}|_{\varepsilon=0} = \int_a^b f_x(t, \hat{x}, \hat{x}')\eta(t) + f_{x'}(t, \hat{x}, \hat{x}')\eta'(t)\, dt = 0\,.$$

If we integrate the second term by parts we get, using $\eta(a) = 0$ and $\eta(b) = 0$,

$$\int_a^b \left(f_x(t, \hat{x}, \hat{x}') - \frac{d}{dt} f_{x'}(t, \hat{x}, \hat{x}') \right) \eta(t)\, dt = 0\,.$$

One can then argue that since $\eta(t)$ was arbitrary and $\hat{x}$ is smooth, we must have the quantity in brackets identically zero. This gives the *Euler-Lagrange* equations:

$$\frac{\partial}{\partial x} f(t, x, x') - \frac{d}{dt}\frac{\partial}{\partial x'} f(t, x, x') = 0\,. \tag{11.52}$$

In general this gives a second-order ordinary differential equation which can be solved to obtain the extremal function $f(x)$. We remark that the Euler–Lagrange equation is a necessary, but not a sufficient, condition for an extremum.

This can be generalized to many variables: Given the functional:

$$I(x) = \int_a^b f(t, x^1(t), x'^1(t), \ldots, x^n(t), x'^n(t))\, dt$$

We have the corresponding Euler-Lagrange equations:

$$\frac{\partial}{\partial x^k} f(t, x^1(t), x'^1(t), \ldots, x^n(t), x'^n(t)) - \frac{d}{dt}\frac{\partial}{\partial x'^k}(t, x^1(t), x'^1(t), \ldots, x^n(t), x'^n(t)) = 0. \tag{11.53}$$

Theorem 11.6.1 A necessary condition to a curve γ be a geodesic is

$$\frac{d^2\gamma^\lambda}{dt^2} + \Gamma^\lambda_{\mu\nu}\frac{d\gamma^\mu}{dt}\frac{d\gamma^\nu}{dt} = 0$$

Proof: The geodesics are the minimum of the functional

$$L(\gamma) = \int_a^b \sqrt{g_{\gamma(x)}(\dot\gamma(x), \dot\gamma(x))}\, dx.$$

Let

$$E = \frac{1}{2} g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}$$

We will write the Euler Lagrange equations.

$$\frac{d}{d\lambda}\frac{\partial L}{\partial(dx^\mu/d\lambda)} = \frac{\partial L}{\partial x^\mu}$$

Developing the right hand side we have:

$$\frac{\partial E}{\partial x^\lambda} = \frac{1}{2}\partial_\lambda g_{\mu\nu}\dot x^\mu \dot x^\nu$$

The first derivative on the left hand side is

$$\frac{\partial L}{\partial \dot x^\lambda} = g_{\mu\lambda}(x(\lambda))\dot x^\mu$$

where we have made the dependence of g on λ clear for the next step. Now we differentiate with respect to the curve parameter:

$$\frac{d}{d\lambda}[g_{\mu\lambda}(x(\lambda))\dot x^\mu] = \partial_\nu g_{\mu\lambda}\dot x^\mu \dot x^\nu + g_{\mu\lambda}\ddot x^\mu = \frac{1}{2}\partial_\nu g_{\mu\lambda}\dot x^\mu \dot x^\nu + \frac{1}{2}\partial_\mu g_{\nu\lambda}\dot x^\mu \dot x^\nu + g_{\mu\lambda}\ddot x^\mu$$

Putting it all together, we obtain

$$g_{\mu\lambda}\ddot x^\mu = -\frac{1}{2}\left(\partial_\nu g_{\mu\lambda} + \partial_\mu g_{\nu\lambda} - \partial_\lambda g_{\mu\nu}\right)\dot x^\mu \dot x^\nu = -\Gamma_{\lambda\mu\nu}\dot x^\mu \dot x^\nu$$

where in the last step we used the definition of the Christoffel symbols with three lower indices. Now contract with the inverse metric to raise the first index and cancel the metric on the left hand side. So

$$\ddot x^\lambda = -\Gamma^\lambda{}_{\mu\nu}\dot x^\mu \dot x^\nu$$

■

V

Applications of Tensor Calculus

CHAPTER 12

Applications of Tensor

12.1 Common Definitions in Tensor Notation

The trace of a matrix $\mathbf{A}$ representing a rank-2 tensor is:

$$\operatorname{tr}(\mathbf{A}) = A_{ii} \tag{12.1}$$

For a 3×3 matrix representing a rank-2 tensor in a 3D space, the determinant is:

$$\det(\mathbf{A}) = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \epsilon_{ijk} A_{1i} A_{2j} A_{3k} = \epsilon_{ijk} A_{i1} A_{j2} A_{k3} \tag{12.2}$$

where the last two equalities represent the expansion of the determinant by row and by column. Alternatively

$$\det(\mathbf{A}) = \frac{1}{3!} \epsilon_{ijk} \epsilon_{lmn} A_{il} A_{jm} A_{kn} \tag{12.3}$$

For an $n \times n$ matrix representing a rank-2 tensor in $\mathbb{R}^n$, the determinant is:

$$\det(\mathbf{A}) = \epsilon_{i_1 \cdots i_n} A_{1 i_1} \ldots A_{n i_n} = \epsilon_{i_1 \cdots i_n} A_{i_1 1} \ldots A_{i_n n} = \frac{1}{n!} \epsilon_{i_1 \cdots i_n} \epsilon_{j_1 \cdots j_n} A_{i_1 j_1} \ldots A_{i_n j_n} \tag{12.4}$$

The inverse of a matrix $\mathbf{A}$ representing a rank-2 tensor is:

$$\left[\mathbf{A}^{-1}\right]_{ij} = \frac{1}{2 \det(\mathbf{A})} \epsilon_{jmn} \epsilon_{ipq} A_{mp} A_{nq} \tag{12.5}$$

The multiplication of a matrix $\mathbf{A}$ by a vector $\mathbf{b}$ as defined in linear algebra is:

$$[\mathbf{A}\mathbf{b}]_i = A_{ij} b_j \tag{12.6}$$

It should be noticed that here we are using matrix notation. The multiplication operation, according to the symbolic notation of tensors, should be denoted by a dot between the tensor and the vector, i.e. $\mathbf{A}\cdot\mathbf{b}$.[1]

The multiplication of two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ as defined in linear algebra is:

$$[\mathbf{AB}]_{ik} = A_{ij}B_{jk} \tag{12.7}$$

Again, here we are using matrix notation; otherwise a dot should be inserted between the two matrices.

The dot product of two vectors is:

$$\mathbf{A}\cdot\mathbf{B} = \delta_{ij}A_iB_j = A_iB_i \tag{12.8}$$

The readers are referred to S 10.3.5 for a more general definition of this type of product that includes higher rank tensors.

The cross product of two vectors is:

$$[\mathbf{A}\times\mathbf{B}]_i = \epsilon_{ijk}A_jB_k \tag{12.9}$$

The scalar triple product of three vectors is:

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} = \epsilon_{ijk}A_iB_jC_k \tag{12.10}$$

The vector triple product of three vectors is:

$$\left[\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\right]_i = \epsilon_{ijk}\epsilon_{klm}A_jB_lC_m \tag{12.11}$$

12.2 Common Differential Operations in Tensor Notation

Here we present the most common differential operations as defined by tensor notation. These operations are mostly based on the various types of interaction between the vector differential operator nabla ∇ with tensors of different ranks as well as interaction with other types of operation like dot and cross products.

The operator ∇ is essentially a spatial partial differential operator defined in Cartesian coordinate systems by:

$$\nabla_i = \frac{\partial}{\partial x_i} \tag{12.12}$$

[1]The matrix multiplication in matrix notation is equivalent to a dot product operation in tensor notation.

The gradient of a differentiable scalar function of position f is a vector given by:

$$[\nabla f]_i = \nabla_i f = \frac{\partial f}{\partial x_i} = \partial_i f = f_{,i} \tag{12.13}$$

The gradient of a differentiable vector function of position $\mathbf{A}$ (which is the outer product, as defined in S 10.3.3, between the ∇ operator and the vector) is a rank-2 tensor defined by:

$$[\nabla \mathbf{A}]_{ij} = \partial_i A_j \tag{12.14}$$

The gradient operation is distributive but not commutative or associative:

$$\nabla (f + h) = \nabla f + \nabla h \tag{12.15}$$

$$\nabla f \neq f \nabla \tag{12.16}$$

$$(\nabla f) h \neq \nabla (fh) \tag{12.17}$$

where f and h are differentiable scalar functions of position.

The divergence of a differentiable vector $\mathbf{A}$ is a scalar given by:

$$\nabla \cdot \mathbf{A} = \delta_{ij} \frac{\partial A_i}{\partial x_j} = \frac{\partial A_i}{\partial x_i} = \nabla_i A_i = \partial_i A_i = A_{i,i} \tag{12.18}$$

The divergence operation can also be viewed as taking the gradient of the vector followed by a contraction. Hence, the divergence of a vector is invariant because it is the trace of a rank-2 tensor.[2]

The divergence of a differentiable rank-2 tensor $\mathbf{A}$ is a vector defined in one of its forms by:

$$[\nabla \cdot \mathbf{A}]_i = \partial_j A_{ji} \tag{12.19}$$

and in another form by

$$[\nabla \cdot \mathbf{A}]_j = \partial_i A_{ji} \tag{12.20}$$

These two different forms can be given, respectively, in symbolic notation by:

$$\nabla \cdot \mathbf{A} \quad \& \quad \nabla \cdot \mathbf{A}^{\mathsf{T}} \tag{12.21}$$

where $\mathbf{A}^{\mathsf{T}}$ is the transpose of $\mathbf{A}$. More generally, the divergence of a tensor of rank $n \geqslant 2$, which is a tensor of rank-$(n-1)$, can be defined in several forms, which are different in general, depending on the combination of the contracted indices.

The divergence operation is distributive but not commutative or associative:

$$\nabla \cdot (\mathbf{A} + \mathbf{B}) = \nabla \cdot \mathbf{A} + \nabla \cdot \mathbf{B} \tag{12.22}$$

$$\nabla \cdot \mathbf{A} \neq \mathbf{A} \cdot \nabla \tag{12.23}$$

[2] It may also be argued that the divergence of a vector is a scalar and hence it is invariant.

$$\nabla \cdot (f\mathbf{A}) \neq \nabla f \cdot \mathbf{A} \tag{12.24}$$

where **A** and **B** are differentiable tensor functions of position.

The curl of a differentiable vector **A** is a vector given by:

$$[\nabla \times \mathbf{A}]_i = \epsilon_{ijk} \frac{\partial A_k}{\partial x_j} = \epsilon_{ijk} \nabla_j A_k = \epsilon_{ijk} \partial_j A_k = \epsilon_{ijk} A_{k,j} \tag{12.25}$$

The curl operation may be generalized to tensors of rank > 1, and hence the curl of a differentiable rank-2 tensor **A** can be defined as a rank-2 tensor given by:

$$[\nabla \times \mathbf{A}]_{ij} = \epsilon_{imn} \partial_m A_{nj} \tag{12.26}$$

The curl operation is distributive but not commutative or associative:

$$\nabla \times (\mathbf{A} + \mathbf{B}) = \nabla \times \mathbf{A} + \nabla \times \mathbf{B} \tag{12.27}$$

$$\nabla \times \mathbf{A} \neq \mathbf{A} \times \nabla \tag{12.28}$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) \neq (\nabla \times \mathbf{A}) \times \mathbf{B} \tag{12.29}$$

The Laplacian scalar operator, also called the harmonic operator, acting on a differentiable scalar f is given by:

$$\Delta f = \nabla^2 f = \delta_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_i} = \nabla_{ii} f = \partial_{ii} f = f_{,ii} \tag{12.30}$$

The Laplacian operator acting on a differentiable vector **A** is defined for each component of the vector similar to the definition of the Laplacian acting on a scalar, that is

$$\left[\nabla^2 \mathbf{A}\right]_i = \partial_{jj} A_i \tag{12.31}$$

The following scalar differential operator is commonly used in science (e.g. in fluid dynamics):

$$\mathbf{A} \cdot \nabla = A_i \nabla_i = A_i \frac{\partial}{\partial x_i} = A_i \partial_i \tag{12.32}$$

where **A** is a vector. As indicated earlier, the order of A_i and ∂_i should be respected.

The following vector differential operator also has common applications in science:

$$[\mathbf{A} \times \nabla]_i = \epsilon_{ijk} A_j \partial_k \tag{12.33}$$

The differentiation of a tensor increases its rank by one, by introducing an extra covariant index, unless it implies a contraction in which case it reduces the rank by one. Therefore the gradient of a scalar is a vector and the gradient of a vector is a rank-2 tensor ($\partial_i A_j$), while the divergence of a vector is a scalar and the divergence of a rank-2 tensor is a vector ($\partial_j A_{ji}$ or $\partial_i A_{ji}$). This may be justified by the fact that ∇ is a vector operator. On the other hand the Laplacian operator does not change the rank since it is a scalar operator; hence the Laplacian of a scalar is a scalar and the Laplacian of a vector is a vector.

12.3 Common Identities in Vector and Tensor Notation

Here we present some of the widely used identities of vector calculus in the traditional vector notation and in its equivalent tensor notation. In the following bullet points, f and h are differentiable scalar fields; $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are differentiable vector fields; and $\mathbf{r} = x_i \mathbf{e}_i$ is the position vector.

$$\begin{aligned} \nabla \cdot \mathbf{r} &= n \\ &\Updownarrow \\ \partial_i x_i &= n \end{aligned} \tag{12.34}$$

where n is the space dimension.

$$\begin{aligned} \nabla \times \mathbf{r} &= \mathbf{0} \\ &\Updownarrow \\ \epsilon_{ijk} \partial_j x_k &= 0 \end{aligned} \tag{12.35}$$

$$\begin{aligned} \nabla (\mathbf{a} \cdot \mathbf{r}) &= \mathbf{a} \\ &\Updownarrow \\ \partial_i (a_j x_j) &= a_i \end{aligned} \tag{12.36}$$

where $\mathbf{a}$ is a constant vector.

$$\begin{aligned} \nabla \cdot (\nabla f) &= \nabla^2 f \\ &\Updownarrow \\ \partial_i (\partial_i f) &= \partial_{ii} f \end{aligned} \tag{12.37}$$

$$\begin{aligned} \nabla \cdot (\nabla \times \mathbf{A}) &= 0 \\ &\Updownarrow \\ \epsilon_{ijk} \partial_i \partial_j A_k &= 0 \end{aligned} \tag{12.38}$$

$$\begin{aligned} \nabla \times (\nabla f) &= \mathbf{0} \\ &\Updownarrow \\ \epsilon_{ijk} \partial_j \partial_k f &= 0 \end{aligned} \tag{12.39}$$

$$\begin{aligned} \nabla(fh) &= f\nabla h + h\nabla f \\ &\Updownarrow \\ \partial_i(fh) &= f\partial_i h + h\partial_i f \end{aligned} \tag{12.40}$$

$$\begin{aligned} \nabla\cdot(f\mathbf{A}) &= f\nabla\cdot\mathbf{A} + \mathbf{A}\cdot\nabla f \\ &\Updownarrow \\ \partial_i(fA_i) &= f\partial_i A_i + A_i\partial_i f \end{aligned} \tag{12.41}$$

$$\begin{aligned} \nabla\times(f\mathbf{A}) &= f\nabla\times\mathbf{A} + \nabla f\times\mathbf{A} \\ &\Updownarrow \\ \epsilon_{ijk}\partial_j(fA_k) &= f\epsilon_{ijk}\partial_j A_k + \epsilon_{ijk}\left(\partial_j f\right)A_k \end{aligned} \tag{12.42}$$

$$\begin{array}{ccccc} \mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = & \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) = & \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) \\ \Updownarrow & \Updownarrow & \\ \epsilon_{ijk}A_iB_jC_k = & \epsilon_{kij}C_kA_iB_j = & \epsilon_{jki}B_jC_kA_i \end{array} \tag{12.43}$$

$$\begin{aligned} \mathbf{A}\times(\mathbf{B}\times\mathbf{C}) &= \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B}) \\ &\Updownarrow \\ \epsilon_{ijk}A_j\epsilon_{klm}B_lC_m &= B_i(A_mC_m) - C_i(A_lB_l) \end{aligned} \tag{12.44}$$

$$\begin{aligned} \mathbf{A}\times(\nabla\times\mathbf{B}) &= (\nabla\mathbf{B})\cdot\mathbf{A} - \mathbf{A}\cdot\nabla\mathbf{B} \\ &\Updownarrow \\ \epsilon_{ijk}\epsilon_{klm}A_j\partial_l B_m &= (\partial_i B_m)A_m - A_l(\partial_l B_i) \end{aligned} \tag{12.45}$$

$$\begin{aligned} \nabla\times(\nabla\times\mathbf{A}) &= \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A} \\ &\Updownarrow \\ \epsilon_{ijk}\epsilon_{klm}\partial_j\partial_l A_m &= \partial_i(\partial_m A_m) - \partial_{ll}A_i \end{aligned} \tag{12.46}$$

$$\begin{aligned}
\nabla\left(\mathbf{A}\cdot\mathbf{B}\right) &= \mathbf{A}\times\left(\nabla\times\mathbf{B}\right)+\mathbf{B}\times\left(\nabla\times\mathbf{A}\right)+\left(\mathbf{A}\cdot\nabla\right)\mathbf{B}+\left(\mathbf{B}\cdot\nabla\right)\mathbf{A} \\
&\Updownarrow \\
\partial_i\left(A_m B_m\right) &= \epsilon_{ijk}A_j\left(\epsilon_{klm}\partial_l B_m\right)+\epsilon_{ijk}B_j\left(\epsilon_{klm}\partial_l A_m\right)+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i
\end{aligned}\tag{12.47}$$

$$\begin{aligned}
\nabla\cdot\left(\mathbf{A}\times\mathbf{B}\right) &= \mathbf{B}\cdot\left(\nabla\times\mathbf{A}\right)-\mathbf{A}\cdot\left(\nabla\times\mathbf{B}\right) \\
&\Updownarrow \\
\partial_i\left(\epsilon_{ijk}A_jB_k\right) &= B_k\left(\epsilon_{kij}\partial_iA_j\right)-A_j\left(\epsilon_{jik}\partial_iB_k\right)
\end{aligned}\tag{12.48}$$

$$\begin{aligned}
\nabla\times\left(\mathbf{A}\times\mathbf{B}\right) &= \left(\mathbf{B}\cdot\nabla\right)\mathbf{A}+\left(\nabla\cdot\mathbf{B}\right)\mathbf{A}-\left(\nabla\cdot\mathbf{A}\right)\mathbf{B}-\left(\mathbf{A}\cdot\nabla\right)\mathbf{B} \\
&\Updownarrow \\
\epsilon_{ijk}\epsilon_{klm}\partial_j\left(A_lB_m\right) &= \left(B_m\partial_m\right)A_i+\left(\partial_mB_m\right)A_i-\left(\partial_jA_j\right)B_i-\left(A_j\partial_j\right)B_i
\end{aligned}\tag{12.49}$$

$$\begin{aligned}
\left(\mathbf{A}\times\mathbf{B}\right)\cdot\left(\mathbf{C}\times\mathbf{D}\right) &= \begin{vmatrix} \mathbf{A}\cdot\mathbf{C} & \mathbf{A}\cdot\mathbf{D} \\ \mathbf{B}\cdot\mathbf{C} & \mathbf{B}\cdot\mathbf{D} \end{vmatrix} \\
&\Updownarrow \\
\epsilon_{ijk}A_jB_k\epsilon_{ilm}C_lD_m &= \left(A_lC_l\right)\left(B_mD_m\right)-\left(A_mD_m\right)\left(B_lC_l\right)
\end{aligned}\tag{12.50}$$

$$\begin{aligned}
\left(\mathbf{A}\times\mathbf{B}\right)\times\left(\mathbf{C}\times\mathbf{D}\right) &= \left[\mathbf{D}\cdot\left(\mathbf{A}\times\mathbf{B}\right)\right]\mathbf{C}-\left[\mathbf{C}\cdot\left(\mathbf{A}\times\mathbf{B}\right)\right]\mathbf{D} \\
&\Updownarrow \\
\epsilon_{ijk}\epsilon_{jmn}A_mB_n\epsilon_{kpq}C_pD_q &= \left(\epsilon_{qmn}D_qA_mB_n\right)C_i-\left(\epsilon_{pmn}C_pA_mB_n\right)D_i
\end{aligned}\tag{12.51}$$

In vector and tensor notations, the condition for a vector field **A** to be solenoidal is:

$$\begin{aligned}
\nabla\cdot\mathbf{A} &= 0 \\
&\Updownarrow \\
\partial_iA_i &= 0
\end{aligned}\tag{12.52}$$

In vector and tensor notations, the condition for a vector field **A** to be irrotational is:

$$\begin{aligned}
\nabla\times\mathbf{A} &= \mathbf{0} \\
&\Updownarrow \\
\epsilon_{ijk}\partial_jA_k &= 0
\end{aligned}\tag{12.53}$$

12.4 Integral Theorems in Tensor Notation

The divergence theorem for a differentiable vector field $\mathbf{A}$ in vector and tensor notation is:

$$\begin{aligned}
\iiint_V \nabla\cdot\mathbf{A}\,d\tau &= \iint_S \mathbf{A}\cdot\mathbf{n}\,d\sigma \\
&\Updownarrow \\
\int_V \partial_i A_i\,d\tau &= \int_S A_i n_i\,d\sigma
\end{aligned} \tag{12.54}$$

where V is a bounded region in an nD space enclosed by a generalized surface S, $d\tau$ and $d\sigma$ are generalized volume and surface elements respectively, $\mathbf{n}$ and n_i are unit normal to the surface and its i^{th} component respectively, and the index i ranges over $1,\ldots,n$.

The divergence theorem for a differentiable rank-2 tensor field $\mathbf{A}$ in tensor notation for the first index is given by:

$$\int_V \partial_i A_{il}\,d\tau = \int_S A_{il} n_i\,d\sigma \tag{12.55}$$

The divergence theorem for differentiable tensor fields of higher ranks $\mathbf{A}$ in tensor notation for the index k is:

$$\int_V \partial_k A_{ij\ldots k\ldots m}\,d\tau = \int_S A_{ij\ldots k\ldots m} n_k\,d\sigma \tag{12.56}$$

Stokes theorem for a differentiable vector field $\mathbf{A}$ in vector and tensor notation is:

$$\begin{aligned}
\iint_S (\nabla\times\mathbf{A})\cdot\mathbf{n}\,d\sigma &= \int_C \mathbf{A}\cdot d\mathbf{r} \\
&\Updownarrow \\
\int_S \epsilon_{ijk}\partial_j A_k n_i\,d\sigma &= \int_C A_i\,dx_i
\end{aligned} \tag{12.57}$$

where C stands for the perimeter of the surface S and $d\mathbf{r}$ is the vector element tangent to the perimeter.

Stokes theorem for a differentiable rank-2 tensor field $\mathbf{A}$ in tensor notation for the first index is:

$$\int_S \epsilon_{ijk}\partial_j A_{kl} n_i\,d\sigma = \int_C A_{il}\,dx_i \tag{12.58}$$

Stokes theorem for differentiable tensor fields of higher ranks $\mathbf{A}$ in tensor notation for the index k is:

$$\int_S \epsilon_{ijk}\partial_j A_{lm\ldots k\ldots n} n_i\,d\sigma = \int_C A_{lm\ldots k\ldots n}\,dx_k \tag{12.59}$$

12.5 Examples of Using Tensor Techniques to Prove Identities

$\nabla\cdot\mathbf{r} = n$:

$$\begin{aligned}
\nabla\cdot\mathbf{r} &= \partial_i x_i && \text{(Eq. 12.18)} \\
&= \delta_{ii} && \text{(Eq. 10.37)} \\
&= n && \text{(Eq. 10.37)}
\end{aligned} \tag{12.60}$$

$\nabla \times \mathbf{r} = \mathbf{0}$:

$$\begin{aligned}
[\nabla \times \mathbf{r}]_i &= \epsilon_{ijk} \partial_j x_k && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk} \delta_{kj} && \text{(Eq. 10.36)} \\
&= \epsilon_{ijj} && \text{(Eq. 10.33)} \\
&= 0 && \text{(Eq. 10.28)}
\end{aligned} \tag{12.61}$$

Since i is a free index the identity is proved for all components.

$\nabla (\mathbf{a} \cdot \mathbf{r}) = \mathbf{a}$:

$$\begin{aligned}
\left[\nabla (\mathbf{a} \cdot \mathbf{r})\right]_i &= \partial_i \left(a_j x_j\right) && \text{(Eqs. 12.13 \& 12.8)} \\
&= a_j \partial_i x_j + x_j \partial_i a_j && \text{(product rule)} \\
&= a_j \partial_i x_j && (a_j \text{ is constant}) \\
&= a_j \delta_{ji} && \text{(Eq. 10.36)} \\
&= a_i && \text{(Eq. 10.33)} \\
&= [\mathbf{a}]_i && \text{(definition of index)}
\end{aligned} \tag{12.62}$$

Since i is a free index the identity is proved for all components.

$\nabla \cdot (\nabla f) = \nabla^2 f$:

$$\begin{aligned}
\nabla \cdot (\nabla f) &= \partial_i [\nabla f]_i && \text{(Eq. 12.18)} \\
&= \partial_i (\partial_i f) && \text{(Eq. 12.13)} \\
&= \partial_i \partial_i f && \text{(rules of differentiation)} \\
&= \partial_{ii} f && \text{(definition of 2nd derivative)} \\
&= \nabla^2 f && \text{(Eq. 12.30)}
\end{aligned} \tag{12.63}$$

$\nabla \cdot (\nabla \times \mathbf{A}) = 0$:

$$\begin{aligned}
\nabla \cdot (\nabla \times \mathbf{A}) &= \partial_i [\nabla \times \mathbf{A}]_i && \text{(Eq. 12.18)} \\
&= \partial_i \left(\epsilon_{ijk} \partial_j A_k\right) && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk} \partial_i \partial_j A_k && (\partial \text{ not acting on } \epsilon) \\
&= \epsilon_{ijk} \partial_j \partial_i A_k && \text{(continuity condition)} \\
&= -\epsilon_{jik} \partial_j \partial_i A_k && \text{(Eq. 10.41)} \\
&= -\epsilon_{ijk} \partial_i \partial_j A_k && \text{(relabeling dummy indices } i \text{ and } j) \\
&= 0 && (\text{since } \epsilon_{ijk} \partial_i \partial_j A_k = -\epsilon_{ijk} \partial_i \partial_j A_k)
\end{aligned} \tag{12.64}$$

This can also be concluded from line three by arguing that: since by the continuity condition ∂_i and ∂_j can change their order with no change in the value of the term while a corresponding change of the order of i and j in ϵ_{ijk} results in a sign change, we see that each term in the sum has its own negative and hence the terms add up to zero (see Eq. 10.51).

$\nabla \times (\nabla f) = \mathbf{0}$:

$$\begin{aligned}
\left[\nabla \times (\nabla f)\right]_i &= \epsilon_{ijk}\partial_j \left[\nabla f\right]_k && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk}\partial_j \left(\partial_k f\right) && \text{(Eq. 12.13)} \\
&= \epsilon_{ijk}\partial_j \partial_k f && \text{(rules of differentiation)} \\
&= \epsilon_{ijk}\partial_k \partial_j f && \text{(continuity condition)} \\
&= -\epsilon_{ikj}\partial_k \partial_j f && \text{(Eq. 10.41)} \\
&= -\epsilon_{ijk}\partial_j \partial_k f && \text{(relabeling dummy indices } j \text{ and } k\text{)} \\
&= 0 && \text{(since } \epsilon_{ijk}\partial_j\partial_k f = -\epsilon_{ijk}\partial_j\partial_k f\text{)}
\end{aligned} \tag{12.65}$$

This can also be concluded from line three by a similar argument to the one given in the previous point. Because $\left[\nabla \times (\nabla f)\right]_i$ is an arbitrary component, then each component is zero.

$\nabla (fh) = f\nabla h + h\nabla f$:

$$\begin{aligned}
\left[\nabla (fh)\right]_i &= \partial_i (fh) && \text{(Eq. 12.13)} \\
&= f\partial_i h + h\partial_i f && \text{(product rule)} \\
&= [f\nabla h]_i + [h\nabla f]_i && \text{(Eq. 12.13)} \\
&= [f\nabla h + h\nabla f]_i && \text{(Eq. \textbf{??})}
\end{aligned} \tag{12.66}$$

Because i is a free index the identity is proved for all components.

$\nabla \cdot (f\mathbf{A}) = f\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla f$:

$$\begin{aligned}
\nabla \cdot (f\mathbf{A}) &= \partial_i [f\mathbf{A}]_i && \text{(Eq. 12.18)} \\
&= \partial_i (fA_i) && \text{(definition of index)} \\
&= f\partial_i A_i + A_i\partial_i f && \text{(product rule)} \\
&= f\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla f && \text{(Eqs. 12.18 \& 12.32)}
\end{aligned} \tag{12.67}$$

$\nabla \times (f\mathbf{A}) = f\nabla \times \mathbf{A} + \nabla f \times \mathbf{A}$:

$$\begin{aligned}
\left[\nabla \times (f\mathbf{A})\right]_i &= \epsilon_{ijk}\partial_j [f\mathbf{A}]_k && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk}\partial_j (fA_k) && \text{(definition of index)} \\
&= f\epsilon_{ijk}\partial_j A_k + \epsilon_{ijk}\left(\partial_j f\right) A_k && \text{(product rule \& commutativity)} \\
&= f\epsilon_{ijk}\partial_j A_k + \epsilon_{ijk}[\nabla f]_j A_k && \text{(Eq. 12.13)} \\
&= [f\nabla \times \mathbf{A}]_i + [\nabla f \times \mathbf{A}]_i && \text{(Eqs. 12.25 \& \textbf{??})} \\
&= [f\nabla \times \mathbf{A} + \nabla f \times \mathbf{A}]_i && \text{(Eq. \textbf{??})}
\end{aligned} \tag{12.68}$$

Because i is a free index the identity is proved for all components.

$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A})$:

$$\begin{aligned}
\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) &= \epsilon_{ijk}A_iB_jC_k && \text{(Eq. ??)}\\
&= \epsilon_{kij}A_iB_jC_k && \text{(Eq. 10.41)}\\
&= \epsilon_{kij}C_kA_iB_j && \text{(commutativity)}\\
&= \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) && \text{(Eq. ??)}\\
&= \epsilon_{jki}A_iB_jC_k && \text{(Eq. 10.41)}\\
&= \epsilon_{jki}B_jC_kA_i && \text{(commutativity)}\\
&= \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) && \text{(Eq. ??)}
\end{aligned} \tag{12.69}$$

The negative permutations of these identities can be similarly obtained and proved by changing the order of the vectors in the cross products which results in a sign change.

$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})$:

$$\begin{aligned}
\left[\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\right]_i &= \epsilon_{ijk}A_j\left[\mathbf{B}\times\mathbf{C}\right]_k && \text{(Eq. ??)}\\
&= \epsilon_{ijk}A_j\epsilon_{klm}B_lC_m && \text{(Eq. ??)}\\
&= \epsilon_{ijk}\epsilon_{klm}A_jB_lC_m && \text{(commutativity)}\\
&= \epsilon_{ijk}\epsilon_{lmk}A_jB_lC_m && \text{(Eq. 10.41)}\\
&= \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)A_jB_lC_m && \text{(Eq. 10.59)}\\
&= \delta_{il}\delta_{jm}A_jB_lC_m - \delta_{im}\delta_{jl}A_jB_lC_m && \text{(distributivity)}\\
&= (\delta_{il}B_l)\left(\delta_{jm}A_jC_m\right) - (\delta_{im}C_m)\left(\delta_{jl}A_jB_l\right) && \text{(commutativity and grouping)}\\
&= B_i(A_mC_m) - C_i(A_lB_l) && \text{(Eq. 10.33)}\\
&= B_i(\mathbf{A}\cdot\mathbf{C}) - C_i(\mathbf{A}\cdot\mathbf{B}) && \text{(Eq. 12.8)}\\
&= \left[\mathbf{B}(\mathbf{A}\cdot\mathbf{C})\right]_i - \left[\mathbf{C}(\mathbf{A}\cdot\mathbf{B})\right]_i && \text{(definition of index)}\\
&= \left[\mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})\right]_i && \text{(Eq. ??)}
\end{aligned} \tag{12.70}$$

Because i is a free index the identity is proved for all components. Other variants of this identity [e.g. $(\mathbf{A}\times\mathbf{B})\times\mathbf{C}$] can be obtained and proved similarly by changing the order of the factors in the external cross product with adding a minus sign.

$\mathbf{A} \times (\nabla \times \mathbf{B}) = (\nabla \mathbf{B}) \cdot \mathbf{A} - \mathbf{A} \cdot \nabla \mathbf{B}$:

$$\begin{aligned}
\left[\mathbf{A} \times (\nabla \times \mathbf{B})\right]_i &= \epsilon_{ijk} A_j \left[\nabla \times \mathbf{B}\right]_k && \text{(Eq. ??)} \\
&= \epsilon_{ijk} A_j \epsilon_{klm} \partial_l B_m && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk} \epsilon_{klm} A_j \partial_l B_m && \text{(commutativity)} \\
&= \epsilon_{ijk} \epsilon_{lmk} A_j \partial_l B_m && \text{(Eq. 10.41)} \\
&= \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right) A_j \partial_l B_m && \text{(Eq. 10.59)} \\
&= \delta_{il}\delta_{jm} A_j \partial_l B_m - \delta_{im}\delta_{jl} A_j \partial_l B_m && \text{(distributivity)} \\
&= A_m \partial_i B_m - A_l \partial_l B_i && \text{(Eq. 10.33)} \\
&= (\partial_i B_m) A_m - A_l (\partial_l B_i) && \text{(commutativity \& grouping)} \\
&= \left[(\nabla \mathbf{B}) \cdot \mathbf{A}\right]_i - \left[\mathbf{A} \cdot \nabla \mathbf{B}\right]_i && \\
&= \left[(\nabla \mathbf{B}) \cdot \mathbf{A} - \mathbf{A} \cdot \nabla \mathbf{B}\right]_i && \text{(Eq. ??)}
\end{aligned} \tag{12.71}$$

Because i is a free index the identity is proved for all components.

$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$:

$$\begin{aligned}
\left[\nabla \times (\nabla \times \mathbf{A})\right]_i &= \epsilon_{ijk} \partial_j \left[\nabla \times \mathbf{A}\right]_k && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk} \partial_j \left(\epsilon_{klm} \partial_l A_m\right) && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk} \epsilon_{klm} \partial_j \left(\partial_l A_m\right) && \text{(∂ not acting on ϵ)} \\
&= \epsilon_{ijk} \epsilon_{lmk} \partial_j \partial_l A_m && \text{(Eq. 10.41 \& definition of derivative)} \\
&= \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right) \partial_j \partial_l A_m && \text{(Eq. 10.59)} \\
&= \delta_{il}\delta_{jm} \partial_j \partial_l A_m - \delta_{im}\delta_{jl} \partial_j \partial_l A_m && \text{(distributivity)} \\
&= \partial_m \partial_i A_m - \partial_l \partial_l A_i && \text{(Eq. 10.33)} \\
&= \partial_i (\partial_m A_m) - \partial_{ll} A_i && \text{(∂ shift, grouping \& Eq. ??)} \\
&= \left[\nabla (\nabla \cdot \mathbf{A})\right]_i - \left[\nabla^2 \mathbf{A}\right]_i && \text{(Eqs. 12.18, 12.13 \& 12.31)} \\
&= \left[\nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}\right]_i && \text{(Eqs. ??)}
\end{aligned} \tag{12.72}$$

Because i is a free index the identity is proved for all components. This identity can also be considered as an instance of the identity before the last one, observing that in the second term on the right hand side the Laplacian should precede the vector, and hence no independent proof is required.

$\nabla (\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A}) + (\mathbf{A} \cdot \nabla) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{A}$:

We start from the right hand side and end with the left hand side

$$\begin{aligned}
\left[\mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A}) + (\mathbf{A} \cdot \nabla) \mathbf{B} + (\mathbf{B} \cdot \nabla) \mathbf{A}\right]_i &= \\
\left[\mathbf{A} \times (\nabla \times \mathbf{B})\right]_i + \left[\mathbf{B} \times (\nabla \times \mathbf{A})\right]_i + \left[(\mathbf{A} \cdot \nabla) \mathbf{B}\right]_i + \left[(\mathbf{B} \cdot \nabla) \mathbf{A}\right]_i &= && \text{(Eq. ??)} \\
\epsilon_{ijk} A_j \left[\nabla \times \mathbf{B}\right]_k + \epsilon_{ijk} B_j \left[\nabla \times \mathbf{A}\right]_k + (A_l \partial_l) B_i + (B_l \partial_l) A_i &= && \text{(Eqs. ??, 12.18 \& indexing)}
\end{aligned}$$

$$
\begin{aligned}
\epsilon_{ijk}A_j\left(\epsilon_{klm}\partial_l B_m\right)+\epsilon_{ijk}B_j\left(\epsilon_{klm}\partial_l A_m\right)+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(Eq. 12.25)}\\
\epsilon_{ijk}\epsilon_{klm}A_j\partial_l B_m+\epsilon_{ijk}\epsilon_{klm}B_j\partial_l A_m+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(commutativity)}\\
\epsilon_{ijk}\epsilon_{lmk}A_j\partial_l B_m+\epsilon_{ijk}\epsilon_{lmk}B_j\partial_l A_m+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(Eq. 10.41)}\\
\left(\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl}\right)A_j\partial_l B_m+\left(\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl}\right)B_j\partial_l A_m+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(Eq. 10.59)}\\
\left(\delta_{il}\delta_{jm}A_j\partial_l B_m-\delta_{im}\delta_{jl}A_j\partial_l B_m\right)+\left(\delta_{il}\delta_{jm}B_j\partial_l A_m-\delta_{im}\delta_{jl}B_j\partial_l A_m\right)+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(distributivity)}\\
\delta_{il}\delta_{jm}A_j\partial_l B_m-A_l\partial_l B_i+\delta_{il}\delta_{jm}B_j\partial_l A_m-B_l\partial_l A_i+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(Eq. 10.33)}\\
\delta_{il}\delta_{jm}A_j\partial_l B_m-\left(A_l\partial_l\right)B_i+\delta_{il}\delta_{jm}B_j\partial_l A_m-\left(B_l\partial_l\right)A_i+\left(A_l\partial_l\right)B_i+\left(B_l\partial_l\right)A_i &= && \text{(grouping)}\\
\delta_{il}\delta_{jm}A_j\partial_l B_m+\delta_{il}\delta_{jm}B_j\partial_l A_m &= && \text{(cancellation)}\\
A_m\partial_i B_m+B_m\partial_i A_m &= && \text{(Eq. 10.33)}\\
\partial_i\left(A_m B_m\right) &= && \text{(product rule)}\\
&= && \left[\nabla\left(\mathbf{A}\cdot\mathbf{B}\right)\right]_i \text{ (Eqs. 12.13 \& 12.18)}
\end{aligned}
\tag{12.73}
$$

Because i is a free index the identity is proved for all components.

$\nabla\cdot(\mathbf{A}\times\mathbf{B})=\mathbf{B}\cdot(\nabla\times\mathbf{A})-\mathbf{A}\cdot(\nabla\times\mathbf{B})$:

$$
\begin{aligned}
\nabla\cdot(\mathbf{A}\times\mathbf{B}) &= \partial_i\left[\mathbf{A}\times\mathbf{B}\right]_i && \text{(Eq. 12.18)}\\
&= \partial_i\left(\epsilon_{ijk}A_j B_k\right) && \text{(Eq. ??)}\\
&= \epsilon_{ijk}\partial_i\left(A_j B_k\right) && (\partial \text{ not acting on } \epsilon)\\
&= \epsilon_{ijk}\left(B_k\partial_i A_j+A_j\partial_i B_k\right) && \text{(product rule)}\\
&= \epsilon_{ijk}B_k\partial_i A_j+\epsilon_{ijk}A_j\partial_i B_k && \text{(distributivity)}\\
&= \epsilon_{kij}B_k\partial_i A_j-\epsilon_{jik}A_j\partial_i B_k && \text{(Eq. 10.41)}\\
&= B_k\left(\epsilon_{kij}\partial_i A_j\right)-A_j\left(\epsilon_{jik}\partial_i B_k\right) && \text{(commutativity \& grouping)}\\
&= B_k\left[\nabla\times\mathbf{A}\right]_k-A_j\left[\nabla\times\mathbf{B}\right]_j && \text{(Eq. 12.25)}\\
&= \mathbf{B}\cdot(\nabla\times\mathbf{A})-\mathbf{A}\cdot(\nabla\times\mathbf{B}) && \text{(Eq. 12.8)}
\end{aligned}
\tag{12.74}
$$

$\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} + (\nabla \cdot \mathbf{B})\mathbf{A} - (\nabla \cdot \mathbf{A})\mathbf{B} - (\mathbf{A} \cdot \nabla)\mathbf{B}$:

$$
\begin{aligned}
\left[\nabla \times (\mathbf{A} \times \mathbf{B})\right]_i &= \epsilon_{ijk}\partial_j [\mathbf{A} \times \mathbf{B}]_k && \text{(Eq. 12.25)} \\
&= \epsilon_{ijk}\partial_j (\epsilon_{klm}A_l B_m) && \text{(Eq. \textbf{??})} \\
&= \epsilon_{ijk}\epsilon_{klm}\partial_j (A_l B_m) && \text{(∂ not acting on ϵ)} \\
&= \epsilon_{ijk}\epsilon_{klm}\left(B_m\partial_j A_l + A_l\partial_j B_m\right) && \text{(product rule)} \\
&= \epsilon_{ijk}\epsilon_{lmk}\left(B_m\partial_j A_l + A_l\partial_j B_m\right) && \text{(Eq. 10.41)} \\
&= \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)\left(B_m\partial_j A_l + A_l\partial_j B_m\right) && \text{(Eq. 10.59)} \\
&= \delta_{il}\delta_{jm}B_m\partial_j A_l + \delta_{il}\delta_{jm}A_l\partial_j B_m - \delta_{im}\delta_{jl}B_m\partial_j A_l - \delta_{im}\delta_{jl}A_l\partial_j B_m && \text{(distributivity)} \\
&= B_m\partial_m A_i + A_i\partial_m B_m - B_i\partial_j A_j - A_j\partial_j B_i && \text{(Eq. 10.33)} \\
&= (B_m\partial_m)A_i + (\partial_m B_m)A_i - \left(\partial_j A_j\right)B_i - \left(A_j\partial_j\right)B_i && \text{(grouping)} \\
&= \left[(\mathbf{B}\cdot\nabla)\mathbf{A}\right]_i + \left[(\nabla\cdot\mathbf{B})\mathbf{A}\right]_i - \left[(\nabla\cdot\mathbf{A})\mathbf{B}\right]_i - \left[(\mathbf{A}\cdot\nabla)\mathbf{B}\right]_i && \text{(Eqs. 12.32 \& 12.18)} \\
&= \left[(\mathbf{B}\cdot\nabla)\mathbf{A} + (\nabla\cdot\mathbf{B})\mathbf{A} - (\nabla\cdot\mathbf{A})\mathbf{B} - (\mathbf{A}\cdot\nabla)\mathbf{B}\right]_i && \text{(Eq. \textbf{??})}
\end{aligned}
\tag{12.75}
$$

Because i is a free index the identity is proved for all components.

$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \begin{vmatrix} \mathbf{A}\cdot\mathbf{C} & \mathbf{A}\cdot\mathbf{D} \\ \mathbf{B}\cdot\mathbf{C} & \mathbf{B}\cdot\mathbf{D} \end{vmatrix}$:

$$
\begin{aligned}
(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) &= [\mathbf{A} \times \mathbf{B}]_i [\mathbf{C} \times \mathbf{D}]_i && \text{(Eq. 12.8)} \\
&= \epsilon_{ijk}A_j B_k \epsilon_{ilm}C_l D_m && \text{(Eq. \textbf{??})} \\
&= \epsilon_{ijk}\epsilon_{ilm}A_j B_k C_l D_m && \text{(commutativity)} \\
&= \left(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\right)A_j B_k C_l D_m && \text{(Eqs. 10.41 \& 10.59)} \\
&= \delta_{jl}\delta_{km}A_j B_k C_l D_m - \delta_{jm}\delta_{kl}A_j B_k C_l D_m && \text{(distributivity)} \\
&= \left(\delta_{jl}A_j C_l\right)(\delta_{km}B_k D_m) - \left(\delta_{jm}A_j D_m\right)(\delta_{kl}B_k C_l) && \text{(commutativity \& grouping)} \\
&= (A_l C_l)(B_m D_m) - (A_m D_m)(B_l C_l) && \text{(Eq. 10.33)} \\
&= (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}) && \text{(Eq. 12.8)} \\
&= \begin{vmatrix} \mathbf{A}\cdot\mathbf{C} & \mathbf{A}\cdot\mathbf{D} \\ \mathbf{B}\cdot\mathbf{C} & \mathbf{B}\cdot\mathbf{D} \end{vmatrix} && \text{(definition of determinant)}
\end{aligned}
\tag{12.76}
$$

$$(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = \left[\mathbf{D}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{C} - \left[\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{D}:$$

$$\begin{aligned}
\left[(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D})\right]_i &= \epsilon_{ijk}\,[\mathbf{A}\times\mathbf{B}]_j\,[\mathbf{C}\times\mathbf{D}]_k && \text{(Eq. ??)}\\
&= \epsilon_{ijk}\epsilon_{jmn}A_mB_n\epsilon_{kpq}C_pD_q && \text{(Eq. ??)}\\
&= \epsilon_{ijk}\epsilon_{kpq}\epsilon_{jmn}A_mB_nC_pD_q && \text{(commutativity)}\\
&= \epsilon_{ijk}\epsilon_{pqk}\epsilon_{jmn}A_mB_nC_pD_q && \text{(Eq. 10.41)}\\
&= \left(\delta_{ip}\delta_{jq}-\delta_{iq}\delta_{jp}\right)\epsilon_{jmn}A_mB_nC_pD_q && \text{(Eq. 10.59)}\\
&= \left(\delta_{ip}\delta_{jq}\epsilon_{jmn}-\delta_{iq}\delta_{jp}\epsilon_{jmn}\right)A_mB_nC_pD_q && \text{(distributivity)}\\
&= \left(\delta_{ip}\epsilon_{qmn}-\delta_{iq}\epsilon_{pmn}\right)A_mB_nC_pD_q && \text{(Eq. 10.33)}\\
&= \delta_{ip}\epsilon_{qmn}A_mB_nC_pD_q-\delta_{iq}\epsilon_{pmn}A_mB_nC_pD_q && \text{(distributivity)}\\
&= \epsilon_{qmn}A_mB_nC_iD_q-\epsilon_{pmn}A_mB_nC_pD_i && \text{(Eq. 10.33)}\\
&= \epsilon_{qmn}D_qA_mB_nC_i-\epsilon_{pmn}C_pA_mB_nD_i && \text{(commutativity)}\\
&= \left(\epsilon_{qmn}D_qA_mB_n\right)C_i-\left(\epsilon_{pmn}C_pA_mB_n\right)D_i && \text{(grouping)}\\
&= \left[\mathbf{D}\cdot(\mathbf{A}\times\mathbf{B})\right]C_i-\left[\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})\right]D_i && \text{(Eq. ??)}\\
&= \Big[\left[\mathbf{D}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{C}\Big]_i-\Big[\left[\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{D}\Big]_i && \text{(definition of index)}\\
&= \Big[\left[\mathbf{D}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{C}-\left[\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})\right]\mathbf{D}\Big]_i && \text{(Eq. ??)}
\end{aligned} \tag{12.77}$$

Because i is a free index the identity is proved for all components.

12.6 The Inertia Tensor

Consider masses m_α with positions $\mathbf{r}_\alpha$, all rotating with angular velocity ω about $\mathbf{0}$. So the velocities are $\mathbf{v}_\alpha = \omega\times\mathbf{r}_\alpha$. The total angular momentum is

$$\begin{aligned}
\mathbf{L} &= \sum_\alpha \mathbf{r}_\alpha\times m_\alpha\mathbf{v}_\alpha\\
&= \sum_\alpha m_\alpha\mathbf{r}_\alpha\times(\omega\times\mathbf{r}_\alpha)\\
&= \sum_\alpha m_\alpha(|\mathbf{r}_\alpha|^2\omega-(\mathbf{r}_\alpha\cdot\omega)\mathbf{r}_\alpha).
\end{aligned}$$

by vector identities. In components, we have

$$L_i = I_{ij}\omega_j,$$

where

Definition 12.6.1 — Inertia tensor. The **inertia tensor** is defined as

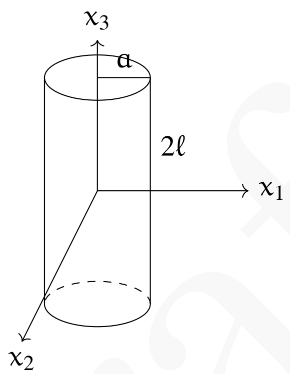
$$I_{ij} = \sum_{\alpha} m_{\alpha}[|\mathbf{r}_{\alpha}|^2 \delta_{ij} - (\mathbf{r}_{\alpha})_i(\mathbf{r}_{\alpha})_j].$$

For a rigid body occupying volume V with mass density $\rho(\mathbf{r})$, we replace the sum with an integral to obtain

$$I_{ij} = \int_V \rho(\mathbf{r})(x_k x_k \delta_{ij} - x_i x_j)\, dV.$$

By inspection, I is a symmetric tensor.

Example 12.6.1 Consider a rotating cylinder with uniform density ρ_0. The total mass is $2\ell\pi a^2 \rho_0$.



Use cylindrical polar coordinate:

$$\begin{aligned}
x_1 &= r\cos\theta \\
x_2 &= r\sin\theta \\
x_3 &= x_3 \\
dV &= r\, dr\, d\theta\, dx_3
\end{aligned}$$

We have

$$\begin{aligned}
I_{33} &= \int_V \rho_0(x_1^2 + x_2^2)\, dV \\
&= \rho_0 \int_0^a \int_0^{2\pi} \int_{-\ell}^{\ell} r^2 (r\, dr\, d\theta\, dx_2) \\
&= \rho_0 \cdot 2\pi \cdot 2\ell \left[\frac{r^4}{4}\right]_0^a \\
&= \varepsilon_0 \pi \ell a^4.
\end{aligned}$$

Similarly, we have

$$
\begin{aligned}
I_{11} &= \int_V \rho_0 (x_2^2 + x_3^2)\, dV \\
&= \rho_0 \int_0^a \int_0^{2\pi} \int_{-\ell}^{\ell} (r^2 \sin^2\theta + x_3^2) r \, dr \, d\theta \, dx_3 \\
&= \rho_0 \int_0^a \int_0^{2\pi} r \left(r^2 \sin^2\theta \, [x_3]_{-\ell}^{\ell} + \left[\frac{x_3^3}{3} \right]_{-\ell}^{\ell} \right) d\theta \, dr \\
&= \rho_0 \int_0^a \int_0^{2\pi} r \left(r^2 \sin^2\theta 2\ell + \frac{2}{3}\ell^3 \right) d\theta \, dr \\
&= \rho_0 \left(2\pi a \cdot \frac{2}{3}\ell^3 + 2\ell \int_0^a r^2 \, dr \int_0^{2\pi} \sin^2\theta \right) \\
&= \rho_0 \pi a^2 \ell \left(\frac{a^2}{2} + \frac{2}{3}\ell^2 \right)
\end{aligned}
$$

By symmetry, the result for I_{22} is the same.

How about the off-diagonal elements?

$$
\begin{aligned}
I_{13} &= -\int_V \rho_0 x_1 x_3 \, dV \\
&= -\rho_0 \int_0^a \int_{-\ell}^{\ell} \int_0^{2\pi} r^2 \cos\theta x_3 \, dr \, dx_3 \, d\theta \\
&= 0
\end{aligned}
$$

Since $\int_0^{2\pi} d\theta \cos\theta = 0$. Similarly, the other off-diagonal elements are all 0. So the non-zero components are

$$
\begin{aligned}
I_{33} &= \frac{1}{2} M a^2 \\
I_{11} = I_{22} &= M \left(\frac{a^2}{4} + \frac{\ell^2}{3} \right)
\end{aligned}
$$

In the particular case where $\ell = \frac{a\sqrt{3}}{2}$, we have $I_{ij} = \frac{1}{2} m a^2 \delta_{ij}$. So in this case,

$$
\mathbf{L} = \frac{1}{2} M a^2 \omega
$$

for rotation about any axis.

Example 12.6.2 — Inertia Tensor of a Cube about the Center of Mass. The high degree of symmetry here means we only need to do two out of nine possible integrals.

$$
\begin{aligned}
I_{xx} &= \int dV\rho(y^2+z^2) && (12.78)\\
&= \rho\int_{-b/2}^{b/2} dx \int_{-b/2}^{b/2} dy \int_{-b/2}^{b/2} dz(y^2+z^2) && (12.79)\\
&= \rho b\int_{-b/2}^{b/2} dy\ (zy^2+\frac{1}{3}z^3)\Big|_{-b/2}^{b/2} && (12.80)\\
&= \rho b\int_{-b/2}^{b/2} dy\left(by^2+\frac{1}{3}\frac{b^3}{4}\right) && (12.81)\\
&= \rho b\left(\frac{1}{3}by^3+\frac{1}{12}b^3y\right)\Bigg|_{-b/2}^{b/2} && (12.82)\\
&= \rho b\left(\frac{1}{12}b^4+\frac{1}{12}b^4\right) && (12.83)\\
&= \frac{1}{6}\rho b^5 = \frac{1}{6}Mb^2. && (12.84)
\end{aligned}
$$

On the other hand, all the off-diagonal moments are zero, for example $I_{xy} = \int dV\rho(-xy)$. This is an odd function of x and y, and our integration is now symmetric about the origin in all directions, so it vanishes identically. So the inertia tensor of the cube about its center is

$$
\bar{I} = \frac{1}{6}Mb^2\begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}.
$$

12.6.1 The Parallel Axis Theorem

The Parallel Axis Theorem relates the inertia tensor about the center of gravity and the inertia tensor about a parallel axis.

For this purpose we consider two coordinate systems: the first $\mathbf{r} = (x,y,z)$ with origin at the center of mass of an arbitrary object, and the second $\mathbf{r}' = (x',y',z')$ offset by some distance. We consider that the object is translated from the origin, but not rotated, by some constant vector $\mathbf{a}$.

In vector form, the coordinates are related as

$$
\mathbf{r}' = \mathbf{a}+\mathbf{r}.
$$

Note that $\mathbf{a}$ points towards the center of mass - the direction is important.

Theorem 12.6.3 If I_{ij} is the inertia tensor calculated in Center of Mass Coordinate, and J_{ij} is the tensor in the translated coordinates, then:

$$J_{ij} = I_{ij} + M(a^2\delta_{ij} - a_i a_j).$$

Example 12.6.4 — Inertia Tensor of a Cube about a corner. The CM inertia tensor was

$$\bar{I} = Mb^2 \begin{pmatrix} 1/6 & 0 & 0 \\ 0 & 1/6 & 0 \\ 0 & 0 & 1/6 \end{pmatrix}$$

If instead we want the tensor about one corner of the cube, the displacement vector is

$$\mathbf{a} = (b/2, b/2, b/2),$$

so $a^2 = (3/4)b^2$. We can construct the difference as a matrix: the off-diagonal components are

$$M\left[\frac{3}{4}B^2 - \left(\frac{1}{2}b\right)\left(\frac{1}{2}b\right)\right] = \frac{1}{2}Mb^2$$

and off-diagonal,

$$M\left[-\left(\frac{1}{2}b\right)\left(\frac{1}{2}b\right)\right] = -\frac{1}{4}Mb^2$$

so the shifted inertia tensor is

$$\bar{J} = Mb^2 \begin{pmatrix} 1/6 & 0 & 0 \\ 0 & 1/6 & 0 \\ 0 & 0 & 1/6 \end{pmatrix} + Mb^2 \begin{pmatrix} 1/2 & -1/4 & -1/4 \\ -1/4 & 1/2 & -1/4 \\ -1/4 & -1/4 & 1/2 \end{pmatrix} \tag{12.85}$$

$$= Mb^2 \begin{pmatrix} 2/3 & -1/4 & -1/4 \\ -1/4 & 2/3 & -1/4 \\ -1/4 & -1/4 & 2/3 \end{pmatrix} \tag{12.86}$$

12.7 Taylor's Theorem

12.7.1 Multi-index Notation

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

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

of natural number. The set of n-tuple is denoted $\mathbb{N}_0^n$.
For multi-indices $\alpha, \beta \in \mathbb{N}_0^n$ and $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ one defines:

❶ **Componentwise sum and difference** $\alpha \pm \beta = (\alpha_1 \pm \beta_1,\ \alpha_2 \pm \beta_2, \dots,\ \alpha_n \pm \beta_n)$

❷ **Partial order** $\alpha < \beta \quad \Leftrightarrow \quad \alpha_i < \beta_i \quad \forall i \in \{1, \dots, n\}$

❸ **Sum of components** $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$

❹ **Factorial** $\alpha! = \alpha_1! \cdot \alpha_2! \cdots \alpha_n!$

❺ **Binomial coefficient** $\binom{\alpha}{\beta} = \binom{\alpha_1}{\beta_1}\binom{\alpha_2}{\beta_2}\cdots\binom{\alpha_n}{\beta_n} = \dfrac{\alpha!}{\beta!(\alpha-\beta)!}$

❻ **Multinomial coefficient** $\binom{k}{\alpha} = \dfrac{k!}{\alpha_1!\alpha_2!\cdots\alpha_n!} = \dfrac{k!}{\alpha!}$

❼ **Power** $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$.

❽ **Higher-order Derivatives**

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n}$$

where $k := |\alpha| \in \mathbb{N}_0$ and $\partial_i^{\alpha_i} := \partial^{\alpha_i}/x_i^{\alpha_i}$

12.7.2 Taylor's Theorem for Multivariate Functions

If the function is $k+1$ times continuously differentiable in the closed ball B, then one can derive an exact formula for the remainder in terms of order partial derivatives of f in this neighborhood. Namely,

Theorem 12.7.1 — Taylor Theorem. Let $f : B \subset \mathbb{R}^n \to \mathbb{R}$ $k+1$ times continuously differentiable in the closed ball B, then:

$$f(\mathbf{x}) = \sum_{|\alpha|<qk} \frac{D^\alpha f(\mathbf{a})}{\alpha!}(\mathbf{x}-\mathbf{a})^\alpha + \sum_{|\beta|=k+1} R_\beta(\mathbf{x})(\mathbf{x}-\mathbf{a})^\beta, \tag{12.87}$$

$$R_\beta(\mathbf{x}) = \frac{|\beta|}{\beta!}\int_0^1 (1-t)^{|\beta|-1} D^\beta f\big(\mathbf{a}+t(\mathbf{x}-\mathbf{a})\big)\,dt. \tag{12.88}$$

We prove the special case, where $f : \mathbb{R}^n \to \mathbb{R}$ has continuous partial derivatives up to the order $k+1$ in some closed ball B with center $\mathbf{a}$.

The strategy of the proof is to apply the one-variable case of Taylor's theorem to the restriction of f to the line segment adjoining $\mathbf{x}$ and $\mathbf{a}$. Parametrize the line segment between $\mathbf{a}$ and $\mathbf{x}$ by $u(t) = a + t(x - a)$

We apply the one-variable version of Taylor's theorem to the function $g(t) = f(u(t))$

$$f(\mathbf{x}) = g(1) = g(0) + \sum_{j=1}^{k} \frac{1}{j!} g^{(j)}(0) \; + \int_0^1 \frac{(1-t)^k}{k!} g^{(k+1)}(t)\, dt.$$

Applying the chain rule for several variables gives

$$g^{(j)}(t) = \frac{d^j}{dt^j} f(u(t)) = \frac{d^j}{dt^j} f(\mathbf{a} + t(\mathbf{x} - \mathbf{a})) \tag{12.89}$$

$$= \sum_{|\alpha|=j} \binom{j}{\alpha} (D^\alpha f)(\mathbf{a} + t(\mathbf{x} - \mathbf{a}))(\mathbf{x} - \mathbf{a})^\alpha \tag{12.90}$$

where $\binom{j}{\alpha}$ is the multinomial coefficient. Since $\frac{1}{j!}\binom{j}{\alpha} = \frac{1}{\alpha!}$, we get

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{|\alpha|<qk} \frac{1}{\alpha!} (D^\alpha f)(\mathbf{a})(\mathbf{x} - \mathbf{a})^\alpha + \sum_{|\alpha|=k+1} \frac{k+1}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha \int_0^1 (1-t)^k (D^\alpha f)(\mathbf{a} + t(\mathbf{x} - \mathbf{a}))\, dt.$$

12.8 Ohm's Law

Ohm's law is an empirical law that states that there is a linear relationship between the electric current j flowing through a material and the electric field E applied to this material. This law can be written as

$$j = \sigma E$$

where the constant of proportionality σ is known as the conductivity (the conductivity is defined as the inverse of resistivity).

One important consequence of equation 12.8 is that the vectors j and E are necessary parallel.

This law is true for some materials, but not for all. For example, if the medium is made of alternate layers of a conductor and an insulator, then the current can only flow along the layers, regardless of the direction of the electric field. It is useful therefore to have an alternative to equation in which j and E do not have to be parallel.

This can be achieved by introducing the **conductivity tensor**, σ_{ik}, which relates j and E through the equation:

$$j_i = \sigma_{ik} E_k$$

We note that as j and E are vectors, it follows from the quotient rule that σ_{ik} is a tensor.

12.9 Equation of Motion for a Fluid: Navier-Stokes Equation

12.9.1 Stress Tensor

The stress tensor consists of nine components σ_{ij} that completely define the state of stress at a point inside a material in the deformed state, placement, or configuration.

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

The stress tensor can be separated into two components. One component is a **hydrostatic** or **dilatational** stress that acts to change the volume of the material only; the other is the **deviator stress** that acts to change the shape only.

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{31} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{23} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \sigma_H & 0 & 0 \\ 0 & \sigma_H & 0 \\ 0 & 0 & \sigma_H \end{pmatrix} + \begin{pmatrix} \sigma_{11} - \sigma_H & \sigma_{12} & \sigma_{31} \\ \sigma_{12} & \sigma_{22} - \sigma_H & \sigma_{23} \\ \sigma_{31} & \sigma_{23} & \sigma_{33} - \sigma_H \end{pmatrix}$$

12.9.2 Derivation of the Navier-Stokes Equations

The Navier-Stokes equations can be derived from the conservation and continuity equations and some properties of fluids. In order to derive the equations of fluid motion, we will first derive the continuity equation, apply the equation to conservation of mass and momentum, and finally combine the conservation equations with a physical understanding of what a fluid is.

The first assumption is that the motion of a fluid are described with the flow velocity of the fluid:

Definition 12.9.1 The flow velocity $\mathbf{v}$ of a fluid is a vector field

$$\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$$

which gives the velocity of an element of fluid at a position $\mathbf{x}$ and time t

Material Derivative

A normal derivative is the rate of change of of an property at a point. For instance, the value $\frac{dT}{dt}$ could be the rate of change of temperature at a point (x, y). However, a material derivative is the rate of change of an property on a particle in a velocity field. It incorporates two things:

- Rate of change of the property, $\frac{dL}{dt}$

- Change in position of of the particle in the velocity field $\mathbf{v}$

Therefore, the material derivative can be defined as

Definition 12.9.2 — Material Derivative. Given a function $u(t,x,y,z)$

$$\frac{Du}{Dt} = \frac{du}{dt} + (\mathbf{v}\cdot\nabla)u.$$

Continuity Equation

An **intensive property** is a quantity whose value does not depend on the amount of the substance for which it is measured. For example, the temperature of a system is the same as the temperature of any part of it. If the system is divided the temperature of each subsystem is identical. The same applies to the density of a homogeneous system; if the system is divided in half, the mass and the volume change in the identical ratio and the density remains unchanged.

The volume will be denoted by U and its bounding surface area is referred to as ∂U. The continuity equation derived can later be applied to mass and momentum.

Reynold's Transport Theorem

The first basic assumption is the Reynold's Transport Theorem:

Theorem 12.9.1 — Reynold's Transport Theorem. Let U be a region in $\mathbb{R}^n$ with a C^1 boundary ∂U. Let $\mathbf{x}(t)$ be the positions of points in the region and let $\mathbf{v}(\mathbf{x},t)$ be the velocity field in the region. Let $\mathbf{n}(\mathbf{x},t)$ be the outward unit normal to the boundary. Let $L(\mathbf{x},t)$ be a C^2 scalar field. Then

$$\frac{d}{dt}\left(\int_U L\, dV\right) = \int_U \frac{\partial L}{\partial t}\, dV + \int_{\partial U} (\mathbf{v}\cdot\mathbf{n}) L\, dA\ .$$

What we will write in a simplified way as

$$\frac{d}{dt}\int_U L\, dV = -\int_{\partial U} L\mathbf{v}\cdot\mathbf{n}\, dA - \int_U Q\, dV. \tag{12.91}$$

The left hand side of the equation denotes the rate of change of the property L contained inside the volume U. The right hand side is the sum of two terms:

- A flux term, $\int_{\partial U} L\mathbf{v}\cdot\mathbf{n}\, dA$, which indicates how much of the property L is leaving the volume by flowing over the boundary ∂U
- A sink term, $\int_U Q\, dV$, which describes how much of the property L is leaving the volume due to sinks or sources inside the boundary

This equation states that the change in the total amount of a property is due to how much flows out through the volume boundary as well as how much is lost or gained through sources or sinks inside the boundary.

If the intensive property we're dealing with is density, then the equation is simply a statement of conservation of mass: the change in mass is the sum of what leaves the boundary and what appears within it; no mass is left unaccounted for.

Divergence Theorem

The Divergence Theorem allows the flux term of the above equation to be expressed as a volume integral. By the Divergence Theorem,

$$\int_{\partial U} L\mathbf{v}\cdot\mathbf{n}\,dA = \int_{U} \nabla\cdot(L\mathbf{v})\,dV.$$

Therefore, we can now rewrite our previous equation as

$$\frac{d}{dt}\int_{U} L\,dV = -\int_{U}\left[\nabla\cdot(L\mathbf{v})+Q\right]\,dV.$$

Deriving under the integral sign, we find that

$$\int_{U}\frac{d}{dt}L\,dV = -\int_{U}\nabla\cdot(L\mathbf{v})+Q\,dV.$$

Equivalently,

$$\int_{U}\frac{d}{dt}L+\nabla\cdot(L\mathbf{v})+Q\,dV = 0.$$

This relation applies to any volume U; the only way the above equality remains true for any volume U is if the integrand itself is zero. Thus, we arrive at the differential form of the continuity equation

$$\frac{dL}{dt}+\nabla\cdot(L\mathbf{v})+Q = 0.$$

Conservation of Mass

Applying the continuity equation to density, we obtain

$$\frac{d\rho}{dt}+\nabla\cdot(\rho\mathbf{v})+Q = 0.$$

This is the conservation of mass because we are operating with a constant volume U. With no sources or sinks of mass $(Q=0)$,

$$\frac{d\rho}{dt}+\nabla\cdot(\rho\mathbf{v}) = 0. \tag{12.92}$$

The equation 12.92 is called conversation of mass.

In certain cases it is useful to simplify it further. For an incompressible fluid, the density is constant. Setting the derivative of density equal to zero and dividing through by a constant ρ, we obtain the simplest form of the equation

$$\nabla \cdot \mathbf{v} = 0.$$

Conversation of Momentum

We start with

$$\mathbf{F} = m\mathbf{a}.$$

Allowing for the body force $\mathbf{F} = \mathbf{a}$ and substituting density for mass, we get a similar equation

$$\mathbf{b} = \rho \frac{d}{dt}\mathbf{v}(x, y, z, t).$$

Applying the chain rule to the derivative of velocity, we get

$$\mathbf{b} = \rho \left(\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathbf{v}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathbf{v}}{\partial z}\frac{\partial z}{\partial t} \right).$$

Equivalently,

$$\mathbf{b} = \rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right).$$

Substituting the value in parentheses for the definition of a material derivative, we obtain

$$\rho \frac{D\mathbf{v}}{Dt} = \mathbf{b}. \tag{12.93}$$

Equations of Motion

The conservation equations derived above, in addition to a few assumptions about the forces and the behaviour of fluids, lead to the equations of motion for fluids.

We assume that the body force on the fluid parcels is due to two components, fluid stresses and other, external forces.

$$\mathbf{b} = \nabla \cdot \sigma + \mathbf{f}. \tag{12.94}$$

Here, σ is the stress tensor, and $\mathbf{f}$ represents external forces. Intuitively, the fluid stress is represented as the divergence of the stress tensor because the divergence is the extent to which the tensor acts like a sink or source; in other words, the divergence of the tensor results in a momentum source or sink, also known as a force. For many applications $\mathbf{f}$ is the gravity force, but for now we will leave the equation in its most general form.

General Form of the Navier-Stokes Equation

We divide the stress tensor σ into the hydrostatic and deviator part. Denoting the stress deviator tensor as T, we can make the substitution

$$\sigma = -pI + T. \tag{12.95}$$

Substituting this into the previous equation, we arrive at the most general form of the Navier-Stokes equation:

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \nabla \cdot T + \mathbf{f}. \tag{12.96}$$

CHAPTER 13

Integration of Forms

13.1 Differential Forms

Definition 13.1.1 A k-**differential form field in** $\mathbb{R}^n$ is an expression of the form

$$\omega = \sum_{1\leqslant j_1 \leqslant j_2 \leqslant \cdots \leqslant j_k \leqslant n} a_{j_1 j_2 \ldots j_k} dx_{j_1} \wedge dx_{j_2} \wedge \cdots dx_{j_k},$$

where the $a_{j_1 j_2 \ldots j_k}$ are differentiable functions in $\mathbb{R}^n$.

A 0-differential form in $\mathbb{R}^n$ is simply a differentiable function in $\mathbb{R}^n$.

Example 13.1.1

$$g(x,y,z,w) = x + y^2 + z^3 + w^4$$

is a 0-form in $\mathbb{R}^4$.

Example 13.1.2 An example of a 1-form field in $\mathbb{R}^3$ is

$$\omega = x dx + y^2 dy + xyz^3 dz.$$

Example 13.1.3 An example of a 2-form field in $\mathbb{R}^3$ is

$$\omega = x^2 dx \wedge dy + y^2 dy \wedge dz + dz \wedge dx.$$

Example 13.1.4 An example of a 3-form field in $\mathbb{R}^3$ is

$$\omega = (x+y+z) dx \wedge dy \wedge dz.$$

We shew now how to multiply differential forms.

Example 13.1.5 The product of the 1-form fields in $\mathbb{R}^3$

$$\omega_1 = y\mathrm{d}x + x\mathrm{d}y,$$

$$\omega_2 = -2x\mathrm{d}x + 2y\mathrm{d}y,$$

is

$$\omega_1 \wedge \omega_2 = (2x^2 + 2y^2)\mathrm{d}x \wedge \mathrm{d}y.$$

Definition 13.1.2 Let $f(x_1, x_2, \ldots, x_n)$ be a 0-form in $\mathbb{R}^n$. The **exterior derivative** $\mathrm{d}f$ of f is

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \mathrm{d}x_i.$$

Furthermore, if

$$\omega = f(x_1, x_2, \ldots, x_n)\mathrm{d}x_{j_1} \wedge \mathrm{d}x_{j_2} \wedge \cdots \wedge \mathrm{d}x_{j_k}$$

is a k-form in $\mathbb{R}^n$, the **exterior derivative** $\mathrm{d}\omega$ of ω is the $(k+1)$-form

$$\mathrm{d}\omega = \mathrm{d}f(x_1, x_2, \ldots, x_n) \wedge \mathrm{d}x_{j_1} \wedge \mathrm{d}x_{j_2} \wedge \cdots \wedge \mathrm{d}x_{j_k}.$$

Example 13.1.6 If in $\mathbb{R}^2$, $\omega = x^3y^4$, then

$$\mathrm{d}(x^3y^4) = 3x^2y^4\mathrm{d}x + 4x^3y^3\mathrm{d}y.$$

Example 13.1.7 If in $\mathbb{R}^2$, $\omega = x^2y\mathrm{d}x + x^3y^4\mathrm{d}y$ then

$$\begin{aligned}
\mathrm{d}\omega &= \mathrm{d}(x^2y\mathrm{d}x + x^3y^4\mathrm{d}y) \\
&= (2xy\mathrm{d}x + x^2\mathrm{d}y) \wedge \mathrm{d}x + (3x^2y^4\mathrm{d}x + 4x^3y^3\mathrm{d}y) \wedge \mathrm{d}y \\
&= x^2\mathrm{d}y \wedge \mathrm{d}x + 3x^2y^4\mathrm{d}x \wedge \mathrm{d}y \\
&= (3x^2y^4 - x^2)\mathrm{d}x \wedge \mathrm{d}y
\end{aligned}$$

Example 13.1.8 Consider the change of variables $x = u + v, y = uv$. Then

$$\mathrm{d}x = \mathrm{d}u + \mathrm{d}v,$$

$$\mathrm{d}y = v\mathrm{d}u + u\mathrm{d}v,$$

whence

$$\mathrm{d}x \wedge \mathrm{d}y = (u - v)\mathrm{d}u \wedge \mathrm{d}v.$$

Example 13.1.9 Consider the transformation of coordinates xyz into uvw coordinates given by

$$u = x + y + z,\ v = \frac{z}{y+z},\ w = \frac{y+z}{x+y+z}.$$

Then

$$du = dx + dy + dz,$$
$$dv = -\frac{z}{(y+z)^2}dy + \frac{y}{(y+z)^2}dz,$$
$$dw = -\frac{y+z}{(x+y+z)^2}dx + \frac{x}{(x+y+z)^2}dy + \frac{x}{(x+y+z)^2}dz.$$

Multiplication gives

$$\begin{aligned} du \wedge dv \wedge dw &= \left(-\frac{zx}{(y+z)^2(x+y+z)^2} - \frac{y(y+z)}{(y+z)^2(x+y+z)^2} \right. \\ &\quad \left. + \frac{z(y+z)}{(y+z)^2(x+y+z)^2} - \frac{xy}{(y+z)^2(x+y+z)^2} \right) dx \wedge dy \wedge dz \\ &= \frac{z^2 - y^2 - zx - xy}{(y+z)^2(x+y+z)^2} dx \wedge dy \wedge dz. \end{aligned}$$

13.2 Integrating Differential Forms

Let

$$\omega = \sum_{i_1 < \cdots < i_k} a_{i_1,\ldots,i_k}(\mathbf{x})\, dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

be a differential form and M a differentiable-manifold over which we wish to integrate, where M has the parameterization

$$M(\mathbf{u}) = (x^1(\mathbf{u}), \ldots, x^k(\mathbf{u}))$$

for in the parameter $\mathbf{u}$ domain D . Then defines the integral of the differential form over as

$$\int_S \omega = \int_D \sum_{i_1 < \cdots < i_k} a_{i_1,\ldots,i_k}(M(\mathbf{u})) \frac{\partial(x^{i_1}, \ldots, x^{i_k})}{\partial(u^1, \ldots, u^k)}\, du^1 \cdots du^k,$$

where the integral on the right-hand side is the standard Riemann integral over D, and

$$\frac{\partial(x^{i_1}, \ldots, x^{i_k})}{\partial(u^1, \ldots, u^k)}$$

is the determinant of the Jacobian.

13.3 Zero-Manifolds

Definition 13.3.1 A 0-dimensional oriented manifold of $\mathbb{R}^n$ is simply a point $\mathbf{x} \in \mathbb{R}^n$, with a choice of the $+$ or $-$ sign. A general oriented 0-manifold is a union of oriented points.

Definition 13.3.2 Let $M = +\{\mathbf{b}\} \cup -\{\mathbf{a}\}$ be an oriented 0-manifold, and let ω be a 0-form. Then

$$\int_M \omega = \omega(\mathbf{b}) - \omega(\mathbf{a}).$$

$-\mathbf{x}$ has opposite orientation to $+\mathbf{x}$ and

$$\int_{-\mathbf{x}} \omega = -\int_{+\mathbf{x}} \omega.$$

Example 13.3.1 Let $M = -\{(1,0,0)\} \cup +\{(1,2,3)\} \cup -\{(0,-2,0)\}$[1] be an oriented 0-manifold, and let $\omega = x + 2y + z^2$. Then

$$\int_M \omega = -\omega((1,0,0)) + \omega(1,2,3) - \omega(0,0,3) = -(1) + (14) - (-4) = 17.$$

13.4 One-Manifolds

Definition 13.4.1 A 1-**dimensional oriented manifold of** $\mathbb{R}^n$ is simply an oriented smooth curve $\Gamma \in \mathbb{R}^n$, with a choice of a $+$ orientation if the curve traverses in the direction of increasing t, or with a choice of a $-$ sign if the curve traverses in the direction of decreasing t. A general oriented 1-manifold is a union of oriented curves.

The curve $-\Gamma$ has opposite orientation to Γ and

$$\int_{-\Gamma} \omega = -\int_{\Gamma} \omega.$$

If $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and if $d\mathbf{r} = \begin{bmatrix} dx \\ dy \end{bmatrix}$, the classical way of writing this is

$$\int_{\Gamma} \mathbf{f} \cdot d\mathbf{r}.$$

We now turn to the problem of integrating 1-forms.

Example 13.4.1 Calculate

$$\int_{\Gamma} xydx + (x+y)dy$$

where Γ is the parabola $y = x^2$, $x \in [-1;2]$ oriented in the positive direction.

[1] Do not confuse, say, $-\{(1,0,0)\}$ with $-(1,0,0) = (-1,0,0)$. The first one means that the point $(1,0,0)$ is given negative orientation, the second means that $(-1,0,0)$ is the additive inverse of $(1,0,0)$.

Solution: ► We parametrise the curve as $x = t, y = t^2$. Then

$$xydx + (x+y)dy = t^3dt + (t+t^2)dt^2 = (3t^3 + 2t^2)dt,$$

whence

$$\begin{aligned}\int_\Gamma \omega &= \int_{-1}^{2} (3t^3 + 2t^2)dt \\ &= \left[\frac{2}{3}t^3 + \frac{3}{4}t^4\right]_{-1}^{2} \\ &= \frac{69}{4}.\end{aligned}$$

What would happen if we had given the curve above a different parametrisation? First observe that the curve travels from $(-1,1)$ to $(2,4)$ on the parabola $y = x^2$. These conditions are met with the parametrisation $x = \sqrt{t} - 1, y = (\sqrt{t} - 1)^2$, $t \in [0;9]$. Then

$$\begin{aligned} xydx + (x+y)dy &= (\sqrt{t}-1)^3 d(\sqrt{t}-1) + ((\sqrt{t}-1) + (\sqrt{t}-1)^2) d(\sqrt{t}-1)^2 \\ &= (3(\sqrt{t}-1)^3 + 2(\sqrt{t}-1)^2) d(\sqrt{t}-1) \\ &= \frac{1}{2\sqrt{t}}(3(\sqrt{t}-1)^3 + 2(\sqrt{t}-1)^2)dt, \end{aligned}$$

whence

$$\begin{aligned}\int_\Gamma \omega &= \int_0^9 \frac{1}{2\sqrt{t}}(3(\sqrt{t}-1)^3 + 2(\sqrt{t}-1)^2)dt \\ &= \left[\tfrac{3t^2}{4} - \tfrac{7t^{3/2}}{3} + \tfrac{5t}{2} - \sqrt{t}\right]_0^9 \\ &= \frac{69}{4},\end{aligned}$$

as before.

◄

It turns out that if two different parametrisations of the same curve have the same orientation, then their integrals are equal. Hence, we only need to worry about finding a suitable parametrisation.

Example 13.4.2 Calculate the line integral

$$\int_\Gamma y \sin x dx + x \cos y dy,$$

where Γ is the line segment from $(0,0)$ to $(1,1)$ in the positive direction.

Solution: ► This line has equation $y = x$, so we choose the parametrisation $x = y = t$. The integral is thus

$$\begin{aligned}\int_\Gamma y\sin x dx + x\cos y dy &= \int_0^1 (t\sin t + t\cos t)dt \\ &= [t(\sin x - \cos t)]_0^1 - \int_0^1 (\sin t - \cos t)dt \\ &= 2\sin 1 - 1,\end{aligned}$$

upon integrating by parts.

◄

Example 13.4.3 Calculate the path integral

$$\int_\Gamma \frac{x+y}{x^2+y^2}\ dy + \frac{x-y}{x^2+y^2}\ dx$$

around the closed square $\Gamma = ABCD$ with $A = (1,1)$, $B = (-1,1)$, $C = (-1,-1)$, and $D = (1,-1)$ in the direction $ABCDA$.

Solution: ► On AB, $y = 1, dy = 0$, on BC, $x = -1, dx = 0$, on CD, $y = -1, dy = 0$, and on DA, $x = 1, dx = 0$. The integral is thus

$$\begin{aligned}\int_\Gamma \omega &= \int_{AB} \omega + \int_{BC} \omega + \int_{CD} \omega + \int_{DA} \omega \\ &= \int_1^{-1} \frac{x-1}{x^2+1}\ dx + \int_1^{-1} \frac{y-1}{y^2+1}\ dy + \int_{-1}^1 \frac{x+1}{x^2+1}\ dx + \int_{-1}^1 \frac{y+1}{y^2+1}\ dy \\ &= 4\int_{-1}^1 \frac{1}{x^2+1}\ dx \\ &= 4\arctan x|_{-1}^1 \\ &= 2\pi.\end{aligned}$$

◄

When the integral is along a closed path, like in the preceding example, it is customary to use the symbol $\oint_\Gamma$ rather than $\int_\Gamma$. The positive direction of integration is that sense that when traversing the path, the area enclosed by the curve is to the left of the curve.

Example 13.4.4 Calculate the path integral

$$\oint_\Gamma x^2 dy + y^2 dx,$$

where Γ is the ellipse $9x^2 + 4y^2 = 36$ traversed once in the positive sense.

Solution: ▶ Parametrise the ellipse as $x = 2\cos t, y = 3\sin t, t \in [0;2\pi]$. Observe that when traversing this closed curve, the area of the ellipse is on the left hand side of the path, so this parametrisation traverses the curve in the positive sense. We have

$$\begin{aligned}\oint_\Gamma \omega &= \int_0^{2\pi} ((4\cos^2 t)(3\cos t) + (9\sin t)(-2\sin t))dt \\ &= \int_0^{2\pi} (12\cos^3 t - 18\sin^3 t)dt \\ &= 0.\end{aligned}$$

◀

Definition 13.4.2 Let Γ be a smooth curve. The integral

$$\int_\Gamma f(\mathbf{x})\|d\mathbf{x}\|$$

is called the **path integral of** f **along** Γ**.**

Example 13.4.5 Find $\int_\Gamma x\|d\mathbf{x}\|$ where Γ is the triangle starting at $A : (-1,-1)$ to $B : (2,-2)$, and ending in $C : (1,2)$.

Solution: ▶ The lines passing through the given points have equations $L_{AB} : y = \frac{-x-4}{3}$, and $L_{BC} : y = -4x+6$. On L_{AB}

$$x\|d\mathbf{x}\| = x\sqrt{(dx)^2 + (dy)^2} = x\sqrt{1+\left(-\frac{1}{3}\right)^2}dx = \frac{x\sqrt{10}dx}{3},$$

and on L_{BC}

$$x\|d\mathbf{x}\| = x\sqrt{(dx)^2 + (dy)^2} = x(\sqrt{1+(-4)^2})dx = x\sqrt{17}dx.$$

Hence

$$\begin{aligned}\int_\Gamma x\|d\mathbf{x}\| &= \int_{L_{AB}} x\|d\mathbf{x}\| + \int_{L_{BC}} x\|d\mathbf{x}\| \\ &= \int_{-1}^{2} \frac{x\sqrt{10}dx}{3} + \int_2^1 x\sqrt{17}dx \\ &= \frac{\sqrt{10}}{2} - \frac{3\sqrt{17}}{2}.\end{aligned}$$

◀

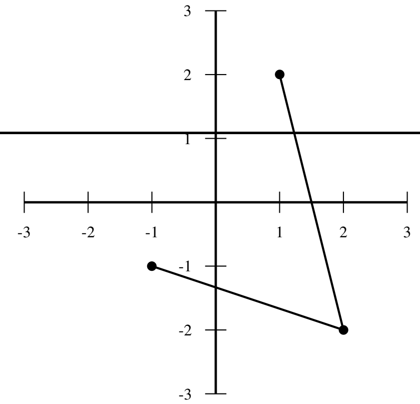


Figure 13.1 Example 13.4.5.

Homework

Problem 13.1 Consider $\int_C x\mathrm{d}x + y\mathrm{d}y$ and $\int_C xy\|\mathrm{d}\mathbf{x}\|$.

1. Evaluate $\int_C x\mathrm{d}x + y\mathrm{d}y$ where C is the straight line path that starts at $(-1,0)$ goes to $(0,1)$ and ends at $(1,0)$, by parametrising this path. Calculate also $\int_C xy\|\mathrm{d}\mathbf{x}\|$ using this parametrisation.

2. Evaluate $\int_C x\mathrm{d}x + y\mathrm{d}y$ where C is the semicircle that starts at $(-1,0)$ goes to $(0,1)$ and ends at $(1,0)$, by parametrising this path. Calculate also $\int_C xy\|\mathrm{d}\mathbf{x}\|$ using this parametrisation.

Problem 13.2 Find $\int_\Gamma x\mathrm{d}x + y\mathrm{d}y$ where Γ is the path shewn in figure **??**, starting at $O(0,0)$ going on a straight line to $A\left(4\cos\frac{\pi}{6}, 4\sin\frac{\pi}{6}\right)$ and continuing on an arc of a circle to $B\left(4\cos\frac{\pi}{5}, 4\sin\frac{\pi}{5}\right)$.

Problem 13.3 Find $\oint_\Gamma z\mathrm{d}x + x\mathrm{d}y + y\mathrm{d}z$ where Γ is the intersection of the sphere $x^2+y^2+z^2=1$ and the plane $x+y=1$, traversed in the positive direction.

13.5 Closed and Exact Forms

Lemma 13.5.1 — Poincaré Lemma. If ω is a p-differential form of continuously differentiable functions in $\mathbb{R}^n$ then

$$\mathrm{d}(\mathrm{d}\omega) = 0.$$

Proof: We will prove this by induction on p. For $p = 0$ if

$$\omega = f(x_1, x_2, \ldots, x_n)$$

then

$$\mathrm{d}\omega = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\mathrm{d}x_k$$

and

$$\begin{aligned}
d(d\omega) &= \sum_{k=1}^{n} d\left(\frac{\partial f}{\partial x_k}\right)\wedge dx_k \\
&= \sum_{k=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_j \partial x_k}\wedge dx_j\right)\wedge dx_k \\
&= \sum_{1\leqslant j\leqslant k\leqslant n}\left(\frac{\partial^2 f}{\partial x_j \partial x_k}-\frac{\partial^2 f}{\partial x_k \partial x_j}\right)dx_j\wedge dx_k \\
&= 0,
\end{aligned}$$

since ω is continuously differentiable and so the mixed partial derivatives are equal. Consider now an arbitrary p-form, $p > 0$. Since such a form can be written as

$$\omega = \sum_{1\leqslant j_1\leqslant j_2\leqslant\cdots\leqslant j_p\leqslant n} a_{j_1 j_2 \ldots j_p} dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p},$$

where the $a_{j_1 j_2 \ldots j_p}$ are continuous differentiable functions in $\mathbb{R}^n$, we have

$$\begin{aligned}
d\omega &= \sum_{1\leqslant j_1\leqslant j_2\leqslant\cdots\leqslant j_p\leqslant n} da_{j_1 j_2 \ldots j_p}\wedge dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p} \\
&= \sum_{1\leqslant j_1\leqslant j_2\leqslant\cdots\leqslant j_p\leqslant n}\left(\sum_{i=1}^{n}\frac{\partial a_{j_1 j_2 \ldots j_p}}{\partial x_i}dx_i\right)\wedge dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p},
\end{aligned}$$

it is enough to prove that for each summand

$$d\left(da\wedge dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right) = 0.$$

But

$$\begin{aligned}
d\left(da\wedge dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right) &= dda\wedge\left(dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right) \\
&\quad + da\wedge d\left(dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right) \\
&= da\wedge d\left(dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right),
\end{aligned}$$

since $dda = 0$ from the case $p = 0$. But an independent induction argument proves that

$$d\left(dx_{j_1}\wedge dx_{j_2}\wedge\cdots dx_{j_p}\right) = 0,$$

completing the proof. ■

Definition 13.5.1 A differential form ω is said to be **exact** if there is a continuously differentiable function F such that

$$dF = \omega.$$

Example 13.5.2 The differential form

$$xdx + ydy$$

is exact, since

$$xdx + ydy = d\left(\frac{1}{2}(x^2 + y^2)\right).$$

Example 13.5.3 The differential form

$$ydx + xdy$$

is exact, since

$$ydx + xdy = d(xy).$$

Example 13.5.4 The differential form

$$\frac{x}{x^2+y^2}dx + \frac{y}{x^2+y^2}dy$$

is exact, since

$$\frac{x}{x^2+y^2}dx + \frac{y}{x^2+y^2}dy = d\left(\frac{1}{2}\log_e(x^2+y^2)\right).$$

Let $\omega = dF$ be an exact form. By the Poincaré Lemma Theorem 13.5.1, $d\omega = ddF = 0$. A result of Poincaré says that for certain domains (called **star-shaped domains**) the converse is also true, that is, if $d\omega = 0$ on a star-shaped domain then ω is exact.

Example 13.5.5 Determine whether the differential form

$$\omega = \frac{2x(1-e^y)}{(1+x^2)^2}dx + \frac{e^y}{1+x^2}dy$$

is exact.

Solution: ► Assume there is a function F such that

$$dF = \omega.$$

By the Chain Rule

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy.$$

This demands that

$$\frac{\partial F}{\partial x} = \frac{2x(1-e^y)}{(1+x^2)^2},$$

$$\frac{\partial F}{\partial y} = \frac{e^y}{1+x^2}.$$

We have a choice here of integrating either the first, or the second expression. Since integrating the second expression (with respect to y) is easier, we find

$$F(x,y) = \frac{e^y}{1+x^2} + \phi(x),$$

where $\phi(x)$ is a function depending only on x. To find it, we differentiate the obtained expression for F with respect to x and find

$$\frac{\partial F}{\partial x} = -\frac{2xe^y}{(1+x^2)^2} + \phi'(x).$$

Comparing this with our first expression for $\frac{\partial F}{\partial x}$, we find

$$\phi'(x) = \frac{2x}{(1+x^2)^2},$$

that is

$$\phi(x) = -\frac{1}{1+x^2} + c,$$

where c is a constant. We then take

$$F(x,y) = \frac{e^y - 1}{1+x^2} + c.$$

◀

Example 13.5.6 Is there a continuously differentiable function such that

$$dF = \omega = y^2z^3dx + 2xyz^3dy + 3xy^2z^2dz \ ?$$

Solution: ► We have

$$\begin{aligned}
d\omega &= (2yz^3dy + 3y^2z^2dz) \wedge dx \\
&\quad + (2yz^3dx + 2xz^3dy + 6xyz^2dz) \wedge dy \\
&\quad + (3y^2z^2dx + 6xyz^2dy + 6xy^2zdz) \wedge dz \\
&= 0,
\end{aligned}$$

so this form is exact in a star-shaped domain. So put

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial z}dz = y^2z^3dx + 2xyz^3dy + 3xy^2z^2dz.$$

Then

$$\begin{aligned}
\frac{\partial F}{\partial x} &= y^2z^3 \implies F = xy^2z^3 + a(y,z), \\
\frac{\partial F}{\partial y} &= 2xyz^3 \implies F = xy^2z^3 + b(x,z), \\
\frac{\partial F}{\partial z} &= 3xy^2z^2 \implies F = xy^2z^3 + c(x,y),
\end{aligned}$$

Comparing these three expressions for F, we obtain $F(x,y,z) = xy^2z^3$. ◀

We have the following equivalent of the Fundamental Theorem of Calculus.

Theorem 13.5.7 Let $U \subseteq \mathbb{R}^n$ be an open set. Assume $\omega = dF$ is an exact form, and Γ a path in U with starting point A and endpoint B. Then

$$\int_\Gamma \omega = \int_A^B dF = F(B) - F(A).$$

In particular, if Γ is a simple closed path, then

$$\oint_\Gamma \omega = 0.$$

Example 13.5.8 Evaluate the integral

$$\oint_\Gamma \frac{2x}{x^2+y^2}dx + \frac{2y}{x^2+y^2}\,dy$$

where Γ is the closed polygon with vertices at $A = (0,0)$, $B = (5,0)$, $C = (7,2)$, $D = (3,2)$, $E = (1,1)$, traversed in the order $ABCDEA$.

Solution: ► Observe that

$$d\left(\frac{2x}{x^2+y^2}dx + \frac{2y}{x^2+y^2}\,dy\right) = -\frac{4xy}{(x^2+y^2)^2}dy \wedge dx - \frac{4xy}{(x^2+y^2)^2}dx \wedge dy = 0,$$

and so the form is exact in a start-shaped domain. By virtue of Theorem 13.5.7, the integral is 0. ◄

Example 13.5.9 Calculate the path integral

$$\oint_\Gamma (x^2-y)dx + (y^2-x)dy,$$

where Γ is a loop of $x^3 + y^3 - 2xy = 0$ traversed once in the positive sense.

Solution: ► Since

$$\frac{\partial}{\partial y}(x^2-y) = -1 = \frac{\partial}{\partial x}(y^2-x),$$

the form is exact, and since this is a closed simple path, the integral is 0. ◄

13.6 Two-Manifolds

Definition 13.6.1 A 2-**dimensional oriented manifold of** $\mathbb{R}^2$ is simply an open set (region) $D \in \mathbb{R}^2$, where the $+$ orientation is counter-clockwise and the $-$ orientation is clockwise. A general oriented 2-manifold is a union of open sets.

The region $-D$ has opposite orientation to D and

$$\int_{-D} \omega = -\int_D \omega.$$

We will often write
$$\int_D f(x,y)dA$$
where dA denotes the **area element**.

In this section, unless otherwise noticed, we will choose the positive orientation for the regions considered. This corresponds to using the area form $dxdy$.

Let $D \subseteq \mathbb{R}^2$. Given a function $f : D \to \mathbb{R}$, the integral
$$\int\limits_D fdA$$
is the sum of all the values of f restricted to D. In particular,
$$\int\limits_D dA$$
is the area of D.

13.7 Three-Manifolds

Definition 13.7.1 A 3-**dimensional oriented manifold of** $\mathbb{R}^3$ is simply an open set (body) $V \in \mathbb{R}^3$, where the $+$ orientation is in the direction of the outward pointing normal to the body, and the $-$ orientation is in the direction of the inward pointing normal to the body. A general oriented 3-manifold is a union of open sets.

The region $-M$ has opposite orientation to M and
$$\int_{-M} \omega = -\int_M \omega.$$
We will often write
$$\int_M fdV$$
where dV denotes the **volume element**.

In this section, unless otherwise noticed, we will choose the positive orientation for the regions considered. This corresponds to using the volume form $dx \wedge dy \wedge dz$.

Let $V \subseteq \mathbb{R}^3$. Given a function $f : V \to \mathbb{R}$, the integral
$$\int\limits_V fdV$$

is the sum of all the values of f restricted to V. In particular,

$$\int_V dV$$

is the oriented volume of V.

Example 13.7.1 Find

$$\int_{[0;1]^3} x^2 y e^{xyz}\, dV.$$

Solution: ► The integral is

$$\begin{aligned}\int_0^1 \left(\int_0^1 \left(\int_0^1 x^2 y e^{xyz}\, dz\right) dy\right) dx &= \int_0^1 \left(\int_0^1 x(e^{xy}-1)\, dy\right) dx \\ &= \int_0^1 (e^x - x - 1)dx \\ &= e - \frac{5}{2}.\end{aligned}$$

◄ s

13.8 Surface Integrals

Definition 13.8.1 A 2-**dimensional oriented manifold of** $\mathbb{R}^3$ is simply a smooth surface $D \in \mathbb{R}^3$, where the $+$ orientation is in the direction of the outward normal pointing away from the origin and the $-$ orientation is in the direction of the inward normal pointing towards the origin. A general oriented 2-manifold in $\mathbb{R}^3$ is a union of surfaces.

The surface $-\Sigma$ has opposite orientation to Σ and

$$\int_{-\Sigma} \omega = -\int_{\Sigma} \omega.$$

In this section, unless otherwise noticed, we will choose the positive orientation for the regions considered. This corresponds to using the ordered basis

$$\{dy \wedge dz,\ dz \wedge dx,\ dx \wedge dy\}.$$

Definition 13.8.2 Let $f:\mathbb{R}^3 \to \mathbb{R}$. The integral of f over the smooth surface Σ (oriented in the positive sense) is given by the expression

$$\int_{\Sigma} f \left\| d^2\mathbf{x} \right\|.$$

Here

$$\left\| d^2\mathbf{x} \right\| = \sqrt{(dx\wedge dy)^2 + (dz\wedge dx)^2 + (dy\wedge dz)^2}$$

is the **surface area element**.

Example 13.8.1 Evaluate $\displaystyle\int_{\Sigma} z \left\| d^2\mathbf{x} \right\|$ where Σ is the outer surface of the section of the paraboloid $z = x^2 + y^2, 0 \leqslant z \leqslant 1$.

Solution: ▶ We parametrise the paraboloid as follows. Let $x = u, y = v, z = u^2 + v^2$. Observe that the domain D of Σ is the unit disk $u^2 + v^2 \leqslant 1$. We see that

$$dx\wedge dy = du\wedge dv,$$

$$dy\wedge dz = -2u du\wedge dv,$$

$$dz\wedge dx = -2v du\wedge dv,$$

and so

$$\left\| d^2\mathbf{x} \right\| = \sqrt{1+4u^2+4v^2}\, du\wedge dv.$$

Now,

$$\int_{\Sigma} z \left\| d^2\mathbf{x} \right\| = \int_{D} (u^2+v^2)\sqrt{1+4u^2+4v^2}\, du dv.$$

To evaluate this last integral we use polar coordinates, and so

$$\begin{aligned}\int_{D} (u^2+v^2)\sqrt{1+4u^2+4v^2}\, du dv &= \int_0^{2\pi}\int_0^1 \rho^3\sqrt{1+4\rho^2}\, d\rho d\theta \\ &= \frac{\pi}{12}(5\sqrt{5}+\frac{1}{5}).\end{aligned}$$

◀

Example 13.8.2 Find the area of that part of the cylinder $x^2 + y^2 = 2y$ lying inside the sphere $x^2 + y^2 + z^2 = 4$.

Solution: ▶ We have

$$x^2 + y^2 = 2y \iff x^2 + (y-1)^2 = 1.$$

We parametrise the cylinder by putting $x = \cos u, y - 1 = \sin u$, and $z = v$. Hence

$$dx = -\sin u du,\ dy = \cos u du,\ dz = dv,$$

whence

$$dx \wedge dy = 0, dy \wedge dz = \cos u du \wedge dv, dz \wedge dx = \sin u du \wedge dv,$$

and so

$$\begin{aligned}\left\|d^2\mathbf{x}\right\| &= \sqrt{(dx \wedge dy)^2 + (dz \wedge dx)^2 + (dy \wedge dz)^2} \\ &= \sqrt{\cos^2 u + \sin^2 u}\, du \wedge dv \\ &= du \wedge dv.\end{aligned}$$

The cylinder and the sphere intersect when $x^2 + y^2 = 2y$ and $x^2 + y^2 + z^2 = 4$, that is, when $z^2 = 4 - 2y$, i.e. $v^2 = 4 - 2(1 + \sin u) = 2 - 2\sin u$. Also $0 \leqslant u \leqslant \pi$. The integral is thus

$$\begin{aligned}\int_\Sigma \left\|d^2\mathbf{x}\right\| &= \int_0^\pi \int_{-\sqrt{2-2\sin u}}^{\sqrt{2-2\sin u}} dv du = \int_0^\pi 2\sqrt{2-2\sin u} du \\ &= 2\sqrt{2} \int_0^\pi \sqrt{1-\sin u}\, du \\ &= 2\sqrt{2}\left(4\sqrt{2} - 4\right).\end{aligned}$$

◄

Example 13.8.3 Evaluate

$$\int_\Sigma x dy dz + (z^2 - zx) dz dx - xy dx dy,$$

where Σ is the top side of the triangle with vertices at $(2,0,0)$, $(0,2,0)$, $(0,0,4)$.

Solution: ► Observe that the plane passing through the three given points has equation $2x + 2y + z = 4$. We project this plane onto the coordinate axes obtaining

$$\int_\Sigma x dy dz = \int_0^4 \int_0^{2-z/2} (2 - y - z/2) dy dz = \frac{8}{3},$$

$$\int_\Sigma (z^2 - zx) dz dx = \int_0^2 \int_0^{4-2x} (z^2 - zx) dz dx = 8,$$

$$-\int_\Sigma xy dx dy = -\int_0^2 \int_0^{2-y} xy dx dy = -\frac{2}{3},$$

and hence

$$\int_\Sigma x dy dz + (z^2 - zx) dz dx - xy dx dy = 10.$$

◄

Homework

Problem 13.4 Evaluate $\int_{\Sigma} y \left\|d^2\mathbf{x}\right\|$ where Σ is the surface $z = x + y^2, 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 2$.

Problem 13.5 Consider the cone $z = \sqrt{x^2 + y^2}$. Find the surface area of the part of the cone which lies between the planes $z = 1$ and $z = 2$.

Problem 13.6 Evaluate $\int_{\Sigma} x^2 \left\|d^2\mathbf{x}\right\|$ where Σ is the surface of the unit sphere $x^2 + y^2 + z^2 = 1$.

Problem 13.7 Evaluate $\int_{S} z \left\|d^2\mathbf{x}\right\|$ over the conical surface $z = \sqrt{x^2 + y^2}$ between $z = 0$ and $z = 1$.

Problem 13.8 You put a perfectly spherical egg through an egg slicer, resulting in n slices of identical height, but you forgot to peel it first! Shew that the amount of egg shell in any of the slices is the same. Your argument must use surface integrals.

Problem 13.9 Evaluate

$$\int_{\Sigma} xy\,dy\,dz - x^2\,dz\,dx + (x+z)\,dx\,dy,$$

where Σ is the top of the triangular region of the plane $2x + 2y + z = 6$ bounded by the first octant.

13.9 Green's, Stokes', and Gauss' Theorems

We are now in position to state the general Stoke's Theorem.

Theorem 13.9.1 — General Stoke's Theorem. Let M be a smooth oriented manifold, having boundary ∂M. If ω is a differential form, then

$$\int_{\partial M} \omega = \int_{M} d\omega.$$

In $\mathbb{R}^2$, if ω is a 1-form, this takes the name of **Green's Theorem.**

Example 13.9.2 Evaluate $\oint_C (x - y^3)dx + x^3 dy$ where C is the circle $x^2 + y^2 = 1$.

Solution: ► We will first use Green's Theorem and then evaluate the integral directly. We have

$$\begin{aligned} d\omega &= d(x - y^3) \wedge dx + d(x^3) \wedge dy \\ &= (dx - 3y^2 dy) \wedge dx + (3x^2 dx) \wedge dy \\ &= (3y^2 + 3x^2) dx \wedge dy. \end{aligned}$$

The region M is the area enclosed by the circle $x^2+y^2=1$. Thus by Green's Theorem, and using polar coordinates,

$$\begin{aligned}\oint_C (x-y^3)dx+x^3dy &= \int_M (3y^2+3x^2)dxdy \\ &= \int_0^{2\pi}\int_0^1 3\rho^2\rho d\rho d\theta \\ &= \frac{3\pi}{2}.\end{aligned}$$

Aliter: We can evaluate this integral directly, again resorting to polar coordinates.

$$\begin{aligned}\oint_C (x-y^3)dx+x^3dy &= \int_0^{2\pi}(\cos\theta-\sin^3\theta)(-\sin\theta)d\theta+(\cos^3\theta)(\cos\theta)d\theta \\ &= \int_0^{2\pi}(\sin^4\theta+\cos^4\theta-\sin\theta\cos\theta)d\theta.\end{aligned}$$

To evaluate the last integral, observe that $1=(\sin^2\theta+\cos^2\theta)^2=\sin^4\theta+2\sin^2\theta\cos^2\theta+\cos^4\theta$, whence the integral equals

$$\begin{aligned}\int_0^{2\pi}(\sin^4\theta+\cos^4\theta-\sin\theta\cos\theta)d\theta &= \int_0^{2\pi}(1-2\sin^2\theta\cos^2\theta-\sin\theta\cos\theta)d\theta \\ &= \frac{3\pi}{2}.\end{aligned}$$

◀

In general, let

$$\omega = f(x,y)dx+g(x,y)dy$$

be a 1-form in $\mathbb{R}^2$. Then

$$\begin{aligned}d\omega &= df(x,y)\wedge dx+dg(x,y)\wedge dy \\ &= \left(\frac{\partial}{\partial x}f(x,y)dx+\frac{\partial}{\partial y}f(x,y)dy\right)\wedge dx+\left(\frac{\partial}{\partial x}g(x,y)dx+\frac{\partial}{\partial y}g(x,y)dy\right)\wedge dy \\ &= \left(\frac{\partial}{\partial x}g(x,y)-\frac{\partial}{\partial y}f(x,y)\right)dx\wedge dy\end{aligned}$$

which gives the classical Green's Theorem

$$\int_{\partial M} f(x,y)dx+g(x,y)dy = \int_M\left(\frac{\partial}{\partial x}g(x,y)-\frac{\partial}{\partial y}f(x,y)\right)dxdy.$$

In $\mathbb{R}^3$, if ω is a 2-form, the above theorem takes the name of **Gauss** or the **Divergence Theorem**.

Example 13.9.3 Evaluate $\int_S (x-y)dydz+zdzdx-ydxdy$ where S is the surface of the sphere

$$x^2+y^2+z^2=9$$

and the positive direction is the outward normal.

Solution: ► The region M is the interior of the sphere $x^2+y^2+z^2=9$. Now,

$$\begin{aligned} d\omega &= (dx-dy)\wedge dy\wedge dz+dz\wedge dz\wedge dx-dy\wedge dx\wedge dy \\ &= dx\wedge dy\wedge dz. \end{aligned}$$

The integral becomes

$$\begin{aligned} \int_M dxdydz &= \frac{4\pi}{3}(27) \\ &= 36\pi. \end{aligned}$$

Aliter: We could evaluate this integral directly. We have

$$\int_\Sigma (x-y)dydz = \int_\Sigma xdydz,$$

since $(x,y,z)\mapsto -y$ is an odd function of y and the domain of integration is symmetric with respect to y. Now,

$$\begin{aligned} \int_\Sigma xdydz &= \int_{-3}^{3}\int_0^{2\pi} |\rho|\sqrt{9-\rho^2}d\rho d\theta \\ &= 36\pi. \end{aligned}$$

Also

$$\int_\Sigma zdzdx = 0,$$

since $(x,y,z)\mapsto z$ is an odd function of z and the domain of integration is symmetric with respect to z. Similarly

$$\int_\Sigma -ydxdy = 0,$$

since $(x,y,z)\mapsto -y$ is an odd function of y and the domain of integration is symmetric with respect to y. ◄

In general, let

$$\omega = f(x,y,z)dy\wedge dz+g(x,y,z)dz\wedge dx+h(x,y,z)dx\wedge dy$$

be a 2-form in $\mathbb{R}^3$. Then

$$\begin{aligned} d\omega &= df(x,y,z)dy\wedge dz+dg(x,y,z)dz\wedge dx+dh(x,y,z)dx\wedge dy \\ &= \left(\frac{\partial}{\partial x}f(x,y,z)dx+\frac{\partial}{\partial y}f(x,y,z)dy+\frac{\partial}{\partial z}f(x,y,z)dz\right)\wedge dy\wedge dz \\ &\quad +\left(\frac{\partial}{\partial x}g(x,y,z)dx+\frac{\partial}{\partial y}g(x,y,z)dy+\frac{\partial}{\partial z}g(x,y,z)dz\right)\wedge dz\wedge dx \\ &\quad +\left(\frac{\partial}{\partial x}h(x,y,z)dx+\frac{\partial}{\partial y}h(x,y,z)dy+\frac{\partial}{\partial z}h(x,y,z)dz\right)\wedge dx\wedge dy \\ &= \left(\frac{\partial}{\partial x}f(x,y,z)+\frac{\partial}{\partial y}g(x,y,z)+\frac{\partial}{\partial z}h(x,y,z)\right)dx\wedge dy\wedge dz, \end{aligned}$$

which gives the classical Gauss's Theorem

$$\int_{\partial M} f(x,y,z)dydz + g(x,y,z)dzdx + h(x,y,z)dxdy = \int_{M} \left(\frac{\partial}{\partial x}f(x,y,z) + \frac{\partial}{\partial y}g(x,y,z) + \frac{\partial}{\partial z}h(x,y,z)\right) dxdydz.$$

Using classical notation, if

$$\mathbf{a} = \begin{bmatrix} f(x,y,z) \\ g(x,y,z) \\ h(x,y,z) \end{bmatrix}, d\mathbf{S} = \begin{bmatrix} dydz \\ dzdx \\ dxdy \end{bmatrix},$$

then

$$\int_{M} (\nabla \cdot \mathbf{a}) dV = \int_{\partial M} \mathbf{a} \cdot d\mathbf{S}.$$

The classical Stokes' Theorem occurs when ω is a 1-form in $\mathbb{R}^3$.

Example 13.9.4 Evaluate $\oint_C y dx + (2x - z)dy + (z - x)dz$ where C is the intersection of the sphere $x^2 + y^2 + z^2 = 4$ and the plane $z = 1$.

Solution: ► We have

$$\begin{aligned} d\omega &= (dy)\wedge dx + (2dx - dz)\wedge dy + (dz - dx)\wedge dz \\ &= -dx\wedge dy + 2dx\wedge dy + dy\wedge dz + dz\wedge dx \\ &= dx\wedge dy + dy\wedge dz + dz\wedge dx. \end{aligned}$$

Since on C, $z = 1$, the surface Σ on which we are integrating is the inside of the circle $x^2 + y^2 + 1 = 4$, i.e., $x^2 + y^2 = 3$. Also, $z = 1$ implies $dz = 0$ and so

$$\int_{\Sigma} d\omega = \int_{\Sigma} dxdy.$$

Since this is just the area of the circular region $x^2 + y^2 \leqslant 3$, the integral evaluates to

$$\int_{\Sigma} dxdy = 3\pi.$$

◄

In general, let

$$\omega = f(x,y,z)dx + g(x,y,z)dy + +h(x,y,z)dz$$

be a 1-form in $\mathbb{R}^3$. Then

$$\begin{aligned}
d\omega &= df(x,y,z)\wedge dx + dg(x,y,z)\wedge dy + dh(x,y,z)\wedge dz \\
&= \left(\frac{\partial}{\partial x}f(x,y,z)dx + \frac{\partial}{\partial y}f(x,y,z)dy + \frac{\partial}{\partial z}f(x,y,z)dz\right)\wedge dx \\
&\quad + \left(\frac{\partial}{\partial x}g(x,y,z)dx + \frac{\partial}{\partial y}g(x,y,z)dy + \frac{\partial}{\partial z}g(x,y,z)dz\right)\wedge dy \\
&\quad + \left(\frac{\partial}{\partial x}h(x,y,z)dx + \frac{\partial}{\partial y}h(x,y,z)dy + \frac{\partial}{\partial z}h(x,y,z)dz\right)\wedge dz \\
&= \left(\frac{\partial}{\partial y}h(x,y,z) - \frac{\partial}{\partial z}g(x,y,z)\right) dy\wedge dz \\
&\quad + \left(\frac{\partial}{\partial z}f(x,y,z) - \frac{\partial}{\partial x}h(x,y,z)\right) dz\wedge dx \\
&\quad \left(\frac{\partial}{\partial x}g(x,y,z) - \frac{\partial}{\partial y}f(x,y,z)\right) dx\wedge dy
\end{aligned}$$

which gives the classical Stokes' Theorem

$$\begin{aligned}
&\int_{\partial M} f(x,y,z)dx + g(x,y,z)dy + h(x,y,z)dz \\
&\qquad = \int_M \left(\frac{\partial}{\partial y}h(x,y,z) - \frac{\partial}{\partial z}g(x,y,z)\right) dydz \\
&\qquad\quad + \left(\frac{\partial}{\partial z}g(x,y,z) - \frac{\partial}{\partial x}f(x,y,z)\right) dxdy \\
&\qquad\quad + \left(\frac{\partial}{\partial x}h(x,y,z) - \frac{\partial}{\partial y}f(x,y,z)\right) dxdy.
\end{aligned}$$

Using classical notation, if

$$\mathbf{a} = \begin{bmatrix} f(x,y,z) \\ g(x,y,z) \\ h(x,y,z) \end{bmatrix}, \quad d\mathbf{r} = \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}, \quad d\mathbf{S} = \begin{bmatrix} dydz \\ dzdx \\ dxdy \end{bmatrix},$$

then

$$\int_M (\nabla\times\mathbf{a})\cdot d\mathbf{S} = \int_{\partial M} \mathbf{a}\cdot d\mathbf{r}.$$

Homework

Problem 13.10 Evaluate $\oint_C x^3 y dx + xy dy$ where C is the square with vertices at $(0,0)$, $(2,0)$, $(2,2)$ and $(0,2)$.

Problem 13.11 Consider the triangle $\triangle$ with vertices $A:(0,0)$, $B:(1,1)$, $C:(-2,2)$.

❶ If L_{PQ} denotes the equation of the line joining P and Q find L_{AB}, L_{AC}, and L_{BC}.

❷ Evaluate
$$\oint_{\triangle} y^2 dx + x dy.$$

❸ Find
$$\int_{D} (1-2y) dx \wedge dy$$
where D is the interior of $\triangle$.

Problem 13.12 Problems 1 through 4 refer to the differential form
$$\omega = x dy \wedge dz + y dz \wedge dx + 2z dx \wedge dy,$$
and the solid M whose boundaries are the paraboloid $z = 1 - x^2 - y^2$, $0 \leqslant z \leqslant 1$ and the disc $x^2 + y^2 \leqslant 1$, $z = 0$. The surface ∂M of the solid is positively oriented upon considering outward normals.

1. Prove that $d\omega = 4 dx \wedge dy \wedge dz$.
2. Prove that in Cartesian coordinates, $\displaystyle\int_{\partial M} \omega = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{0}^{1-x^2-y^2} 4 dz dy dx$.
3. Prove that in cylindrical coordinates, $\displaystyle\int_{M} d\omega = \int_{0}^{2\pi} \int_{0}^{1} \int_{0}^{1-r^2} 4r dz dr d\theta$.
4. Prove that $\displaystyle\int_{\partial M} x dy dz + y dz dx + 2z dx dy = 2\pi$.

Problem 13.13 Problems 1 through 4 refer to the box
$$M = \{(x, y, z) \in \mathbb{R}^3 : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 2\},$$
the upper face of the box
$$U = \{(x, y, z) \in \mathbb{R}^3 : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ z = 2\},$$
the boundary of the box without the upper top $S = \partial M \setminus U$, and the differential form
$$\omega = (\arctan y - x^2) dy \wedge dz + (\cos x \sin z - y^3) dz \wedge dx + (2zx + 6zy^2) dx \wedge dy.$$

1. Prove that $d\omega = 3y^2 dx \wedge dy \wedge dz$.
2. Prove that $\displaystyle\int_{\partial M} (\arctan y - x^2) dy dz + (\cos x \sin z - y^3) dz dx + (2zx + 6zy^2) dx dy = \int_0^2 \int_0^1 \int_0^1 3y^2 dx dy dz = 2$. Here the boundary of the box is positively oriented considering outward normals.

3. Prove that the integral on the upper face of the box is $\int_U (\arctan y - x^2)dydz + (\cos x \sin z - y^3)dzdx + (2zx + 6zy^2)dxdy = \int_0^1 \int_0^1 4x + 12y^2 dxdy = 6.$

4. Prove that the integral on the open box is $\int_{\partial M\setminus U} (\arctan y - x^2)dydz + (\cos x \sin z - y^3)dzdx + (2zx + 6zy^2)dxdy = -4.$

Problem 13.14 Problems 1 through 3 refer to a triangular surface T in $\mathbb{R}^3$ and a differential form ω. The vertices of T are at $A(6,0,0)$, $B(0,12,0)$, and $C(0,0,3)$. The boundary of of the triangle ∂T is oriented positively by starting at A, continuing to B, following to C, and ending again at A. The surface T is oriented positively by considering the top of the triangle, as viewed from a point far above the triangle. The differential form is

$$\omega = (2xz + \arctan e^x)\,dx + \left(xz + (y+1)^y\right)dy + \left(xy + \frac{y^2}{2} + \log(1+z^2)\right)dz.$$

1. Prove that the equation of the plane that contains the triangle T is $2x + y + 4z = 12$.

2. Prove that $d\omega = ydy \wedge dz + (2x - y)\,dz \wedge dx + zdx \wedge dy$.

3. Prove that $\int_{\partial T} (2xz + \arctan e^x)\,dx + \left(xz + (y+1)^y\right)dy + \left(xy + \frac{y^2}{2} + \log(1+z^2)\right)dz = \int_0^3 \int_0^{12-4z} ydydz + \int_0^6 \int_0^{3-x/2} 2xdzdx = 108.$

Problem 13.15 Use Green's Theorem to prove that

$$\int_\Gamma (x^2 + 2y^3)dy = 16\pi,$$

where Γ is the circle $(x-2)^2 + y^2 = 4$. Also, prove this directly by using a path integral.

Problem 13.16 Let Γ denote the curve of intersection of the plane $x + y = 2$ and the sphere $x^2 - 2x + y^2 - 2y + z^2 = 0$, oriented clockwise when viewed from the origin. Use Stoke's Theorem to prove that

$$\int\limits_\Gamma ydx + zdy + xdz = -2\pi\sqrt{2}.$$

Prove this directly by parametrising the boundary of the surface and evaluating the path integral.

Problem 13.17 Use Green's Theorem to evaluate

$$\oint_C (x^3 - y^3)dx + (x^3 + y^3)dy,$$

where C is the positively oriented boundary of the region between the circles $x^2 + y^2 = 2$ and $x^2 + y^2 = 4$.

V

Appendix

CHAPTER 1

ofs of the Inverse and the Implicit Functions Theorems

A.1 Banach's Fixed Point Theorem

Definition A.1.1 Let M be a subset of $\mathbb{R}^n$. A function $f: M \to M$ is a **strict contraction** if there exists a constant $0 \leqslant \lambda < 1$ such that

$$\forall x, y \in M : d(f(x), f(y)) \leqslant \lambda d(x, y).$$

Theorem A.1.1 — Banach's Fixed Point. Let M be a closed subset[a] of $\mathbb{R}^n$ and let $f: M \to M$ be a **strict contraction**. Then f has a unique **fixed point**, which means that there is a unique $z \in M$ such that $f(z) = z$. Furthermore, if we start with a completely arbitrary point $y \in M$, then the sequence

$$y, f(y), f(f(y)), f(f(f(y))), \ldots$$

converges to z.

[a] Tthe Banach Fixed Point can be proved for a complete metric space (M, d)

Proof: We start proving the uniqueness of the fixed point.

Assume that x, y are fixed points. Then

$$d(x, y) = d(f(x), f(y)) \leqslant \lambda d(x, y) \Rightarrow (1 - \lambda) d(x, y) = 0.$$

Since $0 \leqslant \lambda < 1$, we have $d(x, y) = 0$ and so $x = y$. So we have the the uniqueness of the fixed point.

Now we prove the existence of the fixed point.

Let z_n be the sequence recursively defined by $z_0 := y$ and $z_{n+1} = f(z_n)$. This sequence $(z_n)_{n\in\mathbb{N}}$ is nothing else but the sequence $y, f(y), f(f(y)), f(f(f(y))), \ldots$.

Let $n \geqslant 0$. We claim that

$$d(z_{n+1}, z_n) \leqslant \lambda^n d(z_1, z_0).$$

Indeed, this follows by induction on n. The case $n = 0$ is trivial, and if the claim is true for n, then $d(z_{n+2}, z_{n+1}) = d(f(z_{n+1}), f(z_n)) \leqslant \lambda d(z_{n+1}, z_n) \leqslant \lambda \cdot \lambda^n d(z_1, z_0)$.

Hence, by the triangle inequality,

$$\begin{aligned} d(z_{n+m}, z_n) &\leqslant \sum_{j=n+1}^{n+m} d(z_j, z_{j-1}) && \text{(A.1)} \\ &\leqslant \sum_{j=n+1}^{n+m} \lambda^{j-1} d(z_1, z_0) && \text{(A.2)} \\ &\leqslant \sum_{j=n+1}^{\infty} \lambda^{j-1} d(z_1, z_0) && \text{(A.3)} \\ &= d(z_1, z_0)\lambda^n \frac{1}{1-\lambda} && \text{(A.4)} \end{aligned}$$

We note that the latter expression goes to zero as $n \to \infty$ and so z_n is a Cauchy sequence. Since M is a closed subset of $\mathbb{R}^n$, the sequence converges to a limit z. This limit is the fixed point:

$$x = \lim_{n\to\infty} z_n = \lim_{n\to\infty} f(z_{n-1}) = f(\lim_{n\to\infty} z_{n-1}) = f(z).$$

■

Lemma A.1.2 Let $g : \overline{B_r(0)} \to \overline{B_r(0)}$ be Lipschitz continuous function with Lipschitz constant less or equal $1/2$ such that $g(0) = 0$. Then the function

$$f : \overline{B_r(0)} \to \mathbb{R}^n, f(x) := g(x) + x$$

is injective and $B_{r/2}(0) \subseteq f(B_r(0))$.

Proof: First, we note that for $y \in B_{r/2}(0)$ the function

$$h : \overline{B_r(0)} \to \mathbb{R}^n, h(z) := y - g(z)$$

is a strict contraction. This fact follows from

$$\|y - g(z) - (y - g(z'))\| = \|g(z') - g(z)\| \leqslant \frac{1}{2}\|z - z'\|.$$

Furthermore, it maps the closed ball $\overline{B_r(0)}$ to itself, since for $z \in \overline{B_r(0)}$

$$\|y - g(z)\| \leqslant \|y\| + \|g(z - 0)\| \leqslant \frac{r}{2} + \frac{1}{2}\|z\| \leqslant r.$$

Hence, the Banach Fixed Point Theorem is applicable to the function h. Now the fact that the point x is a fixed point of the function h is equivalent to

$$f(x) = y,$$

and thus $B_{r/2}(0) \subseteq f(B_r(0))$ follows from the existence of fixed points. Furthermore, if $f(x) = f(x')$, then

$$\frac{1}{2}\|x - x'\| \geqslant \|g(x) - g(x')\| = \|f(x) - x - (f(x') - x')\| = \|x - x'\|$$

and so $x = x'$. So the function is injective. ■

A.2 The Inverse Function Theorem

Theorem A.2.1 Let $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function in a neighborhood $\mathbf{x}_0 \in \mathbb{R}^n$ such that $\mathbf{f}'(\mathbf{x}_0)$ is invertible. Then there exists an open set $U \subseteq \mathbb{R}^n$ with $\mathbf{x}_0 \in U$ such that $\mathbf{f}|_U$ is a bijective function with an inverse $\mathbf{f}^{-1}: \mathbf{f}(U) \to U$ which is differentiable at $\mathbf{x}_0$ and satisfies

$$(\mathbf{f}^{-1})'(\mathbf{x}_0) = (\mathbf{f}'(\mathbf{x}_0))^{-1}.$$

Proof: We first reduce the analysis to the case $\mathbf{f}(\mathbf{x}_0) = 0$, $\mathbf{x}_0 = 0$ and $\mathbf{f}'(\mathbf{x}_0) = I$. Indeed, suppose for all those functions the theorem holds, and let now h be an arbitrary function satisfying the requirements of the theorem (where the differentiability is given at $\mathbf{x}_0$). We set

$$\tilde{h}(x) := h'(\mathbf{x}_0)^{-1}(h(\mathbf{x}_0 - x) - h(\mathbf{x}_0))$$

and obtain that $\tilde{h}$ is differentiable at 0 with differential Id and $\tilde{h}(0) = 0$; the first property follows since we multiply both the function and the linear approximation by $h'(\mathbf{x}_0)^{-1}$ and only shift the function, and the second one is seen from inserting $x = 0$. Hence, we obtain an inverse of $\tilde{h}$ with it's differential at $\tilde{h}(0) = 0$, and if we now set

$$h^{-1}(y) := (\tilde{h}^{-1}(h'(\mathbf{x}_0)^{-1}(y - h(\mathbf{x}_0))) - \mathbf{x}_0),$$

it can be seen that h^{-1} is an inverse of h with all the required properties (which is a bit of a tedious exercise, but involves nothing more than the definitions).

Thus let $\mathbf{f}$ be a function such that $\mathbf{f}(0) = 0$, $\mathbf{f}$ is invertible at 0 and $\mathbf{f}'(0) = \mathrm{Id}$. We define

$$g(x) := \mathbf{f}(x) - x.$$

The differential of this function is zero. Since the function g is also continuously differentiable at a small neighborhood of 0, we find $\delta > 0$ such that

$$\frac{\partial g}{\partial x_j}(x) < \frac{1}{2n^2}$$

for all $j \in \{1, \ldots, n\}$ and $x \in B_\delta(0)$. Since further $g(0) = \mathbf{f}(0) - 0 = 0$, the general mean-value theorem and Cauchy's inequality imply that for $k \in \{1, \ldots, n\}$ and $x \in B_\delta(0)$,

$$|g_k(x)| = |\langle x, \frac{\partial g}{\partial x_j}(t_k x)\rangle| \leqslant \|x\| n \frac{1}{2n^2}$$

for suitable $t_k \in [0, 1]$. Hence,

$$\|g(x)\| \leqslant |g_1(x)| + \cdots + |g_n(x)| \leqslant \frac{1}{2}\|x\|$$

and thus, we note that the lemma is applicable, and $\mathbf{f}$ is a bijection on $\overline{B_\delta(0)}$, whose image is contained within the open set $\overline{B_{\delta/2}(0)}$. So we may pick $U := \mathbf{f}^{-1}(B_{\delta/2}(0))$, which is open due to the continuity of f.

Thus, the most important part of the theorem is already done. All that is left to do is to prove differentiability of $\mathbf{f}^{-1}$ at 0. Now we even prove the slightly stronger claim that the differential of $\mathbf{f}^{-1}$ at $\mathbf{x}_0$ is given by the identity, although this would also follow from the chain rule once differentiability is proven.

Note now that the contraction identity for g implies the following bounds on $\mathbf{f}$:

$$\frac{1}{2}\|x\| \leqslant \|\mathbf{f}(x)\| \leqslant \frac{3}{2}\|x\|.$$

The second bound follows from

$$\|\mathbf{f}(x)\| \leqslant \|\mathbf{f}(x) - x\| + \|x\| = \|g(x)\| + \|x\| \leqslant \frac{3}{2}\|x\|,$$

and the first bound follows from

$$\|\mathbf{f}(x)\| \geqslant |\|\mathbf{f}(x) - x\| - \|x\|| = \big|\|g(x)\| - \|x\|\big| \geqslant \frac{1}{2}\|x\|.$$

Now for the differentiability at 0. We have, by substitution of limits (as f is continuous and $f(0) = 0$):

$$\lim_{\mathbf{h}\to 0} \frac{\|\mathbf{f}^{-1}(\mathbf{h}) - \mathbf{f}^{-1}(0) - \mathrm{Id}(\mathbf{h} - 0)\|}{\|\mathbf{h}\|} = \lim_{\mathbf{h}\to 0} \frac{\|\mathbf{f}^{-1}(\mathbf{f}(\mathbf{h})) - \mathbf{f}(\mathbf{h})\|}{\|\mathbf{f}(\mathbf{h})\|} \tag{A.5}$$

$$= \lim_{\mathbf{h}\to 0} \frac{\|\mathbf{h} - f(\mathbf{h})\|}{\|f(\mathbf{h})\|}, \tag{A.6}$$

where the last expression converges to zero due to the differentiability of **f** at 0 with differential the identity, and the sandwhich criterion applied to the expressions

$$\frac{\|\mathbf{h} - \mathbf{f}(\mathbf{h})\|}{\frac{3}{2}\|\mathbf{h}\|}$$

and

$$\frac{\|\mathbf{h} - f(\mathbf{h})\|}{\frac{1}{2}\|\mathbf{h}\|}.$$

■

A.3 The Implicit Function Theorem

Theorem A.3.1 Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and consider the set

$$S := \{(x_1, \ldots, x_n) \in \mathbb{R}^n | \mathbf{f}(x_1, \ldots, x_n) = 0\}.$$

If we are given some $y \in S$ such that $\partial_n \mathbf{f}(y) \neq 0$, then we find $U \subseteq \mathbb{R}^{n-1}$ open with $(y_1, \ldots, y_{n-1}) \in U$ and $g : U \to S$ such that

$$y = g(y_1, \ldots, y_{n-1})$$

and $\{(z_1, \ldots, z_{n-1}, g(z_1, \ldots, z_{n-1})) | (z_1, \ldots, z_{n-1}) \in U\} \subseteq S$,

Furthermore, g is a differentiable function.

Proof:

We define a new function

$$F : \mathbb{R}^n \to \mathbb{R}^n, F(x_1, \ldots, x_n) := (x_1, \ldots, x_{n-1}, f(x_1, \ldots, x_n)).$$

The differential of this function looks like this:

$$F'(x) = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ 0 & \cdots & 0 & 1 & 0 \\ \partial_1 f(x) & & \cdots & & \partial_n f(x) \end{pmatrix}$$

Since we assumed that $\partial_n f(y) \neq 0$, $F'(y)$ is invertible, and hence the Inverse Function Theorem implies the existence of a small open neighbourhood $\tilde{V} \subseteq \mathbb{R}^n$ containing y such that *restricted to that neighbourhood* F is invertible, with a differentiable inverse F^{-1}, which is itself defined on an open set $\tilde{U}$ containing $F(y)$. Now set first

$$U := \{(x_1, \ldots, x_{n-1}) | (x_1, \ldots, x_{n-1}, 0) \in \tilde{U}\},$$

) and then

$$g : U \to \mathbb{R}, g(x_1, \ldots, x_{n-1}) := \pi_n(F^{-1}(x_1, \ldots, x_{n-1}, 0)),$$

the n-th component of $F^{-1}(x_1, \ldots, x_{n-1}, 0)$. We claim that g has the desired properties.

Indeed, we first note that $F^{-1}(x_1, \ldots, x_{n-1}, 0) = (x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1}))$, since applying F leaves the first $n-1$ components unchanged, and thus we get the identity by observing $F(F^{-1}(x)) = x$. Let thus $(z_1, \ldots, z_{n-1}) \in U$. Then

$$f(z_1, \ldots, z_{n-1}, g(z_1, \ldots, z_{n-1})) = (\pi_n \circ F)(F^{-1}(z_1, \ldots, z_{n-1}, 0)) \tag{A.7}$$

$$= \pi_n((F \circ F^{-1})(z_1, \ldots, z_{n-1}, 0)) = 0. \tag{A.8}$$

Furthermore, the set

$$\{(z_1, \ldots, z_{n-1}, g(z_1, \ldots, z_{n-1})) | (z_1, \ldots, z_{n-1}) \in U\} = S \cap \tilde{V}.$$

For $\subseteq$, we first note that the set on the left hand side is in S, since all points in it are mapped to zero by f. Further,

$$F(z_1, \ldots, z_{n-1}, g(z_1, \ldots, z_{n-1})) = (z_1, \ldots, z_{n-1}, 0) \in \tilde{U}$$

and hence $\subseteq$ is completed when applying F^{-1}. For the other direction, let a point $(x_1, \ldots, x_n)$ in $S \cap \tilde{V}$ be given, apply F to get

$$F((x_1, \ldots, x_n)) = (x_1, \ldots, x_{n-1}, 0) \in \tilde{U}$$

and hence $(x_1, \ldots, x_{n-1}) \in U$; further

$$(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) = (x_1, \ldots, x_n)$$

by applying F to both sides of the equation.

Now g is automatically differentiable since g is the component of a differentiable function.

■

Chapter 2

Determinants

B.1 Permutations

Definition B.1.1 Let S be a finite set with $n \geqslant 1$ elements. A **permutation** is a bijective function $\tau : S \to S$. It is easy to see that there are $n!$ permutations from S onto itself.

Since we are mostly concerned with the *action* that τ exerts on S rather than with the particular names of the elements of S, we will take S to be the set $S = \{1, 2, 3, \dots, n\}$. We indicate a permutation τ by means of the following convenient diagram

$$\tau = \begin{bmatrix} 1 & 2 & \cdots & n \\ \tau(1) & \tau(2) & \cdots & \tau(n) \end{bmatrix}.$$

Definition B.1.2 The notation S_n will denote the set of all permutations on $\{1,2,3,\ldots,n\}$. Under this notation, the composition of two permutations $(\tau,\sigma)\in S_n^2$ is

$$\begin{aligned}\tau\circ\sigma &= \begin{bmatrix}1 & 2 & \cdots & n\\ \tau(1) & \tau(2) & \cdots & \tau(n)\end{bmatrix}\circ\begin{bmatrix}1 & 2 & \cdots & n\\ \sigma(1) & \sigma(2) & \cdots & \sigma(n)\end{bmatrix}\\ &= \begin{bmatrix}1 & 2 & \cdots & n\\ (\tau\circ\sigma)(1) & (\tau\circ\sigma)(2) & \cdots & (\tau\circ\sigma)(n)\end{bmatrix}\end{aligned}.$$

The k-fold composition of τ is

$$\underbrace{\tau\circ\cdots\circ\tau}_{k \text{ compositions}} = \tau^k.$$

We usually do away with the $\circ$ and write $\tau\circ\sigma$ simply as $\tau\sigma$. This "product of permutations" is thus simply function composition.

Given a permutation $\tau: S\to S$, since τ is bijective,

$$\tau^{-1}: S\to S$$

exists and is also a permutation. In fact if

$$\tau = \begin{bmatrix}1 & 2 & \cdots & n\\ \tau(1) & \tau(2) & \cdots & \tau(n)\end{bmatrix},$$

then

$$\tau^{-1} = \begin{bmatrix}\tau(1) & \tau(2) & \cdots & \tau(n)\\ 1 & 2 & \cdots & n\end{bmatrix}.$$

Example B.1.1 The set S_3 has $3! = 6$ elements, which are given below.

1. $\mathbf{Id} : \{1,2,3\}\to\{1,2,3\}$ with

$$\mathbf{Id} = \begin{bmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{bmatrix}.$$

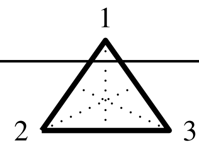


Figure B.1 S_3 are rotations and reflexions.

2. $\tau_1 : \{1,2,3\} \to \{1,2,3\}$ with
$$\tau_1 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}.$$
3. $\tau_2 : \{1,2,3\} \to \{1,2,3\}$ with
$$\tau_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$
4. $\tau_3 : \{1,2,3\} \to \{1,2,3\}$ with
$$\tau_3 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}.$$
5. $\sigma_1 : \{1,2,3\} \to \{1,2,3\}$ with
$$\sigma_1 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}.$$
6. $\sigma_2 : \{1,2,3\} \to \{1,2,3\}$ with
$$\sigma_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Example B.1.2 The compositions $\tau_1 \circ \sigma_1$ and $\sigma_1 \circ \tau_1$ can be found as follows.
$$\tau_1 \circ \sigma_1 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} \circ \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \tau_2.$$
(We read from right to left $1 \to 2 \to 3$ (“1 goes to 2, 2 goes to 3, so 1 goes to 3”), etc. Similarly
$$\sigma_1 \circ \tau_1 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} = \tau_3.$$

Observe in particular that $\sigma_1 \circ \tau_1 \neq \tau_1 \circ \sigma_1$. Finding all the other products we deduce the following "multiplication table" (where the "multiplication" operation is really composition of functions).

$\circ$	$\mathbf{Id}$	τ_1	τ_2	τ_3	σ_1	σ_2
$\mathbf{Id}$	$\mathbf{Id}$	τ_1	τ_2	τ_3	σ_1	σ_2
τ_1	τ_1	$\mathbf{Id}$	σ_1	σ_2	τ_2	τ_3
τ_2	τ_2	σ_2	$\mathbf{Id}$	σ_1	τ_3	τ_1
τ_3	τ_3	σ_1	σ_2	$\mathbf{Id}$	τ_1	τ_2
σ_2	σ_2	τ_2	τ_3	τ_1	$\mathbf{Id}$	σ_1
σ_1	σ_1	τ_3	τ_1	τ_2	σ_2	$\mathbf{Id}$

The permutations in example B.1.1 can be conveniently interpreted as follows. Consider an equilateral triangle with vertices labelled 1, 2 and 3, as in figure B.1. Each τ_a is a reflexion ("flipping") about the line joining the vertex a with the midpoint of the side opposite a. For example τ_1 fixes 1 and flips 2 and 3. Observe that two successive flips return the vertices to their original position and so $(\forall a \in \{1,2,3\})(\tau_a^2 = \mathbf{Id}\ )$. Similarly, σ_1 is a rotation of the vertices by an angle of $120°$. Three successive rotations restore the vertices to their original position and so $\sigma_1^3 = \mathbf{Id}$.

Example B.1.3 To find τ_1^{-1} take the representation of τ_1 and exchange the rows:

$$\tau_1^{-1} = \begin{bmatrix} 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

This is more naturally written as

$$\tau_1^{-1} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}.$$

Observe that $\tau_1^{-1} = \tau_1$.

Example B.1.4 To find σ_1^{-1} take the representation of σ_1 and exchange the rows:

$$\sigma_1^{-1} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

This is more naturally written as

$$\sigma_1^{-1} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Observe that $\sigma_1^{-1} = \sigma_2$.

B.2 Cycle Notation

We now present a shorthand notation for permutations by introducing the idea of a *cycle*. Consider in S_9 the permutation

$$\tau = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 1 & 3 & 6 & 9 & 7 & 8 & 4 & 5 \end{bmatrix}.$$

We start with 1. Since 1 goes to 2 and 2 goes back to 1, we write (12). Now we continue with 3. Since 3 goes to 3, we write (3). We continue with 4. As 4 goes 6, 6 goes to 7, 7 goes 8, and 8 goes back to 4, we write (4678). We consider now 5 which goes to 9 and 9 goes back to 5, so we write (59). We have written τ as a product of disjoint cycles

$$\tau = (12)(3)(4678)(59).$$

This prompts the following definition.

Definition B.2.1 Let $l \geqslant 1$ and let $i_1, \dots, i_l \in \{1, 2, \dots n\}$ be distinct. We write $(i_1\ i_2\ \dots\ i_l)$ for the element $\sigma \in S_n$ such that $\sigma(i_r) = i_{r+1}$, $1 \leqslant r < l$, $\sigma(i_l) = i_1$ and $\sigma(i) = i$ for $i \notin \{i_1, \dots, i_l\}$. We say that $(i_1\ i_2\ \dots\ i_l)$ is a **cycle of length** l. The **order** of a cycle is its length. Observe that if τ has order l then $\tau^l = \mathbf{Id}$.

Observe that $(i_2\ \dots\ i_l\ i_1) = (i_1\ \dots\ i_l)$ etc., and that $(1) = (2) = \cdots = (n) = \mathbf{Id}$. In fact, we have

$$(i_1\ \dots\ i_l) = (j_1\ \dots\ j_m)$$

if and only if (1) $l = m$ and if (2) $l > 1$: $\exists a$ such that $\forall k$: $i_k = j_{k+a \mod l}$. Two cycles $(i_1, \dots, i_l)$ and $(j_1, \dots, j_m)$ are disjoint if $\{i_1, \dots, i_l\} \cap \{j_1, \dots, j_m\} = \varnothing$. Disjoint cycles commute and if $\tau = \sigma_1 \sigma_2 \cdots \sigma_t$ is the product of disjoint cycles of length $l_1, l_2, \dots, l_t$ respectively, then τ has order

$$\mathrm{lcm}(l_1, l_2, \dots, l_t).$$

Example B.2.1 A cycle decomposition for $\alpha \in S_9$,

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 8 & 7 & 6 & 2 & 3 & 4 & 5 & 9 \end{bmatrix}$$

is

$$(285)(3746).$$

The order of α is $\mathrm{lcm}(3, 4) = 12$.

Example B.2.2 The cycle decomposition $\beta = (123)(567)$ in S_9 arises from the permutation

$$\beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 1 & 4 & 6 & 7 & 5 & 8 & 9 \end{bmatrix}.$$

Its order is $\mathrm{lcm}(3,3) = 3$.

Example B.2.3 Find a shuffle of a deck of 13 cards that requires 42 repeats to return the cards to their original order.

Solution: ▶ Here is one (of many possible ones). Observe that $7+6 = 13$ and $7 \times 6 = 42$. We take the permutation

$$(1\ 2\ 3\ 4\ 5\ 6\ 7)(8\ 9\ 10\ 11\ 12\ 13)$$

which has order 42. This corresponds to the following shuffle: For

$$i \in \{1,2,3,4,5,6,8,9,10,11,12\},$$

take the ith card to the $(i+1)$th place, take the 7th card to the first position and the 13th card to the 8th position. Query: Of all possible shuffles of 13 cards, which one takes the longest to restitute the cards to their original position? ◀

Example B.2.4 Let a shuffle of a deck of 10 cards be made as follows: The top card is put at the bottom, the deck is cut in half, the bottom half is placed on top of the top half, and then the resulting bottom card is put on top. How many times must this shuffle be repeated to get the cards in the initial order? Explain.

Solution: ▶ Putting the top card at the bottom corresponds to

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 1 \end{bmatrix}.$$

Cutting this new arrangement in half and putting the lower half on top corresponds to

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 8 & 9 & 10 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}.$$

Putting the bottom card of this new arrangement on top corresponds to

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 8 & 9 & 10 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} = (1\ 6)(2\ 7)(3\ 8)(4\ 9)(5\ 10).$$

The order of this permutation is $\mathrm{lcm}(2,2,2,2,2) = 2$, so in 2 shuffles the cards are restored to their original position. ◀ The above examples illustrate the general case, given in the following theorem.

Theorem B.2.5 Every permutation in S_n can be written as a product of disjoint cycles.

Proof: Let $\tau \in S_n, a_1 \in \{1,2,\ldots,n\}$. Put $\tau^k(a_1) = a_{k+1}, k \geqslant 0$. Let $a_1, a_2, \ldots, a_s$ be the longest chain with no repeats. Then we have $\tau(a_s) = a_1$. If the $\{a_1, a_2, \ldots, a_s\}$ exhaust $\{1,2,\ldots,n\}$ then we have $\tau = (a_1\ a_2\ \ldots\ a_s)$. If not, there exist $b_1 \in \{1,2,\ldots,n\} \setminus \{a_1, a_2, \ldots, a_s\}$. Again, we find the longest chain of distinct $b_1, b_2, \ldots, b_t$ such that $\tau(b_k) = b_{k+1}, k = 1, \ldots, t-1$ and $\tau(b_t) = b_1$. If the $\{a_1, a_2, \ldots, a_s, b_1, b_2, \ldots, b_t\}$ exhaust all the $\{1,2,\ldots,n\}$ we have $\tau = (a_1\ a_2\ \ldots\ a_s)(b_1\ b_2\ \ldots\ b_t)$. If not we continue the process and find

$$\tau = (a_1\ a_2\ \ldots\ a_s)(b_1\ b_2\ \ldots\ b_t)(c_1 \ldots)\ldots.$$

This process stops because we have only n elements. ■

Definition B.2.2 A **transposition** is a cycle of length 2.[a]

[a] A cycle of length 2 should more appropriately be called a **bicycle**.

Example B.2.6 The cycle (23468) can be written as a product of transpositions as follows

$$(23468) = (28)(26)(24)(23).$$

Notice that this decomposition as the product of transpositions is not unique. Another decomposition is

$$(23468) = (23)(34)(46)(68).$$

Lemma B.2.7 Every permutation is the product of transpositions.

Proof: It is enough to observe that

$$(a_1\ a_2\ \ldots\ a_s) = (a_1\ a_s)(a_1\ a_{s-1})\cdots(a_1\ a_2)$$

and appeal to Theorem B.2.5. ■ Let $\sigma \in S_n$ and let $(i,j) \in \{1,2,\ldots,n\}^2,\ \ i \neq j$. Since σ is a permutation, $\exists (a,b) \in \{1,2,\ldots,n\}^2,\ \ a \neq b$, such that $\sigma(j) - \sigma(i) = b - a$. This means that

$$\left| \prod_{1 \leqslant i < j \leqslant n} \frac{\sigma(i) - \sigma(j)}{i - j} \right| = 1.$$

Definition B.2.3 Let $\sigma \in S_n$. We define the **sign** $\operatorname{sgn}\sigma$ of σ as

$$\operatorname{sgn}(\sigma) = \prod_{1 \leqslant i < j \leqslant n} \frac{\sigma(i) - \sigma(j)}{i - j} = (-1)^{\sigma}.$$

If $\operatorname{sgn}(\sigma) = 1$, then we say that σ is an **even permutation**, and if $\operatorname{sgn}(\sigma) = -1$ we say that σ is an **odd permutation**.

Notice that in fact

$$\operatorname{sgn}(\sigma) = (-1)^{\mathbf{I}(\sigma)},$$

where $\mathbf{I}(\sigma) = \#\{(i, j) \,|\, 1 \leqslant i < j \leqslant n \text{ and } \sigma(i) > \sigma(j)\}$, i.e., $\mathbf{I}(\sigma)$ is the number of inversions that σ effects to the identity permutation $\mathbf{Id}$.

Example B.2.8 The transposition $(1\ 2)$ has one inversion.

Lemma B.2.9 For any transposition $(k\ l)$ we have $\operatorname{sgn}\left((k\ l)\right) = -1$.

Proof: Let τ be transposition that exchanges k and l, and assume that $k < l$:

$$\tau = \begin{bmatrix} 1 & \dots & k-1 & k & \dots & l-1 & l & l+1 & \dots & n \\ 1 & \dots & k-1 & l & \dots & l-1 & k & l+1 & \dots & n \end{bmatrix}$$

Let us count the number of inversions of τ:

- The pairs (i, j) with $i \in \{1, 2, \dots, k-1\} \cup \{l, l+1, \dots, n\}$ and $i < j$ do not suffer an inversion;
- The pair (k, j) with $k < j$ suffers an inversion if and only if $j \in \{k+1, k+2, \dots, l\}$, making $l - k$ inversions;
- If $i \in \{k+1, k+2, \dots, l-1\}$ and $i < j$, (i, j) suffers an inversion if and only if $j = l$, giving $l - 1 - k$ inversions.

This gives a total of $\mathbf{I}(\tau) = (l - k) + (l - 1 - k) = 2(l - k - 1) + 1$ inversions when $k < l$. Since this number is odd, we have $\operatorname{sgn}(\tau) = (-1)^{\mathbf{I}(\tau)} = -1$. In general we see that the transposition $(k\ l)$ has $2|k - l| - 1$ inversions. ■

Theorem B.2.10 Let $(\sigma, \tau) \in S_n^2$. Then

$$\operatorname{sgn}(\tau\sigma) = \operatorname{sgn}(\tau) \operatorname{sgn}(\sigma).$$

Proof: We have

$$\begin{aligned} \operatorname{sgn}(\sigma\tau) &= \prod_{1\leqslant i<j\leqslant n} \frac{(\sigma\tau)(i)-(\sigma\tau)(j)}{i-j} \\ &= \left(\prod_{1\leqslant i<j\leqslant n} \frac{\sigma(\tau(i))-\sigma(\tau(j))}{\tau(i)-\tau(j)}\right)\cdot\left(\prod_{1\leqslant i<j\leqslant n} \frac{\tau(i)-\tau(j)}{i-j}\right). \end{aligned}$$

The second factor on this last equality is clearly $\operatorname{sgn}(\tau)$, we must shew that the first factor is $\operatorname{sgn}(\sigma)$. Observe now that for $1 \leqslant a < b \leqslant n$ we have

$$\frac{\sigma(a)-\sigma(b)}{a-b} = \frac{\sigma(b)-\sigma(a)}{b-a}.$$

Since σ and τ are permutations, $\exists b \neq a$, $\tau(i) = a, \tau(j) = b$ and so $\sigma\tau(i) = \sigma(a), \sigma\tau(j) = b$. Thus

$$\frac{\sigma(\tau(i))-\sigma(\tau(j))}{\tau(i)-\tau(j)} = \frac{\sigma(a)-\sigma(b)}{a-b}$$

and so

$$\prod_{1\leqslant i<j\leqslant n} \frac{\sigma(\tau(i))-\sigma(\tau(j))}{\tau(i)-\tau(j)} = \prod_{1\leqslant a<b\leqslant n} \frac{\sigma(a)-\sigma(b)}{a-b} = \operatorname{sgn}(\sigma).$$

■

Corollary B.2.11 The identity permutation is even. If $\tau \in S_n$, then $\operatorname{sgn}(\tau) = \operatorname{sgn}\left(\tau^{-1}\right)$.

Proof: Since there are no inversions in $\mathbf{Id}$, we have $\operatorname{sgn}(\mathbf{Id}) = (-1)^0 = 1$. Since $\tau\tau^{-1} = \mathbf{Id}$, we must have $1 = \operatorname{sgn}(\mathbf{Id}) = \operatorname{sgn}\left(\tau\tau^{-1}\right) = \operatorname{sgn}(\tau)\operatorname{sgn}\left(\tau^{-1}\right) = (-1)^{\tau}(-1)^{\tau^{-1}}$ by Theorem B.2.10. Since the values on the righthand of this last equality are ± 1, we must have $\operatorname{sgn}(\tau) = \operatorname{sgn}\left(\tau^{-1}\right)$.
■

Lemma B.2.12 We have $\operatorname{sgn}\left(1\ 2\ \ldots\ l\right) = (-1)^{l-1}$.

Proof: Simply observe that the number of inversions of $(1\ 2\ \ldots\ l)$ is $l-1$. ■

Lemma B.2.13 Let $(\tau, (i_1\ \ldots\ i_l) \in S_n^2$. Then

$$\tau(i_1\ \ldots\ i_l)\tau^{-1} = (\tau(i_1)\ \ldots\ \tau(i_l)),$$

and if $\sigma \in S_n$ is a cycle of length l then

$$\operatorname{sgn}(\sigma) = (-1)^{l-1}$$

.

Proof: For $1 \leqslant k < l$ we have $(\tau(i_1\ \dots\ i_l)\tau^{-1})(\tau(i_k)) = \tau((i_1\ \dots\ i_l)(i_k)) = \tau(i_{k+1})$. On a $(\tau(i_1\ \dots\ i_l)\tau^{-1})(\tau(i_l)) = \tau((i_1\ \dots\ i_l)(i_l)) = \tau(i_1)$. For $i \not\in \{\tau(i_1)\ \dots\ \tau(i_l)\}$ we have $\tau^{-1}(i) \not\in \{i_1\ \dots\ i_l\}$ whence $(i_1\ \dots\ i_l)(\tau^{-1}(i)) = \tau^{-1}(i)$ etc.

Furthermore, write $\sigma = (i_1\ \dots\ i_l)$. Let $\tau \in S_n$ be such that $\tau(k) = i_k$ for $1 \leqslant k \leqslant l$. Then $\sigma = \tau(1\ 2\ \dots\ l)\tau^{-1}$ and so we must have $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}\big((1\ 2\ \dots\ l)\big)\operatorname{sgn}\left(\tau^{-1}\right)$, which equals $\operatorname{sgn}\big((1\ 2\ \dots\ l)\big)$ by virtue of Theorem B.2.10 and Corollary B.2.11. The result now follows by appealing to Lemma B.2.12 ■

Corollary B.2.14 Let $\sigma = \sigma_1\sigma_2\cdots\sigma_r$ be a product of disjoint cycles, each of length $l_1, \dots, l_r$, respectively. Then

$$\operatorname{sgn}(\sigma) = (-1)^{\sum_{i=1}^{r}(l_i - 1)}.$$

Hence, the product of two even permutations is even, the product of two odd permutations is even, and the product of an even permutation and an odd permutation is odd.

Proof: This follows at once from Theorem B.2.10 and Lemma B.2.13. ■

Example B.2.15 The cycle (4678) is an odd cycle; the cycle (1) is an even cycle; the cycle (12345) is an even cycle.

Corollary B.2.16 Every permutation can be decomposed as a product of transpositions. This decomposition is not necessarily unique, but its parity is unique.

Proof: This follows from Theorem B.2.5, Lemma B.2.7, and Corollary B.2.14. ■

Example B.2.17 — The 15 puzzle. Consider a grid with 16 squares, as shewn in (B.1), where 15 squares are numbered 1 through 15 and the 16th slot is empty.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

(B.1)

In this grid we may successively exchange the empty slot with any of its neighbours, as for example

1	2	3	4
5	6	7	8
9	10	11	12
13	14		15

. (B.2)

We ask whether through a series of valid moves we may arrive at the following position.

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

(B.3)

Solution: ► Let us shew that this is impossible. Each time we move a square to the empty position, we make transpositions on the set $\{1,2,\ldots,16\}$. Thus at each move, the permutation is multiplied by a transposition and hence it changes sign. Observe that the permutation corresponding to the square in (B.3) is $(14\ 15)$ (the positions 14th and 15th are transposed) and hence it is an odd permutation. But we claim that the empty slot can only return to its original position after an even permutation. To see this paint the grid as a checkerboard:

B	R	B	R
R	B	R	B
B	R	B	R
R	B	R	B

(B.4)

We see that after each move, the empty square changes from black to red, and thus after an odd number of moves the empty slot is on a red square. Thus the empty slot cannot return to its original position in an odd number of moves. This completes the proof. ◄

Homework

Remark B.2.18 Decompose the permutation

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 4 & 1 & 5 & 8 & 6 & 7 & 9 \end{bmatrix}$$

as a product of disjoint cycles and find its order.

B.3 Determinants

There are many ways of developing the theory of determinants. We will choose a way that will allow us to deduce the properties of determinants with ease, but has the drawback of being computationally cumbersome. In the next section we will shew that our way of defining determinants is equivalent to a more computationally friendly one.

It may be pertinent here to quickly review some properties of permutations. Recall that if $\sigma \in S_n$ is a cycle of length l, then its signum $\operatorname{sgn}(\sigma) = \pm 1$ depending on the parity of $l-1$. For example, in S_7,

$$\sigma = (1\ 3\ 4\ 7\ 6)$$

has length 5, and the parity of $5-1=4$ is even, and so we write $\operatorname{sgn}(\sigma) = +1$. On the other hand,

$$\tau = (1\ 3\ 4\ 7\ 6\ 5)$$

has length 6, and the parity of $6-1=5$ is odd, and so we write $\operatorname{sgn}(\tau) = -1$.

Recall also that if $(\sigma, \tau) \in S_n^2$, then

$$\operatorname{sgn}(\tau\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma).$$

Thus from the above two examples

$$\sigma\tau = (1\ 3\ 4\ 7\ 6)(1\ 3\ 4\ 7\ 6\ 5)$$

has signum $\operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) = (+1)(-1) = -1$. Observe in particular that for the identity permutation $\mathbf{Id} \in S_n$ we have $\operatorname{sgn}(\mathbf{Id}) = +1$.

Definition B.3.1 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ be a square matrix. The **determinant of** A is defined and denoted by the sum

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

The determinantal sum has $n!$ summands.

Example B.3.1 If $n = 1$, then S_1 has only one member, $\mathbf{Id}$, where $\mathbf{Id}(1) = 1$. Since $\mathbf{Id}$ is an even permutation, $\operatorname{sgn}(\mathbf{Id}) = (+1)$ Thus if $A = (a_{11})$, then

$$\det A = a_{11}$$

.

Example B.3.2 If $n = 2$, then S_2 has $2! = 2$ members, $\mathbf{Id}$ and $\sigma = (1\ 2)$. Observe that $\operatorname{sgn}(\sigma) = -1$. Thus if

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then

$$\det A = \operatorname{sgn}(\mathbf{Id})\, a_{1\mathbf{Id}(1)} a_{2\mathbf{Id}(2)} + \operatorname{sgn}(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} = a_{11}a_{22} - a_{12}a_{21}.$$

Example B.3.3 From the above formula for 2×2 matrices it follows that

$$\begin{aligned} \det A &= \det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\ &= (1)(4) - (3)(2) = -2, \\ \det B &= \det \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix} (-1)(4) - (3)(2) \\ &= -10, \end{aligned}$$

and

$$\det(A + B) = \det \begin{bmatrix} 0 & 4 \\ 6 & 8 \end{bmatrix} = (0)(8) - (6)(4) = -24.$$

Observe in particular that $\det(A + B) \neq \det A + \det B$.

Example B.3.4 If $n = 3$, then S_2 has $3! = 6$ members:

$$\mathbf{Id}, \tau_1 = (2\ 3), \tau_2 = (1\ 3), \tau_3 = (1\ 2), \sigma_1 = (1\ 2\ 3), \sigma_2 = (1\ 3\ 2).$$

. Observe that $\mathbf{Id}, \sigma_1, \sigma_2$ are even, and τ_1, τ_2, τ_3 are odd. Thus if

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

then

$$\begin{aligned} \det A &= \operatorname{sgn}(\mathbf{Id})\, a_{1\mathbf{Id}(1)} a_{2\mathbf{Id}(2)} a_{3\mathbf{Id}(3)} + \operatorname{sgn}(\tau_1)\, a_{1\tau_1(1)} a_{2\tau_1(2)} a_{3\tau_1(3)} \\ &\quad + \operatorname{sgn}(\tau_2)\, a_{1\tau_2(1)} a_{2\tau_2(2)} a_{3\tau_2(3)} + \operatorname{sgn}(\tau_3)\, a_{1\tau_3(1)} a_{2\tau_3(2)} a_{3\tau_3(3)} \\ &\quad + \operatorname{sgn}(\sigma_1)\, a_{1\sigma_1(1)} a_{2\sigma_1(2)} a_{3\sigma_1(3)} + \operatorname{sgn}(\sigma_2)\, a_{1\sigma_2(1)} a_{2\sigma_2(2)} a_{3\sigma_2(3)} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} \\ &\quad - a_{13}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}. \end{aligned}$$

Theorem B.3.5 — Row-Alternancy of Determinants. Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$. If $B \in \mathbf{M}_{n\times n}(\mathbb{R}), B = [b_{ij}]$ is the matrix obtained by interchanging the s-th row of A with its t-th row, then $\det B = -\det A$.

Proof: Let τ be the transposition

$$\tau = \begin{bmatrix} s & t \\ \tau(t) & \tau(s) \end{bmatrix}.$$

Then $\sigma\tau(a) = \sigma(a)$ for $a \in \{1, 2, \ldots, n\} \setminus \{s, t\}$. Also, $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) = -\operatorname{sgn}(\sigma)$. As σ ranges through all permutations of S_n, so does $\sigma\tau$, hence

$$\begin{aligned}
\det B &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\, b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{s\sigma(s)} \cdots b_{t\sigma(t)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{st} \cdots a_{ts} \cdots a_{n\sigma(n)} \\
&= -\sum_{\sigma\in S_n} \operatorname{sgn}(\sigma\tau)\, a_{1\sigma\tau(1)} a_{2\sigma\tau(2)} \cdots a_{s\sigma\tau(s)} \cdots a_{t\sigma\tau(t)} \cdots a_{n\sigma\tau(n)} \\
&= -\sum_{\lambda\in S_n} \operatorname{sgn}(\lambda)\, a_{1\lambda(1)} a_{2\lambda(2)} \cdots a_{n\lambda(n)} \\
&= -\det A.
\end{aligned}$$

■

Corollary B.3.6 If $A_{(r:k)}, 1 \leqslant k \leqslant n$ denote the rows of A and $\sigma \in S_n$, then

$$\det \begin{bmatrix} A_{(r:\sigma(1))} \\ A_{(r:\sigma(2))} \\ \vdots \\ A_{(r:\sigma(n))} \end{bmatrix} = (\operatorname{sgn}(\sigma)) \det A.$$

An analogous result holds for columns.

Proof: Apply the result of Theorem B.3.5 multiple times. ■

Theorem B.3.7 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$. Then

$$\det A^T = \det A.$$

Proof: Let $C = A^T$. By definition

$$\begin{aligned}
\det A^T &= \det C \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}.
\end{aligned}$$

But the product $a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$ also appears in $\det A$ with the same signum $\operatorname{sgn}(\sigma)$, since the permutation

$$\begin{bmatrix} \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ 1 & 2 & \cdots & n \end{bmatrix}$$

is the inverse of the permutation

$$\begin{bmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{bmatrix}.$$

■

Corollary B.3.8 — Column-Alternancy of Determinants. Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$. If $C \in \mathbf{M}_{n\times n}(\mathbb{R}), C = [c_{ij}]$ is the matrix obtained by interchanging the s-th column of A with its t-th column, then $\det C = -\det A$.

Proof: This follows upon combining Theorem B.3.5 and Theorem B.3.7. ■

Theorem B.3.9 — Row Homogeneity of Determinants. Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ and $\alpha \in \mathbb{R}$. If $B \in \mathbf{M}_{n\times n}(\mathbb{R}), B = [b_{ij}]$ is the matrix obtained by multiplying the s-th row of A by α, then

$$\det B = \alpha \det A.$$

Proof: Simply observe that

$$\operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots \alpha a_{s\sigma(s)} \cdots a_{n\sigma(n)} = \alpha \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{s\sigma(s)} \cdots a_{n\sigma(n)}.$$

■

Corollary B.3.10 — Column Homogeneity of Determinants. If $C \in \mathbf{M}_{n \times n}(\mathbb{R}), C = (C_{ij})$ is the matrix obtained by multiplying the s-th column of A by α, then

$$\det C = \alpha \det A.$$

Proof: This follows upon using Theorem B.3.7 and Theorem B.3.9. ■

It follows from Theorem B.3.9 and Corollary B.3.10 that if a row (or column) of a matrix consists of $0_{\mathbb{R}}$s only, then the determinant of this matrix is $0_{\mathbb{R}}$.

Example B.3.11

$$\det \begin{bmatrix} x & 1 & a \\ x^2 & 1 & b \\ x^3 & 1 & c \end{bmatrix} = x \det \begin{bmatrix} 1 & 1 & a \\ x & 1 & b \\ x^2 & 1 & c \end{bmatrix}.$$

Corollary B.3.12

$$\det(\alpha A) = \alpha^n \det A.$$

Proof: Since there are n columns, we are able to pull out one factor of α from each one. ■

Example B.3.13 Recall that a matrix A is **skew-symmetric** if $A = -A^T$. Let $A \in \mathbf{M}_{2001}(\mathbb{R})$ be skew-symmetric. Prove that $\det A = 0$.

Solution: ► We have

$$\det A = \det(-A^T) = (-1)^{2001} \det A^T = -\det A,$$

and so $2\det A = 0$, from where $\det A = 0$. ◄

Lemma B.3.14 — Row-Linearity and Column-Linearity of Determinants. Let $A \in \mathbf{M}_{n \times n}(\mathbb{R}), A =$

$[a_{ij}]$. For a fixed row s, suppose that $a_{sj} = b_{sj} + c_{sj}$ for each $j \in [1;n]$. Then

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ b_{s1}+c_{s1} & b_{s2}+c_{s2} & \cdots & b_{sn}+c_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix}$$

$$= \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ b_{s1} & b_{s2} & \cdots & b_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix}$$

$$+ \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ c_{s1} & c_{s2} & \cdots & c_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix}.$$

An analogous result holds for columns.

Proof: Put

$$S = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ b_{s1}+c_{s1} & b_{s2}+c_{s2} & \cdots & b_{sn}+c_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix},$$

$$T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ b_{s1} & b_{s2} & \cdots & b_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix}$$

and

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{(s-1)1} & a_{(s-1)2} & \cdots & a_{(s-1)n} & \\ c_{s1} & c_{s2} & \cdots & c_{sn} & \\ a_{(s+1)1} & a_{(s+1)2} & \cdots & a_{(s+1)n} & \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{bmatrix}.$$

Then

$$\begin{aligned} \det S &= \textstyle\sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{(s-1)\sigma(s-1)}(b_{s\sigma(s)} \\ &\qquad +c_{s\sigma(s)})a_{(s+1)\sigma(s+1)}\cdots a_{n\sigma(n)} \\ &= \textstyle\sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{(s-1)\sigma(s-1)}b_{s\sigma(s)}a_{(s+1)\sigma(s+1)}\cdots a_{n\sigma(n)} \\ &\quad +\textstyle\sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{(s-1)\sigma(s-1)}c_{s\sigma(s)}a_{(s+1)\sigma(s+1)}\cdots a_{n\sigma(n)} \\ &= \det T + \det U. \end{aligned}$$

By applying the above argument to A^T, we obtain the result for columns.

■

Lemma B.3.15 If two rows or two columns of $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ are identical, then $\det A = 0_{\mathbb{R}}$.

Proof: Suppose $a_{sj} = a_{tj}$ for $s \neq t$ and for all $j \in [1;n]$. In particular, this means that for any $\sigma \in S_n$ we have $a_{s\sigma(t)} = a_{t\sigma(t)}$ and $a_{t\sigma(s)} = a_{s\sigma(s)}$. Let τ be the transposition

$$\tau = \begin{bmatrix} s & t \\ \tau(t) & \tau(s) \end{bmatrix}.$$

Then $\sigma\tau(a) = \sigma(a)$ for $a \in \{1,2,\ldots,n\}\setminus\{s,t\}$. Also, $\mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau) = -\mathrm{sgn}(\sigma)$. As σ

runs through all even permutations, $\sigma\tau$ runs through all odd permutations, and viceversa. Therefore

$$
\begin{aligned}
\det A &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{s\sigma(s)}\cdots a_{t\sigma(t)}\cdots a_{n\sigma(n)} \\
&= \sum_{\substack{\sigma\in S_n\\ \operatorname{sgn}(\sigma)=1}} \Big(\operatorname{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{s\sigma(s)}\cdots a_{t\sigma(t)}\cdots a_{n\sigma(n)} \\
&\quad +\operatorname{sgn}(\sigma\tau)\, a_{1\sigma\tau(1)}a_{2\sigma\tau(2)}\cdots a_{s\sigma\tau(s)}\cdots a_{t\sigma\tau(t)}\cdots a_{n\sigma\tau(n)}\Big) \\
&= \sum_{\substack{\sigma\in S_n\\ \operatorname{sgn}(\sigma)=1}} \operatorname{sgn}(\sigma)\Big(a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{s\sigma(s)}\cdots a_{t\sigma(t)}\cdots a_{n\sigma(n)} \\
&\quad - a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{s\sigma(t)}\cdots a_{t\sigma(s)}\cdots a_{n\sigma(n)}\Big) \\
&= \sum_{\substack{\sigma\in S_n\\ \operatorname{sgn}(\sigma)=1}} \operatorname{sgn}(\sigma)\Big(a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{s\sigma(s)}\cdots a_{t\sigma(t)}\cdots a_{n\sigma(n)} \\
&\quad - a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{t\sigma(t)}\cdots a_{s\sigma(s)}\cdots a_{n\sigma(n)}\Big) \\
&= 0_{\mathbb{R}}.
\end{aligned}
$$

Arguing on A^T will yield the analogous result for the columns. ■

Corollary B.3.16 If two rows or two columns of $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ are proportional, then $\det A = 0_{\mathbb{R}}$.

Proof: Suppose $a_{sj} = \alpha a_{tj}$ for $s \neq t$ and for all $j \in [1;n]$. If B is the matrix obtained by pulling out the factor α from the s-th row then $\det A = \alpha \det B$. But now the s-th and the t-th rows in B are identical, and so $\det B = 0_{\mathbb{R}}$. Arguing on A^T will yield the analogous result for the columns.

■

Example B.3.17

$$\det \begin{bmatrix} 1 & a & b \\ 1 & a & c \\ 1 & a & d \end{bmatrix} = a \det \begin{bmatrix} 1 & 1 & b \\ 1 & 1 & c \\ 1 & 1 & d \end{bmatrix} = 0,$$

since on the last determinant the first two columns are identical.

Theorem B.3.18 — Multilinearity of Determinants. Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ and $\alpha \in \mathbb{R}$. If $X \in \mathbf{M}_{n\times n}(\mathbb{R}), X = (x_{ij})$ is the matrix obtained by the row transvection $R_s + \alpha R_t \to R_s$ then $\det X = \det A$. Similarly, if $Y \in \mathbf{M}_{n\times n}(\mathbb{R}), Y = (y_{ij})$ is the matrix obtained by the column transvection $C_s + \alpha C_t \to C_s$ then $\det Y = \det A$.

Proof: For the row transvection it suffices to take $b_{sj} = a_{sj}$, $c_{sj} = \alpha a_{tj}$ for $j \in [1;n]$ in Lemma

B.3.14. With the same notation as in the lemma, $T = A$, and so,

$$\det X = \det T + \det U = \det A + \det U.$$

But U has its s-th and t-th rows proportional ($s \neq t$), and so by Corollary B.3.16 $\det U = 0_{\mathbb{R}}$. Hence $\det X = \det A$. To obtain the result for column transvections it suffices now to also apply Theorem B.3.7. ■

Example B.3.19 Demonstrate, *without actually calculating the determinant* that

$$\det \begin{bmatrix} 2 & 9 & 9 \\ 4 & 6 & 8 \\ 7 & 4 & 1 \end{bmatrix}$$

is divisible by 13.

Solution: ► Observe that $299, 468$ and 741 are all divisible by 13. Thus

$$\det \begin{bmatrix} 2 & 9 & 9 \\ 4 & 6 & 8 \\ 7 & 4 & 1 \end{bmatrix} \overset{C_3+10C_2+100C_1 \to C_3}{=} \det \begin{bmatrix} 2 & 9 & 299 \\ 4 & 6 & 468 \\ 7 & 4 & 741 \end{bmatrix} = 13 \det \begin{bmatrix} 2 & 9 & 23 \\ 4 & 6 & 36 \\ 7 & 4 & 57 \end{bmatrix},$$

which shews that the determinant is divisible by 13. ◄

Theorem B.3.20 The determinant of a triangular matrix (upper or lower) is the product of its diagonal elements.

Proof: Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ be a triangular matrix. Observe that if $\sigma \neq \mathbf{Id}$ then $a_{i\sigma(i)} a_{\sigma(i)\sigma^2(i)} = 0_{\mathbb{R}}$ occurs in the product

$$a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Thus

$$\begin{aligned} \det A &= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \mathrm{sgn}(\mathbf{Id}\,)\, a_{1\mathbf{Id}\,(1)} a_{2\mathbf{Id}\,(2)} \cdots a_{n\mathbf{Id}\,(n)} = a_{11} a_{22} \cdots a_{nn}. \end{aligned}$$

■

Example B.3.21 The determinant of the $n \times n$ identity matrix $\mathbf{I}_n$ over a field $\mathbb{R}$ is

$$\det \mathbf{I}_n = 1_{\mathbb{R}}.$$

Example B.3.22 Find

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Solution: ► We have

$$\begin{aligned} \det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} & \overset{\substack{C_2-2C_1\to C_2 \\ C_3-3C_1\to C_3}}{\longrightarrow} & \det \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \\ 7 & -6 & -12 \end{bmatrix} \\ & = & (-3)(-6)\det \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 1 \\ 7 & 2 & 2 \end{bmatrix} \\ & = & 0, \end{aligned}$$

since in this last matrix the second and third columns are identical and so Lemma B.3.15 applies. ◄

Theorem B.3.23 Let $(A, B) \in (\mathbf{M}_{n\times n}(\mathbb{R}))^2$. Then

$$\det(AB) = (\det A)(\det B).$$

Proof: Put $D = AB, D = (d_{ij}), d_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. If $A_{(c:k)}, D_{(c:k)}, 1 \leqslant k \leqslant n$ denote the columns of A and D, respectively, observe that

$$D_{(c:k)} = \sum_{l=1}^{n} b_{lk}A_{(c:l)}, \quad 1 \leqslant k \leqslant n.$$

Applying Corollary B.3.10 and Lemma B.3.14 multiple times, we obtain

$$\begin{aligned} \det D & = & \det(D_{(c:1)}, D_{(c:2)}, \ldots, D_{(c:n)}) \\ & = & \sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_n=1}^{n} b_{1j_1}b_{2j_2}\cdots b_{nj_n} \\ & & \cdot\det(A_{(c:j_1)}, A_{(c:j_2)}, \ldots, A_{(c:j_n)}). \end{aligned}$$

By Lemma B.3.15, if any two of the $A_{(c:j_l)}$ are identical, the determinant on the right vanishes. So each one of the j_l is different in the non-vanishing terms and so the map

$$\sigma: \begin{array}{ccc} \{1,2,\ldots,n\} & \to & \{1,2,\ldots,n\} \\ l & \mapsto & j_l \end{array}$$

is a permutation. Here $j_l = \sigma(l)$. Therefore, for the non-vanishing

$$\det(A_{(c:j_1)}, A_{(c:j_2)}, \ldots, A_{(c:j_n)})$$

we have in view of Corollary B.3.6,

$$\begin{aligned} \det(A_{(c:j_1)}, A_{(c:j_2)}, \ldots, A_{(c:j_n)}) &= (\operatorname{sgn}(\sigma)) \det(A_{(c:1)}, A_{(c:2)}, \ldots, A_{(c:n)}) \\ &= (\operatorname{sgn}(\sigma)) \det A. \end{aligned}$$

We deduce that

$$\begin{aligned} \det(AB) &= \det D \\ &= \sum_{j_n=1}^{n} b_{1j_1} b_{2j_2} \cdots b_{nj_n} \det(A_{(c:j_1)}, A_{(c:j_2)}, \ldots, A_{(c:j_n)}) \\ &= (\det A) \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma)) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} \\ &= (\det A)(\det B), \end{aligned}$$

as we wanted to shew. ■ By applying the preceding theorem multiple times we obtain

Corollary B.3.24 If $A \in \mathbf{M}_{n\times n}(\mathbb{R})$ and if k is a positive integer then

$$\det A^k = (\det A)^k.$$

Corollary B.3.25 If $A \in \mathbf{GL}_n(\mathbb{R})$ and if k is a positive integer then $\det A \neq 0_{\mathbb{R}}$ and

$$\det A^{-k} = (\det A)^{-k}.$$

Proof: We have $AA^{-1} = \mathbf{I}_n$ and so by Theorem B.3.23 $(\det A)(\det A^{-1}) = 1_{\mathbb{R}}$, from where the result follows. ■

Homework

Remark B.3.26 Let

$$\Omega = \det \begin{bmatrix} bc & ca & ab \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix}.$$

Without expanding either determinant, prove that

$$\Omega = \det \begin{bmatrix} 1 & 1 & 1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix}.$$

Remark B.3.27 Demonstrate that

$$\Omega = \det \begin{bmatrix} a-b-c & 2a & 2a \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{bmatrix} = (a+b+c)^3.$$

Remark B.3.28 After the indicated column operations on a 3×3 matrix A with $\det A = -540$, matrices $A_1, A_2, \ldots, A_5$ are successively obtained:

$$A \overset{C_1+3C_2\to C_1}{\rightarrow} A_1 \overset{C_2\leftrightarrow C_3}{\rightarrow} A_2 \overset{3C_2-C_1\to C_2}{\rightarrow} A_3 \overset{C_1-3C_2\to C_1}{\rightarrow} A_4 \overset{2C_1\to C_1}{\rightarrow} A_5$$

Determine the numerical values of $\det A_1, \det A_2, \det A_3, \det A_4$ and $\det A_5$.

Remark B.3.29 Prove, without actually expanding the determinant, that

$$\det \begin{bmatrix} 1 & 2 & 3 & 7 & 0 \\ 6 & 1 & 5 & 14 & 1 \\ 8 & 6 & 1 & 21 & 3 \\ 7 & 3 & 8 & 7 & 1 \\ 2 & 4 & 6 & 0 & 4 \end{bmatrix}$$

is divisible by 1722.

Remark B.3.30 Let A, B, C be 3×3 matrices with $\det A = 3, \det B^3 = -8, \det C = 2$. Compute (i) $\det ABC$, (ii) $\det 5AC$, (iii) $\det A^3B^{-3}C^{-1}$. Express your answers as fractions.

Remark B.3.31 Shew that $\forall A \in \mathbf{M}_{n\times n}(\mathbb{R})$,

$$\exists (X,Y) \in (\mathbf{M}_{n\times n}(\mathbb{R}))^2, (\det X)(\det Y) \neq 0$$

such that

$$A = X + Y.$$

That is, any square matrix over $\mathbb{R}$ can be written as a sum of two matrices whose determinant is not zero.

Remark B.3.32 Prove or disprove! The set $X = \{A \in \mathbf{M}_{n\times n}(\mathbb{R}) : \det A = 0_{\mathbb{R}}\}$ is a vector subspace of $\mathbf{M}_{n\times n}(\mathbb{R})$.

B.4 Laplace Expansion

We now develop a more computationally convenient approach to determinants.

Put

$$C_{ij} = \sum_{\substack{\sigma \in S_n \\ \sigma(i)=j}} (\operatorname{sgn}(\sigma)) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Then

$$\begin{aligned} \det A &= \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma)) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{i=1}^{n} a_{ij} \sum_{\substack{\sigma \in S_n \\ \sigma(i)=j}} (\operatorname{sgn}(\sigma)) a_{1\sigma(1)} a_{2\sigma(2)} \\ &\qquad \cdots a_{(i-1)\sigma(i-1)} a_{(i+1)\sigma(i+1)} \cdots a_{n\sigma(n)} \\ &= \sum_{i=1}^{n} a_{ij} C_{ij}, \end{aligned} \tag{B.5}$$

is the expansion of $\det A$ along the j-th column. Similarly,

$$\begin{aligned} \det A &= \sum_{\sigma \in S_n} (\operatorname{sgn}(\sigma)) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{j=1}^{n} a_{ij} \sum_{\substack{\sigma \in S_n \\ \sigma(i)=j}} (\operatorname{sgn}(\sigma)) a_{1\sigma(1)} a_{2\sigma(2)} \\ &\qquad \cdots a_{(i-1)\sigma(i-1)} a_{(i+1)\sigma(i+1)} \cdots a_{n\sigma(n)} \\ &= \sum_{j=1}^{n} a_{ij} C_{ij}, \end{aligned}$$

is the expansion of $\det A$ along the i-th row.

Definition B.4.1 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$. The ij-th minor $A_{ij} \in \mathbf{M}_{n-1}(\mathbb{R})$ is the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column from A.

Example B.4.1 If

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

then, for example,

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix}, \quad A_{33} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}.$$

Theorem B.4.2 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. Then

$$\det A = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det A_{ij} = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det A_{ij}.$$

Proof: It is enough to shew, in view of B.5 that

$$(-1)^{i+j}\det A_{ij} = C_{ij}.$$

Now,

$$\begin{aligned} C_{nn} &= \sum_{\substack{\sigma\in S_n \\ \sigma(n)=n}} \operatorname{sgn}(\sigma)\, a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{(n-1)\sigma(n-1)} \\ &= \sum_{\tau\in S_{n-1}} \operatorname{sgn}(\tau)\, a_{1\tau(1)}a_{2\tau(2)}\cdots a_{(n-1)\tau(n-1)} \\ &= \det A_{nn}, \end{aligned}$$

since the second sum shewn is the determinant of the submatrix obtained by deleting the last row and last column from A.

To find C_{ij} for general ij we perform some row and column interchanges to A in order to bring a_{ij} to the nn-th position. We thus bring the i-th row to the n-th row by a series of transpositions, first swapping the i-th and the $(i+1)$-th row, then swapping the new $(i+1)$-th row and the $(i+2)$-th row, and so forth until the original i-th row makes it to the n-th row. We have made thereby $n-i$ interchanges. To this new matrix we perform analogous interchanges to the j-th column, thereby making $n-j$ interchanges. We have made a total of $2n-i-j$ interchanges. Observe that $(-1)^{2n-i-j} = (-1)^{i+j}$. Call the analogous quantities in the resulting matrix A', C'_{nn}, A'_{nn}. Then

$$C_{ij} = C'_{nn} = \det A'_{nn} = (-1)^{i+j}\det A_{ij},$$

by virtue of Corollary B.3.6.

■

It is irrelevant which row or column we choose to expand a determinant of a square matrix. We always obtain the same result. The sign pattern is given by

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \vdots \\ + & - & + & - & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Example B.4.3 Find

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

by expanding along the first row.

Solution: ▶ We have

$$\begin{aligned} \det A &= 1(-1)^{1+1} \det \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} + 2(-1)^{1+2} \det \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} + 3(-1)^{1+3} \det \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} \\ &= 1(45-48) - 2(36-42) + 3(32-35) = 0. \end{aligned}$$

◀

Example B.4.4 Evaluate the *Vandermonde* determinant

$$\det \begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix}.$$

Solution: ▶

$$\begin{aligned} \det \begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix} &= \det \begin{bmatrix} 1 & 0 & 0 \\ a & b-a & c-a \\ a^2 & b^2-a^2 & c^2-a^2 \end{bmatrix} \\ &= \det \begin{bmatrix} b-a & c-a \\ b^2-c^2 & c^2-a^2 \end{bmatrix} \\ &= (b-a)(c-a) \det \begin{bmatrix} 1 & 1 \\ b+a & c+a \end{bmatrix} \\ &= (b-a)(c-a)(c-b). \end{aligned}$$

◀

Example B.4.5 Evaluate the determinant

$$\det A = \det\begin{bmatrix} 1 & 2 & 3 & 4 & \cdots & 2000 \\ 2 & 1 & 2 & 3 & \cdots & 1999 \\ 3 & 2 & 1 & 2 & \cdots & 1998 \\ 4 & 3 & 2 & 1 & \cdots & 1997 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 2000 & 1999 & 1998 & 1997 & \cdots & 1 \end{bmatrix}.$$

Solution: ► Applying $R_n - R_{n+1} \to R_n$ for $1 \leqslant n \leqslant 1999$, the determinant becomes

$$\det\begin{bmatrix} -1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ -1 & -1 & 1 & 1 & \cdots & 1 & 1 \\ -1 & -1 & -1 & 1 & \cdots & 1 & 1 \\ -1 & -1 & -1 & -1 & \cdots & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & -1 & -1 & \cdots & -1 & 1 \\ 2000 & 1999 & 1998 & 1997 & \cdots & 2 & 1 \end{bmatrix}.$$

Applying now $C_n + C_{2000} \to C_n$ for $1 \leqslant n \leqslant 1999$, we obtain

$$\det\begin{bmatrix} 0 & 2 & 2 & 2 & \cdots & 2 & 1 \\ 0 & 0 & 2 & 2 & \cdots & 2 & 1 \\ 0 & 0 & 0 & 2 & \cdots & 2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 2001 & 2000 & 1999 & 1998 & \cdots & 3 & 1 \end{bmatrix}.$$

This last determinant we expand along the first column. We have

$$2001\det\begin{bmatrix} 2 & 2 & 2 & \cdots & 2 & 1 \\ 0 & 2 & 2 & \cdots & 2 & 1 \\ 0 & 0 & 2 & \cdots & 2 & 1 \\ 0 & 0 & 0 & \cdots & 2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = 2001(2^{1998}).$$

◀

Definition B.4.2 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. The **classical adjoint** or **adjugate** of A is the $n\times n$ matrix $\operatorname{adj} A$ whose entries are given by

$$[\operatorname{adj} A]_{ij} = (-1)^{i+j} \det A_{ji},$$

where A_{ji} is the ji-th minor of A.

Theorem B.4.6 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. Then

$$(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)\mathbf{I}_n.$$

Proof: We have

$$\begin{aligned} [A(\operatorname{adj} A)]_{ij} &= \textstyle\sum_{k=1}^{n} a_{ik}[\operatorname{adj} A]_{kj} \\ &= \textstyle\sum_{k=1}^{n} a_{ik}(-1)^{i+k}\det A_{jk}. \end{aligned}$$

Now, this last sum is $\det A$ if $i=j$ by virtue of Theorem B.4.2. If $i\neq j$ it is 0, since then the j-th row is identical to the i-th row and this determinant is $0_{\mathbb{R}}$ by virtue of Lemma B.3.15. Thus on the diagonal entries we get $\det A$ and the off-diagonal entries are $0_{\mathbb{R}}$. This proves the theorem. ■ The

next corollary follows immediately.

Corollary B.4.7 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. Then A is invertible if and only $\det A \neq 0_{\mathbb{R}}$ and

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}.$$

Homework

Remark B.4.8 Find

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

by expanding along the second column.

Remark B.4.9 Prove that $\det \begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix} = a^3 + b^3 + c^3 - 3abc$. This type of matrix is called a circulant matrix.

Remark B.4.10 Compute the determinant

$$\det \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 0 & 0 & 1 \\ 666 & -3 & -1 & 1000000 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Remark B.4.11 Prove that

$$\det \begin{bmatrix} x+a & b & c \\ a & x+b & c \\ a & b & x+c \end{bmatrix} = x^2(x+a+b+c).$$

Remark B.4.12 If

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ x & a & 0 & 0 \\ x & 0 & b & 0 \\ x & 0 & 0 & c \end{bmatrix} = 0,$$

and $xabc \neq 0$, prove that

$$\frac{1}{x} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c}.$$

Remark B.4.13 Consider the matrix

$$A = \begin{bmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{bmatrix}.$$

❶ Compute $A^T A$.

❷ Use the above to prove that

$$\det A = (a^2 + b^2 + c^2 + d^2)^2.$$

Remark B.4.14 Prove that

$$\det \begin{bmatrix} 0 & a & b & 0 \\ a & 0 & b & 0 \\ 0 & a & 0 & b \\ 1 & 1 & 1 & 1 \end{bmatrix} = 2ab(a-b).$$

Remark B.4.15 Demonstrate that

$$\det \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix} = (ad - bc)^2.$$

Remark B.4.16 Use induction to shew that

$$\det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} = (-1)^{n+1}.$$

Remark B.4.17 Let

$$A = \begin{bmatrix} 1 & n & n & n & \cdots & n \\ n & 2 & n & n & \vdots & n \\ n & n & 3 & n & \cdots & n \\ n & n & n & 4 & \cdots & n \\ \vdots & \vdots & \vdots & \cdots & \vdots & \\ n & n & n & n & n & n \end{bmatrix},$$

that is, $A \in \mathbf{M}_{n\times n}(\mathbb{R}), A = [a_{ij}]$ is a matrix such that $a_{kk} = k$ and $a_{ij} = n$ when $i \neq j$. Find $\det A$.

Remark B.4.18 Let $n \in \mathbb{N}, n > 1$ be an odd integer. Recall that the binomial coefficients $\binom{n}{k}$ satisfy $\binom{n}{n} = \binom{n}{0} = 1$ and that for $1 \leqslant k \leqslant n$,

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Prove that

$$\det \begin{bmatrix} 1 & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n-1} & 1 \\ 1 & 1 & \binom{n}{1} & \cdots & \binom{n}{n-2} & \binom{n}{n-1} \\ \binom{n}{n-1} & 1 & 1 & \cdots & \binom{n}{n-3} & \binom{n}{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \binom{n}{1} & \binom{n}{2} & \binom{n}{3} & \cdots & 1 & 1 \end{bmatrix} = (1 + (-1)^n)^n.$$

Remark B.4.19 Let $A \in \mathbf{GL}_n(\mathbb{R}), n > 1$. Prove that $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

Remark B.4.20 Let $(A, B, S) \in (\mathbf{GL}_n(\mathbb{R}))^3$. Prove that

❶ $\operatorname{adj}\operatorname{adj} A = (\det A)^{n-2} A$.

❷ $\operatorname{adj} AB = \operatorname{adj} A \operatorname{adj} B$.

❸ $\operatorname{adj} SAS^{-1} = S(\operatorname{adj} A)S^{-1}$.

Remark B.4.21 If $A \in \mathbf{GL}_2(\mathbb{R})$, , and let k be a positive integer. Prove that $\det(\underbrace{\operatorname{adj} \cdots \operatorname{adj}}_{k}(A)) = \det A$.

Remark B.4.22 Find the determinant

$$\det \begin{bmatrix} (b+c)^2 & ab & ac \\ ab & (a+c)^2 & bc \\ ac & bc & (a+b)^2 \end{bmatrix}$$

by hand, making explicit all your calculations.

Remark B.4.23 The matrix

$$\begin{bmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix}$$

is known as a **circulant matrix**. Prove that its determinant is $(a+b+c+d)(a-b+c-d)((a-c)^2+(b-d)^2)$.

B.5 Determinants and Linear Systems

Theorem B.5.1 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. The following are all equivalent

❶ $\det A \neq 0_{\mathbb{R}}$.

❷ A is invertible.

❸ There exists a unique solution $X \in \mathbf{M}_{n\times 1}(\mathbb{R})$ to the equation $AX = Y$.

❹ If $AX = \mathbf{0}_{n\times 1}$ then $X = \mathbf{0}_{n\times 1}$.

Proof: We prove the implications in sequence:
$\Longrightarrow$: follows from Corollary B.4.7
$\Longrightarrow$: If A is invertible and $AX = Y$ then $X = A^{-1}Y$ is the unique solution of this equation.
$\Longrightarrow$: follows by putting $Y = \mathbf{0}_{n\times 1}$
$\Longrightarrow$: Let R be the row echelon form of A. Since $RX = \mathbf{0}_{n\times 1}$ has only $X = \mathbf{0}_{n\times 1}$ as a solution, every entry on the diagonal of R must be non-zero, R must be triangular, and hence $\det R \neq 0_{\mathbb{R}}$. Since $A = PR$ where P is an invertible $n \times n$ matrix, we deduce that $\det A = \det P \det R \neq 0_{\mathbb{R}}$.
■ The contrapositive form of the implications ❶ and ❹ will be used later. Here it is for future reference.

Corollary B.5.2 Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$. If there is $X \neq \mathbf{0}_{n\times 1}$ such that $AX = \mathbf{0}_{n\times 1}$ then $\det A = 0_{\mathbb{R}}$.

Homework

Remark B.5.3 For which a is the matrix $\begin{bmatrix} -1 & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{bmatrix}$ singular (non-invertible)?

3

CHAPTER

Answers and Hints

2.2 Since polynomials are continuous functions and the image of a connected set is connected for a continuous function, the image must be an interval of some sort. If the image were a finite interval, then $f(x,kx)$ would be bounded for every constant k, and so the image would just be the point $f(0,0)$. The possibilities are thus

1. a single point (take for example, $p(x,y)=0$),
2. a semi-infinite interval with an endpoint (take for example $p(x,y)=x^2$ whose image is $[0;+\infty[$),
3. a semi-infinite interval with no endpoint (take for example $p(x,y)=(xy-1)^2+x^2$ whose image is $]0;+\infty[$),
4. all real numbers (take for example $p(x,y)=x$).

2.8 0

2.9 2

2.10 $c=0$.

2.11 0

2.14 By AM-GM,

$$\frac{x^2y^2z^2}{x^2+y^2+z^2}\leqslant\frac{(x^2+y^2+z^2)^3}{27(x^2+y^2+z^2)}=\frac{(x^2+y^2+z^2)^2}{27}\to 0$$

as $(x,y,z)\to(0,0,0)$.

2.21 0

2.22 2

2.23 $c = 0$.

2.24 0

2.27 By AM-GM,

$$\frac{x^2y^2z^2}{x^2+y^2+z^2} \leqslant \frac{(x^2+y^2+z^2)^3}{27(x^2+y^2+z^2)} = \frac{(x^2+y^2+z^2)^2}{27} \to 0$$

as $(x, y, z) \to (0, 0, 0)$.

3.4 We have

$$\begin{aligned} F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) &= (\mathbf{x}+\mathbf{h}) \times L(\mathbf{x}+\mathbf{h}) - \mathbf{x} \times L(\mathbf{x}) \\ &= (\mathbf{x}+\mathbf{h}) \times (L(\mathbf{x}) + L(\mathbf{h})) - \mathbf{x} \times L(\mathbf{x}) \\ &= \mathbf{x} \times L(\mathbf{h}) + \mathbf{h} \times L(\mathbf{x}) + \mathbf{h} \times L(\mathbf{h}) \end{aligned}$$

Now, we will prove that $||\mathbf{h} \times L(\mathbf{h})|| = \mathbf{o}\left(||\mathbf{h}||\right)$ as $\mathbf{h} \to \mathbf{0}$. For let

$$\mathbf{h} = \sum_{k=1}^{n} h_k \mathbf{e}_k,$$

where the $\mathbf{e}_k$ are the standard basis for $\mathbb{R}^n$. Then

$$L(\mathbf{h}) = \sum_{k=1}^{n} h_k L(\mathbf{e}_k),$$

and hence by the triangle inequality, and by the Cauchy-Bunyakovsky-Schwarz inequality,

$$\begin{aligned} ||L(\mathbf{h})|| &\leqslant \sum_{k=1}^{n} |h_k| ||L(\mathbf{e}_k)|| \\ &\leqslant \left(\sum_{k=1}^{n} |h_k|^2\right)^{1/2} \left(\sum_{k=1}^{n} ||L(\mathbf{e}_k)||^2\right)^{1/2} \\ &= ||\mathbf{h}|| (\sum_{k=1}^{n} ||L(\mathbf{e}_k)||^2)^{1/2}, \end{aligned}$$

whence, again by the Cauchy-Bunyakovsky-Schwarz Inequality,

$$||\mathbf{h} \times L(\mathbf{h})|| \leqslant ||\mathbf{h}|| ||L(\mathbf{h})| \leqslant ||\mathbf{h}||^2 |||L(\mathbf{e}_k)||^2)^{1/2} = \mathbf{o}\left(||\mathbf{h}||\right),$$

as it was to be shewn.

3.5 Assume that $\mathbf{x} \neq \mathbf{0}$. We use the fact that $(1+t)^{1/2} = 1 + \frac{t}{2} + \mathbf{o}(t)$ as $t \to 0$. We have

$$\begin{aligned} \mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) &= ||\mathbf{x}+\mathbf{h}|| - ||\mathbf{x}|| \\ &= \sqrt{(\mathbf{x}+\mathbf{h})\bullet(\mathbf{x}+\mathbf{h})} - ||\mathbf{x}|| \\ &= \sqrt{||\mathbf{x}||^2 + 2\mathbf{x}\bullet\mathbf{h} + ||\mathbf{h}||^2} - ||\mathbf{x}|| \\ &= \frac{2\mathbf{x}\bullet\mathbf{h} + ||\mathbf{h}||^2}{\sqrt{||\mathbf{x}||^2 + 2\mathbf{x}\bullet\mathbf{h} + ||\mathbf{h}||^2} + ||\mathbf{x}||}. \end{aligned}$$

As $\mathbf{h}\to\mathbf{0}$,

$$\sqrt{\|\mathbf{x}\|^2+2\mathbf{x}\bullet\mathbf{h}+\|\mathbf{h}\|^2}+\|\mathbf{x}\|\to 2\|\mathbf{x}\|.$$

Since $\|\mathbf{h}\|^2=\mathbf{o}\left(\|\mathbf{h}\|\right)$ as $\mathbf{h}\to\mathbf{0}$, we have

$$\frac{2\mathbf{x}\bullet\mathbf{h}+\|\mathbf{h}\|^2}{\sqrt{\|\mathbf{x}\|^2+2\mathbf{x}\bullet\mathbf{h}+\|\mathbf{h}\|^2}+\|\mathbf{x}\|}\to\frac{\mathbf{x}\bullet\mathbf{h}}{\|\mathbf{h}\|}+\mathbf{o}\left(\|\mathbf{h}\|\right),$$

proving the first assertion.

To prove the second assertion, assume that there is a linear transformation $D_{\mathbf{0}}(f)=L$, $L:\mathbb{R}^n\to\mathbb{R}$ such that

$$\|\mathbf{f}(\mathbf{0}+\mathbf{h})-\mathbf{f}(\mathbf{0})-L(\mathbf{h})\|=\mathbf{o}\left(\|\mathbf{h}\|\right),$$

as $\|\mathbf{h}\|\to 0$. Recall that by theorem **??**, $L(\mathbf{0})=\mathbf{0}$, and so by example 3.3.2, $D_{\mathbf{0}}(L)(\mathbf{0})=L(\mathbf{0})=\mathbf{0}$. This implies that $\dfrac{L(\mathbf{h})}{\|\mathbf{h}\|}\to D_{\mathbf{0}}(L)(\mathbf{0})=\mathbf{0}$, as $\|\mathbf{h}\|\to 0$. Since $\mathbf{f}(\mathbf{0})=\|0\|=0, \mathbf{f}(\mathbf{h})=\|\mathbf{h}\|$ this would imply that

$$\Big|\,\|\mathbf{h}\|-L(\mathbf{h})\Big|=\mathbf{o}\left(\|\mathbf{h}\|\right),$$

or

$$\left|\left|1-\frac{L(\mathbf{h})}{\|\mathbf{h}\|}\right|\right|=\mathbf{o}\left(1\right).$$

But the left side $\to 1$ as $\mathbf{h}\to\mathbf{0}$, and the right side $\to 0$ as $\mathbf{h}\to\mathbf{0}$. This is a contradiction, and so, such linear transformation L does not exist at the point $\mathbf{0}$.

3.7 Observe that

$$\mathbf{f}(x,y)=\begin{cases} x & \text{if } x\leqslant y^2\\ y^2 & \text{if } x>y^2\end{cases}$$

Hence

$$\frac{\partial}{\partial x}\mathbf{f}(x,y)=\begin{cases} 1 & \text{if } x>y^2\\ 0 & \text{if } x>y^2\end{cases}$$

and

$$\frac{\partial}{\partial y}\mathbf{f}(x,y)=\begin{cases} 0 & \text{if } x>y^2\\ 2y & \text{if } x>y^2\end{cases}$$

3.8 Observe that

$$\mathbf{g}(1,0,1)=(30),\qquad f'(x,y)=\begin{bmatrix} y^2 & 2xy\\ 2xy & x^2\end{bmatrix},\qquad g'(x,y)=\begin{bmatrix} 1 & -1 & 2\\ y & x & 0\end{bmatrix},$$

and hence

$$g'(1,0,1)=\begin{bmatrix} 1 & -1 & 2\\ 0 & 1 & 0\end{bmatrix},\qquad f'(\mathbf{g}(1,0,1))=f'(3,0)=\begin{bmatrix} 0 & 0\\ 0 & 9\end{bmatrix}.$$

This gives, via the Chain-Rule,

$$(f\circ g)'(1,0,1) = f'(\mathbf{g}(1,0,1))g'(1,0,1) = \begin{bmatrix} 0 & 0 \\ 0 & 9 \end{bmatrix}\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \end{bmatrix}.$$

The composition $g\circ \mathbf{f}$ is undefined. For, the output of $\mathbf{f}$ is $\mathbb{R}^2$, but the input of $\mathbf{g}$ is in $\mathbb{R}^3$.

3.9 Since $\mathbf{f}(0,1) = (01)$, the Chain Rule gives

$$(g\circ f)'(0,1) = (g'(\mathbf{f}(0,1)))(f'(0,1)) = (g'(0,1))(f'(0,1)) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 0 \\ 2 & 1 \end{bmatrix}$$

3.13 We have

$$\frac{\partial}{\partial x}(x+z)^2 + \frac{\partial}{\partial x}(y+z)^2 = \frac{\partial}{\partial x}8 \implies 2(1+\frac{\partial z}{\partial x})(x+z) + 2\frac{\partial z}{\partial x}(y+z) = 0.$$

At $(1,1,1)$ the last equation becomes

$$4(1+\frac{\partial z}{\partial x}) + 4\frac{\partial z}{\partial x} = 0 \implies \frac{\partial z}{\partial x} = -\frac{1}{2}.$$

3.14 a) Here $\nabla T = (y+z)\mathbf{i} + (x+z)\mathbf{j} + (y+x)\mathbf{k}$. The maximum rate of change at $(1,1,1)$ is $|\nabla T(1,1,1)| = 2\sqrt{3}$ and direction cosines are

$$\frac{\nabla T}{|\nabla T|} = \frac{1}{\sqrt{3}}\mathbf{i} + \frac{1}{\sqrt{3}}\mathbf{j} + \frac{1}{\sqrt{3}}\mathbf{k} = \cos\alpha\mathbf{i} + \cos\beta\mathbf{j} + \cos\gamma\mathbf{k}$$

b) The required derivative is

$$\nabla T(1,1,1)\bullet\frac{3\mathbf{i}-4\mathbf{k}}{|3\mathbf{i}-4\mathbf{k}|} = -\frac{2}{5}$$

3.15 a) Here $\nabla\phi = \mathbf{F}$ requires $\nabla\times\mathbf{F} = 0$ which is not the case here, so no solution.

b) Here $\nabla\times\mathbf{F} = 0$ so that

$$\phi(x,y,z) = x^2y + y^2z + z + c$$

3.16 $\nabla f(x,y,z) = (e^{yz}, xze^{yz}, xye^{yz}) \implies (\nabla f)(2,1,1) = (e, 2e, 2e)$.

3.17 $(\nabla\times f)(x,y,z) = (0, x, ye^{xy}) \implies (\nabla\times f)(2,1,1) = \left(0, 2, e^2\right)$.

3.19 The vector $(1,-7,0)$ is perpendicular to the plane. Put $\mathbf{f}(x,y,z) = x^2 + y^2 - 5xy + xz - yz + 3$. Then $(\nabla f)(x,y,z) = (2x - 5y + z, 2y - 5x - zx - y)$. Observe that $\nabla f(x,y,z)$ is parallel to the vector $(1,-7,0)$, and hence there exists a constant a such that

$$(2x-5y+z, 2y-5x-zx-y) = a\,(1,-7,0) \implies x = a, \quad y = a, \quad z = 4a.$$

Since the point is on the plane

$$x - 7y = -6 \implies a - 7a = -6 \implies a = 1.$$

Thus $x = y = 1$ and $z = 4$.

3.22 Observe that

$$\mathbf{f}(0,0)=1, \quad f_x(x,y)=(\cos 2y)e^{x\cos 2y} \implies f_x(0,0)=1,$$

$$f_y(x,y)=-2x\sin 2y e^{x\cos 2y} \implies f_y(0,0)=0.$$

Hence

$$\mathbf{f}(x,y)\approx \mathbf{f}(0,0)+f_x(0,0)(x-0)+f_y(0,0)(y-0) \implies \mathbf{f}(x,y)\approx 1+x.$$

This gives $\mathbf{f}(0.1,-0.2)\approx 1+0.1=1.1$.

3.23 This is essentially the product rule: $duv = udv + vdu$, where ∇ acts the differential operator and $\times$ is the product. Recall that when we defined the volume of a parallelepiped spanned by the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, we saw that

$$\mathbf{a}\bullet(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\times\mathbf{b})\bullet\mathbf{c}.$$

Treating $\nabla=\nabla_{\mathbf{u}}+\nabla_{\mathbf{v}}$ as a vector, first keeping $\mathbf{v}$ constant and then keeping $\mathbf{u}$ constant we then see that

$$\nabla_{\mathbf{u}}\bullet(\mathbf{u}\times\mathbf{v})=(\nabla\times\mathbf{u})\bullet\mathbf{v}, \qquad \nabla_{\mathbf{v}}\bullet(\mathbf{u}\times\mathbf{v})=-\nabla\bullet(\mathbf{v}\times\mathbf{u})=-(\nabla\times\mathbf{v})\bullet\mathbf{u}.$$

Thus

$$\nabla\bullet(\mathbf{u}\times\mathbf{v})=(\nabla_{\mathbf{u}}+\nabla_{\mathbf{v}})\bullet(\mathbf{u}\times\mathbf{v})=\nabla_{\mathbf{u}}\bullet(\mathbf{u}\times\mathbf{v})+\nabla_{\mathbf{v}}\bullet(\mathbf{u}\times\mathbf{v})=(\nabla\times\mathbf{u})\bullet\mathbf{v}-(\nabla\times\mathbf{v})\bullet\mathbf{u}.$$

3.26 An angle of $\frac{\pi}{6}$ with the x-axis and $\frac{\pi}{3}$ with the y-axis.

5.1 Let $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a point on S. If this point were on the xz plane, it would be on the ellipse, and its distance to the axis of rotation would be $|x|=\frac{1}{2}\sqrt{1-z^2}$. Anywhere else, the distance from $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ to the z-axis is the distance of this point to the point $\begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$: $\sqrt{x^2+y^2}$. This distance is the same as the length of the segment on the xz-plane going from the z-axis. We thus have

$$\sqrt{x^2+y^2}=\frac{1}{2}\sqrt{1-z^2},$$

or

$$4x^2+4y^2+z^2=1.$$

5.2 Let $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a point on S. If this point were on the xy plane, it would be on the line, and its distance to the axis of rotation would be $|x| = \frac{1}{3}|1-4y|$. Anywhere else, the distance of $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ to the axis of rotation is the same as the distance of $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ to $\begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}$, that is $\sqrt{x^2+z^2}$. We must have

$$\sqrt{x^2+z^2} = \frac{1}{3}|1-4y|,$$

which is to say

$$9x^2+9z^2-16y^2+8y-1=0.$$

5.3 A spiral staircase.

5.4 A spiral staircase.

5.6 The planes $A: x+z=0$ and $B: y=0$ are secant. The surface has equation of the form $f(A,B) = e^{A^2+B^2}-A=0$, and it is thus a cylinder. The directrix has direction $\mathbf{i}-\mathbf{k}$.

5.7 Rearranging,

$$(x^2+y^2+z^2)^2-\frac{1}{2}((x+y+z)^2-(x^2+y^2+z^2))-1=0,$$

and so we may take $A: x+y+z=0, S: x^2+y^2+z^2=0$, shewing that the surface is of revolution. Its axis is the line in the direction $\mathbf{i}+\mathbf{j}+\mathbf{k}$.

5.8 Considering the planes $A: x-y=0, B: y-z=0$, the equation takes the form

$$f(A,B) = \frac{1}{A}+\frac{1}{B}-\frac{1}{A+B}-1=0,$$

thus the equation represents a cylinder. To find its directrix, we find the intersection of the planes $x=y$ and $y=z$. This gives $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. The direction vector is thus $\mathbf{i}+\mathbf{j}+\mathbf{k}$.

5.9 Rearranging,

$$(x+y+z)^2-(x^2+y^2+z^2)+2(x+y+z)+2=0,$$

so we may take $A: x+y+z=0, S: x^2+y^2+z^2=0$ as our plane and sphere. The axis of revolution is then in the direction of $\mathbf{i}+\mathbf{j}+\mathbf{k}$.

5.10 After rearranging, we obtain
$$(z-1)^2 - xy = 0,$$
or
$$-\frac{x}{z-1}\frac{y}{z-1} + 1 = 0.$$
Considering the planes
$$A: x = 0,\ B: y = 0,\ \ C: z = 1,$$
we see that our surface is a cone, with apex at $(0,0,1)$.

5.11 The largest circle has radius b. Parallel cross sections of the ellipsoid are similar ellipses, hence we may increase the size of these by moving towards the centre of the ellipse. Every plane through the origin which makes a circular cross section must intersect the yz-plane, and the diameter of any such cross section must be a diameter of the ellipse $x = 0, \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$. Therefore, the radius of the circle is at most b. Arguing similarly on the xy-plane shews that the radius of the circle is at least b. To shew that circular cross section of radius b actually exist, one may verify that the two planes given by $a^2(b^2 - c^2)z^2 = c^2(a^2 - b^2)x^2$ give circular cross sections of radius b.

5.12 Any hyperboloid oriented like the one on the figure has an equation of the form
$$\frac{z^2}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1.$$
When $z = 0$ we must have
$$4x^2 + y^2 = 1 \implies a = \frac{1}{2},\ b = 1.$$
Thus
$$\frac{z^2}{c^2} = 4x^2 + y^2 - 1.$$
Hence, letting $z = \pm 2$,
$$\frac{4}{c^2} = 4x^2 + y^2 - 1 \implies \frac{1}{c^2} = x^2 + \frac{y^2}{4} - \frac{1}{4} = 1 - \frac{1}{4} = \frac{3}{4},$$
since at $z = \pm 2$, $x^2 + \frac{y^2}{4} = 1$. The equation is thus
$$\frac{3z^2}{4} = 4x^2 + y^2 - 1.$$

13.1

1. Let $L_1 : y = x + 1$, $L_2 : -x + 1$. Then
$$\begin{aligned}\int_C x\mathrm{d}x + y\mathrm{d}y &= \int_{L_1} x\mathrm{d}x + y\mathrm{d}y + \int_{L_2} x\mathrm{d}x + y\mathrm{d}y \\ &= \int_{-1}^{1} x\mathrm{d}x(x+1)\mathrm{d}x + \int_0^1 x\mathrm{d}x - (-x+1)\mathrm{d}x \\ &= 0.\end{aligned}$$

Also, both on L_1 and on L_2 we have $\|d\mathbf{x}\| = \sqrt{2}dx$, thus

$$\begin{aligned}\int_C xy\|d\mathbf{x}\| &= \int_{L_1} xy\|d\mathbf{x}\| + \int_{L_2} xy\|d\mathbf{x}\| \\ &= \sqrt{2}\int_{-1}^{1} x(x+1)dx - \sqrt{2}\int_0^1 x(-x+1)dx \\ &= 0.\end{aligned}$$

2. We put $x = \sin t$, $y = \cos t$, $t \in \left[-\frac{\pi}{2}\,;\frac{\pi}{2}\right]$. Then

$$\begin{aligned}\int_C xdx + ydy &= \int_{-\pi/2}^{\pi/2} (\sin t)(\cos t)dt - (\cos t)(\sin t)dt \\ &= 0.\end{aligned}$$

Also, $\|d\mathbf{x}\| = \sqrt{(\cos t)^2 + (-\sin t)^2}dt = dt$, and thus

$$\begin{aligned}\int_C xy\|d\mathbf{x}\| &= \int_{-\pi/2}^{\pi/2} (\sin t)(\cos t)dt \\ &= \left.\frac{(\sin t)^2}{2}\right|_{-\pi/2}^{\pi/2} \\ &= 0.\end{aligned}$$

13.2 Let Γ_1 denote the straight line segment path from O to $A = (2\sqrt{3}, 2)$ and Γ_2 denote the arc of the circle centred at $(0,0)$ and radius 4 going counterclockwise from $\theta = \dfrac{\pi}{6}$ to $\theta = \dfrac{\pi}{5}$.

Observe that the Cartesian equation of the line $\overleftrightarrow{\mathbf{O}A}$ is $y = \dfrac{x}{\sqrt{3}}$. Then on Γ_1

$$xdx + ydy = xdx + \frac{x}{\sqrt{3}}d\frac{x}{\sqrt{3}} = \frac{4}{3}xdx.$$

Hence

$$\int_{\Gamma_1} xdx + ydy = \int_0^{2\sqrt{3}} \frac{4}{3}xdx = 8.$$

On the arc of the circle we may put $x = 4\cos\theta$, $y = 4\sin\theta$ and integrate from $\theta = \dfrac{\pi}{6}$ to $\theta = \dfrac{\pi}{5}$. Observe that there

$$xdx + ydy = (\cos\theta)d\cos\theta + (\sin\theta)d\sin\theta = -\sin\theta\cos\theta d\theta + \sin\theta\cos\theta d\theta = 0,$$

and since the integrand is 0, the integral will be zero.

Assembling these two pieces,

$$\int_\Gamma xdx + ydy = \int_{\Gamma_1} xdx + ydy + \int_{\Gamma_2} xdx + ydy = 8 + 0 = 8.$$

Using the parametrisations from the solution of problem **??**, we find on Γ_1 that

$$x\|d\mathbf{x}\| = x\sqrt{(dx)^2 + (dy)^2} = x\sqrt{1 + \frac{1}{3}}dx = \frac{2}{\sqrt{3}}xdx,$$

whence

$$\int_{\Gamma_1} x\|d\mathbf{x}\| = \int_0^{2\sqrt{3}} \frac{2}{\sqrt{3}} x dx = 4\sqrt{3}.$$

On Γ_2 that

$$x\|d\mathbf{x}\| = x\sqrt{(dx)^2+(dy)^2} = 16\cos\theta\sqrt{\sin^2\theta+\cos^2\theta}d\theta = 16\cos\theta d\theta,$$

whence

$$\int_{\Gamma_2} x\|d\mathbf{x}\| = \int_{\pi/6}^{\pi/5} 16\cos\theta d\theta = 16\sin\frac{\pi}{5} - 16\sin\frac{\pi}{6} = 4\sin\frac{\pi}{5} - 8.$$

Assembling these we gather that

$$\int_{\Gamma} x\|d\mathbf{x}\| = \int_{\Gamma_1} x\|d\mathbf{x}\| + \int_{\Gamma_2} x\|d\mathbf{x}\| = 4\sqrt{3} - 8 + 16\sin\frac{\pi}{5}.$$

13.3 The curve lies on the sphere, and to parametrise this curve, we dispose of one of the variables, y say, from where $y = 1-x$ and $x^2+y^2+z^2 = 1$ give

$$\begin{aligned} x^2+(1-x)^2+z^2 = 1 \quad &\implies \quad 2x^2-2x+z^2 = 0 \\ &\implies \quad 2\left(x-\tfrac{1}{2}\right)^2+z^2 = \tfrac{1}{2} \\ &\implies \quad 4\left(x-\tfrac{1}{2}\right)^2+2z^2 = 1. \end{aligned}$$

So we now put

$$x = \frac{1}{2}+\frac{\cos t}{2}, \quad z = \frac{\sin t}{\sqrt{2}}, \quad y = 1-x = \frac{1}{2}-\frac{\cos t}{2}.$$

We must integrate on the side of the plane that can be viewed from the point $(1,1,0)$ (observe that the vector $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ is normal to the plane). On the zx-plane, $4\left(x-\frac{1}{2}\right)^2+2z^2 = 1$ is an ellipse. To obtain a positive parametrisation we must integrate from $t = 2\pi$ to $t = 0$ (this is because when you look at the ellipse from the point $(1,1,0)$ the positive x-axis is to your left, and not your right). Thus

$$\begin{aligned} \oint_{\Gamma} zdx+xdy+ydz \quad &= \quad \int_{2\pi}^{0} \frac{\sin t}{\sqrt{2}} d\left(\frac{1}{2}+\frac{\cos t}{2}\right) \\ &\quad + \int_{2\pi}^{0} \left(\frac{1}{2}+\frac{\cos t}{2}\right) d\left(\frac{1}{2}-\frac{\cos t}{2}\right) \\ &\quad + \int_{2\pi}^{0} \left(\frac{1}{2}-\frac{\cos t}{2}\right) d\left(\frac{\sin t}{\sqrt{2}}\right) \\ &= \quad \int_{2\pi}^{0} \left(\frac{\sin t}{4}+\frac{\cos t}{2\sqrt{2}}+\frac{\cos t\sin t}{4}-\frac{1}{2\sqrt{2}}\right) dt \\ &= \quad \frac{\pi}{\sqrt{2}}. \end{aligned}$$

13.4 We parametrise the surface by letting $x = u, y = v, z = u+v^2$. Observe that the domain D of Σ is the square $[0;1]\times[0;2]$. Observe that

$$dx\wedge dy = du\wedge dv,$$

$$dy \wedge dz = -du \wedge dv,$$

$$dz \wedge dx = -2v du \wedge dv,$$

and so

$$\left\|d^2\mathbf{x}\right\| = \sqrt{2+4v^2} du \wedge dv.$$

The integral becomes

$$\begin{aligned}\int_\Sigma y \left\|d^2\mathbf{x}\right\| &= \int_0^2 \int_0^1 v\sqrt{2+4v^2} du dv \\ &= \left(\int_0^1 du\right)\left(\int_0^2 y\sqrt{2+4v^2} dv\right) \\ &= \frac{13\sqrt{2}}{3}.\end{aligned}$$

13.5 Using $x = r\cos\theta$, $y = r\sin\theta$, $1 \leqslant r \leqslant 2$, $0 \leqslant \theta \leqslant 2\pi$, the surface area is

$$\sqrt{2}\int_0^{2\pi}\int_1^2 r dr d\theta = 3\pi\sqrt{2}.$$

13.6 We use spherical coordinates, $(x,y,z) = (\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi)$. Here $\theta \in [0;2\pi]$ is the latitude and $\phi \in [0;\pi]$ is the longitude. Observe that

$$dx \wedge dy = \sin\phi\cos\phi d\phi \wedge d\theta,$$

$$dy \wedge dz = \cos\theta\sin^2\phi d\phi \wedge d\theta,$$

$$dz \wedge dx = -\sin\theta\sin^2\phi d\phi \wedge d\theta,$$

and so

$$\left\|d^2\mathbf{x}\right\| = \sin\phi d\phi \wedge d\theta.$$

The integral becomes

$$\begin{aligned}\int_\Sigma x^2 \left\|d^2\mathbf{x}\right\| &= \int_0^{2\pi}\int_0^{\pi} \cos^2\theta\sin^3\phi d\phi d\theta \\ &= \frac{4\pi}{3}.\end{aligned}$$

13.7 Put $x = u, y = v, z^2 = u^2 + v^2$. Then

$$dx = du,\ dy = dv,\ zdz = udu + vdv,$$

whence

$$dx \wedge dy = du \wedge dv, dy \wedge dz = -\frac{u}{z} du \wedge dv, dz \wedge dx = -\frac{v}{z} du \wedge dv,$$

and so

$$\begin{aligned}\left\|d^2\mathbf{x}\right\| &= \sqrt{(dx\wedge dy)^2 + (dz\wedge dx)^2 + (dy\wedge dz)^2} \\ &= \sqrt{1 + \frac{u^2+v^2}{z^2}}\ du \wedge dv \\ &= \sqrt{2}\ du \wedge dv.\end{aligned}$$

Hence

$$\int_\Sigma z\left\|d^2\mathbf{x}\right\| = \int_{u^2+v^2 \leqslant 1} \sqrt{u^2+v^2}\ \sqrt{2}\ du dv = \sqrt{2}\int_0^{2\pi}\int_0^1 \rho^2\ d\rho d\theta = \frac{2\pi\sqrt{2}}{3}.$$

13.8 If the egg has radius R, each slice will have height $2R/n$. A slice can be parametrised by $0 \leqslant \theta \leqslant 2\pi$, $\phi_1 \leqslant \phi \leqslant \phi_2$, with

$$R\cos\phi_1 - R\cos\phi_2 = 2R/n.$$

The area of the part of the surface of the sphere in slice is

$$\int_0^{2\pi}\int_{\phi_1}^{\phi_2} R^2 \sin\phi d\phi d\theta = 2\pi R^2(\cos\phi_1 - \cos\phi_2) = 4\pi R^2/n.$$

This means that each of the n slices has identical area $4\pi R^2/n$.

13.9 We project this plane onto the coordinate axes obtaining

$$\int_\Sigma xydydz = \int_0^6\int_0^{3-z/2}(3-y-z/2)ydydz = \frac{27}{4},$$

$$-\int_\Sigma x^2dzdx = -\int_0^3\int_0^{6-2x} x^2dzdx = -\frac{27}{2},$$

$$\int_\Sigma (x+z)dxdy = \int_0^3\int_0^{3-y}(6-x-2y)dxdy = \frac{27}{2},$$

and hence

$$\int_\Sigma xydydz - x^2dzdx + (x+z)dxdy = \frac{27}{4}.$$

13.10 Evaluating this directly would result in evaluating four path integrals, one for each side of the square. We will use Green's Theorem. We have

$$\begin{aligned} d\omega &= d(x^3y)\wedge dx + d(xy)\wedge dy \\ &= (3x^2ydx + x^3dy)\wedge dx + (ydx + xdy)\wedge dy \\ &= (y - x^3)dx\wedge dy. \end{aligned}$$

The region M is the area enclosed by the square. The integral equals

$$\begin{aligned} \oint_C x^3ydx + xydy &= \int_0^2\int_0^2 (y-x^3)dxdy \\ &= -4. \end{aligned}$$

13.11 We have

❶ L_{AB} is $y = x$; L_{AC} is $y = -x$, and L_{BC} is clearly $y = -\frac{1}{3}x + \frac{4}{3}$.

❷ We have

$$\begin{aligned} \int_{AB} y^2dx + xdy &= \int_0^1 (x^2+x)dx &&= \frac{5}{6} \\ \int_{BC} y^2dx + xdy &= \int_1^{-2}\left(\left(-\frac{1}{3}x+\frac{4}{3}\right)^2 - \frac{1}{3}x\right)dx &&= -\frac{15}{2} \\ \int_{CA} y^2dx + xdy &= \int_{-2}^0 (x^2-x)dx &&= \frac{14}{3} \end{aligned}$$

Adding these integrals we find

$$\oint_{\triangle} y^2 dx + x dy = -2.$$

❸ We have

$$\begin{aligned}\int_D (1-2y)dx \wedge dy &= \int_{-2}^{0}\left(\int_{-x}^{-x/3+4/3}(1-2y)dy\right)dx \\ &\quad + \int_0^1\left(\int_x^{-x/3+4/3}(1-2y)dy\right)dx \\ &= -\frac{44}{27}-\frac{10}{27} \\ &= -2.\end{aligned}$$

13.15 Observe that

$$d(x^2+2y^3)\wedge dy = 2xdx\wedge dy.$$

Hence by the generalised Stokes' Theorem the integral equals

$$\int_{\{(x-2)^2+y^2\leqslant 4\}} 2xdx\wedge dy = \int_{-\pi/2}^{\pi/2}\int_0^{4\cos\theta} 2\rho^2\cos\theta d\rho\wedge d\theta = 16\pi.$$

To do it directly, put $x-2 = 2\cos t, y = 2\sin t, 0 \leqslant t \leqslant 2\pi$. Then the integral becomes

$$\begin{aligned}\int_0^{2\pi}((2+2\cos t)^2+16\sin^3 t)d2\sin t &= \int_0^{2\pi}(8\cos t+16\cos^2 t \\ &\qquad +8\cos^3 t+32\cos t\sin^3 t)dt \\ &= 16\pi.\end{aligned}$$

13.16 At the intersection path

$$0 = x^2+y^2+z^2-2(x+y) = (2-y)^2+y^2+z^2-4 = 2y^2-4y+z^2 = 2(y-1)^2+z^2-2,$$

which describes an ellipse on the yz-plane. Similarly we get $2(x-1)^2+z^2 = 2$ on the xz-plane. We have

$$d(ydx+zdy+xdz) = dy\wedge dx+dz\wedge dy+dx\wedge dz = -dx\wedge dy-dy\wedge dz-dz\wedge dx.$$

Since $dx\wedge dy = 0$, by Stokes' Theorem the integral sought is

$$-\int_{2(y-1)^2+z^2\leqslant 2} dydz - \int_{2(x-1)^2+z^2\leqslant 2} dzdx = -2\pi(\sqrt{2}).$$

(To evaluate the integrals you may resort to the fact that the area of the elliptical region $\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}\leqslant 1$ is πab).

If we were to evaluate this integral directly, we would set

$$y = 1+\cos\theta,\ z = \sqrt{2}\sin\theta, x = 2-y = 1-\cos\theta.$$

The integral becomes

$$\int_0^{2\pi} (1+\cos\theta)d(1-\cos\theta)+\sqrt{2}\sin\theta d(1+\cos\theta)+(1-\cos\theta)d(\sqrt{2}\sin\theta)$$

which in turn

$$=\int_0^{2\pi} \sin\theta+\sin\theta\cos\theta-\sqrt{2}+\sqrt{2}\cos\theta d\theta=-2\pi\sqrt{2}.$$

B.2.18 This is clearly $(1\ 2\ 3\ 4)(6\ 8\ 7)$ of order $\mathrm{lcm}(4,3)=12$.

B.3.26 Multiplying the first column of the given matrix by a, its second column by b, and its third column by c, we obtain

$$abc\Omega=\begin{bmatrix} abc & abc & abc \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix}.$$

We may factor out abc from the first row of this last matrix thereby obtaining

$$abc\Omega=abc\det\begin{bmatrix} 1 & 1 & 1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix}.$$

Upon dividing by abc,

$$\Omega=\det\begin{bmatrix} 1 & 1 & 1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix}.$$

B.3.27 Performing $R_1+R_2+R_3\to R_1$ we have

$$\Omega=\det\begin{bmatrix} a-b-c & 2a & 2a \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{bmatrix}$$

$$=\det\begin{bmatrix} a+b+c & a+b+c & a+b+c \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{bmatrix}.$$

Factorising $(a+b+c)$ from the first row of this last determinant, we have

$$\Omega=(a+b+c)\det\begin{bmatrix} 1 & 1 & 1 \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{bmatrix}.$$

Performing $C_2 - C_1 \to C_2$ and $C_3 - C_1 \to C_3$,

$$\Omega = (a+b+c)\det \begin{bmatrix} 1 & 0 & 0 \\ 2b & -b-c-a & 0 \\ 2c & 0 & -c-a-b \end{bmatrix}.$$

This last matrix is triangular, hence

$$\Omega = (a+b+c)(-b-c-a)(-c-a-b) = (a+b+c)^3,$$

as wanted.

B.3.28 $\det A_1 = \det A = -540$ by multilinearity. $\det A_2 = -\det A_1 = 540$ by alternancy. $\det A_3 = 3\det A_2 = 1620$ by both multilinearity and homogeneity from one column. $\det A_4 = \det A_3 = 1620$ by multilinearity, and $\det A_5 = 2\det A_4 = 3240$ by homogeneity from one column.

B.3.30 From the given data, $\det B = -2$. Hence

$$\det ABC = (\det A)(\det B)(\det C) = -12,$$

$$\det 5AC = 5^3 \det AC = (125)(\det A)(\det C) = 750,$$

$$(\det A^3 B^{-3} C^{-1}) = \frac{(\det A)^3}{(\det B)^3 (\det C)} = -\frac{27}{16}.$$

B.3.31 Pick $\lambda \in \mathbb{R} \setminus \{0, a_{11}, a_{22}, \ldots, a_{nn}\}$. Put

$$X = \begin{bmatrix} a_{11}-\lambda & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22}-\lambda & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33}-\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}-\lambda \end{bmatrix}$$

and

$$Y = \begin{bmatrix} \lambda & a_{12} & a_{13} & \vdots & a_{1n} \\ 0 & \lambda & a_{23} & \vdots & a_{2n} \\ 0 & 0 & \lambda & \vdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \lambda \end{bmatrix}$$

Clearly $A = X + Y$, $\det X = (a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda) \neq 0$, and $\det Y = \lambda^n \neq 0$. This completes the proof.

B.3.32 No.

B.4.8 We have

$$\begin{aligned}\det A &= 2(-1)^{1+2}\det\begin{bmatrix}4 & 6\\ 7 & 9\end{bmatrix}+5(-1)^{2+2}\det\begin{bmatrix}1 & 3\\ 7 & 9\end{bmatrix}+8(-1)^{2+3}\det\begin{bmatrix}1 & 3\\ 4 & 6\end{bmatrix}\\ &= -2(36-42)+5(9-21)-8(6-12)=0.\end{aligned}$$

B.4.9 Simply expand along the first row

$$a\det\begin{bmatrix}a & b\\ c & a\end{bmatrix}-b\det\begin{bmatrix}c & b\\ b & a\end{bmatrix}+c\det\begin{bmatrix}c & a\\ b & c\end{bmatrix}=a(a^2-bc)-b(ca-b^2)+c(c^2-ab)=a^3+b^3+c^3-3abc.$$

B.4.10 Since the second column has three 0's, it is advantageous to expand along it, and thus we are reduced to calculate

$$-3(-1)^{3+2}\det\begin{bmatrix}1 & -1 & 1\\ 2 & 0 & 1\\ 1 & 0 & 1\end{bmatrix}$$

Expanding this last determinant along the second column, the original determinant is thus

$$-3(-1)^{3+2}(-1)(-1)^{1+2}\det\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}=-3(-1)(-1)(-1)(1)=3.$$

B.4.12 Expanding along the first column,

$$\begin{aligned}0 &= \det\begin{bmatrix}1 & 1 & 1 & 1\\ x & a & 0 & 0\\ x & 0 & b & 0\\ x & 0 & 0 & c\end{bmatrix}\\ &= \det\begin{bmatrix}a & 0 & 0\\ 0 & b & 0\\ 0 & 0 & c\end{bmatrix}-x\det\begin{bmatrix}1 & 1 & 1\\ 0 & b & 0\\ 0 & 0 & c\end{bmatrix}\\ &\quad +x\det\begin{bmatrix}1 & 1 & 1\\ a & 0 & 0\\ 0 & 0 & c\end{bmatrix}-x\det\begin{bmatrix}1 & 1 & 1\\ a & 0 & 0\\ 0 & b & 0\end{bmatrix}\\ &= xabc-xbc+x\det\begin{bmatrix}1 & 1 & 1\\ a & 0 & 0\\ 0 & 0 & c\end{bmatrix}-x\det\begin{bmatrix}1 & 1 & 1\\ a & 0 & 0\\ 0 & b & 0\end{bmatrix}.\end{aligned}$$

Expanding these last two determinants along the third row,

$$\begin{aligned}
0 &= abc - xbc + x\det\begin{bmatrix}1 & 1 & 1\\ a & 0 & 0\\ 0 & 0 & c\end{bmatrix} - x\det\begin{bmatrix}1 & 1 & 1\\ a & a & 0\\ 0 & b & 0\end{bmatrix}\\
&= abc - xbc + xc\det\begin{bmatrix}1 & 1\\ a & 0\end{bmatrix} + xb\det\begin{bmatrix}1 & 1\\ a & 0\end{bmatrix}\\
&= abc - xbc - xca - xab.
\end{aligned}$$

It follows that

$$abc = x(bc + ab + ca),$$

whence

$$\frac{1}{x} = \frac{bc+ab+ca}{abc} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c},$$

as wanted.

B.4.14 Expanding along the first row the determinant equals

$$\begin{aligned}
-a\det\begin{bmatrix}a & b & 0\\ 0 & 0 & b\\ 1 & 1 & 1\end{bmatrix} + b\det\begin{bmatrix}a & 0 & 0\\ 0 & a & b\\ 1 & 1 & 1\end{bmatrix} &= ab\det\begin{bmatrix}a & b\\ 1 & 1\end{bmatrix} + ab\det\begin{bmatrix}a & b\\ 1 & 1\end{bmatrix}\\
&= 2ab(a-b),
\end{aligned}$$

as wanted.

B.4.15 Expanding along the first row, the determinant equals

$$a\det\begin{bmatrix}a & 0 & b\\ 0 & d & 0\\ c & 0 & d\end{bmatrix} + b\det\begin{bmatrix}0 & a & b\\ c & 0 & 0\\ 0 & c & d\end{bmatrix}.$$

Expanding the resulting two determinants along the second row, we obtain

$$ad\det\begin{bmatrix}a & b\\ c & d\end{bmatrix} + b(-c)\det\begin{bmatrix}a & b\\ c & d\end{bmatrix} = ad(ad-bc) - bc(ad-bc) = (ad-bc)^2,$$

as wanted.

B.4.16 For $n = 1$ we have $\det(1) = 1 = (-1)^{1+1}$. For $n = 2$ we have

$$\det\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix} = -1 = (-1)^{2+1}.$$

Assume that the result is true for $n-1$. Expanding the determinant along the first column

$$
\begin{aligned}
\det\begin{bmatrix}
1 & 1 & 1 & \cdots & 1 & 1\\
1 & 0 & 0 & \vdots & 0 & 0\\
0 & 1 & 0 & \cdots & 0 & 0\\
0 & 0 & 1 & \cdots & 0 & 0\\
\vdots & \vdots & \cdots & \vdots & \vdots & \\
0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix}
&= 1\det\begin{bmatrix}
0 & 0 & \vdots & 0 & 0\\
1 & 0 & \cdots & 0 & 0\\
0 & 1 & \cdots & 0 & 0\\
\vdots & \cdots & \vdots & \vdots & \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}\\
&\quad -1\det\begin{bmatrix}
1 & 1 & \cdots & 1 & 1\\
1 & 0 & \cdots & \vdots & 0\\
0 & 1 & \cdots & \cdots & 0\\
0 & 0 & \cdots & \cdots & 0\\
\vdots & \vdots & \cdots & \vdots & \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}\\
&= 1(0)-(1)(-1)^n\\
&= (-1)^{n+1},
\end{aligned}
$$

giving the result.

B.4.17 Perform $C_k - C_1 \to C_k$ for $k \in [2;n]$. Observe that these operations do not affect the value of the determinant. Then

$$
\det A = \det\begin{bmatrix}
1 & n-1 & n-1 & n-1 & \cdots & n-1\\
n & 2-n & 0 & 0 & \vdots & 0\\
n & 0 & 3-n & 0 & \cdots & 0\\
n & 0 & 0 & 4-n & \cdots & 0\\
\vdots & \vdots & \vdots & \cdots & \vdots & \\
n & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

Expand this last determinant along the n-th row, obtaining,

$$
\begin{aligned}
\det A &= (-1)^{1+n} n \det \begin{bmatrix} n-1 & n-1 & n-1 & \cdots & n-1 & n-1 \\ 2-n & 0 & 0 & \vdots & 0 & 0 \\ 0 & 3-n & 0 & \cdots & 0 & 0 \\ 0 & 0 & 4-n & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix} \\
&= (-1)^{1+n} n(n-1)(2-n)(3-n) \\
&\qquad \cdots (-2)(-1) \det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \\
&= -(n!) \det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \\
&= -(n!)(-1)^n \\
&= (-1)^{n+1} n!,
\end{aligned}
$$

upon using the result of problem B.4.16.

B.4.18 Recall that $\binom{n}{k} = \binom{n}{n-k}$,

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

and

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0, \quad \text{if } n > 0.$$

Assume that n is odd. Observe that then there are $n+1$ (an even number) of columns and that on the same row, $\binom{n}{k}$ is on a column of opposite parity to that of $\binom{n}{n-k}$. By performing $C_1 - C_2 + C_3 - C_4 + \cdots + C_n - C_{n+1} \to C_1$, the first column becomes all 0's, whence the determinant if 0 if n is odd.

B.4.22 I will prove that

$$\det\begin{bmatrix}(b+c)^2 & ab & ac\\ ab & (a+c)^2 & bc\\ ac & bc & (a+b)^2\end{bmatrix}=2abc(a+b+c)^3.$$

Using permissible row and column operations,

$$\begin{aligned}\det\begin{bmatrix}(b+c)^2 & ab & ac\\ ab & (a+c)^2 & bc\\ ac & bc & (a+b)^2\end{bmatrix} &= \det\begin{bmatrix}b^2+2bc+c^2 & ab & ac\\ ab & a^2+2ca+c^2 & bc\\ ac & bc & a^2+2ab+b^2\end{bmatrix}\\ &\stackrel{C_1+C_2+C_3\to C_1}{=} \det\begin{bmatrix}b^2+2bc+c^2+ab+ac & ab & ac\\ ab+a^2+2ca+c^2+bc & a^2+2ca+c^2 & bc\\ ac+bc+a^2+2ab+b^2 & bc & a^2+2ab+b^2\end{bmatrix}\\ &= \det\begin{bmatrix}(b+c)(a+b+c) & ab & ac\\ (a+c)(a+b+c) & a^2+2ca+c^2 & bc\\ (a+b)(a+b+c) & bc & a^2+2ab+b^2\end{bmatrix}\end{aligned}$$

Pulling out a factor, the above equals

$$(a+b+c)\det\begin{bmatrix}b+c & ab & ac\\ a+c & a^2+2ca+c^2 & bc\\ a+b & bc & a^2+2ab+b^2\end{bmatrix}$$

and performing $R_1+R_2+R_3\to R_1$, this is

$$(a+b+c)\det\begin{bmatrix}2a+2b+2c & ab+a^2+2ca+c^2+bc & ac+bc+a^2+2ab+b^2\\ a+c & a^2+2ca+c^2 & bc\\ a+b & bc & a^2+2ab+b^2\end{bmatrix}$$

Factoring this is

$$(a+b+c)\det\begin{bmatrix}2(a+b+c) & (a+c)(a+b+c) & (a+b)(a+b+c)\\ a+c & a^2+2ca+c^2 & bc\\ a+b & bc & a^2+2ab+b^2\end{bmatrix},$$

which in turn is

$$(a+b+c)^2\det\begin{bmatrix}2 & a+c & a+b\\ a+c & a^2+2ca+c^2 & bc\\ a+b & bc & a^2+2ab+b^2\end{bmatrix}$$

Performing $C_2-(a+c)C_1 \to C_2$ and $C_3-(a+b)C_1 \to C_3$ we obtain

$$(a+b+c)^2 \det \begin{bmatrix} 2 & -a-c & -a-b \\ a+c & 0 & -a^2-ab-ac \\ a+b & -a^2-ab-ac & 0 \end{bmatrix}$$

This last matrix we will expand by the second column, obtaining that the original determinant is thus

$$(a+b+c)^2 \left((a+c)\det \begin{bmatrix} a+c & -a^2-ab-ac \\ a+b & 0 \end{bmatrix} + (a^2+ab+ac)\det \begin{bmatrix} 2 & -a-b \\ a+c & -a^2-ab-ac \end{bmatrix} \right)$$

This simplifies to

$$\begin{aligned} (a+b+c)^2 \Big((a+c)(a+b)(a^2+ab+ac) & \\ +(a^2+ab+ac)(-a^2-ab-ac+bc) \Big) &= a(a+b+c)^3((a+c)(a+b)-a^2-ab-ac+bc) \\ &= 2abc(a+b+c)^3, \end{aligned}$$

as claimed.

B.4.23 We have

$$\begin{aligned} \det \begin{bmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix} & \overset{R_1+R_2+R_3+R_4 \to R_1}{\underset{=}{\longrightarrow}} \det \begin{bmatrix} a+b+c+d & a+b+c+d & a+b+c+d & a+b+c+d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix} \\ &= (a+b+c+d)\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix} \\ & \overset{C_4-C_3+C_2-C_1 \to C_4}{\underset{=}{\longrightarrow}} (a+b+c+d) \begin{bmatrix} 1 & 1 & 1 & 0 \\ d & a & b & c-b+a-d \\ c & d & a & b-a+d-c \\ b & c & d & a-d+c-b \end{bmatrix} \\ &= (a+b+c+d)(a-b+c-d) \begin{bmatrix} 1 & 1 & 1 & 0 \\ d & a & b & 1 \\ c & d & a & -1 \\ b & c & d & 1 \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&\overset{R_2+R_3\to R_2,\ R_4+R_3\to R_4}{\overset{\longrightarrow}{=}} && (a+b+c+d)(a-b+c-d)\begin{bmatrix} 1 & 1 & 1 & 0 \\ d+c & a+d & b+a & 0 \\ c & d & a & -1 \\ b+c & c+d & a+d & 0 \end{bmatrix} \\
&= && (a+b+c+d)(a-b+c-d)\begin{bmatrix} 1 & 1 & 1 \\ d+c & a+d & b+a \\ b+c & c+d & a+d \end{bmatrix} \\
&\overset{C_1-C_3\to C_1,\ C_2-C_3\to C_2}{\overset{\longrightarrow}{=}} && (a+b+c+d)(a-b+c-d)\begin{bmatrix} 0 & 0 & 1 \\ d+c-b-a & d-b & b+a \\ b+c-a-d & c-a & a+d \end{bmatrix} \\
&= && (a+b+c+d)(a-b+c-d)\begin{bmatrix} d+c-b-a & d-b \\ b+c-a-d & c-a \end{bmatrix} \\
&= && (a+b+c+d)(a-b+c-d)(d+c-b-a)(c-a)-(d-b)(b+c-a-d \\
&= && (a+b+c+d)(a-b+c-d) \\
& && \quad ((c-a)(c-a)+(c-a)(d-b)-(d-b)(c-a)-(d-b)(b-d)) \\
&= && (a+b+c+d)(a-b+c-d)((a-c)^2+(b-d)^2).
\end{aligned}
$$

Since

$$(a-c)^2+(b-d)^2=(a-c+i(b-d))(a-c-i(b-d)),$$

the above determinant is then

$$(a+b+c+d)(a-b+c-d)(a+ib-c-id)(a-ib-c+id).$$

Generalisations of this determinant are possible using roots of unity.

CHAPTER 4

GNU Free Documentation License

Version 1.2, November 2002

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

Bibliography

[1] E. Abbott. *Flatland*. 7th edition. New York: Dover Publications, Inc., 1952.

[2] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, tensor analysis, and applications*. Volume 75. Springer Science & Business Media, 2012.

[3] H. Anton and C. Rorres. *Elementary Linear Algebra: Applications Version*. 8th edition. New York: John Wiley & Sons, 2000.

[4] T. M. Apostol. *Calculus, Volume I*. John Wiley & Sons, 2007.

[5] T. M. Apostol. *Calculus, Volume II*. John Wiley & Sons, 2007.

[6] V. Arnold. *Mathematical Methods of Classical Mechanics*. New York: Springer-Verlag, 1978.

[7] M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming: Theory and Algorithms*. 2nd edition. New York: John Wiley & Sons, 1993.

[8] R. L. Bishop and S. I. Goldberg. *Tensor analysis on manifolds*. Courier Corporation, 2012.

[9] A. I. Borisenko, I. E. Tarapov, and P. L. Balise. "Vector and tensor analysis with applications". In: *Physics Today* 22.2 (1969), pages 83–85.

[10] F. Bowman. *Introduction to Elliptic Functions, with Applications*. New York: Dover, 1961.

[11] M. A. P. Cabral. *Curso de Cálculo de Uma Variável*. 2013.

[12] M. P. do Carmo Valero. *Riemannian geometry*. 1992.

[13] A. Chorin and J. Marsden. *A Mathematical Introduction to Fluid Mechanics*. New York: Springer-Verlag, 1979.

[14] M. Corral et al. *Vector Calculus*. Citeseer, 2008.

[15] R. Courant. *Differential and Integral Calculus*. Volume 2. John Wiley & Sons, 2011.

[16] P. Dawkins. *Paul's Online Math Notes*. (Visited on 12/02/2015).

[17] B. Demidovitch. *Problemas e Exercícios de Análise Matemática*. 1977.

[18] T. P. Dence and J. B. Dence. *Advanced Calculus: A Transition to Analysis*. Academic Press, 2009.

[19] M. P. Do Carmo. *Differential forms and applications*. Springer Science & Business Media, 2012.

[20] M. P. Do Carmo and M. P. Do Carmo. *Differential geometry of curves and surfaces*. Volume 2. Prentice-hall Englewood Cliffs, 1976.

[21] C. H. Edwards and D. E. Penney. *Calculus and Analytic Geometry*. Prentice-Hall, 1982.

[22] H. M. Edwards. *Advanced calculus: a differential forms approach*. Springer Science & Business Media, 2013.

[23] t. b. T. H. Euclid. *Euclid's Elements*. Santa Fe, NM: Green Lion Press, 2002.

[24] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*. 2nd edition. San Diego, CA: Academic Press, 1990.

[25] P. Fitzpatrick. *Advanced Calculus*. Volume 5. American Mathematical Soc., 2006.

[26] W. H. Fleming. *Functions of several variables*. Springer Science & Business Media, 2012.

[27] D. Guichard, N. Koblitz, and H. J. Keisler. *Calculus: Early Transcendentals*. Whitman College, 2014.

[28] H. L. Guidorizzi. *Um curso de Cálculo, vol. 2*. Grupo Gen-LTC, 2000.

[29] H. L. Guidorizzi. *Um Curso de Cálculo*. Livros Técnicos e Científicos Editora, 2001.

[30] D. Halliday and R. Resnick. *Physics: Parts 1&2 Combined*. 3rd edition. New York: John Wiley & Sons, 1978.

[31] G. Hartman. *APEX Calculus II*. 2015.

[32] E. Hecht. *Optics*. 2nd edition. Reading, MA: Addison-Wesley Publishing Co., 1987.

[33] P. Hoel, S. Port, and C. Stone. *Introduction to Probability Theory*. Boston, MA: Houghton Mifflin Co., 1971 (cited on page 146).

[34] J. H. Hubbard and B. B. Hubbard. *Vector calculus, linear algebra, and differential forms: a unified approach*. Matrix Editions, 2015.

[35] J. Jackson. *Classical Electrodynamics*. 2nd edition. New York: John Wiley & Sons, 1975 (cited on page 45).

[36] L. G. Kallam and M. Kallam. "An Investigation into a Problem-Solving Strategy for Indefinite Integration and Its Effect on Test Scores of General Calculus Students." In: (1996).

[37] D. Kay. *Schaum's Outline of Tensor Calculus*. McGraw Hill Professional, 1988.

[38] M. Kline. *Calculus: An Intuitive and Physical Approach*. Courier Corporation, 1998.

[39] S. G. Krantz. *The Integral: a Crux for Analysis*. Volume 4. 1. Morgan & Claypool Publishers, 2011, pages 1–105.

[40] S. G. Krantz and H. R. Parks. *The implicit function theorem: history, theory, and applications*. Springer Science & Business Media, 2012.

[41] J. Kuipers. *Quaternions and Rotation Sequences*. Princeton, NJ: Princeton University Press, 1999.

[42] S. Lang. *Calculus of several variables*. Springer Science & Business Media, 1987.

[43] L. Leithold. *The Calculus with Analytic Geometry*. Volume 1. Harper & Row, 1972.

[44] E. L. Lima. *Análise Real Volume 1*. 2008.

[45] E. L. Lima. *Analise real, volume 2: funções de n variáveis*. Impa, 2013.

[46] I. Malta, S. Pesco, and H. Lopes. *Cálculo a Uma Variável*. 2002.

[47] J. Marion. *Classical Dynamics of Particles and Systems*. 2nd edition. New York: Academic Press, 1970.

[48] J. E. Marsden and A. Tromba. *Vector calculus*. Macmillan, 2003.

[49] P. C. Matthews. *Vector calculus*. Springer Science & Business Media, 2012.

[50] P. F. McLoughlin. "When Does a Cross Product on Rn Exist". In: *arXiv preprint arXiv:1212.3515* (2012) (cited on page 46).

[51] P. R. Mercer. *More Calculus of a Single Variable*. Springer, 2014.

[52] C. Misner, K. Thorne, and J. Wheeler. *Gravitation*. New York: W.H. Freeman & Co., 1973.

[53] J. R. Munkres. *Analysis on manifolds*. Westview Press, 1997.

[54] S. M. Musa and D. A. Santos. *Multivariable and Vector Calculus: An Introduction*. Mercury Learning & Information, 2015.

[55] O. de Oliveira et al. "The Implicit and Inverse Function Theorems: Easy Proofs". In: *Real Analysis Exchange* 39.1 (2013), pages 207–218.

[56] B. O'Neill. *Elementary Differential Geometry*. New York: Academic Press, 1966.

[57] A. Pogorelov. *Analytical Geometry*. Moscow: Mir Publishers, 1980.

[58] J. Powell and B. Crasemann. *Quantum Mechanics*. Reading, MA: Addison-Wesley Publishing Co., 1961.

[59] M. Protter and C. Morrey. *Analytic Geometry*. 2nd edition. Reading, MA: Addison-Wesley Publishing Co., 1975.

[60] J. Reitz, F. Milford, and R. Christy. *Foundations of Electromagnetic Theory*. 3rd edition. Reading, MA: Addison-Wesley Publishing Co., 1979.

[61] W. Rudin. *Principles of Mathematical Analysis*. 3rd edition. New York: McGraw-Hill, 1976.

[62] H. Schey. *Div, Grad, Curl, and All That: An Informal Text on Vector Calculus*. New York: W.W. Norton & Co., 1973.

[63] A. H. Schoenfeld. *Presenting a Strategy for Indefinite Integration*. JSTOR, 1978, pages 673–678.

[64] R. Sharipov. "Quick introduction to tensor analysis". In: *arXiv preprint math/0403252* (2004).

[65] G. F. Simmons. *Calculus with Analytic Geometry*. Volume 10. 1985, page 12.

[66] M. Spivak. *Calculus on manifolds*. Volume 1. WA Benjamin New York, 1965.

[67] M. Spivak. *Calculus*. 1984.

[68] M. Spivak. *The Hitchhiker's Guide to Calculus*. Mathematical Assn of America, 1995.

[69] J. Stewart. *Calculus: Early Transcendentals*. Cengage Learning, 2015.

[70] E. W. Swokowski. *Calculus with Analytic Geometry*. Taylor & Francis, 1979.

[71] S. Tan. *Calculus: Early Transcendentals*. Cengage Learning, 2010.

[72] A. Taylor and W. Mann. *Advanced Calculus*. 2nd edition. New York: John Wiley & Sons, 1972.

[73] J. Uspensky. *Theory of Equations*. New York: McGraw-Hill, 1948.

[74] H. Weinberger. *A First Course in Partial Differential Equations*. New York: John Wiley & Sons, 1965.

[75] S. Winitzki. *Linear algebra via exterior products*. Sergei Winitzki, 2010.