Lectures on Vector Calculus

Paul Renteln

Department of Physics

California State University

San Bernardino, CA 92407

March, 2009; Revised March, 2011

Contents

List of Figures

1 Vector Algebra and Index Notation

1.1 Orthonormality and the Kronecker Delta

We begin with three dimensional Euclidean space $\mathbb{R}^3$. In $\mathbb{R}^3$ we can define three special coordinate vectors $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$, and $\hat{\boldsymbol{e}}_3$. [1] We choose these vectors to be *orthonormal*, which is to say, both orthogonal and normalized (to unity). We may express these conditions mathematically by means of the *dot product* or *scalar product* as follows:

$$
\begin{aligned}
\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_2 &= \hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}_1 = 0 \\
\hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}_3 &= \hat{\boldsymbol{e}}_3 \cdot \hat{\boldsymbol{e}}_2 = 0 \qquad \text{(orthogonality)} \\
\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_3 &= \hat{\boldsymbol{e}}_3 \cdot \hat{\boldsymbol{e}}_1 = 0
\end{aligned}
\tag{1.1}
$$

and

$$
\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}_2 = \hat{\boldsymbol{e}}_3 \cdot \hat{\boldsymbol{e}}_3 = 1 \qquad \text{(normalization)}.
\tag{1.2}
$$

To save writing, we will abbreviate these equations using *dummy indices* instead. (They are called 'indices' because they index something, and they are called 'dummy' because the exact letter used is irrelevant.) In index notation, then, I claim that the conditions (1.1) and (1.2) may be written

$$
\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \delta_{ij}.
\tag{1.3}
$$

How are we to understand this equation? Well, for starters, this equation is really nine equations rolled into one! The index i can assume the values 1, 2, or 3, so we say "i runs from 1 to 3", and similarly for j. **The equation is**

---

[1]These vectors are also denoted $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, and $\hat{\boldsymbol{k}}$, or $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$. We will use all three notations interchangeably.

valid for all possible choices of values for the indices. So, if we pick, say, $i = 1$ and $j = 2$, (1.3) would read

$$\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_2 = \delta_{12}. \tag{1.4}$$

Or, if we chose $i = 3$ and $j = 1$, (1.3) would read

$$\hat{\boldsymbol{e}}_3 \cdot \hat{\boldsymbol{e}}_1 = \delta_{31}. \tag{1.5}$$

Clearly, then, as i and j each run from 1 to 3, there are nine possible choices for the values of the index pair i and j on each side, hence nine equations.

The object on the right hand side of (1.3) is called the *Kronecker delta*. It is defined as follows:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.6}$$

The Kronecker delta assumes nine possible values, depending on the choices for i and j. For example, if $i = 1$ and $j = 2$ we have $\delta_{12} = 0$, because i and j are not equal. If $i = 2$ and $j = 2$, then we get $\delta_{22} = 1$, and so on. A convenient way of remembering the definition (1.6) is to imagine the Kronecker delta as a 3 by 3 matrix, where the first index represents the row number and the second index represents the column number. Then we could write (abusing notation slightly)

$$\delta_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.7}$$

Finally, then, we can understand Equation (1.3): it is just a shorthand way of writing the nine equations (1.1) and (1.2). For example, if we choose $i = 2$ and $j = 3$ in (1.3), we get

$$\hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}_3 = 0, \tag{1.8}$$

(because $\delta_{23} = 0$ by definition of the Kronecker delta). This is just one of the equations in (1.1). Letting i and j run from 1 to 3, we get all the nine orthornormality conditions on the basis vectors $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$ and $\hat{\boldsymbol{e}}_3$.

Remark. It is easy to see from the definition (1.6) or from (1.7) that the Kronecker delta is what we call *symmetric*. That is

$$\delta_{ij} = \delta_{ji}. \tag{1.9}$$

Hence we could have written Equation (1.3) as

$$\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \delta_{ji}. \tag{1.10}$$

(In general, you must pay careful attention to the order in which the indices appear in an equation.)

Remark. We could have written Equation (1.3) as

$$\hat{\boldsymbol{e}}_a \cdot \hat{\boldsymbol{e}}_b = \delta_{ab}, \tag{1.11}$$

which employs the letters a and b instead of i and j. The meaning of the equation is exactly the same as before. The only difference is in the labels of the indices. This is why they are called 'dummy' indices.

Remark. We **cannot** write, for instance,

$$\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_a = \delta_{ij}, \tag{1.12}$$

as this equation makes no sense. Because all the dummy indices appearing in (1.3) are what we call *free* (see below), they must match exactly on both sides. Later we will consider what happens when the indices are not all free.

1.2   Vector Components and Dummy Indices

Let $\boldsymbol{A}$ be a vector in $\mathbb{R}^3$. As the set $\{\hat{\boldsymbol{e}}_i\}$ forms a basis for $\mathbb{R}^3$, the vector $\boldsymbol{A}$ may be written as a linear combination of the $\hat{\boldsymbol{e}}_i$:

$$\boldsymbol{A} = A_1\hat{\boldsymbol{e}}_1 + A_2\hat{\boldsymbol{e}}_2 + A_3\hat{\boldsymbol{e}}_3. \tag{1.13}$$

The three numbers A_i, $i = 1, 2, 3$, are called the (Cartesian) *components* of the vector $\boldsymbol{A}$.

We may rewrite Equation (1.13) using indices as follows:

$$\boldsymbol{A} = \sum_{i=1}^{3} A_i\hat{\boldsymbol{e}}_i. \tag{1.14}$$

As we already know that i runs from 1 to 3, we usually omit the limits from the summation symbol and just write

$$\boldsymbol{A} = \sum_i A_i\hat{\boldsymbol{e}}_i. \tag{1.15}$$

Later we will abbreviate this expression further.

Using indices allows us to shorten many computations with vectors. For

example, let us prove the following formula for the components of a vector:

$$A_j = \hat{\boldsymbol{e}}_j \cdot \boldsymbol{A}. \tag{1.16}$$

We proceed as follows:

$$\hat{\boldsymbol{e}}_j \cdot \boldsymbol{A} = \hat{\boldsymbol{e}}_j \cdot \left(\sum_i A_i \hat{\boldsymbol{e}}_i \right) \tag{1.17}$$

$$= \sum_i A_i (\hat{\boldsymbol{e}}_j \cdot \hat{\boldsymbol{e}}_i) \tag{1.18}$$

$$= \sum_i A_i \delta_{ij} \tag{1.19}$$

$$= A_j. \tag{1.20}$$

In Equation (1.17) we simply substituted Equation (1.15). In Equation (1.18) we used the linearity of the dot product, which basically says that we can distribute the dot product over addition, and scalars pull out. That is, dot products are products between vectors, so any scalars originally multiplying vectors just move out of the way, and only multiply the final result. Equation (1.19) employed Equation (1.3) and the symmetry of δ_{ij}.

It is Equation (1.20) that sometimes confuses the beginner. To see how the transition from (1.19) to (1.20) works, let us look at it in more detail. The equation reads

$$\sum_i A_i \delta_{ij} = A_j. \tag{1.21}$$

Notice that the left hand side is a sum over i, and not i and j. We say that the index j in this equation is "free", because it is not summed over. As j is free, we are free to choose any value for it, from 1 to 3. Hence (1.21) is really three equations in one (as is Equation (1.16)). Suppose we choose $j = 1$.

Then written out in full, Equation (1.21) becomes

$$A_1 \delta_{11} + A_2 \delta_{21} + A_3 \delta_{31} = A_1. \tag{1.22}$$

Substituting the values of the Kronecker delta yields the identity $A_1 = A_1$, which is correct. You should convince yourself that the other two cases work out as well. That is, no matter what value of j we choose, the left hand side of (1.21) (which involved the sum with the Kronecker delta) always equals the right hand side.

Looking at (1.21) again, we say that the Kronecker delta together with the summation has effected an "index substitution", allowing us to replace the i index on the A_i with a j. In what follows we will often make this kind of index substitution without commenting. If you are wondering what happened to an index, you may want to revisit this discussion.

Observe that I could have written Equation (1.16) as follows:

$$A_i = \hat{\boldsymbol{e}}_i \cdot \boldsymbol{A}, \tag{1.23}$$

using an i index rather than a j index. The equation remains true, because i, like j, can assume all the values from 1 to 3. However, the *proof* of (1.23) must now be different. Let's see why.

Repeating the proof line for line, but with an index i instead gives us the

6

following

$$\hat{\boldsymbol{e}}_i \cdot \boldsymbol{A} = \hat{\boldsymbol{e}}_i \cdot \left(\sum_i A_i \hat{\boldsymbol{e}}_i \right) \tag{1.24}$$

$$= \sum_i A_i (\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_i)?? \tag{1.25}$$

$$= \sum_i A_i \delta_{ii}?? \tag{1.26}$$

$$= A_i. \tag{1.27}$$

Unfortunately, the whole thing is nonsense. Well, not the whole thing. Equation (1.24) is correct, but a little confusing, as an index i now appears both inside and outside the summation. Is i a free index or not? Well, there is an ambiguity, which is why you **never** want to write such an expression. The reason for this can be seen in (1.25), which is a mess. There are now *three i* indices, and it is never the case that you have a simultaneous sum over three indices like this. The sum, written out, reads

$$A_1(\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_1) + A_2(\hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}_2) + A_3(\hat{\boldsymbol{e}}_3 \cdot \hat{\boldsymbol{e}}_3)?? \tag{1.28}$$

Equation (1.2) would allow us to reduce this expression to

$$A_1 + A_2 + A_3?? \tag{1.29}$$

which is definitely not equal to A_i under any circumstances. Equation (1.26) is equally nonsense.

What went wrong? Well, the problem stems from using too many i indices. We can fix the proof, but we have to be a little more clever. The left hand side of (1.24) is fine. But instead of expressing $\boldsymbol{A}$ as a sum over i,

we can replace it by a sum over j! After all, the indices are just dummies. If we were to do this (which, by the way, we call "switching dummy indices"), the (correct) proof of (1.23) would now be

$$\hat{e}_i \cdot \boldsymbol{A} = \hat{e}_i \cdot \left(\sum_j A_j \hat{e}_j \right) \tag{1.30}$$

$$= \sum_j A_j (\hat{e}_i \cdot \hat{e}_j) \tag{1.31}$$

$$= \sum_j A_j \delta_{ij} \tag{1.32}$$

$$= A_i. \tag{1.33}$$

You should convince yourself that every step in this proof is legitimate!

1.3   Vector Algebra I: Dot Product

Vector algebra refers to doing addition and multiplication of vectors. Addition is easy, but perhaps unfamiliar using indices. Suppose we are given two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$, and define

$$\boldsymbol{C} := \boldsymbol{A} + \boldsymbol{B}. \tag{1.34}$$

Then

$$C_i = (\boldsymbol{A} + \boldsymbol{B})_i = A_i + B_i. \tag{1.35}$$

That is, the components of the sum are just the sums of the components of the addends.

Dot products are also easy. I claim

$$\boldsymbol{A} \cdot \boldsymbol{B} = \sum_i A_i B_i. \tag{1.36}$$

The proof of this from (1.3) and (1.16) is as follows:

$$\boldsymbol{A} \cdot \boldsymbol{B} = \left(\sum_i A_i \hat{\boldsymbol{e}}_i \right) \cdot \left(\sum_j B_j \hat{\boldsymbol{e}}_j \right) \tag{1.37}$$

$$= \sum_{ij} A_i B_j (\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j) \tag{1.38}$$

$$= \sum_{ij} A_i B_j \delta_{ij} \tag{1.39}$$

$$= \sum_i A_i B_i. \tag{1.40}$$

A few observations are in order. First, (1.36) could be taken as the *definition* of the dot product of two vectors, from which we could derive the properties of the dot products of the basis vectors. We chose to do it this way to illustrate the computational power of index notation. Second, in Equation (1.38) the sum over the pair ij means the double sum over i and j separately. All we have done there is use the linearity of the dot product again to pull the scalars to the front and leave the vectors to multiply *via* the dot product. Third, in Equation (1.39) we used Equation (1.3), while in (1.40) we used the substitution property of the Kronecker delta under a sum. In this case we summed over j and left i alone. This changed the j to an i. We could have equally well summed over i and left the j alone. Then the final expression would have been

$$\sum_j A_j B_j. \tag{1.41}$$

But, of course,

$$\sum_i A_i B_i = \sum_j A_j B_j = A_1 B_1 + A_2 B_2 + A_3 B_3, \qquad (1.42)$$

so it would not have mattered. Dummy indices again!

Lastly, notice that we would have gotten into big trouble had we used an i index in the sum for $\boldsymbol{B}$ instead of a j index. We would have been very confused as to which i belonged with which sum! In this case I chose an i and a j, but when you do computations like this you will have to be alert and choose your indices wisely.

1.4 The Einstein Summation Convention

You can already see that more involved computations will require more indices, and the formulas can get a little crowded. This happened often to Einstein. Being the lazy guy he was, he wanted to simplify the writing of his formulas, so he invented a new kind of notation. He realized that he could simply *erase* the summation symbols, because it was always clear that, whenever two identical dummy indices appeared on the same side of an equation they were always summed over. Removing the summation symbol leaves behind an expression with what we call an "implicit sum". The sum is still there, but it is hiding.

As an example, let us rewrite the proof of (1.36):

$$\boldsymbol{A} \cdot \boldsymbol{B} = (A_i \hat{\boldsymbol{e}}_i) \cdot (B_j \hat{\boldsymbol{e}}_j) \tag{1.43}$$

$$= A_i B_j (\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j) \tag{1.44}$$

$$= A_i B_j \delta_{ij} \tag{1.45}$$

$$= A_i B_i. \tag{1.46}$$

The only thing that has changed is that we have dropped the sums! We just have to tell ourselves that the sums are still there, so that any time we see two identical indices on the *same* side of an equation, we have to sum over them. As we were careful to use different dummy indices for the expansions of $\boldsymbol{A}$ and $\boldsymbol{B}$, we never encounter any trouble doing these sums. But note that two identical indices on *opposite* sides of an equation are never summed.

Having said this, I must say that there are rare instances when it becomes necessary to not sum over repeated indices. If the Einstein summation convention is in force, one must explicitly say "no sum over repeated indices". I do not think we shall encounter any such computations in this course, but you never know.

For now we will continue to write out the summation symbols. Later we will use the Einstein convention.

1.5   Dot Products and Lengths

The (Euclidean) **length** of a vector $\boldsymbol{A} = A_1 \hat{\boldsymbol{e}}_1 + A_2 \hat{\boldsymbol{e}}_2 + A_3 \hat{\boldsymbol{e}}_3$ is, by definition,

$$A = |\boldsymbol{A}| = \sqrt{A_1^2 + A_2^2 + A_3^2}. \tag{1.47}$$

Hence, the squared length of $\boldsymbol{A}$ is

$$A^2 = \boldsymbol{A} \cdot \boldsymbol{A} = \sum_i A_i^2. \tag{1.48}$$

Observe that, in this case, the Einstein summation convention can be confusing, because the right hand side would become simply A_i^2, and we would not know whether we mean the square of the single component A_i or the sum of squares of the A_i's. But the former interpretation would be nonsensical in this context, because A^2 is clearly not the same as the square of one of its components. That is, there is only one way to interpret the equation $A^2 = A_i^2$, and that is as an implicit sum. Nevertheless, confusion still sometimes persists, so under these circumstances it is usually best to either write $A^2 = A_i A_i$, in which case the presence of the repeated index i clues in the reader that there is a suppressed summation sign, or else to simply restore the summation symbol.

1.6   Dot Products and Angles

Let $\boldsymbol{A}$ be a vector in the plane inclined at an angle of θ to the horizontal. Then from elementary trigonometry we know that $A_1 = \hat{\boldsymbol{e}}_1 \cdot \boldsymbol{A} = A \cos \theta$ where A is the length of $\boldsymbol{A}$. It follows that if $\boldsymbol{B}$ is a vector of length B along the x axis, then $\boldsymbol{B} = B \hat{\boldsymbol{e}}_1$, and

$$\boldsymbol{A} \cdot \boldsymbol{B} = AB \cos \theta. \tag{1.49}$$

But now we observe that this relation must hold in general, no matter which way $\boldsymbol{A}$ and $\boldsymbol{B}$ are pointing, because we can always rotate the coordinate system until the two vectors lie in a plane with $\boldsymbol{B}$ along one axis.
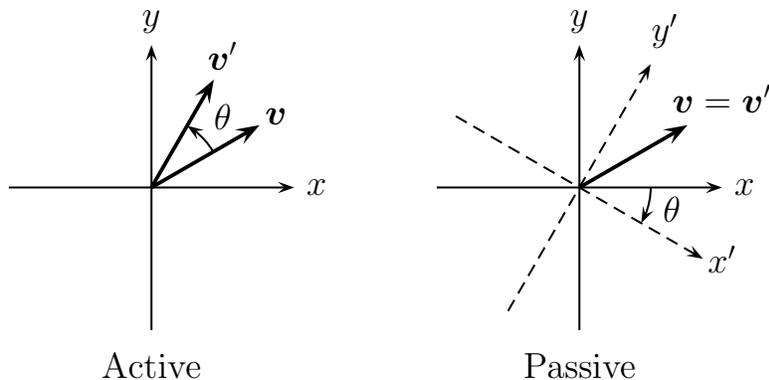
Figure 1: Active *versus* passive rotations in the plane

1.7 Angles, Rotations, and Matrices

This brings us naturally to the subject of rotations. There are many ways to understand rotations. A physicist understands rotations intuitively, whereas a mathematician requires a bit more rigor. We will begin with the intuitive approach, and later discuss the more rigorous version.

Physicists speak of transformations as being either **active** or **passive**. Consider the rotation of a vector $\boldsymbol{v}$ in the plane. According to the active point of view, we rotate the vector and leave the coordinate system alone, whereas according to the passive point of view we leave the vector alone but rotate the coordinate system. This is illustrated in Figure 1. The two operations are physically equivalent, and we can choose whichever point of view suits us.

Consider the passive point of view for a moment. How are the components of the vector $\boldsymbol{v}$ in the *new* coordinate system related to those in the *old* coordinate system? In two dimensions we can write

$$\boldsymbol{v} = v_1\hat{\boldsymbol{e}}_1 + v_2\hat{\boldsymbol{e}}_2 = v_1'\hat{\boldsymbol{e}}_1' + v_2'\hat{\boldsymbol{e}}_2' = \boldsymbol{v}'. \tag{1.50}$$

13

By taking dot products we find

$$v'_1 = \boldsymbol{v} \cdot \hat{\boldsymbol{e}}'_1 = v_1(\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}'_1) + v_2(\hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}'_1) \tag{1.51}$$

and

$$v'_2 = \boldsymbol{v} \cdot \hat{\boldsymbol{e}}'_2 = v_1(\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}'_2) + v_2(\hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}'_2). \tag{1.52}$$

It is convenient to express these equations in terms of matrices. Recall that we multiply two matrices using 'row-column' multiplication. If M is an m by p matrix and N is a p by n matrix, then the product matrix $Q := MN$ is an m by n matrix whose ij^{th} entry is the dot product of the ith row of M and the jth column of N. [2] Using indices we can express matrix multiplication as follows:

$$Q_{ij} = (MN)_{ij} = \sum_{k=1}^{n} M_{ik} N_{kj}. \tag{1.53}$$

You should verify that this formula gives the correct answer for matrix multiplication.

With this as background, observe that we can combine (1.51) and (1.52) into a single matrix equation:

$$\begin{pmatrix} v'_1 \\ v'_2 \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}'_1 & \hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}'_1 \\ \hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}'_2 & \hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{e}}'_2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{1.54}$$

The 2×2 matrix appearing in (1.54), which we call R, is an example of a *rotation matrix*. Letting $\boldsymbol{v}$ and $\boldsymbol{v}'$ denote the column vectors on either side of R, we can rewrite (1.54) as

$$\boldsymbol{v}' = R\boldsymbol{v}. \tag{1.55}$$

---

[2]This is why M must have the same number of columns as N has rows.

In terms of components, (1.55) becomes

$$v_i' = \sum_j R_{ij} v_j. \qquad (1.56)$$

The matrix R is the mathematical representation of the planar rotation. Examining Figure 1, we see from (1.49) and (1.54) that the entries of R are simply related to the angle of rotation by

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \qquad (1.57)$$

According to the active point of view, R represents a rotation of all the vectors through an angle θ in the counterclockwise direction. In this case the vector $\boldsymbol{v}$ is rotated to a new vector $\boldsymbol{v}'$ with components v_1' and v_2' *in the old coordinate system.*

According to passive point of view, R represents a rotation of the coordinate system through an angle θ in the *clockwise* direction. In this case the vector $\boldsymbol{v}$ remains unchanged, and the numbers v_1' and v_2' represent the components of $\boldsymbol{v}$ *in the new coordinate system.*

Again, it makes no difference which interpretation you use, but to avoid confusion you should stick to one interpretation for the duration of any problem! (In fact, as long as you just stick to the mathematics, you can usually avoid committing yourself to one interpretation or another.)

We note two important properties of the rotation matrix in (1.57):

$$R^T R = I \qquad (1.58)$$

$$\det R = 1 \qquad (1.59)$$

15

Equation (1.59) just means that the matrix has unit determinant. [3] In (1.58) R^T means the **transpose** of R, which is the matrix obtained from R by flipping it about the diagonal running from NW to SE, and I denotes the **identity** matrix, which consists of ones along the diagonal and zeros elsewhere.

It turns out that these two properties are satisfied by *any* rotation matrix. To see this, we must finally define what we mean by a rotation. The definition is best understood by thinking of a rotation as an active transformation.

Definition. *A* **rotation** *is a linear map taking vectors to vectors that preserves lengths, angles, and handedness.*

The handedness condition says that a rotation must map a right handed coordinate system to a right handed coordinate system. The first two properties can be expressed mathematically by saying that *rotations leave the dot product of two vectors invariant.* For, if $\boldsymbol{v}$ is mapped to $\boldsymbol{v}'$ by a rotation R and $\boldsymbol{w}$ is mapped to $\boldsymbol{w}'$ by R, then we must have

$$\boldsymbol{v}' \cdot \boldsymbol{w}' = \boldsymbol{v} \cdot \boldsymbol{w}. \tag{1.60}$$

This is because, if we set $\boldsymbol{w} = \boldsymbol{v}$ then (1.60) says that $v'^2 = v^2$ (where $v' = |\boldsymbol{v}'|$ and $v = |\boldsymbol{v}|$), so the length of $\boldsymbol{v}'$ is the same as the length of $\boldsymbol{v}$ (and similarly, the length of $\boldsymbol{w}'$ is the same as the length of $\boldsymbol{w}$), and if $\boldsymbol{w} \neq \boldsymbol{v}$ then (1.60) says that $v'w'\cos\theta' = vw\cos\theta$, which, because the lengths are the same, implies that the angle between $\boldsymbol{v}'$ and $\boldsymbol{w}'$ is the same as the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$.

Let's see where the condition (1.60) leads. In terms of components we

---

[3]For a review of the determinant and its properties, consult Appendix B.

have

$$\sum_i v_i' w_i' = \sum_i v_i w_i$$

$$\implies \quad \sum_{ijk}(R_{ij}v_j)(R_{ik}w_k) = \sum_{jk}\delta_{jk}v_j w_k$$

$$\implies \quad \sum_{jk}\left(\sum_i R_{ij}R_{ik} - \delta_{jk}\right)v_j w_k = 0$$

As the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ are arbitrary, we can conclude

$$\sum_i R_{ij}R_{ik} = \delta_{jk}. \tag{1.61}$$

Note that the components of the transposed matrix R^T are obtained from those of R by switching indices. That is, $(R^T)_{ij} = R_{ji}$. Hence (1.61) can be written

$$\sum_i (R^T)_{ji}R_{ik} = \delta_{jk}. \tag{1.62}$$

Comparing this equation to (1.53) we see that it can be written

$$R^T R = I. \tag{1.63}$$

Thus we see that the condition (1.63) is just another way of saying that lengths and angles are preserved by a rotation. Incidentally, yet another way of expressing (1.63) is

$$R^T = R^{-1}, \tag{1.64}$$

where R^{-1} is the **matrix inverse** of R. [4]

---

[4]The inverse A^{-1} of a matrix A satisfies $AA^{-1} = A^{-1}A = I$.

Now, it is a fact that, for any two square matrices A and B,

$$\det AB = \det A \det B. \tag{1.65}$$

and

$$\det A^T = \det A, \tag{1.66}$$

(see Appendix B). Applying the two properties (1.65) and (1.66) to (1.63) gives

$$(\det R)^2 = 1 \quad \Rightarrow \quad \det R = \pm 1. \tag{1.67}$$

Thus, if R preserves lengths and angles then it is *almost* a rotation. It is a rotation if $\det R = 1$, which is the condition of preserving handedness, and it is a *roto-reflection* (product of a rotation and a reflection) if $\det R = -1$. The set of all linear transformations R satisfying (1.63) is called the **orthogonal group**, and the subset satisfying $\det R = 1$ is called the **special orthogonal group**.

1.8 Vector Algebra II: Cross Products and the Levi Civita Symbol

We have discussed the dot product, which is a way of forming a scalar from two vectors. There are other sorts of vector products, two of which are particularly relevant to physics. They are the *vector* or *cross* product, and the *dyadic* or *tensor* product.

First we discuss the cross product. Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be given, and define the

cross product $\boldsymbol{B} \times \boldsymbol{C}$ in terms of the following determinant: [5]

$$\boldsymbol{B} \times \boldsymbol{C} = \begin{vmatrix} \hat{\boldsymbol{e}}_1 & \hat{\boldsymbol{e}}_2 & \hat{\boldsymbol{e}}_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}$$

$$= (B_2 C_3 - B_3 C_2)\hat{\boldsymbol{e}}_1 + (B_3 C_1 - B_1 C_3)\hat{\boldsymbol{e}}_2 + (B_1 C_2 - B_2 C_1)\hat{\boldsymbol{e}}_3$$

$$= (B_2 C_3 - B_3 C_2)\hat{\boldsymbol{e}}_1 + \text{cyclic}. \tag{1.68}$$

It is clear from the definition that the cross product is **antisymmetric**, meaning that it flips sign if you flip the vectors:

$$\boldsymbol{B} \times \boldsymbol{C} = -\boldsymbol{C} \times \boldsymbol{B}. \tag{1.69}$$

Just as the dot product admits a geometric interpretation, so does the cross product: the length of $\boldsymbol{B} \times \boldsymbol{C}$ is the area of the parallelogram spanned by the vectors $\boldsymbol{B}$ and $\boldsymbol{C}$, and $\boldsymbol{B} \times \boldsymbol{C}$ points orthogonally to the parallelogram in the direction given by the right hand rule. [6] We see this as follows. Let θ be the angle between $\boldsymbol{B}$ and $\boldsymbol{C}$. We can always rotate our vectors (or else our coordinate system) so that $\boldsymbol{B}$ lies along the x-axis and $\boldsymbol{C}$ lies somewhere in the xy plane. Then we have (see Figure 2):

$$\boldsymbol{B} = B\hat{\boldsymbol{e}}_1 \qquad \text{and} \qquad \boldsymbol{C} = C(\cos\theta\hat{\boldsymbol{e}}_1 + \sin\theta\hat{\boldsymbol{e}}_2), \tag{1.70}$$

so that

$$\boldsymbol{B} \times \boldsymbol{C} = BC\hat{\boldsymbol{e}}_1 \times (\cos\theta\hat{\boldsymbol{e}}_1 + \sin\theta\hat{\boldsymbol{e}}_2) = BC\sin\theta\hat{\boldsymbol{e}}_3. \tag{1.71}$$

---

[5]The word 'cyclic' means that the other terms are obtained from the first term by successive cyclic permutation of the indices $1 \to 2 \to 3$. For a brief discussion of permutations, see Appendix A.

[6]To apply the right hand rule, point your hand in the direction of $\boldsymbol{B}$ and close it in the direction of $\boldsymbol{C}$. Your thumb will then point in the direction of $\boldsymbol{B} \times \boldsymbol{C}$.
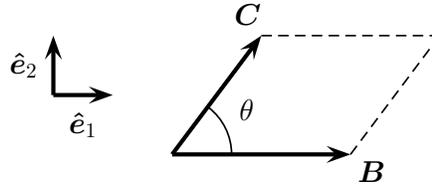
Figure 2: Two vectors spanning a parallelogram
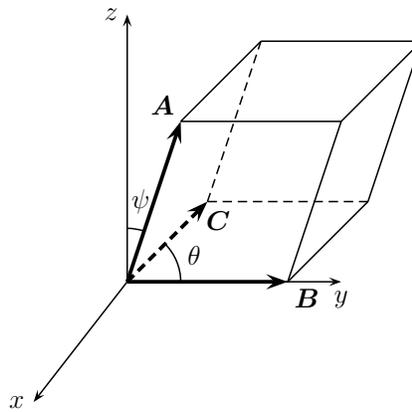
Figure 3: Three vectors spanning a parallelepiped

The direction is consistent with the right hand rule, and the magnitude,

$$|\boldsymbol{B} \times \boldsymbol{C}| = BC \sin \theta, \qquad (1.72)$$

is precisely the area of the parallelogram spanned by $\boldsymbol{B}$ and $\boldsymbol{C}$, as promised.

We can now combine the geometric interpretation of the dot and cross products to get a geometric interpretation of the **triple product $\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$**: it is the volume of the parallelepiped spanned by all three vectors. Suppose $\boldsymbol{A}$ lies in the yz-plane and is inclined at an angle ψ relative to the z-axis, and that $\boldsymbol{B}$ and $\boldsymbol{C}$ lie in the xy-plane, separated by an angle θ, as shown in Figure 3. Then

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = A(\cos\psi\,\hat{\boldsymbol{e}}_3 + \sin\psi\,\hat{\boldsymbol{e}}_1) \cdot BC\sin\theta\,\hat{\boldsymbol{e}}_3$$

$$= ABC\sin\theta\cos\psi$$

$$= \text{volume of parallelepiped}$$

$$= \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}, \tag{1.73}$$

where the last equality follows by taking the dot product of $\boldsymbol{A}$ with the cross product $\boldsymbol{B} \times \boldsymbol{C}$ given in (1.68). Since the determinant flips sign if two rows are interchanged, the triple product is invariant under cyclic permutations:

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B} \cdot (\boldsymbol{C} \times \boldsymbol{A}) = \boldsymbol{C} \cdot (\boldsymbol{A} \times \boldsymbol{B}). \tag{1.74}$$

It turns out to be convenient, when dealing with cross products, to define a new object that packages all the minus signs of a determinant in a convenient fashion. This object is called the *Levi Civita Alternating Symbol*. (It is also called a *permutation symbol* or the *epsilon symbol*. We will use any of these terms as suits us.)

Formally, the Levi Civita alternating symbol ε_{ijk} is a three-indexed object with the following two defining properties:

i) $\varepsilon_{123} = 1$.

ii) ε_{ijk} changes sign whenever any two indices are interchanged.

These two properties suffice to fix every value of the epsilon symbol. *A priori* there are 27 possible values for ε_{ijk}, one for each choice of i, j, and k, each of which runs from 1 to 3. But the defining conditions eliminate most of them. For example, consider ε_{122}. By property (ii) above, it should flip sign when

we flip the last two indices. But then we have $\varepsilon_{122} = -\varepsilon_{122}$, and the only number that is equal to its negative is zero. Hence $\varepsilon_{122} = 0$. Similarly, it follows that ε_{ijk} is zero whenever any two indices are the same.

This means that, of the 27 possible values we started with, only 6 of them can be nonzero, namely those whose indices are permutations of (123). These nonzero values are determined by properties (i) and (ii) above. So, for example, $\varepsilon_{312} = 1$, because we can get from ε_{123} to ε_{312} by two index flips:

$$\varepsilon_{312} = -\varepsilon_{132} = +\varepsilon_{123} = +1. \tag{1.75}$$

A moment's thought should convince you that the epsilon symbol gives us the sign of the permutation of its indices, where the sign of a permutation is just -1 raised to the power of the number of flips of the permuation from the identity permutation (123). This explains its name 'permutation symbol'.

The connection between cross products and the alternating symbol is *via* the following formula:

$$(\boldsymbol{A} \times \boldsymbol{B})_i = \sum_{jk} \varepsilon_{ijk} A_j B_k. \tag{1.76}$$

To illustrate, let us choose $i = 1$. Then, written out in full, (1.76) reads

$$\begin{aligned}
(\boldsymbol{A} \times \boldsymbol{B})_1 &= \varepsilon_{111} A_1 B_1 + \varepsilon_{112} A_1 B_2 + \varepsilon_{113} A_1 B_3 \\
&\quad + \varepsilon_{121} A_2 B_1 + \varepsilon_{122} A_2 B_2 + \varepsilon_{123} A_2 B_3 \\
&\quad + \varepsilon_{131} A_3 B_1 + \varepsilon_{132} A_3 B_2 + \varepsilon_{133} A_3 B_3 \\
&= A_2 B_3 - A_3 B_2, \tag{1.77}
\end{aligned}$$

where the last equality follows by substituting in the values of the epsilon

symbols. You should check that the other two components of the cross product are given correctly as well. Observe that, using the summation convention, (1.76) would be written

$$(\boldsymbol{A} \times \boldsymbol{B})_i = \varepsilon_{ijk} A_j B_k. \tag{1.78}$$

Note also that, due to the symmetry properties of the epsilon symbol, we could also write

$$(\boldsymbol{A} \times \boldsymbol{B})_i = \varepsilon_{jki} A_j B_k. \tag{1.79}$$

1.9 Products of Epsilon Symbols

There are four important product identities involving epsilon symbols. They are (using the summation convention throughout):

$$\varepsilon_{ijk} \varepsilon_{mnp} = \begin{vmatrix} \delta_{im} & \delta_{in} & \delta_{ip} \\ \delta_{jm} & \delta_{jn} & \delta_{jp} \\ \delta_{km} & \delta_{kn} & \delta_{kp} \end{vmatrix} \tag{1.80}$$

$$\varepsilon_{ijk} \varepsilon_{mnk} = \delta_{im} \delta_{jn} - \delta_{in} \delta_{jm} \tag{1.81}$$

$$\varepsilon_{ijk} \varepsilon_{mjk} = 2 \delta_{im} \tag{1.82}$$

$$\varepsilon_{ijk} \varepsilon_{ijk} = 3!. \tag{1.83}$$

The proofs of these identities are left as an exercise. To get you started, let's prove (1.82). To begin, you must figure out which indices are free and which are summed. Well, j and k are repeated on the left hand side, so they are summed over, while i and m are both on opposite sides of the equation, so they are free. This means (1.82) represents nine equations, one for each possible pair of values for i and m. To prove the formula, we have to show

23

that, no matter what values of i and m we choose, the left side is equal to the right side.

So let's pick $i = 1$ and $m = 2$, say. Then by the definition of the Kronecker delta, the right hand side is zero. This means we must show the left hand side is also zero. For clarity, let us write out the left hand side in this case (remember, j and k are summed over, while i and m are fixed):

$$\varepsilon_{111}\varepsilon_{211} + \varepsilon_{112}\varepsilon_{212} + \varepsilon_{113}\varepsilon_{213}$$
$$+ \varepsilon_{121}\varepsilon_{221} + \varepsilon_{122}\varepsilon_{222} + \varepsilon_{123}\varepsilon_{223}$$
$$+ \varepsilon_{131}\varepsilon_{231} + \varepsilon_{132}\varepsilon_{232} + \varepsilon_{133}\varepsilon_{233}.$$

If you look carefully at this expression, you will see that it is always zero! The reason is that, in order to get something nonzero, at least one summand must be nonzero. But each summand is the product of two epsilon symbols, and because i and m are different, these two epsilon symbols are never simultaneously nonzero. The only time the first epsilon symbol in a term is nonzero is when the pair (j, k) is $(2, 3)$ or $(3, 2)$. But then the second epsilon symbol must vanish, as it has at least two 2s. A similar argument shows that the left hand side vanishes whenever i and m are different, and as the right hand side also vanishes under these circumstances, the two sides are always equal whenever i and m are different.

What if $i = m$? In that case the left side is 2, because the sum includes precisely two nonzero summands, each of which has the value 1. For example, if $i = m = 1$, the two nonzero terms in the sum are $\varepsilon_{123}\varepsilon_{123}$ and $\varepsilon_{132}\varepsilon_{132}$, each of which is 1. But the right hand side is also 2, by the properties of the Kronecker delta. Hence the equation holds.

In general, this is a miserable way to prove the identities above, because

you have to consider all these cases. The better way is to derive (1.81), (1.82), and (1.83) from (1.80) (which I like to call "the mother of all epsilon identities"). (The derivation of (1.80) proceeds by comparing the symmetry properties of both sides.) To demonstrate how this works, consider obtaining (1.82) from (1.81). Observe that (1.81) represents $3^4 = 81$ equations, as i, j, m, and n are free (only k is summed). We want to somehow relate it to (1.82). This means we need to set n equal to j and sum over j. We are able to do this because (1.81) remains true for any values of j and n. So it certainly is true if $n = j$, and summing true equations produces another true equation. If we do this (which, by the way, is called *contracting the indices j and n*) we get the left hand side of (1.82). So we must show that doing the same thing to the right hand side of (1.81) (namely, setting $n = j$ and summing over j) yields the right hand side of (1.82). If we can do this we will have completed our proof that (1.82) follows from (1.81).

So, we must show that

$$\delta_{im}\delta_{jj} - \delta_{jm}\delta_{ij} = 2\delta_{im}. \tag{1.84}$$

Perhaps it would be a little more clear if we restored the summation symbols, giving

$$\sum_j \delta_{im}\delta_{jj} - \sum_j \delta_{jm}\delta_{ij} = 2\delta_{im}. \tag{1.85}$$

The first sum is over j, so we may pull out the δ_{im} term, as it is independent of j. Using the properties of the Kronecker delta, we see that $\sum_j \delta_{jj} = 3$. So the first term is just $3\delta_{im}$. The second term is just δ_{im}, using the substitution property of the Kronecker delta. Hence the two sides are equal, as desired.

Example 1   The following computation illustrates the utility of the formulae

(1.80)-(1.83). The objective is to prove the vector identity

$$\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B}), \tag{1.86}$$

the so-called "BAC minus CAB rule".

We proceed as follows (summation convention in force):

$$(\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}))_i = \varepsilon_{ijk} A_j (\boldsymbol{B} \times \boldsymbol{C})_k \tag{1.87}$$

$$= \varepsilon_{ijk} A_j \varepsilon_{klm} B_l C_m \tag{1.88}$$

$$= (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}) A_j B_l C_m \tag{1.89}$$

$$= A_j B_i C_j - A_j B_j C_i \tag{1.90}$$

$$= (\boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B}))_i. \tag{1.91}$$

and we are done.

This was a little fast, perhaps. So let us fill in a few of the steps. Observe that we choose to prove that the left and right hand sides of (1.86) are the same by proving their components are the same. This makes sense according to the way in which we introduced cross products *via* epsilon symbols.

Equation (1.87) is obtained from (1.78), leaving $\boldsymbol{B} \times \boldsymbol{C}$ temporarily unexpanded. In (1.88) we apply (1.78) again, this time to $\boldsymbol{B} \times \boldsymbol{C}$. Notice that we had to choose different dummy indices for the second epsilon expansion, otherwise we would have gotten into trouble, as we have emphasized previously. In (1.89) we did a few things all at once. First, we commuted the A_j and ε_{klm} terms. We can always do this because, for any value of the indices, these two quantities are just numbers, and numbers always commute. Second, we permuted some indices in our head in order to bring the index structure of the epsilon product into the form exhibited in (1.81). In particular, we substituted ε_{lmk} for ε_{klm}, which we can do by virtue of the symmetry properties of the epsilon symbol. Third, we applied

(1.81) to the product $\varepsilon_{ijk}\varepsilon_{lmk}$. To get (1.90) we used the substitution property of the Kronecker delta. Finally, we recognized that $A_j C_j$ is just $\boldsymbol{A} \cdot \boldsymbol{C}$, and $A_j B_j$ is $\boldsymbol{A} \cdot \boldsymbol{B}$. The equality of (1.90) and (1.91) is precisely the definition of the ith component of the right hand side of (1.86). The result then follows because two vectors are equal if and only if their components are equal.

1.10   Determinants and Epsilon Symbols

Given the close connection between cross products and determinants, it should come as no surprise that there are formulas relating determinants to epsilon symbols. Consider again the triple product (1.73). Using the epsilon symbol we can write

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \sum_k A_k (\boldsymbol{B} \times \boldsymbol{C})_k = \sum_{ijk} \varepsilon_{kij} A_k B_i C_j = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} \quad (1.92)$$

Thus,

$$\det A = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \sum_{ijk} \varepsilon_{ijk} A_{1i} A_{2j} A_{3k}. \quad (1.93)$$

We could just as well multiply on the left by ε_{123}, because $\varepsilon_{123} = 1$, in which case (1.93) would read

$$\varepsilon_{123} \det A = \sum_{ijk} \varepsilon_{ijk} A_{1i} A_{2j} A_{3k}. \quad (1.94)$$

As the determinant changes sign whenever any two rows of the matrix are

switched, it follows that the right hand side has exactly the same symmetries as the left hand side under any interchange of 1, 2, and 3. Hence we may write

$$\varepsilon_{mnp} \det A = \sum_{ijk} \varepsilon_{ijk} A_{mi} A_{nj} A_{pk}. \tag{1.95}$$

Again, using our summation convention, this would be written

$$\varepsilon_{mnp} \det A = \varepsilon_{ijk} A_{mi} A_{nj} A_{pk}. \tag{1.96}$$

Finally, we can transform (1.96) into a more symmetric form by using property (1.83). Multiply both sides by ε_{mnp}, sum over m, n, and p, and divide by 3! to get [7]

$$\det A = \frac{1}{3!} \varepsilon_{mnp} \varepsilon_{ijk} A_{mi} A_{nj} A_{pk}. \tag{1.97}$$

1.11   Vector Algebra III: Tensor Product

So, what is a tensor anyway? There are many different ways to introduce the notion of a tensor, varying from what some mathematicians amusingly call "low brow" to "high brow". In keeping with the discursive nature of these notes, I will restrict the discussion to the "low brow" approach, reserving a more advanced treatment for later work.

To start, we define a new kind of vector product called the **tensor product**, usually denoted by the symbol $\otimes$. Given two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$, we can form their tensor product $\boldsymbol{A} \otimes \boldsymbol{B}$. $\boldsymbol{A} \otimes \boldsymbol{B}$ is called a *tensor of order 2.* [8] The tensor product is **not** generally commutative—order matters. So

---

[7]Because we have restricted attention to the three dimensional epsilon symbol, the formulae in this section work only for 3×3 matrices. One can write formulae for higher determinants using higher dimensional epsilon symbols, but we shall not do so here.

[8]N.B. Many people use the word 'rank' interchangeably with the word 'order', so that $\boldsymbol{A} \otimes \boldsymbol{B}$ is then called a tensor of rank 2. The problem with this terminology is that it

$\boldsymbol{B} \otimes \boldsymbol{A}$ is generally different from $\boldsymbol{A} \otimes \boldsymbol{B}$. We can form higher order tensors by repeating this procedure. So, for example, given another vector $\boldsymbol{C}$, we have $\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C}$, a third order tensor. (The tensor product is associative, so we need not worry about parentheses.) Order zero tensors are just scalars, while order one tensors are just vectors.

In older books, the tensor $\boldsymbol{A} \otimes \boldsymbol{B}$ is sometimes called a *dyadic* product (of the vectors $\boldsymbol{A}$ and $\boldsymbol{B}$), and is written $\boldsymbol{A}\boldsymbol{B}$. That is, the tensor product symbol $\otimes$ is simply dropped. This generally leads to no confusion, as the only way to understand the proximate juxtaposition of two vectors is as a tensor product. We will use either notation as it suits us.

The set of all tensors forms a mathematical object called a **graded algebra**. This just means that you can add and multiply as usual. For example, if α and β are numbers and $\boldsymbol{S}$ and $\boldsymbol{T}$ are both tensors of order s, then $\alpha \boldsymbol{T} + \beta \boldsymbol{S}$ is a tensor of order s. If $\boldsymbol{R}$ is a tensor of order r then $\boldsymbol{R} \otimes \boldsymbol{S}$ is a tensor of order $r + s$. In addition, scalars pull through tensor products

$$\boldsymbol{T} \otimes (\alpha \boldsymbol{S}) = (\alpha \boldsymbol{T}) \otimes \boldsymbol{S} = \alpha(\boldsymbol{T} \otimes \boldsymbol{S}), \qquad (1.98)$$

and tensor products are distributive over addition:

$$\boldsymbol{R} \otimes (\boldsymbol{S} + \boldsymbol{T}) = \boldsymbol{R} \otimes \boldsymbol{S} + \boldsymbol{R} \otimes \boldsymbol{T}. \qquad (1.99)$$

Just as a vector has components in some basis, so does a tensor. Let

---

conflicts with another standard usage. In linear algebra the rank of a matrix is the number of linearly independent rows (or columns). If we consider the components of the tensor $\boldsymbol{A} \otimes \boldsymbol{B}$, namely $A_i B_j$, to be the components of a matrix, then this matrix only has rank 1! (The rows are all multiples of each other.) To avoid this problem, one usually says that a tensor of the form $\boldsymbol{A}_1 \otimes \boldsymbol{A}_2 \otimes \cdots$ has rank 1. Any tensor is a sum of rank 1 tensors, and we say that the rank of the tensor is the minimum number of rank 1 tensors needed to write it as such a sum.

$\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \hat{\boldsymbol{e}}_3$ be the canonical basis of $\mathbb{R}^3$. Then the canonical basis for the vector space $\mathbb{R}^3 \otimes \mathbb{R}^3$ of order 2 tensors on $\mathbb{R}^3$ is given by the set $\hat{\boldsymbol{e}}_i \otimes \hat{\boldsymbol{e}}_j$, as i and j run from 1 to 3. Written out in full, these basis elements are

$$
\begin{aligned}
&\hat{\boldsymbol{e}}_1 \otimes \hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{e}}_1 \otimes \hat{\boldsymbol{e}}_2 \quad \hat{\boldsymbol{e}}_1 \otimes \hat{\boldsymbol{e}}_3 \\
&\hat{\boldsymbol{e}}_2 \otimes \hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{e}}_2 \otimes \hat{\boldsymbol{e}}_2 \quad \hat{\boldsymbol{e}}_2 \otimes \hat{\boldsymbol{e}}_3. \\
&\hat{\boldsymbol{e}}_3 \otimes \hat{\boldsymbol{e}}_1 \quad \hat{\boldsymbol{e}}_3 \otimes \hat{\boldsymbol{e}}_2 \quad \hat{\boldsymbol{e}}_3 \otimes \hat{\boldsymbol{e}}_3
\end{aligned}
\tag{1.100}
$$

The most general second order tensor on $\mathbb{R}^3$ is a linear combination of these basis tensors:

$$
\boldsymbol{T} = \sum_{ij} T_{ij} \hat{\boldsymbol{e}}_i \otimes \hat{\boldsymbol{e}}_j.
\tag{1.101}
$$

Almost always the basis is understood and fixed throughout. For this reason, tensors are often identified with their components. So, for example, we often do not distinguish between the vector $\boldsymbol{A}$ and its components A_i. Similarly, we often call T_{ij} a tensor, when it is really just the components of a tensor in some basis. This terminology drives mathematicians crazy, but it works for most physicists. This is the reason why we have already referred to the Kronecker delta δ_{ij} and the epsilon tensor ε_{ijk} as 'tensors'.

As an example, let us find the components of the tensor $\boldsymbol{A} \otimes \boldsymbol{B}$. We have

$$
\boldsymbol{A} \otimes \boldsymbol{B} = \left(\sum_i A_i \hat{\boldsymbol{e}}_i\right) \otimes \left(\sum_j B_j \hat{\boldsymbol{e}}_j\right)
\tag{1.102}
$$

$$
= \sum_{ij} A_i B_j \hat{\boldsymbol{e}}_i \otimes \hat{\boldsymbol{e}}_j,
\tag{1.103}
$$

so the components of $\boldsymbol{A} \otimes \boldsymbol{B}$ are just $A_i B_j$. This works in general, so that, for example, the components of $\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C}$ are just $A_i B_j C_k$.

It is perhaps worth observing that a tensor of the form $\boldsymbol{A} \otimes \boldsymbol{B}$ for some
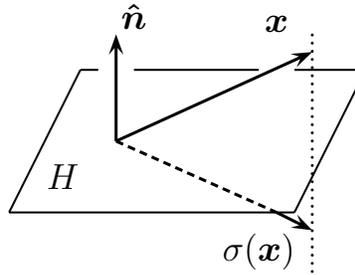
Figure 4: Reflection through a plane

vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is **not** the most general order two tensor. The reason is that the most general order two tensor has 9 independent components, whereas $A_i B_j$ has only 6 independent components (three from each vector).

1.12   Problems

1) The *Cauchy-Schwarz inequality* states that, for any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$:
$$(\boldsymbol{u} \cdot \boldsymbol{v})^2 \leq (\boldsymbol{u} \cdot \boldsymbol{u})(\boldsymbol{v} \cdot \boldsymbol{v}),$$
with equality holding if and only if $\boldsymbol{u} = \lambda \boldsymbol{v}$ for some $\lambda \in \mathbb{R}$. Prove the Cauchy-Schwarz inequality. [Hint: Use angles.]

2) Show that the equation $\boldsymbol{a} \cdot \boldsymbol{r} = a^2$ defines a two dimensional plane in three dimensional space, where $\boldsymbol{a}$ is the minimal length vector from the origin to the plane. [Hint: A plane is the translate of the linear span of two vectors. The Cauchy-Schwarz inequality may come in handy.]

3) A *reflection* σ through a plane H with unit normal vector $\hat{\boldsymbol{n}}$ is a linear map satisfying (i) $\sigma(\boldsymbol{x}) = \boldsymbol{x}$, for $\boldsymbol{x} \in H$, and (ii) $\sigma(\hat{\boldsymbol{n}}) = -\hat{\boldsymbol{n}}$. (See Figure 4.) Find an expression for $\sigma(\boldsymbol{x})$ in terms of $\boldsymbol{x}$, $\hat{\boldsymbol{n}}$, and the dot product. Verify that $\sigma^2 = 1$, as befits a reflection.

4) The volume of a tetrahedron is $V = bh/3$, where b is the area of a base and h is the height (distance from base to apex). Consider a tetrahedron with one vertex at the origin and the other three vertices at positions $\boldsymbol{A}$, $\boldsymbol{B}$ and

$\boldsymbol{C}$. Show that we can write

$$V = \frac{1}{6}\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}).$$

This demonstrates that the volume of such a tetrahedron is one sixth of the volume of the parallelepiped defined by the vectors $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$.

5) Prove Equation (1.80) by the following method. First, show that both sides have the same symmetry properties by showing that both sides are anti-symmetric under the interchange of a pair of $\{ijk\}$ or a pair of $\{mnp\}$, and that both sides are left invariant if you exchange the sets $\{ijk\}$ and $\{mnp\}$. Next, show that both sides agree when $(i, j, k, m, n, p) = (1, 2, 3, 1, 2, 3)$.

6) Using index notation, prove Lagrange's identity:

$$(\boldsymbol{A} \times \boldsymbol{B}) \cdot (\boldsymbol{C} \times \boldsymbol{D}) = (\boldsymbol{A} \cdot \boldsymbol{C})(\boldsymbol{B} \cdot \boldsymbol{D}) - (\boldsymbol{A} \cdot \boldsymbol{D})(\boldsymbol{B} \cdot \boldsymbol{C}).$$

7) For any two matrices A and B, show that $(AB)^T = B^T A^T$ and $(AB)^{-1} = B^{-1}A^{-1}$. [Hint: You may wish to use indices for the first equation, but for the second use the uniqueness of the inverse.]

8) Let $R(\theta)$ and $R(\psi)$ be planar rotations through angles θ and ψ, respectively. By explicitly multiplying the matrices together, show that $R(\theta)R(\psi) = R(\theta + \psi)$.

[Remark: This makes sense physically, because it says that if we first rotate a vector through an angle ψ and then rotate it through an angle θ, that the result is the same as if we simply rotated it through a total angle of $\theta + \psi$. Incidentally, this shows that planar rotations *commute*, which means that we get the same result whether we first rotate through ψ then θ, or first rotate through θ then ψ, as one would expect. This is no longer true for rotations in three and higher dimensions where the order of rotations matters, as you can see by performing successive rotations about different axes, first in one order and then in the opposite order.]

2 Vector Calculus I

It turns out that the laws of physics are most naturally expressed in terms of tensor fields, which are simply fields of tensors. We have already seen many
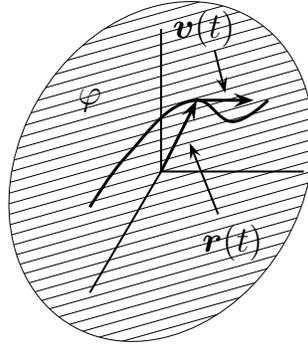
Figure 5: An observer moving along a curve through a scalar field

examples of this in the case of scalar fields and vector fields, and tensors are just a natural generalization. But in physics we are not just interested in how things *are*, we are also interested in how things *change*. For that we need to introduce the language of change, namely calculus. This leads us to the topic of tensor calculus. However, we will restrict ourselves here to tensor fields of order 0 and 1 (scalar fields and vector fields) and leave the general case for another day.

2.1   Fields

A **scalar field** $\varphi(\boldsymbol{r})$ is a field of scalars. This means that, to every point $\boldsymbol{r}$ we associate a scalar quantity $\varphi(\boldsymbol{r})$. A physical example is the electrostatic potential. Another example is the temperature.

A **vector field** $\boldsymbol{A}(\boldsymbol{r})$ is a field of vectors. This means that, to every point $\boldsymbol{r}$ we associate a vector $\boldsymbol{A}(\boldsymbol{r})$. A physical example is the electric field. Another example is the gravitational field.

2.2 The Gradient

Consider an observer moving through space along a parameterized curve $\boldsymbol{r}(t)$ in the presence of a scalar field $\varphi(\boldsymbol{r})$. According to the observer, how fast is φ changing? For convenience we work in Cartesian coordinates, so that the position of the observer at time t is given by

$$\boldsymbol{r}(t) = (x(t), y(t), z(t)). \tag{2.1}$$

At this instant the observer measures the value

$$\varphi(t) := \varphi(\boldsymbol{r}(t)) = \varphi(x(t), y(t), z(t)) \tag{2.2}$$

for the scalar field φ. Thus $d\varphi/dt$ measures the rate of change of φ along the curve. By the chain rule this is

$$\begin{aligned}
\frac{d\varphi(t)}{dt} &= \frac{\partial\varphi}{\partial x}\frac{dx}{dt} + \frac{\partial\varphi}{\partial y}\frac{dy}{dt} + \frac{\partial\varphi}{\partial z}\frac{dz}{dt} \\
&= \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \cdot \left(\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}, \frac{\partial\varphi}{\partial y}\right) \\
&= \boldsymbol{v}\cdot\boldsymbol{\nabla}\varphi,
\end{aligned} \tag{2.3}$$

where

$$\boldsymbol{v}(t) = \frac{d\boldsymbol{r}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \tag{2.4}$$

is the **velocity vector** of the particle. The quantity $\boldsymbol{\nabla}$ is called the **gradient operator**. We interpret Equation (2.3) by saying that the rate of change of φ in the direction $\boldsymbol{v}$ is

$$\frac{d\varphi}{dt} = \boldsymbol{v}\cdot(\boldsymbol{\nabla}\varphi) = (\boldsymbol{v}\cdot\boldsymbol{\nabla})\varphi. \tag{2.5}$$
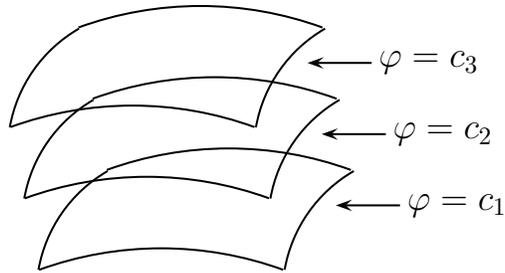
Figure 6: Some level surfaces of a scalar field φ

The latter expression is called the **directional derivative** of φ in the direction $\boldsymbol{v}$.

We can understand the gradient operator in another way.

Definition. *A **level surface** (or **equipotential surface**) of a scalar field $\varphi(\boldsymbol{r})$ is the locus of points $\boldsymbol{r}$ for which $\varphi(\boldsymbol{r})$ is constant. (See Figure 6.)*

With this definition we make the following

Claim 2.1. $\boldsymbol{\nabla}\varphi$ *is a vector field that points everywhere orthogonal to the level surfaces of φ and in the direction of fastest increase of φ.*

Proof. Pick a point in a level surface and suppose that $\boldsymbol{\nabla}\varphi$ fails to be orthogonal to the level surface at that point. Consider moving along a curve lying within the level surface (Figure 7). Then $\boldsymbol{v} = d\boldsymbol{r}/dt$ is tangent to the surface, which implies that

$$\frac{d\varphi}{dt} = \boldsymbol{v} \cdot \boldsymbol{\nabla}\varphi \neq 0,$$

a contradiction. Also, $d\varphi/dt$ is positive when moving from low φ to high φ, so $\boldsymbol{\nabla}\varphi$ must point in the direction of increase of φ. $\qquad\square$
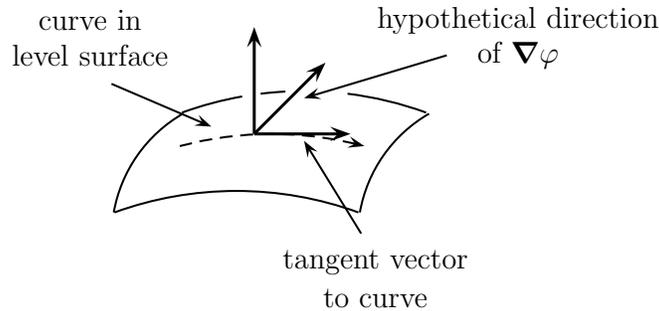
Figure 7: Gradients and level surfaces

Example 2 Let $T = x^2 - y^2 + z^2 - 2xy + 2yz + 273$. Suppose you are at the point $(3, 1, 4)$. Which way does it feel hottest? What is the rate of increase of the temperature in the direction $(1, 1, 1)$ at this point?

We have

$$\left.\boldsymbol{\nabla} T\right|_{(3,14)} = \left.\left(2(x - y), -2y - 2x + 2z, 2z + 2y\right)\right|_{(3,1,4)} = (4, 0, 10).$$

Since

$$\frac{dT}{dt} = \boldsymbol{v} \cdot \boldsymbol{\nabla}\varphi,$$

the rate of increase in the temperature depends on the speed of the observer. If we want to compute the rate of temperature increase independent of the speed of the observer we must normalize the direction vector. This gives, for the rate of increase

$$\frac{1}{\sqrt{3}}(1, 1, 1) \cdot (4, 0, 10) = 8.1.$$

If temperature were measured in Kelvins and distance in meters this last answer would be in K/m.
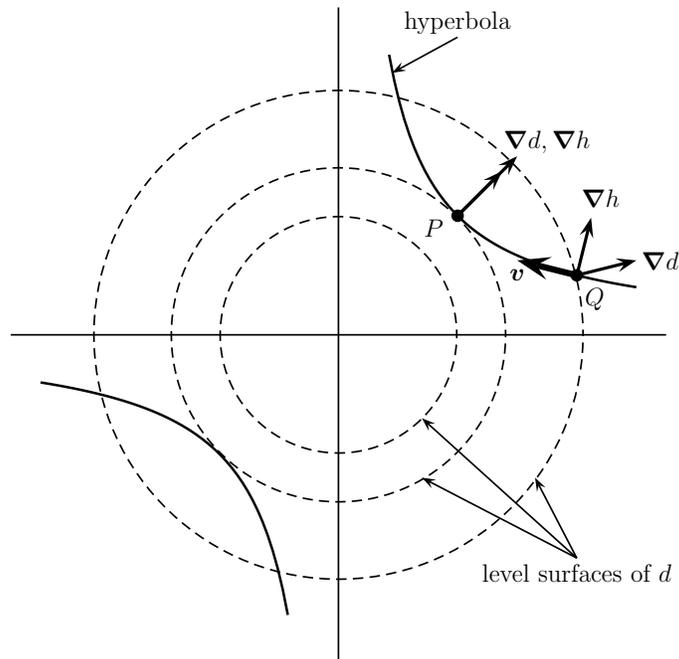
Figure 8: A hyperbola meets some level surfaces of d

2.3   Lagrange Multipliers

One important application of the gradient operator is to constrained optimization problems. Let's consider a simple example first. We would like to find the point (or points) on the hyperbola $xy = 4$ closest to the origin. (See Figure 8.) Of course, this is geometrically obvious, but we will use the method of Lagrange multipliers to illustrate the general method, which is applicable in more involved cases.

The distance from any point (x, y) to the origin is given by the function $d(x, y) = \sqrt{x^2 + y^2}$. Define $h(x, y) = xy$. Then we want the solution (or solutions) to the problem

minimize $d(x, y)$ subject to the constraint $h(x, y) = 4$.

$d(x, y)$ is called the **objective** function, while $h(x, y)$ is called the **constraint** function.

We can interpret the problem geometrically as follows. The level surfaces of d are circles about the origin, and the direction of fastest increase in d is parallel to $\boldsymbol{\nabla}d$, which is orthogonal to the level surfaces. Now imagine walking along the hyperbola. At a point Q on the hyperbola where $\boldsymbol{\nabla}h$ is not parallel to $\boldsymbol{\nabla}d$, $\boldsymbol{v}$ has a component parallel to $\boldsymbol{\nabla}d$, [9] so we can continue to walk in the direction of the vector $\boldsymbol{v}$ and cause the value of d to decrease. Hence d was not a minimum at Q. Only when $\boldsymbol{\nabla}h$ and $\boldsymbol{\nabla}d$ are parallel (at P) do we reach the minimum of d subject to the constraint. Of course, we have to require $h = 4$ as well (otherwise we might be on some other level surface of h by accident). Hence, the minimum of d subject to the constraint is achieved at a point $\boldsymbol{r}_0 = (x_0, y_0)$, where

$$\boldsymbol{\nabla}d|_{\boldsymbol{r}_0} = \lambda\boldsymbol{\nabla}h|_{\boldsymbol{r}_0} \tag{2.6}$$

$$\text{and} \quad h(\boldsymbol{r}_0) = 4, \tag{2.7}$$

and where λ is some unknown constant, called a **Lagrange multiplier**.

At this point we invoke a small simplification, and change our objective function to $f(x, y) = [d(x, y)]^2 = x^2 + y^2$, because it is easy to see that $d(x, y)$ and $f(x, y)$ are minimized at the same points. So, we want to solve the equations

$$\boldsymbol{\nabla}f = \lambda\boldsymbol{\nabla}h \tag{2.8}$$

$$\text{and} \quad h = 4. \tag{2.9}$$

---

[9]Equivalently, $\boldsymbol{v}$ is not tangent to the level surfaces of d.

38

In our example, these equations become

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial h}{\partial x} \qquad\qquad \Rightarrow 2x = \lambda y \qquad\qquad (2.10)$$

$$\frac{\partial f}{\partial y} = \lambda \frac{\partial h}{\partial y} \qquad\qquad \Rightarrow 2y = \lambda x \qquad\qquad (2.11)$$

$$h = 4 \qquad\qquad \Rightarrow xy = 4. \qquad\qquad (2.12)$$

Solving them (and discarding the unphysical complex valued solution) yields $(x, y) = (2, 2)$ and $(x, y) = (-2, -2)$. Hardly a surprise.

Remark. The method of Lagrange multipliers does not tell you whether you have a maximum, minimum, or saddle point for your objective function, so you need to check this by other means.

In higher dimensions the mathematics is similar—we just add variables. If we have more than one constraint, though, we need to impose more conditions. Suppose we have one objective function f, but m constraints, $h_1 = c_1$, $h_2 = c_2$, ..., $h_m = c_m$. If $\boldsymbol{\nabla} f$ had a component tangent to every one of the constraint surfaces at some point $\boldsymbol{r}_0$, then we could move a bit in that direction and change f while maintaining all the constraints. But then $\boldsymbol{r}_0$ would not be an extreme point of f. So $\boldsymbol{\nabla} f$ must be orthogonal to at least some (possibly all) of the constraint surfaces at that point. This means that $\boldsymbol{\nabla} f$ must be a linear combination of the gradient vectors $\boldsymbol{\nabla} h_i$. Together with the constraint equations themselves, the conditions now read

$$\boldsymbol{\nabla} f = \sum_{i=1}^{m} \lambda_i \boldsymbol{\nabla} h_i \qquad\qquad (2.13)$$

$$h_i = c_i \qquad (i = 1, \ldots, m). \qquad\qquad (2.14)$$

Remark. Observe that this is the correct number of equations. If we are in $\mathbb{R}^n$, there are n variables and the gradient operator is a vector of length n, so (2.13) gives n equations. (2.14) gives m more equations, for a total of $m + n$, and this is precisely the number of unknowns (namely, $x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m$).

Remark. The desired equations can be packaged more neatly using the **Lagrangian function** for the problem. In the preceding example, the Lagrangian function is

$$F = f - \sum_i \lambda_i h_i. \tag{2.15}$$

If we define the **augmented gradient operator** to be the vector operator given by

$$\boldsymbol{\nabla}' = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}, \frac{\partial}{\partial \lambda_1}, \frac{\partial}{\partial \lambda_2}, \ldots, \frac{\partial}{\partial \lambda_m} \right),$$

then Equations (2.13) and (2.14) are equivalent to the single equation

$$\boldsymbol{\nabla}' F = 0. \tag{2.16}$$

This is sometimes a convenient way to remember the optimization equations.

Remark. Let's give a physicists' proof of the correctness of the method of Lagrange multipliers for the simple case of one constraint. The general case follows similarly. We want to extremize $f(\boldsymbol{r})$ subject to the constraint $h(\boldsymbol{r}) = c$. Let $\boldsymbol{r}(t)$ be a curve lying in the level surface $\Sigma := \{\boldsymbol{r} | h(\boldsymbol{r}) = c\}$, and set $\boldsymbol{r}(0) = \boldsymbol{r}_0$. Then $d\boldsymbol{r}(t)/dt|_{t=0}$ is tangent to Σ at $\boldsymbol{r}_0$. Now restrict f to Σ and suppose $f(\boldsymbol{r}_0)$ is an extremum. Then $f(\boldsymbol{r}(t))$ is extremized at $t = 0$. But this implies that

$$0 = \left. \frac{df(\boldsymbol{r}(t))}{dt} \right|_{t=0} = \left. \frac{d\boldsymbol{r}(t)}{dt} \right|_{t=0} \cdot \boldsymbol{\nabla} f|_{\boldsymbol{r}_0}. \tag{2.17}$$

Hence $\boldsymbol{\nabla} f|_{\boldsymbol{r}_0}$ is orthogonal to Σ, so $\boldsymbol{\nabla} f|_{\boldsymbol{r}_0}$ and $\boldsymbol{\nabla} h|_{\boldsymbol{r}_0}$ are proportional. $\qquad$ $\square$

2.4 The Divergence

Definition.   *In Cartesian coordinates, the **divergence** of a vector field $\boldsymbol{A} = (A_x, A_y, A_z)$ is the scalar field given by*

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}. \tag{2.18}$$

Example 3   If $\boldsymbol{A} = (3xz, 2y^2 x, 4xy)$ then $\boldsymbol{\nabla} \cdot \boldsymbol{A} = 3z + 4xy$.

N.B. The gradient operator takes scalar fields to vector fields, while the divergence operator takes vector fields to scalar fields. Try not to confuse the two.

2.5 The Laplacian

Definition.   *The **Laplacian** of a scalar field φ is the divergence of the gradient of φ. In Cartesian coordinates we have*

$$\begin{aligned} \nabla^2 \varphi &= \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \varphi \\ &= \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2}. \end{aligned} \tag{2.19}$$

Example 4 If $\varphi = x^2y^2z^3 + xz^4$ then $\nabla^2\varphi = 2y^2z^3 + 2x^2z^3 + 6x^2y^2z + 12xz^2$.

2.6 The Curl

Definition. *The **curl** of a vector field $\boldsymbol{A}$ is another vector field given by*

$$\boldsymbol{\nabla} \times \boldsymbol{A} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \partial_x & \partial_y & \partial_z \\ A_x & A_y & A_z \end{vmatrix} \tag{2.20}$$

$$= \hat{\boldsymbol{\imath}}\,(\partial_y A_z - \partial_z A_y) + cyclic. \tag{2.21}$$

In this definition (and in all that follows) we employ the notation

$$\partial_x := \frac{\partial}{\partial x}, \quad \partial_y := \frac{\partial}{\partial y}, \quad \text{and} \quad \partial_z := \frac{\partial}{\partial z}. \tag{2.22}$$

Example 5 Let $\boldsymbol{A} = (3xz, 2xy^2, 4xy)$. Then

$$\boldsymbol{\nabla} \times \boldsymbol{A} = \hat{\boldsymbol{\imath}}(4x - 0) + \hat{\boldsymbol{\jmath}}(3x - 4y) + \hat{\boldsymbol{k}}(2y^2 - 0) = (4x, 3x - 4y, 2y^2).$$

Definition. $\boldsymbol{A}$ *is **solenoidal** (or **divergence-free**) if $\boldsymbol{\nabla} \cdot \boldsymbol{A} = 0$. $\boldsymbol{A}$ is* **irrotational** *(or **curl-free**) if $\boldsymbol{\nabla} \times \boldsymbol{A} = 0$.*

Claim 2.2. *DCG$\equiv$0.* i.e.,

  i) $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) \equiv 0$.

ii) $\boldsymbol{\nabla} \times \boldsymbol{\nabla}\varphi \equiv 0.$

Proof. We illustrate the proof for (i). The proof of (ii) is similar. We have

$$
\begin{aligned}
\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) &= \boldsymbol{\nabla} \cdot (\partial_y A_z - \partial_z A_y, \partial_z A_x - \partial_x A_z, \partial_x A_y - \partial_y A_x) \\
&= \partial_x (\partial_y A_z - \partial_z A_y) + \partial_y (\partial_z A_x - \partial_x A_z) + \partial_z (\partial_x A_y - \partial_y A_x) \\
&= \partial_x \partial_y A_z - \partial_x \partial_z A_y + \partial_y \partial_z A_x - \partial_y \partial_x A_z + \partial_z \partial_x A_y - \partial_z \partial_y A_x \\
&= 0,
\end{aligned}
$$

where we used the crucial fact that mixed partial derivatives commute

$$
\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}, \tag{2.23}
$$

for any twice differentiable function. □

2.7 Vector Calculus with Indices

Remark. In this section we employ the summation convention without comment.

Recall that the gradient operator $\boldsymbol{\nabla}$ in Cartesian coordinates is the vector differential operator given by

$$
\boldsymbol{\nabla} := \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) = (\partial_x, \partial_y, \partial_z). \tag{2.24}
$$

It follows that the ith component of the gradient of a scalar field φ is just

$$
(\boldsymbol{\nabla}\varphi)_i = \partial_i \varphi. \tag{2.25}
$$

43

Similarly, the divergence of a vector field $\boldsymbol{A}$ is written

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = \partial_i A_i. \tag{2.26}$$

The Laplacian operator may be viewed as the divergence of the gradient, so (2.25) and (2.26) together yield the Laplacian of a scalar field φ:

$$\boldsymbol{\nabla}^2 \varphi = \partial_i \partial_i \varphi. \tag{2.27}$$

Finally, the curl becomes

$$(\boldsymbol{\nabla} \times \boldsymbol{A})_i = \varepsilon_{ijk} \partial_j A_k. \tag{2.28}$$

Once again, casting formulae into index notation greatly simplifies some proofs. As a simple example we demonstrate the fact that the divergence of a curl is always zero:

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) = \partial_i(\varepsilon_{ijk} \partial_j A_k) = \varepsilon_{ijk} \partial_i \partial_j A_k = 0. \tag{2.29}$$

(Compare this to the proof given in Section 2.6.) The first equality is true by definition, while the second follows from the fact that the epsilon tensor is constant (0, 1, or -1), so it pulls out of the derivative. We say the last equality holds "by inspection", because (i) mixed partial derivatives commute (*cf.*, (2.23)) so $\partial_i \partial_j A_k$ is *symmetric* under the interchange of i and j, and (ii) the contracted product of a symmetric and an antisymmetric tensor is identically zero.

The proof of (ii) goes as follows. Let A_{ij} be a symmetric tensor and B_{ij}

be an antisymmetric tensor. This means that

$$A_{ij} = A_{ji} \qquad\qquad (2.30)$$

and

$$B_{ij} = -B_{ji}, \qquad\qquad (2.31)$$

for all pairs i and j. Then

$$
\begin{aligned}
A_{ij}B_{ij} &= A_{ji}B_{ij} & \text{(using (2.30))} && (2.32)\\
&= -A_{ji}B_{ji} & \text{(using (2.31))} && (2.33)\\
&= -A_{ij}B_{ij} & \text{(switching dummy indices } i \text{ and } j) && (2.34)\\
&= 0. &&& (2.35)
\end{aligned}
$$

Be sure you understand each step in the sequence above. The tricky part is switching the dummy indices in step three. We can always do this in a sum, provided we are careful to change *all* the indices of the same kind with the same letter. For example, given two vectors $\boldsymbol{C}$ and $\boldsymbol{D}$, their dot product can be written as either C_iD_i or C_jD_j, because both expressions are equal to $C_1D_1 + C_2D_2 + C_3D_3$. It does not matter whether we use i as our dummy index or whether we use j—the sum is the same. But note that it would not be true if the indices were not summed.

The same argument shows that $\varepsilon_{ijk}\partial_i\partial_j A_k = 0$, because the epsilon tensor is antisymmetric under the interchange of i and j, while the partial derivatives are symmetric under the same interchange. (The k index just goes along for the ride; alternatively, the expression vanishes for each of $k = 1, 2, 3$, so the sum over k also vanishes.)

Let us do one more vector calculus identity for the road. This time, we

prove the identity:

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{A}) - \boldsymbol{\nabla}^2 \boldsymbol{A}. \tag{2.36}$$

Consider what is involved in proving this the old-fashioned way. We first have to expand the curl of $\boldsymbol{A}$, and then take the curl of that. So we the first few steps of a demonstration along these lines would look like this:

$$
\begin{aligned}
\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) &= \boldsymbol{\nabla} \times (\partial_y A_z - \partial_z A_y, \partial_z A_x - \partial_x A_z, \partial_x A_y - \partial_y A_x) \\
&= (\partial_y (\partial_x A_y - \partial_y A_x) - \partial_z (\partial_z A_x - \partial_x A_z), \dots) \\
&= \dots.
\end{aligned}
$$

We would then have to do all the derivatives and collect terms to show that we get the right hand side of (2.36). You can do it this way, but it is unpleasant.

A more elegant proof using index notation proceeds as follows:

$$
\begin{aligned}
[\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A})]_i &= \varepsilon_{ijk}\partial_j(\boldsymbol{\nabla} \times \boldsymbol{A})_k && \text{(using (2.28))} \\
&= \varepsilon_{ijk}\partial_j(\varepsilon_{klm}\partial_l A_m) && \text{(using (2.28) again)} \\
&= \varepsilon_{ijk}\varepsilon_{klm}\partial_j(\delta_l A_m) && \text{(as ε_{klm} is constant)} \\
&= (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\partial_j\partial_l A_m && \text{(from (1.81))} \\
&= \partial_i\partial_j A_j - \partial_j\partial_j A_i && \text{(substitution property of δ)} \\
&= \partial_i(\boldsymbol{\nabla} \cdot \boldsymbol{A}) - \boldsymbol{\nabla}^2 A_i, && \text{((2.26) and (2.27))}
\end{aligned}
$$

and we are finished. (You may want to compare this with the proof of (1.86).)

46

2.8 Problems

1) Write down equations for the tangent plane and normal line to the surface $x^2y + y^2z + z^2x + 1 = 0$ at the point $(1, 2, -1)$.

2) Old postal regulations dictated that the maximum size of a rectangular box that could be sent parcel post was $108''$, measured as length plus girth. (If the box length is z, say, then the girth is $2x + 2y$, where x and y are the lengths of the other sides.) What is the maximum volume of such a package?

3) Either directly or using index methods, show that, for any scalar field φ and vector field $\boldsymbol{A}$,

   (a) $\boldsymbol{\nabla} \cdot (\varphi\boldsymbol{A}) = \boldsymbol{\nabla}\varphi \cdot \boldsymbol{A} + \varphi\boldsymbol{\nabla} \cdot \boldsymbol{A}$,

   (b) $\boldsymbol{\nabla} \times (\varphi\boldsymbol{A}) = \boldsymbol{\nabla}\varphi \times \boldsymbol{A} + \varphi\boldsymbol{\nabla} \times \boldsymbol{A}$, and

   (c) $\boldsymbol{\nabla}^2(\varphi\psi) = (\boldsymbol{\nabla}^2\varphi)\psi + 2\boldsymbol{\nabla}\varphi \cdot \boldsymbol{\nabla}\psi + \varphi\boldsymbol{\nabla}^2\psi$.

4) Show that, for $r \neq 0$,

   (a) $\boldsymbol{\nabla} \cdot \hat{\boldsymbol{r}} = 2/r$, and

   (b) $\boldsymbol{\nabla} \times \hat{\boldsymbol{r}} = 0$.

5) A function $f(\boldsymbol{r}) = f(x, y, z)$ is *homogeneous* of degree k if

$$f(a\boldsymbol{r}) = a^k f(\boldsymbol{r}) \tag{2.37}$$

   for any nonzero constant a. Prove *Euler's Theorem*, which states that, for any homogeneous function f of degree k,

$$(\boldsymbol{r} \cdot \boldsymbol{\nabla})f = kf. \tag{2.38}$$

   [Hint: Differentiate both sides of (2.37) with respect to a and use the chain rule, then evaluate at $a = 1$.]

6) A function φ satisfying $\boldsymbol{\nabla}^2\varphi = 0$ is called *harmonic*.

   (a) Using Cartesian coordinates, show that $\varphi = 1/r$ is harmonic, where $r = (x^2 + y^2 + z^2)^{1/2} \neq 0$.

(b) Let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ be a vector of nonnegative integers, and define $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3$. Let ∂_α be the differential operator $\partial_x^{\alpha_1} \partial_y^{\alpha_2} \partial_z^{\alpha_3}$. Prove that any function of the form $\varphi = r^{2|\alpha|+1} \partial_\alpha(1/r)$ is harmonic. [Hint: Use vector calculus identities to expand out $\boldsymbol{\nabla}^2(r^n f)$ where $f := \partial_\alpha(1/r)$ and use Euler's theorem and the fact that mixed partials commute.]

7) Using index methods, prove the following vector calculus identities:

(a) $\boldsymbol{\nabla} \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{B} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) - \boldsymbol{A} \cdot (\boldsymbol{\nabla} \times \boldsymbol{B})$.

(b) $\boldsymbol{\nabla}(\boldsymbol{A} \cdot \boldsymbol{B}) = (\boldsymbol{B} \cdot \boldsymbol{\nabla})\boldsymbol{A} + (\boldsymbol{A} \cdot \boldsymbol{\nabla})\boldsymbol{B} + \boldsymbol{B} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) + \boldsymbol{A} \times (\boldsymbol{\nabla} \times \boldsymbol{B})$.

3 Vector Calculus II: Other Coordinate Systems

3.1 Change of Variables from Cartesian to Spherical Polar

So far we have dealt exclusively with Cartesian coordinates. But for many problems it is more convenient to analyse the problem using a different coordinate system. Here we see what is involved in translating the vector operators to spherical coordinates, leaving the task for other coordinate systems to the reader.

First we recall the relationship between Cartesian coordinates and spherical polar coordinates (see Figure 9):

$$
\begin{aligned}
x &= r \sin \theta \cos \phi & r &= (x^2 + y^2 + z^2)^{1/2} \\
y &= r \sin \theta \sin \phi & \theta &= \cos^{-1} \frac{z}{(x^2 + y^2 + z^2)^{1/2}} \\
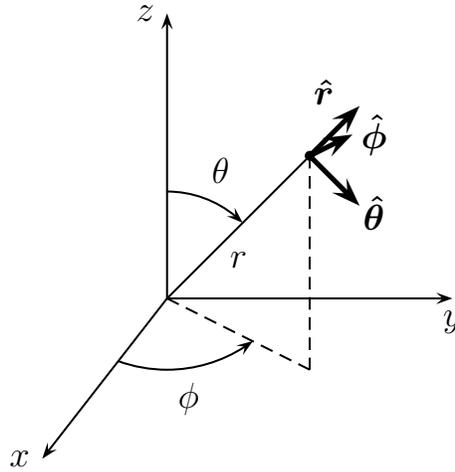z &= r \cos \theta & \phi &= \tan^{-1} \frac{y}{x}.
\end{aligned}
\tag{3.1}
$$

Figure 9: Spherical polar coordinates and corresponding unit vectors

3.2 Vector Fields and Derivations

Next we need the equations relating the Cartesian unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$, to the spherical polar unit vectors $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$. To do this we introduce a new idea, namely the idea of **vector field as derivation**. We have already encountered the basic idea above. Suppose you walk along a curve $\boldsymbol{r}(t)$ in the presence of a scalar field φ. Then the rate of change of φ along the curve is

$$\frac{d\varphi(t)}{dt} = (\boldsymbol{v} \cdot \boldsymbol{\nabla})\varphi. \tag{3.2}$$

On the left side of this expression we have the derivative of φ along the curve, while on the right side we have the directional derivative of φ in a direction tangent to the curve. We can dispense with φ altogether, and simply write

$$\frac{d}{dt} = \boldsymbol{v} \cdot \boldsymbol{\nabla}. \tag{3.3}$$

49

That is, d/dt, the derivative with respect to t, the parameter along the curve, is the same thing as directional derivative in the $\boldsymbol{v}$ direction. This allows us to identify the **derivation** d/dt and the vector field $\boldsymbol{v}$. [10] To every vector field there is a derivation, namely the directional derivative in the direction of the vector field, and *vice versa*, so mathematicians often identify the two concepts.

For example, let us walk along the x axis with some speed v. Then

$$\frac{d}{dt} = \frac{dx}{dt}\frac{\partial}{\partial x} = v\frac{\partial}{\partial x}, \tag{3.4}$$

so (3.3) becomes

$$v\frac{\partial}{\partial x} = v\hat{\boldsymbol{x}}\cdot\boldsymbol{\nabla}. \tag{3.5}$$

Dividing both sides by v gives

$$\frac{\partial}{\partial x} = \hat{\boldsymbol{x}}\cdot\boldsymbol{\nabla}, \tag{3.6}$$

which is consistent with our previous results. Hence we write

$$\frac{\partial}{\partial x} \longleftrightarrow \hat{\boldsymbol{x}} \tag{3.7}$$

to indicate that the derivation on the left corresponds to the vector field on the right. Clearly, an analogous result holds for $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$. Note also that (3.6) is an *equality* whereas (3.7) is an *association*. Keep this distinction in mind to avoid confusion.

Suppose instead that we were to move along a longitude in the direction

---

[10] A *derivation* D is a linear operator obeying the Leibniz rule. That is, $D(\phi + \psi) = D\phi + D\psi$, and $D(\phi\psi) = (D\phi)\psi + \phi D\psi$.

of increasing θ. Then we would have

$$\frac{d}{dt} = \frac{d\theta}{dt}\frac{\partial}{\partial\theta}, \tag{3.8}$$

and (3.3) would become

$$\frac{d\theta}{dt}\frac{\partial}{\partial\theta} = v\hat{\boldsymbol{\theta}} \cdot \boldsymbol{\nabla}. \tag{3.9}$$

But now $d\theta/dt$ is *not* the speed. Instead,

$$v = r\frac{d\theta}{dt}, \tag{3.10}$$

so (3.9) yields

$$\frac{1}{r}\frac{\partial}{\partial\theta} = \hat{\boldsymbol{\theta}} \cdot \boldsymbol{\nabla}. \tag{3.11}$$

This allows us to identify

$$\frac{1}{r}\frac{\partial}{\partial\theta} \longleftrightarrow \hat{\boldsymbol{\theta}}. \tag{3.12}$$

We can avoid reference to the speed of the observer by the following method. From the chain rule

$$\begin{aligned}
\frac{\partial}{\partial r} &= \left(\frac{\partial x}{\partial r}\right)\frac{\partial}{\partial x} + \left(\frac{\partial y}{\partial r}\right)\frac{\partial}{\partial y} + \left(\frac{\partial z}{\partial r}\right)\frac{\partial}{\partial z} \\
\frac{\partial}{\partial\theta} &= \left(\frac{\partial x}{\partial\theta}\right)\frac{\partial}{\partial x} + \left(\frac{\partial y}{\partial\theta}\right)\frac{\partial}{\partial y} + \left(\frac{\partial z}{\partial\theta}\right)\frac{\partial}{\partial z} \\
\frac{\partial}{\partial\phi} &= \left(\frac{\partial x}{\partial\phi}\right)\frac{\partial}{\partial x} + \left(\frac{\partial y}{\partial\phi}\right)\frac{\partial}{\partial y} + \left(\frac{\partial z}{\partial\phi}\right)\frac{\partial}{\partial z}.
\end{aligned} \tag{3.13}$$

Using (3.1) gives

$$\frac{\partial x}{\partial r} = \sin\theta\cos\phi \qquad \frac{\partial y}{\partial r} = \sin\theta\sin\phi \qquad \frac{\partial z}{\partial r} = \cos\theta$$

$$\frac{\partial x}{\partial \theta} = r\cos\theta\cos\phi \qquad \frac{\partial y}{\partial \theta} = r\cos\theta\sin\phi \qquad \frac{\partial z}{\partial \theta} = -r\sin\theta \qquad (3.14)$$

$$\frac{\partial x}{\partial \phi} = -r\sin\theta\sin\phi \qquad \frac{\partial y}{\partial \phi} = r\sin\theta\cos\phi \qquad \frac{\partial z}{\partial \phi} = 0,$$

so

$$\frac{\partial}{\partial r} = \sin\theta\cos\phi\,\frac{\partial}{\partial x} \quad +\sin\theta\sin\phi\,\frac{\partial}{\partial y} \quad +\cos\theta\,\frac{\partial}{\partial z} \qquad (3.15)$$

$$\frac{\partial}{\partial \theta} = r\cos\theta\cos\phi\,\frac{\partial}{\partial x} \quad +r\cos\theta\sin\phi\,\frac{\partial}{\partial y} \quad -r\sin\theta\,\frac{\partial}{\partial z} \qquad (3.16)$$

$$\frac{\partial}{\partial \phi} = -r\sin\theta\sin\phi\,\frac{\partial}{\partial x} \quad +r\sin\theta\cos\phi\,\frac{\partial}{\partial y}. \qquad (3.17)$$

Now we just identify the derivations on the left with multiples of the corresponding unit vectors. For example, if we write

$$\frac{\partial}{\partial \theta} \longleftrightarrow \alpha\hat{\boldsymbol{\theta}}, \qquad (3.18)$$

then from (3.16) we get

$$\alpha\hat{\boldsymbol{\theta}} = r\cos\theta\cos\phi\,\hat{\boldsymbol{x}} + r\cos\theta\sin\phi\,\hat{\boldsymbol{y}} - r\sin\theta\,\hat{\boldsymbol{z}}. \qquad (3.19)$$

The vector on the right side of (3.19) has length r, which means that $\alpha = r$, and we recover (3.12) from (3.18). Furthermore, we also conclude that

$$\hat{\boldsymbol{\theta}} = \cos\theta\cos\phi\,\hat{\boldsymbol{x}} + \cos\theta\sin\phi\,\hat{\boldsymbol{y}} - \sin\theta\,\hat{\boldsymbol{z}}. \qquad (3.20)$$

Continuing in this way gives

$$\hat{\boldsymbol{r}} \longleftrightarrow \frac{\partial}{\partial r} \tag{3.21}$$

$$\hat{\boldsymbol{\theta}} \longleftrightarrow \frac{1}{r}\frac{\partial}{\partial \theta} \tag{3.22}$$

$$\hat{\boldsymbol{\phi}} \longleftrightarrow \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi} \tag{3.23}$$

and

$$\hat{\boldsymbol{r}} = \sin\theta\cos\phi\,\hat{\boldsymbol{x}} + \sin\theta\sin\phi\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}} \tag{3.24}$$

$$\hat{\boldsymbol{\theta}} = \cos\theta\cos\phi\,\hat{\boldsymbol{x}} + \cos\theta\sin\phi\,\hat{\boldsymbol{y}} - \sin\theta\,\hat{\boldsymbol{z}} \tag{3.25}$$

$$\hat{\boldsymbol{\phi}} = -\sin\phi\,\hat{\boldsymbol{x}} + \cos\phi\,\hat{\boldsymbol{y}}. \tag{3.26}$$

If desired, we could now use (3.1) to express the above equations in terms of Cartesian coordinates.

3.3   Derivatives of Unit Vectors

The reason why vector calculus is simpler in Cartesian coordinates than in any other coordinate system is that the unit vectors $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$ are *constant*. This means that, no matter where you are in space, these vectors never change length or direction. But it is immediately apparent from (3.24)-(3.26) (see also Figure 9) that the spherical polar unit vectors $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ vary in direction (though not in length) as we move around. It is this difference that makes vector calculus in spherical coordinates a bit of a mess.

Thus, we must compute how the spherical polar unit vectors change as

we move around. A little calculation yields

$$
\begin{array}{lll}
\dfrac{\partial \hat{\boldsymbol{r}}}{\partial r} = 0 & \dfrac{\partial \hat{\boldsymbol{r}}}{\partial \theta} = \hat{\boldsymbol{\theta}} & \dfrac{\partial \hat{\boldsymbol{r}}}{\partial \phi} = \sin\theta\,\hat{\boldsymbol{\phi}} \\[2ex]
\dfrac{\partial \hat{\boldsymbol{\theta}}}{\partial r} = 0 & \dfrac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta} = -\hat{\boldsymbol{r}} & \dfrac{\partial \hat{\boldsymbol{\theta}}}{\partial \phi} = \cos\theta\,\hat{\boldsymbol{\phi}} \\[2ex]
\dfrac{\partial \hat{\boldsymbol{\phi}}}{\partial r} = 0 & \dfrac{\partial \hat{\boldsymbol{\phi}}}{\partial \theta} = 0 & \dfrac{\partial \hat{\boldsymbol{\phi}}}{\partial \phi} = -(\sin\theta\,\hat{\boldsymbol{r}} + \cos\theta\,\hat{\boldsymbol{\theta}})
\end{array} \tag{3.27}
$$

3.4 Vector Components in a Non-Cartesian Basis

We began these notes by observing that the Cartesian components of a vector can be found by computing inner products. For example, the x component of a vector $\boldsymbol{A}$ is just $\hat{\boldsymbol{x}} \cdot \boldsymbol{A}$. Similarly, the spherical polar components of the vector $\boldsymbol{A}$ are defined by

$$
\boldsymbol{A} = A_r\hat{\boldsymbol{r}} + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}. \tag{3.28}
$$

Equivalently,

$$
\boldsymbol{A} = \hat{\boldsymbol{r}}(\hat{\boldsymbol{r}} \cdot \boldsymbol{A}) + \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} \cdot \boldsymbol{A}) + \hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}} \cdot \boldsymbol{A}). \tag{3.29}
$$

3.5 Vector Operators in Spherical Coordinates

We are finally ready to find expressions for the gradient, divergence, curl, and Laplacian in spherical polar coordinates. We begin with the gradient operator. According to (3.29) we have

$$
\boldsymbol{\nabla} = \hat{\boldsymbol{r}}(\hat{\boldsymbol{r}} \cdot \boldsymbol{\nabla}) + \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} \cdot \boldsymbol{\nabla}) + \hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}} \cdot \boldsymbol{\nabla}). \tag{3.30}
$$

In this formula the unit vectors are followed by the derivations in the direction of the unit vectors. But the latter are precisely what we computed in (3.21)-

(3.23), so we get

$$\boldsymbol{\nabla} = \hat{\boldsymbol{r}}\partial_r + \hat{\boldsymbol{\theta}}\frac{1}{r}\partial_\theta + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\partial_\phi. \tag{3.31}$$

Example 6 If $\varphi = r^2\sin^2\theta\sin\phi$ then $\boldsymbol{\nabla}\varphi = 2r\sin^2\theta\sin\phi\,\hat{\boldsymbol{r}} + 2r\sin\theta\cos\theta\sin\phi\,\hat{\boldsymbol{\theta}} + r\sin\theta\cos\phi\,\hat{\boldsymbol{\phi}}$.

The divergence is a bit trickier. Now we have

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = (\hat{\boldsymbol{r}}\partial_r + \hat{\boldsymbol{\theta}}\frac{1}{r}\partial_\theta + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\partial_\phi) \cdot (A_r\hat{\boldsymbol{r}} + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}). \tag{3.32}$$

To compute this expression we must act *first* with the derivatives, and *then* take the dot products. This gives

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{A} = {}& \hat{\boldsymbol{r}} \cdot \partial_r(A_r\hat{\boldsymbol{r}} + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \\
& + \frac{1}{r}\hat{\boldsymbol{\theta}} \cdot \partial_\theta(A_r\hat{\boldsymbol{r}} + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \\
& + \frac{1}{r\sin\theta}\hat{\boldsymbol{\phi}} \cdot \partial_\phi(A_r\hat{\boldsymbol{r}} + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}).
\end{aligned} \tag{3.33}$$

With a little help from (3.27) we get

$$\begin{aligned}
\partial_r(A_r\hat{\boldsymbol{r}} & + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \\
& = (\partial_r A_r)\hat{\boldsymbol{r}} + A_r(\partial_r\hat{\boldsymbol{r}}) + (\partial_r A_\theta)\hat{\boldsymbol{\theta}} + A_\theta(\partial_r\hat{\boldsymbol{\theta}}) + (\partial_r A_\phi)\hat{\boldsymbol{\phi}} + A_\phi(\partial_r\hat{\boldsymbol{\phi}}) \\
& = (\partial_r A_r)\hat{\boldsymbol{r}} + (\partial_r A_\theta)\hat{\boldsymbol{\theta}} + (\partial_r A_\phi)\hat{\boldsymbol{\phi}},
\end{aligned} \tag{3.34}$$

$$\begin{aligned}
\partial_\theta(A_r\hat{\boldsymbol{r}} & + A_\theta\hat{\boldsymbol{\theta}} + A_\phi\hat{\boldsymbol{\phi}}) \\
& = (\partial_\theta A_r)\hat{\boldsymbol{r}} + A_r(\partial_\theta\hat{\boldsymbol{r}}) + (\partial_\theta A_\theta)\hat{\boldsymbol{\theta}} + A_\theta(\partial_\theta\hat{\boldsymbol{\theta}}) + (\partial_\theta A_\phi)\hat{\boldsymbol{\phi}} + A_\phi(\partial_\theta\hat{\boldsymbol{\phi}}) \\
& = (\partial_\theta A_r)\hat{\boldsymbol{r}} + A_r\hat{\boldsymbol{\theta}} + (\partial_\theta A_\theta)\hat{\boldsymbol{\theta}} - A_\theta\hat{\boldsymbol{r}} + (\partial_\theta A_\phi)\hat{\boldsymbol{\phi}},
\end{aligned} \tag{3.35}$$

and

$$\partial_\phi (A_r \hat{\boldsymbol{r}} + A_\theta \hat{\boldsymbol{\theta}} + A_\phi \hat{\boldsymbol{\phi}})$$

$$= (\partial_\phi A_r)\hat{\boldsymbol{r}} + A_r(\partial_\phi \hat{\boldsymbol{r}}) + (\partial_\phi A_\theta)\hat{\boldsymbol{\theta}} + A_\theta(\partial_\phi \hat{\boldsymbol{\theta}}) + (\partial_\phi A_\phi)\hat{\boldsymbol{\phi}} + A_\phi(\partial_\phi \hat{\boldsymbol{\phi}})$$

$$= (\partial_\phi A_r)\hat{\boldsymbol{r}} + A_r(\sin\theta\,\hat{\boldsymbol{\phi}}) + (\partial_\phi A_\theta)\hat{\boldsymbol{\theta}}$$

$$+ A_\theta(\cos\theta\,\hat{\boldsymbol{\phi}}) + (\partial_\phi A_\phi)\hat{\boldsymbol{\phi}} - A_\phi(\sin\theta\,\hat{\boldsymbol{r}} + \cos\theta\,\hat{\boldsymbol{\theta}}). \tag{3.36}$$

Taking the dot products and combining terms gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = \partial_r A_r + \frac{1}{r}(A_r + \partial_\theta A_\theta) + \frac{1}{r\sin\theta}(A_r \sin\theta + A_\theta \cos\theta + \partial_\phi A_\phi)$$

$$= \left(\frac{\partial A_r}{\partial r} + \frac{A_r}{r}\right) + \left(\frac{1}{r}\frac{\partial A_\theta}{\partial\theta} + \frac{A_\theta \cos\theta}{r\sin\theta}\right) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 A_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi}. \tag{3.37}$$

Well, that was fun. Similar computations, which are left to the reader :-), yield the curl:

$$\boldsymbol{\nabla} \times \boldsymbol{A} = \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta A_\phi) - \frac{\partial A_\theta}{\partial\phi}\right]\hat{\boldsymbol{r}}$$

$$+ \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{\partial}{\partial r}(rA_\phi)\right]\hat{\boldsymbol{\theta}}$$

$$+ \frac{1}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right]\hat{\boldsymbol{\phi}}, \tag{3.38}$$

and the Laplacian

$$\boldsymbol{\nabla}^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{3.39}$$

Example 7 Let $\boldsymbol{A} = r^2\sin\theta\,\hat{\boldsymbol{r}} + 4r^2\cos\theta\,\hat{\boldsymbol{\theta}} + r^2\tan\theta\,\hat{\boldsymbol{\phi}}$. Then $\boldsymbol{\nabla} \cdot \boldsymbol{A} =$

$4r \cos^2 \theta / \sin \theta$, and $\boldsymbol{\nabla} \times \boldsymbol{A} = -r\,\hat{\boldsymbol{r}} - 3r \tan \theta\, \hat{\boldsymbol{\theta}} + 11r \cos \theta\, \hat{\boldsymbol{\phi}}$.

3.6 Problems

1) The transformation relating Cartesian and cylindrical coordinates is

$$
\begin{aligned}
x &= \rho \cos \theta & \rho &= (x^2 + y^2)^{1/2} \\
y &= \rho \sin \theta & \theta &= \tan^{-1} \frac{y}{x} \\
z &= z & z &= z
\end{aligned}
\tag{3.40}
$$

Using the methods of this section, show that the gradient operator in cylindrical coordinates is given by

$$
\boldsymbol{\nabla} = \hat{\boldsymbol{\rho}} \partial_\rho + \frac{1}{\rho} \hat{\boldsymbol{\theta}} \partial_\theta + \hat{\boldsymbol{z}} \partial_z .
\tag{3.41}
$$

4 Vector Calculus III: Integration

Integration is the flip side of differentiation—you cannot have one without the other. We begin with line integrals, then continue on to surface and volume integrals and the relations between them.

4.1 Line Integrals

There are many different types of line integrals, but the most important type arises as the inverse of the gradient function. Given a parameterized curve $\gamma(t)$ and a vector field $\boldsymbol{A}$, the line [11] integral of $\boldsymbol{A}$ along the curve $\gamma(t)$ is usually written

$$
\int_\gamma \boldsymbol{A} \cdot d\boldsymbol{\ell},
\tag{4.1}
$$

---

[11] The word 'line' is a bit of a misnomer in this context, because we really mean a 'curve' integral, but we will follow standard terminology.

where $d\boldsymbol{\ell}$ is the infinitessimal tangent vector to the curve. This notation fails to specify where the curve begins and ends. If the curve starts at a and ends at b, the same integral is usually written

$$\int_a^b \boldsymbol{A} \cdot d\boldsymbol{\ell}, \tag{4.2}$$

but the problem here is that the curve is not specified. The best notation would be

$$\int_{a;\gamma}^b \boldsymbol{A} \cdot d\boldsymbol{\ell}, \tag{4.3}$$

or some such, but unfortunately, no one does this. Thus, one usually has to decide *from context* what is going on.

As written, the line integral is merely a formal expression. We give it meaning by the mathematical operation of 'pullback', which basically means using the parameterization to write it as a conventional integral over a line segment in Euclidean space. Thus, if we write $\gamma(t) = (x(t), y(t), z(t))$, then $d\boldsymbol{\ell}/dt = d\gamma(t)/dt = \boldsymbol{v}$ is the velocity with which the curve is traversed, so

$$\int_\gamma \boldsymbol{A} \cdot d\boldsymbol{\ell} = \int_{t_0}^{t_1} \boldsymbol{A}(\gamma(t)) \cdot \frac{d\boldsymbol{\ell}}{dt} \, dt = \int_{t_0}^{t_1} \boldsymbol{A}(\gamma(t)) \cdot \boldsymbol{v} \, dt. \tag{4.4}$$

This last expression is independent of the parameterization used, which means it depends only on the curve. Suppose $t^* = t^*(t)$ were some other

parameterization. Then we would have

$$\int_{t_0^*}^{t_1^*} \boldsymbol{A}(\gamma(t^*)) \cdot \boldsymbol{v}(t^*)\, dt^* = \int_{t_0}^{t_1} \boldsymbol{A}(\gamma(t^*(t))) \cdot \frac{d\gamma(t^*(t))}{dt^*}\, dt^*$$

$$= \int_{t_0}^{t_1} \boldsymbol{A}(\gamma(t)) \cdot \frac{d\gamma(t)}{dt} \frac{dt}{dt^*}\, dt^*$$

$$= \int_{t_0}^{t_1} \boldsymbol{A}(\gamma(t)) \cdot \frac{d\gamma(t)}{dt}\, dt.$$

Example 8   Let $\boldsymbol{A} = (4xy, -8yz, 2xz)$, and let γ be the straight line path from $(1, 2, 6)$ to $(5, 3, 5)$. Every straight line segment from a to b can be parameterized in a natural way by $\gamma(t) = (b-a)t + a$. This is clearly a line segment which begins at a when $t = 0$ and ends up at b when $t = 1$. In our case we have

$$\gamma(t) = [(5, 3, 5) - (1, 2, 6)]t + (1, 2, 6) = (4t + 1, t + 2, -t + 6),$$

which implies

$$\boldsymbol{v}(t) = \dot{\gamma}(t) = (4, 1, -1).$$

Thus

$$\int_\gamma \boldsymbol{A} \cdot d\boldsymbol{\ell}$$

$$= \int_0^1 (4(4t + 1)(t + 2), -8(t + 2)(-t + 6), 2(4t + 1)(-t + 6)) \cdot (4, 1, -1)\, dt$$

$$= \int_0^1 (16(4t + 1)(t + 2) - 8(t + 2)(-t + 6) - 2(4t + 1)(-t + 6))\, dt$$

$$= \int_0^1 (80t^2 + 66t - 76)\, dt = \left. \frac{80}{3}t^3 + 33t^2 - 76t \right|_0^1 = -\frac{49}{3}.$$

Physicists usually simplify the notation by writing

$$d\boldsymbol{\ell} = \boldsymbol{v}\, dt = (dx, dy, dz). \tag{4.5}$$

Although this notation is somewhat ambiguous, it can be used to good effect under certain circumstances. Let $\boldsymbol{A} = \boldsymbol{\nabla}\varphi$ for some scalar field φ, and let γ be some curve. Then

$$\begin{aligned}
\int_{\gamma} \boldsymbol{\nabla}\varphi \cdot d\boldsymbol{\ell} &= \int_{\gamma} \left(\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}, \frac{\partial\varphi}{\partial z} \right) \cdot (dx, dy, dz) \\
&= \int_{\gamma} d\varphi \\
&= \varphi(b) - \varphi(a).
\end{aligned}$$

This clearly demonstrates that the line integral (4.4) is indeed the inverse operation to the gradient, in the same way that one dimensional integration is the inverse operation to one dimensional differentiation.

Recall that a vector field $\boldsymbol{A}$ is **conservative** if the line integral $\int \boldsymbol{A} \cdot d\boldsymbol{\ell}$ is **path independent**. That is, for *any* two curves γ_1 and γ_2 joining the points a and b, we have

$$\int_{\gamma_1} \boldsymbol{A} \cdot d\boldsymbol{\ell} = \int_{\gamma_2} \boldsymbol{A} \cdot d\boldsymbol{\ell}. \tag{4.6}$$

We can express this result another way. As $d\boldsymbol{\ell} \to -d\boldsymbol{\ell}$ when we reverse directions, if traverse the curve in the opposite direction we get the negative of the original path integral:

$$\int_{\gamma^{-1}} \boldsymbol{A} \cdot d\boldsymbol{\ell} = - \int_{\gamma} \boldsymbol{A} \cdot d\boldsymbol{\ell}, \tag{4.7}$$

where γ^{-1} represents the same curve γ traced backwards. Combining (4.6) and (4.7) we can write the condition for path independence as

$$\int_{\gamma_1} \boldsymbol{A} \cdot d\boldsymbol{\ell} = -\int_{\gamma_2^{-1}} \boldsymbol{A} \cdot d\boldsymbol{\ell} \tag{4.8}$$

$$\Rightarrow \quad \int_{\gamma_1} \boldsymbol{A} \cdot d\boldsymbol{\ell} + \int_{\gamma_2^{-1}} \boldsymbol{A} \cdot d\boldsymbol{\ell} = 0 \tag{4.9}$$

$$\Rightarrow \quad \oint_{\gamma} \boldsymbol{A} \cdot d\boldsymbol{\ell} = 0, \tag{4.10}$$

where $\gamma = \gamma_1 + \gamma_2^{-1}$ is the *closed* curve obtained by following γ_1 from a to b and then γ_2^{-1} from b back to a. [12] The line integral of a vector field $\boldsymbol{A}$ is usually called the **circulation** of $\boldsymbol{A}$, so $\boldsymbol{A}$ is conservative if the circulation of $\boldsymbol{A}$ vanishes around every closed curve.

Example 9 Consider the vector field $\boldsymbol{A} = (x^2 - y^2, 2xy)$ in the plane. Let γ_1 be the curve given by $y = 2x^2$ and let γ_2 be the curve given by $y = 2\sqrt{x}$. Let the endpoints be $a = (0,0)$ and $b = (1,2)$. This situation is sufficiently simple that a parameterization is unnecessary. We compute as follows:

$$
\begin{aligned}
\int_{\gamma_1} \boldsymbol{A} \cdot d\boldsymbol{\ell} &= \int_{\gamma_1} (A_x \, dx + A_y \, dy) = \int_{\gamma_1} (x^2 - y^2) \, dx + 2xy \, dy \\
&= \int_0^1 (x^2 - 4x^4) \, dx + 4x^3 \cdot 4x \, dx \\
&= \left[\frac{x^3}{3} - \frac{4x^5}{5} + \frac{16x^5}{5} \right] \Big|_0^1 = \frac{1}{3} + \frac{12}{5} = \frac{41}{15}.
\end{aligned}
$$

In the computation above we substituted in y as a function of x along the curve, then used the x limits. We could just as well have solved for x in terms of y and then solved the integral in the variable y instead. We do this in the next integral

---

[12]The circle on the integral sign merely serves to remind us that the integral is taken around a *closed* curve.

to avoid messy square roots. Since $x = (y/2)^2$ along γ_2, we get

$$\int_{\gamma_2} \boldsymbol{A} \cdot d\boldsymbol{\ell} = \int_{\gamma_2} (A_x \, dx + A_y \, dy) = \int_{\gamma_2} (x^2 - y^2) \, dx + 2xy \, dy$$

$$= \int_0^2 \left(\frac{y^4}{16} - y^2 \right) \cdot \frac{y}{2} \, dy + \frac{y^3}{2} \, dy$$

$$= \left[\frac{y^6}{160} \right] \Big|_0^2 = \frac{2}{5}.$$

Evidently, these are not equal. Hence the vector field $\boldsymbol{A} = (x^2 - y^2, 2xy)$ is not conservative.

But suppose we began with the vector field $\boldsymbol{A} = (x^2 - y^2, -2xy)$ instead. Now carrying out the same procedure as above would give $\int_{\gamma_1} \boldsymbol{A} \cdot d\boldsymbol{\ell} = -11/3 = \int_{\gamma_2} \boldsymbol{A} \cdot d\boldsymbol{\ell}$. Can we conclude from this that $\boldsymbol{A}$ is conservative? No! The reason is that we have only shown that the line integral of $\boldsymbol{A}$ is the same along these two curves between these two endpoints. But we must show that we get the same answer no matter which curve and which endpoints we pick. Now, this vector field is sufficiently simple that we can actually tell that it is indeed conservative. We do this by observing that $\boldsymbol{A} \cdot d\boldsymbol{\ell}$ is an **exact differential**, which means that it can be written as $d\varphi$ for some function φ. In our case the function $\varphi = (x^3/3) - xy^2$. (See the discussion below.) Hence

$$\int_{\gamma} \boldsymbol{A} \cdot d\boldsymbol{\ell} = \int_{(0,0)}^{(1,2)} d\varphi = \varphi(1,2) - \varphi(0,0) = -\frac{11}{3}, \tag{4.11}$$

which demonstrates the path independence of the integral. Comparing this analysis to our discussion above shows that the reason why $\boldsymbol{A} \cdot d\boldsymbol{\ell}$ is an exact differential is because $\boldsymbol{A} = \boldsymbol{\nabla} \varphi$.

Example 9 illustrates the fact that any vector field that can be written as the gradient of a scalar field is conservative. This brings us naturally to the question of determining when such a situation holds. An obvious necessary

condition is that the curl of the vector field must vanish (because the curl of a gradient is identically zero). It turns out that this condition is also sufficient. That is, if $\boldsymbol{\nabla} \times \boldsymbol{A} = 0$ for some vector field $\boldsymbol{A}$ then $\boldsymbol{A} = \boldsymbol{\nabla}\varphi$ for some φ. This follows from a lemma of Poincaré that we will not discuss here. [13]

Example 10 Consider again the two vector fields from Example 9. The first one, namely $\boldsymbol{A} = (x^2 - y^2, 2xy, 0)$, satisfies $\boldsymbol{\nabla} \times \boldsymbol{A} = 4y\,\hat{\boldsymbol{k}}$, and so is non-conservative, whereas the second one, namely $\boldsymbol{A} = (x^2 - y^2, -2xy, 0)$, satisfies $\boldsymbol{\nabla} \times \boldsymbol{A} = 0$ and so is conservative, as previously demonstrated.

This gives us an easy criterion to test for conservative vector fields, but it does not produce the corresponding scalar field for us. To find this, we use **partial integration**. Suppose we are given a vector field $\boldsymbol{A} = (A_x, A_y, A_z)$. If $\boldsymbol{A} = \boldsymbol{\nabla}\varphi$ for some φ, then

$$\frac{\partial \varphi}{\partial x} = A_x, \quad \frac{\partial \varphi}{\partial y} = A_y, \quad \text{and} \quad \frac{\partial \varphi}{\partial z} = A_z. \tag{4.12}$$

Partially integrating these equations gives

$$\varphi = \int A_x \, dx + f(y, z), \tag{4.13}$$

$$\varphi = \int A_y \, dy + g(x, z), \quad \text{and} \tag{4.14}$$

$$\varphi = \int A_z \, dz + h(x, y), \tag{4.15}$$

where f, g, and h are unknown functions of the given variables. [14] If (4.13)-

---

[13]There is one *caveat*, which is that the conclusion only holds if the region over which the curl vanishes is *simply connected*. Roughly speaking, this means the region has no 'holes'.

[14]We are integrating the vector field components with respect to one variable only, which is why it is called *partial* integration.

(4.15) can be solved consistently for a function φ, then $\boldsymbol{A} = \boldsymbol{\nabla}\varphi$.

Example 11 Let $\boldsymbol{A} = (x^2 - y^2, -2xy, 0)$ as in Example 9. To show that it can be written as the gradient of a scalar field we partially integrate the components to get

$$\varphi = \int (x^2 - y^2)\, dx + f(y, z) = \frac{1}{3}x^3 - xy^2 + f(y, z),$$

$$\varphi = \int (-2xy)\, dy + g(x, z) = -xy^2 + g(x, z), \quad \text{and}$$

$$\varphi = 0 + h(x, y).$$

These equations can be made consistent if we choose

$$f = 0, \qquad g = \frac{1}{3}x^3, \qquad \text{and} \qquad h = \frac{1}{3}x^3 - xy^2. \tag{4.16}$$

so $\varphi = (x^3/3) - xy^2$ is the common solution. The reader should verify that this procedure fails for the nonconservative vector field $\boldsymbol{A} = (x^2 - y^2, 2xy, 0)$.

4.2   Surface Integrals

Let S be a two dimensional surface in space, and let $\boldsymbol{A}$ be a vector field. Then the **flux** of $\boldsymbol{A}$ through the surface S is

$$\int_S \boldsymbol{A} \cdot d\boldsymbol{S}. \tag{4.17}$$

In this formula $d\boldsymbol{S} = \hat{\boldsymbol{n}}dS$, where dS is the infinitessimal area element on the surface and $\hat{\boldsymbol{n}}$ points orthogonally to the surface. As before, this integral is *defined* in terms of a parameterization. A surface is a two dimensional object, so it depends on two parameters, which we will usually denote by u and v. Let $\sigma(u, v) : \mathbb{R}^2 \to \mathbb{R}^3$ be such a parameterization. As u and v
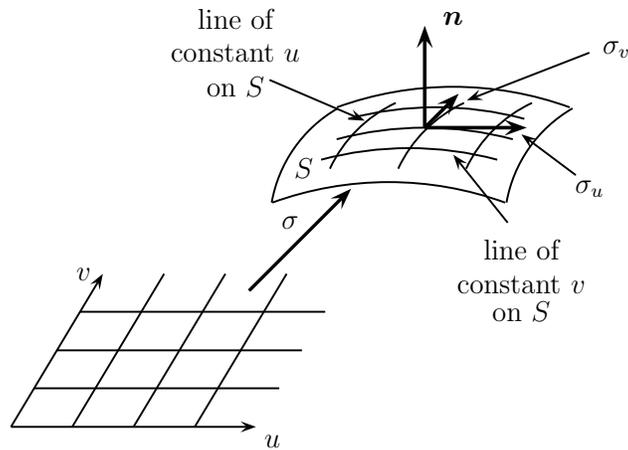
Figure 10: A parameterized surface

vary over some domain $D \subseteq \mathbb{R}^2$, $\sigma(u, v)$ traces out the surface S in $\mathbb{R}^3$. (See Figure 10.)

If we fix v and let u vary, then we get a line of constant v on S. The tangent vector field to this line is just $\sigma_u := \partial\sigma/\partial u$. Similarly, $\sigma_v := \partial\sigma/\partial v$ gives the tangent vector field to the lines of constant u. The normal vector to the surface [15] is therefore

$$\boldsymbol{n} = \sigma_u \times \sigma_v. \tag{4.18}$$

---

[15] When we say 'the' normal vector, we really mean 'a' normal vector, because the vector depends on the parametrization. Even if we normalize the vector to have length one there is still some ambiguity, because a surface has two sides, and therefore two unit normal vectors at every point. If one is interested in, say, the flux through a surface in one direction, then one must select the normal vector in that direction. If the surface is closed, then we usually choose the outward pointing normal in order to be consistent with Gauss' theorem. (See the discussion below.)

Now we *define* the surface integral in terms of its parameterization:

$$\int_S \boldsymbol{A} \cdot d\boldsymbol{S} = \int_D \boldsymbol{A}(\sigma(u,v)) \cdot \boldsymbol{n}\, du dv. \qquad (4.19)$$

Once again, one can show the integral is independent of the parameterization by changing variables.

Example 12 Let $\boldsymbol{A} = (y, 2y, xz)$ and let S be the paraboloid of revolution obtained by rotating the curve $z = 2x^2$ about the z axis, where $0 \leq z \leq 3$. To compute the flux integral we must first parameterize the surface. One possible parameterization is

$$\sigma = (u, v, 2(u^2 + v^2)),$$

while another is

$$\sigma = (u \cos v, u \sin v, 2u^2).$$

Let us choose the latter. Then the domain D is $0 \leq u \leq \sqrt{3/2}$ and $0 \leq v \leq 2\pi$. Also,

$$\sigma_u = (\cos v, \sin v, 4u) \qquad \text{and} \qquad \sigma_v = (-u \sin v, u \cos v, 0),$$

so

$$\boldsymbol{n} = \sigma_u \times \sigma_v = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \cos v & \sin v & 4u \\ -u \sin v & u \cos v & 0 \end{vmatrix} = (-4u^2 \cos v, -4u^2 \sin v, u).$$

(Note that this normal points inward towards the z axis. If you were asked for the

66

flux in the other direction you would have to use $-\boldsymbol{n}$ instead.) Thus

$$
\begin{aligned}
\int_S \boldsymbol{A} \cdot d\boldsymbol{S} &= \int_D \boldsymbol{A}(\sigma(u,v)) \cdot \boldsymbol{n}\, dudv \\
&= \int_D (u \sin v, 2(u \sin v), (u \cos v)(2u^2)) \cdot (-4u^2 \cos v, -4u^2 \sin v, u)\, dudv \\
&= \int_D (-4u^3 \sin v \cos v - 8u^3 \sin^2 v + 2u^4 \cos v) dudv
\end{aligned}
$$

The first and third terms disappear when integrated over v, leaving only

$$
\int_S \boldsymbol{A} \cdot d\boldsymbol{S} = -8 \int_0^{\sqrt{3/2}} u^3\, du \int_0^{2\pi} \sin^2 v\, dv = -2u^4 \Big|_0^{\sqrt{3/2}} \cdot \pi = -\frac{9}{2}\pi.
$$

4.3   Volume Integrals

We will usually consider only volume integrals of scalar fields of the form

$$
\int_V f\, d\tau,
$$

where $d\tau$ is the infinitessimal volume element. For example, in Cartesian coordinates we would compute

$$
\int_V f(x,y,z)\, dx\, dy\, dz.
$$

In any other coordinate system we must use the change of variables theorem.

Theorem 4.1. *Let $F : X \to Y$ be a map from $X \subseteq \mathbb{R}^n$ to $Y \subseteq \mathbb{R}^n$, and let f be an integrable function on Y. Then*

$$
\int_Y f\, dy_1 \ldots dy_n = \int_X (f \circ F)\, |\det J|\, dx_1 \ldots dx_n, \qquad (4.20)
$$

where J is the Jacobian matrix of partial derivatives:

$$J_{ij} = \frac{\partial y_i}{\partial x_j}. \tag{4.21}$$

The map F is the change of variables map, given explicitly by the functions $y_i = F_i(x_1, x_2, \ldots, x_n)$. Geometrically, the theorem says that the integral of f over Y is not equal to the integral of $f \circ F$ over X, because the mapping F distorts volumes. The Jacobian factor compensates precisely for this distortion.

Example 13   Let $f = x^2 y^2 z^2$, and let V be a sphere of radius 2. It makes sense to change variables from Cartesian to spherical polar coordinates because then the domain of integration becomes much simpler. In the above theorem the 'x' coordinates are (r, θ, ϕ) and the 'y' coordinates are (x, y, z). The Jacobian factor can be computed from (3.14):

$$|J| = \begin{vmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{vmatrix} = r^2\sin\theta.$$

Hence

$$\int_V f(x, y, z)\, dx\, dy\, dz = \int_0^2 r^2\, dr \int_0^\pi \sin\theta\, d\theta \int_0^{2\pi} d\phi\ f(r, \theta, \phi)$$

$$= \int_0^2 r^2\, dr \int_0^\pi \sin\theta\, d\theta \int_0^{2\pi} d\phi\ (r^2 \sin^2\theta\cos^2\phi)$$

$$\cdot (r^2 \sin^2\theta\sin^2\phi)(r^2\cos^2\theta).$$

The r integral is

$$\int_0^2 r^8\, dr = \frac{1}{9}r^9 = \frac{512}{9}.$$

The θ integral is

$$\int_0^\pi \sin\theta(\sin^4\theta\cos^2\theta)\,d\theta = -\int_0^\pi (1-\cos^2\theta)^2(\cos^2\theta)d(\cos\theta)$$
$$= \left[-\frac{\cos^3\theta}{3} + \frac{\cos^5\theta}{5}\right]\Big|_0^\pi = \frac{2}{3} - \frac{2}{5} = \frac{4}{15}.$$

The ϕ integral is

$$\int_0^{2\pi} \sin^2\phi\cos^2\phi\,d\phi = \frac{1}{4}\int_0^{2\pi} \sin^2 2\phi\,d\phi = \frac{1}{8}\int_0^{4\pi} \sin^2\phi'\,d\phi' = \frac{\pi}{4}.$$

Putting everything together gives

$$\int_V f\,d\tau = \frac{512}{9}\cdot\frac{4}{15}\cdot\frac{\pi}{4} = \frac{512}{135}\pi.$$

4.4 Problems

1) Verify that each of the following vector fields $\boldsymbol{F}$ is conservative in two ways: first by showing that $\boldsymbol{\nabla}\times\boldsymbol{F} = 0$, and second by finding a function φ such that $\boldsymbol{F} = \boldsymbol{\nabla}\varphi$.

  (a) $\boldsymbol{F} = (1, -z, -y)$.

  (b) $\boldsymbol{F} = (3x^2yz - 3y, x^3z - 3x, x^3y + 2z)$.

  (c) $\boldsymbol{F} = \left(\dfrac{y}{\sqrt{1-x^2y^2}}, \dfrac{x}{\sqrt{1-x^2y^2}}, 0\right)$.

2) By explicitly evaluating the line integrals, calculate the work done by the force field $\boldsymbol{F} = (1, -z, -y)$ on a particle when it is moved from $(1,0,0)$ to $(-1, 0, \pi)$ (i) along the helix $(\cos t, \sin t, t)$, and (ii) along the straight line joining the two points. (iii) Do you expect your answers to (i) and (ii) to be the same? Explain.

69

3) Let $\boldsymbol{A} = (y^2, 2x, 1)$. Evaluate the line integral

$$\int_\gamma \boldsymbol{A} \cdot d\boldsymbol{\ell}$$

between $(0,0,0)$ and $(1,1,1)$, where

(a) γ is the piecewise linear path from $(0,0,0)$ to $(1,0,0)$ to $(1,0,1)$ to $(1,1,1)$.

(b) γ is the path going from $(0,0,0)$ to $(1,1,0)$ along an arc of the circle $x^2 + y^2 - 2y = 0$, and then from $(1,1,0)$ to $(1,1,1)$ along a straight line segment.

(c) Should the answers to (a) and (b) have been the same? Explain.

4) The *helicoid* admits the parameterization $\sigma = (u\cos v, u\sin v, av)$. Compute the area of the helicoid over the domain $0 \le u \le 1$ and $0 \le v \le 2\pi$.

5) Compute the surface integral $\int_S \boldsymbol{F} \cdot d\boldsymbol{S}$, where $\boldsymbol{F} = (1, x^2, xyz)$ and the surface S is given by $z = xy$, with $0 \le x \le y$ and $0 \le y \le 1$.

5   Integral Theorems

Let f be a differentiable function on the interval $[a, b]$. Then, by the Fundamental Theorem of Calculus

$$\int_a^b f'(x)\, dx = f(b) - f(a). \tag{5.1}$$

In this section we discuss a generalization of this theorem to functions of many variables. The best formulation of this theorem is expressed in the language of manifolds and differential forms, which are, unfortunately, slightly beyond the scope of these lectures. Therefore we will have to content ourselves with rather more pedestrian formulations.

5.1 Green's Theorem

The simplest generalization of the Fundamental Theorem of Calculus to two dimensions is **Green's Theorem**. [16] It relates an area integral over a region to a line integral over the boundary of the region.

Let R be a region in the plane with a simple closed curve boundary ∂R.[17] Then we have

Theorem 5.1 (Green's Theorem). *For any differentiable functions P and Q in the plane,*

$$\int_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\, dy = \oint_{\partial R} (P\, dx + Q\, dy), \tag{5.2}$$

where the boundary ∂R is traversed counterclockwise.

Sketch of Proof. The proof of Green's theorem is included in most vector calculus textbooks, but it is worth pointing out some of the basic ideas involved. Consider a square in the plane with lower left corner at (a, a) and upper right corner at (b, b). Then

$$\int_R \frac{\partial P}{\partial y}\, dx\, dy = \int_a^b dx \int_a^b \frac{\partial P(x, y)}{\partial y}\, dy$$
$$= \int_a^b (P(x, b) - P(x, a))\, dx,$$

where the last equality follows from the Fundamental Theorem of Calculus. The boundary ∂R consists of the four sides of the square, oriented as follows: γ_1 from (a, a) to (b, a), γ_2 from (b, a) to (b, b), γ_3 from (b, b) to (a, b), and γ_4

---

[16]Named after the British mathematician and physicist George Green (1793-1841).

[17]In this context, the notation ∂R does *not* mean 'derivative'. Instead it represents the curve that bounds the region R. A *simple closed curve* is a closed curve that has no self-intersections.

from (a, b) to (a, a). Considering the meaning of the line integrals involved, we see that

$$\int_{\gamma_1} P \, dx = \int_a^b P(x, a) \, dx, \qquad \int_{\gamma_3} P \, dx = \int_b^a P(x, b) \, dx,$$

and (since x is fixed along γ_2 and γ_4),

$$\int_{\gamma_2} P \, dx = \int_{\gamma_4} P \, dx = 0,$$

from which it follows that

$$\oint_{\partial R} P \, dx = \oint_{\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4} P \, dx = \int_a^b (P(x, a) - P(x, b)) \, dx.$$

Thus we have shown that

$$\int_R \frac{\partial P}{\partial y} \, dx \, dy = - \oint_{\partial R} P \, dx.$$

A similar argument yields

$$\int_R \frac{\partial Q}{\partial x} \, dx \, dy = \oint_{\partial R} Q \, dy.$$

Adding these two results shows that the theorem is true for squares.

Now consider a rectangular region R' consisting of two squares R_1 and R_2 sharing an edge e. By definition of the integral as a sum,

$$\int_{R'} f \, dx \, dy = \int_{R_1} f \, dx \, dy + \int_{R_2} f \, dx \, dy$$

for any function f. But also

$$\oint_{\partial R'} (g\, dx + h\, dy) = \oint_{\partial R_1} (g\, dx + h\, dy) + \oint_{\partial R_2} (g\, dx + h\, dy)$$

for any functions g and h, because the contribution to the line integral over ∂R_1 coming from e is exactly canceled by the contribution to the line integral over ∂R_2 coming from e, since e is traversed one direction in ∂R_1 and the opposite direction in ∂R_2. It follows that Green's theorem holds for the rectangle R', and, by extension, for any region that can be obtained by pasting together squares along their boundaries. By taking small enough squares, any region in the plane can be built this way, so Green's theorem holds in general. $\qquad\square$

5.2 Stokes' Theorem

Consider a three dimensional vector field of the form $\boldsymbol{A} = P\hat{\boldsymbol{\imath}} + Q\hat{\boldsymbol{\jmath}}$. Note that

$$\boldsymbol{\nabla} \times \boldsymbol{A} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \partial_x & \partial_y & \partial_z \\ P & Q & 0 \end{vmatrix} = (\partial_x Q - \partial_y P)\hat{\boldsymbol{k}}. \tag{5.3}$$

Let S be a region in the xy plane with boundary ∂S. Let $d\boldsymbol{S} = \hat{\boldsymbol{k}}\, dx\, dy$ be the area element on S. Then Green's theorem can be written

$$\int_S (\boldsymbol{\nabla} \times \boldsymbol{A}) \cdot d\boldsymbol{S} = \oint_{\partial S} \boldsymbol{A} \cdot d\boldsymbol{\ell}. \tag{5.4}$$

By a similar argument to that given above in the proof of Green's theorem, this formula holds for *any* reasonable surface S in three dimensional space, provided ∂S is traversed in such a way that the surface normal points 'up-

73

wards on the left' at all times. (Just paste together infinitessimal squares to form S.) In the general case it is known as **Stokes' Theorem**. [18]

Note that this is consistent with the results of Section 4.1. If the vector field $\boldsymbol{A}$ is conservative, then the right side of (5.4) vanishes for every closed curve. Hence the left side of (5.4) vanishes for every open surface S. The only way this can happen is if the integrand vanishes everywhere, which means that $\boldsymbol{A}$ is irrotational. Thus, to test whether a vector field is conservative we need only check whether its curl vanishes.

5.3   Gauss' Theorem

Yet another integral formula, called **Gauss' Theorem** [19] or the **divergence theorem** has the following statement. Let V be a bounded three dimensional region with two dimensional boundary ∂V oriented so that its normal vector points everywhere outward from the volume. Then, for any well behaved vector field $\boldsymbol{A}$,

$$\int_V (\boldsymbol{\nabla} \cdot \boldsymbol{A})\, d\tau = \oint_{\partial V} \boldsymbol{A} \cdot d\boldsymbol{S}, \tag{5.5}$$

where $d\tau$ is the infinitessimal volume element of V.

5.4   The Generalized Stokes' Theorem

There is a clear pattern in all the integral formulae (5.1), (5.4), and (5.5). In each case we have the integral of a derivative of something over an oriented n dimensional region equals the integral of that same something over the

---

[18]Named after the British mathematician and physicist Sir George Gabriel Stokes (1819-1903).

[19]Not to be confused with Gauss' Law.

oriented $n-1$ dimensional boundary of the region. [20] This idea is made rigorous by the theory of **differential forms**. Basically, a differential form ω is something that you integrate. Although the theory of differential forms is beyond the scope of these lectures, I cannot resist giving the elegant generalization and unification of all the results we have discussed so far, just to whet your appetite to investigate the matter more thoroughly on your own:

Theorem 5.2 (Generalized Stokes' Theorem). *If ω is any smooth $n-1$ form with compact support on a smooth oriented n-dimensional surface M, and if the boundary ∂M is given the induced orientation, then*

$$\int_M d\omega = \int_{\partial M} \omega. \qquad (5.6)$$

5.5   Problems

1) Evaluate $\oint_S \boldsymbol{A} \cdot d\boldsymbol{S}$ using Gauss' theorem, where

$$\boldsymbol{A} = (x^2 - y^2, 2xyz, -xz^2),$$

and the surface S bounds the part of a ball of radius 4 that lies in the first octant. (The ball has equation $x^2 + y^2 + z^2 \leq 16$, and the first octant is the region with $x \geq 0$, $y \geq 0$, and $z \geq 0$.)

---

[20]In the case of (5.1) we have an integral over a line segment $[a, b]$ (thought of as oriented from a to b) of the derivative of a function f equals the integral over a zero dimensional region (thought of as oriented positively at b and negatively at a) of f itself, namely $f(b) - f(a)$.

A   Permutations

Let $X = \{1, 2, \ldots, n\}$ be a set of n elements. Informally, a **permutation** of X is just a choice of ordering for the elements of X. More formally, a permutation of X is a bijection [21] $\sigma : X \to X$. The collection of all permutations of X is called the **symmetric group** on n elements and is denoted $\mathfrak{S}_n$. It contains $n!$ elements.

Permutations can be represented in many different ways, but the simplest is just to write down the elements in order. So, for example, if $\sigma(1) = 2$, $\sigma(2) = 4$, $\sigma(3) = 3$, and $\sigma(4) = 1$ then we write $\sigma = 2431$. The **identity** permutation, sometimes denoted e, is just the one satisfying $\sigma(i) = i$ for all i. For example, the identity permutation of $\mathfrak{S}_4$ is $e = 1234$. If σ and τ are two permutations, then the product permutation $\sigma\tau$ is the composite map $\sigma \circ \tau$. That is, $(\sigma\tau)(i) = \sigma(\tau(i))$. For example, if $\tau(1) = 4$, $\tau(2) = 2$, $\tau(3) = 3$, and $\tau(4) = 1$, then $\tau = 4231$ and $\sigma\tau = 2431$. The **inverse** of a permutation σ is just the inverse map σ^{-1}, which satisfies $\sigma\sigma^{-1} = \sigma^{-1}\sigma = e$.

A **transposition** is a permutation that switches two numbers and leaves the rest fixed. For example, the permutation 4231 is a transposition, because it flips 1 and 4 and leaves 2 and 3 alone. It is not too difficult to see that $\mathfrak{S}_n$ is generated by transpositions. This means that any permutation σ may be written as the product of transpositions.

Definition. *A permutation σ is **even** if it can be expressed as the product of an even number of transpositions, otherwise it is **odd**. The **sign** of a permutation σ, written $(-1)^\sigma$, is $+1$ if it is even and -1 if it is odd.*

---

[21] A **bijection** is a map that is one-to-one (so that $i \neq j \Rightarrow \sigma(i) \neq \sigma(j)$) and onto (so that for every k there is an i such that $\sigma(i) = k$).

Example 14   One can show that the sign of a permutation is the number of transpositions required to transform it back to the identity permutation. So 2431 is an even permutation (sign $+1$) because we can get back to the identity permutation in two steps: $2431 \xrightarrow{1 \leftrightarrow 2} 1432 \xrightarrow{2 \leftrightarrow 4} 1234$.

Although a given permutation σ can be written in many different ways as a product of transpositions, it turns out that the sign of σ is always the same. Furthermore, as the notation is meant to suggest, $(-1)^{\sigma\tau} = (-1)^{\sigma}(-1)^{\tau}$. Both these claims require proof, which we omit.

B   Determinants

Definition.   *Let A by an $n \times n$ matrix. The* **determinant** *of A, written* $\det A$ *or* $|A|$*, is the scalar given by*

$$\det A := \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\sigma} A_{1\sigma(1)} A_{2\sigma(2)} \dots A_{n\sigma(n)}. \tag{B.1}$$

Remark.   In general, determinants are difficult to compute because the above sum has $n!$ terms. There are tricks for special kinds of determinants, but few techniques for general matrices. One general method that works nicely in a wide variety of circumstances is called **Dodgson condensation**, named after Charles Lutwidge Dodgson, also known as Lewis Carroll, the inventor of *Alice in Wonderland*. (Look it up.)

Example 15

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{B.2}$$

Example 16 $\det I = 1$ because the only term contributing to the sum in (B.1) is the one in which σ is the identity permutation, and its sign is $+1$.

Definition. *The **transpose** A^T of the matrix A has components*

$$(A^T)_{ij} := A_{ji}. \tag{B.3}$$

Remark. The transpose matrix is obtained simply by flipping the matrix about the main diagonal, which runs from A_{11} to A_{nn}.

Lemma B.1.

$$\det A^T = \det A. \tag{B.4}$$

Proof. An arbitrary term of the expansion of $\det A$ is of the form

$$(-1)^\sigma A_{1\sigma(1)} A_{2\sigma(2)} \ldots A_{n\sigma(n)}. \tag{B.5}$$

As each number from 1 to n appears precisely once among the set $\sigma(1)$, $\sigma(2)$, $\ldots$, $\sigma(n)$, the product may be rewritten (after some rearrangement) as

$$(-1)^\sigma A_{\sigma^{-1}(1)1} A_{\sigma^{-1}(2)2} \ldots A_{\sigma^{-1}(n)n}, \tag{B.6}$$

where σ^{-1} is the inverse permutation to σ. For example, suppose $\sigma(5) = 1$. Then there would be a term in (B.5) of the form $A_{5\sigma(5)} = A_{51}$. This term appears first in (B.6), as $\sigma^{-1}(1) = 5$. Since a permutation and its inverse both have the same sign (because $\sigma\sigma^{-1} = e$ implies $(-1)^\sigma(-1)^{\sigma^{-1}} = 1$), Equation (B.6) may be written

$$(-1)^{\sigma^{-1}} A_{\sigma^{-1}(1)1} A_{\sigma^{-1}(2)2} \dots A_{\sigma^{-1}(n)n}. \tag{B.7}$$

Hence

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\sigma^{-1}} A_{\sigma^{-1}(1)1} A_{\sigma^{-1}(2)2} \dots A_{\sigma^{-1}(n)n}. \tag{B.8}$$

As σ runs over all the elements of S_n, so does σ^{-1}, so (B.8) may be written

$$\det A = \sum_{\sigma^{-1} \in \mathfrak{S}_n} (-1)^{\sigma^{-1}} A_{\sigma^{-1}(1)1} A_{\sigma^{-1}(2)2} \dots A_{\sigma^{-1}(n)n}. \tag{B.9}$$

But this is just $\det A^T$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark. Equation (B.9) shows that we may also write $\det A$ as

$$\det A := \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma A_{\sigma(1)1} A_{\sigma(2)2} \dots A_{\sigma(n)n}. \tag{B.10}$$

B.1 The Determinant as a Multilinear Map

Recall that a map $T : \mathbb{R}^n \to \mathbb{R}$ is **linear** if $T(a\boldsymbol{v} + b\boldsymbol{w}) = aT\boldsymbol{v} + bT\boldsymbol{w}$ for all vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ and all scalars a and b. A map $S : \mathbb{R}^n \times \mathbb{R}^n \times \dots \times \mathbb{R}^n \to \mathbb{R}$

is **multilinear** if S is linear on each entry. That is, we have

$$S(\ldots, a\boldsymbol{v} + b\boldsymbol{w}, \ldots) = aS(\ldots, \boldsymbol{v}, \ldots) + bS(\ldots, \boldsymbol{w}, \ldots). \qquad \text{(B.11)}$$

Theorem B.2. *The determinant, considered as a map on the rows or columns of the matrix, is multilinear.*

Proof. We show that the determinant is linear on the first row of A. A similar argument then shows that it is linear on all the rows or columns, which means it is a multilinear function. Let $A(a\boldsymbol{v} + b\boldsymbol{w}, \ldots)$ be the matrix obtained by replacing the first row of A by the vector $a\boldsymbol{v} + b\boldsymbol{w}$. From (B.1), we have

$$\begin{aligned}
\det A(a\boldsymbol{v} + b\boldsymbol{w}, \ldots) &= \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma (av_{\sigma(1)} + bw_{\sigma(1)}) A_{2\sigma(2)} \cdots A_{n\sigma(n)} \\
&= a \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma v_{\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} \\
&\quad + b \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma w_{\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} \\
&= a \det A(\boldsymbol{v}, \ldots) + b \det A(\boldsymbol{w}, \ldots).
\end{aligned}$$

$\square$

Lemma B.3. *(1) The determinant changes sign whenever any two rows or columns are interchanged. (2) The determinant vanishes if any two rows or columns are equal. (3) The determinant is unchanged if we add a multiple of any row to another row or a multiple of any column to another column.*

Proof. Let B be the matrix A except with rows 1 and 2 flipped. Then

$$
\begin{aligned}
\det B &= \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma B_{1\sigma(1)} B_{2\sigma(2)} \cdots B_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma A_{2\sigma(1)} A_{1\sigma(2)} \cdots A_{n\sigma(n)} \\
&= \sum_{\sigma'\tau \in \mathfrak{S}_n} (-1)^{\sigma'\tau} A_{2(\sigma'\tau)(1)} A_{1(\sigma'\tau)(2)} \cdots A_{n(\sigma'\tau)(n)}. \quad\quad\text{(B.12)}
\end{aligned}
$$

In the last sum we have written the permutation σ in the form $\sigma'\tau$, where τ is the transposition that flips 1 and 2 and σ' is some other permutation. By definition, the action of $\sigma'\tau$ on the numbers $(1, 2, 3, \ldots, n)$ is the same as the action of σ' on the numbers $(2, 1, 3, \ldots, n)$. But by the properties of the sign, $(-1)^{\sigma'\tau} = (-1)^{\sigma'}(-1)^\tau = -(-1)^{\sigma'}$, because all transpositions are odd. Also, σ' ranges over all permutations of $\mathfrak{S}_n$ as $\sigma'\tau$ does, because the map from $\mathfrak{S}_n$ to $\mathfrak{S}_n$ given by right multiplication by τ is bijective. Putting all this together (and switching the order of two of the A terms in (B.12)) gives

$$
\begin{aligned}
\det B &= - \sum_{\sigma' \in \mathfrak{S}_n} (-1)^{\sigma'} A_{1\sigma'(1)} A_{2\sigma'(2)} \cdots A_{n\sigma'(n)} \\
&= -\det A. \quad\quad\text{(B.13)}
\end{aligned}
$$

The same argument holds for columns by starting with (B.10) instead. This proves property (1). Property (2) then follows immediately, because if B is obtained from A by switching two identical rows (or columns), then $B = A$, so $\det B = \det A$. But by Property 1, $\det B = -\det A$, so $\det A = -\det A = 0$. Property (3) now follows by the multilinearity of the determinant. Let $\boldsymbol{v}$ be

the first row of A and let $\boldsymbol{w}$ be any another row of A. Then, for any scalar b,

$$\det A(\boldsymbol{v} + b\boldsymbol{w}, \dots) = \det A(\boldsymbol{v}, \dots) + b \det A(\boldsymbol{w}, \dots) = \det A, \qquad \text{(B.14)}$$

because the penultimate determinant has two identical rows ($\boldsymbol{w}$ appears in the first row and its original row) and so vanishes by Property (2). The same argument works for any rows or columns. □

B.2   Cofactors and the Adjugate

We now wish to derive another way to compute the determinant. To this end, let us investigate the coefficient of A_{11} in $\det A$. By (B.1) it must be

$$\sum_{\sigma' \in \mathfrak{S}_n} (-1)^{\sigma'} A_{2\sigma'(2)} \dots A_{n\sigma'(n)}, \qquad \text{(B.15)}$$

where σ' means a general permutation in S_n that fixes $\sigma(1) = 1$. But this means the sum in (B.15) extends over all permutations of the numbers $\{2, 3, \dots, n\}$, of which there are $(n-1)!$. A moment's reflection reveals that (B.15) is nothing more than the determinant of the matrix obtained from A by removing the first row and first column. As this idea reappears later, we introduce some convenient notation. The $n-1$ by $n-1$ matrix obtained from A by deleting the ith row and j^{th} column is denoted $A(i|j)$.

Definition.   *Let A_{ij} be an element of a matrix A. The* **minor** *of A_{ij} is* $\det A(i|j)$.

By the previous discussion, the coefficient of A_{11} appearing in $\det A$ is precisely its minor, namely $\det A(1|1)$. Now consider a general element A_{ij}. What is its coefficient in $\det A$? Well, consider the matrix A' obtained from A by moving the ith row up to the first row. To get A' we must execute

82

$i-1$ adjacent row flips, so by Lemma B.3, $\det A' = (-1)^{i-1} \det A$. Now consider the matrix A'' obtained from A' by moving the jth column left to the first column. Again by Lemma B.3 we have $\det A'' = (-1)^{j-1} \det A'$. So $\det A'' = (-1)^{i+j} \det A$. Now the element A_{ij} appears in the (11) position in A'', so by the reasoning used above, its coefficient in $\det A''$ is the determinant of $A''(1|1)$. But this is just $\det A(i|j)$. Hence the coefficient of A_{ij} in $\det A$ is $(-1)^{i+j} \det A(i|j)$. This leads to another

Definition. *The* **signed minor** *or* **cofactor** *of A_{ij} is the number given by* $(-1)^{i+j} \det A(i|j)$. *We denote this number by $\overline{A}_{ij}$*

We conclude that the coefficient of A_{ij} in $\det A$ is just its cofactor $\overline{A}_{ij}$.

Now consider the expression

$$A_{11}\overline{A}_{11} + A_{12}\overline{A}_{12} + \cdots + A_{1n}\overline{A}_{1n}. \tag{B.16}$$

Thinking of the A_{ij} as independent variables, each term in (B.16) is distinct (because, for example, only the first term contains A_{11}, *etc.*). Moreover, each term appears in (B.16) precisely as it appears in $\det A$ (with the correct sign and correct products of elements of A). Finally, (B.16) contains $n(n-1)! = n!$ terms, which is the number that appear in $\det A$. So (B.16) must *be* $\det A$. Equation (B.16) is called the **(Laplace) expansion of det A by the first row**.

Thinking back over the argument of the previous paragraph we see there is nothing particularly special about the first row. We could have written a corresponding expression for *any* row or column. Hence we have proved the following

Lemma B.4. *The determinant of A may be written*

$$\det A = \sum_{j=1}^{n} A_{ij}\overline{A}_{ij}, \tag{B.17}$$

for any i, or

$$\det A = \sum_{i=1}^{n} A_{ij}\overline{A}_{ij}, \tag{B.18}$$

for any j.

Remark. This proposition allows us to write the following odd looking but often useful formula for the derivative of the determinant of a matrix with respect to one of its elements (treating them all as independent variables):

$$\frac{\partial}{\partial A_{ij}}\det A = \overline{A}_{ij}. \tag{B.19}$$

We may derive another very useful formula from the following considerations. Suppose we begin with a matrix A and substitute for the ith row a new row of elements labeled B_{ij}, where j runs from 1 to n. Now, the cofactors of the B_{ij} in the new matrix are obviously the same as those of the A_{ij} in the old matrix, so we may write the determinant of the new matrix as, for instance,

$$B_{i1}\overline{A}_{i1} + B_{i2}\overline{A}_{i2} + \cdots + B_{in}\overline{A}_{in}. \tag{B.20}$$

Of course, we could have substituted a new j^{th} column instead, with similar results.

Now suppose we let the B_{ij} be the elements of any row of A other than the ith. Then the expression in Equation (B.20) will vanish, as the determinant

of any matrix with two identical rows is zero. This gives us the following result:

$$A_{k1}\overline{A}_{i1} + A_{k2}\overline{A}_{i2} + \cdots + A_{kn}\overline{A}_{in} = 0, \qquad k \neq i. \tag{B.21}$$

Again, a similar result holds for columns. We call the cofactors appearing in (B.20) **alien cofactors**, because they are the cofactors properly corresponding to the elements A_{ij}, $j = 1, \ldots, n$, of the ith row of A rather than the kth row. We may summarize (B.21) by saying that expansions in terms of alien cofactors vanish identically.

Now we have the following

Definition. *The **adjugate** matrix of A, written $\operatorname{adj} A$ is the transpose of the matrix of cofactors. That is, $(\operatorname{adj} A)_{ij} = \overline{A}_{ji}$.*

Lemma B.5. *For any matrix A we have*

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)I. \tag{B.22}$$

Proof. Consider the ik^{th} element of $A(\operatorname{adj} A)$:

$$[A(\operatorname{adj} A)]_{ik} = \sum_{j=1}^{n} A_{ij}(\operatorname{adj} A)_{jk} = \sum_{j=1}^{n} A_{ij}\overline{A}_{kj}. \tag{B.23}$$

If $i \neq k$ this is an expansion in terms of alien cofactors and vanishes. If $i = k$ then this is just the determinant of A. Hence $[A(\operatorname{adj} A)]_{ik} = (\det A)\delta_{ik}$. This proves the first half. To prove the second half, note that $(\operatorname{adj} A)^T = (\operatorname{adj} A^T)$. That is, the transpose of the adjugate is the adjugate of the transpose. (Just trace back the definitions.) Hence, using the result (whose proof is left to the reader) that $(AB)^T = B^T A^T$ for any matrices A and B,

$$[(\operatorname{adj} A)A)]^T = A^T(\operatorname{adj} A)^T = A^T \operatorname{adj} A^T = (\det A^T)I = (\det A)I. \tag{B.24}$$

□

Definition. *A matrix A is* **singular** *if* $\det A = 0$ *and* **non-singular** *otherwise.*

Lemma B.6. *A matrix A is invertible if and only if it is non-singular. If it is non-singular, its inverse is given by the expression*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A. \tag{B.25}$$

Proof. Follows immediately from Lemma B.5. □

B.3   The Determinant as Multiplicative Homomorphism

Theorem B.7. *Let $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{w}_j\}$ be two collections of n vectors each, related by a matrix A according to*

$$\boldsymbol{w}_j = \sum_{i=1}^{n} A_{ij} \boldsymbol{v}_i. \tag{B.26}$$

Let $D(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be the determinant of the $n \times n$ matrix whose rows are the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. [22] *Then*

$$D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n) = (\det A) \; D(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n). \tag{B.27}$$

Proof. By hypothesis

$$D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$$
$$= D(A_{11}\boldsymbol{v}_1 + A_{21}\boldsymbol{v}_2 + \cdots + A_{n1}\boldsymbol{v}_n, \ldots, A_{1n}\boldsymbol{v}_1 + A_{2n}\boldsymbol{v}_2 + \cdots + A_{nn}\boldsymbol{v}_n).$$

---

[22]Later this expression will mean the determinant of the $n \times n$ matrix whose *columns* are the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. This will not affect any of the results, only the arguments.

Expanding out the right hand side using the multilinearity property of the determinant gives a sum of terms of the form

$$A_{\sigma(1)1}A_{\sigma(2)2}\ldots A_{\sigma(n)n}D(\boldsymbol{v}_{\sigma(1)},\ldots,\boldsymbol{v}_{\sigma(n)}),$$

where σ is an arbitrary map of $\{1,2,\ldots,n\}$ to itself. If σ is not a bijection (*i.e.,* a permutation) then two vector arguments of D will be equal and the entire term will vanish. Hence the only terms that will appear in the sum are those for which $\sigma \in \mathfrak{S}_n$. But now, by a series of transpositions of the arguments, we may write

$$D(\boldsymbol{v}_{\sigma(1)},\ldots,\boldsymbol{v}_{\sigma(n)}) = (-1)^\sigma D(\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_n),$$

where $(-1)^\sigma$ is the sign of the permutation σ. Hence

$$D(\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_n) = \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma A_{\sigma(1)1}A_{\sigma(2)2}\ldots A_{\sigma(n)n}D(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n).$$

$\square$

This brings us to the main theorem of this section, which is a remarkable multiplicative property of determinants. [23]

---

[23]Let X and Y be sets, each equipped with a natural multiplication operation. So, for example, given two elements x_1 and x_2 in X, their product $x_1 x_2$ also belongs to X (and similarly for Y). If ϕ maps elements of X to elements of Y, and if $\phi(x_1 x_2) = \phi(x_1)\phi(x_2)$, then we say that ϕ is a *multiplicative homomorphism* from X to Y. (The word *homomorphism* comes from the Greek 'ομος' ('homos'), meaning the same, and 'μορφη' ('morphe'), meaning shape or form.) Equation (B.28) is expressed mathematically by saying that the determinant is a multiplicative homomorphism from the set of matrices to the set of scalars. Incidentally, both matrices and scalars also come equipped with an addition operation, which makes them into objects called *rings*. A homomorphism that respects both the additive and multiplicative properties of a ring is called a *ring homomorphism*. But the determinant map from matrices to scalars is *nonlinear* (that is, $\det(A+B) \neq \det A + \det B$), so the determinant fails to be a ring homomorphism.

Theorem B.8. *For any two matrices A and B,*

$$\det (AB) = (\det A)(\det B). \tag{B.28}$$

Proof. Choose $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) = (\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_n)$ where $\hat{\boldsymbol{e}}_i$ is the ith canonical basis vector of $\mathbb{R}^n$, and let

$$\boldsymbol{w}_j = \sum_{i=1}^{n} (AB)_{ij} \boldsymbol{v}_i. \tag{B.29}$$

Let $D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ be the determinant of the matrix whose rows are the vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$. Then by Theorem B.7,

$$D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n) = (\det (AB)) D(\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_n). \tag{B.30}$$

On the other hand, if

$$\boldsymbol{u}_k = \sum_{i=1}^{n} A_{ik} \boldsymbol{v}_i, \tag{B.31}$$

then, again by Theorem B.7,

$$D(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n) = (\det A) D(\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_n). \tag{B.32}$$

But expanding (B.29) and using (B.31) gives

$$\boldsymbol{w}_j = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{kj} \boldsymbol{v}_i \tag{B.33}$$

$$= \sum_{k=1}^{n} B_{kj} \boldsymbol{u}_k. \tag{B.34}$$

So using Theorem B.7 again we have, from (B.34)

$$D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n) = (\det B)D(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n). \tag{B.35}$$

Combining (B.32) and (B.35) gives

$$D(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n) = (\det B)(\det A)D(\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_n). \tag{B.36}$$

The proposition now follows by comparing (B.30) and (B.36) and using the fact that $D(\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_n) = 1$. $\qquad\square$

Corollary B.9. *If A is invertible we have*

$$\det (A^{-1}) = (\det A)^{-1}. \tag{B.37}$$

Proof. Just use Theorem B.8:

$$1 = \det I = \det (AA^{-1}) = (\det A)(\det A^{-1}). \tag{B.38}$$

$\qquad\square$

B.4   Cramer's Rule

A few other related results are worth recording.

Theorem B.10 (Cramer's Rule). *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be n column vectors. Let $x_1, \ldots, x_n \in k$ be given, and define*

$$\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{v}_j. \tag{B.39}$$

Then, for each i we have

$$x_i D(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) = D(\boldsymbol{v}_1, \ldots, \underbrace{\boldsymbol{v}}_{i^{\text{th}}\ \text{place}}, \ldots, \boldsymbol{v}_n). \tag{B.40}$$

Proof. Say $i = 1$. By multilinearity we have

$$D(\boldsymbol{v}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) = \sum_{j=1}^{n} x_j D(\boldsymbol{v}_j, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n). \tag{B.41}$$

By Lemma B.3 every term on the right hand side is zero except the term with $j = 1$. □

Remark. Cramer's rule allows us to solve simultaneous sets of linear equations (although there are easier ways).

Example 17 Consider the following system of linear equations:

$$3x_1 + 5x_2 + x_3 = -6$$
$$x_1 - x_2 + 11x_3 = 4$$
$$7x_2 - x_3 = 1.$$

We may write this as

$$x_1 \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 5 \\ -1 \\ 7 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 11 \\ -1 \end{pmatrix} = \begin{pmatrix} -6 \\ 4 \\ 1 \end{pmatrix}.$$

The three column vectors on the left correspond to the vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ above, and constitute the three columns of a matrix we shall call A. By Theorem B.10 this system has the following solution:

$$x_1 = \frac{1}{\det A} \begin{vmatrix} -6 & 5 & 1 \\ 4 & -1 & 11 \\ 1 & 7 & -1 \end{vmatrix}, \quad x_2 = \frac{1}{\det A} \begin{vmatrix} 3 & -6 & 1 \\ 1 & 4 & 11 \\ 0 & 1 & -1 \end{vmatrix}, \quad x_3 = \frac{1}{\det A} \begin{vmatrix} 3 & 5 & -6 \\ 1 & -1 & 4 \\ 0 & 7 & 1 \end{vmatrix}.$$

(Of course, one still has to evaluate all the determinants.)

Corollary B.11. *Let $v_1, v_2, \ldots, v_n$ be the n column vectors of an n by n matrix A. Then these vectors are linearly dependent if and only if*

$$D(v_1, v_2, \ldots, v_n) = \det A = 0.$$

Proof. Suppose the v_i are linearly dependent. Then $w := \sum_i c_i v_i = 0$ for some constants $\{c_i\}_{i=1}^n$, not all of which vanish. Suppose $c_i \neq 0$. Then

$$0 = D(v_1, v_2, \ldots, \underbrace{w}_{i^{\text{th}} \text{ place}}, \ldots, v_n) = c_i \det A,$$

where the first equality follows because a determinant vanishes if one entire column vanishes and the second equality follows from Theorem B.10. Conversely, suppose the v_i are linearly independent. Then we may write $\hat{e}_i = \sum_{j=1}^n B_{ji} v_j$ for some matrix B, where the $\hat{e}_i$ are the canonical basis vectors of $\mathbb{R}^n$. Then by Theorem B.7,

$$1 = D(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n) = (\det B)D(v_1, v_2, \ldots, v_n) = (\det B)(\det A). \quad \text{(B.42)}$$

Hence $\det A$ cannot vanish. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark. Corollary B.11 shows that a set $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ of vectors is a basis for $\mathbb{R}^n$ if and only if $D(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) \neq 0$.