Vector Calculus

Lecture Notes

Adolfo J. Rumbos

Contents

Chapter 1

Motivation for the course

We start with the statement of the Fundamental Theorem of Calculus (FTC) in one–dimension:

Theorem 1.0.1 (Fundamental Theorem of Calculus)**.** *Let $f\colon I \to \mathbb{R}$ denote a continuous*[1] *function defined on an open interval, I, which contains the closed interval $[a,b]$, where $a, b \in \mathbb{R}$ with $a < b$. Suppose that there exists a differentiable*[2] *function $F\colon I \to \mathbb{R}$ such that*

$$F'(x) = f(x) \quad \text{for all } x \in I.$$

Then

$$\int_a^b f(x)\ dx = F(b) - F(a). \tag{1.1}$$

The main goal of this course is to extend this result to higher dimensions. To indicate how we intend to do so, we first rewrite the integral in (1.1) as follows: First, denote the interval $[a,b]$ by M; then, its boundary, denoted by ∂M, consists of the end–points a and b of the interval; thus,

$$\partial M = \{a, b\}.$$

Since $F' = f$, the expression $f(x)\ dx$ is $F'(x)\ dx$, or the *differential* of F, denoted by dF. We therefore may write the integral in (1.1) as

$$\int_a^b f(x)\ dx = \int_M dF.$$

The reason for doing this change in notation is so that, in subsequent sections in these notes, we can talk about integrals over regions M in Euclidean space, and

[1] Recall that a function $f\colon I \to \mathbb{R}$ is continuous at $c \in I$, if (i) $f(c)$ is defined, (ii) $\lim\limits_{x\to c} f(x)$ exists, and (iii) $\lim\limits_{x\to c} f(x) = f(c)$.

[2] Recall that a function $f\colon I \to \mathbb{R}$ is differentiable at $c \in I$, if $\lim\limits_{x\to c} \dfrac{f(x) - f(c)}{x - c}$ exists.

not just integrals over intervals. Thus, the concept of the integral will also have to be expanded. To see how this might come about, we discuss briefly how the right–hand side the expression in (1.1) might also be expressed as an integral.

Rewrite the right–hand side of (1.1) as the sum

$$(-1)F(a) + (+1)F(b);$$

thus, we are adding the values of the function F on the boundary of M taking into account the convention that, as we do the integration on the left–hand side of (1.1), we go from left to right along the interval $[a, b]$; hence, as we integrate, "we leave a" (this explains the -1 in front of $F(a)$) and "we enter b" (hence the $+1$ in front of $F(b)$). Since integration of a function is, in some sense, the sum of its values over a certain region, we are therefore led to suggesting that the right–hand side in (1.1) may be written as:

$$\int_{\partial M} F.$$

Thus the result of the Fundamental Theorem of Calculus in equation (1.1) may now be written in a more general form as

$$\int_M dF = \int_{\partial M} F. \tag{1.2}$$

This is known as the Generalized Stokes' Theorem and a precise statement of this theorem will be given in a subsequent section in these notes. It says that under certain conditions on the sets M and ∂M, and the "integrands," also to be made precise later in these notes, integrating the "differential" of "something" over some "set," is the same as integrating that "something" over the boundary of the set. Before we get to the stage at which we can state and prove this generalized form of the Fundamental Theorem of Calculus, we will need to introduce concepts and theory that will make the terms "something," "set" and "integration on sets" make sense. This will motivate the topics that we will discuss in this course. Here is a broad outline of what we will be studying.

- The sets M and ∂M are instances of what is known as *differentiable manifolds.* In this course, they will be subsets of n–dimensional Euclidean space satisfying certain properties that will allow us to define integration and differentiation on them.

- The manifolds M and ∂M live in n–dimensional Euclidean space and therefore we will be spending some time studying the essential properties of Euclidean space.

- The generalization of the integrands F and $\mathrm{d}F$ will lead to the study of vector valued functions (paths and vector fields) and *differential forms.*

Chapter 2

Euclidean Space

2.1 Definition of n–Dimensional Euclidean Space

Euclidean space of dimension n, denoted by $\mathbb{R}^n$, is the vector space of column vectors with real entries of the form

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where $x_1, x_2, \ldots, x_n$ are real numbers.

Remark 2.1.1 (Interpretations of Vectors)**.** In some texts on Vector Calculus, elements of $\mathbb{R}^n$ are denoted by row–vectors; in the lectures and homework assignments, we will use column vectors. The convention that I will try to follow in the lectures is that if we are interested in locating a point in space, we will use a row vector; for instance, a point P in $\mathbb{R}^n$ will be indicated by $P(x_1, x_2, \ldots, x_n)$, where $x_1, x_2, \ldots, x_n$ are the coordinates of the point. Vectors in $\mathbb{R}^n$ can also be used to locate points; for instance, the point $P(x_1, x_2, \ldots, x_n)$ is located by the vector

$$\overrightarrow{OP} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where O denotes the origin, or zero vector, in n dimensional Euclidean space. This is illustrated in Figure 2.1.1 in three–dimensional Euclidean space $\mathbb{R}^3$. In this case, we picture $\overrightarrow{OP}$ as a directed line segment ("an arrow") starting at O and ending at P.

On the other hand, the vector $\overrightarrow{OP}$ can also be used to indicate the direction of the line segment and its length; in this case, the directed line segment can be

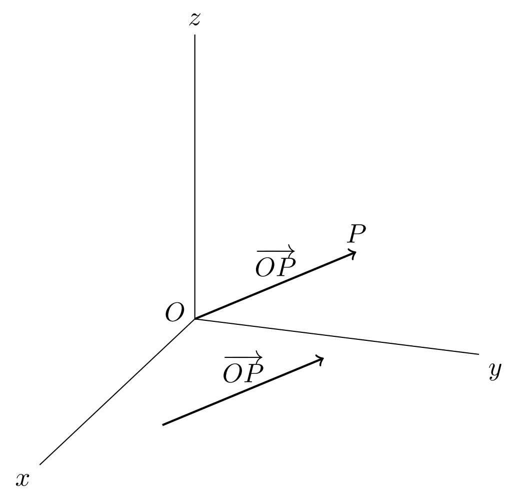

Figure 2.1.1: A point P in $\mathbb{R}^3$ and the vector OP

drawn as emanating from any point (see the illustration in Figure 2.1.1). The direction and length of the segment are what matter in the latter case.

We will say that a vector v in $\mathbb{R}^n$ is in **standard position** if the directed line segment defining v starts at the origin.

Remark 2.1.2 (Vector Notation)**.** In some texts, vectors are denoted with an arrow over the symbol for the vector; for instance, $\overrightarrow{v}$, $\overrightarrow{r}$, etc. In other texts, vectors are denoted in bold face type, $\mathbf{v}$, $\mathbf{r}$, etc. For the most part, we will do away with arrows over symbols and bold face type in these notes, lectures, and homework assignments (an exception will be the directed line segment from a point P to a point Q, denoted $\overrightarrow{PQ}$). The context will make clear whether a given symbol represents a point, a number, a vector, or a matrix.

As a vector space, $\mathbb{R}^n$ is endowed with the algebraic operations of

- **Vector Addition**

  Given $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, the vector sum $v + w$ or v and w is

$$v + w = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}. \tag{2.1}$$

Example 2.1.3. This is a 2–dimensional example.

Let $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $w = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Then,

$$v + w = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+3 \\ 2+1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

Figure 2.1.2 shows a pictorial representation of the vector addition in Example 2.1.3.

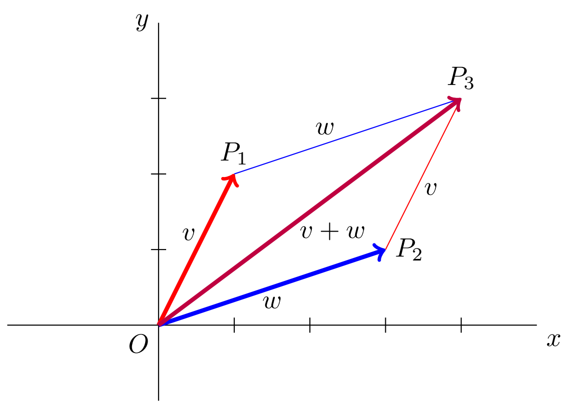

Figure 2.1.2: Parallelogram Rule

The vectors v, w and $v+w$ are drawn in standard position in Figure 2.1.2. The tip of v is then at $P_1(1,2)$, the tip of w is at $P_2(3,1)$, and the tip of $v+w$ is at $P_3(4,3)$. Figure 2.1.2 also shows the vector w translated so that its starting point is at P_1, and the vector v translated so that its starting point is at P_2. We then see that $w = \overrightarrow{P_1P_3}$ and $v = \overrightarrow{P_2P_3}$ in the figure. Consequently, O,P_1, P_2 and P_3 are the vertices of a parallelogram, since $\overrightarrow{OP_1}$ and $\overrightarrow{P_2P_3}$ are parallel to each other, and $\overrightarrow{OP_2}$ and $\overrightarrow{P_1P_3}$ are parallel to each other. Thus, vector addition is also called the **Parallelogram Rule**. This rule also shows that

$$v + w = w + v, \tag{2.2}$$

since, to go from O to P_3 along the parallelogram in Figure 2.1.2, one can go from O to P_1 and then from P_1 to P_3 (this is $v+w$), or from O to P_2 and then from P_2 from P_3 (this is $w+v$).

The parallelogram rule illustrated in Figure 2.1.2 provides a geometric interpretation of vector addition: To compute $v+w$, sketch v in standard position with its tip at point P_1. Then, translate the vector w and place it with starting point at the point P_1 and its tip at the point P_3. The vector sum is the vector from O to P_3. This is a vector along a diagonal of the parallelogram shown in Figure 2.1.2.

The vector expression in (2.2) is called the commutative property of vector addition. It can also be shown algebraically using the definition of vector addition in (2.1).

- **Scalar Multiplication**

  Given a real number t, also called a *scalar*, and a vector $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ the scaling of v by t, denoted by tv, is given by

$$tv = \begin{pmatrix} tx_1 \\ tx_2 \\ \vdots \\ tx_n \end{pmatrix}. \tag{2.3}$$

Example 2.1.4 (Scalar multiples of a vector in standard position)**.** Let

$$L = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; \begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} a \\ b \end{pmatrix}, \ t \in \mathbb{R} \right\}; \tag{2.4}$$

That is, L is the set of scalar multiples of v, where

$$v = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Assume that $a^2 + b^2 \neq 0$; that is, $\begin{pmatrix} a \\ b \end{pmatrix}$ is not the zero vector in $\mathbb{R}^2$.

Now, a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is in L, according to the definition of L in (2.4), if and only if

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} at \\ bt \end{pmatrix}, \quad \text{for some } t \in \mathbb{R},$$

from which we get the pair of equations

$$\left\{ \begin{array}{rcl} x & = & at; \\ y & = & bt, \end{array} \right. \quad \text{for } t \in \mathbb{R}. \tag{2.5}$$

The equations in (2.5) are known as the **parametric equations of the line** L. The variable t is known as a parameter. Thus, the equations in (2.5) are a parametrization of a straight line from through the origin $(0, 0)$ and the point (a, b) in $\mathbb{R}^2$. Thus, L is a straight line in the direction of the vector v. This is shown in Figure 2.1.3. Hence, all the multiples of v lie in a line through the

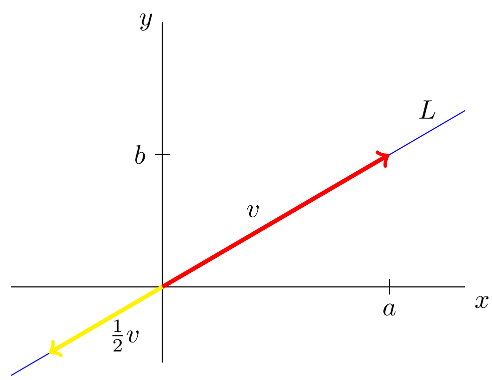

Figure 2.1.3: Line generated by v

origin along the vector v; that is, the line through the points $(0,0)$ and (a,b). We note that, if $t>0$, tv lies along the direction of v; and, if $t<0$, tv points in the opposite direction to that of v. The sketch in Figure 2.1.3 shows the vector $-\dfrac{1}{2}v$, for the case in which both a and b are assumed to be positive.

Properties of vector addition and scalar multiplication can be shown algebraically using the definitions of those operations in (2.1) and (2.3), respectively. We state some of those properties in the following proposition.

Proposition 2.1.5 (Properties of Vector Addition and Sclar Multiplication)**.** Let u, v, w denote vectors in $\mathbb{R}^n$ and t and s be scalars.

(i) **Commutativity of Vector Addition**

$$v+w=w+v.$$

(ii) **Associativity of Vector Addition**

$$(u+v)+w=u+(v+w).$$

(iii) **Existence of an Additive Identity**

The vector $\mathbf{0}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}$ in $\mathbb{R}^n$ has the property that

$$v+\mathbf{0}=\mathbf{0}+v=v \quad \text{for all } v \text{ in } \mathbb{R}^n.$$

This follows from the fact that $x+0=x$ for all real numbers x.

(iv) **Existence of an Additive Inverse**

Given $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbb{R}^n$, the vector w defined by $v = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}$ has the property that

$$v + w = \mathbf{0}.$$

The vector w is called an additive inverse of v, and is denoted by $-v$.

(v) **Associativity of Scalar Multiplication**

Given scalars t and s and a vector v in $\mathbb{R}^n$,

$$t(sv) = (ts)v.$$

(vi) **Identity in Scalar Multiplication**

The scalar 1 has the property that

$$1v = v \quad \text{for all } v \in \mathbb{R}^n.$$

(vii) **Distributive Properties**

Given vectors v and w in $\mathbb{R}^n$, and scalars t and s,

(a) $t(v + w) = tv + tw$;

(b) $(t + s)v = tv + sv$.

2.2 Spans, Lines and Planes

The span of a single vector v in $\mathbb{R}^n$ is the set of all scalar multiples of v:

$$\text{span}\{v\} = \{tv \mid t \in \mathbb{R}\}.$$

Geometrically, if v is not the zero vector in $\mathbb{R}^n$, span$\{v\}$ is the line through the origin on $\mathbb{R}^n$ in the direction of the vector v.

If P is a point in $\mathbb{R}^n$ and v is a non–zero vector also in $\mathbb{R}^n$, then the line through P in the direction of v is the set

$$\overrightarrow{OP} + \text{ span}\{v\} = \{\overrightarrow{OP} + tv \mid t \in \mathbb{R}\}.$$

A possible picture of the line $\overrightarrow{OP} +$ span$\{v\}$ is shown in Figure 2.4.8.

Example 2.2.1 (Parametric Equations of a line in $\mathbb{R}^3$). *Let* $v = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$ *be a vector in* $\mathbb{R}^3$ *and* P *the point with coordinates* $(1, 0 - 1)$. *Find the line through* P *in the direction of* v.

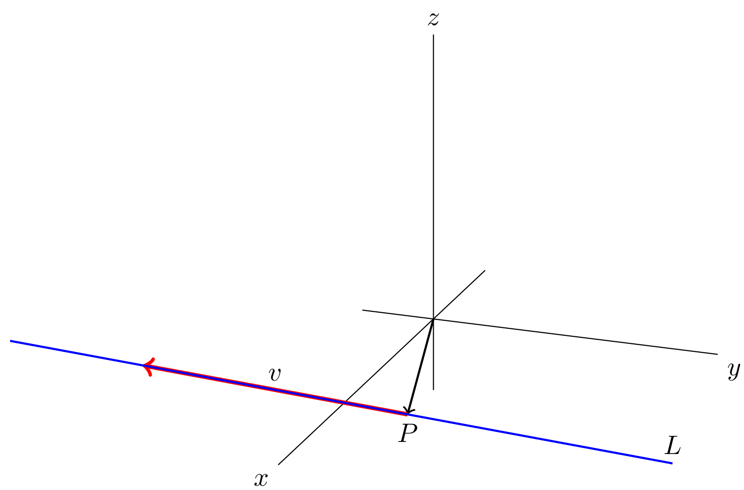

Figure 2.2.4: Line through P in the direction of v

Solution: The line through P in the direction of v is the set

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix},\ t \in \mathbb{R} \right\}$$

or

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1+2t \\ -3t \\ -1+t \end{pmatrix},\ t \in \mathbb{R} \right\}$$

Thus, for a point $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ to be on the line, x, y and z must satisfy the equations

$$\left\{ \begin{aligned} x &= 1 + 2t; \\ y &= -3t; \\ z &= -1 + t, \end{aligned} \right. \qquad \text{for } t \in \mathbb{R}. \tag{2.6}$$

A sketch of the line parametrized by the equations in (2.6) is shown in Figure 2.4.8 under the label L. $\square$

In general, the parametric equations of a line through $P(b_1, b_2, \ldots, b_n)$ in the direction of a vector $v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ in $\mathbb{R}^n$ are

$$\begin{cases} x_1 = b_1 + a_1 t \\ x_2 = b_2 + a_2 t \\ \qquad \vdots \\ x_n = b_n + a_n t \end{cases}$$

In some cases we are interested in the directed line segment from a point $P_1(x_1, x_2, \ldots, x_n)$ to a point $P_1(x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$. We will denote this set by $[P_1P_2]$; so that,

$$[P_1P_2] = \{\overrightarrow{OP_1} + t\overrightarrow{P_1P_2} \mid 0 \leqslant t \leqslant 1\}.$$

The span of two linearly independent vectors, v_1 and v_2, in $\mathbb{R}^n$ is a two–dimensional subspace of $\mathbb{R}^n$. In three–dimensional Euclidean space, $\mathbb{R}^3$, span$\{v_1, v_2\}$ is a plane through the origin containing the points located by the vectors v_1 and v_2.

If P is a point in $\mathbb{R}^3$, the plane through P generated by the linearly independent vectors v_1 and v_2, also in $\mathbb{R}^3$, is given by

$$\overrightarrow{OP} + \text{ span}\{v_1, v_2\} = \{\overrightarrow{OP} + tv_1 + sv_2 \mid t, s \in \mathbb{R}\}.$$

Example 2.2.2 (Equations of planes $\mathbb{R}^3$)**.** *Let* $v_1 = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$ *and* $v_2 = \begin{pmatrix} 6 \\ 2 \\ -3 \end{pmatrix}$ *be vectors in* $\mathbb{R}^3$ *and* P *the point with coordinates* $(1, 0 - 1)$. *Give the equation of the plane through* P *spanned by the vectors* v_1 *and* v_2.

Solution: The plane through P spanned by the vectors v_1 and v_2 is the set

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} + s \begin{pmatrix} 6 \\ 2 \\ -3 \end{pmatrix}, \ t, s \in \mathbb{R} \right\}$$

This leads to the parametric equations

$$\begin{cases} x = 1 + 2t + 6s \\ y = -3t + 2s \\ z = -1 + t - 3s. \end{cases}$$

We can write this set of parametric equations as single equation involving only x, y and z. We do this by first solving the system

$$\begin{cases} 2t + 6s = x - 1 \\ -3t + 2s = y \\ t - 3s = z + 1 \end{cases}$$

for t and s.

Using Gaussian elimination, we get can determine conditions on x, y and z that will allows us to solve for t and s:

$$\begin{pmatrix} 2 & 6 & | & x-1 \\ -3 & 2 & | & y \\ 1 & -3 & | & z+1 \end{pmatrix} \to \begin{pmatrix} 1 & 3 & | & \frac{x-1}{2} \\ -3 & 2 & | & y \\ 1 & -3 & | & z+1 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 3 & | & \frac{x-1}{2} \\ 0 & 11 & | & \frac{3}{2}(x-1)+y \\ 0 & -6 & | & -\frac{1}{2}(x-1)+(z+1) \end{pmatrix} \to \begin{pmatrix} 1 & 3 & | & \frac{x-1}{2} \\ 0 & 1 & | & \frac{3}{22}(x-1)+\frac{1}{11}y \\ 0 & -1 & | & -\frac{1}{12}(x-1)+\frac{1}{6}(z+1) \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 3 & | & \frac{x-1}{2} \\ 0 & 1 & | & \frac{3}{22}(x-1)+\frac{1}{11}y \\ 0 & 0 & | & \frac{7}{132}(x-1)+\frac{1}{11}y+\frac{1}{6}(z+1) \end{pmatrix}$$

Thus, for the system to be solvable for t and s, the third row must be a row of zeros. We therefore get the equation

$$\frac{7}{132}(x-1)+\frac{1}{11}y+\frac{1}{6}(z+1)=0$$

or

$$7(x-1)+12(y-0)+22(z+1)=0.$$

This is the equation of the plane. $\square$

In general, the equation

$$a(x-x_o)+b(y-y_o)c(z-z_o)=0$$

represents a plain in $\mathbb{R}^3$ through the point $P(x_o, y_o, z_o)$. We will see in a later section that a, b and c are the components of a vector perpendicular to the plane.

2.3 Dot Product and Euclidean Norm

Definition 2.3.1 (Euclidean Inner Product)**.** Given vectors $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, the *inner product*, or *dot product,* of v and w is the real number (or scalar), denoted by $v \cdot w$, obtained as follows

$$v \cdot w = v^T w = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

The superscript T in the above definition indicates that the column vector v has been *transposed* into a row vector.

The inner product, or dot product, defined above satisfies the following properties, which can be verified using Definition 2.3.1.

Proposition 2.3.2 (Properties of the Euclidean Inner Product)**.** Let v, v_1, v_2, w, w_1 and w_2 denote vectors in $\mathbb{R}^n$, and c_1 and c_2 be scalars. denote scalars.

(i) Positive Definiteness: $v \cdot v \geqslant 0$ for all $v \in \mathbb{R}^n$ and $v \cdot v = 0$ if and only if v is the zero vector.

(ii) Symmetry: $v \cdot w = w \cdot v$.

(iii) Bi–Linearity:
$$(c_1 v_1 + c_2 v_2) \cdot w = c_1 v_1 \cdot w + c_2 v_2 \cdot w,$$
and
$$v \cdot (c_1 w_1 + c_2 w_2) = c_1 v \cdot w_1 + c_2 v \cdot w_2,$$

for scalars c_1 and c_2.

Thus, the Euclidean inner product is a real–valued, **bi–linear form** that is **symmetric** and **positive definite**.

Given the Euclidean inner product in $\mathbb{R}^n$ defined in Definition 2.3.1, we can define the Euclidean norm as follows.

Definition 2.3.3 (Euclidean Norm in $\mathbb{R}^n$)**.** For any vector $v \in \mathbb{R}^n$, its Euclidean norm, denoted $\|v\|$, is defined by
$$\|v\| = \sqrt{v \cdot v}. \tag{2.7}$$

Observe that, by the positive definiteness of the inner product, this definition makes sense. Note also that we have defined the norm of a vector to be the *positive* square root of the the inner product of the vector with itself. Thus, the norm of any vector is always non–negative.

If P is a point in $\mathbb{R}^n$ with coordinates $(x_1, x_2, \ldots, x_n)$, the norm of the vector $\overrightarrow{OP}$ that goes from the origin to P is the distance from P to the origin; that is,
$$\text{dist}(O, P) = \|\overrightarrow{OP}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

If $P_1(x_1, x_2, \ldots, x_n)$ and $P_2(y_1, y_2, \ldots, y_n)$ are any two points in $\mathbb{R}^n$, then the distance from P_1 to P_2 is given by
$$\text{dist}(P_1, P_2) = \|\overrightarrow{OP_2} - \overrightarrow{OP_1}\| = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}.$$

Figure 2.3.5 illustrates this situation in two–dimensional Euclidean space. The vector $\overrightarrow{OP_2} - \overrightarrow{OP_1}$ is shown in the figure as the directed line segment going from P_1 to P_2. This is justified by the parallelogram rule of vector addition, since
$$\overrightarrow{OP_1} + (\overrightarrow{OP_2} - \overrightarrow{OP_1}) = \overrightarrow{OP_2},$$

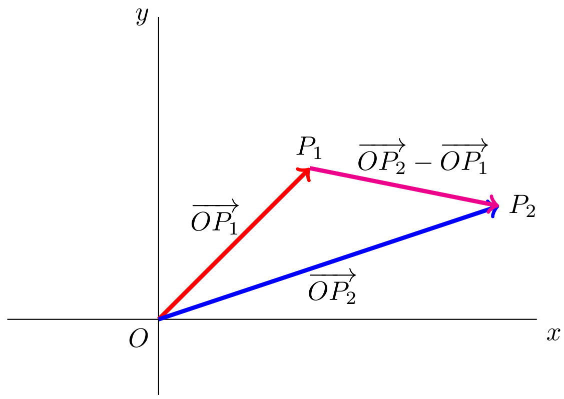

Figure 2.3.5: Distance between two points

by the properties of vector addition. The norm of $\overrightarrow{OP_2} - \overrightarrow{OP_1}$ is the length of the straight line segment connecting P_1 to P_2; consequently, $\|\overrightarrow{OP_2} - \overrightarrow{OP_1}\|$ gives the distance from P_1 to P_2.

As a consequence of the properties of the Euclidean inner product in Proposition 2.3.2, we obtain the following properties of the Euclidean norm:

Proposition 2.3.4 (Properties of the Euclidean Norm)**.** Let v denote a vector in $\mathbb{R}^n$ and c a scalar. Then,

(i) $\|v\| \geqslant 0$ and $\|v\| = 0$ if and only if v is the zero vector.

(ii) $\|cv\| = |c|\|v\|$.

We also have the following very important inequality

Theorem 2.3.5 (The Cauchy–Schwarz Inequality)**.** Let v and w denote vectors in $\mathbb{R}^n$; then,

$$|v \cdot w| \leqslant \|v\|\|w\|. \tag{2.8}$$

Furthermore, equality in (2.8) holds if and only if v and w are linearly dependent.

Proof: Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(t) = \|v - tw\|^2 \quad \text{for all} \quad t \in \mathbb{R}. \tag{2.9}$$

Using the definition of the norm, we can write

$$f(t) = (v - tw) \cdot (v - tw).$$

We can now use the properties of the Euclidean inner product in Proposition 2.3.2 to expand this expression and get

$$f(t) = \|v\|^2 - 2tv \cdot w + t^2\|w\|^2.$$

Thus, $f(t)$ is a quadratic polynomial in t that is always non–negative. Therefore, it can have at most one real root, or solution of the equation $f(t) = 0$. It then follows that the discriminant, $(2v \cdot w)^2 - 4\|w\|^2\|v\|^2$, cannot be positive; hence,

$$(2v \cdot w)^2 - 4\|w\|^2\|v\|^2 \leqslant 0,$$

from which we get

$$(v \cdot w)^2 \leqslant \|w\|^2\|v\|^2.$$

Taking square roots on both sides yields the inequality in (2.8).

Equality in (2.8) occurs when there exists $t_o \in \mathbb{R}$ for which $f(t_o) = 0$, where f is the function defined in (2.9). Consequently, $\|v - t_o w\|^2 = 0$, from which we get that $v = t_o w$, which shows that v and w are linearly dependent.

Conversely, assume that v and w are linearly dependent; so that, there exists a scalar c such that $v = cw$. Then,

$$v \cdot w = (cw) \cdot w = cw \cdot w = c\|w\|^2.$$

Consequently,

$$|v \cdot w| = |c|\|w\|^2 = |c|\|w\|\|w\| = \|cw\|\|w\|;$$

so that,

$$|v \cdot w| = \|v\|\|w\|,$$

and equality in (2.8) holds true. □

The Cauchy–Schwarz inequality, together with the properties of the inner product and the definition of the norm, yields the following inequality known as the *Triangle Inequality.*

Proposition 2.3.6 (The Triangle Inequality)**.** *For any v and w in $\mathbb{R}^n$,*

$$\|w + w\| \leqslant \|v\| + \|w\|.$$

Proof: This is an Exercise. □

Example 2.3.7 (Geometric Interpretation of the Euclidean Inner Product)**.** Let v and w denote two linearly independent vectors in $\mathbb{R}^n$ in standard position. Figure 2.3.6 pictures the special case in which the vectors v and w are in $\mathbb{R}^2$ and are in standard position. The symbol θ in the figure denotes the measure angle determined by the vectors v and w (the acute angle between the two vectors). The tips of the vectors v and w, together with the origin O, form a triangle whose sides have lengths $\|v\|$, $\|w\|$ and $\|v - w\|$. Thus, the Law of Cosines applies to yield the equation

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\| \cos(\theta). \tag{2.10}$$

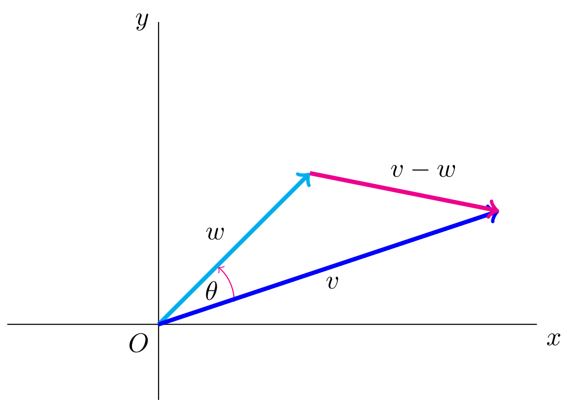

Figure 2.3.6: Law of Cosines: $\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\|\cos(\theta)$

The left–hand side in (2.10) can be expanded using the definition of the Euclidean norm in (2.7) and the properties of the Euclidean inner product in Proposition 2.3.2 to yield

$$\begin{aligned} \|v-w\|^2 &= (v-w)\cdot(v-w) \\ &= v\cdot v - v\cdot w - w\cdot v + w\cdot w \\ &= \|v\|^2 - 2v\cdot w + \|w\|^2; \end{aligned}$$

so that,

$$\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2v\cdot w. \tag{2.11}$$

Comparing the equations in (2.10) and (2.11) we then see that

$$v\cdot w = \|v\|\|w\|\cos(\theta), \tag{2.12}$$

where θ is the angle between v and w.

For the case in which v and w are nonzero vectors, the expression in (2.12) yields

$$\cos(\theta) = \frac{v\cdot w}{\|v\|\|w\|}. \tag{2.13}$$

The formula in (2.13) for two nonzero vectors, v and w, shows that the Euclidean inner-product is related to the angle, θ, between the vectors v and w. In particular, if $v\cdot w = 0$, then $\cos(\theta) = 0$; so that, $\theta = 90°$, $\theta = 270°$. In either case, the vector w lies along a line that is perpendicular to the direction of v. We say that v and w are **orthogonal**.

Definition 2.3.8 (Orthogonality)**.** Two vectors v and w in $\mathbb{R}^n$ are said to be *orthogonal,* or perpendicular, if

$$v\cdot w = 0.$$

Remark 2.3.9 (Angle between two vectors convention)**.** The formula in (2.12) gives another way for computing the the–dot product of vectors, v and w, in $\mathbb{R}^n$. It requires knowing the norms (or lengths) of the vectors, $\|v\|$ and '$\|w\|$, and the angle, θ, between the vectors v and w (see Figure 2.3.6). What do we mean by the expression "the angle between v and w?"

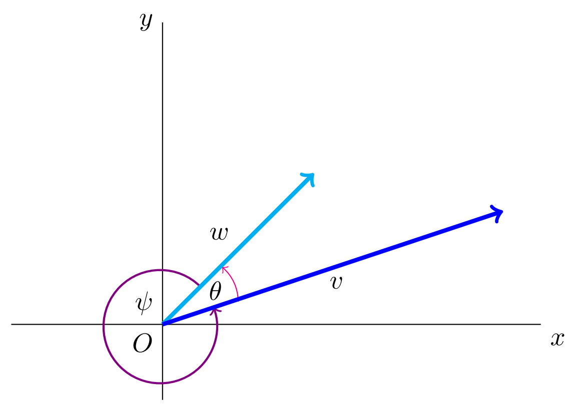

Figure 2.3.7: Angle between two vectors in $\mathbb{R}^2$

Figure 2.3.7 depicts the situation in the Cartesian plane. We see that we have two options. There is the angle θ that is swept by the arrow depicting v as we imagine it rotating towards the direction of the vector w in the counterclockwise sense. There is also the angle ψ that is swept by the arrow depicting w as it rotates in the counterclockwise sense towards the arrow depicting v. By convention, we take the angle between v and w to be the smaller of θ and ψ. For the situation pictured in Figure 2.3.7, we take θ to be the angle between v and w.

2.4 Orthogonality and Projections

We begin this section with the following geometric example.

Example 2.4.1 (Distance from a point to a line)**.** Let v denote a non–zero vector in $\mathbb{R}^n$; then, $\operatorname{span}\{v\}$ is a line through the origin in the direction of v. Given a point P in $\mathbb{R}^3$ that is not in the span of v, we would like to find the distance from P to the line; in other words, the shortest distance from P to any point on the line. There are two parts in the solution of this problem:

- first, locate the point, tv, on the line that is closest to P, and
- second, compute the distance from that point to P.

Figure 2.4.8 shows a sketch of the line in $\mathbb{R}^3$ representing $\operatorname{span}\{v\}$.

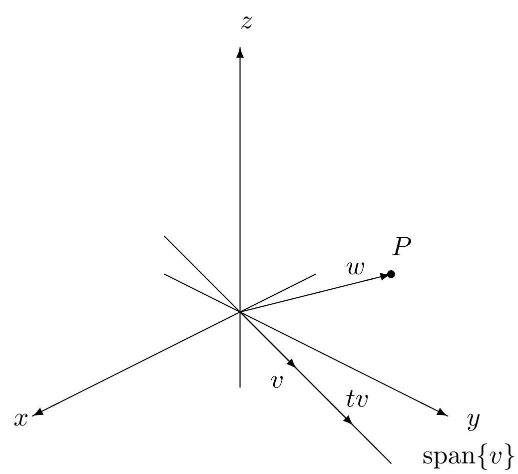

Figure 2.4.8: Line in $\mathbb{R}^3$

Let $w = \overrightarrow{OP}$ denote the vector from the origin to P (see sketch in Figure 2.4.8), and define the function $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(t) = \|w - tv\|^2, \quad \text{for any} \quad t \in \mathbb{R};$$

that is, $f(t)$ is the square of the distance from P to any point on the line through O in the direction of v. We wish to minimize this function.

Observe that $f(t)$ can be written in terms of the dot product as

$$f(t) = (w - tv) \cdot (w - tv),$$

which can be expanded by virtue of the properties of the inner product and the definition of the Euclidean norm into

$$f(t) = \|w\|^2 - 2tv \cdot w + t^2\|v\|^2.$$

Thus, $f(t)$ is a quadratic polynomial in t that can be shown to have an absolute minimum when

$$t = \frac{v \cdot w}{\|v\|^2}.$$

Thus, the point on $\operatorname{span}\{v\}$ that is the closest to P is the point

$$\frac{v \cdot w}{\|v\|^2} v,$$

where $w = \overrightarrow{OP}$.

The distance form P to the line (i.e., the shortest distance) is then

$$\left\| w - \frac{v \cdot w}{\|v\|^2} v \right\|.$$

Remark 2.4.2. *The argument of the previous example can be used to show that the point on the line*

$$\overrightarrow{OP_o} + \ span\{v\},$$

for a given point P_o, which is the closest to P is given by

$$\overrightarrow{OP_o} + \frac{v \cdot w}{\|v\|^2} v,$$

where $w = \overrightarrow{P_oP}$, and the distance from P to the line is

$$\left\| \overrightarrow{OP_o} + \frac{v \cdot w}{\|v\|^2} v - w \right\|.$$

Definition 2.4.3 (Orthogonal Projection)**.** *The vector*

$$\frac{v \cdot w}{\|v\|^2} v$$

is called the orthogonal projection *of w onto v. We denote it by $P_v(w)$. Thus,*

$$P_v(w) = \frac{(v \cdot w)}{\|v\|^2}\ v.$$

$P_v(w)$ is called the orthogonal projection of $w = \overrightarrow{OP}$ onto v because it lies along a line through P that is perpendicular to the direction of v. To see why this is the case compute

$$\begin{aligned}
(w - P_v(w)) \cdot v &= w \cdot v - P_v(w) \cdot v \\
&= w \cdot v - \frac{w \cdot v}{\|v\|^2} v \cdot v \\
&= w \cdot v - \frac{w \cdot v}{\|v\|^2} \|v\|^2 \\
&= w \cdot v - w \cdot v \\
&= 0;
\end{aligned}$$

so that, the vector $w - P_v(w)$ is perpendicular to v. We have therefore shown that the point in the line span($\{v\}$) that is closest to P lies on line through P that is perpendicular to the line through O in the direction of v. See the sketch in Figure 2.4.9 for an illustration of this result in two–dimensional Euclidean space.

Note that any vector $w \in \mathbb{R}^n$ can be written as

$$w = P_v(w) + (w - P_v(w)). \tag{2.14}$$

Thus, by the previous calculations we also see that any vector, w, in $\mathbb{R}^n$ can be expressed as the sum of a vector that is parallel to v and another vector that is perpendicular to v. See the sketch in Figure 2.4.9. The expression in (2.14) is known as the **orthogonal decomposition** of w with respect to v.

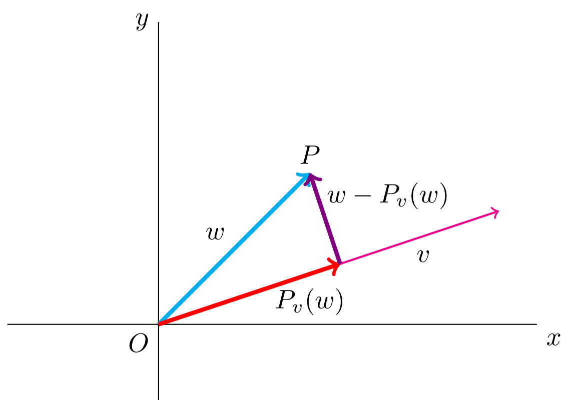

Figure 2.4.9: Orthogonal Projection

Example 2.4.4. Let L denote the line given parametrically by the equations

$$\left\{\begin{array}{rcl} x & = & 1-t \\ y & = & 2t \\ z & = & 2+t, \end{array}\right. \tag{2.15}$$

for $t \in \mathbb{R}$. Find the point on the line, L, which is closest to the point $P(1,2,0)$ and compute the distance from P to L.
Solution: Let P_o be the point on L with coordinates $(1,0,2)$ (note that P_o is the point in $\mathbb{R}^3$ corresponding to $t=0$). Put

$$w = \overrightarrow{P_oP} = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix}.$$

Let $v = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$; that is, v is the direction of the line, L. Hence, any point on L is of the form $\overrightarrow{OP_o} + tv$, for some t in $\mathbb{R}$.

The point on the line L that is the closest to P is

$$\overrightarrow{OP_o} + P_v(w),$$

where $P_v(w)$ is the orthogonal projection of w onto v; that is,

$$P_v(w) = \frac{(v \cdot w)}{\|v\|^2}\, v = \frac{2}{6} v = \frac{1}{3} v.$$

Thus, the point on L that is closest to P corresponds to $t = 1/3$ in (2.15); that is, the point $Q(2/3, 2/3, 7/3)$ is the point on L that is closest to P.

The distance form P to the line L is

$$\begin{aligned} \operatorname{dist}(P, L) &= \operatorname{dist}(P, Q) \\ &= \|\overrightarrow{OP} - \overrightarrow{OQ}\|, \end{aligned}$$

so that

$$\begin{aligned} \operatorname{dist}(P, L) &= \left\| \begin{pmatrix} 1/3 \\ 4/3 \\ -7/3 \end{pmatrix} \right\| \\ &= \frac{1}{3} \left\| \begin{pmatrix} 1 \\ 4 \\ -7 \end{pmatrix} \right\| \\ &= \frac{1}{3}\sqrt{1+16+49} = \frac{\sqrt{66}}{3}. \end{aligned}$$

□

Definition 2.4.5 (Unit Vectors)**.** A vector $u \in \mathbb{R}^n$ is said to be a **unit vector** if $\|u\| = 1$; that is, u has unit length.

If u is a unit vector in $\mathbb{R}^n$, then the orthogonal projection of $w \in \mathbb{R}^n$ onto u is given by

$$P_u(w) = (w \cdot u)u.$$

We call this vector the orthogonal component of w in the direction of u.

If v is a non–zero vector in $\mathbb{R}^n$, we can scale v to obtain a unit vector in the direction of v as follows: $\dfrac{1}{\|v\|}v$.

Denote this vector by $\widehat{v}$; then, $\widehat{v} = \dfrac{1}{\|v\|}v$ and

$$\|\widehat{v}\| = \left\| \frac{1}{\|v\|}v \right\| = \frac{1}{\|v\|}\|v\| = 1;$$

so that, $\widehat{v}$ is indeed a unit vector.

As a convention, we will always denote unit vectors in a given direction with a hat upon the symbol denoting the direction vector.

Example 2.4.6. The vectors $\widehat{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\widehat{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\widehat{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are unit vectors in $\mathbb{R}^3$. Observe also that they are mutually orthogonal; that is

$$\widehat{i} \cdot \widehat{j} = 0, \quad \widehat{i} \cdot \widehat{k} = 0, \quad \text{and} \quad \widehat{j} \cdot \widehat{k} = 0.$$

Note also that every vector v in $\mathbb{R}^3$ can be written us

$$v = (v \cdot \widehat{i})\widehat{i} + (v \cdot \widehat{j})\widehat{j} + (v \cdot \widehat{k})\widehat{k}.$$

This is known as the orthogonal decomposition of v with respect to the basis $\{\widehat{i}, \widehat{j}, \widehat{k}\}$ in $\mathbb{R}^3$.

Example 2.4.7 (Normal Direction to a Plane in $\mathbb{R}^3$)**.** We have seen that the set of points, (x, y, z), satisfying the equation

$$ax + by + cz = d \tag{2.16}$$

forms a plane in $\mathbb{R}^3$. We call (2.16) the equation of a plane determined by the real constants a, b, c and d.

Suppose that $P_o(x_o, y_o, z_o)$ is a point on the plane represented by the equation in (2.16). Then,

$$ax_o + by_o + cz_o = d. \tag{2.17}$$

Similarly, if $P(x, y, y)$ is another point on the plane, then

$$ax + by + c = d. \tag{2.18}$$

Subtracting equation (2.17) from equation (2.18) we then obtain that

$$a(x - x_o) + b(y - y_o) + c(z - z_o) = 0.$$

This is the general equation of a plane derived in a previous example. This equation can be interpreted as saying that the dot product of the vector $n = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ with the vector $\overrightarrow{P_oP} = \begin{pmatrix} x - x_o \\ y - y_o \\ z - z_o \end{pmatrix}$ is zero. Thus, the vector n is orthogonal, or perpendicular, to any vector lying on the plane. We then say that n is *normal vector* to the plane. In the next section we will see how to obtain a normal vector to the plane determined by three non–collinear points in three–dimensional space.

Example 2.4.8 (Distance from a point to a plane)**.** Let H denote the plane in $\mathbb{R}^3$ given by

$$H = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; ax + by + cz = d \right\}.$$

Let P denote a point which is not on the plane H. Find the shortest distance from the point P to H.

Solution: Let $P_o(x_o, y_o, z_o)$ be any point in the plane, H, and define the vector, $w = \overrightarrow{P_oP}$, which goes from the point P_o to the point P. The shortest distance from P to the plane will be the norm of the projection of w onto the orthogonal direction vector,

$$n = \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

to the plane H. Then,

$$\operatorname{dist}(P,H) = \|P_n(w)\|,$$

where $P_n(w) = \dfrac{w\cdot n}{\|n\|^2}n$. □

Example 2.4.9. Let H be the plane in $\mathbb{R}^3$ given by the equation

$$2x + 3y + 6z = 6.$$

Find the distance from H to $P(0,2,2)$.
Solution: Let P_o denote the z–intercept of the plane; namely, $P_o(0,0,1)$, and put

$$w = \overrightarrow{P_oP} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}.$$

Then, according to the result of Example 2.4.8,

$$\operatorname{dist}(P,H) = \frac{|w\cdot n|}{\|n\|},$$

where

$$n = \begin{pmatrix} 2 \\ 3 \\ 6 \end{pmatrix},$$

so that

$$w\cdot n = 12,$$

and

$$\|n\| = \sqrt{4+9+36} = 7.$$

Consequently,

$$\operatorname{dist}(P,H) = \frac{12}{7}.$$

□

2.5 The Cross Product in $\mathbb{R}^3$

We begin this section by first showing how to compute the area of parallelogram determined by two linearly independent vectors in $\mathbb{R}^2$.

Example 2.5.1 (Area of a Parallelogram)**.** Let v and w denote two linearly independent vectors in $\mathbb{R}^2$ given by

$$v = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Figure 2.5.10 shows a sketch of the arrows representing v and w for the special case in which they lie in the first quadrant of the xy–plane.

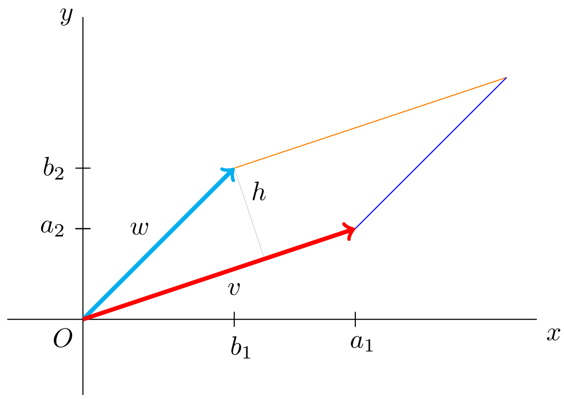

Figure 2.5.10: Area of a parallelogram

We would like to compute the area of the parallelogram, $\mathcal{P}(v, w)$, determined by v and w. This may be computed as follows:

$$\text{area}(\mathcal{P}(v, w)) = \|v\| h,$$

where h is given by $\|w - P_v(w)\|$; that is, the distance from w to its orthogonal projection along v.

Squaring both sides of the previous equation we have that

$$\begin{aligned}
(\text{area}(\mathcal{P}(v, w)))^2 &= \|v\|^2 \|w - P_v(w)\|^2 \\
&= \|v\|^2 (w - P_v(w)) \cdot (w - P_v(w)) \\
&= \|v\|^2 (\|w\|^2 - 2w \cdot P_v(w) + \|P_v(w)\|^2) \\
&= \|v\|^2 \left(\|w\|^2 - 2w \cdot \frac{(v \cdot w)}{\|v\|^2} v + \frac{(v \cdot w)^2}{\|v\|^2} \right) \\
&= \|v\|^2 \left(\|w\|^2 - 2 \frac{(v \cdot w)}{\|v\|^2} w \cdot v + \frac{(v \cdot w)^2}{\|v\|^2} \right) \\
&= \|v\|^2 \left(\|w\|^2 - 2 \frac{(v \cdot w)^2}{\|v\|^2} + \frac{(v \cdot w)^2}{\|v\|^2} \right).
\end{aligned}$$

We then obtain that

$$(\text{area}(\mathcal{P}(v, w)))^2 = \|v\|^2 \|w\|^2 - (v \cdot w)^2. \tag{2.19}$$

We note that the expression for the area of the parallelogram determined by vectors v and w given in (2.19) also holds true for vectors in $\mathbb{R}^n$.

Writing the expression in (2.19) in terms of the coordinates of v and w, we then have that

$$\begin{aligned}
(\text{area}(\mathcal{P}(v,w)))^2 &= \|v\|^2\|w\|^2 - (v\cdot w)^2 \\
&= (a_1^2+a_2^2)(b_1^2+b_2^2) - (a_1b_1+a_2b_2)^2 \\
&= a_1^2b_1^2 + a_1^2b_2^2 + a_2^2b_1^2 + a_2^2b_2^2 - (a_1^2b_1^2 + 2a_1b_1a_2b_2 + a_2^2b_2^2) \\
&= a_1^2b_2^2 + a_2^2b_1^2 - 2a_1b_1a_2b_2 \\
&= a_1^2b_2^2 - 2(a_1b_2)(a_2b_1) + a_2^2b_1^2;
\end{aligned}$$

so that,

$$(\text{area}(\mathcal{P}(v,w)))^2 = (a_1b_2 - a_2b_1)^2. \tag{2.20}$$

Taking square roots on both sides of (2.20), we get

$$\text{area}(\mathcal{P}(v,w)) = |a_1b_2 - a_2b_1|.$$

Observe that the expression in the absolute value on the right-hand side of the previous equation is the determinant of the matrix:

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

We then have that the area of the parallelogram determined by v and w is the absolute value of the determinant of a 2×2 matrix whose columns are the vectors v and w. If we denote the matrix by $[v\ w]$, then we obtain the formula

$$\text{area}(\mathcal{P}(v,w)) = |\det([v\ w])|.$$

Observe that this formula works even in the case in which v and w are not linearly independent. In that case we get that the area of the parallelogram determined by the two vectors is 0.

2.5.1 Defining the Cross–Product

Given two linearly independent vectors, v and w, in $\mathbb{R}^3$, we would like to associate to them a vector, denoted by $v\times w$ and called the *cross product* of v and w, satisfying the following properties:

- $v\times w$ is perpendicular to the plane spanned by v and w.
- There are two choices for a perpendicular direction to the span of v and w. The direction for $v\times w$ is determined according to the so called "right–hand rule":

  > *With the fingers of your right hand, follow the direction of v while curling them towards the direction of w. The thumb will point in the direction of $v\times w$.*

- The norm of $v \times w$ is the area of the parallelogram, $\mathcal{P}(v,w)$, determined by the vectors v and w.

These properties imply that the cross product is not a symmetric operation; in fact, it is antisymmetric:

$$w \times v = -v \times w \quad \text{for all} \quad v, w \in \mathbb{R}^3.$$

From this property we immediately get that

$$v \times v = \mathbf{0} \quad \text{for all} \quad v \in \mathbb{R}^3,$$

where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^3$.

Putting the properties defining the cross product together we get that

$$v \times w = \pm \text{area}(\mathcal{P}(v,w))\widehat{n}, \tag{2.21}$$

where $\widehat{n}$ is a unit vector perpendicular to the plane determined by v and w. The sign on the right–hand side of (2.21) is determined by the right hand rule and, according to 2.19, area($\mathcal{P}(v,w)$ satisfies the equation

$$(\text{area}(\mathcal{P}(v,w)))^2 = \|v\|^2\|w\|^2 - (v \cdot w)^2. \tag{2.22}$$

To compute $v \times w$, we first consider the special case in which v and w lie along the xy–plane. More specifically, suppose that

$$v = \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}.$$

Figure 2.5.10 illustrates the situation in which v and w lie on the first quadrant of the xy–plane.

For the situation shown in Figure 2.5.10, according to (2.21), $v \times w$ is in the direction of $\widehat{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. We then have, in view of (2.21), that

$$v \times w = \pm \text{area}(\mathcal{P}(v,w))\widehat{k}, \tag{2.23}$$

where the area of the parallelogram $\mathcal{P}(v,w)$ is computed as in Example 2.5.1 to obtain

$$\text{area}(\mathcal{P}(v,w)) = \left| \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \right|$$

To determine the sign on the right–hand side of (2.23), we apply the right–hand rule. It turns out that the sign of the determinant of the matrix

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix},$$

where the entries in the first column correspond to v and those in the second column to w, agree with the sign convention dictated by the right–hand–rule. We then have that

$$v \times w = \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \widehat{k}. \tag{2.24}$$

To simplify notation in (2.24), we will write $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ for $\det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$. Thus,

$$v \times w = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \widehat{k}. \tag{2.25}$$

Observe that, since the determinant of the transpose of a matrix is the same as that of the matrix, we can also write (2.25) as

$$v \times w = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \widehat{k}, \tag{2.26}$$

for vectors

$$v = \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$$

lying in the xy–plane.

In general, the cross product of the vectors

$$v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

in $\mathbb{R}^3$ is the vector

$$v \times w = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \widehat{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \widehat{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \widehat{k}, \tag{2.27}$$

where $\widehat{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\widehat{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\widehat{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are the *standard basis* vectors in $\mathbb{R}^3$.

Observe that if $a_3 = b_3 = 0$ in definition on $v \times w$ in (2.27), we recover the expression in (2.26),

$$v \times w = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \widehat{k}$$

for the cross product of vectors lying entirely in the xy–plane.

In the remainder of this section, we verify that the cross product of two vectors, v and w, in $\mathbb{R}^3$ defined in the (2.27) does indeed satisfies the properties listed at the beginning of the section.

To check that $v \times w$ is orthogonal to the plane spanned by v and w, write

$$v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

and compute the dot product of v and $v \times w$,

$$\begin{aligned} v \cdot (v \times w) &= v^T (v \times w) \\ &= \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \\ \\ -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \\ \\ \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \end{pmatrix} \end{aligned}$$

so that

$$v \cdot (v \times w) = a_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - a_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + a_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}. \tag{2.28}$$

We recognize in the right–hand side of equation (2.28) the expansion along the first row of the determinant

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix},$$

which is 0 because the first two rows are the same. Thus,

$$v \cdot (v \times w) = 0,$$

and therefore $v \times w$ is orthogonal to v. Similarly, we can compute

$$w \cdot (v \times w) = \begin{vmatrix} b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = 0,$$

which shows that $v \times w$ is also orthogonal to w. Hence, $v \times w$ is orthogonal to the span of v and w.

Next, we need to see that $\|v \times w\|$ gives the area of the parallelogram spanned by v and w. According to (2.22), we need to show that

$$\|v \times w\|^2 = \|v\|^2 \|w\|^2 - (v \cdot w)^2. \tag{2.29}$$

To establish (2.29), use the definition of $v \times w$ in (2.27) to compute

$$\begin{aligned}
\|v \times w\|^2 &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2 \\
&= (a_2b_3 - a_3b_2)^2 + (a_1b_3 - a_3b_1)^2 + (a_1b_2 - a_2b_1)^2 \\
&= a_2^2b_3^2 - 2a_2b_3a_3b_2 + a_3^2b_2^2 \\
&\qquad + a_1^2b_3^2 - 2a_1b_3a_3b_1 + a_3^2b_1^2 \\
&\qquad\quad + a_1^2b_2^2 - 2a_1b_2a_2b_1 + a_2^2b_1^2.
\end{aligned}$$

Next, add and subtract the binomials $a_1^2b_1^2$, $a_2^2b_2^2$ and $a_3^2b_3^2$ on the right–hand side of the previous expression, and simplify to obtain

$$\begin{aligned}
\|v \times w\|^2 &= a_1^2(b_1^2 + b_2^2 + b_3^2) \\
&\qquad + a_2^2(b_1^2 + b_2^2 + b_3^2) \\
&\qquad\quad + a_3^2(b_1^2 + b_2^2 + b_3^2) \\
&\qquad\qquad - 2a_2b_3a_3b_2 - 2a_1b_3a_3b_1 - 2a_1b_2a_2b_1 \\
&\qquad\qquad\quad - a_1^2b_1^2 - a_2^2b_2^2 - a_3^2b_3^2,
\end{aligned}$$

or

$$\begin{aligned}
\|v \times w\|^2 &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) \\
&\qquad - (a_1^2b_1^2 + a_2^2b_2^2 + a_3^2b_3^2 + 2a_2b_3a_3b_2 + 2a_1b_3a_3b_1 + 2a_1b_2a_2b_1),
\end{aligned}$$

which simplifies to

$$\|v \times w\|^2 = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_2)^2, \tag{2.30}$$

Noticing on the right–hand side of (2.30) the product of the square of the norms of v and w minus the square of the dot product of v and w, we see that the expression in (2.30) implies the statement in (2.29). Consequently, in view of (2.22), we see that (2.29) can we written as

$$\|v \times w\|^2 = [\text{area}(\mathcal{P}(v, w)]^2,$$

from which it follows that

$$\|v \times w\| = \text{area}(\mathcal{P}(v, w)). \tag{2.31}$$

Consequently, the norm of the cross–product of vectors v and w in $\mathbb{R}^3$ gives the area of the parallelogram determined by the vectors v and w, which was to be shown.

2.5.2 Triple Scalar Product

Example 2.5.2 (Volume of a Parallelepiped). Three linearly independent vectors, u, v and w, in $\mathbb{R}^3$ determine a solid figure called a parallelepiped (see Figure 2.5.11 on page 33). In this section, we see how to compute the volume of that object, which we shall denote by $\mathcal{P}(u, v, w)$.

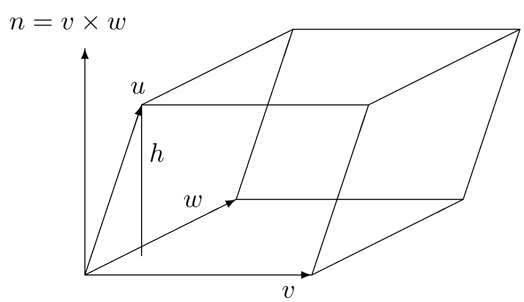

Figure 2.5.11: Volume of a Parallelepiped

First, observe that the volume of the parallelepiped, $\mathcal{P}(v, w, u)$, drawn in Figure 2.5.11 is the area of the parallelogram spanned by v and w times the height, h, of the parallelepiped:

$$\text{volume}(P(v, w, u)) = \text{area}(P(v, w)) \cdot h, \tag{2.32}$$

where h can be obtained by projecting u onto the cross–product, $v \times w$, of v and w; that is

$$h = \|P_n(u)\| = \left\| \frac{u \cdot n}{\|n\|^2}\, n \right\|,$$

where

$$n = v \times w.$$

We then have that

$$h = \frac{|u \cdot (v \times w)|}{\|v \times w\|}.$$

Consequently, since area$(\mathcal{P}(v, w)) = \|v \times w\|$, according to (2.31), we get from (2.32) that

$$\text{volume}(P(v, w, u)) = |u \cdot (v \times w)|. \tag{2.33}$$

The scalar, $u \cdot (v \times w)$, in the right–hand side of the equation in (2.33) is called the triple scalar product of u, v and w.

Given three vectors

$$u = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, \quad v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

in $\mathbb{R}^3$, the *triple scalar product* of u, v and w is given by

$$u \cdot (v \times w) = c_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix},$$

or

$$u \cdot (v \times w) = \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix};$$

that is, $u \cdot (v \times w)$ is the the determinant of the 3×3 matrix whose rows are the vector u, v and w, in that order. Since the determinant of the transpose of a matrix is the same as the determinant of the original matrix, we may also write

$$u \cdot (v \times w) = \det[\, u \quad v \quad w \,],$$

the determinant of the 3×3 matrix whose columns are the vector u, v and w, in that order.

Chapter 3

Functions

3.1 Types of Functions in Euclidean Space

Given a subset D of n–dimensional Euclidean space, $\mathbb{R}^n$, we are interested in functions that map D to m–dimensional Euclidean space, $\mathbb{R}^m$, where n and m could possibly be the same. We write

$$F\colon D \to \mathbb{R}^m$$

and call D the *domain* of F; that is, the set where the function is defined.

Example 3.1.1. *The function f given by*

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

is defined over the set

$$D = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\},$$

or the open unit disc in $\mathbb{R}^2$. In this case, $n = 2$ and $m = 1$.

There are different types of functions that we will be studying in this course. Some of the types have received traditional names, and we present them here.

- *Vector Fields.* If $m = n > 1$, then the map

  $$F\colon D \to \mathbb{R}^n$$

  is called a vector field on D. The idea here is that each point in D gets assigned a vector. A picture for this is provided by a model of fluid flow in which each point in region where fluid is flowing gets assigned a vector giving the velocity of the flow at that particular point.

- *Scalar Fields.* For the case in which $m = 1$ and $n > 1$, every point in D now gets assigned a scalar (a real number). An example of this in applications would be the temperature distribution over a region in space. Scalar fields in this course will usually be denoted by lower case letters (f, g, etc.). The value of a scalar field

  $$f\colon D \to \mathbb{R}$$

  at a point $P(x_1, x_2, \ldots, x_n)$ in D will be denoted by

  $$f(x_1, x_2, \ldots, x_n).$$

  If D is a region in the xy–plane, we simply write

  $$f(x, y) \quad \text{for} \quad (x, y) \in D.$$

- *Paths.* If $n = 1$, $m > 1$ and D is an interval, I, of real line, then the map

  $$\sigma\colon I \to \mathbb{R}^m$$

  is called a path in $\mathbb{R}^m$.

  Example 3.1.2. *Let $\sigma(t) = (\cos t, \sin t)$ for $t \in (-\pi, \pi]$, then*

  $$\sigma\colon (-\pi, \pi] \to \mathbb{R}^2$$

  is a path in $\mathbb{R}^2$. A picture of this map would a particle in the xy–plane moving along the unit circle in the counterclockwise direction.

3.2 Open Subsets of Euclidean Space

In Example 3.1.1 we saw that the function f given by

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

has the open unit disc, $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$, as its domain. The set D is an example of what is known as an *open set.*

Definition 3.2.1 (Open Balls)**.** *Given $u \in \mathbb{R}^n$, the open ball of radius $r > 0$ in $\mathbb{R}^n$ about u is defined to be the set*

$$B_r(u) = \{v \in \mathbb{R}^n \mid \|v - u\| < r\}.$$

That is, $B_r(u)$ is the set of points in $\mathbb{R}^n$ that are within a distance of r from u.

Definition 3.2.2 (Open Sets)**.** *A set $U \subseteq \mathbb{R}^n$ is said to be open if and only if for every $u \in U$ there exists $r > 0$ such that*

$$B_r(u) \subseteq U.$$

The empty set, $\emptyset$, is considered to be open.

Example 3.2.3. *For any $R > 0$, the open ball $B_R(O) = \{v \in \mathbb{R}^n \mid \|v\| < R\}$ is an open set.*

Proof. Let u be an arbitrary point in $B_R(O)$; then $\|u\| < R$. Put $r = R - \|u\| > 0$ and consider the open ball $B_r(u)$. If $v \in B_r(u)$, then, by the triangle inequality,

$$\|v\| = \|v - u + u\| \leqslant \|v - u\| + \|u\| < r + \|u\| = R,$$

which shows that $v \in B_R(O)$. Consequently,

$$B_r(u) \subseteq B_R(O).$$

It then follows that $B_R(O)$ is open by Definition 3.2.2. □

Example 3.2.4. *The set $A = \{(x, y) \in \mathbb{R}^2 \mid y = 0\}$ is not an open subset of $\mathbb{R}^2$. To see why this is the case, observe that for any $r > 0$, the ball $B_r((0, 0))$ is is not a subset of A, since, for instance, the point $(0, r/2)$ is in $B_r((0, 0))$, but it is not an element of A.*

3.3 Continuous Functions

In single variable Calculus you learned that a real valued function, $f\colon (a, b) \to \mathbb{R}$, defined in the open interval (a, b), is continuous at $c \in (a, b)$ if

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

We may rewrite the last expression as

$$\lim_{|x-c| \to 0} |f(x) - f(c)| = 0.$$

This is the expression that we will use to generalize the notion of continuity of vector–valued functions at points in subsets of Euclidean space. We simply replace the absolute values with norms, when needed

Definition 3.3.1. *Let U be an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a vector–valued map on U. F is said to be continuous at $u \in U$ if*

$$\lim_{\|v-u\| \to 0} \|F(v) - F(u)\| = 0.$$

If F is continuous at every u in U, we say that F is continuous on U.

We present several examples of continuous functions.

Example 3.3.2. Let $T\colon \mathbb{R}^n \to \mathbb{R}$ be a linear transformation. Then, T is continuous on $\mathbb{R}^n$.

Proof: Since T is linear, there exists a vector, w, in $\mathbb{R}^n$ such that

$$T(v) = w \cdot v \quad \text{for all} \quad v \in \mathbb{R}^n. \tag{3.1}$$

This fact was shown to be true in Problem 4 of Assignment #3.

It follows from (3.1) that, for any u and v in $\mathbb{R}^n$,

$$\begin{aligned} T(v) - T(u) &= w \cdot v - w \cdot u \\ &= w \cdot (v - u); \end{aligned}$$

so that

$$|T(v) - T(u)| = |w \cdot (v - u)|, \quad \text{for } v, u \in \mathbb{R}^n. \tag{3.2}$$

Thus, applying the Cauchy–Schwartz inequality on the right-hand side of (3.2) then yields

$$|T(v) - T(u)| \leqslant \|w\|\|v - u\||, \quad \text{for } v, u \in \mathbb{R}^n. \tag{3.3}$$

Hence, applying the Squeeze (or Sandwich) Theorem in single–variable Calculus, we obtain from (3.3) that

$$\lim_{\|v-u\|\to 0} |T(v) - T(u)| = 0,$$

and so T is continuous at u. Since u is any element of $\mathbb{R}^n$, it follows that T is continuous on $\mathbb{R}^n$. □

Example 3.3.3. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = xy, \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

Prove that f is continuous at every $(x_o, y_o) \in \mathbb{R}^2$.

Solution: We want to show that, for every $(x_o, y_o) \in \mathbb{R}^2$,

$$\lim_{\|(x,y)-(x_o,y_o)\|\to 0} |f(x, y) - f(x_o, y_o)| = 0. \tag{3.4}$$

First, write

$$f(x, y) - f(x_o, y_o) = xy - x_o y_o = xy - x_o y + x_o y - x_o y_o,$$

or

$$f(x, y) - f(x_o, y_o) = y(x - x_o) + x_o(y - y_o). \tag{3.5}$$

Taking absolute values on both sides of (3.5) and applying the triangle inequality yields that

$$|f(x, y) - f(x_o, y_o)| \leqslant |y||x - x_o| + |x_o||y - y_o|. \tag{3.6}$$

Restricting to values of (x, y) such that

$$\|(x, y) - (x_o, y_o)\| \leqslant 1, \tag{3.7}$$

we see that

$$|y - y_o| = \sqrt{(y - y_o)^2} \leqslant \sqrt{(x - x_o)^2 + (y - y_o)^2} \leqslant 1;$$

so that, using the triangle inequality again,

$$|y| = |y - y_o + y_o| \leqslant |y - y_o| + |y_o| \leqslant 1 + |y_o|, \tag{3.8}$$

provided that (3.7) holds true. Thus, using the estimate in (3.8) in (3.6), we obtain that, if (x, y) satisfies (3.7),

$$|f(x, y) - f(x_o, y_o)| \leqslant (1 + |y_o|)|x - x_o| + |x_o||y - y_o|. \tag{3.9}$$

Next, apply the Cauchy–Schwarz inequality in (2.8) to the vectors

$$v = \begin{pmatrix} 1 + |y_o| \\ |x_o| \end{pmatrix} \qquad \text{and} \qquad w = \begin{pmatrix} |x - x_o| \\ |y - y_o| \end{pmatrix}$$

to obtain

$$(1 + |y_o|)|x - x_o| + |x_o||y - y_o| \leqslant \sqrt{(1 + |y_o|)^2 + x_o^2}\sqrt{(x - x_o)^2 + (y - y_o)^2}.$$

We can therefore estimate the right–hand side of (3.9) to obtain

$$|f(x, y) - f(x_o, y_o)| \leqslant \sqrt{(1 + |y_o|)^2 + x_o^2}\sqrt{(x - x_o)^2 + (y - y_o)^2},$$

or

$$|f(x, y) - f(x_o, y_o)| \leqslant C_o\|(x, y) - (x_o, y_o)\|,$$

for values of (x, y) within 1 of (x_o, y_o), where $C_o = \sqrt{(1 + |y_o|)^2 + x_o^2}$. We then have that, if $\|(x, y) - (x_o, y_o)\| \leqslant 1$, then

$$0 \leqslant |f(x, y) - f(x_o, y_o)| \leqslant C_o\|(x, y) - (x_o, y_o)\|. \tag{3.10}$$

The claim in (3.4) now follows by applying the Squeeze Theorem to the expressions in (3.10) because the rightmost term in (3.10) goes to 0 as

$$\|(x, y) - (x_o, y_o)\| \to 0.$$

□

Example 3.3.4. Let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 \\ -y \end{pmatrix}, \qquad \text{for all } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Prove that F is continuous at every $\begin{pmatrix} x_o \\ y_o \end{pmatrix} \in \mathbb{R}^2$.

Solution: First, estimate

$$\begin{aligned} \left\|F\begin{pmatrix} x \\ y \end{pmatrix} - F\begin{pmatrix} x_o \\ y_o \end{pmatrix}\right\|^2 &= \left\|\begin{pmatrix} x^2 - x_o^2 \\ -y + y_o \end{pmatrix}\right\|^2 \\ &= (x^2 - x_o^2)^2 + (y - y_o)^2, \end{aligned}$$

which may be written as

$$\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|^2=(x+x_o)^2(x-x_o)^2+(y-y_o)^2, \tag{3.11}$$

after factoring.

Next, restrict to values of $\begin{pmatrix}x\\y\end{pmatrix}\in\mathbb{R}^2$ such that

$$\left\|\begin{pmatrix}x\\y\end{pmatrix}-\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|\leqslant 1. \tag{3.12}$$

It follows from (3.12) that

$$|x-x_o|=\sqrt{(x-x_o)^2}\leqslant\sqrt{(x-x_o)^2+(y-y_o)^2}\leqslant 1.$$

Consequently, if (3.12) holds, then

$$|x|=|x-x_o+x_o|\leqslant|x-x_o|+|x_o|<1+|x_o|, \tag{3.13}$$

where we have used the triangle inequality. It follows from the last inequality in (3.13) that

$$|x+x_o|\leqslant|x|+|x_o|\leqslant 1+2|x_o|, \tag{3.14}$$

where we have, again, used the triangle inequality. Applying the estimate in (3.14) to the equation in (3.11), we obtain

$$\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|^2\leqslant(1+2|x_o|)^2(x-x_o)^2+(y-y_o)^2,$$

which implies that

$$\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|^2\leqslant(1+2|x_o|)^2[(x-x_o)^2+(y-y_o)^2]. \tag{3.15}$$

Taking the positive square root on both sides of the inequality in (3.15) then yields

$$\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|\leqslant(1+2|x_o|)\sqrt{(x-x_o)^2+(y-y_o)^2}. \tag{3.16}$$

From (3.16) we get that, if (3.12) holds, then

$$0\leqslant\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|\leqslant(1+2|x_o|)\left\|\begin{pmatrix}x\\y\end{pmatrix}-\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|. \tag{3.17}$$

Applying the Squeeze Theorem to the inequality in (3.17) we see that, since the rightmost expression in (3.17) goes to 0 as $\left\|\begin{pmatrix}x\\y\end{pmatrix}-\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|$ goes to 0,

$$\lim_{\left\|\begin{pmatrix}x\\y\end{pmatrix}-\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|\to 0}\left\|F\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x_o\\y_o\end{pmatrix}\right\|=0.$$

Hence, F is continuous at $\begin{pmatrix} x_o \\ y_o \end{pmatrix}$. □

The function in Example 3.3.4 is an instance of more general situation addressed by the following proposition.

Proposition 3.3.5. Let U denote an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a vector valued function defined on U and given by

$$F(v) = \begin{pmatrix} f_1(v) \\ f_2(v) \\ \vdots \\ f_m(v) \end{pmatrix}, \quad \text{for all } v \in U,$$

where

$$f_j\colon U \to \mathbb{R}, \quad \text{for } j = 1, 2, \ldots m,$$

are real valued functions defined on U. The vector valued function, F, is continuous at $u \in U$ if and only if each one of its components, f_j, for $j = 1, 2, \ldots m$, is continuous at u.

Proof: F is continuous at $u \in U$ if and only if

$$\lim_{\|v-u\|\to 0} \|F(v) - F(u)\|^2 = 0,$$

if and only if

$$\lim_{\|v-u\|\to 0} \left(\sum_{j=1}^{m} |f_j(v) - f_j(u)|^2 \right) = 0,$$

if and only if

$$\sum_{j=1}^{m} \lim_{\|v-u\|\to 0} |f_j(v) - f_j(u)|^2 = 0,$$

if and only if

$$\lim_{\|v-u\|\to 0} |f_j(v) - f_j(u)|^2 = 0, \quad \text{for all } j = 1, 2, \ldots, m,$$

if and only if

$$\lim_{\|v-u\|\to 0} |f_j(v) - f_j(u)| = 0, \quad \text{for all } j = 1, 2, \ldots, m,$$

if and only if each f_j is continuous at u, for $i = 1, 2, \ldots, m$. □

As an application of Proposition 3.3.5, we obtain the following results for paths.

Example 3.3.6 (Continuous Paths)**.** *Let (a, b) denote the open interval from a to b. A path $\sigma(a, b) \to \mathbb{R}^m$, defined by*

$$\sigma(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{pmatrix}, \quad \textit{for all} \;\; t \in (a, b),$$

where each x_i, for $i = 1, 2, \ldots, m$, denotes a real valued function defined on (a, b), is continuous if and only if each x_i is continuous.

Proof. Let t_o denote an arbitrary element in (a, b). By Proposition 3.3.5, σ is continuous at t_o if and only if each $x_i \colon (a, b) \to \mathbb{R}$ is continuous at t_o. Since, this is true for every $t_o \in (a, b)$, the result follows. □

A particular instance of the previous example is the path in $\mathbb{R}^2$ given by

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

for all t in some interval (a, b) of real numbers. Since the sine and cosine functions are continuous everywhere on $\mathbb{R}$, it follows that the path is continuous.

Next, we use Proposition 3.3.5 to prove that linear functions are continuous.

Example 3.3.7 (Linear Functions are Continuous)**.** Let $F \colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then F is continuous on $\mathbb{R}^n$; that is, F is continuous at every $u \in \mathbb{R}^n$.

Proof: Let $F \colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then, F has the matrix representation

$$F(v) = Av, \quad \text{for all } v \in \mathbb{R}^n, \tag{3.18}$$

with respect to the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, where A is an $m \times n$ matrix,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{3.19}$$

We can therefore write the expression for F in (3.18) as

$$F(v) = \begin{pmatrix} w_1^T v \\ w_2^T v \\ \vdots \\ v_m^T v \end{pmatrix}, \quad \text{for all } v \in \mathbb{R}^n, \tag{3.20}$$

where $w_1^T, w_2^T, \ldots, w_m^T$ are the rows of the matrix A in (3.19). We then have from (3.20) that

$$F(v) = \begin{pmatrix} f_1(v) \\ f_2(v) \\ \vdots \\ f_m(v) \end{pmatrix}, \quad \text{for all } v \in \mathbb{R}^n, \tag{3.21}$$

where

$$f_j(v) = w_j \cdot v, \quad \text{for all } v \in \mathbb{R}^n, \tag{3.22}$$

and $j = 1, 2, \ldots, m$. As shown in Example 3.3.2, each f_j defined in (3.22) is continuous at every $u \in \mathbb{R}^n$. Thus, all the components of the function F as given in (3.21) are continuous. Consequently, using Proposition 3.3.5, we conclude that F is continuous at every $u \in \mathbb{R}^n$. $\square$

The functions in the following examples are continuous because they are linear functions, and linear functions are continuous as shown in Example 3.3.7.

Example 3.3.8. Define $f \colon \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, x_2, \ldots, x_n) = x_i$, for a fixed i in $\{1, 2, \ldots, n\}$. Show that f is continuous on $\mathbb{R}$.
***Solution**:* Observe that f is linear. In fact, note that

$$f(v) = e_i \cdot v, \quad \text{for all } v \in \mathbb{R}^n,$$

where e_i is the i^{th} vector in the standard basis of $\mathbb{R}^n$. It follows from the result of Example 3.3.7 that f is continuous on $\mathbb{R}^n$. $\square$

Example 3.3.9 (Orthogonal Projections are Continuous)**.** Let $\widehat{u}$ denote a unit vector in $\mathbb{R}^n$ and define $P_{\widehat{u}} \colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$P_{\widehat{u}}(v) = (v \cdot \widehat{u})\widehat{u}, \quad \text{for all } v \in \mathbb{R}^n.$$

Prove that $P_{\widehat{u}}$ is continuous on $\mathbb{R}^n$.
***Solution**:* Observe that $P_{\widehat{u}}$ is linear. In fact, for any $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in \mathbb{R}^n$,

$$\begin{aligned} P_{\widehat{u}}(c_1 v_1 + c_2 v_2) &= [(c_1 v_1 + c_2 v_2) \cdot \widehat{u}]\widehat{u} \\ &= (c_1 v_1 \cdot \widehat{u} + c_2 v_2 \cdot \widehat{u})\widehat{u} \\ &= (c_1 v_1 \cdot \widehat{u})\widehat{u} + (c_2 v_2 \cdot \widehat{u})\widehat{u} \\ &= c_1 (v_1 \cdot \widehat{u})\widehat{u} + c_2 (v_2 \cdot \widehat{u})\widehat{u} \\ &= c_1 P_{\widehat{u}}(v_1) + c_2 P_{\widehat{u}}(v_2). \end{aligned}$$

It then follows from the result of Example 3.3.7 that $P_{\widehat{u}}$ is continuous on $\mathbb{R}^n$. $\square$

3.3.1 Images and Pre–Images

Let U denote and open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a map.

Definition 3.3.10. *Given $A \subseteq U$, we define the* image of A under F *to be the set*

$$F(A) = \{y \in \mathbb{R}^m \mid y = F(x), \quad \textit{for some} \;\; x \in A\}.$$

Given $B \subseteq \mathbb{R}^m$, we define the pre–image of B under F *to be the set*

$$F^{-1}(B) = \{x \in U \mid F(x) \in B\}.$$

Example 3.3.11. *Let $\sigma\colon \mathbb{R} \to \mathbb{R}^2$ be given by $\sigma(t) = (\cos t, \sin t)$ for all $t \in \mathbb{R}$. If $A = (0, 2\pi]$, then the image of A under σ is the unit circle around the origin in the xy–plane, or*

$$\sigma((0, 2\pi]) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

Example 3.3.12. *Let σ be as in the previous example, and $A = (0, \pi/2)$. Then,*

$$\sigma(A) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1, 0 < x < 1, 0 < y < 1\}.$$

Example 3.3.13. *Let $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$, the open unit disc in $\mathbb{R}^2$, and $f\colon D \to \mathbb{R}$ be given by*

$$f(x, y) = \sqrt{1 - x^2 - y^2}, \quad \textit{for} \;\; (x, y) \in D$$

Find the pre–image of $B = \{0\}$ under f.
Solution:

$$f^{-1}(0) = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\}.$$

Now, $f(x, y) = 0$ if and only if

$$\sqrt{1 - x^2 - y^2} = 0$$

if and only if

$$x^2 + y^2 = 1.$$

Thus,

$$f^{-1}(0) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\},$$

or the unit circle around the origin in $\mathbb{R}^2$. □

3.3.2 An alternate definition of continuity

In this section we will prove the following proposition

Proposition 3.3.14. *Let U denote an open subset of $\mathbb{R}^n$. A map $F\colon U \to \mathbb{R}^m$ is continuous on U if and only if the pre–image of any open subset of $\mathbb{R}^m$ under F is an open subset of U.*

Proof. Suppose that F is continuous on U. Then, according to Definition 3.3.1, for every $u \in U$,

$$\lim_{\|v-u\|\to 0} \|F(v) - F(u)\| = 0.$$

In other words, $F(v)$ can be made arbitrarily close to $F(u)$ by making v sufficiently close to u.

Let V denote an arbitrary open subset of $\mathbb{R}^m$ and consider

$$F^{-1}(V) = \{u \in U \mid F(u) \in V\}.$$

We show that $F^{-1}(V)$ is open.

To see why this assertion is true, let $u \in F^{-1}(V)$. Then, $F(u) \in V$. Therefore, since V is open, there exists $\varepsilon > 0$ such that

$$B_\varepsilon(F(u)) \subseteq V.$$

This implies that, any $w \in \mathbb{R}^n$ satisfying $\|w - F(u)\| < \varepsilon$ is also an element of V.

Now, by the continuity of F at u, we can make $\|F(v) - F(u)\| < \varepsilon$ bay making $\|v - u\|$ sufficiently small; say, smaller than some $\delta > 0$. It then follows that

$$\|v - u\| < \delta \text{ implies that } \|F(v) - F(u)\| < \varepsilon,$$

which in turn implies that $F(v) \in V$, or $v \in F^{-1}(V)$. We then have that

$$v \in B_\delta(u) \text{ implies that } v \in F^{-1}(V).$$

In other words,

$$B_\delta(u) \subseteq F^{-1}(V).$$

Therefore, $F^{-1}(V)$ is open; thus, the assertion is proved to be true.

Conversely, assume that for any open subset, V, of $\mathbb{R}^m$, $F^{-1}(V)$ is open. We show that this implies that F is continuous at any $u \in U$. To see this, suppose that $u \in U$ and let $\varepsilon > 0$ be arbitrary. Now, since $B_\varepsilon(F(u))$, the open ball of radius ε around $F(u)$, is an open subset of $\mathbb{R}^m$, it follows that

$$F^{-1}(B_\varepsilon(F(u)))$$

is open, by the assumption we are making in this part of the proof. Hence, since $u \in F^{-1}(B_\varepsilon(F(u)))$, there exists $\delta > 0$ such that

$$B_\delta(u) \subseteq F^{-1}(B_\varepsilon(F(u))).$$

This is equivalent to saying that

$$\|v - u\| < \delta \text{ implies that } v \in F^{-1}(B_\varepsilon(F(u))),$$

or

$$\|v - u\| < \delta \text{ implies that } F(v) \in B_\varepsilon(F(u)),$$

or

$$\|v-u\| < \delta \text{ implies that } \|F(v)-F(u)\| < \varepsilon.$$

Thus, given an arbitrary $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|v-u\| < \delta \text{ implies that } \|F(v)-F(u)\| < \varepsilon.$$

This is precisely the definition of

$$\lim_{\|v-u\|\to 0} \|F(v)-F(u)\| = 0.$$

□

3.3.3 Compositions of Continuous Functions

Proposition 3.3.14 provides another definition of continuity: A map is continuous if and only if the pre–image of any open set under the map is open. We will now use this alternate definition prove that a composition of continuous functions is continuous.

Let U be an open subset of $\mathbb{R}^n$ and Q an open subset of $\mathbb{R}^m$. Suppose that we are given two maps, $F\colon U \to \mathbb{R}^m$ and $G\colon Q \to \mathbb{R}^k$. Recall that in order to define the composition of G and F, we must require that the image of U under F is contained in the domain, Q, of G; that is,

$$F(U) \subseteq Q.$$

If this is the case, we define the composition of G and F, denoted $G \circ F$, by

$$G \circ F(v) = G(F(v)), \quad \text{for all } \ v \in U.$$

This yields a map

$$G \circ F\colon U \to \mathbb{R}^k.$$

Proposition 3.3.15. *Let U be an open subset of $\mathbb{R}^n$ and Q an open subset of $\mathbb{R}^m$. Suppose that the maps $F\colon U \to \mathbb{R}^m$ and $G\colon Q \to \mathbb{R}^k$ are continuous on their respective domains and that $F(U) \subseteq Q$. Then, the composition $G\circ F\colon U \to \mathbb{R}^k$ is continuous on U.*

Proof. According to Proposition 3.3.14, it suffices to prove that, for any open set $V \subseteq \mathbb{R}^k$, the pre–image $(G \circ F)^{-1}(V)$ is an open subset of U. Thus, let $V \subseteq \mathbb{R}^k$ be open and observe that

$$\begin{aligned} v \in (G\circ F)^{-1}(V) \quad & \text{iff} \quad (G\circ F)(v) \in V \\ & \text{iff} \quad G(F(v)) \in V \\ & \text{iff} \quad F(v) \in G^{-1}(V) \\ & \text{iff} \quad v \in F^{-1}(G^{-1}(V)), \end{aligned}$$

so that

$$(G\circ F)^{-1}(V) = F^{-1}(G^{-1}(V)).$$

Now, G is continuous; consequently, since V is open, $G^{-1}(V)$ is an open subset of Q by Proposition 3.3.14. Similarly, since F is continuous, it follows again from Proposition 3.3.14 that $F^{-1}(G^{-1}(V))$ is open. Thus, $(G \circ F)^{-1}(V)$ is open. Since, V was an arbitrary open subset of $\mathbb{R}^k$, it follows from Proposition 3.3.14 that $G \circ F$ is continuous on U. □

Example 3.3.16 (Evaluating scalar fields on paths)**.** Let (a, b) denote an open interval of real numbers and $\sigma\colon (a, b) \to \mathbb{R}^n$ be a path. Let U denote an open subset of $\mathbb{R}^n$ and let $f\colon U \to \mathbb{R}$ be a scalar field defined on U. Suppose that $\sigma(t) \in U$ for all $t \in (a, b)$. Then, we can define the composition

$$f \circ \sigma\colon (a, b) \to \mathbb{R}$$

by $f \circ \sigma(t) = f(\sigma(t))$ for all $t \in (a, b)$. Thus, $f \circ \sigma$ is a real valued function of a single variable like those studied in Calculus I and II. An example of a composition $f \circ \sigma$ is provided by evaluating the electrostatic potential, f, along the path of a particle moving according to $\sigma(t)$, where t denotes time.

According to Proposition 3.3.15, if both f and σ are continuous, then so is the function $f \circ \sigma$. Therefore, if $\lim\limits_{t \to t_o} \sigma(t) = u_o$ for some $t_o \in (a, b)$ and $u_o \in U$, then

$$\lim_{t \to t_o} f(\sigma(t)) = f(u_o).$$

The point here is that, if f is continuous at u_o, the limit of f along **any** continuous path that approaches u_o must yield the same value of $f(u_o)$.

3.3.4 Limits and Continuity

Let U denote and open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a vector values function. At end of the proof of Proposition 3.3.14 we remarked that the statement

$$\begin{gathered} \text{for every } \varepsilon > 0, \quad \text{there exists } \delta > 0, \text{ such that} \\ \|v - u\| < \delta \implies \|F(v) - F(u)\| < \varepsilon \end{gathered} \tag{3.23}$$

is equivalent to the statement

$$\lim_{\|v-u\| \to 0} \|F(v) - F(u)\| = 0. \tag{3.24}$$

The statement in (3.24) is usually written as

$$\lim_{v \to u} F(v) = F(u). \tag{3.25}$$

The statements in (3.24) and (3.25) are equivalent. These statements in turn are equivalent to the statement in (3.23).

By replacing $F(u)$ by w, where w denotes a vector in $\mathbb{R}^m$, in the statement in (3.23), we can give meaning to the statement

$$\lim_{v \to u} F(v) = w. \tag{3.26}$$

Thus, we say that the statement in (3.26) is true if and only if

$$\begin{aligned} &\text{for every } \varepsilon > 0, \quad \text{there exists } \delta > 0, \text{ such that} \\ &\|v - u\| < \delta \implies \|F(v) - w\| < \varepsilon. \end{aligned} \tag{3.27}$$

Another way of writing the statement in (3.26) is

$$\lim_{\|v-u\|\to 0} \|F(v) - w\| = 0. \tag{3.28}$$

Any of the assertions in (3.26), (3.27) or (3.28) is equivalent to saying that the function $F\colon U \to \mathbb{R}^m$ has limit $w \in \mathbb{R}^m$ as v approaches u in U.

We note that the a functions F needs not be defined at u for the limit in (3.26) to exist. Here is an example to illustrate this point.

Example 3.3.17. Let $U = \mathbb{R}^2\backslash\{(0,0)\}$; that is,

$$U = \{(x,y) \in \mathbb{R}^2 \mid (x,y) \neq (0,0)\}.$$

Define $f\colon U \to \mathbb{R}$ by

$$f(x,y) = \frac{xy}{\sqrt{x^2+y^2}}, \qquad \text{for } (x,y) \in U. \tag{3.29}$$

In this example we show that

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0. \tag{3.30}$$

To prove the assertion in (3.30) we show that

$$\lim_{\|(x,y)\|\to 0} |f(x,y)| = 0, \tag{3.31}$$

where, according to (3.29),

$$|f(x,y)| = \frac{|xy|}{\sqrt{x^2+y^2}}, \qquad \text{for } (x,y) \neq (0,0). \tag{3.32}$$

Using the estimate

$$|xy| \leqslant \frac{1}{2}(x^2+y^2), \qquad \text{for all } x,y \in \mathbb{R},$$

(see Problem 3 in Assignment #5), we obtain from (3.32) that

$$|f(x,y)| \leqslant \frac{1}{2}\sqrt{x^2+y^2}, \qquad \text{for } (x,y) \neq (0,0),$$

or

$$0 \leqslant |f(x,y)| \leqslant \frac{1}{2}\|(x,y)\|, \qquad \text{for } (x,y) \neq (0,0). \tag{3.33}$$

We can now see that the assertion in (3.31) follows from the estimate in (3.33) and the squeeze theorem.

Remark 3.3.18. In Example 3.3.16 we saw that if a scalar field, f, is continuous at a point $u_o \in U$, where U is an open subset of $\mathbb{R}^n$, then, for any continuous path σ with the property that $\sigma(t) \to u_o$ as $t \to t_o$,

$$\lim_{t \to t_o} f(\sigma(t)) = f(u_o).$$

In other words, taking the limit along any continuous path approaching u_o as $t \to t_o$ must yield one, and only one, value.

In the next examples, we use the observation in Remark 3.3.18 to show that certain limits do not exist.

Example 3.3.19. Let $U = \mathbb{R}^2 \backslash \{(0,0)\}$; that is,

$$U = \{(x,y) \in \mathbb{R}^2 \mid (x,y) \neq (0,0)\}.$$

Define $f \colon U \to \mathbb{R}$ by

$$f(x,y) = \frac{xy}{x^2+y^2}, \qquad \text{for } (x,y) \in U. \tag{3.34}$$

In this example we show that

$$\lim_{(x,y)\to(0,0)} f(x,y) \tag{3.35}$$

does not exist.

We argue by contradiction.

Suppose the limit in (3.35) exists and denote it by L; so that,

$$\lim_{(x,y)\to(0,0)} f(x,y) = L. \tag{3.36}$$

Define the function $g \colon \mathbb{R}^2 \to \mathbb{R}$ as follows:

$$g(x,y) = \begin{cases} f(x,y), & \text{if } (x,y) \neq (0,0); \\ L, & \text{if } (x,y) = (0,0), \end{cases} \tag{3.37}$$

where L is given in (3.36).

It follows from (3.36) and the definition of g in (3.37) that

$$\lim_{(x,y)\to(0,0)} g(x,y) = g(0,0);$$

so that, g is continuous at $(0,0)$. Thus, in view of the observation in Remark 3.3.18, for any continuous path, σ, with the property that $\sigma(t) \to (0,0)$ as $t \to 0$, we would have that

$$\lim_{t\to 0} g(\sigma(t)) = g(0,0) = L, \tag{3.38}$$

since $g \circ \sigma$ would be continuous by Proposition 3.3.15.

Observe that the path $\sigma_1\colon \mathbb{R} \to \mathbb{R}^2$ given by $\sigma_1(t) = (t, 0)$, for $t \in \mathbb{R}$, is continuous and $\sigma_1(t) \to (0,0)$ as $t \to 0$. Also, for $t \neq 0$,

$$\begin{aligned} g(\sigma_1(t)) &= g(t,0) \\ &= f(t,0), \end{aligned}$$

since $t \neq 0$; so that, using the definition of f in (3.34),

$$g(\sigma_1(t)) = 0, \qquad \text{for all } t \neq 0,$$

from which it follows that

$$\lim_{t\to 0} g(\sigma_1(t)) = 0.$$

Consequently, in view of (3.38),

$$L = 0. \tag{3.39}$$

On the other hand, the path $\sigma_2(t) = (t,t)$, for $t \in \mathbb{R}$, is also continuous with $\sigma_2(t) \to (0,0)$ as $t \to 0$. Now, for $t \neq 0$,

$$\begin{aligned} g(\sigma_1(t)) &= g(t,t) \\ &= f(t,t) \\ &= \frac{t^2}{t^2+t^2} \\ &= \frac{1}{2}; \end{aligned}$$

so that,

$$\lim_{t\to 0} g(\sigma_2(t)) = \frac{1}{2}.$$

It then follows from (3.38) that

$$L = \frac{1}{2},$$

which is in direct contradiction with (3.39). This contradiction shows that the assertion in (3.36) cannot be true. Therefore, the limit in (3.35) does not exist.

Remark 3.3.20. The essential part of the argument used in Example 3.3.19 consists of finding two paths, σ_1 and σ_2, such that $\sigma_1(t) \to (x_o, y_o)$ and $\sigma_2(t) \to (x_o, y_o)$ as $t \to t_o$ for some $t_o \in \mathbb{R}$, along which the values $f(\sigma_1(t))$ and $f(\sigma_2(t))$ have different limits as $t \to t_o$. This is enough to conclude that

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

does not exist. We use this reasoning in the following example.

Example 3.3.21. Let $f\colon \mathbb{R}^2\backslash\{(0,0)\} \to \mathbb{R}$ be given by

$$f(x,y) = \frac{|x|}{\sqrt{x^2+y^2}}, \quad \text{for} \ \ (x,y) \neq (0,0).$$

Show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist.

Solution: Let $\sigma_1\colon \mathbb{R} \to \mathbb{R}^2$ be given by $\sigma_1(t) = (0,t)$ for $t \in \mathbb{R}$. Then, $\sigma_1(t) \to (0,0)$ as $t \to 0$ and, for $t \neq 0$,

$$f(\sigma_1(t)) \quad = \quad f(0,t) = 0;$$

so that,

$$\lim_{t\to 0} f(\sigma_1(t)) = 0. \tag{3.40}$$

On the other hand, if $\sigma_2\colon \mathbb{R} \to \mathbb{R}$ is given by $\sigma_2(t) = (t,0)$ for $t \in \mathbb{R}$, then $\sigma_2(t) \to (0,0)$ as $t \to 0$ and, for $t \neq 0$,

$$\begin{aligned} f(\sigma_2(t)) &= f(t,0) \\ &= \frac{|t|}{\sqrt{t^2+0^2}} \\ &= 1; \end{aligned}$$

so that,

$$\lim_{t\to 0} f(\sigma_2(t)) = 1. \tag{3.41}$$

The fact that the limits in (3.40) and (3.41) are different implies that

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

cannot exist, since both $\sigma_1(t)$ and $\sigma_2(t)$ tend to $(0,0)$ as $t \to 0$. $\square$

Chapter 4

Differentiability

In single variable Calculus, a real valued function, $f\colon I \to \mathbb{R}$, defined on an an open interval I, is said to be differentiable at a point $a \in I$ if the limit

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a}$$

exists. If this limit exists, we denote it by $f'(a)$ and call it the *derivative of f at a*. We then have that

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = f'(a).$$

The last expression is equivalent to

$$\lim_{x\to a} \left| \frac{f(x)-f(a)}{x-a} - f'(a) \right| = 0,$$

which we can rewrite as

$$\lim_{x\to a} \frac{|f(x)-f(a)-f'(a)(x-a)|}{|x-a|} = 0. \tag{4.1}$$

Expression (4.1) has the familiar geometric interpretation learned in a first course in Calculus: If f is differentiable at a, then the graph of $y = f(x)$ can be approximated by that of the tangent line,

$$L_a(x) = f(x) + f'(a)(x-a), \quad \text{for all} \ \ x \in \mathbb{R},$$

in the sense that, if

$$E_a(x-a) = f(x) - L_a(x)$$

is the error in the approximation, then

$$\lim_{x\to a} \frac{|E_a(x-a)|}{|x-a|} = 0;$$

that is, the error of the linear approximation to f at a goes to 0 more rapidly than $|x-a|$ goes to 0 as x gets closer to a.

If we are interested in differentiability of f at a variable point $x \in I$, and not a fixed point a, then we can rewrite (4.1) more generally as

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

or

$$\lim_{|y-x| \to 0} \frac{|f(y) - f(x) - f'(x)(y-x)|}{|y-x|} = 0. \tag{4.2}$$

The limit expression in (4.2) is the one we are going to be able to extend to higher dimensions for a vector–valued function $F\colon U \to \mathbb{R}^m$ defined on an open subset, U, of $\mathbb{R}^n$. The symbols x and y will represent vectors in U, and the absolute values will turn into norms. To see how the expression $f'(x)(y-x)$ can be generalized to higher dimensions, let $f'(x) = m_x$, the slope of the tangent line to the graph of f at x, and $y = x + w$; then,

$$f(x+w) - f(x) = m_x w + E_x(w),$$

where

$$\lim_{w \to 0} \frac{|E_a(w)|}{|w|} = 0.$$

Observe that the map

$$w \mapsto m_x w$$

defines a linear map from $\mathbb{R}$ to $\mathbb{R}$. We then conclude that if f is differentiable at x, there exists a linear map such that the linear map approximates the difference $f(x+w) - f(x)$ in the sense that the error in the approximation goes to 0 as $w \to 0$ at a faster rate than $|w|$ approaches 0. This notion of using linear maps to approximate functions locally is the key to extending the concept of differentiability to higher dimensions.

4.1 Definition of Differentiability

Definition 4.1.1 (Differentiability)**.** *Let U denote an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a vector–valued map defined on U. The function F is said to be differentiable at $u \in U$ if and only if there exists a linear transformation $T_u\colon \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{\|v-u\| \to 0} \frac{\|F(v) - F(u) - T_u(v-u)\|}{\|v-u\|} = 0. \tag{4.3}$$

Thus, F is differentiable at $u \in U$ if and only if it can be approximated by a linear function for vectors sufficiently close to u.

Rewrite the expression in (4.3) by putting $v = u+w$, then F is differentiable at $u \in U$ iff there exists a linear transformation $T_u\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\|w\|\to 0} \frac{\|F(u+w) - F(u) - T_u(w)\|}{\|w\|} = 0. \tag{4.4}$$

We can also say that $F\colon U \to \mathbb{R}^m$ is differentiable at $x \in U$ iff there exists a linear transformation $T_u\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(u+w) = F(u) + T_u(w) + E_u(w), \quad \text{for } w \in B_r(u), \tag{4.5}$$

for $r > 0$ sufficiently small, where $E_u(w)$, the *error term,* has the property that

$$\lim_{\|w\|\to 0} \frac{\|E_u(w)\|}{\|w\|} = 0. \tag{4.6}$$

4.2 The Derivative

Proposition 4.2.1 (Uniqueness of the Linear Approximation)**.** *Let U denote an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a map. If F is differentiable at $u \in U$, then the linear transformation, T_u, given in Definition 4.1.1 is unique.*

Proof. Suppose there is another linear transformation, $T\colon \mathbb{R}^n \to \mathbb{R}^m$, given by Definition 4.1.1 in addition to T_u. We show that T and T_u are the same transformation.

From (4.5) and (4.6) we get that

$$F(u+w) = F(u) + T_u(w) + E_u(w), \quad \text{for } w \in B_r(u),$$

for $r > 0$ sufficiently small, where

$$\lim_{\|w\|\to 0} \frac{\|E_u(w)\|}{\|w\|} = 0.$$

Similarly,

$$F(u+w) = F(u) + T(w) + E(w), \quad \text{for } w \in B_r(u),$$

for $r > 0$ sufficiently small, where

$$\lim_{\|w\|\to 0} \frac{\|E(w)\|}{\|w\|} = 0.$$

It then follows that

$$T(w) + E(w) = T_u(w) + E_u(w), \quad \text{for } w \in B_r(u), \tag{4.7}$$

for $r > 0$ sufficiently close to 0.

Let $\widehat{u}$ denote a unit vector and put $w = t\widehat{u}$ in (4.7) for $t \in \mathbb{R}$ sufficiently close to 0. Then, by the linearity of T and T_u,

$$tT(\widehat{u}) + E(t\widehat{u}) = tT_u(\widehat{u}) + E_x(t\widehat{u}), \quad \text{for } |t| < r,$$

where $r > 0$ is small enough.

Dividing by $t \neq 0$ we then get that

$$T(\widehat{u}) + \frac{E(t\widehat{u})}{t} = T_u(\widehat{u}) + \frac{E_u(t\widehat{u})}{t}, \quad \text{for } 0 < |t| < r, \tag{4.8}$$

for $r > 0$ sufficiently small.

Next, observe that

$$\lim_{|t|\to 0} \frac{\|E_u(t\widehat{u})\|}{|t|} = \lim_{\|t\widehat{u}\|\to 0} \frac{\|E_x(t\widehat{u})\|}{\|t\widehat{u}\|} = 0$$

by (4.6). Similarly,

$$\lim_{|t|\to 0} \frac{\|E(t\widehat{u})\|}{|t|} = 0.$$

Thus, letting $t \to 0$ in (4.8) we get that

$$T(\widehat{u}) = T_u(\widehat{u}).$$

Hence T agrees with T_u on any unit vector $\widehat{u}$. Therefore, T and T_u agree on the standard basis $\{e_1, e_2, \ldots, e_n\}$ of $\mathbb{R}^n$. Consequently, since T and T_u are linear

$$T(v) = T_u(v) \quad \text{for all} \quad v \in \mathbb{R}^n;$$

that is, T and T_u are the same linear transformation. □

Proposition 4.2.1 allows as to talk about *the derivative* of $F\colon U \to \mathbb{R}^m$ at a vector $u \in U$.

Definition 4.2.2 (Derivative of a Map)**.** *Let U denote an open subset of $\mathbb{R}^n$ and $F\colon U \to \mathbb{R}^m$ be a map. If F is differentiable at $u \in U$, then the unique linear transformation, T_u, given in Definition 4.1.1 is called the derivative of F at u and is denoted by $DF(u)$. We then have that, if F is differentiable at $u \in U$, there exists a unique linear transformation, $DF(u)\colon \mathbb{R}^n \to \mathbb{R}^m$, such that*

$$F(u + w) = F(u) + DF(u)w + E_u(w), \quad \textit{for } w \in B_r(u),$$

for $r > 0$ sufficiently small, where

$$\lim_{\|w\|\to 0} \frac{\|E_u(w)\|}{\|w\|} = 0.$$

4.3 Example: Differentiable Scalar Fields

Let U denote an open subset of $\mathbb{R}^n$ and let $f\colon U \to \mathbb{R}$ be a scalar field on U. If f is differentiable at $u \in U$, there exists a unique linear map $Df(u)\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$f(u+w) = f(u) + Df(u)w + E_u(w), \quad \text{for } w \in B_r(u), \tag{4.9}$$

where $r > 0$ is sufficiently small, and

$$\lim_{\|w\|\to 0} \frac{|E_u(w)|}{\|w\|} = 0. \tag{4.10}$$

Now, since $Df(u)$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$, there exists a vector v in $\mathbb{R}^n$ with the property that

$$Df(u)w = v \cdot w \quad \text{for all } \ w \in \mathbb{R}^n; \tag{4.11}$$

see the result of Problem 4 in Assignment #3. Thus, according to (4.11), $Df(u)w$ is the dot–product of v an w.

We would like to know what the differentiability of f implies about the components of the vector v in (4.11). Write the vector v as

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

where $a_1, a_2, \ldots . a_n$ are real numbers, or

$$v = a_1\widehat{e}_1 + a_2\widehat{e}_2 + \cdots + a_n\widehat{e}_n, \tag{4.12}$$

where $\{\widehat{e}_1, \widehat{e}_2, \ldots, \widehat{e}_n\}$ is the standard basis in $\mathbb{R}^n$.

Next, apply (4.9) to the case in which $w = t\widehat{e}_j$, where $t \in \mathbb{R}$ is sufficiently close to 0 and $\widehat{e}_j$ is the j^{th} vector in the standard basis for $\mathbb{R}^n$, to get that

$$f(u+t\widehat{e}_j) = f(u) + Df(u)(t\widehat{e}_j) + E_u(t\widehat{e}_j), \quad \text{for } |t| < r, \tag{4.13}$$

where $r > 0$ is sufficiently small.

Using the linearity of $Df(u)$ and (4.11) we get from (4.13) that

$$f(u+t\widehat{e}_j) - f(u) = tv \cdot \widehat{e}_j + E_u(t\widehat{e}_j), \quad \text{for } |t| < r,$$

for $r > 0$ is sufficiently small, where v is the vector in (4.12).

Dividing by $t \neq 0$ we then get that

$$\frac{f(u+t\widehat{e}_j) - f(u)}{t} = a_j + \frac{E_u(t\widehat{e}_j)}{t}, \quad \text{for } 0 < |t| < r, \tag{4.14}$$

for $r > 0$ is sufficiently small.

It follows from (4.10) that

$$\lim_{t\to 0}\frac{|E_u(t\widehat{e}_j)|}{|t|} = \lim_{|t|\to 0}\frac{|E_u(t\widehat{e}_j)|}{\|t\widehat{e}_j\|} = 0,$$

and therefore, we get from (4.14) that

$$\lim_{t\to 0}\frac{f(u+t\widehat{e}_j)-f(u)}{t} = a_j. \tag{4.15}$$

Definition 4.3.1 (Partial Derivatives)**.** Let U be an open subset of $\mathbb{R}^n$,

$$f\colon U\to\mathbb{R}$$

denote a scalar field, and $u\in U$. If

$$\lim_{t\to 0}\frac{f(u+t\widehat{e}_j)-f(u)}{t}$$

exists, we call it *the partial derivative of f at u with respect to x_j* and denote it by $\dfrac{\partial f}{\partial x_j}(u)$.

The argument leading up to equation (4.15) then shows that if the scalar field $f\colon U\to\mathbb{R}$ is differentiable at $u\in U$, then its partial derivatives at u exist and they are the components of the vector v in (4.12) used in the definition of the map $Df(u)\colon\mathbb{R}^n\to\mathbb{R}$ given in (4.11). Thus, in view of (4.12) and (4.15),

$$v = \frac{\partial f}{\partial x_1}(u)\ \widehat{e}_1 + \frac{\partial f}{\partial x_2}(u)\ \widehat{e}_2 + \cdots + \frac{\partial f}{\partial x_n}(u)\ \widehat{e}_2. \tag{4.16}$$

The vector on the right–hand side of (4.16) is called the **gradient** of the function f at u.

Definition 4.3.2 (Gradient)**.** Suppose that the partial derivatives of a scalar field $f\colon U\to\mathbb{R}$ exist at $u\in U$. The expression

$$\frac{\partial f}{\partial x_1}(u)\ \widehat{e}_1 + \frac{\partial f}{\partial x_2}(u)\widehat{e}_2 + \cdots + \frac{\partial f}{\partial x_n}(u)\ \widehat{e}_n$$

is called the **gradient** of f at u. The gradient of f at u is denoted by the symbol $\nabla f(u)$. We then have that

$$\nabla f(u) = \frac{\partial f}{\partial x_1}(u)\ \widehat{e}_1 + \frac{\partial f}{\partial x_2}(u)\ \widehat{e}_2 + \cdots + \frac{\partial f}{\partial x_n}(u)\ \widehat{e}_n.$$

Remark 4.3.3. If the partial derivatives of $f\colon U\to\mathbb{R}$ exist, the gradient is sometimes written as a row vector

$$\nabla f(u) = \left(\frac{\partial f}{\partial x_1}(u), \frac{\partial f}{\partial x_2}(u), \ldots, \frac{\partial f}{\partial x_n}(u)\right).$$

The preceding discussion establishes the following proposition.

Proposition 4.3.4. If $f\colon U \to \mathbb{R}$ is differentiable at $u \in U$, the partial derivatives of f at u exist and the derivative map $Df(u)\colon \mathbb{R}^n \to \mathbb{R}$ is computed according to the formula in (4.11) with $v = \nabla f(u)$; so that,

$$Df(u)w = \nabla f(u) \cdot w, \quad \text{for all} \;\; w \in \mathbb{R}^n; \tag{4.17}$$

that is, $Df(u)w$ is the dot–product of the gradient of f at u with the vector w.

Remark 4.3.5 (Notation)**.** In the case in which $U \subseteq \mathbb{R}^2$ and $f\colon U \to \mathbb{R}$, we will denote the gradient of f at $(x,y) \in \mathbb{R}^2$ by

$$\nabla f(x,y) = \frac{\partial f}{\partial x}(x,y)\, \widehat{\imath} + \frac{\partial f}{\partial y}(x,y)\, \widehat{\jmath},$$

where, according Definition 4.3.1,

$$\frac{\partial f}{\partial x}(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}, \tag{4.18}$$

provided that the limit on the right–hand side of (4.18) exists, and

$$\frac{\partial f}{\partial y}(x,y) = \lim_{k\to 0} \frac{f(x,y+k) - f(x,y)}{k}. \tag{4.19}$$

provided that the limit on the right–hand side of (4.19) exists.

Remark 4.3.6 (Computing partial derivatives)**.** For the case of a real–values function of a single variable, $g\colon I \to \mathbb{R}$, where I is an open interval of real numbers, and $x \in I$, if

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

exists, we call it the derivative of g at x and denote it by $g'(x)$; so that,

$$g'(x) = \lim_{h\to 0} \frac{g(x+h) - g(x)}{h}. \tag{4.20}$$

Comparison of the right–side in (4.18) with the right–hand side in (4.20) suggests that, to compute the partial derivative of f with respect, we may think of the value of y as fixed (constant), and proceed by computing an ordinary derivative with respect to x while holding y constant. The following examples illustrate this procedure.

Example 4.3.7. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = xy, \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

Compute the partial derivatives of f with respect to x and with respect to y.

Solution: Thinking of y as constant, we take the derivative with respect to x to get

$$\frac{\partial f}{\partial x}(x,y) = y, \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

Similarly, thinking of x as fixed and taking the derivative with respect to y, we get

$$\frac{\partial f}{\partial y}(x,y) = x, \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

□

Example 4.3.8. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = e^{-x^2-y^2}, \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

Compute the partial derivatives of f with respect to x and with respect to y.
Solution: Thinking of y as constant, we take the derivative with respect to x to get

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{\partial}{\partial x}\left[e^{-x^2-y^2}\right] \\ &= e^{-x^2-y^2} \cdot \frac{\partial}{\partial x}\left[-x^2-y^2\right], \end{aligned}$$

where we have used the chain rule.

Consequently,

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

for $(x,y) \in \mathbb{R}^2$.

Similarly, thinking of x as fixed and taking the derivative with respect to y, we get

$$\frac{\partial f}{\partial y}(x,y) = -2ye^{-x^2-y^2}, \qquad \text{for all } (x,y) \in \mathbb{R}^2.$$

□

Example 4.3.9. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \sqrt{x^2+y^2}, \qquad \text{for } (x,y) \in \mathbb{R}^2. \tag{4.21}$$

Determine whether or not the partial derivatives of f at $(0,0)$ exist.
Solution: For $h \neq 0$, use the formula for f in (4.21) to compute

$$\begin{aligned} \frac{f(0+h,0)-f(0,0)}{h} &= \frac{f(h,0)}{h} \\ &= \frac{\sqrt{h^2}}{h}; \end{aligned}$$

so that

$$\frac{f(0+h,0)-f(0,0)}{h}=\frac{|h|}{h}, \quad \text{for } h\neq 0. \tag{4.22}$$

Observe that

$$\lim_{h\to 0^+}\frac{|h|}{h}=1,$$

while

$$\lim_{h\to 0^-}\frac{|h|}{h}=-1.$$

thus, in view of (4.22), we conclude that $\lim\limits_{h\to 0}\dfrac{f(0+h,0)-f(0,0)}{h}$ does not exist; so, $\dfrac{\partial f}{\partial x}(0,0)$ does not exist. Similarly, $\dfrac{\partial f}{\partial y}(0,0)$ does not exist. It then follows from Proposition 4.3.4 that the function $f\colon \mathbb{R}^2\to\mathbb{R}$ defined in (4.21) cannot be differentiable at $(0,0)$. □

Example 4.3.10. Let $f\colon \mathbb{R}^2\to\mathbb{R}$ be given by

$$f(x,y)=\begin{cases}\dfrac{xy}{\sqrt{x^2+y^2}}, & \text{if } (x,y)\neq(0,0)\\ 0, & \text{if } (x,y)\neq(0,0).\end{cases} \tag{4.23}$$

Compute the partial derivatives of f and its gradient at $(0,0)$. Is f differentiable at $(0,0)$?

***Solution**:* Use the formula for f in (4.23) to compute

$$\frac{f(0+h,0)-f(0,0)}{h}=\frac{f(h,0)}{h}=0, \quad \text{for } h\neq 0,$$

from which we get that

$$\lim_{h\to 0}\frac{f(0+h,0)-f(0,0)}{h}=0;$$

so that,

$$\frac{\partial f}{\partial x}(0,0)=0.$$

Similarly,

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

It then follows that

$$\nabla f(0,0)=0\,\widehat{i}+0\,\widehat{j}. \tag{4.24}$$

Next, we show that f is not differentiable at $(0,0)$.

Arguing by contradiction, assume that f is differentiable at $(0,0)$. Then, there exists a linear map $Df(0,0)\colon \mathbb{R}^2\to\mathbb{R}$ such that

$$f(w)=f(0,0)+Df(0,0)w+E(w), \quad \text{for } w\in D_r(0,0), \tag{4.25}$$

where $D_r(0,0)$ is the open disc of radius $r > 0$ around the origin in $\mathbb{R}^2$,

$$D_r(0,0) = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < r^2\},$$

for some $r > 0$, and where

$$\lim_{\|w\|\to 0} \frac{|E(w)|}{\|w\|} = 0. \tag{4.26}$$

It follows from Proposition 4.3.4 that

$$Df(0,0)w = \nabla f(0,0) \cdot w, \qquad \text{for all } w \in \mathbb{R}^2;$$

so that, in view of (4.24),

$$Df(0,0)w = 0, \qquad \text{for all } w \in \mathbb{R}^2.$$

Thus, we can rewrite (4.25) as

$$f(w) = E(w), \qquad \text{for } w \in D_r(0,0), \tag{4.27}$$

for some $r > 0$, where we have used the fact that $f(0,0) = 0$, according to the definition of f in (4.23).

Letting $w = (t,t)$, for $t \neq 0$, with $|t|\sqrt{2} < r$, we get from (4.27) and the definition of f in (4.23) that

$$\frac{E(t,t)}{\|(t,t)\|} = \frac{t^2}{t^2 + t^2}, \qquad \text{for } 0 < |t| < \frac{r}{\sqrt{2}},$$

or

$$\frac{E(t,t)}{\|(t,t)\|} = \frac{1}{2}, \qquad \text{for } 0 < |t| < \frac{r}{\sqrt{2}}. \tag{4.28}$$

It follows from (4.28) that

$$\lim_{t\to 0} \frac{|E(t,t)|}{\|(t,t)\|} = \frac{1}{2},$$

which is in direct contradiction with the assertion in (4.26). This contradiction shows that f is not differentiable at $(0,0)$, □

Example 4.3.11. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} e^{-\dfrac{1}{x^2+y^2}} & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) \neq (0,0). \end{cases}$$

Compute the partial derivatives of f and its gradient. Is f differentiable at $(0,0)$?

Solution: According to Definition 4.3.1,

$$\frac{\partial f}{\partial x}(x,y) = \lim_{t\to 0} \frac{f(x+t,y) - f(x,y)}{t}.$$

Thus, we compute the rate of change of f as x changes while y is fixed. For the case in which $(x,y) \neq (0,0)$, we may compute $\partial f/\partial x$ as follows:

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= \frac{\partial}{\partial x}\left(e^{-\dfrac{1}{x^2+y^2}} \right) \\
&= e^{-\dfrac{1}{x^2+y^2}} \cdot \frac{\partial}{\partial x}\left(-\frac{1}{x^2+y^2} \right) \\
&= e^{-\dfrac{1}{x^2+y^2}} \cdot \frac{2x}{(x^2+y^2)^2} \\
&= \frac{2x}{(x^2+y^2)^2} \cdot e^{-\dfrac{1}{x^2+y^2}}.
\end{aligned}$$

That is, we took the one dimensional derivative with respect to x and thought of y as a constant (or fixed with respect to x). Notice that we used the Chain Rule twice in the previous calculation. A similar calculation shows that

$$\frac{\partial f}{\partial x}(x,y) = \frac{2y}{(x^2+y^2)^2} \cdot e^{-\dfrac{1}{x^2+y^2}}$$

for $(x,y) \neq (0,0)$.

To compute the partial derivatives at $(0,0)$, we must compute the limit in Definition 4.3.1. For instance,

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,0) &= \lim_{t\to 0} \frac{f(t,0) - f(0,0)}{t} \\
&= \lim_{t\to 0} \frac{e^{-\dfrac{1}{t^2}}}{t} \\
&= \lim_{t\to 0} \frac{1/t}{e^{1/t^2}}.
\end{aligned}$$

Applying L'Hospital's Rule we then have that

$$\begin{aligned}\frac{\partial f}{\partial x}(0,0) &= \lim_{t\to 0}\frac{1/t^2}{2/t^3 e^{1/t^2}}\\ &= \frac{1}{2}\lim_{t\to 0}\frac{t}{e^{1/t^2}}\\ &= 0.\end{aligned}$$

Similarly, $\dfrac{\partial f}{\partial y}(0,0) = 0$. It then follows that

$$\nabla f(0,0) = (0,0),$$

or the zero vector, and, for $(x,y)\neq(0,0)$,

$$\nabla f(x,y) = \frac{2e^{-\frac{1}{x^2+y^2}}}{(x^2+y^2)^2}(x,y),$$

or

$$\nabla f(x,y) = \frac{2e^{-\frac{1}{x^2+y^2}}}{(x^2+y^2)^2}(x\,\widehat{i} + y\,\widehat{j}).$$

To show that f is differentiable at $(0,0)$, we show that

$$f(x,y) = f(0,0) + T(x,y) + E(x,y),$$

where

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

and T is the zero linear transformation from $\mathbb{R}^2$ to $\mathbb{R}$.

In this case

$$E(x,y) = e^{-\frac{1}{x^2+y^2}} \quad \text{if } (x,y)\neq(0,0).$$

Thus, for $(x,y)\neq(0,0)$,

$$\frac{|E(x,y)|}{\sqrt{x^2+y^2}} = \frac{e^{-\frac{1}{x^2+y^2}}}{\sqrt{x^2+y^2}} = \frac{e^{-\frac{1}{u^2}}}{u},$$

where we have set $u = \sqrt{x^2+y^2}$. Thus,

$$\lim_{(x,y)\to(0,0)}\frac{|E(x,y)|}{\sqrt{x^2+y^2}} = \lim_{u\to 0}\frac{e^{-\frac{1}{u^2}}}{u} = 0,$$

by the same calculation involving L'Hospital's Rule that was used to compute $\partial f/\partial x$ at $(0,0)$. Consequently, f is differentiable at $(0,0)$ and its derivative is the zero map. □

We have seen that if a scalar field $f\colon U \to \mathbb{R}$ is differentiable at $v \in U$, then

$$f(v+w) = f(v) + \nabla f(v)\cdot w + E_v(w), \qquad \text{for } w \in B_r(v),$$

where $r > 0$ with sufficiently small, $\nabla f(v)$ is the gradient of f at $v \in U$, and

$$\lim_{\|w\|\to 0} \frac{|E_v(w)|}{\|w\|} = 0.$$

Applying this to the case where $w = t\widehat{u}$, for a unit vector $\widehat{u}$, and $|t| < r$, we get that

$$f(v+t\widehat{u}) - f(v) = t\nabla f(v)\cdot \widehat{u} + E_v(t\widehat{u})$$

for $t \in \mathbb{R}$ sufficiently close to 0. Dividing by $t \neq 0$ and letting $t \to 0$ leads to

$$\lim_{t\to 0} \frac{f(v+t\widehat{u}) - f(v)}{t} = \nabla f(v)\cdot \widehat{u},$$

where we have used (4.10).

Definition 4.3.12 (Directional Derivatives)**.** *Let $f\colon U \to \mathbb{R}$ denote a scalar field defined on an open subset U of $\mathbb{R}^n$, and let $\widehat{u}$ be a unit vector in $\mathbb{R}^n$. If the limit*

$$\lim_{t\to 0} \frac{f(v+t\widehat{u}) - f(v)}{t}$$

exists, we call it the directional derivative of f at v in the direction of the unit vector $\widehat{u}$. *We denote it by $D_{\widehat{u}}f(v)$.*

We have then shown that if the scalar field $f\colon U \to \mathbb{R}$ is differentiable at $v \in U$, then its directional derivative at v in the direction of a unit vector $\widehat{u}$ is given by

$$D_{\widehat{u}}f(v) = \nabla f(v)\cdot \widehat{u};$$

that is, the dot–product of the gradient of f at v with the unit vector $\widehat{u}$. In other words, the directional derivative on f at $v \in U$ in the direction of a unit vector $\widehat{u}$ is the orthogonal projection of $\nabla f(v)$ along the direction of $\widehat{u}$.

4.4 Example: Differentiable Paths

Example 4.4.1. Let I denote an open interval in $\mathbb{R}$, and suppose that the path $\sigma\colon I \to \mathbb{R}^n$ is differentiable at $t \in I$. It then follows that there exists a linear map $D\sigma(t)\colon \mathbb{R} \to \mathbb{R}^n$ such that

$$\sigma(t+h) = \sigma(t) + D\sigma(t)(h) + E_t(h), \qquad \text{for } |h| < r, \tag{4.29}$$

where $r > 0$ is sufficiently small, and

$$\lim_{h\to 0} \frac{\|E_t(h)\|}{|h|} = 0. \tag{4.30}$$

(a) Show that the linear map $D\sigma(t)\colon \mathbb{R} \to \mathbb{R}^n$ is of the form

$$D\sigma(t)(h) = hv(t) \quad \text{for all} \quad h \in \mathbb{R},$$

where the vector $v(t)$ is obtained from

$$v(t) = D\sigma(t)(1);$$

that is, $v(t)$ is the image of the real number 1 under the linear transformation $D\sigma(t)$.

Solution: Let h denote any real number; then, by the linearity of $D\sigma(t)$,

$$D\sigma(t)(h) = D\sigma(t)(h \cdot 1) = hD\sigma(t)(1) = hv(t).$$

$\square$

(b) Write

$$\sigma(t) = x_1(t)\ \widehat{e}_1 + x_2(t)\ \widehat{e}_2 + \cdots + x_n(t)\ \widehat{e}_n, \qquad \text{for all } t \in I, \tag{4.31}$$

where $\{\widehat{e}_1, \widehat{e}_2, \ldots, \widehat{e}_n\}$ is the standard basis in $\mathbb{R}^n$.

Show that if $\sigma\colon I \to \mathbb{R}^n$ is differentiable at $t \in I$ and $v(t) = D\sigma(t)(1)$, then each function $x_j\colon I \to \mathbb{R}$, for $j = 1, 2, \ldots, n$, in the definition of σ in (4.31), is differentiable at t, and

$$x_j'(t) = v_j(t), \qquad \text{for all } t \in I, \tag{4.32}$$

where $v_1, v_2, \ldots, v_n$ are the components of the vector $v(t)$; that is,

$$v(t) = v_1(t)\ \widehat{e}_1 + v_2(t)\ \widehat{e}_2 + \cdots + v_n(t)\ \widehat{e}_n, \qquad \text{for all } t \in I. \tag{4.33}$$

Solution: Writing σ and $v(t)$ as a column vector, equation (4.29) takes the form

$$\begin{pmatrix} x_1(t+h) \\ x_2(t+h) \\ \vdots \\ x_n(t+h) \end{pmatrix} - \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = h \begin{pmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_n(t) \end{pmatrix} + E_t(h), \qquad \text{for } |h| < r,$$

or, after division by $h \neq 0$,

$$\begin{pmatrix} \dfrac{x_1(t+h) - x_1(t)}{h} \\ \dfrac{x_2(t+h) - x_2(t)}{h} \\ \vdots \\ \dfrac{x_n(t+h) - x_n(t)}{h} \end{pmatrix} = \begin{pmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_n(t) \end{pmatrix} + \frac{E_t(h)}{h}, \qquad \text{for } 0 < |h| < r. \tag{4.34}$$

Now, it follows from (4.30) that

$$\lim_{h\to 0}\frac{E_t(h)}{h}=\mathbf{0}, \tag{4.35}$$

where $\mathbf{0}$ is the zero vector in $\mathbb{R}^n$. Consequently, in view of (4.35), we obtain from (4.34) that

$$\lim_{h\to 0}\frac{x_j(t+h)-x_j(t)}{h}=v_j(t) \quad \text{for each} \quad j=1,2,\ldots n,$$

which shows that each $x_j\colon I\to\mathbb{R}$ in (4.31) is differentiable at t with

$$x_j'(t)=v_j(t), \qquad \text{for all } t\in I,$$

for each $j=1,2,\ldots,n$, which is the assertion in (4.32). □

Notation: If $\sigma\colon I\to\mathbb{R}^n$ is differentiable at every $t\in I$, the vector valued function $v\colon I\to\mathbb{R}^n$ given by $v(t)=D\sigma(t)(1)$ is called the *velocity* of the path σ, and is usually denoted by $\sigma'(t)$. We then have that

$$D\sigma(t)(h)=h\sigma'(t) \quad \text{for all} \quad h\in\mathbb{R}$$

and all t at which the path σ is differentiable. We can then rewrite (4.29) as

$$\sigma(t+h)=\sigma(t)+h\sigma'(t)+E_t(h), \qquad \text{for } |h|<r, \tag{4.36}$$

for some $r>0$, where

$$\lim_{h\to 0}\frac{\|E_t(h)\|}{|h|}=0. \tag{4.37}$$

Rewriting the expression in (4.36) once more, by replacing t by t_o and $t+h$ by t, we have that

$$\sigma(t)=\sigma(t_o)+(t-t_o)\sigma'(t_o)+E_{t_o}(t-t_o), \qquad \text{for } |t-t_o|<r, \tag{4.38}$$

for some $r>0$, where, according to (4.37),

$$\lim_{t\to t_o}\frac{\|E_{t_o}(t-t_o)\|}{|t-t_o|}=0. \tag{4.39}$$

The expression

$$\ell(t)=\sigma(t_o)+(t-t_o)\sigma'(t_o), \qquad \text{for } t\in\mathbb{R}, \tag{4.40}$$

on the right–hand side of (4.38) gives the vector–parametric equation of a straight line through $\sigma(t_o)$ in the direction of the velocity vector, $\sigma'(t_o)$, of the path $\sigma(t)$ at the t_o. Thus, (4.38) and (4.39) yield the following interpretation of differentiability of a path $\sigma(t)$ at t_o:

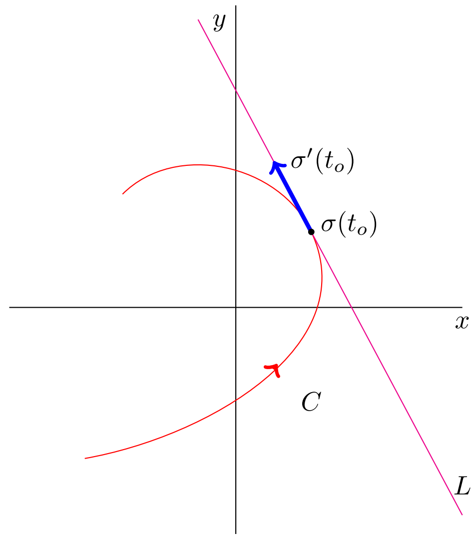

Figure 4.4.1: Sketch of graphs of σ and its tangent line at $\sigma(t_o)$

> If a path $\sigma\colon I \to \mathbb{R}^n$ is differentiable at the t_o, then it can be approximated by a straight line through $\sigma(t_o)$ in the direction of the velocity vector $\sigma'(t_o)$.

Figure 4.4.1 shows this interpretation geometrically in two dimensions. In this case the graph of σ is a curve C in the plane and the graph of ℓ in (4.40) is a line L that is called the tangent line approximation to $\sigma(t)$ for t near t_o.

Definition 4.4.2 (Tangent line to a path)**.** The straight line given parimetrically by the vector equation in (4.40) is called the *the tangent line* to the path $\sigma(t)$ at the point $\sigma(t_o)$.

Example 4.4.3. *Give the tangent line to the path*

$$\sigma(t) = (\cos t, t, \sin t) \quad \textit{for} \quad t \in \mathbb{R}$$

when $t_o = \pi/4$.

Solution: The equation of the tangent line is given by

$$r(t) = \sigma(t_o) + (t - t_o)\sigma'(t_o),$$

where $\sigma'(t) = (-\sin t, 1, \cos t)$; so that, for $t_o = \pi/4$, we get that

$$r(t) = \left(\frac{\sqrt{2}}{2}, \quad \frac{\pi}{4}, \quad \frac{\sqrt{2}}{2}\right) + \left(t - \frac{\pi}{4}\right)\left(-\frac{\sqrt{2}}{2}, \quad 1, \quad \frac{\sqrt{2}}{2}\right)$$

for $t \in \mathbb{R}$.

Writing (x, y, z) for the vector $r(t)$, we obtain the parametric equations for the tangent line:

$$\begin{cases} x = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}\left(t - \frac{\pi}{4}\right) \\ y = \frac{\pi}{4} + t \\ z = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\left(t - \frac{\pi}{4}\right) \end{cases}$$

□

4.5 Sufficient Condition for Differentiability

4.5.1 Differentiability of Paths

Let I be an open interval of real numbers and $\sigma\colon I \to \mathbb{R}^n$ denote a path in $\mathbb{R}^n$. Write

$$\sigma(t) = x_1(t)\,\widehat{e}_1 + x_2(t)\,\widehat{e}_2 + \cdots + +x_n(t)\,\widehat{e}_n, \quad \text{for all } t \in I,$$

and suppose that the functions $x_1, x_2, \ldots, x_n$ are all differentiable functions of t in I. We show that the path σ is differentiable according to Definition 4.1.1.

Let $t \in I$ and $h \in \mathbb{R}$ be such that $t + h \in I$. Since each $x_i\colon I \to \mathbb{R}$ is differentiable at t, we can write

$$x_j(t+h) = x_i(t) + x_j'(t)h + E_j(t,h), \quad \text{for all } j = 1, 2, \ldots n. \tag{4.41}$$

where

$$\lim_{h\to 0} \frac{|E_j(t,h)|}{|h|} = 0, \quad \text{for all } j = 1, 2, \ldots n. \tag{4.42}$$

It follows from (4.41) that

$$x_j(t+h) - x_j(t) - hx_j'(t) = E_j(t,h) \quad \text{for} \quad j = 1, 2, \ldots, n. \tag{4.43}$$

Putting

$$\sigma'(t) = \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{pmatrix}, \quad \text{for } t \in I, \tag{4.44}$$

we obtain from the equations in (4.43) that

$$\begin{aligned} \sigma(t+h) - \sigma(t) - h\sigma'(t) &= \begin{pmatrix} x_1(t+h) - x_1(t) - hx_1'(t) \\ x_2(t+h) - x_2(t) - hx_2'(t) \\ \vdots \\ x_n(t+h) - x_n(t) - hx_n'(t) \end{pmatrix} \\ &= \begin{pmatrix} E_1(t,h) \\ E_2(t,h) \\ \vdots \\ E_n(t,h) \end{pmatrix}, \end{aligned}$$

where $E_1(t,h), E_2(t,h), \ldots, E_n(t,h)$ are given in (4.41) and satisfy (4.42). It then follows that, for $h \neq 0$ and $|h|$ small enough,

$$\frac{1}{h}(\sigma(t+h) - \sigma(t) - h\sigma'(t)) = \begin{pmatrix} E_1(t,h)/h \\ E_2(t,h)/h \\ \vdots \\ E_n(t,h)/h \end{pmatrix}.$$

Taking the square of the norm on both sides we get that

$$\frac{\|\sigma(t+h) - \sigma(t) - h\sigma'(t)\|^2}{|h|^2} = \sum_{j=1}^{n} \left| \frac{E_j(t,h)}{h} \right|^2.$$

Hence, by virtue of (4.42),

$$\lim_{h\to 0} \frac{\|\sigma(t+h) - \sigma(t) - h\sigma'(t)\|}{|h|} = 0,$$

which shows that σ is differentiable at t. Furthermore, $D\sigma(t)\colon \mathbb{R} \to \mathbb{R}^n$ is given by

$$D\sigma(t)h = h\sigma'(t), \qquad \text{for all } h \in \mathbb{R},$$

where $\sigma'(t)$ is given in (4.44).

4.5.2 Differentiability of Scalar Fields

Let U denote an open subset of $\mathbb{R}^n$ and $f\colon U \to \mathbb{R}$ be a scalar field defined on U. Suppose also that the partial derivatives of f,

$$\frac{\partial f}{\partial x_1}(u),\ \frac{\partial f}{\partial x_2}(u),\ \ldots,\ \frac{\partial f}{\partial x_n}(u),$$

exist for all $u \in U$. We show in this section that, if the partial derivatives of f are continuous on U, then the scalar field f is differentiable according to Definition 4.1.1.

Observe that ∇f defines a map from U to $\mathbb{R}^n$ by

$$\nabla f(u) = \frac{\partial f}{\partial x_1}(u)\ \widehat{e}_1 + \frac{\partial f}{\partial x_2}(u)\ \widehat{e}_2 + \cdots + \frac{\partial f}{\partial x_n}(u)\ \widehat{e}_n, \quad \text{for all } \ u \in U.$$

Note that, if the partial derivatives of f are continuous on U, then the vector field

$$\nabla f\colon U \to \mathbb{R}^n$$

is a continuous map.

Proposition 4.5.1. Let U denote an open subset of $\mathbb{R}^n$ and $f\colon U \to \mathbb{R}$ be a scalar field defined on U. Suppose that the partial derivatives of f are continuous on U. Then the scalar field f is differentiable.

Proof: We present the proof here for the case $n = 2$. In this case we may write

$$\nabla f(x,y) = \begin{pmatrix} \dfrac{\partial f}{\partial x}(x,y) \\ \\ \dfrac{\partial f}{\partial y}(x,y) \end{pmatrix},$$

where we are assuming that the functions $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ are continuous on U.

Let $(x,y) \in U$; then, since U is open, there exists $r > 0$ such that $B_r(x,y) \subseteq U$. It then follows that, for $(h,k) \in B_r(0,0)$, $(x+h, y+k) \in U$. For $(h,k) \in B_r(0,0)$ we define

$$E(h,k) = f(x+h, y+k) - f(x,y) - \nabla f(x,y) \cdot (h,k). \tag{4.45}$$

We prove that

$$\lim_{(h,k)\to(0,0)} \frac{|E(h,k)|}{\sqrt{h^2+k^2}} = 0 \tag{4.46}$$

Assume that $h > 0$ and $k > 0$ (the other cases can be treated in an analogous manner). By the mean value theorem, there are real numbers θ and η such that $0 < \theta < 1$ and $0 < \eta < 1$ and

$$f(x+h, y+k) - f(x, y+k) = \frac{\partial f}{\partial x}(x+\theta h, y+k) \cdot h,$$

and

$$f(x, y+k) - f(x,y) = \frac{\partial f}{\partial y}(x, y+\eta k) \cdot k.$$

Consequently,

$$f(x+h, y+k) - f(x,y) = \frac{\partial f}{\partial x}(x+\theta h, y+k) \cdot h + \frac{\partial f}{\partial y}(x, y+\eta k) \cdot k.$$

Thus, in view of (4.45), we see that

$$E(h,k) = \left(\frac{\partial f}{\partial x}(x+\theta h, y+k) - \frac{\partial f}{\partial x}(x,y)\right) h + \left(\frac{\partial f}{\partial y}(x, y+\eta k) - \frac{\partial f}{\partial x}(x,y)\right) k.$$

Thus, $E(h,k)$ is the dot product of the vector $v(h,k)$, given by

$$v(h,k) = \left(\frac{\partial f}{\partial x}(x+\theta h, y+k) - \frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x, y+\eta k) - \frac{\partial f}{\partial x}(x,y)\right),$$

and the vector (h,k). Consequently, by the Cauchy–Schwarz inequality,

$$|E(h,k)| \leqslant \|v(h,k)\| \|(h,k)\|.$$

Dividing by $\|(h,k)\|$ for $(h,k) \neq (0,0)$ we then get

$$\frac{|E(h,k)|}{\sqrt{h^2+k^2}} \leqslant \|v(h,k)\|, \tag{4.47}$$

where

$$\|v(h,k)\| = \sqrt{\left(\frac{\partial f}{\partial x}(x+\theta h, y+k) - \frac{\partial f}{\partial x}(x,y)\right)^2 + \left(\frac{\partial f}{\partial y}(x, y+\eta k) - \frac{\partial f}{\partial x}(x,y)\right)^2}$$

tends to 0 as $(h,k) \to (0,0)$ since the partial derivatives of f are continuous on U. It then follows from the estimate in (4.47) and the squeeze theorem that

$$\lim_{(h,k)\to(0,0)} \frac{|E(h,k)|}{\sqrt{h^2+k^2}} = 0,$$

which is (4.46). This shows that f is differentiable at (x,y). Since (x,y) was arbitrary, the result follows. $\square$

4.5.3 C^1 Maps and Differentiability

Definition 4.5.2 (C^1 Maps)**.** Let U denote an open subset of $\mathbb{R}^n$. The vector valued map

$$F(u) = \begin{pmatrix} f_1(u) \\ f_2(u) \\ \vdots \\ f_m(u) \end{pmatrix} \quad \text{for all} \quad u \in U,$$

where $f_i \colon U \to \mathbb{R}$ are scalar fields on U, is said to be of class C^1, or a C^1 map, if the partial derivatives

$$\frac{\partial f_i}{\partial x_j} \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n,$$

are continuous on U.

Proposition 4.5.1 then says that a C^1 scalar field must be differentiable. Thus, being a C^1 scalar field is sufficient for the map being differentiable. However, it is not necessary. For example, the function

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\dfrac{1}{x^2+y^2}\right), & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0) \end{cases}$$

is differentiable at $(0,0)$; however, the partial derivatives are not continuous at the origin (This is shown in Problem 5 of Assignment #9).

The result of Proposition 4.5.1 applies more generally to C^1 vector–valued maps:

Proposition 4.5.3 (C^1 implies Differentiability)**.** Let U denote an open subset of $\mathbb{R}^n$ and $F \colon U \to \mathbb{R}^m$ be a vector field on U defined by

$$F(u) = \begin{pmatrix} f_1(u) \\ f_2(u) \\ \vdots \\ f_m(u) \end{pmatrix} \quad \text{for all} \quad u \in U,$$

where the scalar fields $f_i\colon U \to \mathbb{R}$ are of class C^1 in U, for $i = 1, 2, \ldots, m$. Then, the vector–valued F is differentiable in U and the matrix representation of the linear transformation

$$DF(u)\colon \mathbb{R}^n \to \mathbb{R}^m$$

is given by

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(u) & \dfrac{\partial f_1}{\partial x_2}(u) & \cdots & \dfrac{\partial f_1}{\partial x_n}(u) \\ \dfrac{\partial f_2}{\partial x_1}(u) & \dfrac{\partial f_2}{\partial x_2}(u) & \cdots & \dfrac{\partial f_2}{\partial x_n}(u) \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(u) & \dfrac{\partial f_m}{\partial x_2}(u) & \cdots & \dfrac{\partial f_m}{\partial x_n}(u) \end{pmatrix}, \quad \text{for } u \in U. \tag{4.48}$$

The matrix of partial derivative of the components of F in equation (4.48) is called the *Jacobian matrix* of the map F at u. It is the matrix that represents the derivative map $DF(u)\colon \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$. We will therefore denote it by $DF(u)$. Hence, $DF(u)w$ can be understood as matrix multiplication of the Jacobian matrix of F at u by the column vector w.

If $m = n$, the determinant of the square matrix $DF(u)$ is called the *Jacobian determinant* of F at u, and is denoted by the symbols $J_F(u)$ or $\dfrac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_n)}$. We then have that

$$J_F(u) = \frac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_n)} = \det DF(u).$$

Example 4.5.4. Let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the map

$$F(x, y) = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} \quad \text{for all} \quad (x, y) \in \mathbb{R}^2.$$

Then, the Jacobian matrix of F is

$$DF(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \quad \text{for all} \quad (x, y) \in \mathbb{R}^2,$$

and the Jacobian determinant is

$$J_F(x, y) = 4(x^2 + y^2).$$

If we let $u = x^2 - y^2$ and $v = 2xy$, we can write the Jacobian determinant as $\dfrac{\partial(u, v)}{\partial(x, y)}$.

4.6 Derivatives of Compositions

The goal of this section is to prove that compositions of differentiable functions are differentiable:

Theorem 4.6.1 (The Chain Rule). *Let U denote an open subset of $\mathbb{R}^n$ and Q and open subset of $\mathbb{R}^m$, and let $F\colon U \to \mathbb{R}^m$ and $G\colon Q \to \mathbb{R}^k$ be maps. Suppose that $F(U) \subseteq Q$. If F is differentiable at $x \in U$ and G is differentiable at $y = F(x) \in Q$, then the composition*

$$G \circ F\colon U \to \mathbb{R}^k$$

is differentiable at x and the derivative map $D(G\circ F)(x)\colon \mathbb{R}^n \to \mathbb{R}^k$ is given by

$$D(G\circ F)(x)w = DG(y)DF(x)w \quad \text{for all} \quad w \in \mathbb{R}^n.$$

Proof. Since F is differentiable at $x \in U$, for $w \in \mathbb{R}^n$ with $\|w\|$ sufficiently small,

$$F(x+w) = F(x) + DF(x)w + E_{F}(w), \tag{4.49}$$

where

$$\lim_{\|w\|\to 0} \frac{\|E_{F}(w)\|}{\|w\|} = 0. \tag{4.50}$$

Similarly, for $v \in \mathbb{R}^m$ with $\|v\|$ sufficiently small,

$$G(y+v) = G(y) + DG(y)v + E_{G}(v), \tag{4.51}$$

where

$$\lim_{\|v\|\to 0} \frac{\|E_{G}(v)\|}{\|v\|} = 0. \tag{4.52}$$

It then follows from (4.49) that, for $w \in \mathbb{R}^n$ with $\|w\|$ sufficiently small,

$$\begin{aligned} (G\circ F)(x+w) &= G(F(x+w)) \\ &= G(F(x) + DF(x)w + E_{F}(w)) \\ &= G(F(x) + v), \end{aligned} \tag{4.53}$$

where we have set

$$v = DF(x)w + E_{F}(w). \tag{4.54}$$

Observe that, by the triangle inequality and the Cauchy–Schwarz inequality,

$$\|v\| \leqslant \|DF(x)\|\|w\| + \|E_{F}(w)\|, \tag{4.55}$$

where

$$\|DF(x)\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}(x)\right)^2};$$

so that, by virtue of (4.50), we can make $\|v\|$ small by making $\|w\|$ small. It then follows from (4.51) and (4.53) that

$$(G\circ F)(x+w) = G(F(x)) + DG(F(x))v + E_{G}(v),$$

where v as given in (4.54) can be made sufficiently small in norm by making $\|w\|$ sufficiently small. It then follows that, for $\|w\|$ sufficiently small,

$$(G\circ F)(x+w) = (G\circ F)(x) + DG(y)DF(x)w + DG(y)E_{_F}(w) + E_{_G}(v). \quad (4.56)$$

Put

$$E(w) = DG(y)E_{_F}(w) + E_{_G}(v) \quad (4.57)$$

for $w \in \mathbb{R}^n$ and v as given in (4.54). The differentiability of $G\circ F$ at x will then follow from (4.56) if we can prove that

$$\lim_{\|w\|\to 0} \frac{\|E(w)\|}{\|w\|} = 0. \quad (4.58)$$

This will also prove that

$$D(G\circ F)(x)w = DG(y)DF(x)w \quad \text{for all} \quad w \in \mathbb{R}^n.$$

To prove (4.58), take the norm of $E(w)$ defined in (4.57), apply the triangle and Cauchy–Schwarz inequalities, and divide by $\|w\|$ to get that

$$\frac{\|E(w)\|}{\|w\|} \quad \leqslant \quad \|DG(y)\|\frac{\|E_{_F}(w)\|}{\|w\|} + \frac{\|E_{_G}(v)\|}{\|v\|}\frac{\|v\|}{\|w\|}, \quad (4.59)$$

where, by virtue of the inequality in (4.55),

$$\frac{\|v\|}{\|w\|} \leqslant \|DF(x)\| + \frac{\|E_{_F}(w)\|}{\|w\|}.$$

The proof of (4.58) will then follow from this last estimate, (4.50), (4.52), (4.59) and the Squeeze Theorem. This completes the proof of the Chain Rule. □

Example 4.6.2. *Let U be an open subset of the xy–plane, $\mathbb{R}^2$, and $f\colon U \to \mathbb{R}$ be a differentiable scalar field. Let Q be an open subset of the uv–plane, $\mathbb{R}^2$, and $\Phi\colon Q \to \mathbb{R}^2$ be a differentiable map such that $\Phi(Q) \subseteq U$. Then, by the Chain Rule, the map*

$$f\circ\Phi\colon Q \to \mathbb{R}$$

is differentiable. Furthermore, putting

$$g(u,v) = (f\circ\Phi)(u,v),$$

where

$$\Phi(u,v) = \begin{pmatrix} x(u,v)\\ y(u,v)\end{pmatrix}, \quad \textit{for} \quad (u,v)\in Q,$$

we have that

$$Dg(u,v) = Df(x(u,v),y(u,v))D\Phi(u,v).$$

Writing this in terms of Jacobian matrices we get

$$\begin{pmatrix} \dfrac{\partial g}{\partial u} & \dfrac{\partial g}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix},$$

from which we get that

$$\frac{\partial g}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}$$

and

$$\frac{\partial g}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.$$

In the previous example, if $\Phi\colon Q \to \mathbb{R}^2$ is a one–to–one map, then Φ is called a *change of variable map.* Writing Φ in terms of a its components we have

$$\begin{aligned} x &= x(u,v) \\ y &= y(u,v), \end{aligned}$$

we see that Φ changes from uv–coordinates to xy–coordinates. As a more concrete example, consider the change to *polar coordinates* maps

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta, \end{aligned}$$

where $0 \leqslant r < \infty$ and $-\pi < \theta \leqslant \pi$. We then have that

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$

and

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}$$

give the partial derivatives of f with respect to the polar variables r and θ in terms of the partial derivatives of f with respect to the Cartesian coordinates x and y and the derivative of the change of variables map

$$\Phi(r,\theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Example 4.6.3. *Let U denote an open subset of $\mathbb{R}^n$ and I an open interval of real numbers. Suppose that $f\colon U \to \mathbb{R}$ is a scalar differentiable field and $\sigma\colon I \to \mathbb{R}^n$ is a differentiable path with $\sigma(I) \subseteq U$. Then, by the Chain Rule, $f(\sigma(t))$ is differentiable for all $t \in I$, and*

$$\frac{d}{dt} f(\sigma(t)) = \nabla f(\sigma(t)) \cdot \sigma'(t) \quad \text{for all} \ \ t \in I.$$

Example 4.6.4 (Tangent plane to a sphere). *Let* $f\colon \mathbb{R}^3 \to \mathbb{R}$ *be given by*

$$f(x, y, x) = x^2 + y^2 + z^2 \quad \text{for all } (x, y, z) \in \mathbb{R}^3.$$

Define the set

$$S = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = 1\}.$$

Then, S is the sphere of radius 1 *around the origin in* $\mathbb{R}^3$*, or the* unit sphere *in* $\mathbb{R}^3$*.*

Let $\sigma\colon I \to \mathbb{R}^3$ *denote a* C^1 *maps that lies entirely on the unit sphere; that is,*

$$f(\sigma(t)) = 1 \quad \text{for all } \ t \in I.$$

Then, differentiating with respect to t on both sides,

$$\frac{d}{dt} f(\sigma(t)) = 0 \quad \text{for all } \ t \in I,$$

and applying the Chain Rule, we obtain that

$$\nabla f(\sigma(t)) \cdot \sigma'(t) = 0 \quad \text{for all } \ t \in I.$$

Thus, the gradient of f is perpendicular to the tangent to the path σ.

For a fixed point, (x_o, y_o, z_o)*, on the sphere S, consider the collection of all* C^1 *paths,* $\sigma\colon I \to \mathbb{R}^3$ *on the sphere, such that* $\sigma(t_o) = (x_o, y_o, z_o)$ *for a fixed* $t_o \in I$*. What we have just derived shows that the tangent vectors to the path at* (x_o, y_o, z_o) *all lie on a plane perpendicular to* $\nabla f(x_o, y_o, z_o)$*. This plane is called the* tangent plane *to S at* (x_o, y_o, z_o)*, and it has* $\nabla f(x_o, y_o, z_o)$ *as its normal vector.*

For example, the tangent plane to S at the point

$$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$$

has normal vector

$$n = \nabla f(1/2, 1/2, 1/\sqrt{2}),$$

where

$$\nabla f(x, y, z) = 2x\ \widehat{i} + 2y\ \widehat{j} + 2z\ \widehat{k};$$

so that

$$n = \widehat{i} + \widehat{j} + \sqrt{2}\ \widehat{k}.$$

Consequently, the tangent plane to S at the point $(1/2, 1/2, 1/\sqrt{2})$ *has equation*

$$(1)\left(x - \frac{1}{2}\right) + (1)\left(y - \frac{1}{2}\right) + (\sqrt{2})\left(z - \frac{1}{\sqrt{2}}\right) = 0,$$

which simplifies to

$$x + y + \sqrt{2}\ z = 2.$$

Chapter 5

Integration

In this chapter we extend the concept of the Riemann integral

$$\int_a^b f(x)\mathrm{d}x$$

for a real valued function, f, defined on a closed and bounded interval $[a,b]$. We begin by defining integrals of scalar fields over curves in $\mathbb{R}^n$ which can be parametrized by C^1 paths.

5.1 Path Integrals

Definition 5.1.1 (Simple Curve). *A curve C in $\mathbb{R}^n$ is said to be a C^1, simple curve if there exists a C^1 path $\sigma\colon I \to \mathbb{R}^n$, for some open interval I containing a closed and bounded interval $[a,b]$, such that*

(i) $\sigma([a,b]) = C$,

(ii) σ *is one–to–one on* $[a,b]$, *and*

(iii) $\sigma'(t)$ *is never the zero vector for all t in I.*

The path σ is called a **parametrization** *of the curve C.*

Example 5.1.2. *Let C denote the arc of the unit circle in $\mathbb{R}^2$ given by*

$$C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1;\ y \geqslant 0;\ 0 \leqslant x \leqslant 1\}.$$

Figure 5.1.1 shows a picture of C. The path $\sigma\colon [0,\pi/2] \to \mathbb{R}^2$ given by

$$\sigma(t) = (\cos t, \sin t) \quad \text{for all} \quad t \in [0,\pi/2]$$

provides a parametrization of C. Observe that σ is a C^1 path defined for all $t \in \mathbb{R}$ since sin *and* cos *are infinitely differentiable functions in all of $\mathbb{R}$. Furthermore, observe that*

$$\sigma'(t) = (-\sin t, \cos t) \quad \text{for all} \quad t \in \mathbb{R}$$

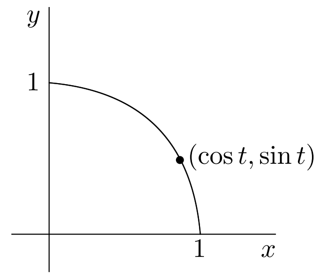

Figure 5.1.1: Curve C

always has norm 1*; thus, condition (iii) in Definition 5.1.1 is satisfied.*

To show that σ *is one–to–one on* $[0, \pi/2]$*, suppose that*

$$\sigma(t_1) = \sigma(t_2)$$

for some t_1 *and* t_2 *in* $[0, \pi/2]$*. Then,*

$$(\cos(t_1), \sin(t_1)) = (\cos(t_2), \sin(t_2))$$

and so

$$\cos(t_1) = \cos(t_2).$$

Since cos *is one–to–one on* $[0, \pi/2]$*, it follows that*

$$t_1 = t_2,$$

and, therefore, σ *is one–to–one. Thus, condition (ii) in Definition 5.1.1 also holds true for* σ*.*

Condition (i) in Definition 5.1.1 is left for the reader to verify.

There are more than one way to parametrize a given simple curve. For instance, in the previous example, we could have used $\gamma\colon [0, \pi] \to \mathbb{R}^2$ given by

$$\gamma(t) = (\cos(t/2), \sin(t/2)) \quad \text{for all} \quad t \in [0, \pi].$$

γ is called a *reparametrization* of the curve C. Observe that, since

$$\|\gamma'(t)\| = \frac{1}{2}, \quad \text{for all} \quad t \in \mathbb{R},$$

this new parametrization of C amounts to traversing the curve C at a slower speed.

Definition 5.1.3. Let $\sigma\colon [a, b] \to \mathbb{R}^n$ be a differentiable, one–to–one path. Suppose also that $\sigma'(t)$, is never the zero vector. Let $h\colon [c, d] \to [a, b]$ be a one–to–one and onto map such that $h'(t) \neq 0$ for all $t \in [c, d]$. Define

$$\gamma(t) = \sigma(h(t)) \quad \text{for all} \quad t \in [c, d].$$

$\gamma\colon [c, d] \to \mathbb{R}^n$ is a called a *reparametrization* of σ

Observe that the path $\sigma\colon [0,1] \to \mathbb{R}^2$ given by

$$\sigma(t) = (t, \sqrt{1-t^2}) \quad \text{for all} \quad t \in [0,1]$$

also parametrizes the quarter circle C in the previous example. However, it is not a C^1 parametrization of C in the sense of Definition 5.1.1 since the derivative map

$$\sigma'(t) = \left(1, -\frac{t}{\sqrt{1-t^2}}\right) \quad \text{for} \quad |t| < 1,$$

does not extend to a continuous map on an open interval containing $[0,1]$ since it is undefined at $t = 1$.

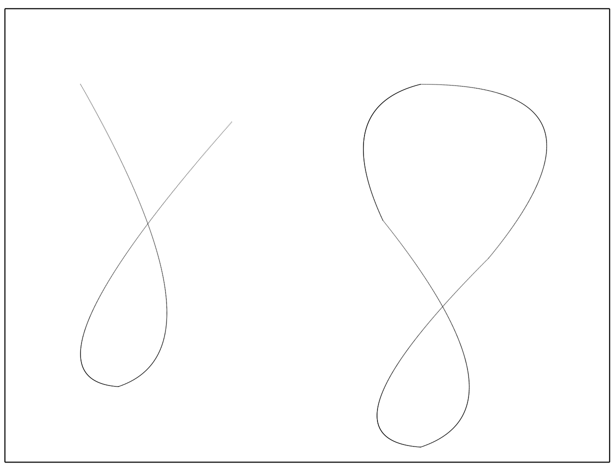

Figure 5.1.2: Curves which are not simple

Definition 5.1.4 (Simple Closed Curve)**.** *A curve C in $\mathbb{R}^n$ is said to be a C^1, simple closed curve if there exists a C^1 parametrization of C, $\sigma\colon [a,b] \to \mathbb{R}^n$, satisfying:*

(i) $\sigma([a,b]) = C$,

(ii) $\sigma(a) = \sigma(b)$,

(iii) σ *is one–to–one on* $[a,b)$*, and*

(iv) $\sigma'(t)$ *is never the zero vector for all* t *where it is defined.*

Example 5.1.5. *The unit circle, C, in $\mathbb{R}^2$ given by*

$$C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\},$$

is a C^1, simple closed curve. The path $\sigma\colon [0, 2\pi] \to \mathbb{R}^2$ given by

$$\sigma(t) = (\cos t, \sin t) \quad \text{for all} \quad t \in [0, 2\pi]$$

provides a C^1 parametrization of C satisfying all the conditions in Definition 5.1.4. The verification of this is left to the reader.

Remark 5.1.6. *Condition (ii) in Definition 5.1.1 and condition (iii) in Definition 5.1.4 guarantee that a simple curve does not have self–intersections or crossings. Thus, the plane curves pictured in Figure 5.1.2 are not simple curves.*

5.1.1 Arc Length

Definition 5.1.7 (Arc Length of a Simple Curve)**.** *Let C denote a simple curve (either closed or otherwise). We define the* arc length *of C, denoted $\ell(C)$, by*

$$\ell(C) = \int_a^b \|\sigma'(t)\|\,dt,$$

where $\sigma\colon [a, b] \to \mathbb{R}^n$ is a C^1 parametrization of C, over a closed and bounded interval $[a, b]$, satisfying the conditions in Definition 5.1.1 (or in Definition 5.1.4 for the case of a simple closed curve).

Example 5.1.8. *Let C denote the quarter of the unit circle in $\mathbb{R}^2$ defined in Example 5.1.2 (see also Figure 5.1.1). In this case,*

$$\sigma(t) = (\cos t, \sin t) \quad \text{for all} \quad t \in [0, \pi/2]$$

provides a C^1 parametrization of C with

$$\sigma'(t) = (-\sin t, \cos t) \quad \text{for all} \quad t \in \mathbb{R};$$

so that $\|\sigma'(t)\| = 1$ for all t and therefore

$$\ell(C) = \int_0^{\pi/2} \|\sigma'(t)\|\,dt = \int_0^{\pi/2} dt = \frac{\pi}{2}.$$

To see why the definition of arc length in Definition 5.1.7 is plausible, consider a simple curve pictured in Figure 5.1.3 and parametrized by the C^1 path

$$\sigma\colon [a, b] \to \mathbb{R}^n.$$

Subdivide the interval $[a, b]$ into N subintervals by means of a partition

$$a = t_o < t_1 < t_2 < \cdots < t_{i-1} < t_i < \cdots < t_i < t_{N-1} < t_N = b.$$

This partition generates a polygon in $\mathbb{R}^n$ constructed by joining $\sigma(t_{i-1})$ to $\sigma(t_i)$ by straight line segments, for $i = 1, 2, \ldots, N$ (see Figure 5.1.3). If we denote the polygon by P, then we can approximate $\ell(C)$ by $\ell(P)$; we then have that

$$\ell(C) \approx \sum_{i=1}^{N} \|\sigma(t_i) - \sigma(t_{i-1})\|.$$

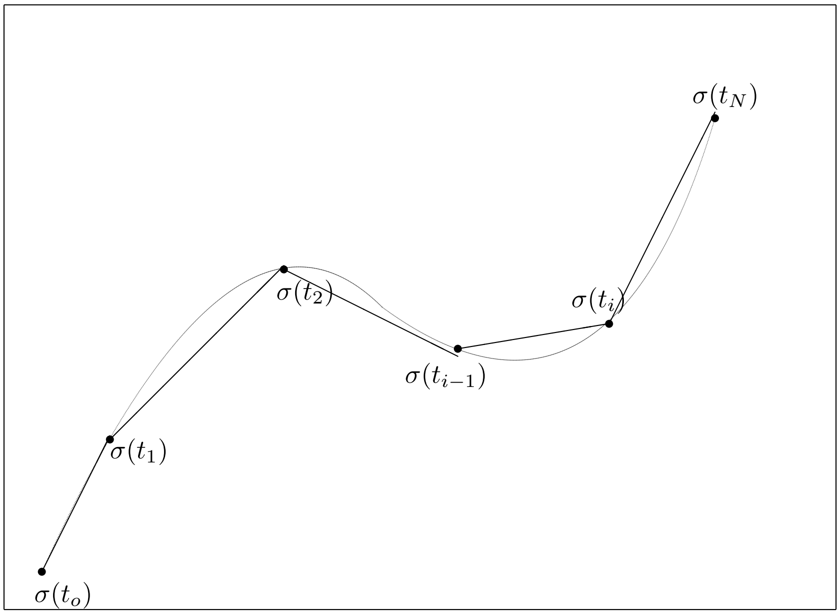

Figure 5.1.3: Approximating arc length

Now, since σ is C^1, and hence differentiable,

$$\sigma(t_i) - \sigma(t_{i-1}) = (t_i - t_{i-1})\sigma'(t_{i-1}) + E_i(t_i - t_{i-1})$$

for each $i = 1, 2, \ldots, N$, where

$$\lim_{h\to 0} \frac{\|E_i(h)\|}{|h|} = 0,$$

for each $i = 1, 2, \ldots, N$. Now, by making N larger and larger, while assuring that the largest of the differences $t_i - t_{i-1}$, for each $i = 1, 2, \ldots, N$, gets smaller and smaller, we can make the further approximation

$$\ell(C) \approx \sum_{i=1}^{N} \|\sigma'(t_{i-1})\|(t_i - t_{i-1}).$$

Observe that the expression

$$\sum_{i=1}^{N} \|\sigma'(t_{i-1})\|(t_i - t_{i-1})$$

is a Riemann sum for the function $\|\sigma'(t)\|$ over the interval $[a, b]$. Now, since we are assuming the σ is of class C^1, it follows that the map $t \mapsto \|\sigma'(t)\|$ is continuous on $[a, b]$. Thus, a theorem from analysis guarantees that the sums

$$\sum_{i=1}^{N} \|\sigma'(t_{i-1})\|(t_i - t_{i-1})$$

converge as $N \to \infty$ while

$$\max_{1 \leqslant i \leqslant N}(t_i - t_{i-1}) \to 0.$$

The limit will be the Riemann integral of $\|\sigma'(t)\|$ over the interval $[a, b]$. Thus, it makes sense to define

$$\ell(C) = \int_a^b \|\sigma'(t)\|\mathrm{d}t.$$

We next see that we will always get the same value of the integral for any C^1 parametrization of σ.

Let $\gamma(t) = \sigma(h(t))$, for all $t \in [c, d]$, be reparametrization of $\sigma\colon [a, b] \to \mathbb{R}^n$; that is, h is a one–to–one, differentiable function from $[c, d]$ to $[a, b]$ with $h'(t) > 0$ for all $t \in (c, d)$. We consider the integral

$$\int_c^d \|\gamma'(t)\|\mathrm{d}t.$$

By the Chain Rule,

$$\gamma'(t) = \frac{d}{dt}[\sigma(h(t))] = h'(t)\sigma'(h(t)).$$

We then have that

$$\begin{aligned} \int_c^d \|\gamma'(t)\|\mathrm{d}t &= \int_c^d \|h'(t)\sigma'(h(t))\|\mathrm{d}t \\ &= \int_c^d \|\sigma'(h(t))\| \; |h'(t)|\mathrm{d}t \\ &= \int_c^d \|\sigma'(h(t))\| \; h'(t)\mathrm{d}t, \end{aligned}$$

since $h'(t) > 0$. Next, make the change of variables $\tau = h(t)$. Then, $\mathrm{d}\tau = h'(t)\mathrm{d}t$ and

$$\int_c^d \|\sigma'(h(t))\|h'(t)\mathrm{d}t = \int_a^b \|\sigma'(\tau)\|\mathrm{d}\tau.$$

It then follows from Definition 5.1.7 that

$$\ell(C) = \int_c^d \|\gamma'(t)\|\mathrm{d}t$$

for any reparametrization $\gamma = \sigma \circ h$ of σ, with $h' > 0$. In the case in which $h' < 0$, we get the same result with the understanding that $h(c) = b$ and $h(d) = a$. Thus, any reparametrization of σ will yield the same value for the integral $\ell(C)$ given in Definition 5.1.7.

It remains to see that any two parametrizations

$$\sigma\colon [a,b] \to \mathbb{R}^n \quad \text{ and } \gamma\colon [c,d] \to \mathbb{R}^n$$

of a simple curve C are reparametrizations of each other. This will be proved in Appendix B.

5.1.2 Defining the Path Integral

Let U be an open subset of $\mathbb{R}^n$ and C be a C^1 simple curve (closed or otherwise) which is entirely contained in U. Suppose that $f\colon U \to \mathbb{R}$ is a continuous scalar field defined on U. We define the integral of f over the curve C, denoted by

$$\int_C f,$$

as follows:

$$\int_C f = \int_a^b f(\sigma(t))\|\sigma'(t)\|\mathrm{d}t, \tag{5.1}$$

where $\sigma\colon [a,b] \to \mathbb{R}^n$ is a C^1 parametrization of C, over a closed and bounded interval $[a,b]$, satisfying the conditions in Definition 5.1.1 (or in Definition 5.1.4 for the case of a simple closed curve).

$\displaystyle\int_C f$ is called the *path integral* of f over C. This integral is guaranteed to exist as a limit of Riemann sums of the function $f(\sigma(t))\|\sigma'(t)\|$ over $[a,b]$ by virtue of the continuity of f and the fact that σ is a C^1 parametrization of C.

Example 5.1.9. *A metal wire is in the shape of the portion of a parabola $y = x^2$ from $x = -1$ to $x = 1$. Suppose the linear mass density along the wire (in grams per centimeter) is proportional to the distance to the y–axis (the axis of the parabola). Compute the mass of the wire.*
Solution: The wire is parametrized by the path

$$\sigma(t) = (t, t^2) \quad \text{for} \quad -1 \leqslant t \leqslant 1.$$

Let C denote the image of σ. Let $f(x,y)$ denote the linear mass density of the wire. Then, $f(x,y) = k|x|$ for some constant of proportionality k. It then follows that the mass of the wire is

$$M = \int_C f = \int_{-1}^{1} k|t|\|\sigma'(t)\|\mathrm{d}t,$$

where

$$\sigma'(t) = (1, 2t),$$

so that

$$\|\sigma'(t)\| = \sqrt{1+4t^2}.$$

Hence, by the symmetry of the wire with respect to the y axis

$$M = \int_C f = 2\int_0^1 kt\sqrt{1+4t^2}\mathrm{d}t.$$

Evaluating this integral yields

$$M = \frac{k}{6}(5\sqrt{5}-1).$$

□

The definition of $\displaystyle\int_C f$ given in (5.1) is based on a choice of parametrization, $\sigma\colon [a,b] \to \mathbb{R}^n$, for C. Thus, in order to see that $\displaystyle\int_C f$ is well defined, we need to show that the value of $\displaystyle\int_C f$ is independent of the choice of parametrization; more precisely, we need to see that if $\gamma\colon [c,d] \to \mathbb{R}^n$ is another parametrization of C, then

$$\int_c^d f(\gamma(t))\|\gamma'(t)\mathrm{d}t = \int_a^b f(\sigma(t))\|\sigma'(t)\|\mathrm{d}t. \tag{5.2}$$

For the case in which γ is a reparametrization of σ; that is, the case in which $\gamma(t) = \sigma(h(t))$, for all $t \in [c,d]$, where h is a one–to–one, differentiable function from $[c,d]$ to $[a,b]$ with $h'(t) > 0$ for all $t \in (c,d)$. We see that (5.2) follows from the Chain Rule and the change of variables: $\tau = h(t)$, for $t \in [c,d]$. In fact we have

$$\gamma'(t) = \frac{d}{dt}[\sigma(h(t))] = h'(t)\sigma'(h(t)),$$

so that

$$\int_c^d f(\gamma(t))\|\gamma'(t)\|\mathrm{d}t \quad = \quad \int_c^d f(\sigma(h(t)))\|\|\sigma'(h(t))\|\ h'(t)\mathrm{d}t,$$

since $h'(t) > 0$. Thus, since $\mathrm{d}\tau = h'(t)\mathrm{d}t$, we can write

$$\int_c^d f(\gamma(t))\|\gamma'(t)\|\mathrm{d}t = \int_a^b f(\sigma(\tau))\|\sigma'(\tau)\|\mathrm{d}\tau,$$

which is (5.2) for the case in which one of the paths is reparametrization of the other. Finally, using the results of Appendix B in this notes, we see that (5.2) holds for any two parametrizations, $\sigma\colon [a,b] \to \mathbb{R}^n$ and $\sigma\colon [c,d] \to \mathbb{R}^n$, of the C^1 simple curve, C.

5.2 Line Integrals

In the previous section we saw how to integrate a scalar field on a C^1, simple curve. In this section we describe how to integrate vector fields on curves. Technically, what we'll be doing is integrating a component (which is a scalar) of a vector field on the given curve. More precisely, let U denote an open subset of $\mathbb{R}^n$ and let $F\colon U \to \mathbb{R}^n$ be a vector field on U. Suppose that there is a curve, C, which is contained in U and which is parametrized by a C^1 path

$$\sigma\colon [a,b] \to \mathbb{R}^n.$$

We have seen that the vector $\sigma'(t)$ gives the tangent direction to the path at $\sigma(t)$. The vector

$$T(t) = \frac{1}{\|\sigma'(t)\|}\sigma'(t)$$

is, therefore, a unit tangent vector to the path. The tangential component of the of the vector field, F, is then given by the dot product of F and T:

$$F \cdot T.$$

The *line integral* of F on the curve C parametrized by σ is given by

$$\int_C F \cdot T \mathrm{d}s = \int_a^b F(\sigma(t)) \cdot T(t)\ \|\sigma'(t)\| \mathrm{d}t.$$

Observe that we can re–write this as

$$\int_C F \cdot T \mathrm{d}s \quad = \quad \int_a^b F(\sigma(t)) \cdot \frac{1}{\|\sigma'(t)\|}\sigma'(t)\ \|\sigma'(t)\| \mathrm{d}t;$$

therefore,

$$\int_C F \cdot T \mathrm{d}s \quad = \quad \int_a^b F(\sigma(t)) \cdot \sigma'(t) \mathrm{d}t. \tag{5.3}$$

Example 5.2.1. *Let* $F\colon \mathbb{R}^2\backslash\{(0,0)\| \to \mathbb{R}^2$ *be given by*

$$F(x,y) = \frac{-y}{x^2+y^2}\ \widehat{i} + \frac{x}{x^2+y^2}\ \widehat{j} \quad \textit{for} \quad (x,y) \neq (0,0),$$

and let C *denote the unit circle traversed in the counterclockwise direction. Evaluate* $\displaystyle\int_C F \cdot T\, ds$.

Solution: The path

$$\sigma(t) = (\cos t, \sin t), \quad \text{for} \quad t \in [0, 2\pi],$$

is a C^1 parametrization for C with

$$\sigma'(t) = (-\sin t, \cos t), \quad \text{for} \quad t \in \mathbb{R}.$$

Applying the definition of the line integral in (5.3) yields

$$\begin{aligned}\int_C F\cdot T\mathrm{d}s &= \int_0^{2\pi} F(\cos t,\sin t)\cdot(-\sin t,\cos t)\mathrm{d}t\\ &= \int_0^{2\pi}(\sin^2 t+\cos^2 t)\mathrm{d}t\\ &= 2\pi.\end{aligned}$$

□

Let

$$F(x,y,z)=P(x,y,z)\,\widehat{\imath}+Q(x,y,z)\,\widehat{\jmath}$$

denote a vector filed defined in a region U of $\mathbb{R}^2$, where P and Q are continuous scalar fields defined on U. Let

$$\sigma(t)=x(t)\,\widehat{\imath}+y(t)\,\widehat{\jmath},\quad \text{for}\quad t\in[a,b],$$

be a C^1 parametrization of a C^1 curve, C, contained in U. Then

$$\sigma'(t)=x'(t)\,\widehat{\imath}+y'(t)\,\widehat{\jmath}\quad \text{for}\quad t\in(a,b),$$

and, applying the definition of the line integral of F on C in (5.3) yields

$$\begin{aligned}\int_C F\cdot T\mathrm{d}s &= \int_a^b (P(x(t),y(t))x'(t)+Q(x(t),y(t))y'(t))\mathrm{d}t\\ &= \int_a^b (P(x(t),y(t))x'(t)\mathrm{d}t+Q(x(t),y(t))y'(t)\mathrm{d}t)\end{aligned}$$

Next, use the notation $\mathrm{d}x=x'(t)\mathrm{d}t$ and $\mathrm{d}y=y'(t)\mathrm{d}t$ for the differentials of x and y, respectively, to re–write the line integral as

$$\int_C F\cdot T\mathrm{d}s=\int_C P\mathrm{d}x+Q\mathrm{d}y. \tag{5.4}$$

Equation (5.4) suggests another way to evaluate the line integral of a 2–dimensional vector field on a plane curve.

Example 5.2.2. *Evaluate the line integral* $\displaystyle\int_C -y\,dx+(x-1)\,dy$, *where C is the simple closed curve made up of the line segment from $(-1,0)$ to $(1,0)$ and the top portion of the unit circle traversed in the counterclockwise direction (see picture in Figure 5.2.4).*

Solution: Observe that C is not a C^1 curve since no tangent vector can be defined at the points $(-1,0)$ and $(1,0)$. However, C can be decomposed into two C^1 curves (see Figure 5.2.4):

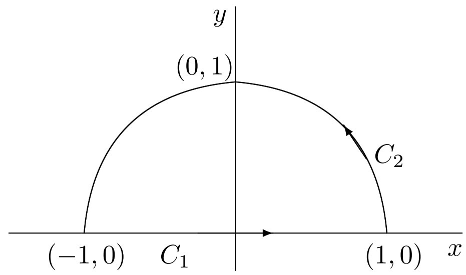

Figure 5.2.4: Example 5.2.2 Picture

(i) C_1: the directed line segment from $(-1, 0)$ to $(1, 0)$, and

(ii) $C_2 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1, y \geqslant 0\}$; the top portion of the unit circle in $\mathbb{R}^2$ traversed in the counterclockwise sense.

Then,

$$\int_C -y\mathrm{d}x + (x-1)\mathrm{d}y = \int_{C_1} -y\mathrm{d}x + (x-1)\mathrm{d}y + \int_{C_2} -y\mathrm{d}x + (x-1)\mathrm{d}y.$$

We evaluate each of the integrals separately.

On C_1: $x = t$ and $y = 0$ for $-1 \leqslant t \leqslant 1$; so that $\mathrm{d}x = \mathrm{d}t$ and $\mathrm{d}y = 0$. Thus,

$$\int_{C_1} -y\mathrm{d}x + (x-1)\mathrm{d}y = 0.$$

On C_2: $x = \cos t$ and $y = \sin t$ for $0 \leqslant t \leqslant \pi$; so that $\mathrm{d}x = -\sin t\mathrm{d}t$ and $\mathrm{d}y = \cos t\mathrm{d}t$. Thus

$$\begin{aligned}
\int_{C_2} -y\mathrm{d}x + (x-1)\mathrm{d}y &= \int_0^\pi (-\sin t(-\sin t)\mathrm{d}t + (\cos t - 1)\cos t\mathrm{d}t) \\
&= \int_0^\pi (\sin^2 t + \cos^2 t - \cos t)\mathrm{d}t \\
&= \int_0^\pi (1 - \cos t)\mathrm{d}t \\
&= [t - \sin t]_0^\pi \\
&= \pi.
\end{aligned}$$

It then follows that

$$\int_C -y\mathrm{d}x + (x-1)\mathrm{d}y = \pi.$$

□

We can obtain an analogous equation to that in (5.4) for the case of a three dimensional field

$$F = P\,\widehat{i} + Q\,\widehat{j} + R\,\widehat{k},$$

where P, Q and R are scalar fields defined in some region U of $\mathbb{R}^3$ which contains the simple curve C:

$$\int_C F \cdot T \mathrm{d}s = \int_C P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z. \tag{5.5}$$

5.3 Gradient Fields

Suppose that a field $F\colon U \to \mathbb{R}^n$ is the gradient of a C^1 scalar field, f, defined on U; that is, $F = \nabla f$. Then, for any C^1 parametrization,

$$\sigma\colon [0,1] \to \mathbb{R}^n,$$

of a curve C in U connecting a point x_o to x_1, also in U,

$$\begin{aligned}
\int_C F \cdot T \mathrm{d}s &= \int_0^1 F(\sigma(t)) \cdot \sigma'(t)\mathrm{d}t \\
&= \int_0^1 \nabla f(\sigma(t)) \cdot \sigma'(t)\mathrm{d}t \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(f(\sigma(t))\right)\mathrm{d}t \\
&= f(\sigma(1)) - f(\sigma(0)) \\
&= f(x_1) - f(x_o).
\end{aligned}$$

Thus, the line integral of $F = \nabla f$ on a curve C is determined by the values of f at the endpoints of the curve.

A field F with the property that $F = \nabla f$, for a C^1 scalar field, f, is called a *gradient field,* and f is called a *potential* for the field F.

Example 5.3.1 (Gravitational Potential)**.** *According to Newton's Law of Universal Gravitation, the earth exerts a gravitational pull on an object of mass m at a point (x, y, z) above the surface of the earth, which is at a distance of*

$$r = \sqrt{x^2 + y^2 + z^2}$$

from the center of the earth (located at the origin of three dimensional space), an is given by

$$F(x, y, z) = -\frac{km}{r^2}\widehat{r}, \tag{5.6}$$

where $\widehat{r}$ is a unit vector in the direction of the vector $\vec{r} = x\,\widehat{i} + y\,\widehat{j} + z\,\widehat{k}$. The minus sign indicates that the force is directed towards the center of the earth.

Show that the field F is a gradient field.

Solution: We claim that $F = \nabla f$, where

$$f(r) = \frac{km}{r} \quad \text{and} \quad r = \sqrt{x^2 + y^2 + z^2} \neq 0. \tag{5.7}$$

To see why this is so, use the Chain Rule to compute

$$\frac{\partial f}{\partial x} = f'(r)\frac{\partial r}{\partial x} = -\frac{km}{r^2}\frac{x}{r}.$$

Similarly,

$$\frac{\partial f}{\partial y} = -\frac{km}{r^2}\frac{y}{r}, \text{ and } \frac{\partial f}{\partial z} = -\frac{km}{r^2}\frac{z}{r}.$$

It then follows that

$$\begin{aligned}
\nabla f &= \frac{\partial f}{\partial x}\,\widehat{i} + \frac{\partial f}{\partial y}\,\widehat{j} + \frac{\partial f}{\partial z}\,\widehat{k} \\
&= -\frac{km}{r^2}\frac{x}{r}\,\widehat{i} - \frac{km}{r^2}\frac{y}{r}\,\widehat{j} - \frac{km}{r^2}\frac{z}{r}\,\widehat{k} \\
&= -\frac{km}{r^2}\left(\frac{x}{r}\,\widehat{i} + \frac{y}{r}\,\widehat{j} + \frac{z}{r}\,\widehat{k}\right) \\
&= -\frac{km}{r^2}\,\frac{1}{r}\left(x\,\widehat{i} + y\,\widehat{j} + z\,\widehat{k}\right) \\
&= -\frac{km}{r^2}\,\widehat{r},
\end{aligned}$$

which is the vector field F defined in (5.6). □

It follows from the fact that the Newtonian gravitational field F defined in (5.6) is a gradient field that the line integral of F along any curve in $\mathbb{R}^3$, which does not go through the origin, connecting $\vec{r}_o = (x_o, y_o, z_o)$ to $\vec{r}_1 = (x_1, y_1, z_1)$, is given by

$$\int_C F \cdot T \mathrm{d}s = f(x_1, y_1, z_1) - f(x_o, y_o, z_o) = \frac{km}{r_1} - \frac{km}{r_o},$$

where $r_o = \sqrt{x_o^2 + y_o^2 + z_o^2}$ and $r_1 = \sqrt{x_1^2 + y_1^2 + z_1^2}$. The function f defined in (5.7) is called the gravitational potential.

5.4 Flux Across Plane Curves

According the Jordan Curve Theorem, a simple closed curve in the plane divides the plane into two connected regions:

(i) a bounded region called the "inside" of the curve, and

(ii) an unbounded region called the "outside" of the curve.

Let C denote a C^1, simple, closed curve in the plane parametrized by the C^1 path

$$\sigma\colon [a,b] \to \mathbb{R}^2.$$

We can then define a unit vector, $\widehat{n}$, perpendicular to to the tangent unit vector, T, to the curve, and pointing towards the outside of the curve. $\widehat{n}$ is called the *outward unit normal* to the curve.

Example 5.4.1. *The outward unit normal to the unit circle, C, parametrized by the path*

$$\sigma(t) = (\cos t, \sin t), \quad \textit{for } t \in [0, \pi],$$

is the vector

$$\widehat{n}(t) = (\cos t, \sin t), \quad \textit{for } t \in [0, \pi].$$

In general, if the parametrization of a C^1, simple, closed curve, C, is given by

$$\sigma(t) = (x(t), y(t)) \quad \text{for} \quad a \leqslant t \leqslant b,$$

where x and y are C^1 functions of t, then the vector

$$\widehat{n}(t) = \pm\frac{1}{\|\sigma'(t)\|}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\,\widehat{\imath} - \frac{\mathrm{d}x}{\mathrm{d}t}\,\widehat{\jmath}\right),$$

where the sign is chosen appropriately, will be the outward unit normal to the curve. We assume, for convenience, that the path σ is always oriented so that the positive sign indicates the outward direction.

Given a vector field, $F = P\,\widehat{\imath} + Q\,\widehat{\jmath}$, defined on a region containing a C^1, simple, closed curve, C, we define the *flux* of F across C to be the integral

$$\begin{aligned}
\int_C F\cdot\widehat{n}\mathrm{d}s &= \int_a^b F(\sigma(t))\cdot\frac{1}{\|\sigma'(t)\|}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\,\widehat{\imath} - \frac{\mathrm{d}x}{\mathrm{d}t}\,\widehat{\jmath}\right)\|\sigma'(t)\|\mathrm{d}t \\
&= \int_a^b (P\,\widehat{\imath} + Q\,\widehat{\jmath})\cdot\left(\frac{\mathrm{d}y}{\mathrm{d}t}\,\widehat{\imath} - \frac{\mathrm{d}x}{\mathrm{d}t}\,\widehat{\jmath}\right)\mathrm{d}t \\
&= \int_a^b \left(P\frac{\mathrm{d}y}{\mathrm{d}t} - Q\frac{\mathrm{d}x}{\mathrm{d}t}\right)\mathrm{d}t
\end{aligned}$$

Thus, using the definitions of the differentials of x and y, we can write the flux of F across the curve C as

$$\int_C F\cdot\widehat{n}\mathrm{d}s = \int_C P\mathrm{d}y - Q\mathrm{d}x. \tag{5.8}$$

Example 5.4.2. *Compute the flux of the field* $F(x,y) = x\,\widehat{\imath} + y\,\widehat{\jmath}$ *across the unit circle*

$$C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

traversed in the counterclockwise direction.
Solution: Parametrize the circle with $x = \cos t$, $y = \sin t$, for $t \in [0, 2\pi]$. Then, $\mathrm{d}x = -\sin t\mathrm{d}t$, $\mathrm{d}y = \cos t$, and, using the definition of flux in (5.8),

$$\begin{aligned}\int_C F \cdot \widehat{n}\mathrm{d}s &= \int_C P\mathrm{d}y - Q\mathrm{d}x \\ &= \int_0^{2\pi} (\cos^2 t + \sin^2 t)\mathrm{d}t \\ &= 2\pi.\end{aligned}$$

□

Remark 5.4.3 (Interpretation of the flux of a vector field)**.** An interpretation of the flux of a vector field is provided by the following situation in *fluid dynamics:* Let $V(x,y)$ denote the velocity field of a plane fluid in some region U in $\mathbb{R}^2$ containing the simple closed curve C. Then, at each point (x,y) in U, $V(x,y)$ gives the velocity of the fluid as it goes through that point in units of length per unit time. Suppose we know the density of the fluid as a function, $\rho(x,y)$, of the position of the fluid in U (this is a scalar field) in units of mass per unit area (since this is a two–dimensional fluid). Then, the vector field

$$F(x,y) = \rho(x,y)V(x,y),$$

in units of mass per unit length per unit time, gives the rate of fluid flow per unit length at the point (x,y). The integrand

$$F \cdot \widehat{n}\mathrm{d}s,$$

in the flux definition in (5.8), is then in units of mass per unit time and measures the amount of fluid that crosses a section of the curve C of length $\mathrm{d}s$ in the outward normal direction. The flux then gives the rate at which the fluid is crossing the curve C from the inside to the outside; in other words, the flux gives the rate of flow of fluid out of the region bounded by C.

Notation 5.4.4. Let U denote an open subset of $\mathbb{R}^2$ and $F\colon U \to \mathbb{R}^2$ be a continuous vector field defined on U. Let $C \subset U$ be a piecewise C^1, simple, closed curve oriented in the counterclockwise sense. We usually denote the flux of F across C by

$$\oint_C F \cdot dn, \tag{5.9}$$

where the circle across the integral sign in (5.9) reminds us that the line integral is computed along a closed curve. Thus, if $F(x,y) = P(x,y)\,\widehat{\imath} + Q(x,y)\,\widehat{\jmath}$, for

$(x, y) \in U$, where P and Q are real valued continuous functions defined in U, we have write

$$\oint_C F \cdot dn = \oint_C P\ dy - Q\ dx, \tag{5.10}$$

to denote the flux of F across C.

If the curve C in (5.10) is not C^1, but is piece–wise C^1, we can still compute the flux of $F = P\widehat{i} + Q\widehat{j}$ across C by decomposing C as a union of finitely many C^1 pieces, $C_1, C_2, \ldots, C_k$, and then computing

$$\oint_C F \cdot dn = \sum_{i=i}^{k} \int_{C_i} P\ dy - Q\ dx, \tag{5.11}$$

where $C = \bigcup_{i=1}^{k} C_i$, and the orientation of each C_i, for $i = 1, 2, \ldots, k$, is consistent with that of C.

Example 5.4.5 (Flux across the boundary of a rectangular region)**.** Let U denote an open subset of $\mathbb{R}^2$ and suppose that $P\colon U \to \mathbb{R}$ and $Q\colon U \to \mathbb{R}$ are C^1, real–valued, functions defined on U. Let (x_o, y_o) denote a point in U and suppose that $h > 0$ and $k > 0$ are sufficiently small so that the rectangle $\left[x_o - \dfrac{h}{2}, x_o + \dfrac{h}{2}\right] \times \left[y_o - \dfrac{k}{2}, y_o + \dfrac{k}{2}\right]$ lies entirely in U, where, denoting the rectangle by R,

$$R = \left\{(x, y) \in \mathbb{R}^2 \middle| x_o - \frac{h}{2} \leqslant x \leqslant x_o + \frac{h}{2} \text{ and } y_o - \frac{k}{2} \leqslant y \leqslant y_o + \frac{k}{2}\right\}. \tag{5.12}$$

Let C denote the boundary of the rectangle R in (5.12) oriented in the counter-clockwise sense; this is pictured in Figure 5.4.5, where P_1, P_2, P_3 and P_4 denote

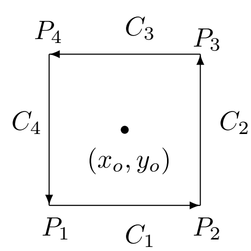

Figure 5.4.5: Boundary of R

the points

$$\left(x_o - \frac{h}{2}, y_o - \frac{k}{2}\right), \left(x_o + \frac{h}{2}, y_o - \frac{k}{2}\right), \left(x_o + \frac{h}{2}, y_o + \frac{k}{2}\right) \text{ and } \left(x_o - \frac{h}{2}, y_o + \frac{k}{2}\right),$$

respectively, and C_1, C_2, C_3 and C_4 are the directed line segments

$$[P_1, P_2],\ [P_2, P_3],\ [P_3, P_4] \text{ and } [P_4, P_1],$$

respectively.

According to (5.11), the flux of F across the boundary of the rectangle R defined in (5.12) is given by

$$\oint_C F \cdot dn = \sum_{i=i}^{4} \int_{C_i} P\ dy - Q\ dx. \tag{5.13}$$

We evaluate each one of the line integrals on the right–hand side of (5.13) separately.

On C_1 we use the parametrization

$$x = t \text{ and } y = y_o - \frac{k}{2}, \quad \text{for } x_o - \frac{h}{2} \leqslant t \leqslant x_o + \frac{h}{2};$$

so that,

$$dx = dt \quad \text{and} \quad dy = 0\ dt.$$

Consequently,

$$\int_{C_1} P\ dy - Q\ dx = -\int_{x_o - h/2}^{x_o + h/2} Q(t, y_o - k/2)\ dt. \tag{5.14}$$

On C_2, we use the parametrization

$$x = x_o + \frac{h}{2} \text{ and } y = t, \quad \text{for } y_o - \frac{k}{2} \leqslant t \leqslant y_o + \frac{kh}{2};$$

so that,

$$dx = 0\ dt \quad \text{and} \quad dy = dt.$$

We then have that

$$\int_{C_2} P\ dy - Q\ dx = \int_{y_o - k/2}^{y_o + k/2} P(x_o + h/2, t)\ dt. \tag{5.15}$$

On C_3, use the parametrization

$$x = x_o + \frac{h}{2} - h\tau \text{ and } y = y_o + \frac{k}{2}, \quad \text{where } 0 \leqslant \tau \leqslant 1;$$

so that,

$$dx = -h\ d\tau \quad \text{and} \quad dy = 0\ d\tau.$$

Thus,

$$\int_{C_3} P\ dy - Q\ dx = \int_0^1 Q(x_o + h/2 - h\tau, y_o + k/2)\ h\ d\tau. \tag{5.16}$$

Next, make the change of variables $t = x_o + h/2 - h\tau$ in the integral on the right–hand side of (5.16) to get

$$\int_{C_3} P\ dy - Q\ dx = -\int_{x_o+h/2}^{x_o-h/2} Q(t, y_o + k/2)\ dt,$$

which can be written as

$$\int_{C_3} P\ dy - Q\ dx = \int_{x_o-h/2}^{x_o+h/2} Q(t, y_o + k/2)\ dt. \tag{5.17}$$

On C_4, use the parametrization

$$x = x_o - \frac{h}{2} \text{ and } y = y_o + k/2 - h\tau, \quad \text{for } 0 \leqslant \tau \leqslant 1;$$

so that,

$$dx = 0\ dt \quad \text{and} \quad dy = -h\ d\tau.$$

We then have that

$$\int_{C_4} P\ dy - Q\ dx = -\int_0^1 P(x_o - h/2, y_o + k/2 - h\tau)h\ d\tau. \tag{5.18}$$

Next, make the change of variables $t = y_o + k/2 - h\tau$ in the integral on the right–hand side of (5.18) to get

$$\int_{C_4} P\ dy - Q\ dx = \int_{y_o+k/2}^{y_o-k/2} P(x_o - h/2, t)\ dt,$$

from which we obtain

$$\int_{C_4} P\ dy - Q\ dx = -\int_{y_o-k/2}^{y_o+k/2} P(x_o - h/2, t)\ dt. \tag{5.19}$$

Putting together the results in (5.14), (5.15), (5.17) and (5.19) into (5.13), we obtain that

$$\begin{aligned}\oint_C F \cdot dn \quad &= \quad \int_{y_o-k/2}^{y_o+k/2} [P(x_o + h/2, t) - P(x_o - h/2, t)]\ dt \\ &\quad + \int_{x_o-h/2}^{x_o+h/2} [Q(t, y_o + k/2) - Q(t, y_o - k/2)]\ dt.\end{aligned} \tag{5.20}$$

Example 5.4.6 (Divergence of a vector field)**.** In this example, we continue with the set up laid out in Example 5.4.5. Up to this point, we have not used the assumptions that P and Q are C^1 scalar fields defined in an open region U that contains the rectangle R defined in (5.12) and pictured in Figure 5.4.5.

Since we are assuming the $P\colon U \to \mathbb{R}$ is a C^1 function, the partial derivative $\dfrac{\partial P}{\partial x}$ is continuous; hence, we can apply the Mean Value Theorem to obtain a real value $\xi_1(x_o, t, h)$ such that

$$x_o - \frac{h}{2} < \xi_1(x_o, t, h) < x_o + \frac{h}{2}, \quad \text{for } y_o - \frac{k}{2} \leqslant t \leqslant y_o + \frac{k}{2}, \tag{5.21}$$

and

$$P(x_o + h/2, t) - P(x_o - h/2, t) = h\frac{\partial P}{\partial x}(\xi_1(x_o, t, h), t), \tag{5.22}$$

for $y_o - \dfrac{k}{2} \leqslant t \leqslant y_o + \dfrac{k}{2}$.

Observe that it follows from (5.21) that

$$\lim_{h\to 0} \xi_1(x_o, t, h) = x_o, \quad \text{for } y_o - \frac{k}{2} \leqslant t \leqslant y_o + \frac{k}{2}. \tag{5.23}$$

Similarly, since $Q\colon U \to \mathbb{R}$ is C^1, there exists $\eta_2(y_o, t, k) \in \mathbb{R}$ such that

$$y_o - \frac{k}{2} < \eta_2(y_o, t, k) < y_o + \frac{k}{2}, \quad \text{for } x_o - \frac{h}{2} \leqslant t \leqslant x_o + \frac{h}{2}, \tag{5.24}$$

and

$$Q(t, y_o + k/2) - Q(t, y_o - k/2) = k\frac{\partial Q}{\partial y}(t, \eta_2(y_o, t, k)), \tag{5.25}$$

for $x_o - \dfrac{h}{2} \leqslant t \leqslant x_o + \dfrac{h}{2}$.

Note that, as a consequence of (5.24),

$$\lim_{k\to 0} \eta_2(y_o, t, k) = y_o, \quad \text{for } x_o - \frac{h}{2} \leqslant t \leqslant x_o + \frac{h}{2}. \tag{5.26}$$

We can now substitute the expressions in (5.22) and (5.25) into the integrands on the right hand side of (5.20) to get

$$\begin{aligned} \oint_C F \cdot dn &= \int_{y_o-k/2}^{y_o+k/2} h\frac{\partial P}{\partial x}(\xi_1(x_o, t, h), t)\; dt \\ &\quad + \int_{x_o-h/2}^{x_o+h/2} k\frac{\partial Q}{\partial y}(t, \eta_2(y_o, t, k))\; dt. \end{aligned} \tag{5.27}$$

Assume for the moment that the integrands on the right–hand side of (5.27) are continuous (recall that we are also assuming that P is a C^1 scalar field); so, we can apply the Mean Value Theorem for Integrals to get that there exists a real value $\eta_1(y_o, k)$ such that

$$y_o - \frac{k}{2} < \eta_1(y_o, k) < y_o + \frac{k}{2}, \tag{5.28}$$

and

$$\int_{y_o-k/2}^{y_o+k/2} \frac{\partial P}{\partial x}(\xi_1(x_o,t,h),t)\; dt = k\frac{\partial P}{\partial x}(\xi_1(x_o,\eta_1,h),\eta_1), \tag{5.29}$$

where we have written η_1 for $\eta_1(y_o,k)$.

We note from (5.28) that

$$\lim_{k\to 0} \eta_1(y_o,k) = y_o. \tag{5.30}$$

Similarly, there exists $\xi_2(x_o,h)\in\mathbb{R}$ such that

$$x_o - \frac{h}{2} < \xi_2(x_o,h) < x_o + \frac{h}{2}, \tag{5.31}$$

and

$$\int_{x_o-h/2}^{x_o+h/2} \frac{\partial Q}{\partial y}(t,\eta_2(y_o,t,k))\; dt = h\frac{\partial Q}{\partial y}(\xi_2,\eta_2(y_o,\xi_2,k)), \tag{5.32}$$

where we have written ξ_2 for $\xi_2(x_o,h)$.

As a consequence of (5.31) we have that

$$\lim_{h\to 0} \xi_2(x_o,h) = x_o. \tag{5.33}$$

Substituting the expressions in (5.29) and (5.32) into the right–hand side of (5.27), we get

$$\oint_C F\cdot dn = hk\left[\frac{\partial P}{\partial x}(\xi_1(x_o,\eta_1,h),\eta_1) + \frac{\partial Q}{\partial y}(\xi_2,\eta_2(y_o,\xi_2,k))\right]. \tag{5.34}$$

Next, assume that $hk>0$ and divide the equation in (5.34) by hk to get

$$\frac{1}{\operatorname{area}(R)}\oint_C F\cdot dn = \frac{\partial P}{\partial x}(\xi_1(x_o,\eta_1,h),\eta_1) + \frac{\partial Q}{\partial y}(\xi_2,\eta_2(y_o,\xi_2,k)), \tag{5.35}$$

where we have written area(R) for hk, the area of the rectangle R defined in (5.12). Note that the simple, closed curve C is the boundary of the rectangle R, which we shall denote by ∂R.

Finally, use the assumptions that P and Q are C^1 scalar fields defined in U, together with the limit facts in (5.23), (5.26), (5.30) and (5.33), to obtain from (5.35) that

$$\lim_{h^2+k^2\to 0}\frac{1}{\operatorname{area}(R)}\oint_{\partial R} F\cdot dn = \frac{\partial P}{\partial x}(x_o,y_o) + \frac{\partial Q}{\partial y}(x_o,y_o). \tag{5.36}$$

The expression on the right–hand side of (5.36) is called the **divergence** of the vector field $F\colon U\to\mathbb{R}^2$ given by

$$F(x,y) = P(x,y)\widehat{\imath} + Q(x,y)\widehat{\jmath}, \quad \text{for } (x,y)\in U, \tag{5.37}$$

where P and Q are assumed to have partial derivatives in U, at the point (x_o,y_o).

Definition 5.4.7 (Divergence of a vector field in $\mathbb{R}^2$)**.** Let U be an open subset of $\mathbb{R}^2$ and $F\colon U \to \mathbb{R}^2$ be a vector field in U given by (5.37), where the components of F, P and Q, have partial derivatives defined in U. The divergence of F in U is a scalar field, denoted by $\mathrm{div}(F)\colon U \to \mathbb{R}$, given by

$$\mathrm{div}(F)(x,y) = \frac{\partial P}{\partial x}(x,y) + \frac{\partial Q}{\partial y}(x,y), \quad \text{for all } (x,y) \in U. \tag{5.38}$$

Remark 5.4.8 (Interpretation of Divergence)**.** For the case in which F is the flow field discussed in Remark 5.4.3 on page 93 in these notes, and F is C^1, the left–hand side of (5.36) measures the rate of fluid, in units of mass per unit time per unit area, emanating from the point (x_o, y_o). Thus, $\mathrm{div}F(x_o, y_o)$ gives the rate of fluid emanating from the point (x_o, y_o) per unit area; hence, the name of divergence for the expression on right–hand side the equations in (5.36) and (5.38).

5.5 Differential Forms

The expression $P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z$ in equation (5.4), where P, Q and R are scalar fields defined in some open region in $\mathbb{R}^3$ is an example of a *differential form;* more precisely, it is called a differential 1–form. The discussion presented in this section parallels the discussion found in Chapter 11 of Baxandall and Liebeck's text.

Let U denote an open subset of $\mathbb{R}^n$. Denote by $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ the space of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}$. The space $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ is also referred to as the dual of $\mathbb{R}^n$ and denoted by $(\mathbb{R}^n)^*$.

Definition 5.5.1 (Preliminary Definition of Differential 1–Forms in U)**.** A differential 1–form, ω, is a map $\omega\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ which assigns to each $p \in U$, a linear transformation $\omega_p\colon \mathbb{R}^n \to \mathbb{R}$.

It was shown in Problem 4 of Assignment 3 that to every linear transformation $\omega_p\colon \mathbb{R}^n \to \mathbb{R}$ there corresponds a unique vector, $w_p \in \mathbb{R}^n$, such that

$$\omega_p(h) = w_p \cdot h, \quad \text{for all } h \in \mathbb{R}^n. \tag{5.39}$$

Denoting the vector w_p by $(F_1(p), F_2(p), \ldots, F_n(p))$, we can then write the expression in (5.39) as

$$\omega_p(h) = F_1(p)h_1 + F_2(p)h_2 + \cdots + F_n(p)h_n, \quad \text{for } (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n. \tag{5.40}$$

Thus, a differential 1–form, ω, defines a vector field $F\colon U \to \mathbb{R}^n$ given by

$$F(p) = (F_1(p), F_2(p), \ldots, F_n(p)), \quad \text{for all } p \in U. \tag{5.41}$$

Conversely, a vector field, $F\colon U \to \mathbb{R}^n$ as in (5.41) gives rise to a differential 1–form, ω, by means of the formula in (5.40). Thus, there is a one–to–one correspondence between differential 1–forms and the space of vector fields on U.

In the final definition of a differentiable 1–form, we will require that the vector field associated to a given form, ω, is at least C^1; in fact, we will require that the field be C^∞, or smooth.

Definition 5.5.2 (Differential 1–Forms in U)**.** A differential 1–form, ω, on U is a (smooth) map $\omega\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ which assigns to each $p \in U$ a linear transformation, $\omega_p\colon \mathbb{R}^n \to \mathbb{R}$, given by

$$\omega_p(h) = F_1(p)h_1 + F_2(p)h_2 + \cdots + F_n(p)h_n,$$

for all $h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$, where the vector field $F = (F_1, F_2, \ldots, F_n)$ is a smooth vector field in U.

Example 5.5.3. Given a smooth function, $f\colon U \to \mathbb{R}$, the vector field $\nabla f\colon U \to \mathbb{R}^n$ gives rise to a differential 1–form denoted by df and defined by

$$df_p(h) = \frac{\partial f}{\partial x_1}(p)\ h_1 + \frac{\partial f}{\partial x_2}(p)\ h_2 + \cdots + \frac{\partial f}{\partial x_n}(p)\ h_n, \tag{5.42}$$

for all $h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

Example 5.5.4. As a special instance of Example 5.5.3, for $j \in \{1, 2, \ldots, n\}$, consider the function $x_j\colon U \to \mathbb{R}$ given by

$$x_j(p) = p_j, \qquad \text{for all } p = (p_1, p_2, \ldots, p_n) \in U.$$

The differential 1–form, dx_j is then given by

$$(dx_j)_p(h) = \frac{\partial x_j}{\partial x_1}(p)\ h_1 + \frac{\partial x_j}{\partial x_2}(p)\ h_2 + \cdots + \frac{\partial x_j}{\partial x_n}(p)\ h_n,$$

for all $h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$; so that,

$$(dx_j)_p(h) = h_j, \qquad \text{for all } h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n. \tag{5.43}$$

Combining the result in (5.42) in Example 5.5.3 with that of (5.43) in Example 5.5.4, we see that for a smooth function $f\colon U \to \mathbb{R}$,

$$df_p(h) = \frac{\partial f}{\partial x_1}(p)\ dx_1(h) + \frac{\partial f}{\partial x_2}(p)\ dx_2(h) + \cdots + \frac{\partial f}{\partial x_n}(p)\ dx_n(h),$$

for all $h \in \mathbb{R}^n$, which can be written as

$$df_p = \frac{\partial f}{\partial x_1}(p)\ dx_1 + \frac{\partial f}{\partial x_2}(p)\ dx_2 + \cdots + \frac{\partial f}{\partial x_n}(p)\ dx_n,$$

for $p \in U$, or

$$df = \frac{\partial f}{\partial x_1}\ dx_1 + \frac{\partial f}{\partial x_2}\ dx_2 + \cdots + \frac{\partial f}{\partial x_n}\ dx_n, \tag{5.44}$$

which gives an interpretation of the differential of a smooth function, f, as a differential 1–form. The expression in (5.44) displays df as a linear combination of the set of differential 1–forms $\{dx_1, dx_2, \ldots, dx_n\}$. In fact, the set

$\{dx_1, dx_2, \ldots, dx_n\}$ is a basis for the space of differential 1–forms. Thus, any differential 1–form, ω, can be written as

$$\omega = F_1\ dx_1 + F_2\ dx_2 + \cdots + F_n\ dx_n, \tag{5.45}$$

where $F = (F_1, F_2, \ldots, F_n)$ is a smooth vector field defined in U.

Differential 1 forms act on oriented, smooth curves, C, by means on integration; we write

$$\omega(C) = \int_C \omega = \int_C F_1\ dx_1 + F_2\ dx_2 + \cdots + F_n\ dx_n.$$

Example 5.5.5 (Action on Directed Line Segments)**.** Given points P_1 and P_2 in $\mathbb{R}^n$, the segment of the line going from P_1 to P_2, denoted by $[P_1, P_2]$, is called the directed line segment from P_1 to P_2. Thus,

$$[P_1, P_2] = \left\{\overrightarrow{OP_1} + t\overrightarrow{P_1P_2} \mid 0 \leqslant t \leqslant 1\right\},$$

where O is the origin in $\mathbb{R}^n$. Thus, $[P_1, P_2]$ is a simple, C^1 curve parametrized by the path

$$\sigma(t) = \overrightarrow{OP_1} + t\overrightarrow{P_1P_2}, \quad 0 \leqslant t \leqslant 1.$$

The action of a differential 1–form, $\omega = F_1\ dx_1 + F_2\ dx_2 + \cdots + F_n\ dx_n$ is then

$$\omega([P_1, P_2]) = \int_{[P_1,P_2]} F \cdot \mathrm{d}\overrightarrow{r}$$

Example 5.5.6. Evaluate the differential 1–form $\omega = yz\mathrm{d}x + xz\mathrm{d}y + xy\mathrm{d}z$ on the directed line segment from the point $P_1(1, 1, 0)$ to the point $P_2(3, 2, 1)$.
Solution: We compute

$$\omega([P_1, P_2]) = \int_{[P_1,P_2]} yz\mathrm{d}x + xz\mathrm{d}y + xy\mathrm{d}z,$$

where $[P_1, P_2]$ is parametrized by

$$\begin{cases} x = 1 + 2t \\ y = 1 + t \\ z = t \end{cases}$$

for $0 \leqslant t \leqslant 1$. Then,

$$\begin{cases} dx = 2\ dt \\ dy = \ dt \\ dz = \ dt, \end{cases}$$

and

$$\begin{aligned}
\int_C yz\mathrm{d}x + xz\mathrm{d}y + xy\mathrm{d}z &= \int_0^1 [2(1+t)t + (1+2t)t + (1+2t)(1+t)]\mathrm{d}t \\
&= \int_0^1 (2t + 2t^2 + t + 2t^2 + 1 + t + 2t + 2t^2)\mathrm{d}t \\
&= \int_0^1 (1 + 6t + 6t^2)\mathrm{d}t \\
&= 6.
\end{aligned}$$

Thus, the differential 1–form, $\omega = yz\mathrm{d}x + xz\mathrm{d}y + xy\mathrm{d}z$ maps the directed line segment $[(1,1,0),(3,2,1)]$ to the real number 6. □

Example 5.5.7. Let $\omega = k_1\ dx_1 + k_2\ dx_2 + \cdots + k_n\ dx_n$, where $k_1, k_2, \ldots, k_n$ are real constants, be a constant differential 1–form. For any two distinct points, P_o and P_1, in $\mathbb{R}^n$, compute $\omega([P_o, P_1])$
Solution: The vector field corresponding to ω is

$$F(x) = (k_1, k_2, \ldots, k_n), \quad \text{for all } x \in \mathbb{R}^n.$$

Compute

$$\begin{aligned}
\omega([P_o, P_1]) &= \int_{[P_o,P_1]} F \cdot \mathrm{d}\overrightarrow{r} \\
&= \int_0^1 F(\sigma(t)) \cdot \sigma'(t)\ \mathrm{d}t,
\end{aligned}$$

where

$$\sigma(t) = \overrightarrow{OP_o} + tv, \quad \text{for } 0 \leqslant t \leqslant 1,$$

where $v = \overrightarrow{P_oP_1}$, the vector that goes from P_o to P_1. Thus,

$$\omega([P_o, P_1]) = K \cdot v,$$

where $K = (k_1, k_2, \ldots, k_n)$ is the constant value of the field F. □

Definition 5.5.8 (Differential 0–Forms)**.** A differential 0–form in $U \subseteq \mathbb{R}^n$ is a C^∞ scalar field $f\colon U \to \mathbb{R}$ which acts on points in U by means of the evaluation the function at those points; that is,

$$f_p = f(p), \quad \text{for all } p \in U.$$

Definition 5.5.9 (Differential of a 0–Form)**.** The differential of a 0 form, f, in U is the differential 1–form given by

$$df = \frac{\partial f}{\partial x_1}\ dx_1 + \frac{\partial f}{\partial x_2}\ dx_2 + \cdots + \frac{\partial f}{\partial x_n}\ dx_n.$$

Example 5.5.10. Given a 0–form f in $\mathbb{R}^n$, evaluate $df([P_1, P_2])$.
Solution: Compute the line integral

$$\begin{aligned}\int_{[P_1,P_2]} \mathrm{d}f &= \int_{[P_1,P_2]} \frac{\partial f}{\partial x_1}\, dx_1 + \frac{\partial f}{\partial x_2}\, dx_2 + \cdots + \frac{\partial f}{\partial x_n}\, dx_n \\ &= \int_0^1 \nabla f(\sigma(t)) \cdot \sigma'(t)\ \mathrm{d}t,\end{aligned}$$

where

$$\sigma(t) = \overrightarrow{OP_1} + t\overrightarrow{P_1P_2}, \quad 0 \leqslant t \leqslant 1.$$

Thus, by the Chain Rule,

$$\begin{aligned}\int_{[P_1,P_2]} df &= \int_0^1 \frac{d}{dt}\left[f(\sigma(t))\right]\ \mathrm{d}t \\ &= f(P_2) - f(P_1).\end{aligned}$$

where we have used the Fundamental Theorem of Calculus. Thus $\mathrm{d}f([P_1, P_2])$ is therefore determined by the values of f at the end–points of the directed line segment $[P_1, P_2]$. □

Example 5.5.11. For two distinct points $P_o(x_o, y_o, z_o)$ and $P_1(x_1, y_1, z_2)$ in $\mathbb{R}^3$, compute $dx([P_o, P_1])$, $dy([P_o, P_1])$ and $dz([P_o, P_1])$.
Solution: Apply the result of the previous example to the function $f(x, y, z) = x$, for all $(x, y, z) \in \mathbb{R}^n$, to obtain that

$$dx([P_o, P_1]) = f(P_1) - f(P_o) = x_1 - x_o.$$

Similarly,

$$dy([P_o, P_1]) = y_1 - y_o,$$

and

$$dz([P_o, P_1]) = z_1 - z_o.$$

□

Next, we define differential 2–forms. Before we give a formal definition, we need to define bilinear, skew–symmetric forms.

Definition 5.5.12 (Bilinear Forms on $\mathbb{R}^n$)**.** A bilinear form on $\mathbb{R}^n$ is a function from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ which is linear in both variables; that is, $B\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is bilinear if

$$B(c_1v_1 + c_2v_2, w) = c_1B(v_1, w) + c_2B(v_2, w),$$

for all $v_1, v_2, w \in \mathbb{R}^n$, and all $c_1, c_2 \in \mathbb{R}$, and

$$B(v, c_1w_1 + c_2w_2) = c_1B(v, w_1) + c_2B(v, w_2),$$

for all $v, w_1, w_2 \in \mathbb{R}^n$, and all $c_1, c_2 \in \mathbb{R}$.

Example 5.5.13. The function $B\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by $B(v,w) = v \cdot w$, the dot–product of v and w, is bilinear.

Definition 5.5.14 (Skew–Symmetric Bilinear Forms on $\mathbb{R}^n$)**.** A bilinear form, $B\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$, is said to be skew–symmetric if

$$B(w,v) = -B(v,w), \quad \text{for all } v, w \in \mathbb{R}^n.$$

Example 5.5.15. For a fixed vector, u, in $\mathbb{R}^3$, define $B\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ by $B(v,w) = u \cdot (v \times w)$, the triple scalar product of u, v and w, for all v and w in $\mathbb{R}^3$. Then, B is skew symmetric.

Example 5.5.16 (Skew–Symmetric Forms in $\mathbb{R}^2$)**.** Let $B\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be a skew–symmetric bilinear form on $\mathbb{R}^2$. We than have that $B(\widehat{i},\widehat{i}) = B(\widehat{j},\widehat{j}) = 0$ and $B(\widehat{j},\widehat{i}) = -B(\widehat{i},\widehat{j})$. Set $\lambda = B(\widehat{i},\widehat{j})$. Then, for any vectors $v = a\widehat{i} + b\widehat{j}$ and $w = c\widehat{i} + d\widehat{j}$ in $\mathbb{R}^2$, we have that

$$\begin{aligned}
B(v,w) &= B(a\widehat{i} + b\widehat{j}, c\widehat{i} + d\widehat{j}) \\
&= ac\ B(\widehat{i},\widehat{i}) + ad\ B(\widehat{i},\widehat{j}) + bc\ B(\widehat{j},\widehat{i}) + bd\ B(\widehat{j},\widehat{j}) \\
&= (ad - bc)\ B(\widehat{i},\widehat{j}) \\
&= \lambda(ad - bc) \\
&= \lambda \det \begin{pmatrix} a & c \\ b & d \end{pmatrix}.
\end{aligned}$$

We have therefore shown that for every skew–symmetric, bilinear form,

$$B\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R},$$

there exists $\lambda \in \mathbb{R}$ such that

$$B(v,w) = \lambda \det[\, v \quad w \,], \quad \text{for all } v, w \in \mathbb{R}^2, \tag{5.46}$$

where $[\, v \quad w \,]$ denotes the 2×2 matrix whose first column are the entries of v, and whose second column are the entries of w.

Example 5.5.17 (Skew–Symmetric Forms in $\mathbb{R}^3$)**.** Let $B\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be a skew–symmetric bilinear form on $\mathbb{R}^3$. Then,

$$B(\widehat{i},\widehat{i}) = B(\widehat{j},\widehat{j}) = B(\widehat{k},\widehat{k}) = 0 \tag{5.47}$$

and

$$\left\{ \begin{aligned}
B(\widehat{j},\widehat{i}) &= -B(\widehat{i},\widehat{j}), \\
B(\widehat{k},\widehat{i}) &= -B(\widehat{i},\widehat{k}), \\
B(\widehat{k},\widehat{j}) &= -B(\widehat{j},\widehat{k}).
\end{aligned} \right. \tag{5.48}$$

Set

$$\begin{cases} \lambda_1 = B(\widehat{j}, \widehat{k}), \\ \lambda_2 = B(\widehat{i}, \widehat{k}), \\ \lambda_3 = B(\widehat{i}, \widehat{j}). \end{cases} \tag{5.49}$$

Then, for any vectors $v = a_1\widehat{i} + a_2\widehat{j} + a_3\widehat{k}$ and $w = b_1\widehat{i} + b_2\widehat{j} + b_3\widehat{k}$ in $\mathbb{R}^3$, we have that

$$\begin{aligned} B(v,w) &= B(a_1\widehat{i} + a_2\widehat{j} + a_3\widehat{k}, b_1\widehat{i} + b_2\widehat{j} + b_3\widehat{k}) \\ &= a_1b_2B(\widehat{i},\widehat{j}) + a_1b_3B(\widehat{i},\widehat{k}) \\ &\quad +a_2b_1B(\widehat{j},\widehat{i}) + a_2b_3B(\widehat{j},\widehat{k}) \\ &\quad\quad +a_3b_1B(\widehat{k},\widehat{i}) + a_3b_2B(\widehat{k},\widehat{j}), \end{aligned}$$

where we have used (5.47). Rearranging terms we obtain

$$\begin{aligned} B(v,w) &= a_2b_3B(\widehat{j},\widehat{k}) + a_3b_2B(\widehat{k},\widehat{j}) \\ &\quad +a_3b_1B(\widehat{k},\widehat{i}) + a_1b_3B(\widehat{i},\widehat{k}) \\ &\quad\quad +a_1b_2B(\widehat{i},\widehat{j}) + a_2b_1B(\widehat{j},\widehat{i}). \end{aligned} \tag{5.50}$$

Next, use (5.48) and (5.49) to rewrite (5.50) as

$$B(v,w) = \lambda_1(a_2b_3 - a_3b_2) - \lambda_2(a_1b_3 - a_3b_1) + \lambda_3(a_1b_2 - a_2b_1),$$

which can be written as

$$B(v,w) = \lambda_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \lambda_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \lambda_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}. \tag{5.51}$$

Recognizing the term on the right–hand side of (5.51) as the triple scalar product of the vector $\Lambda = \lambda_1\widehat{i} + \lambda_2\widehat{j} + \lambda_3\widehat{k}$ and the vectors v and w, we see that we have shown that for every skew–symmetric, bilinear form, $B\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, there exists a vector $\Lambda \in \mathbb{R}^3$ such that

$$B(v,w) = \Lambda \cdot (v \times w), \quad \text{for all } v, w \in \mathbb{R}^3. \tag{5.52}$$

Let $\mathcal{A}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$ denote the space of skew–symmetric bilinear forms in $\mathbb{R}^n$.

Definition 5.5.18 (Differential 2–Forms)**.** Let U denote an open subset of $\mathbb{R}^n$. A differential 2–form in U is a smooth map, $\omega\colon U \to \mathcal{A}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$, which assigns to each $p \in U$, a skew–symmetric, bilinear form, $\omega_p\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

Example 5.5.19 (Differential 2–forms in $\mathbb{R}^2$)**.** Let U denote an open subset of $\mathbb{R}^2$ and $\omega\colon U \to \mathcal{A}(\mathbb{R}^2 \times \mathbb{R}^2, \mathbb{R})$ be a differential 2–form. Then, by Definition 5.5.18, for each $p \in U$, w_p is a skew–symmetric, bilinear form in $\mathbb{R}^2$. By the result in Example 5.5.16 expressed in equation (5.46), for each $p \in U$, there exists a scalar, $f(p)$, such that

$$\omega_p(v, w) = f(p) \det[\, v \quad w \,], \qquad \text{for all } w, w \in \mathbb{R}^2. \tag{5.53}$$

In order to fulfill the smoothness condition in Definition 5.5.18, we require that the scalar field $f : U \to \mathbb{R}$ given in (5.53) be smooth.

Example 5.5.20 (Differential 2–forms in $\mathbb{R}^3$)**.** Let U denote an open subset of $\mathbb{R}^3$ and $\omega : U \to \mathcal{A}(\mathbb{R}^3 \times \mathbb{R}^3, \mathbb{R})$ be a differential 2–form. Then, by Definition 5.5.18, for each $p \in U$, w_p is a skew–symmetric, bilinear form in $\mathbb{R}^3$. Thus, using the representation formula in (5.52) of Example 5.5.17, for each $p \in U$, there exists a vector, $F(p) \in \mathbb{R}^3$, such that

$$\omega_p(v, w) = F(p) \cdot (v \times w), \qquad \text{for all } v, w \in \mathbb{R}^3. \tag{5.54}$$

The smoothness condition in Definition 5.5.18 requires that the vector field $F : U \to \mathbb{R}^3$ given in (5.54) be smooth.

Definition 5.5.21 (Wedge Product of 1–Forms)**.** Given two differential 1–forms, ω and η, in some open subset, U, of $\mathbb{R}^n$, we can define a differential 2–form in U, denoted by $\omega \wedge \eta$, as follows

$$(\omega \wedge \eta)_p(v, w) = \omega_p(v)\eta_p(w) - \omega_p(w)\eta_p(v), \qquad \text{for } p \in U, \text{ and } v, w \in \mathbb{R}^n. \tag{5.55}$$

To see that the expression for $(\omega \wedge \eta)_p$ given in (5.55) does define a bilinear form, compute

$$\begin{aligned}
(\omega \wedge \eta)_p(c_1v_1 + c_2v_2, w) &= \omega_p(c_1v_1 + c_2v_2)\eta_p(w) - \omega_p(w)\eta_p(c_1v_1 + c_2v_2) \\
&= [c_1\omega_p(v_1) + c_2\omega_p(v_2)]\eta_p(w) \\
&\quad -\omega_p(w)[c_1\eta_p(v_1) + c_2\eta_p(v_2)] \\
&= c_1\omega_p(v_1)\eta_p(w) + c_2\omega_p(v_2)\eta_p(w) \\
&\quad -c_1\omega_p(w)\eta_p(v_1) - c_2\omega_p(w)\eta_p(v_2),
\end{aligned}$$

so that

$$\begin{aligned}
(\omega \wedge \eta)_p(c_1v_1 + c_2v_2, w) &= c_1[\omega_p(v_1)\eta_p(w) - \omega_p(w)\eta_p(v_1)] \\
&\quad +c_2[\omega_p(v_2)\eta_p(w) - \omega_p(w)\eta_p(v_2)] \\
&= c_1(\omega \wedge \eta)_p(v_1, w) + c_2(\omega \wedge \eta)_p(v_2, w),
\end{aligned}$$

for all $v_1, v_2, w \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$. A similar calculation shows that

$$(\omega \wedge \eta)_p(v, c_1w_1 + c_2w_2) = c_1(\omega \wedge \eta)_p(v, w_1) + c_2(\omega \wedge \eta)_p(v, w_2),$$

for all $v, w_1, w_2 \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$.

Similarly, to see that $(\omega \wedge \eta)_p \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is skew–symmetric, compute

$$\begin{aligned}
(\omega \wedge \eta)_p(w, v) &= \omega_p(w)\eta_p(v) - \omega_p(v)\eta_p(w) \\
&= -[\omega_p(v)\eta_p(w) - \omega_p(w)\eta_p(v)] \\
&= -(\omega \wedge \eta)_p(v, w)
\end{aligned}$$

Proposition 5.5.22 (Properties of the Wedge Product)**.** Let ω, η and γ denote 1–forms in U, an open subset of $\mathbb{R}^n$. Then,

(i) $\omega \wedge \eta = -\eta \wedge \omega$;

(ii) $\omega \wedge \omega = 0$, where 0 denotes the bilinear form that maps every pair of vectors to 0;

(iii) $(\omega + \eta) \wedge \gamma = \omega \wedge \gamma + \eta \wedge \gamma$;

(iv) $\omega \wedge (\eta + \gamma) = \omega \wedge \eta + \omega \wedge \gamma$.

Example 5.5.23. Let $P_o(x_o, y_o)$, $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ denote three non–collinear points in the xy–plane. Put

$$v = \overrightarrow{P_oP_1} = (x_1 - x_o)\widehat{i} + (y_1 - y_o)\widehat{j}$$

and

$$w = \overrightarrow{P_oP_2} = (x_2 - x_o)\widehat{i} + (y_2 - y_o)\widehat{j}.$$

Then, according to (5.55) in Definition 5.5.21,

$$\begin{aligned}
(dx \wedge dy)(v, w) &= dx(v)\ dy(w) - dx(w)\ dy(v) \\
&= (x_1 - x_o)(y_2 - y_o) - (x_2 - x_o)(y_1 - y_o),
\end{aligned}$$

where we have used the result of Example 5.5.11. We then have that

$$(dx \wedge dy)(v, w) = \begin{vmatrix} x_1 - x_o & x_2 - x_o \\ y_1 - y_o & y_2 - y_o \end{vmatrix},$$

which is the determinant of the 2×2 matrix, $[v\ \ w]$, whose columns are the vectors v and w. In other words,

$$(dx \wedge dy)(v, w) = \det[v \quad w]. \tag{5.56}$$

We have therefore shown that the $(dx \wedge dy)(v, w)$ gives the signed area of the parallelogram determined by the vectors v and w.

Example 5.5.24. Let $P_o(x_o, y_o, z_o)$, $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ denote three non–collinear points in $\mathbb{R}^3$. Put

$$v = \overrightarrow{P_oP_1} = (x_1 - x_o)\widehat{i} + (y_1 - y_o)\widehat{j} + (z_1 - z_o)\widehat{k}$$

and

$$w = \overrightarrow{P_oP_2} = (x_2 - x_o)\widehat{i} + (y_2 - y_o)\widehat{j} + (z_1 - z_o)\widehat{k}.$$

Then, as in Example 5.5.23, we compute

$$(dx \wedge dy)(v, w) = \begin{vmatrix} x_1 - x_o & x_2 - x_o \\ y_1 - y_o & y_2 - y_o \end{vmatrix},$$

which we can also write as

$$(dx \wedge dy)(v, w) = \begin{vmatrix} x_1 - x_o & y_1 - y_o \\ x_2 - x_o & y_2 - y_o \end{vmatrix}, \tag{5.57}$$

Similarly, we compute

$$(dy \wedge dz)(v, w) = \begin{vmatrix} y_1 - y_o & z_1 - z_o \\ y_2 - y_o & z_2 - z_o \end{vmatrix}, \tag{5.58}$$

and

$$(dz \wedge dx)(v, w) = \begin{vmatrix} z_1 - z_o & x_1 - x_o \\ z_2 - z_o & x_2 - x_o \end{vmatrix},$$

or

$$(dz \wedge dx)(v, w) = -\begin{vmatrix} x_1 - x_o & z_1 - z_o \\ x_2 - x_o & z_2 - z_o \end{vmatrix}. \tag{5.59}$$

We recognize in (5.58), (5.59) and (5.57) the components of the cross product of the vectors v and w,

$$v \times w = \begin{vmatrix} y_1 - y_o & z_1 - z_o \\ y_2 - y_o & z_2 - z_o \end{vmatrix} \widehat{i} - \begin{vmatrix} x_1 - x_o & z_1 - z_o \\ x_2 - x_o & z_2 - z_o \end{vmatrix} \widehat{j} + \begin{vmatrix} x_1 - x_o & y_1 - y_o \\ x_2 - x_o & y_2 - y_o \end{vmatrix} \widehat{k}.$$

We can therefore write

$$(dy \wedge dz)(v, w) = (v \times w) \cdot \widehat{i}, \tag{5.60}$$

$$(dz \wedge dx)(v, w) = (v \times w) \cdot \widehat{j}, \tag{5.61}$$

and

$$(dx \wedge dy)(v, w) = (v \times w) \cdot \widehat{k}. \tag{5.62}$$

Differential 0–forms act on points. Differential 1–forms act on directed line segments and, more generally, on oriented curves. We will next see how to define the action of differential 2–forms on *oriented triangles.* We first define oriented triangles in the plane.

Definition 5.5.25 (Oriented Triangles in $\mathbb{R}^2$)**.** Given three non–collinear points P_o, P_1 and P_2 in the plane, we denote by $T = [P_o, P_1, P_2]$ the triangle with vertices P_o, P_1 and P_2. T is a 2–dimensional object consisting of the simple curve generated by the directed line segments $[P_o, P_1]$, $[P_1, P_2]$, and $[P_2, P_o]$ as well as the interior of the curve. If the curve is traversed in the counterclockwise sense, then T has positive orientation; if the curve is traversed in the clockwise sense, then T has negative orientation.

Definition 5.5.26 (Action of a Differential 2–Form on an Oriented Triangle in $\mathbb{R}^2$)**.** The differential 2–form, $\mathrm{d}x \wedge \mathrm{d}y$, acts on an oriented triangle T by evaluating its area, if T has a positive orientation, and the negative of the area if T has a negative orientation:

$$\mathrm{d}x \wedge \mathrm{d}y(T) = \pm \text{ area}(T).$$

We denote this by

$$\int_T \mathrm{d}x \wedge \mathrm{d}y = \text{signed area of } \ T. \tag{5.63}$$

According to the formula (5.56) in Example 5.5.23, the expression in (5.63) may also be obtained by computing

$$\int_T \mathrm{d}x \wedge \mathrm{d}y = \frac{1}{2}(dx \wedge dy)(\overrightarrow{P_oP_1}, \overrightarrow{P_oP_2}), \tag{5.64}$$

since $(dx \wedge dy)(\overrightarrow{P_oP_1}, \overrightarrow{P_oP_1})$ gives the signed area of the parallelogram generated by the vectors $\overrightarrow{P_oP_1}$ and $\overrightarrow{P_oP_2}$. By embedding the vectors $\overrightarrow{P_oP_1}$ and $\overrightarrow{P_oP_2}$ in the xy–coordinate plane in $\mathbb{R}^3$, we may also use the formula in (5.62) to obtain that

$$\int_{[P_oP_1P_2]} \mathrm{d}x \wedge \mathrm{d}y = \frac{1}{2}(\overrightarrow{P_oP_1} \times \overrightarrow{P_oP_2}) \cdot \widehat{k}. \tag{5.65}$$

Example 5.5.27. Let $P_o(0, 0)$, $P_1(1, 2)$ and $P_2(2, 1)$ and let $T = [P_o, P_1, P_2]$ denote the oriented triangle generated by those points. Evaluate $\displaystyle\int_T \mathrm{d}x \wedge \mathrm{d}y$.

Solution: Embed the points P_o, P_1 and P_2 in $\mathbb{R}^3$ by appending 0 as the last coordinate, and let

$$v = \overrightarrow{P_oP_1} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad \text{and} \quad w = \overrightarrow{P_oP_2} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Then $\displaystyle\int_T \mathrm{d}x \wedge \mathrm{d}y$ is the component of the vector $\dfrac{1}{2} v \times w$ along the direction of $\widehat{k}$; that is,

$$\int_T \mathrm{d}x \wedge \mathrm{d}y = \frac{1}{2}(v \times w) \cdot \widehat{k},$$

where

$$v \times w = \begin{vmatrix} \widehat{i} & \widehat{j} & \widehat{k} \\ 1 & 2 & 0 \\ 2 & 1 & 0 \end{vmatrix} = (1-4)\,\widehat{k} = -3\,\widehat{k}.$$

It then follows that

$$\int_T \mathrm{d}x \wedge \mathrm{d}y = -\frac{3}{2}.$$

We see that $\frac{1}{2}(v \times w) \cdot \widehat{k}$ gives the appropriate sign for the $\mathrm{d}x\mathrm{d}y(T)$ since in this case T has negative orientation. □

In general, for non–collinear points P_o, P_1 and P_2 in $\mathbb{R}^3$, the value of $\mathrm{d}x \wedge \mathrm{d}y$ on $T = [P_o, P_1, P_2]$ is obtained by the formula in (5.65); namely,

$$\mathrm{d}x \wedge \mathrm{d}y(T) = \int_T dx \wedge dy = \frac{1}{2}(v \times w) \cdot \widehat{k},$$

where

$$v = \overrightarrow{P_oP_1} \quad \text{and} \quad w = \overrightarrow{P_oP_2}.$$

This gives the signed area of the orthogonal projection of the triangle T onto the xy–plane. Similarly, using the formulas in (5.60) and (5.61), we obtain the values of the differential 2–forms $\mathrm{d}y \wedge \mathrm{d}z$ and $\mathrm{d}z \wedge \mathrm{d}x$ on the oriented triangle $T = [P_oP_1P_2]$:

$$\mathrm{d}y \wedge \mathrm{d}z(T) = \int_T \mathrm{d}y \wedge \mathrm{d}z = \frac{1}{2}(v \times w) \cdot \widehat{i},$$

and

$$\mathrm{d}z \wedge \mathrm{d}x(T) = \int_T \mathrm{d}z \wedge \mathrm{d}x = \frac{1}{2}(v \times w) \cdot \widehat{j}.$$

Example 5.5.28. Evaluate $\int_T \mathrm{d}y \wedge \mathrm{d}z$, $\int_T \mathrm{d}z \wedge \mathrm{d}x$, and $\int_T \mathrm{d}x \wedge \mathrm{d}y$, where $T = [P_o, P_1, P_2]$ for the points

$$P_o(-1,1,2),\ P_1(3,2,1) \text{ and } P_2(4,7,1).$$

Solution: Set

$$v = \overrightarrow{P_oP_1} = \begin{pmatrix} 4 \\ 1 \\ -1 \end{pmatrix} \quad \text{and} \quad w = \overrightarrow{P_oP_2} = \begin{pmatrix} 5 \\ 6 \\ -1 \end{pmatrix},$$

and compute

$$v \times w = \begin{vmatrix} \widehat{i} & \widehat{j} & \widehat{k} \\ 4 & 1 & -1 \\ 5 & 6 & -2 \end{vmatrix} = (-2+6)\,\widehat{i} - (-8+5)\,\widehat{j} + (24-5)\,\widehat{k} = 4\,\widehat{i} + 3\,\widehat{j} + 19\,\widehat{k}.$$

It then follows that

$$\int_T \mathrm{d}y \wedge \mathrm{d}z = 2,$$

$$\int_T \mathrm{d}z \wedge \mathrm{d}x = \frac{3}{2}$$

and

$$\int_T \mathrm{d}x \wedge \mathrm{d}y = \frac{19}{2}.$$

□

We end this section by showing that, in $\mathbb{R}^3$, the space of differential 2–forms in an open subset U of $\mathbb{R}^3$ is generated by the set

$$\{dy \wedge dz,\ dz \wedge dx,\ dx \wedge dy\}, \tag{5.66}$$

in the sense that, for every differential 2–from, ω, in U, there exists a smooth vector field $F\colon U \to \mathbb{R}^3$,

$$F(x, y, z) = F_1(x, y, z)\ \widehat{i} + F_2(x, y, z)\ \widehat{j} + F_3(x, y, z)\ \widehat{k},$$

such that

$$\omega_p = F_1(p)\ dy \wedge dz + F_2(p)\ dz \wedge dx + F_3(p)\ dx \wedge dy, \qquad \text{for all } p \in U.$$

Let $\omega\colon U \to \mathcal{A}(\mathbb{R}^3 \times \mathbb{R}^3, \mathbb{R})$ be a differential 2–form in an open subset, U, of $\mathbb{R}^3$. We consider vectors

$$v = a_1\ \widehat{i} + a_2\ \widehat{j} + a_3\ \widehat{k}$$

and

$$w = b_1\ \widehat{i} + b_2\ \widehat{j} + b_3\ \widehat{k}$$

in $\mathbb{R}^3$. For each $p \in U$, we compute

$$\begin{aligned} \omega_p(v, w) &= \omega_p(a_1\ \widehat{i} + a_2\ \widehat{j} + a_3, b_1\ \widehat{i} + b_2\ \widehat{j} + b_3\ \widehat{k}) \\ &= a_1 b_2 \omega_p(\widehat{i}, \widehat{j}) + a_1 b_3 \omega_p(\widehat{i}, \widehat{k}) \\ &\quad\ a_2 b_1 \omega_p(\widehat{j}, \widehat{i}) + a_2 b_3 \omega_p(\widehat{j}, \widehat{k}) + \\ &\quad\quad a_3 b_1 \omega_p(\widehat{k}, \widehat{i}) + a_3 b_2 \omega_p(\widehat{k}, \widehat{j}), \end{aligned} \tag{5.67}$$

where we have used the fact that

$$\omega_p(\widehat{i}, \widehat{i}) = \omega_p(\widehat{j}, \widehat{j}) = \omega_p(\widehat{k}, \widehat{k}) = 0,$$

which follows from the skew–symmetry of the form $\omega_p\colon \mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}$. Using the skew–symmetry again, we obtain from (5.67) that

$$\begin{aligned}\omega_p(v,w) &= a_2b_3\omega_p(\widehat{j},\widehat{k}) + a_3b_2\omega_p(\widehat{k},\widehat{j})\\ &\quad +a_1b_3\omega_p(\widehat{i},\widehat{k}) + a_3b_1\omega_p(\widehat{k},\widehat{i})\\ &\quad +a_1b_2\omega_p(\widehat{i},\widehat{j}) + a_2b_1\omega_p(\widehat{j},\widehat{i})\\ &= \omega_p(\widehat{j},\widehat{k})(a_2b_3-a_3b_2)\\ &\quad +\omega_p(\widehat{k},\widehat{i})(a_3b_1-a_1b_3)\\ &\quad +\omega_p(\widehat{i},\widehat{j})(a_1b_2-a_2b_1).\end{aligned} \tag{5.68}$$

Next, use Definition 5.5.21 to compute

$$\begin{aligned}dy\wedge dz(v,w) &= dy(v)dz(w)-dy(w)dz(v)\\ &= a_2b_3-b_2a_3,\end{aligned} \tag{5.69}$$

$$\begin{aligned}dz\wedge dx(v,w) &= dz(v)dx(w)-dz(w)dx(v)\\ &= a_3b_1-b_3a_1,\end{aligned} \tag{5.70}$$

and

$$\begin{aligned}dx\wedge dy(v,w) &= dx(v)dy(w)-dx(w)dy(v)\\ &= a_1b_2-b_1a_2.\end{aligned} \tag{5.71}$$

Substituting the expressions obtained in (5.69)–(5.71) into the last expression on the right–hand side of (5.68) yields

$$\begin{aligned}\omega_p(v,w) &= \omega_p(\widehat{j},\widehat{k})\ dy\wedge dz(v,w)\\ &\quad +\omega_p(\widehat{k},\widehat{i})\ dz\wedge dx(v,w)\\ &\quad +\omega_p(\widehat{i},\widehat{j})\ dx\wedge dy(v,w),\end{aligned}$$

from which we get that

$$\omega_p = \omega_p(\widehat{j},\widehat{k})\ dy\wedge dz + \omega_p(\widehat{k},\widehat{i})\ dz\wedge dx + \omega_p(\widehat{i},\widehat{j})\ dx\wedge dy. \tag{5.72}$$

Setting

$$\begin{aligned}F_1(p) &= \omega_p(\widehat{j},\widehat{k}),\\ F_2(p) &= \omega_p(\widehat{k},\widehat{i}),\\ F_3(p) &= \omega_p(\widehat{i},\widehat{j}),\end{aligned}$$

we see from (5.72) that

$$\omega_p = F_1(p)\ dy\wedge dz + F_2(p)\ dz\wedge dx + F_3(p)\ dx\wedge dy, \tag{5.73}$$

which shows that every differential 2–form in $\mathbb{R}^3$ is in the span of the set in (5.66).

To show that the representation in (5.73) is unique, assume that

$$F_1(p)\ dy \wedge dz + F_2(p)\ dz \wedge dx + F_3(p)\ dx \wedge dy = 0, \tag{5.74}$$

the differential 2–form that maps every pair of vectors $(v, w) \in \mathbb{R}^3 \times \mathbb{R}^3$ to the real number 0. Then, applying the form in (5.74) to the pair $(\widehat{j}, \widehat{k})$ we obtain that

$$F_1(p)\ dy \wedge dz(\widehat{j}, \widehat{k}) + F_2(p)\ dz \wedge dx(\widehat{j}, \widehat{k}) + F_3(p)\ dx \wedge dy(\widehat{j}, \widehat{k}) = 0,$$

which implies that

$$F_1(p) = 0,$$

in view of the results of the calculations in (5.69)–(5.71). Similarly, applying (5.74) to $(\widehat{k}, \widehat{i})$ and $(\widehat{i}, \widehat{j})$, successively, leads to

$$F_2(p) = 0 \text{ and } F_3(p) = 0,$$

respectively. Thus, the set in (5.66) is also linearly independent; hence, the representation in (5.73) is unique.

5.6 Calculus of Differential Forms

Proposition 5.5.22 on page 107 in these notes lists some of the algebraic properties of the wedge product of differential 1–forms defined in Definition 5.5.21. Properties (i) and (ii) in Proposition 5.5.22 can be verified for the differential 1–forms dx and dy directly from the definition and the results in Example (5.5.11). In fact, for non–collinear points $P_o(x_o, y_o)$, $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ in $\mathbb{R}^2$, using Definition 5.5.21 we compute

$$\begin{aligned}
(dy \wedge dx)(\overrightarrow{P_oP_1}, \overrightarrow{P_oP_2}) &= dy(\overrightarrow{P_oP_1})dx(\overrightarrow{P_oP_2}) - dy(\overrightarrow{P_oP_2})dx(\overrightarrow{P_oP_1}) \\
&= -[dx(\overrightarrow{P_oP_1})dy(\overrightarrow{P_oP_2}) - dx(\overrightarrow{P_oP_2})dy(\overrightarrow{P_oP_1})] \\
&= -(dx \wedge dy)(\overrightarrow{P_oP_1}, \overrightarrow{P_oP_2}).
\end{aligned}$$

Consequently,

$$dy \wedge dx = -dx \wedge dy. \tag{5.75}$$

From this we can deduce that

$$dx \wedge dx = 0. \tag{5.76}$$

Thus, the wedge product of differential 1–forms is anti–symmetric.

We can also multiply 0–forms and 1–forms; for instance, the differential 1–form,

$$P(x, y)\ dx,$$

where $P\colon U \to \mathbb{R}$ is a smooth function on an open subset, U, of $\mathbb{R}^2$, is the product of a 0–form and a differential 1–form.

The differential 1–form, $P\ dx$, can be added to another 1–form, $Q\ dy$, to obtain the differential 1–form for example,

$$P\ dx + Q\ dy, \tag{5.77}$$

where P and Q are smooth scalar fields. We can also multiply the differential 1–from in (5.77) by the 1–form dx:

$$(P\ dx + Q\ dy) \wedge dx = P\ dx \wedge dx + Q\ dy \wedge \mathrm{dx} = -Q\ dx \wedge dy,$$

where we have used (5.75) and (5.76).

We have already seen how to obtain a differential 1–form from a differential 0–form, f, by computing the differential of f:

$$df = \frac{\partial f}{\partial x}\ dx + \frac{\partial f}{\partial y}\ dy + \frac{\partial f}{\partial z}\ dz.$$

This defines an operator, d, from the class of 0–forms to the class of 1–forms. This operator, d, also acts on the 1–form

$$\omega = P(x,y)\ dx + Q(x,y)\ dy$$

in $\mathbb{R}^2$, where P and Q are smooth scalar fields, as follows:

$$\begin{aligned}
\mathrm{d}\omega &= (dP) \wedge dx + (dQ) \wedge dy \\
&= \left(\frac{\partial P}{\partial x}\ dx + \frac{\partial P}{\partial y}\ dy\right) \wedge dx + \left(\frac{\partial Q}{\partial x}\ dx + \frac{\partial Q}{\partial y} \wedge \mathrm{dy}\right)\ dy \\
&= \frac{\partial P}{\partial x}\ dx \wedge dx + \frac{\partial P}{\partial y}\ dy \wedge dx + \frac{\partial Q}{\partial x}\ dx \wedge dy + \frac{\partial Q}{\partial y}\ dy \wedge dy \\
&= \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ dx \wedge dy,
\end{aligned}$$

where we have used (5.75) and (5.76). Thus, the differential of the 1–form

$$\omega = P\ dx + Q\ dy$$

in $\mathbb{R}^2$ is the differential 2–form

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ dx \wedge dy.$$

Thus, the differential, $d\omega$, of the 1–form, ω, acts on oriented triangles,

$$T = [P_1, P_2, P_3],$$

in $\mathbb{R}^2$. By analogy with what happens to the differential, df, of a 0–form, f, when it is integrated over a directed line segment, we expect that

$$\int_T d\omega$$

is completely determined by the action of ω on the boundary, ∂T, of T, which is a simple, closed curve made up of the directed line segments $[P_1, P_2]$, $[P_2, P_3]$ and $[P_3, P_1]$. More specifically, if T has positive orientation, we expect that

$$\int_T d\omega = \int_{\partial T} \omega. \tag{5.78}$$

This is the Fundamental Theorem of Calculus in two–dimensions for the special case of oriented triangles, and we will prove it in the following sections. We will first see how to evaluate the 2–form $d\omega$ on oriented triangles.

5.7 Evaluating 2–forms: Double Integrals

Given an oriented triangle, $T = [P_1, P_2, P_3]$, in the xy–plane and with positive orientation, we would like to evaluate the 2–form $f(x, y)\mathrm{d}x \wedge \mathrm{d}y$ on T, for a given continuous scalar field f; that is, we would like to evaluate

$$\int_T f(x, y)\ dx \wedge dy.$$

For the case in which T has a positive orientation, we will denote the value of $\displaystyle\int_T f(x, y)\mathrm{d}x \wedge \mathrm{d}y$ by

$$\int_T f(x, y)\ dxdy \tag{5.79}$$

and call it the **double integral** of f over T. In this sense, we then have that

$$\int_T f(x, y)\mathrm{d}y \wedge \mathrm{d}x = -\int_T f(x, y)\ dxdy,$$

for the case in which T has a negative orientation.

We first see how to evaluate the double integral in (5.79) for the case in which T is the *unit triangle* $U = [(0, 0), (1, 0), (0, 1)]$ in Figure 5.7.6, which is oriented in the positive direction. We evaluate $\displaystyle\int_T f(x, y)\mathrm{d}x\mathrm{d}y$ by computing two *iterated integrals* as follows

$$\int_U f(x, y)\mathrm{d}x\mathrm{d}y = \int_0^1 \left\{\int_0^{1-x} f(x, y)\ \mathrm{d}y\right\}\ \mathrm{d}x. \tag{5.80}$$

Observe that the "inside" integral,

$$\int_0^{1-x} f(x, y)\ \mathrm{d}y,$$

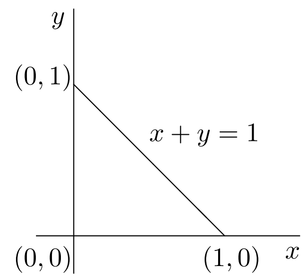

Figure 5.7.6: Unit Triangle U

yields a function of x for $x \in [0,1]$; call this function g; that is,

$$g(x) = \int_0^{1-x} f(x,y)\ \mathrm{d}y \quad \text{for all} \ \ x \in [0,1];$$

Then,

$$\int_U f(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1 g(x)\ \mathrm{d}x.$$

We could also do the integration with respect to x first, then integrate with respect to y:

$$\int_U f(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1 \left\{ \int_0^{1-y} f(x,y)\ \mathrm{d}x \right\}\ \mathrm{d}y. \tag{5.81}$$

In this case the inner integral yields a function of y which can then be integrated from 0 to 1.

Observe that the iterated integrals in (5.80) and (5.81) correspond to alternate descriptions of U as

$$U = \{(x,y) \in \mathbb{R}^2 \mid 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1 - x\}$$

or

$$U = \{(x,y) \in \mathbb{R}^2 \mid 0 \leqslant x \leqslant 1 - y,\ 0 \leqslant y \leqslant 1\},$$

respectively.

The fact that the iterated integrals in equations (5.80) and (5.81) yield the same value, at least for the case in which f is continuous on a region containing U, is a special case of a theorem in Advanced Calculus or Real Analysis known as Fubini's Theorem.

Example 5.7.1. *Evaluate* $\int_U x\ dxdy$.

Solution: Using the iterated integral in (5.80) we get

$$\begin{aligned}\int_U x\ \mathrm{d}x\mathrm{d}y &= \int_0^1 \left\{\int_0^{1-x} x\ \mathrm{d}y\right\} \mathrm{d}x \\ &= \int_0^1 \Big[xy\Big]_0^{1-x} \mathrm{d}x \\ &= \int_0^1 x(1-x)\ \mathrm{d}x \\ &= \int_0^1 (x-x^2)\ \mathrm{d}x \\ &= \frac{1}{6}.\end{aligned}$$

We could have also used the iterated integral in (5.81):

$$\begin{aligned}\int_U x\ \mathrm{d}x\mathrm{d}y &= \int_0^1 \left\{\int_0^{1-y} x\ \mathrm{d}x\right\} \mathrm{d}y \\ &= \int_0^1 \left[\frac{1}{2}x^2\right]_0^{1-y} \mathrm{d}y \\ &= \frac{1}{2}\int_0^1 (1-y)^2\ \mathrm{d}y \\ &= -\frac{1}{2}\int_1^0 u^2\ \mathrm{d}x \\ &= \frac{1}{2}\int_0^1 u^2\ \mathrm{d}u \\ &= \frac{1}{6}.\end{aligned}$$

□

Iterated integrals can be used to evaluate double–integrals over plane regions other than triangles. For instance, suppose a region, R, is bounded by the vertical lines $x = a$ and $x = b$, where $a < b$, and by the graphs of two functions $g_1(x)$ and $g_2(x)$, where $g_1(x) \leqslant g_2(x)$ for $a \leqslant x \leqslant b$; that is

$$R = \{(x, y) \in \mathbb{R}^2 \mid g_1(x) \leqslant y \leqslant g_2(x), a \leqslant x \leqslant b\};$$

then,

$$\int_R f(x, y)\ \mathrm{d}x\mathrm{d}y = \int_a^b \left\{\int_{g_1(x)}^{g_2(x)} f(x, y)\ \mathrm{d}y\right\} \mathrm{d}x.$$

Example 5.7.2. *Let R denote the region in the first quadrant bounded by the unit circle, $x^2+y^2=1$; that is, R is the quarter unit disc. Evaluate $\int_R y\ dx\,dy$.*

Solution: In this case, the region R is described by

$$R=\{(x,y)\in\mathbb{R}^2 \mid 0\leqslant y\leqslant\sqrt{1-x^2}, 0\leqslant x\leqslant 1\},$$

so that

$$\begin{aligned}\int_R y\ \mathrm{d}x\mathrm{d}y &= \int_0^1\int_0^{\sqrt{1-x^2}} y\ \mathrm{d}y\mathrm{d}x\\ &= \int_0^1 \frac{1}{2}y^2\Big|_0^{\sqrt{1-x^2}}\ \mathrm{d}x\\ &= \frac{1}{2}\int_0^1(1-x^2)\ \mathrm{d}x\\ &= \frac{1}{3}\end{aligned}$$

□

Alternatively, the region R can be described by

$$R=\{(x,y)\in\mathbb{R}^2 \mid h_1(y)\leqslant x\leqslant h_2(y), c\leqslant y\leqslant d\},$$

where $h_1(y)\leqslant h_2(y)$ for $c\leqslant y\leqslant d$. In this case,

$$\int_R f(x,y)\ \mathrm{d}x\mathrm{d}y=\int_c^d\left\{\int_{h_1(y)}^{h_2(y)} f(x,y)\ \mathrm{d}x\right\}\mathrm{d}y.$$

Example 5.7.3. *Identify the region, R, in the plane in which the following iterated integral*

$$\int_0^1\int_y^1\frac{1}{\sqrt{1+x^2}}\,dx\,dy$$

is computed. Change the order of integration and then evaluate the double integral

$$\int_R\frac{1}{\sqrt{1+x^2}}\ dx\,dy.$$

Solution: In this case, the region R is

$$R=\{(x,y)\in\mathbb{R}^2 \mid y\leqslant x\leqslant 1, 1\leqslant y\leqslant 1\}.$$

This is also represented by

$$R=\{(x,y)\in\mathbb{R}^2 \mid 0\leqslant x\leqslant 1, 1\leqslant y\leqslant x\};$$

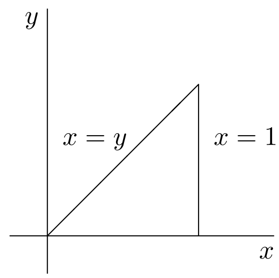

Figure 5.7.7: Region R in example 5.7.3

see picture in Figure 5.7.7. It then follows that

$$\begin{aligned}
\int_R \frac{1}{\sqrt{1+x^2}}\ \mathrm{d}x\mathrm{d}y &= \int_0^1 \int_0^x \frac{1}{\sqrt{1+x^2}}\ \mathrm{d}y\mathrm{d}x \\
&= \int_0^1 \frac{1}{\sqrt{1+x^2}}\ y\Big|_0^x\ \mathrm{d}x \\
&= \int_0^1 \frac{1}{\sqrt{1+x^2}}\ x\ \mathrm{d}x \\
&= \int_0^1 \frac{1}{2\sqrt{1+x^2}}\ 2x\ \mathrm{d}x \\
&= \int_1^2 \frac{1}{2\sqrt{u}}\ \mathrm{d}u \\
&= \sqrt{u}\Big|_1^2 \\
&= \sqrt{2}-1.
\end{aligned}$$

□

If R is a bounded region of $\mathbb{R}^2$, and $f(x,y) \geqslant 0$ for all $(x,y) \in R$, then

$$\int_R f(x,y)\ \mathrm{d}x\mathrm{d}y$$

gives the volume of the three dimensional solid that lies below the graph of the surface $z = f(x,y)$ and above the region R.

Example 5.7.4. *Let a, b and c be positive real numbers. Compute the volume of the tetrahedron whose base is the triangle $T = [(0,0),(a,0),(0,b)]$ and which* *lies below the plane*

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1.$$

Solution: We need to evaluate $\displaystyle\int_T z\ \mathrm{d}x\mathrm{d}y$, where

$$z = c\left(1 - \frac{x}{a} - \frac{y}{b}\right).$$

Then,

$$\begin{aligned}
\int_T z\ \mathrm{d}x\mathrm{d}y &= c\int_T \left(1 - \frac{x}{a} - \frac{y}{b}\right)\mathrm{d}x\mathrm{d}y \\
&= c\int_0^a \int_0^{b(1-x/a)} \left(1 - \frac{x}{a} - \frac{y}{b}\right)\mathrm{d}y\mathrm{d}x \\
&= c\int_0^a \left[y - \frac{xy}{a} - \frac{y^2}{2b}\right]_0^{b(1-x/a)} \mathrm{d}x \\
&= c\int_0^a \left[b\left(1-\frac{x}{a}\right) - \frac{x}{a}b\left(1-\frac{x}{a}\right) - \frac{1}{2b}b^2\left(1-\frac{x}{a}\right)^2\right] \mathrm{d}x \\
&= bc\int_0^a \left(\frac{1}{2} - \frac{x}{a} + \frac{x^2}{2a^2}\right) \mathrm{d}x \\
&= bc\left[\frac{a}{2} - \frac{a}{2} + \frac{a}{6}\right] \\
&= \frac{abc}{6}.
\end{aligned}$$

□

5.8 Fundamental Theorem of Calculus in $\mathbb{R}^2$

In this section we prove the Fundamental Theorem of Calculus in two dimensions expressed in (5.78). More precisely, we have the following theorem:

Proposition 5.8.1 (Fundamental Theorem of Calculus for Oriented Triangles in $\mathbb{R}^2$)**.** *Let ω be a C^1 1–form defined on some plane region containing a positively oriented triangle T. Then,*

$$\int_T d\omega = \int_{\partial T} \omega. \tag{5.82}$$

More specifically, let $\omega = P\mathrm{d}x + Q\mathrm{d}y$ be a differential 1–form for which P and Q are C^1 scalar fields defined in some region containing a positively oriented triangle T. Then

$$\int_T \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}x\mathrm{d}y = \int_{\partial T} P\mathrm{d}x + Q\mathrm{d}y. \tag{5.83}$$

This version of the Fundamental Theorem of Calculus is known as **Green's Theorem**.

Proof of Green's Theorem for the Unit Triangle in $\mathbb{R}^2$. We shall first prove Proposition 5.8.1 for the unit triangle $U = [(0,0),(1,0),(0,1)] = [P_1, P_2, P_3]$:

$$\int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}x\mathrm{d}y = \int_{\partial U} P\mathrm{d}x + Q\mathrm{d}y, \tag{5.84}$$

where P and Q are C^1 scalar fields defined on some region containing U, and ∂U is made up of the directed line segments $[P_1, P_2]$, $[P_2, P_3]$ and $[P_3, P_1]$ traversed in the counterclockwise sense.

We will prove separately that

$$\int_U \frac{\partial Q}{\partial x} \, \mathrm{d}x\mathrm{d}y = \int_{\partial U} Q\mathrm{d}y, \tag{5.85}$$

and

$$-\int_U \frac{\partial P}{\partial y} \, \mathrm{d}x\mathrm{d}y = \int_{\partial U} P\mathrm{d}x. \tag{5.86}$$

Together, (5.85) and (5.86) will establish (5.84).

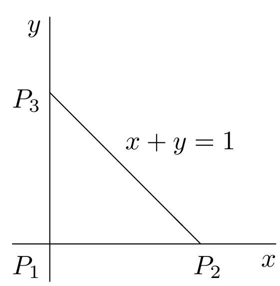

Figure 5.8.8: Unit Triangle U

Evaluating the double integral in (5.85) we get

$$\int_U \frac{\partial Q}{\partial x} \, \mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^{1-y} \frac{\partial Q}{\partial x} \, \mathrm{d}x\mathrm{d}y.$$

Using the Fundamental Theorem of Calculus to evaluate the inner integral we then obtain that

$$\int_U \frac{\partial Q}{\partial x} \, \mathrm{d}x\mathrm{d}y = \int_0^1 [Q(1-y,y) - Q(0,y)] \, \mathrm{d}y. \tag{5.87}$$

Next, we evaluate the line integral in (5.85) to get

$$\int_{\partial U} Q\mathrm{d}y = \int_{[P_1,P_2]} Q\mathrm{d}y + \int_{[P_2,P_3]} Q\mathrm{d}y + \int_{[P_3,P_1]} Q\mathrm{d}y$$

or

$$\int_{\partial U} Q\mathrm{d}y = \int_{[P_2,P_3]} Q\mathrm{d}y + \int_{[P_3,P_1]} Q\mathrm{d}y, \tag{5.88}$$

since $\mathrm{d}y = 0$ on $[P_1, P_2]$.

Now, parametrize $[P_2, P_3]$ by

$$\begin{cases} x = 1 - y \\ y = y, \end{cases}$$

for $0 \leqslant y \leqslant 1$. It then follows that

$$\int_{[P_2,P_3]} Q\mathrm{d}y = \int_0^1 Q(1-y, y)\mathrm{d}y. \tag{5.89}$$

Parametrizing $[P_3, P_1]$ by

$$\begin{cases} x = 0 \\ y = 1 - t, \end{cases}$$

for $0 \leqslant t \leqslant 1$, we get that

$$\begin{cases} \mathrm{d}x = 0\mathrm{d}t \\ \mathrm{d}y = -\mathrm{d}t, \end{cases}$$

and

$$\int_{[P_3,P_1]} Q\mathrm{d}y = -\int_0^1 Q(0, 1-t)\mathrm{d}t,$$

which we can re-write as

$$\int_{[P_3,P_1]} Q\mathrm{d}y = -\int_1^0 Q(0, y)(-\mathrm{d}y) = -\int_0^1 Q(0, y)\mathrm{d}y. \tag{5.90}$$

Substituting (5.90) and (5.89) into (5.88) yields

$$\int_{\partial U} Q\mathrm{d}y = \int_0^1 Q(1-y, y)\mathrm{d}y - \int_0^1 Q(0, y)\mathrm{d}y \tag{5.91}$$

Comparing the left–hand sides on the equations (5.91) and (5.87), we see that (5.85) is true. A similar calculation shows that (5.86) is also true. Hence, Proposition 5.8.1 is proved for the unit triangle U. □

In subsequent sections, we show how to extend the proof of Green's Theorem to arbitrary triangles (which are positively oriented) and then for arbitrary bounded regions which are bounded by positively oriented simple curves.

5.9 Changing Variables

We would like to express the integral of a scalar field, $f(x, y)$, over an arbitrary triangle, T, in the xy–plane,

$$\int_T f(x, y)\ \mathrm{d}x\mathrm{d}y, \tag{5.92}$$

as an integral over the unit triangle, U, in the uv–plane,

$$\int_U g(u, v)\ \mathrm{d}u\mathrm{d}v,$$

where the function g will be determined by f and an appropriate change of coordinates that takes U to T.

We first consider the case of the triangle $T = [(0, 0), (a, 0), (0, b)]$, pictured in Figure 5.9.9, where a and b are positive real numbers.

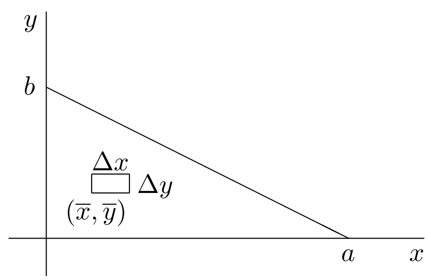

Figure 5.9.9: Triangle $[(0, 0), (a, 0), (0, b)]$

Observe that the vector field

$$\Phi\colon \mathbb{R}^2 \to \mathbb{R}^2$$

defined by

$$\Phi\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au \\ bv \end{pmatrix}, \quad \text{for all} \quad \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2,$$

maps the unit triangle, U, in the uv–plane pictured in Figure 5.9.10, to the triangle T in the xy–plane. The reason for this is that the line segment $[(0, 0), (1, 0)]$ in the uv–plane, parametrized by

$$\begin{cases} u = t \\ v = 0, \end{cases}$$

for $0 \leqslant t \leqslant 1$, gets mapped to

$$\begin{cases} x = at \\ y = 0, \end{cases}$$

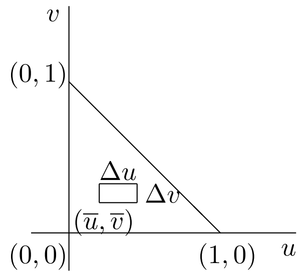

Figure 5.9.10: Unit Triangle, U, in the uv–plane

for $0 \leqslant t \leqslant 1$, which is a parametrization of the line segment $[(0,0),(a,0)]$ in the xy–plane.

Similarly, the line segment $[(1,0),(0,1)]$ in the uv–plane, parametrized by

$$\begin{cases} u = 1-t \\ v = t, \end{cases}$$

for $0 \leqslant t \leqslant 1$, gets mapped to

$$\begin{cases} x = a(1-t) \\ v = bt, \end{cases}$$

for $0 \leqslant t \leqslant 1$, which is a parametrization of the line segment $[(a,0),(0,b)]$ in the xy–plane.

Similar considerations show that $[(0,1),(0,0)]$ gets mapped to $[(0,b),(0,0)]$ under the action of Φ on $\mathbb{R}^2$.

Writing

$$\begin{pmatrix} x(u,v) \\ y(u,v) \end{pmatrix} = \Phi \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{for all} \quad \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2,$$

we can express the integrand in the double integral in (5.92) as a function of u and v:

$$f(x(u,v),y(u,v)) \quad \text{for} \quad (u,v) \quad \text{in} \quad U.$$

We presently see how the differential 2–form $\mathrm{d}x\mathrm{d}y$ can be expressed in terms of $\mathrm{d}u\mathrm{d}v$. To do this consider the small rectangle of area $\Delta u\Delta v$ and lower left–hand corner at $(\overline{u},\overline{v})$ pictured in Figure 5.9.10. We see where the vector field Φ maps this rectangle in the xy–plane. In this case, it happens to be a rectangle with lower–left hand corner $\Phi(\overline{u},\overline{v}) = (\overline{x},\overline{y})$ and dimensions $a\Delta u \times b\Delta v$. In general, however, the image of the $\Delta u \times \Delta v$ rectangle under a change of coordinates Φ will be a plane region bounded by curves like the one pictured in Figure 5.9.11. In the general case, we approximate the area of the image region by the area of the parallelogram spanned by vectors tangent to the image curves of the line

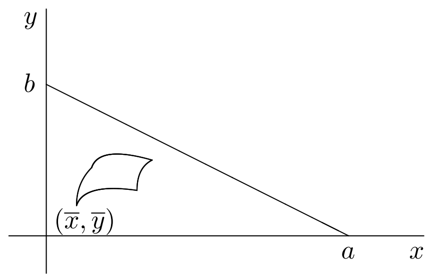

Figure 5.9.11: Image of Rectangle under Φ

segments $[(\overline{u}, \overline{v}), (\overline{u} + \Delta u, \overline{v})]$ and $[(\overline{u}, \overline{v}), (\overline{u}, \overline{v} + \Delta v)]$ under the map Φ at the point $(\overline{u}, \overline{v})$. The curves are given parametrically by

$$\sigma(u) = \Phi(v, \overline{v}) = (x(u, \overline{v}), y(u, \overline{v})) \quad \text{for} \quad \overline{u} \leqslant u \leqslant \overline{u} + \Delta u,$$

and

$$\gamma(v) = \Phi(\overline{u}, v) = (x(\overline{u}, v), y(\overline{u}, v)) \quad \text{for} \quad \overline{v} \leqslant v \leqslant \overline{v} + \Delta v.$$

The tangent vectors the the point $(\overline{u}, \overline{v})$ are, respectively,

$$\Delta u \ \sigma'(\overline{u}) = \Delta u \left(\frac{\partial x}{\partial u} \ \widehat{i} + \frac{\partial y}{\partial u} \ \widehat{j} \right),$$

and

$$\Delta v \ \gamma'(\overline{v}) = \Delta v \left(\frac{\partial x}{\partial v} \ \widehat{i} + \frac{\partial y}{\partial v} \ \widehat{j} \right),$$

where we have scaled by Δu and Δv, respectively, by virtue of the linear approximation provided by the derivative maps $D\sigma(\overline{u})$ and $D\gamma(\overline{v})$, respectively. The area of the image rectangle can then be approximated by the norm of the cross product of the tangent vectors:

$$\begin{aligned} \Delta x \Delta y &\approx \|\Delta u \ \sigma'(\overline{u}) \times \Delta v \ \gamma'(\overline{v})\| \\ &= \|\sigma'(\overline{u}) \times \gamma'(\overline{v})\| \Delta u \Delta v \end{aligned}$$

Evaluating the cross–product $\sigma'(\overline{u}) \times \gamma'(\overline{v})$ yields

$$\begin{aligned}
\sigma'(\overline{u}) \times \gamma'(\overline{v}) &= \left(\frac{\partial x}{\partial u}\,\widehat{i} + \frac{\partial y}{\partial u}\,\widehat{j}\right) \times \left(\frac{\partial x}{\partial v}\,\widehat{i} + \frac{\partial y}{\partial v}\,\widehat{j}\right) \\
&= \frac{\partial x}{\partial u}\frac{\partial y}{\partial v}\,\widehat{i} \times \widehat{j} + \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\,\widehat{j} \times \widehat{i} \\
&= \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\right)\widehat{k} \\
&= \frac{\partial(x,y)}{\partial(u,v)}\widehat{k},
\end{aligned}$$

where $\dfrac{\partial(x,y)}{\partial(u,v)}$ denotes the determinant of the Jacobian matrix of the Φ at $(\overline{u}, \overline{v})$. It then follows that

$$\Delta x \Delta y \approx \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \Delta u \Delta v,$$

which translates in terms of differential forms to

$$\mathrm{d}x\mathrm{d}y = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \mathrm{d}u\mathrm{d}v.$$

We therefore obtain the *Change of Variables Formula*

$$\int_T f(x,y)\ \mathrm{d}x\mathrm{d}y = \int_U f(x(u,v), y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right|\ \mathrm{d}u\mathrm{d}v. \tag{5.93}$$

This formula works for any regions R and D in the plane for which there is a change of coordinates $\Phi\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $\Phi(D) = R$:

$$\int_R f(x,y)\ \mathrm{d}x\mathrm{d}y = \int_D f(x(u,v), y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right|\ \mathrm{d}u\mathrm{d}v. \tag{5.94}$$

Example 5.9.1. *For the case in which $T = [(0,0),(a,0),(0,b)]$ and U is the unit triangle in $\mathbb{R}^2$, and Φ is given by*

$$\Phi\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au \\ bv \end{pmatrix} \quad \textit{for all} \quad \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2,$$

The Change of Variables Formula (5.93) yields

$$\int_T f(x,y)\ dxdy = ab\int_U f(au,bv)\ dudv.$$

Example 5.9.2. *Let $R = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leqslant 1\}$. Evaluate*

$$\int_R e^{-x^2-y^2}\ dxdy.$$

Solution: Let $D = \{(r,\theta) \in \mathbb{R}^2 \mid 0 \leqslant r \leqslant 1, 0 \leqslant \theta < 2\pi\}$ and consider the change of variables

$$\Phi \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \quad \text{for all} \quad \begin{pmatrix} r \\ \theta \end{pmatrix} \in \mathbb{R}^2,$$

or

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta. \end{cases}$$

The change of variables formula (5.94) in this case then reads

$$\int_R f(x,y)\ \mathrm{d}x\mathrm{d}y = \int_D f(r\cos\theta, r\sin\theta) \left| \frac{\partial(x,y)}{\partial(y,\theta)} \right|\ \mathrm{d}r\mathrm{d}\theta,$$

where $f(x,y) = e^{-x^2-y^2}$, and

$$\begin{aligned} \frac{\partial(x,y)}{\partial(y,\theta)} &= \det \begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{pmatrix} \\ &= \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \\ &= r. \end{aligned}$$

Hence,

$$\begin{aligned} \int_R e^{-x^2-y^2}\ \mathrm{d}x\mathrm{d}y &= \int_D e^{-r^2} r\ \mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 e^{-r^2} r\ \mathrm{d}r\mathrm{d}\ \theta \\ &= \int_0^{2\pi} \left[-\frac{1}{2} e^{-r^2} \right]_0^1\ \mathrm{d}\theta \\ &= \frac{1}{2} \int_0^{2\pi} \left(1 - e^{-1}\right)\ \mathrm{d}\theta \\ &= \pi\left(1 - e^{-1}\right). \end{aligned}$$

□

Example 5.9.3 (Green's Theorem for Arbitrary Triangles in $\mathbb{R}^2$)**.**

Appendix A

The Mean Value Theorem in Convex Sets

Definition A.0.1 (Convex Sets)**.** *A subset, A, of* $\mathbb{R}^n$ *is said to be* **convex** *if given any two points x and y in A, the straight line segment connecting them is entirely contained in A; in symbols,*

$$\{x + t(y - x) \in \mathbb{R}^n \mid 0 \leq t \leqslant 1\} \subseteq A$$

Example A.0.2. *Prove that the ball* $B_r(O) = \{x \in \mathbb{R}^n \mid \|x\| < r\}$ *is a convex subset of* $\mathbb{R}^n$*.*

Solution*:* Let x and y be in $B_r(O)$; then, $\|x\| < r$ and $\|y\| < r$. For $0 \leqslant t \leqslant 1$, consider

$$x + t(y - x) = (1 - t)x + ty.$$

Thus, taking the norm and using the triangle inequality

$$\begin{aligned} \|x + t(y - x)\| &= \|(1 - t)x + ty\| \\ &\leqslant (1 - t)\|x\| + t\|y\| \\ &< (1 - t)r + tr = r. \end{aligned}$$

Thus, $x + t(y - x) \in B_r(O)$ for any $t \in [0, 1]$. Since this is true for any $x, y \in B_r(O)$, it follows that $B_r(O)$ is convex. □

In fact, any ball in $\mathbb{R}^n$ is convex.

Proposition A.0.3 (Mean Value Theorem for Scalar Fields on Convex Sets)**.** *Let B denote and open, convex subset of* $\mathbb{R}^n$*, and let* $f : B \to \mathbb{R}$ *be a scalar field. Suppose that f is differentiable on B. Then, for any pair of points x and y in B, there exists a point z is the line segment connecting x to y such that*

$$f(y) - f(x) = D_{\widehat{u}} f(z)\|y - x\|,$$

where $\widehat{u}$ *is the unit vector in the direction of the vector* $y - x$*; that is,*

$$\widehat{u} = \frac{1}{\|y - x\|}(y - x).$$

Proof. Assume that $x \neq y$, for if $x = y$ the equality certainly holds true.

Define $g\colon [0,1] \to \mathbb{R}$ by

$$g(t) = f(x + t(y - x)) \quad \text{for} \quad 0 \leqslant t \leqslant 1.$$

We first show that g is differentiable on $(0,1)$ and that

$$g'(t) = \nabla f(x + t(y - x)) \cdot (y - x) \quad \text{for} \quad 0 < t < 1.$$

(This has been proved in Exercise 4 of Assignment #10).

Now, by the Mean Value Theorem, there exists $\tau \in (0,1)$ such that

$$g(1) - g(0) = g'(\tau)(1 - 0) = g'(\tau).$$

It then follows that

$$f(y) - f(x) = \nabla f(x + \tau(y - x)) \cdot (y - x).$$

Put $z = x + \tau(y - x)$; then, z is a point in the line segment connecting x to y, and

$$\begin{aligned} f(y) - f(x) &= \nabla f(z) \cdot (y - x) \\ &= \nabla f(z) \cdot \frac{y - x}{\|y - x\|} \|y - x\| \\ &= \nabla f(z) \cdot \widehat{u}\ \|y - x\| \\ &= D_{\widehat{u}} f(z) \|y - x\|, \end{aligned}$$

where $\widehat{u} = \dfrac{1}{\|y - x\|}(y - x)$. □

Appendix B

Reparametrizations

In this appendix we prove that any two parameterizations of a C^1 simple curve are reparametrizations of each other; more precisely,

Theorem B.0.1. Let I and J denote open intervals of real numbers containing closed and bounded intervals $[a, b]$ and $[c, d]$, respectively, and $\gamma_1 \colon I \to \mathbb{R}^n$ and $\gamma_2 \colon J \to \mathbb{R}^n$ be C^1 paths. Suppose that $C = \gamma_1([a, b]) = \gamma_2([c, d])$ is a C^1 simple curve parametrized by γ_1 and γ_2. Then, there exists a differentiable function, $\tau \colon J \to I$, such that

(i) $\tau'(t) > 0$ for all $t \in J$;

(ii) $\tau(c) = a$ and $\tau(d) = b$; and

(iii) $\gamma_2(t) = \gamma_1(\tau(t))$ for all $t \in J$.

In order to prove Theorem B.0.1, we need to develop the notion of a **tangent space** to a C^1 curve at a given point. We begin with a preliminary definition.

Definition B.0.2 (Tangent Space (Preliminary Definition))**.** Let C denote a C^1 simple curve parameterized by a C^1 path, $\sigma \colon I \to \mathbb{R}^n$, where I is an open interval containing 0, and such that and $\sigma(0) = p$. We define the **tangent space**, $T_p(C)$, of C at p to be the span of the nonzero vector $\sigma'(0)$; that is,

$$T_p(C) = \operatorname{span}\{\sigma'(0)\}.$$

Remark B.0.3. Observe that the set $p + T_p(C)$ is the tangent line to the curve C at p, hence the name "tangent space" for $T_p(C)$.

The notion of tangent space is important because it allows us to define the derivative at p of a map $g \colon C \to \mathbb{R}$ which is solely defined on the curve C. The idea is to consider the composition $g \circ \sigma \colon I \to \mathbb{R}$ and to require that the real valued function $g \circ \sigma$ be differentiable at $t = 0$. For the case of a C^1 scalar field,

f, which is defined on an open region containing C, the Chain Rule implies that $f \circ \sigma$ is differentiable at 0 and

$$(f \circ \sigma)'(0) = \nabla f(\sigma(0)) \cdot \sigma'(0) = \nabla f(p) \cdot v,$$

where $v = \sigma'(0) \in T_p(C)$. Observe that the map

$$v \mapsto \nabla f(p) \cdot v \quad \text{ for } \ v \in T_p(C)$$

defines a linear map on the tangent space of C at p. We will denote this linear map by df_p; that is, $df_p \colon T_p(C) \to \mathbb{R}$ is given by

$$df_p(v) = \nabla f(p) \cdot v, \quad \text{ for } \ v \in T_p(C).$$

Observe that we can then write, for $h \in \mathbb{R}$ with $|h|$ sufficiently small,

$$(f \circ \sigma)(0 + h) = f(\sigma(0)) + df_p(h\sigma'(0)) + E_0(h),$$

where

$$\lim_{h \to 0} \frac{|E_0(h)|}{|h|} = 0.$$

Definition B.0.4. Let C denote a C^1 curve parametrized by a C^1 path, $\sigma \colon I \to \mathbb{R}^n$, where J is an open interval containing 0 and such that $\sigma(0) = p \in C$. We say that the function $g \colon C \to \mathbb{R}$ is differentiable at p if there exists a linear function

$$dg_p \colon T_p(C) \to \mathbb{R}$$

such that

$$(g \circ \sigma)(h) = g(p) + dg_p(h\sigma'(0)) + E_p(h),$$

where

$$\lim_{h \to 0} \frac{|E_p(h)|}{|h|} = 0.$$

We see from Definition B.0.4 that, if $g \colon C \to \mathbb{R}$ is differentiable at p, then

$$\lim_{h \to 0} \frac{g(\sigma(h)) - g(p)}{h}$$

exists and equals $dg_p(\sigma'(0))$. We have already seen that if f is a C^1 scalar field defined in an open region containing C, then

$$df_p(\sigma'(0)) = \nabla f(p) \cdot \sigma'(0).$$

If the only information we have about a function g is what it does to points on C, then we see why Definition B.0.4 is relevant. In the general case it might not make sense to talk about the gradient of g.

An example of a function, g, which is only defined on C is the inverse of a C^1 parametrization, $\gamma \colon J \to \mathbb{R}^n$, of C where J is an interval containing 0 in its interior with $\gamma(0) = p$. Here we are assuming that γ is one–to–one and onto C, so that

$$g = \gamma^{-1}\colon C \to J$$

is defined. We claim that, since $\gamma'(0) \neq \mathbf{0}$, according to the definition of C^1 parametrization in Definition 5.1.1 on page 79 in these notes, the function g is differentiable at p according to Definition B.0.4. In order to prove this, we first show that g is continuous at p; that is,

Lemma B.0.5. Let C be a C^1 curve parametrized by a C^1 map, $\sigma\colon I \to \mathbb{R}^n$, where I is an interval of real numbers containing 0 in its interior with $\sigma(0) = p$. Let $\gamma\colon J \to \mathbb{R}^n$ denote another C^1 parametrization of C, where J is an interval of real numbers containing 0 in its interior with $\gamma(0) = p$. For every $q \in C$, define $g(q) = \tau$ if and only if $\gamma(\tau) = q$. Then,

$$\lim_{h\to 0} g(\sigma(h)) = 0. \tag{B.1}$$

Proof: Write

$$\tau(h) = g(\sigma(h)), \quad \text{for } h \in I. \tag{B.2}$$

We will show that

$$\lim_{h\to 0} \tau(h) = 0; \tag{B.3}$$

this will prove (B.1).

From (B.2) and the definition of g we obtain that

$$\gamma(\tau(h)) = \sigma(h), \quad \text{for } h \in I. \tag{B.4}$$

Letting $h = 0$ in (B.4) we see that

$$\gamma(\tau(0)) = p, \tag{B.5}$$

from which we get that

$$\tau(0) = 0, \tag{B.6}$$

since $\gamma\colon J \to \mathbb{R}^n$ is a parametrization of C with $\gamma(0) = p$.

Write

$$\sigma(t) = (x_1(t), x_2(t), \ldots, x_n(t)), \quad \text{for all } t \in I, \tag{B.7}$$

$$\gamma(\tau) = (y_1(\tau), y_2(\tau), \ldots, y_n(\tau)), \quad \text{for all } \tau \in J, \tag{B.8}$$

and

$$p = (p_1, p_2, \ldots, p_n). \tag{B.9}$$

Since $\gamma'(\tau) \neq \mathbf{0}$ for all $\tau \in J$, there exists $j \in \{1, 2, \ldots, n\}$ such that

$$y_j'(0) \neq 0.$$

Consequently, there exists $\delta > 0$ such that

$$|\tau| \leqslant \delta \Rightarrow |y_j'(\tau)| \geqslant \frac{|y_j'(0)|}{2}. \tag{B.10}$$

It follows from (B.4), (B.7) and (B.8) that

$$y_j(\tau(h)) = x_j(h), \quad \text{for } h \in I. \tag{B.11}$$

Next, use the differentiability of the function $y_j \colon J \to \mathbb{R}$ and the mean value theorem to obtain $\theta \in (0,1)$ such that

$$y_j(\tau(h)) - p_j = \tau(h) y_j'(\theta\tau(h)), \tag{B.12}$$

where we have used (B.5), (B.6) and (B.9). Thus, for

$$|\tau(h)| \leqslant \delta,$$

it follows from (B.10) and (B.12) that

$$m\ |\tau(h)| \leqslant |y_j(\tau(h)) - p_j|, \tag{B.13}$$

where we have set

$$m = \frac{|y_j'(0)|}{2} > 0. \tag{B.14}$$

On the other hand, it follows from (B.11) and the differentiability of x_j that

$$y_j(\tau(h)) = p_j + h x_j'(0) + E_j(h), \quad \text{for } h \in I, \tag{B.15}$$

where

$$\lim_{h \to 0} \frac{E_j(h)}{h} = 0. \tag{B.16}$$

Consequently, using (B.13) and (B.15), if $\tau(h)| \leqslant \delta$,

$$m\ |\tau(h)| \leqslant |h||x_j'(0)| + |E_j(h)|. \tag{B.17}$$

The statement in (B.3) now follows from (B.17) and (B.16), since $m > 0$ by virtue of (B.14). □

Lemma B.0.6. Let C, $\sigma \colon I \to \mathbb{R}^n$ and $\gamma \colon J \to \mathbb{R}^n$ be as in Lemma B.0.5. For every $q \in C$, define $g(q) = \tau$ if and only if $\gamma(\tau) = q$. Then, the function $\tau \colon I \to J$ is differentiable at 0. Consequently, the function $g \colon C \to J$ is differentiable at p and

$$dg_p(\sigma'(0)) = \lim_{h \to 0} \frac{g(\sigma(h)) - g(p)}{h} = \tau'(0).$$

Furthermore,

$$\gamma'(0) = \frac{1}{\tau'(0)} \sigma'(0). \tag{B.18}$$

Proof: As in the proof of Lemma B.0.5, let $j \in \{1, 2, \ldots, n\}$ be such that

$$y_j'(0) \neq 0. \tag{B.19}$$

Using the differentiability of γ and σ, we obtain from (B.11) that

$$p_j + \tau(h) y_j'(0) + E_1(\tau(h)) = p_j + h x_j'(0) + E_2(h), \tag{B.20}$$

where

$$\lim_{\tau(h)\to 0} \frac{E_1(\tau(h))}{\tau(h)} = 0 \quad \text{and} \quad \lim_{h\to 0} \frac{E_2(h)}{h} = 0. \tag{B.21}$$

We obtain from (B.20) and (B.19) that

$$\frac{\tau(h)}{h}\left[1 + \frac{1}{y_j'(0)}\frac{E_1(\tau(h))}{\tau(h)}\right] = \frac{x_j'(0)}{y_j'(0)} + \frac{1}{y_j'(0)}\frac{E_2(h)}{h},$$

from which we get

$$\frac{\tau(h)}{h} = \frac{\dfrac{x_j'(0)}{y_j'(0)} + \dfrac{1}{y_j'(0)}\dfrac{E_2(h)}{h}}{1 + \dfrac{1}{y_j'(0)}\dfrac{E_1(\tau(h))}{\tau(h)}}. \tag{B.22}$$

Next, apply Lemma B.0.5 and (B.21) to obtain from (B.22) that

$$\lim_{h\to 0} \frac{\tau(h)}{h} = \frac{x_j'(0)}{y_j'(0)},$$

which shows that τ is differentiable at 0, in view of (B.6).

Finally, applying the Chain Rule to the expression in (B.4) we obtain

$$\tau'(0)\gamma'(0) = \sigma'(0),$$

which yields (B.18). □

The expression in (B.18) in the statement of Lemma B.0.6 allows us to expand the preliminary definition of the tangent space of C at p given in Definition B.0.2 as follows:

Definition B.0.7 (Tangent Space)**.** Let C denote a C^1 simple curve in $\mathbb{R}^n$ and $p \in C$. We define the **tangent space**, $T_p(C)$, of C at p to be

$$T_p(C) = \text{span}\{\sigma'(0)\},$$

where $\sigma\colon (-\varepsilon, \varepsilon) \to C$ is any C^1 map defined on $(-\varepsilon, \varepsilon)$, for some $\varepsilon > 0$, such that $\sigma'(t) \neq \mathbf{0}$ for all $t \in (-\varepsilon, \varepsilon)$ and $\sigma(0) = p$.

Indeed, if $\gamma\colon (-\varepsilon, \varepsilon) \to C$ is another C^1 map with the properties that $\gamma'(t) \neq \mathbf{0}$ for all $t \in (-\varepsilon, \varepsilon)$ and $\gamma(0) = p$, it follows from (B.18) in Lemma B.0.6 that

$$\text{span}\{\sigma'(0)\} = \text{span}\{\gamma'(0)\}.$$

Thus, the definition of T_pC in Definition B.0.7 is independent of the choice of parametrization, σ.

Next, let $\gamma\colon J \to C$ be a parametrization of a C^1 curve, C. We note for future reference that $\gamma'(t) \in T_{\gamma(t)}(C)$ for all $t \in J$. To see why this is the case, let $\varepsilon > 0$ be sufficiently small so that $(t-\varepsilon, t+\varepsilon) \subset J$, and define $\sigma\colon (-\varepsilon, \varepsilon) \to C$ by

$$\sigma(\tau) = \gamma(t+\tau), \quad \text{for all } \tau \in (-\varepsilon, \varepsilon).$$

Then, σ is a C^1 map satisfying $\sigma'(\tau) = \gamma'(\tau + t) \neq \mathbf{0}$ for all $\tau \in (-\varepsilon, \varepsilon)$ and $\sigma(0) = \gamma(t)$. Observe also that $\sigma'(0) = \gamma'(t)$. It then follows by Definition B.0.7 that $\gamma'(t) \in T_{\gamma(t)}C$ for all $t \in J$.

Proposition B.0.8 (Chain Rule)**.** Let C be a C^1 simple curve parametrized by a C^1 path, $\gamma\colon J \to \mathbb{R}^n$. Suppose that $g\colon C \to \mathbb{R}$ is a differentiable function defined on C. Then, the map $g \circ \gamma\colon J \to \mathbb{R}$ is a differentiable function and

$$\frac{d}{dt}[g(\gamma(t))] = dg_{\gamma(t)}(\gamma'(t)), \quad \text{for all } t \in J. \tag{B.23}$$

Proof: Put $\sigma(h) = \gamma(t+h)$, for $|h|$ sufficiently small. By Definition B.0.4,

$$g(\gamma(t+h)) = g(\gamma(t)) + d_{\gamma(t)}(h\gamma'(t)) + E_{\gamma(t)}(h), \tag{B.24}$$

where

$$\lim_{h\to 0} \frac{|E_{\gamma(t)}(h)|}{|h|} = 0. \tag{B.25}$$

The statement in (B.23) now follows from (B.24), (B.25), and the linearity of the map $dg_{\gamma(t)}\colon T_{\gamma(t)}(C) \to \mathbb{R}$. □

Proof of Theorem B.0.1: Let I and J denote open intervals of real numbers containing closed and bounded intervals $[a, b]$ and $[c, d]$, respectively, and $\gamma_1\colon I \to \mathbb{R}^n$ and $\gamma_2\colon J \to \mathbb{R}^n$ be C^1 paths. Suppose that $C = \gamma_1([a, b]) = \gamma_2([c, d])$ is a C^1 simple curve parametrized by γ_1 and γ_2. Define $\tau\colon J \to I$ by $\tau = g \circ \gamma_2$, where $g = \gamma_1^{-1}$, the inverse of γ_1. By Lemma B.0.6, $g\colon C \to I$ is differentiable on C. It therefore follows by the Chain Rule (Proposition B.0.8) that τ is differentiable and

$$\tau'(t) = dg_{\gamma_1(t)}(\gamma_2'(t)), \quad \text{for all } t \in J.$$

In addition, we have that

$$\gamma_1(\tau(t)) = \gamma_2(t), \quad \text{for all } t \in J.$$

Thus, by the Chain Rule,

$$\tau'(t)\gamma_1'(\tau(t)) = \gamma_2'(t), \quad \text{for all } t \in J. \tag{B.26}$$

Taking norms on both sides of (B.26), and using the fact that γ_1 and γ_2 are parametrizations, we obtain from (B.26) that

$$|\tau'(t)| = \frac{\|\gamma_2'(t)\|}{\|\gamma_1'(\tau(t))\|}, \quad \text{for all } t \in J. \tag{B.27}$$

Since $\gamma_2'(t) \neq \mathbf{0}$ for all $t \in J$, it follows from (B.27) that

$$\tau'(t) \neq 0, \quad \text{for all } t \in J.$$

Thus, either

$$\tau'(t) > 0, \qquad \text{for all } t \in J, \tag{B.28}$$

or

$$\tau'(t) < 0, \qquad \text{for all } t \in J. \tag{B.29}$$

If (B.28) holds true, then the proof of Theorem B.0.1 is complete. If (B.29) is true, consider the function $\widetilde{\tau}\colon J \to I$ given by

$$\widetilde{\tau}(t) = \tau(b + a - t), \qquad \text{for all } t \in J.$$

Then, $\widetilde{\tau}$ satisfies the the properties in the conclusion of the theorem. $\square$