

Multilinear Mappings and Tensors

In this chapter we generalize our earlier discussion of bilinear forms, which leads in a natural manner to the concepts of tensors and tensor products. While we are aware that our approach is not the most general possible (which is to say it is not the most abstract), we feel that it is more intuitive, and hence the best way to approach the subject for the first time. In fact, our treatment is essentially all that is ever needed by physicists, engineers and applied mathematicians. More general treatments are discussed in advanced courses on abstract algebra.

The basic idea is as follows. Given a vector space V with basis $\{e_i\}$, we defined the dual space V^* (with basis $\{\omega^i\}$) as the space of linear functionals on V . In other words, if $\phi = \sum_i \phi_i \omega^i \in V^*$ and $v = \sum_j v^j e_j \in V$, then

$$\begin{aligned}\phi(v) &= \langle \phi, v \rangle = \langle \sum_i \phi_i \omega^i, \sum_j v^j e_j \rangle = \sum_{i,j} \phi_i v^j \delta^i_j \\ &= \sum_i \phi_i v^i .\end{aligned}$$

Next we defined the space $\mathcal{B}(V)$ of all bilinear forms on V (i.e., bilinear mappings on $V \times V$), and we showed (Theorem 9.10) that $\mathcal{B}(V)$ has a basis given by $\{f^{ij} = \omega^i \otimes \omega^j\}$ where

$$f^{ij}(u, v) = \omega^i \otimes \omega^j(u, v) = \omega^i(u)\omega^j(v) = u^i v^j .$$

It is this definition of the f^{ij} that we will now generalize to include linear functionals on spaces such as, for example, $V^* \times V^* \times V^* \times V \times V$.

11.1 DEFINITIONS

Let V be a finite-dimensional vector space over $\mathcal{F}$, and let V^r denote the r -fold Cartesian product $V \times V \times \cdots \times V$. In other words, an element of V^r is an r -tuple $(v_1, \dots, v_r)$ where each $v_i \in V$. If W is another vector space over $\mathcal{F}$, then a mapping $T: V^r \rightarrow W$ is said to be **multilinear** if $T(v_1, \dots, v_r)$ is linear in each variable. That is, T is multilinear if for each $i = 1, \dots, r$ we have

$$\begin{aligned} T(v_1, \dots, av_i + bv'_i, \dots, v_r) \\ = aT(v_1, \dots, v_i, \dots, v_r) + bT(v_1, \dots, v'_i, \dots, v_r) \end{aligned}$$

for all $v_i, v'_i \in V$ and $a, b \in \mathcal{F}$. In the particular case that $W = \mathcal{F}$, the mapping T is variously called an **r -linear form** on V , or a **multilinear form of degree r** on V , or an **r -tensor** on V . The set of all r -tensors on V will be denoted by $\mathcal{T}_r(V)$. (It is also possible to discuss multilinear mappings that take their values in W rather than in $\mathcal{F}$. See Section 11.5.)

As might be expected, we define addition and scalar multiplication on $\mathcal{T}_r(V)$ by

$$\begin{aligned} (S + T)(v_1, \dots, v_r) &= S(v_1, \dots, v_r) + T(v_1, \dots, v_r) \\ (aT)(v_1, \dots, v_r) &= aT(v_1, \dots, v_r) \end{aligned}$$

for all $S, T \in \mathcal{T}_r(V)$ and $a \in \mathcal{F}$. It should be clear that $S + T$ and aT are both r -tensors. With these operations, $\mathcal{T}_r(V)$ becomes a vector space over $\mathcal{F}$. Note that the particular case of $r = 1$ yields $\mathcal{T}_1(V) = V^*$, i.e., the dual space of V , and if $r = 2$, then we obtain a bilinear form on V .

Although this definition takes care of most of what we will need in this chapter, it is worth going through a more general (but not really more difficult) definition as follows. The basic idea is that a tensor is a scalar-valued multilinear function with variables in both V and V^* . Note also that by Theorem 9.4, the space of linear functions on V^* is V^{**} which we view as simply V . For example, a tensor could be a function on the space $V^* \times V \times V$. By convention, we will always write all V^* variables before all V variables, so that, for example, a tensor on $V \times V^* \times V$ will be replaced by a tensor on $V^* \times V \times V$. (However, not all authors adhere to this convention, so the reader should be very careful when reading the literature.)

Without further ado, we define a **tensor** T on V to be a multilinear map on $V^{*s} \times V^r$:

$$T: V^{*s} \times V^r = \underbrace{V^* \times \cdots \times V^*}_{s \text{ copies}} \times \underbrace{V \times \cdots \times V}_{r \text{ copies}} \rightarrow \mathcal{F}$$

where r is called the **covariant order** and s is called the **contravariant order** of T . We shall say that a tensor of covariant order r and contravariant order s is of **type** (or **rank**) $\binom{s}{r}$. If we denote the set of all tensors of type $\binom{s}{r}$ by $\mathcal{T}_r^s(V)$, then defining addition and scalar multiplication exactly as above, we see that $\mathcal{T}_r^s(V)$ forms a vector space over $\mathcal{F}$. A tensor of type $\binom{0}{0}$ is defined to be a scalar, and hence $\mathcal{T}_0^0(V) = \mathcal{F}$. A tensor of type $\binom{0}{1}$ is called a **contravariant vector**, and a tensor of type $\binom{1}{0}$ is called a **covariant vector** (or simply a **covector**). In order to distinguish between these types of vectors, we denote the basis vectors for V by a subscript (e.g., e_i), and the basis vectors for V^* by a superscript (e.g., ω^j). Furthermore, we will generally leave off the V and simply write $\mathcal{T}_r$ or $\mathcal{T}_r^s$.

At this point we are virtually forced to introduce the so-called **Einstein summation convention**. This convention says that we are to sum over repeated indices in any vector or tensor expression where one index is a superscript and one is a subscript. Because of this, we write the vector components with indices in the opposite position from that of the basis vectors. This is why we have been writing $v = \sum_i v^i e_i \in V$ and $\phi = \sum_j \phi_j \omega^j \in V^*$. Thus we now simply write $v = v^i e_i$ and $\phi = \phi_j \omega^j$ where the summation is to be understood. Generally the limits of the sum will be clear. However, we will revert to the more complete notation if there is any possibility of ambiguity.

It is also worth emphasizing the trivial fact that the indices summed over are just “dummy indices.” In other words, we have $v^i e_i = v^j e_j$ and so on. Throughout this chapter we will be relabelling indices in this manner without further notice, and we will assume that the reader understands what we are doing.

Suppose $T \in \mathcal{T}_r$, and let $\{e_1, \dots, e_n\}$ be a basis for V . For each $i = 1, \dots, r$ we define a vector $v_i = e_j a_i^j$ where, as usual, $a_i^j \in \mathcal{F}$ is just the j th component of the vector v_i . (Note that here the subscript i is not a tensor index.) Using the multilinearity of T we see that

$$T(v_1, \dots, v_r) = T(e_{j_1} a_1^{j_1}, \dots, e_{j_r} a_r^{j_r}) = a_1^{j_1} \cdots a_r^{j_r} T(e_{j_1}, \dots, e_{j_r}) .$$

The n^r scalars $T(e_{j_1}, \dots, e_{j_r})$ are called the **components** of T relative to the basis $\{e_i\}$, and are denoted by $T_{j_1 \dots j_r}$. This terminology implies that there exists a basis for $\mathcal{T}_r$ such that the $T_{j_1 \dots j_r}$ are just the components of T with

respect to this basis. We now construct this basis, which will prove that $\mathcal{T}_r$ is of dimension n^r .

(We will show formally in Section 11.10 that the Kronecker symbols δ_j^i are in fact the components of a tensor, and that these components are the same in any coordinate system. However, for all practical purposes we continue to use the δ_j^i simply as a notational device, and hence we place no importance on the position of the indices, i.e., $\delta_j^i = \delta_i^j$ etc.)

For each collection $\{i_1, \dots, i_r\}$ (where $1 \leq i_k \leq n$), we define the tensor $\Omega^{i_1 \dots i_r}$ (not simply the components of a tensor Ω) to be that element of $\mathcal{T}_r$ whose values on the basis $\{e_i\}$ for V are given by

$$\Omega^{i_1 \dots i_r}(e_{j_1}, \dots, e_{j_r}) = \delta_{j_1}^{i_1} \dots \delta_{j_r}^{i_r}$$

and whose values on an arbitrary collection $\{v_1, \dots, v_r\}$ of vectors are given by multilinearity as

$$\begin{aligned} \Omega^{i_1 \dots i_r}(v_1, \dots, v_r) &= \Omega^{i_1 \dots i_r}(e_{j_1} a^{j_1}_{1_1}, \dots, e_{j_r} a^{j_r}_{r_r}) \\ &= a^{j_1}_{1_1} \dots a^{j_r}_{r_r} \Omega^{i_1 \dots i_r}(e_{j_1}, \dots, e_{j_r}) \\ &= a^{j_1}_{1_1} \dots a^{j_r}_{r_r} \delta_{j_1}^{i_1} \dots \delta_{j_r}^{i_r} \\ &= a^{i_1}_{1_1} \dots a^{i_r}_{r_r} . \end{aligned}$$

That this does indeed define a tensor is guaranteed by this last equation which shows that each $\Omega^{i_1 \dots i_r}$ is in fact linear in each variable (since $v_1 + v'_1 = (a^{j_1}_{1_1} + a'^{j_1}_{1_1})e_{j_1}$ etc.). To prove that the n^r tensors $\Omega^{i_1 \dots i_r}$ form a basis for $\mathcal{T}_r$, we must show that they linearly independent and span $\mathcal{T}_r$.

Suppose that $\alpha_{i_1} \dots \alpha_{i_r} \Omega^{i_1 \dots i_r} = 0$ where each $\alpha_{i_1} \dots \alpha_{i_r} \in \mathcal{F}$. From the definition of $\Omega^{i_1 \dots i_r}$, we see that applying this to any r -tuple $(e_{j_1}, \dots, e_{j_r})$ of basis vectors yields $\alpha_{j_1} \dots \alpha_{j_r} = 0$. Since this is true for every such r -tuple, it follows that $\alpha_{i_1} \dots \alpha_{i_r} = 0$ for every r -tuple of indices $(i_1, \dots, i_r)$, and hence the $\Omega^{i_1 \dots i_r}$ are linearly independent.

Now let $T_{i_1 \dots i_r} = T(e_{i_1}, \dots, e_{i_r})$ and consider the tensor

$$T_{i_1 \dots i_r} \Omega^{i_1 \dots i_r}$$

in $\mathcal{T}_r$. Using the definition of $\Omega^{i_1 \dots i_r}$, we see that both $T_{i_1 \dots i_r} \Omega^{i_1 \dots i_r}$ and T yield the same result when applied to any r -tuple $(e_{j_1}, \dots, e_{j_r})$ of basis vectors, and hence they must be equal as multilinear functions on V^r . This shows that $\{\Omega^{i_1 \dots i_r}\}$ spans $\mathcal{T}_r$.

While we have treated only the space $\mathcal{T}_r$, it is not any more difficult to treat the general space $\mathcal{T}_r^s$. Thus, if $\{e_i\}$ is a basis for V , $\{\omega^j\}$ is a basis for V^* and $T \in \mathcal{T}_r^s$, we define the **components** of T (relative to the given bases) by

$$T^{i_1 \cdots i_s}_{j_1 \cdots j_r} = T(\omega^{i_1}, \dots, \omega^{i_s}, e_{j_1}, \dots, e_{j_r}) .$$

Defining the n^{r+s} analogous tensors $\Omega_{i_1^1 \cdots i_s^s}^{j_1^1 \cdots j_r^r}$, it is easy to mimic the above procedure and hence prove the following result.

Theorem 11.1 The set $\mathcal{T}_r^s$ of all tensors of type $(\overset{s}{r})$ on V forms a vector space of dimension n^{r+s} .

Proof This is Exercise 11.1.1. ■

Since a tensor $T \in \mathcal{T}_r^s$ is a function on $V^{*s} \times V^r$, it would be nice if we could write a basis (e.g., $\Omega_{i_1^1 \cdots i_s^s}^{j_1^1 \cdots j_r^r}$) for $\mathcal{T}_r^s$ in terms of the bases $\{e_i\}$ for V and $\{\omega^j\}$ for V^* . We now show that this is easy to accomplish by defining a product on $\mathcal{T}_r^s$, called the tensor product. The reader is cautioned not to be intimidated by the notational complexities, since the concepts involved are really quite simple.

Suppose that $S \in \mathcal{T}_{r_1}^{s_1}$ and $T \in \mathcal{T}_{r_2}^{s_2}$. Let $u_1, \dots, u_{r_1}, v_1, \dots, v_{r_2}$ be vectors in V , and $\alpha^1, \dots, \alpha^{s_1}, \beta^1, \dots, \beta^{s_2}$ be covectors in V^* . Note that the product

$$S(\alpha^1, \dots, \alpha^{s_1}, u_1, \dots, u_{r_1}) T(\beta^1, \dots, \beta^{s_2}, v_1, \dots, v_{r_2})$$

is linear in each of its $r_1 + s_1 + r_2 + s_2$ variables. Hence we define the **tensor product** $S \otimes T \in \mathcal{T}_{r_1+r_2}^{s_1+s_2}$ (read “ S tensor T ”) by

$$(S \otimes T)(\alpha^1, \dots, \alpha^{s_1}, \beta^1, \dots, \beta^{s_2}, u_1, \dots, u_{r_1}, v_1, \dots, v_{r_2}) = \\ S(\alpha^1, \dots, \alpha^{s_1}, u_1, \dots, u_{r_1}) T(\beta^1, \dots, \beta^{s_2}, v_1, \dots, v_{r_2}) .$$

It is easily shown that the tensor product is both associative and distributive (i.e., bilinear in both factors). In other words, for any scalar $a \in \mathcal{F}$ and tensors R, S and T such that the following formulas make sense, we have

$$\begin{aligned} (R \otimes S) \otimes T &= R \otimes (S \otimes T) \\ R \otimes (S + T) &= R \otimes S + R \otimes T \\ (R + S) \otimes T &= R \otimes T + S \otimes T \\ (aS) \otimes T &= S \otimes (aT) = a(S \otimes T) \end{aligned}$$

(see Exercise 11.1.2). Because of the associativity property (which is a consequence of associativity in $\mathcal{F}$), we will drop the parentheses in expressions such as the top equation and simply write $R \otimes S \otimes T$. This clearly extends to any finite product of tensors. It is important to note, however, that the tensor product is most certainly *not* commutative.

Now let $\{e_1, \dots, e_n\}$ be a basis for V , and let $\{\omega^j\}$ be its dual basis. We claim that the set $\{\omega^{j_1} \otimes \dots \otimes \omega^{j_r}\}$ of tensor products where $1 \leq j_k \leq n$ forms a basis for the space $\mathcal{T}_r$ of covariant tensors. To see this, we note that from the definitions of tensor product and dual space, we have

$$\omega^{j_1} \otimes \dots \otimes \omega^{j_r}(e_{i_1}, \dots, e_{i_r}) = \omega^{j_1}(e_{i_1}) \dots \omega^{j_r}(e_{i_r}) = \delta^{j_1}_{i_1} \dots \delta^{j_r}_{i_r}$$

so that $\omega^{j_1} \otimes \dots \otimes \omega^{j_r}$ and $\Omega^{j_1 \dots j_r}$ take the same values on the r -tuples $(e_{i_1}, \dots, e_{i_r})$, and hence they must be equal as multilinear functions on V^r . Since we showed above that $\{\Omega^{j_1 \dots j_r}\}$ forms a basis for $\mathcal{T}_r$, we have proved that $\{\omega^{j_1} \otimes \dots \otimes \omega^{j_r}\}$ also forms a basis for $\mathcal{T}_r$.

The method of the previous paragraph is readily extended to the space $\mathcal{T}_r^s$. We must recall however, that we are treating V^{**} and V as the same space. If $\{e_i\}$ is a basis for V , then the dual basis $\{\omega^j\}$ for V^* was defined by $\omega^j(e_i) = \langle \omega^j, e_i \rangle = \delta^j_i$. Similarly, given a basis $\{\omega^j\}$ for V^* , we define the basis $\{e_i\}$ for $V^{**} = V$ by $e_i(\omega^j) = \omega^j(e_i) = \delta^j_i$. In fact, using tensor products, it is now easy to repeat Theorem 11.1 in its most useful form. Note also that the next theorem shows that a tensor is determined by its values on the bases $\{e_i\}$ and $\{\omega^j\}$.

Theorem 11.2 Let V have basis $\{e_1, \dots, e_n\}$, and let V^* have the corresponding dual basis $\{\omega^1, \dots, \omega^n\}$. Then a basis for $\mathcal{T}_r^s$ is given by the collection

$$\{e_{i_1} \otimes \dots \otimes e_{i_s} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_r}\}$$

where $1 \leq j_1, \dots, j_r, i_1, \dots, i_s \leq n$, and hence $\dim \mathcal{T}_r^s = n^{r+s}$.

Proof In view of Theorem 11.1, all that is needed is to show that

$$e_{i_1} \otimes \dots \otimes e_{i_s} \otimes \omega^{j_1} \otimes \dots \otimes \omega^{j_r} = \Omega^{j_1 \dots j_r}_{i_1 \dots i_s}.$$

The details are left to the reader (see Exercise 11.1.1). ■

Since the components of a tensor T are defined with respect to a particular basis (and dual basis), we might ask about the relationship between the com-

ponents of T relative to two different bases. Using the multilinearity of tensors, this is a simple problem to solve.

First, let $\{e_i\}$ be a basis for V and let $\{\omega^j\}$ be its dual basis. If $\{\bar{e}_i\}$ is another basis for V , then there exists a nonsingular transition matrix $A = (a^j_i)$ such that

$$\bar{e}_i = e_j a^j_i . \quad (1)$$

(We emphasize that a^j_i is only a matrix, not a tensor. Note also that our definition of the matrix of a linear transformation given in Section 5.3 shows that a^j_i is the element of A in the j th row and i th column.) Using $\langle \omega^i, e_j \rangle = \delta^i_j$, we have

$$\langle \omega^i, \bar{e}_k \rangle = \langle \omega^i, e_j a^j_k \rangle = a^j_k \langle \omega^i, e_j \rangle = a^j_k \delta^i_j = a^i_k .$$

Let us denote the inverse of the matrix $A = (a^j_i)$ by $A^{-1} = B = (b^i_j)$. In other words, $a^i_j b^j_k = \delta^i_k$ and $b^i_j a^j_k = \delta^i_k$. Multiplying $\langle \omega^i, \bar{e}_k \rangle = a^i_k$ by b^j_i and summing on i yields

$$\langle b^j_i \omega^i, \bar{e}_k \rangle = b^j_i a^i_k = \delta^j_k .$$

But the basis $\{\bar{\omega}^i\}$ dual to $\{\bar{e}_i\}$ also must satisfy $\langle \bar{\omega}^j, \bar{e}_k \rangle = \delta^j_k$, and hence comparing this with the previous equation shows that the dual basis vectors transform as

$$\bar{\omega}^j = b^j_i \omega^i \quad (2)$$

The reader should compare this carefully with (1). We say that the dual basis vectors transform **oppositely** (i.e., use the inverse transformation matrix) to the basis vectors. It is also worth emphasizing that if the nonsingular transition matrix from the basis $\{e_i\}$ to the basis $\{\bar{e}_i\}$ is given by A , then (according to the same convention given in Section 5.4) the corresponding nonsingular transition matrix from the basis $\{\omega^i\}$ to the basis $\{\bar{\omega}^i\}$ is given by $B^T = (A^{-1})^T$. We leave it to the reader to write out equations (1) and (2) in matrix notation to show that this is true (see Exercise 11.1.3).

We now return to the question of the relationship between the components of a tensor in two different bases. For definiteness, we will consider a tensor $T \in \mathcal{T}_1^2$. The analogous result for an arbitrary tensor in $\mathcal{T}_r^s$ will be quite obvious. Let $\{e_i\}$ and $\{\omega^j\}$ be a basis and dual basis for V and V^* respectively. Now consider another pair of bases $\{\bar{e}_i\}$ and $\{\bar{\omega}^j\}$ where $\bar{e}_i = e_j a^j_i$ and $\bar{\omega}^i = b^i_j \omega^j$. Then we have $T^{ij}_k = T(\omega^i, \omega^j, e_k)$ as well as $\bar{T}^{pq}_r = T(\bar{\omega}^p, \bar{\omega}^q, \bar{e}_r)$, and therefore

$$\bar{T}^{p\ q}_{\ r} = T(\bar{\omega}^p, \bar{\omega}^q, \bar{e}_r) = b^p_i b^q_j a^k_r T(\omega^i, \omega^j, e_k) = b^p_i b^q_j a^k_r T^{ij}_k .$$

This is the **classical law of transformation** of the components of a tensor of type $(\overset{2}{1})$. It should be kept in mind that (a^i_j) and (b^i_j) are inverse matrices to each other. (In fact, this equation is frequently taken as the *definition* of a tensor (at least in older texts). In other words, according to this approach, any quantity with this transformation property is defined to be a tensor.)

In particular, the components v^i of a vector $v = v^i e_i$ transform as

$$\bar{v}^i = b^i_j v^j$$

while the components α_i of a covector $\alpha = \alpha_i \omega^i$ transform as

$$\bar{\alpha}_i = \alpha_j a^j_i .$$

We leave it to the reader to verify that these transformation laws lead to the self-consistent formulas $v = v^i e_i = \bar{v}^j \bar{e}_j$ and $\alpha = \alpha_i \omega^i = \bar{\alpha}_j \bar{\omega}^j$ as we should expect (see Exercise 11.1.4).

We point out that these transformation laws are the origin of the terms “contravariant” and “covariant.” This is because the components of a vector transform oppositely (“contravariant”) to the basis vectors e_i , while the components of dual vectors transform the same as (“covariant”) these basis vectors.

It is also worth mentioning that many authors use a prime (or some other method such as a different type of letter) for distinguishing different bases. In other words, if we have a basis $\{e_i\}$ and we wish to transform to another basis which we denote by $\{e'_i\}$, then this is accomplished by a transformation matrix $(a^{i'}_j)$ so that $e_{i'} = e_j a^j_{i'}$. In this case, we would write $\omega^{i'} = a^{i'}_j \omega^j$ where $(a^{i'}_j)$ is the inverse of $(a^i_{j'})$. In this notation, the transformation law for the tensor T used above would be written as

$$T^{p'q'}_{\ r'} = b^{p'}_i b^{q'}_j a^k_{r'} T^{ij}_k .$$

Note that specifying the components of a tensor with respect to one coordinate system allows the determination of its components with respect to any other coordinate system. Because of this, we shall frequently refer to a tensor by its “generic” components. In other words, we will refer to e.g., T^{ij}_k , as a “tensor” and not the more accurate description as the “components of the tensor T .”

Example 11.1 For those readers who may have seen a classical treatment of tensors and have had a course in advanced calculus, we will now show how our more modern approach agrees with the classical.

If $\{x^i\}$ is a local coordinate system on a differentiable manifold X , then a (tangent) vector field $v(x)$ on X is defined as the derivative function $v = v^i(\partial/\partial x^i)$, so that $v(f) = v^i(\partial f/\partial x^i)$ for every smooth function $f: X \rightarrow \mathbb{R}$ (and where each v^i is a function of position $x \in X$, i.e., $v^i = v^i(x)$). Since every vector at $x \in X$ can in this manner be written as a linear combination of the $\partial/\partial x^i$, we see that $\{\partial/\partial x^i\}$ forms a basis for the tangent space at x .

We now define the **differential** df of a function by $df(v) = v(f)$ and thus $df(v)$ is just the directional derivative of f in the direction of v . Note that

$$dx^i(v) = v(x^i) = v^j(\partial x^i/\partial x^j) = v^j \delta_j^i = v^i$$

and hence $df(v) = v^i(\partial f/\partial x^i) = (\partial f/\partial x^i)dx^i(v)$. Since v was arbitrary, we obtain the familiar elementary formula $df = (\partial f/\partial x^i)dx^i$. Furthermore, we see that

$$dx^i(\partial/\partial x^j) = \partial x^i/\partial x^j = \delta_j^i$$

so that $\{dx^i\}$ forms the basis dual to $\{\partial/\partial x^i\}$. In summary then, relative to the local coordinate system $\{x^i\}$, we define a basis $\{e_i = \partial/\partial x^i\}$ for a (tangent) space V along with the dual basis $\{\omega^j = dx^j\}$ for the (cotangent) space V^* .

If we now go to a new coordinate system $\{\bar{x}^i\}$ in the same coordinate patch, then from calculus we obtain

$$\partial/\partial \bar{x}^i = (\partial x^j/\partial \bar{x}^i)\partial/\partial x^j$$

so that the expression $\bar{e}_i = e_j a_j^i$ implies $a_j^i = \partial x^j/\partial \bar{x}^i$. Similarly, we also have

$$d\bar{x}^i = (\partial \bar{x}^i/\partial x^j)dx^j$$

so that $\bar{\omega}^i = b_j^i \omega^j$ implies $b_j^i = \partial \bar{x}^i/\partial x^j$. Note that the chain rule from calculus shows us that

$$a_k^i b_j^k = (\partial x^i/\partial \bar{x}^k)(\partial \bar{x}^k/\partial x^j) = \partial x^i/\partial x^j = \delta_j^i$$

and thus (b_j^i) is indeed the inverse matrix to (a_j^i) .

Using these results in the above expression for $\bar{T}^{pq}_r$, we see that

$$\bar{T}^{pq}{}_r = \frac{\partial \bar{x}^p}{\partial x^i} \frac{\partial \bar{x}^q}{\partial x^j} \frac{\partial x^k}{\partial \bar{x}^r} T^{ij}{}_k$$

which is just the classical definition of the transformation law for a tensor of type $\binom{2}{1}$.

We also remark that in older texts, a contravariant vector is *defined* to have the same transformation properties as the expression $d\bar{x}^i = (\partial \bar{x}^i / \partial x^j) dx^j$, while a covariant vector is *defined* to have the same transformation properties as the expression $\partial / \partial \bar{x}^i = (\partial x^j / \partial \bar{x}^i) \partial / \partial x^j$. //

Finally, let us define a simple classical tensor operation that is frequently quite useful. To begin with, we have seen that the result of operating on a vector $v = v^i e_i \in V$ with a dual vector $\alpha = \alpha_j \omega^j \in V^*$ is just $\langle \alpha, v \rangle = \alpha_j v^i \langle \omega^j, e_i \rangle = \alpha_j v^i \delta^j_i = \alpha_i v^i$. This is sometimes called the **contraction** of α with v . We leave it to the reader to show that the contraction is independent of the particular coordinate system used (see Exercise 11.1.5).

If we start with tensors of higher order, then we can perform the same sort of operation. For example, if we have $S \in \mathcal{T}_2^1$ with components $S^i{}_{jk}$ and $T \in \mathcal{T}^2$ with components T^{pq} , then we can form the $\binom{2}{1}$ tensor with components $S^i{}_{jk} T^{jq}$, or a different $\binom{2}{1}$ tensor with components $S^i{}_{jk} T^{pj}$ and so forth. This operation is also called **contraction**. Note that if we start with a $\binom{1}{1}$ tensor T , then we can contract the components of T to obtain the scalar $T^i{}_i$. This is called the **trace** of T .

Exercises

1. (a) Prove Theorem 11.1.
(b) Prove Theorem 11.2.
2. Prove the four associative and distributive properties of the tensor product given in the text following Theorem 11.1.
3. If the nonsingular transition matrix from a basis $\{e_i\}$ to a basis $\{\bar{e}_i\}$ is given by $A = (a^i{}_j)$, show that the transition matrix from the corresponding dual bases $\{\omega^i\}$ and $\{\bar{\omega}^i\}$ is given by $(A^{-1})^T$.
4. Using the transformation matrices $(a^i{}_j)$ and $(b^i{}_j)$ for the bases $\{e_i\}$ and $\{\bar{e}_i\}$ and the corresponding dual bases $\{\omega^i\}$ and $\{\bar{\omega}^i\}$, verify that $v = v^i e_i = \bar{v}^j \bar{e}_j$ and $\alpha = \alpha_i \omega^i = \bar{\alpha}_j \bar{\omega}^j$.

5. If $v \in V$ and $\alpha \in V^*$, show that $\langle \alpha, v \rangle$ is independent of the particular basis chosen for V . Generalize this to arbitrary tensors.
6. Let A_i be a covariant vector field (i.e., $A_i = A_i(x)$) with the transformation rule

$$\bar{A}_i = \frac{\partial x^j}{\partial \bar{x}^i} A_j .$$

Show that the quantity $\partial_j A_i = \partial A_i / \partial x^j$ does not define a tensor, but that $F_{ij} = \partial_i A_j - \partial_j A_i$ is in fact a second-rank tensor.

11.2 SPECIAL TYPES OF TENSORS

In order to obtain some of the most useful results concerning tensors, we turn our attention to the space $\mathcal{T}_r$ of covariant tensors on V . Generalizing our earlier definition for bilinear forms, we say that a tensor $S \in \mathcal{T}_r$ is **symmetric** if for each pair (i, j) with $1 \leq i, j \leq r$ and all $v_i \in V$ we have

$$S(v_1, \dots, v_i, \dots, v_j, \dots, v_r) = S(v_1, \dots, v_j, \dots, v_i, \dots, v_r) .$$

Similarly, $A \in \mathcal{T}_r$ is said to be **antisymmetric** (or **skew-symmetric** or **alternating**) if

$$A(v_1, \dots, v_i, \dots, v_j, \dots, v_r) = -A(v_1, \dots, v_j, \dots, v_i, \dots, v_r) .$$

Note this definition implies that $A(v_1, \dots, v_r) = 0$ if any two of the v_i are identical. In fact, this was the original definition of an alternating bilinear form. Furthermore, we also see that $A(v_1, \dots, v_r) = 0$ if any v_i is a linear combination of the rest of the v_j . In particular, this means that we must always have $r \leq \dim V$ if we are to have a nonzero antisymmetric tensor of type $\binom{0}{r}$ on V .

It is easy to see that if $S_1, S_2 \in \mathcal{T}_r$ are symmetric, then so is $aS_1 + bS_2$ where $a, b \in \mathcal{F}$. Similarly, $aA_1 + bA_2$ is antisymmetric. Therefore the symmetric tensors form a subspace of $\mathcal{T}_r$ which we denote by $\Sigma^r(V)$, and the antisymmetric tensors form another subspace of $\mathcal{T}_r$ which is denoted by $\Lambda^r(V)$ (some authors denote this space by $\Lambda^r(V^*)$). Elements of $\Lambda^r(V)$ are generally called **exterior r-forms**, or simply **r-forms**. According to this terminology, the basis vectors $\{\omega^i\}$ for V^* are referred to as **basis 1-forms**. Note that the only element common to both of these subspaces is the zero tensor.

A particularly important example of an antisymmetric tensor is the determinant function $\det \in \mathcal{T}_n(\mathbb{R}^n)$ (see Theorem 4.9 and the discussion preceding it). Note also that the definition of a symmetric tensor translates into the obvious requirement that (e.g., in the particular case of $\mathcal{T}_2$) $S_{ij} = S_{ji}$, while an antisymmetric tensor obeys $A_{ij} = -A_{ji}$. These definitions can also be extended to include contravariant tensors, although we shall have little need to do so.

It will be extremely convenient for us to now incorporate the treatment of permutation groups given in Section 1.2. In terms of any permutation $\sigma \in S_r$, we may rephrase the above definitions as follows. We say that $S \in \mathcal{T}_r$ is **symmetric** if for every collection $v_1, \dots, v_r \in V$ and each $\sigma \in S_r$ we have

$$S(v_1, \dots, v_r) = S(v_{\sigma 1}, \dots, v_{\sigma r}) .$$

Similarly, we say that $A \in \mathcal{T}_r$ is **antisymmetric** (or **alternating**) if either

$$A(v_1, \dots, v_r) = (\operatorname{sgn} \sigma) A(v_{\sigma 1}, \dots, v_{\sigma r})$$

or

$$A(v_{\sigma 1}, \dots, v_{\sigma r}) = (\operatorname{sgn} \sigma) A(v_1, \dots, v_r)$$

where the last equation follows from the first since $(\operatorname{sgn} \sigma)^2 = 1$. Note that even if $S, T \in \Sigma^r(V)$ are both symmetric, it need not be true that $S \otimes T$ be symmetric (i.e., $S \otimes T \notin \Sigma^{r+r}(V)$). For example, if $S_{ij} = S_{ji}$ and $T_{pq} = T_{qp}$, it does not necessarily follow that $S_{ij}T_{pq} = S_{ip}T_{jq}$. It is also clear that if $A, B \in \bigwedge^r(V)$, then we do not necessarily have $A \otimes B \in \bigwedge^{r+r}(V)$.

Example 11.2 Suppose $\alpha \in \bigwedge^n(V)$, let $\{e_1, \dots, e_n\}$ be a basis for V , and for each $i = 1, \dots, n$ let $v_i = e_j a_i^j$ where $a_i^j \in \mathcal{F}$. Then, using the multilinearity of α , we may write

$$\alpha(v_1, \dots, v_n) = a_1^{j_1} \cdots a_n^{j_n} \alpha(e_{j_1}, \dots, e_{j_n})$$

where the sums are over all $1 \leq j_k \leq n$. But $\alpha \in \bigwedge^n(V)$ is antisymmetric, and hence $(e_{j_1}, \dots, e_{j_n})$ must be a permutation of $(e_1, \dots, e_n)$ in order that the e_{j_k} all be distinct (or else $\alpha(e_{j_1}, \dots, e_{j_n}) = 0$). This means that we are left with

$$\alpha(v_1, \dots, v_n) = \sum_J a_1^{j_1} \cdots a_n^{j_n} \alpha(e_{j_1}, \dots, e_{j_n})$$

where $\sum_J$ denotes the fact that we are summing over only those values of j_k such that $(j_1, \dots, j_n)$ is a permutation of $(1, \dots, n)$. In other words, we have

$$\alpha(v_1, \dots, v_n) = \sum_{\sigma \in S_n} a^{\sigma 1}_1 \cdots a^{\sigma n}_n \alpha(e_{\sigma 1}, \dots, e_{\sigma n}) .$$

But now, by the antisymmetry of α , we see that $\alpha(e_{\sigma 1}, \dots, e_{\sigma n}) = (\text{sgn } \sigma) \alpha(e_1, \dots, e_n)$ and hence we are left with

$$\alpha(v_1, \dots, v_n) = \sum_{\sigma \in S_n} (\text{sgn } \sigma) a^{\sigma 1}_1 \cdots a^{\sigma n}_n \alpha(e_1, \dots, e_n) . \quad (*)$$

Using the definition of determinant and the fact that $\alpha(e_1, \dots, e_n)$ is just some scalar, we finally obtain

$$\alpha(v_1, \dots, v_n) = \det(a^j_i) \alpha(e_1, \dots, e_n) .$$

Referring back to Theorem 4.9, let us consider the special case where $\alpha(e_1, \dots, e_n) = 1$. Note that if $\{\omega^j\}$ is a basis for V^* , then

$$\omega^{\sigma j}(v_i) = \omega^{\sigma j}(e_k a^k_i) = a^k_i \omega^{\sigma j}(e_k) = a^k_i \delta^{\sigma j}_k = a^{\sigma j}_i .$$

Using the definition of tensor product, we can therefore write (*) as

$$\det(a^j_i) = \alpha(v_1, \dots, v_n) = \sum_{\sigma \in S_n} (\text{sgn } \sigma) \omega^{\sigma 1} \otimes \cdots \otimes \omega^{\sigma n}(v_1, \dots, v_n)$$

which implies that the determinant function is given by

$$\alpha = \sum_{\sigma \in S_n} (\text{sgn } \sigma) \omega^{\sigma 1} \otimes \cdots \otimes \omega^{\sigma n} .$$

In other words, if A is a matrix with columns given by $v_1, \dots, v_n$ then $\det A = \alpha(v_1, \dots, v_n)$.

While we went through many detailed manipulations in arriving at these equations, we will assume from now on that the reader understands what was done in this example, and henceforth leave out some of the intermediate steps in such calculations. //

At the risk of boring some readers, let us very briefly review the meaning of the binomial coefficient $\binom{n}{r} = n!/[r!(n-r)!]$. The idea is that we want to know the number of ways of picking r distinct objects out of a collection of n distinct objects. In other words, how many combinations of n things taken r at a time are there? Well, to pick r objects, we have n choices for the first, then $n-1$ choices for the second, and so on down to $n-(r-1) = n-r+1$ choices for the r th. This gives us

$$n(n-1) \cdots (n-r+1) = n!/(n-r)!$$

as the number of ways of picking r objects out of n if we take into account the order in which the r objects are chosen. In other words, this is the number of injections $\text{INJ}(r, n)$ (see Section 4.6). For example, to pick three numbers in order out of the set $\{1, 2, 3, 4\}$, we might choose $(1, 3, 4)$, or we could choose $(3, 1, 4)$. It is this kind of situation that we must take into account. But for each distinct collection of r objects, there are $r!$ ways of arranging these, and hence we have over-counted each collection by a factor of $r!$. Dividing by $r!$ then yields the desired result.

If $\{e_1, \dots, e_n\}$ is a basis for V and $T \in \bigwedge^r(V)$, then T is determined by its values $T(e_{i_1}, \dots, e_{i_r})$ for $i_1 < \dots < i_r$. Indeed, following the same procedure as in Example 11.2, we see that if $v_i = e_{j_i} a_i^{j_i}$ for $i = 1, \dots, r$ then

$$T(v_1, \dots, v_r) = a^{i_1}_1 \cdots a^{i_r}_r T(e_{i_1}, \dots, e_{i_r})$$

where each sum is over $1 \leq i_k \leq n$. Furthermore, each collection $\{e_{i_1}, \dots, e_{i_r}\}$ must consist of distinct basis vectors in order that $T(e_{i_1}, \dots, e_{i_r}) \neq 0$. But the antisymmetry of T tells us that for any $\sigma \in S_r$, we must have

$$T(e_{\sigma i_1}, \dots, e_{\sigma i_r}) = (\text{sgn } \sigma) T(e_{i_1}, \dots, e_{i_r})$$

where we may choose $i_1 < \dots < i_r$. Thus, since the number of ways of choosing r distinct basis vectors $\{e_{i_1}, \dots, e_{i_r}\}$ out of the basis $\{e_1, \dots, e_n\}$ is $\binom{n}{r}$, it follows that

$$\dim \bigwedge^r(V) = \binom{n}{r} = n!/[r!(n-r)!] .$$

We will prove this result again when we construct a specific basis for $\bigwedge^r(V)$ (see Theorem 11.8 below).

In order to define linear transformations on $\mathcal{T}_r$ that preserve symmetry (or antisymmetry), we define the **symmetrizing mapping** $\mathcal{S}: \mathcal{T}_r \rightarrow \mathcal{T}_r$ and **alternation mapping** $\mathcal{A}: \mathcal{T}_r \rightarrow \mathcal{T}_r$ by

$$(\mathcal{S}T)(v_1, \dots, v_r) = (1/r!) \sum_{\sigma \in S_r} T(v_{\sigma 1}, \dots, v_{\sigma r})$$

and

$$(\mathcal{A}T)(v_1, \dots, v_r) = (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) T(v_{\sigma 1}, \dots, v_{\sigma r})$$

where $T \in \mathcal{T}_r(V)$ and $v_1, \dots, v_r \in V$. That these are in fact linear transformations on $\mathcal{T}_r$ follows from the observation that the mapping T_σ defined by

$$T_\sigma(v_1, \dots, v_r) = T(v_{\sigma 1}, \dots, v_{\sigma r})$$

is linear, and any linear combination of such mappings is again a linear trans-

formation.

Given any $\sigma \in S_r$, it will be convenient for us to define the mapping $\hat{\sigma}: V^r \rightarrow V^r$ by

$$\hat{\sigma}(v_1, \dots, v_r) = (v_{\sigma 1}, \dots, v_{\sigma r}) .$$

This mapping permutes the *order* of the vectors in its argument, not the labels (i.e. the indices), and hence its argument must always be $(v_1, v_2, \dots, v_r)$ or $(w_1, w_2, \dots, w_r)$ and so forth. Then for any $T \in \mathcal{T}_r(V)$ we define $\sigma T \in \mathcal{T}_r(V)$ by

$$\sigma T = T \circ \hat{\sigma}$$

which is the mapping T_σ defined above. It should be clear that $\sigma(T_1 + T_2) = \sigma T_1 + \sigma T_2$. Note also that if we write

$$\hat{\sigma}(v_1, \dots, v_r) = (v_{\sigma 1}, \dots, v_{\sigma r}) = (w_1, \dots, w_r)$$

then $w_i = v_{\sigma i}$ and therefore for any other $\tau \in S_r$ we have

$$\begin{aligned} \hat{\tau} \circ \hat{\sigma}(v_1, \dots, v_r) &= \hat{\tau}(w_1, \dots, w_r) \\ &= (w_{\tau 1}, \dots, w_{\tau r}) \\ &= (v_{\sigma \tau 1}, \dots, v_{\sigma \tau r}) \\ &= \widehat{\sigma \circ \tau}(v_1, \dots, v_r) . \end{aligned}$$

This shows that

$$\hat{\tau} \circ \hat{\sigma} = \widehat{\sigma \circ \tau}$$

and hence

$$\sigma(\tau T) = \sigma(T \circ \hat{\tau}) = T \circ (\hat{\tau} \circ \hat{\sigma}) = T \circ (\widehat{\sigma \circ \tau}) = (\sigma \circ \tau)T .$$

Note also that in this notation, the alternation mapping is defined as

$$\mathcal{A}T = (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma)(\sigma T) .$$

Theorem 11.3 The linear mappings $\mathcal{A}$ and $\mathcal{S}$ have the following properties:

- (a) $T \in \bigwedge^r(V)$ if and only if $\mathcal{A}T = T$, and $T \in \Sigma^r(V)$ if and only if $\mathcal{S}T = T$.
- (b) $\mathcal{A}(\mathcal{T}_r(V)) = \bigwedge^r(V)$ and $\mathcal{S}(\mathcal{T}_r(V)) = \Sigma^r(V)$.
- (c) $\mathcal{A}^2 = \mathcal{A}$ and $\mathcal{S}^2 = \mathcal{S}$, i.e., $\mathcal{A}$ and $\mathcal{S}$ are projections.

Proof Since the mapping $\mathcal{A}$ is more useful, we will prove the theorem only for this case, and leave the analogous results for $\mathcal{S}$ to the reader (see Exercise 11.2.1). Furthermore, all three statements of the theorem are interrelated, so

we prove them together.

First suppose that $T \in \bigwedge^r(V)$. From the definition of antisymmetric tensor we have $T(v_{\sigma 1}, \dots, v_{\sigma r}) = (\text{sgn } \sigma)T(v_1, \dots, v_r)$, and thus using the fact that the order of S_r is $r!$, we see that

$$\begin{aligned} \mathcal{A}T(v_1, \dots, v_r) &= (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) T(v_{\sigma 1}, \dots, v_{\sigma r}) \\ &= (1/r!) \sum_{\sigma \in S_r} T(v_1, \dots, v_r) \\ &= T(v_1, \dots, v_r) . \end{aligned}$$

This shows that $T \in \bigwedge^r(V)$ implies $\mathcal{A}T = T$.

Next, let T be any element of $\mathcal{T}_r(V)$. We may fix any particular element $\theta \in S_r$ and then apply $\mathcal{A}T$ to the vectors $v_{\theta 1}, \dots, v_{\theta r}$ to obtain

$$\begin{aligned} \mathcal{A}T(v_{\theta 1}, \dots, v_{\theta r}) &= \mathcal{A}T_{\theta}(v_1, \dots, v_r) \\ &= (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) T_{\theta}(v_{\sigma 1}, \dots, v_{\sigma r}) \\ &= (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) T(v_{\sigma \theta 1}, \dots, v_{\sigma \theta r}) . \end{aligned}$$

Now note that $\text{sgn } \sigma = (\text{sgn } \sigma)(\text{sgn } \theta)(\text{sgn } \theta) = (\text{sgn } \sigma\theta)(\text{sgn } \theta)$, and that $S_r = \{\phi = \sigma\theta : \sigma \in S_r\}$ (this is essentially what was done in Theorem 4.3). We now see that the right hand side of the above equation is just

$$\begin{aligned} (1/r!)(\text{sgn } \theta) \sum_{\sigma \in S_r} (\text{sgn } \sigma\theta) T(v_{\sigma \theta 1}, \dots, v_{\sigma \theta r}) \\ &= (1/r!)(\text{sgn } \theta) \sum_{\phi \in S_r} (\text{sgn } \phi) T(v_{\phi 1}, \dots, v_{\phi r}) \\ &= (\text{sgn } \theta) \mathcal{A}T(v_1, \dots, v_r) \end{aligned}$$

which shows (by definition) that $\mathcal{A}T$ is antisymmetric. In other words, this shows that $\mathcal{A}T \in \bigwedge^r(V)$ for any $T \in \mathcal{T}_r(V)$, or $\mathcal{A}(\mathcal{T}_r(V)) \subset \bigwedge^r(V)$.

Since the result of the earlier paragraph showed that $T = \mathcal{A}T \in \mathcal{A}(\mathcal{T}_r(V))$ for every $T \in \bigwedge^r(V)$, we see that $\bigwedge^r(V) \subset \mathcal{A}(\mathcal{T}_r(V))$, and therefore $\mathcal{A}(\mathcal{T}_r(V)) = \bigwedge^r(V)$. This also shows that if $\mathcal{A}T = T$, then T is necessarily an element of $\bigwedge^r(V)$. It then follows that for any $T \in \mathcal{T}_r(V)$ we have $\mathcal{A}T \in \bigwedge^r(V)$, and hence $\mathcal{A}^2T = \mathcal{A}(\mathcal{A}T) = \mathcal{A}T$ so that $\mathcal{A}^2 = \mathcal{A}$. ■

Suppose $A_{i_1} \dots i_r$ and $T^{i_1} \dots i_r$ (where $r \leq n = \dim V$ and $1 \leq i_k \leq n$) are both antisymmetric tensors, and consider their contraction $A_{i_1} \dots i_r T^{i_1} \dots i_r$. For any particular set of indices $i_1, \dots, i_r$ there will be $r!$ different ordered sets $(i_1, \dots, i_r)$. But by antisymmetry, the values of $A_{i_1} \dots i_r$ corresponding to each ordered set will differ only by a sign, and similarly for $T^{i_1} \dots i_r$. This

means that the product of $A_{i_1 \dots i_r}$ times $T^{i_1 \dots i_r}$ summed over the $r!$ ordered sets $(i_1, \dots, i_r)$ is the same as $r!$ times a single product which we choose to be the indices $i_1, \dots, i_r$ taken in increasing order. In other words, we have

$$A_{i_1 \dots i_r} T^{i_1 \dots i_r} = r! A_{|i_1 \dots i_r|} T^{i_1 \dots i_r}$$

where $|i_1 \dots i_r|$ denotes the fact that we are summing over increasing sets of indices only. For example, if we have antisymmetric tensors A_{ijk} and T^{ijk} in $\mathbb{R}^3$, then

$$A_{ijk} T^{ijk} = 3! A_{|ijk|} T^{ijk} = 6 A_{123} T^{123}$$

(where, in this case of course, A_{ijk} and T^{ijk} can only differ by a scalar).

There is a simple but extremely useful special type of antisymmetric tensor that we now wish to define. Before doing so however, it is first convenient to introduce another useful notational device. Note that if $T \in \mathcal{T}_r$ and we replace $v_1, \dots, v_r$ in the definitions of $\mathcal{S}$ and $\mathcal{A}$ by basis vectors e_i , then we obtain an expression in terms of components as

$$\mathcal{S}T_{1 \dots r} = (1/r!) \sum_{\sigma \in S_r} T_{\sigma 1 \dots \sigma r}$$

and

$$\mathcal{A}T_{1 \dots r} = (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) T_{\sigma 1 \dots \sigma r} .$$

We will write $T_{(1 \dots r)} = \mathcal{S}T_{1 \dots r}$ and $T_{[1 \dots r]} = \mathcal{A}T_{1 \dots r}$. For example, we have

$$T_{(ij)} = (1/2!)(T_{ij} + T_{ji})$$

and

$$T_{[ij]} = (1/2!)(T_{ij} - T_{ji}) .$$

A similar definition applies to any mixed tensor such as

$$\begin{aligned} T^{k(pq)}_{[ij]} &= (1/2!)\{T^{k(pq)}_{ij} - T^{k(pq)}_{ji}\} \\ &= (1/4)\{T^{kpq}_{ij} + T^{kqp}_{ij} - T^{kpq}_{ji} - T^{kqp}_{ji}\} . \end{aligned}$$

Note that if $T \in \Sigma^r(V)$, then $T_{(i_1 \dots i_r)} = T_{i_1 \dots i_r}$, while if $T \in \Lambda^r(V)$, then $T_{[i_1 \dots i_r]} = T_{i_1 \dots i_r}$.

Now consider the vector space $\mathbb{R}^3$ with the standard orthonormal basis $\{e_1, e_2, e_3\}$. We define the antisymmetric tensor $\varepsilon \in \Lambda^3(\mathbb{R}^3)$ by the requirement that

$$\varepsilon_{123} = \varepsilon(e_1, e_2, e_3) = +1 .$$

Since $\dim \bigwedge^3(\mathbb{R}^3) = 1$, this defines all the components of ε by antisymmetry: $\varepsilon_{213} = -\varepsilon_{231} = \varepsilon_{321} = -1$ etc. If $\{\bar{e}_i = e_j a^j_i\}$ is any other orthonormal basis for $\mathbb{R}^3$ related to the first basis by an (orthogonal) transition matrix $A = (a^j_i)$ with determinant equal to $+1$, then it is easy to see that

$$\varepsilon(\bar{e}_1, \bar{e}_2, \bar{e}_3) = \det A = +1$$

also. This is because $\varepsilon(e_i, e_j, e_k) = \text{sgn } \sigma$ where σ is the permutation that takes $(1, 2, 3)$ to (i, j, k) . Since $\varepsilon \in \bigwedge^3(\mathbb{R}^3)$, we see that $\varepsilon_{[ijk]} = \varepsilon_{ijk}$. The tensor ε is frequently called the **Levi-Civita tensor**. However, we stress that in a non-orthonormal coordinate system, it will not generally be true that $\varepsilon_{123} = +1$.

While we have defined the ε_{ijk} as the components of a tensor, it is just as common to see the **Levi-Civita** (or **permutation**) **symbol** ε_{ijk} defined simply as an antisymmetric *symbol* with $\varepsilon_{123} = +1$. In fact, from now on we shall use it in this way as a convenient notation for the sign of a permutation. For notational consistency, we also define the permutation symbol ε^{ijk} to have the same values as ε_{ijk} . A simple calculation shows that $\varepsilon_{ijk} \varepsilon^{ijk} = 3! = 6$.

It should be clear that this definition can easily be extended to an arbitrary number of dimensions. In other words, we define

$$\varepsilon_{i_1 \dots i_s} = \begin{cases} +1 & \text{if } (i_1, \dots, i_s) \text{ is an even permutation of } (1, 2, \dots, n) \\ -1 & \text{if } (i_1, \dots, i_s) \text{ is an odd permutation of } (1, 2, \dots, n) \\ 0 & \text{otherwise} \end{cases}.$$

This is just another way of writing $\text{sgn } \sigma$ where $\sigma \in S_n$. Therefore, using this symbol, we have the convenient notation for the determinant of a matrix $A = (a^i_j) \in M_n(\mathcal{F})$ as

$$\det A = \varepsilon_{i_1 \dots i_n} a^{i_1}_1 \dots a^{i_n}_n.$$

We now wish to prove a very useful result. To keep our notation from getting too cluttered, it will be convenient for us to write $\delta^{ijk}_{pqr} = \delta^i_p \delta^j_q \delta^k_r$. Now note that $\varepsilon_{pqr} = 6\delta^{[123]}_{pqr}$. To see that this is true, simply observe that both sides are antisymmetric in (p, q, r) , and are equal to 1 if $(p, q, r) = (1, 2, 3)$. (This also gives us a formula for ε_{pqr} as a 3×3 determinant with entries that are all Kronecker delta's. See Exercise 11.2.4) Using $\varepsilon^{123} = 1$ we may write this as $\varepsilon^{123} \varepsilon_{pqr} = 6\delta^{[123]}_{pqr}$. But now the antisymmetry in $(1, 2, 3)$ yields the general result

$$\varepsilon^{ijk} \varepsilon_{pqr} = 6\delta^{[ijk]}_{pqr} \quad (1)$$

which is what we wanted to show. It is also possible to prove this in another manner that serves to illustrate several useful techniques in tensor algebra.

Example 11.3 Suppose we have an arbitrary tensor $A \in \bigwedge^3(\mathbb{R}^3)$, and hence $A_{[ijk]} = A_{ijk}$. As noted above, the fact that $\dim \bigwedge^3(\mathbb{R}^3) = 1$ means that we must have $A_{ijk} = \lambda \varepsilon_{ijk}$ for some scalar $\lambda \in \mathbb{R}$. Then

$$A_{ijk} \varepsilon^{ijk} \varepsilon_{pqr} = \lambda \varepsilon_{ijk} \varepsilon^{ijk} \varepsilon_{pqr} = 6\lambda \varepsilon_{pqr} = 6\lambda \varepsilon_{ijk} \delta_p^i \delta_q^j \delta_r^k .$$

Because $\varepsilon_{ijk} = \varepsilon_{[ijk]}$, we can antisymmetrize over the indices i, j, k on the right hand side of the above equation to obtain $6\lambda \varepsilon_{ijk} \delta_{pqr}^{[ijk]}$ (write out the 6 terms if you do not believe it, and see Exercise 11.2.7). This gives us

$$A_{ijk} \varepsilon^{ijk} \varepsilon_{pqr} = 6\lambda \varepsilon_{ijk} \delta_{pqr}^{[ijk]} = 6A_{ijk} \delta_{pqr}^{[ijk]} .$$

Since the antisymmetric tensor A_{ijk} is contracted with another antisymmetric tensor on both sides of this equation, the discussion following the proof of Theorem 11.3 shows that we may write

$$A_{[ijk]} \varepsilon^{ijk} \varepsilon_{pqr} = 6A_{[ijk]} \delta_{pqr}^{[ijk]}$$

or

$$A_{123} \varepsilon^{123} \varepsilon_{pqr} = 6A_{123} \delta_{pqr}^{[123]} .$$

Cancelling A_{123} then yields $\varepsilon^{123} \varepsilon_{pqr} = 6\delta_{pqr}^{[123]}$ as we had above, and hence (1) again follows from this.

In the particular case that $p = k$, we leave it to the reader to show that

$$\varepsilon^{ijk} \varepsilon_{kqr} = 6\delta_{kqr}^{[ijk]} = \delta_{qr}^{ij} - \delta_{qr}^{ji} = \delta_q^i \delta_r^j - \delta_q^j \delta_r^i$$

which is very useful in manipulating vectors in $\mathbb{R}^3$. As an example, consider the vectors $\vec{A}, \vec{B}, \vec{C} \in \mathbb{R}^3$ equipped with a *Cartesian* coordinate system. Abusing our notation for simplicity (alternatively, we will see formally in Example 11.12 that $A^i = A_i$ for such an $\vec{A}$), we have

$$(\vec{B} \times \vec{C})^i = \varepsilon_{ijk} B^j C^k$$

and hence

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = A^i \varepsilon_{ijk} B^j C^k = +\varepsilon_{jki} B^j C^k A^i = \vec{B} \cdot (\vec{C} \times \vec{A}) .$$

Other examples are given in the exercises. //

Exercises

1. Prove Theorem 11.3 for the symmetrizing mapping $\mathcal{S}$.
2. Using the Levi-Civita symbol, prove the following vector identities in $\mathbb{R}^3$ equipped with a Cartesian coordinate system (where the vectors are actually vector fields where necessary, f is differentiable, and $\nabla^i = \partial_i = \partial/\partial x^i$):
 - (a) $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$
 - (b) $(\vec{A} \times \vec{B}) \cdot (\vec{C} \times \vec{D}) = (\vec{A} \cdot \vec{C})(\vec{B} \cdot \vec{D}) - (\vec{A} \cdot \vec{D})(\vec{B} \cdot \vec{C})$
 - (c) $\nabla \times \nabla f = 0$
 - (d) $\nabla \cdot (\nabla \times \vec{A}) = 0$
 - (e) $\nabla \times (\nabla \times \vec{A}) = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}$
 - (f) $\nabla \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot (\nabla \times \vec{A}) - \vec{A} \cdot (\nabla \times \vec{B})$
 - (g) $\nabla \times (\vec{A} \times \vec{B}) = \vec{A}(\nabla \cdot \vec{B}) - \vec{B}(\nabla \cdot \vec{A}) + (\vec{B} \cdot \nabla)\vec{A} - (\vec{A} \cdot \nabla)\vec{B}$
3. Using the divergence theorem ($\int_V \nabla \cdot \vec{A} \, d^3x = \int_S \vec{A} \cdot \vec{n} \, da$), prove that

$$\int_V \nabla \times \vec{A} \, d^3x = \int_S \vec{n} \times \vec{A} \, da \quad .$$

[Hint: Let $\vec{C}$ be a constant vector and show that

$$\vec{C} \cdot \int_V \nabla \times \vec{A} \, d^3x = \int_S (\vec{n} \times \vec{A}) \cdot \vec{C} \, da = \vec{C} \cdot \int_S \vec{n} \times \vec{A} \, da \quad .]$$

4. (a) Find the expression for ϵ_{pqr} as a 3×3 determinant with all Kronecker delta's as entries.
 (b) Write $\epsilon^{ijk}\epsilon_{pqr}$ as a 3×3 determinant with all Kronecker delta's as entries.
5. Suppose $V = \mathbb{R}^3$ and let A_{ij} be antisymmetric and S^{ij} be symmetric. Show that $A_{ij}S^{ij} = 0$ in two ways:
 - (a) Write out all terms in the sum and show that they cancel in pairs.
 - (b) Justify each of the following equalities:

$$A_{ij}S^{ij} = A_{ij}S^{ji} = -A_{ji}S^{ji} = -A_{ij}S^{ij} = 0 \quad .$$

6. Show that a second-rank tensor T_{ij} can be written in the form $T_{ij} = T_{(ij)} + T_{[ij]}$, but that a third-rank tensor can not. (The complete solution for ten-

sors of rank higher than two requires a detailed study of representations of the symmetric group and Young diagrams.)

7. Let $A_{i_1 \dots i_r}$ be antisymmetric, and suppose $T^{i_1 \dots i_r \dots}$ is an arbitrary tensor. Show that

$$A_{i_1 \dots i_r} T^{i_1 \dots i_r \dots} = A_{i_1 \dots i_r} T^{[i_1 \dots i_r] \dots} .$$

8. (a) Let $A = (a^i_j)$ be a 3×3 matrix. Show

$$\varepsilon_{ijk} a^i_p a^j_q a^k_r = (\det A) \varepsilon_{pqr} .$$

- (b) Let A be a linear transformation, and let $y_{(i)} = Ax_{(i)}$. Show

$$\det[y_{(1)}, \dots, y_{(n)}] = (\det A) \det[x_{(1)}, \dots, x_{(n)}] .$$

9. Show that under orthogonal transformations $A = (a^i_j)$ in $\mathbb{R}^3$, the vector cross product $\bar{x} = \bar{y} \times \bar{z}$ transforms as $\bar{x}^i = \varepsilon_{ijk} \bar{y}^j \bar{z}^k = (\det A) a^i_j x^j$. Discuss the difference between the cases $\det A = +1$ and $\det A = -1$.

11.3 THE EXTERIOR PRODUCT

We have seen that the tensor product of two elements of $\bigwedge^r(V)$ is not generally another element of $\bigwedge^{r+r}(V)$. However, using the mapping $\mathcal{A}$ we can define another product on $\bigwedge^r(V)$ that turns out to be of great use. We adopt the convention of denoting elements of $\bigwedge^r(V)$ by Greek letters such as α , β etc., which should not be confused with elements of the permutation group S_r .

If $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$, we define their **exterior product** (or **wedge product**) $\alpha \wedge \beta$ to be the mapping from $\bigwedge^r(V) \times \bigwedge^s(V) \rightarrow \bigwedge^{r+s}(V)$ given by

$$\alpha \wedge \beta = \frac{(r+s)!}{r!s!} \mathcal{A}(\alpha \otimes \beta) .$$

In other words, the wedge product is just an antisymmetrized tensor product. The reader may notice that the numerical factor is just the binomial coefficient $\binom{r+s}{r} = \binom{r+s}{s}$. It is also worth remarking that many authors leave off this coefficient entirely. While there is no fundamental reason for following either convention, our definition has the advantage of simplifying expressions involving volume elements as we shall see later.

A very useful formula for computing exterior products for small values of r and s is given in the next theorem. By way of terminology, a permutation $\sigma \in S_{r+s}$ such that $\sigma 1 < \cdots < \sigma r$ and $\sigma(r+1) < \cdots < \sigma(r+s)$ is called an (r, s) -**shuffle**. The proof of the following theorem should help to clarify this definition.

Theorem 11.4 Suppose $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$. Then for any collection of $r+s$ vectors $v_i \in V$ (with $r+s \leq \dim V$), we have

$$\alpha \wedge \beta(v_1, \dots, v_{r+s}) = \sum^* (\text{sgn } \sigma) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)})$$

where $\sum^*$ denotes the sum over all permutations $\sigma \in S_{r+s}$ such that $\sigma 1 < \cdots < \sigma r$ and $\sigma(r+1) < \cdots < \sigma(r+s)$ (i.e., over all (r, s) -shuffles).

Proof The proof is simply a careful examination of the terms in the definition of $\alpha \wedge \beta$. By definition, we have

$$\begin{aligned} \alpha \wedge \beta(v_1, \dots, v_{r+s}) &= [(r+s)!/r!s!] \mathcal{A}(\alpha \otimes \beta)(v_1, \dots, v_{r+s}) \\ &= [1/r!s!] \sum_{\sigma} (\text{sgn } \sigma) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}) \end{aligned} \quad (*)$$

where the sum is over all $\sigma \in S_{r+s}$. Now note that there are only $\binom{r+s}{r}$ *distinct* collections $\{\sigma 1, \dots, \sigma r\}$, and hence there are also only $\binom{r+s}{s} = \binom{r+s}{r}$ distinct collections $\{\sigma(r+1), \dots, \sigma(r+s)\}$. Let us call the set $\{v_{\sigma 1}, \dots, v_{\sigma r}\}$ the “ α -variables,” and the set $\{v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}\}$ the “ β -variables.” For any of the $\binom{r+s}{r}$ distinct collections of α - and β -variables, there will be $r!$ ways of ordering the α -variables within themselves, and $s!$ ways of ordering the β -variables within themselves. Therefore, there will be $r!s!$ possible arrangements of the α - and β -variables within themselves for each of the $\binom{r+s}{r}$ distinct collections. Let $\sigma \in S_{r+s}$ be a permutation that yields one of these distinct collections, and assume it is the one with the property that $\sigma 1 < \cdots < \sigma r$ and $\sigma(r+1) < \cdots < \sigma(r+s)$. The proof will be finished if we can show that all the rest of the $r!s!$ members of this collection are the same.

Let T denote the term in $(*)$ corresponding to our chosen σ . Then T is given by

$$T = (\text{sgn } \sigma) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}) .$$

This means that every other term t in the distinct collection containing T will be of the form

$$t = (\text{sgn } \theta) \alpha(v_{\theta 1}, \dots, v_{\theta r}) \beta(v_{\theta(r+1)}, \dots, v_{\theta(r+s)})$$

where the permutation $\theta \in S_{r+s}$ is such that the set $\{\theta 1, \dots, \theta r\}$ is the same as the set $\{\sigma 1, \dots, \sigma r\}$ (although possibly in a different order), and similarly, the set $\{\theta(r+1), \dots, \theta(r+s)\}$ is the same as the set $\{\sigma(r+1), \dots, \sigma(r+s)\}$. Thus the α - and β -variables are permuted within themselves. But we may then write $\theta = \phi\sigma$ where $\phi \in S_{r+s}$ is again such that the two sets $\{\sigma 1, \dots, \sigma r\}$ and $\{\sigma(r+1), \dots, \sigma(r+s)\}$ are permuted within themselves. Because none of the transpositions that define the permutation ϕ interchange α - and β -variables, we may use the antisymmetry of α and β separately to obtain

$$\begin{aligned} t &= (\text{sgn } \phi\sigma) \alpha(v_{\phi\sigma 1}, \dots, v_{\phi\sigma r}) \beta(v_{\phi\sigma(r+1)}, \dots, v_{\phi\sigma(r+s)}) \\ &= (\text{sgn } \phi\sigma) (\text{sgn } \phi) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}) \\ &= (\text{sgn } \sigma) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}) \\ &= T. \end{aligned}$$

(It was in bringing out only a single factor of $\text{sgn } \phi$ that we used the fact that there is no mixing of α - and β -variables.) In other words, the original sum over all $(r+s)!$ possible permutations $\sigma \in S_{r+s}$ has been reduced to a sum over $(r+s)!/r!s!$ distinct terms, each one of which is repeated $r!s!$ times. We are thus left with

$$\alpha \wedge \beta(v_1, \dots, v_{r+s}) = \sum^* (\text{sgn } \sigma) \alpha(v_{\sigma 1}, \dots, v_{\sigma r}) \beta(v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)})$$

where the sum is over the $(r+s)!/r!s!$ distinct collections $\{v_{\sigma 1}, \dots, v_{\sigma r}\}$ and $\{v_{\sigma(r+1)}, \dots, v_{\sigma(r+s)}\}$ subject to the requirements $\sigma 1 < \dots < \sigma r$ and $\sigma(r+1) < \dots < \sigma(r+s)$. ■

Let us introduce some convenient notation for handling multiple indices. Instead of writing the ordered set $(i_1, \dots, i_r)$, we simply write I where the exact range will be clear from the context. Furthermore, we write $\underline{I}$ to denote the increasing sequence $(i_1 < \dots < i_r)$. Similarly, we shall also write v_I instead of $(v_{i_1}, \dots, v_{i_r})$. To take full advantage of this notation, we first define the **generalized permutation symbol** ε by

$$\varepsilon_{i_1 \dots i_r}^{j_1 \dots j_r} = \begin{cases} +1 & \text{if } (j_1, \dots, j_r) \text{ is an even permutation of } (i_1, \dots, i_r) \\ -1 & \text{if } (j_1, \dots, j_r) \text{ is an odd permutation of } (i_1, \dots, i_r) \\ 0 & \text{otherwise} \end{cases}.$$

For example, $\varepsilon_{235}^{352} = +1$, $\varepsilon_{341}^{431} = -1$, $\varepsilon_{231}^{142} = 0$ etc. In particular, if $A = (a^j_i)$ is an $n \times n$ matrix, then

$$\det A = \varepsilon_{1 \dots n}^{i_1 \dots i_n} a^1_{i_1} \dots a^n_{i_n} = \varepsilon_{i_1 \dots i_n}^{1 \dots n} a^{i_1}_1 \dots a^{i_n}_n$$

because

$$\varepsilon_{1 \dots n}^{j_1 \dots j_n} = \varepsilon_{j_1 \dots j_n} = \varepsilon^{j_1 \dots j_n}.$$

Using this notation and Theorem 11.4, we may write the wedge product of α and β as

$$\alpha \wedge \beta(v_{i_1}, \dots, v_{i_{r+s}}) = \sum_{\underline{J}, \underline{K}} \varepsilon_{i_1 \dots i_r \dots i_{r+s}}^{j_1 \dots j_r k_1 \dots k_s} \alpha(v_{j_1}, \dots, v_{j_r}) \beta(v_{k_1}, \dots, v_{k_s})$$

or most simply in the following form, which we state as a corollary to Theorem 11.4 for easy reference.

Corollary Suppose $\alpha \in \Lambda^r(V)$ and $\beta \in \Lambda^s(V)$. Then for any collection of $r + s$ vectors $v_i \in V$ (with $r + s \leq \dim V$) we have

$$\alpha \wedge \beta(v_I) = \sum_{\underline{J}, \underline{K}} \varepsilon_I^{JK} \alpha(v_J) \beta(v_K).$$

Example 11.4 Suppose $\dim V = 5$ and $\{e_1, \dots, e_5\}$ is a basis for V . If $\alpha \in \Lambda^2(V)$ and $\beta \in \Lambda^1(V)$, then

$$\begin{aligned} \alpha \wedge \beta(e_5, e_2, e_3) &= \sum_{j_1 < j_2, k} \varepsilon_{523}^{j_1 j_2 k} \alpha(e_{j_1}, e_{j_2}) \beta(e_k) \\ &= \varepsilon_{523}^{235} \alpha(e_2, e_3) \beta(e_5) + \varepsilon_{523}^{253} \alpha(e_2, e_5) \beta(e_3) + \varepsilon_{523}^{352} \alpha(e_3, e_5) \beta(e_2) \\ &= \alpha(e_2, e_3) \beta(e_5) - \alpha(e_2, e_5) \beta(e_3) + \alpha(e_3, e_5) \beta(e_2). \quad // \end{aligned}$$

Our next theorem is a useful result in many computations. It is simply a contraction of indices in the permutation symbols.

Theorem 11.5 Let $I = (i_1, \dots, i_q)$, $J = (j_1, \dots, j_{r+s})$, $K = (k_1, \dots, k_r)$ and $L = (l_1, \dots, l_s)$. Then

$$\sum_{\underline{J}} \varepsilon_{1 \dots q+r+s}^{IJ} \varepsilon_J^{KL} = \varepsilon_{1 \dots q+r+s}^{IKL}$$

where I , K and L are fixed quantities, and J is summed over all increasing subsets $j_1 < \dots < j_{r+s}$ of $\{1, \dots, q + r + s\}$.

Proof The only nonvanishing terms on the left hand side can occur when J is a permutation of KL (or else $\varepsilon_J^{KL} = 0$), and of these possible permutations, we only have one in the sum, and that is for the increasing set $\underline{J}$. If J is an even permutation of KL , then $\varepsilon_J^{KL} = +1$, and $\varepsilon_{\underline{J}}^{IJ \dots q+r+s} = \varepsilon_{\underline{J}}^{IKL} \dots q+r+s$ since an even

number of permutations is required to go from J to KL . If J is an odd permutation of KL , then $\varepsilon_J^{KL} = -1$, and $\varepsilon_I^{J \dots q+r+s} = -\varepsilon_I^{KL \dots q+r+s}$ since an odd number of permutations is required to go from J to KL . The conclusion then follows immediately. ■

Note that we could have let $J = (j_1, \dots, j_r)$ and left out L entirely in Theorem 11.5. The reason we included L is shown in the next example.

Example 11.5 Let us use Theorem 11.5 to give a simple proof of the associativity of the wedge product. In other words, we want to show that

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$$

for any $\alpha \in \bigwedge^q(V)$, $\beta \in \bigwedge^r(V)$ and $\gamma \in \bigwedge^s(V)$. To see this, let $I = (i_1, \dots, i_q)$, $J = (j_1, \dots, j_{r+s})$, $K = (k_1, \dots, k_r)$ and $L = (l_1, \dots, l_s)$. Then we have

$$\begin{aligned} \alpha \wedge (\beta \wedge \gamma)(v_1, \dots, v_{q+r+s}) &= \sum_{I, J} \varepsilon_{1 \dots q+r+s}^{IJ} \alpha(v_I) (\beta \wedge \gamma)(v_J) \\ &= \sum_{I, J} \varepsilon_{1 \dots q+r+s}^{IJ} \alpha(v_I) \sum_{K, L} \varepsilon_J^{KL} \beta(v_K) \gamma(v_L) \\ &= \sum_{I, K, L} \varepsilon_{1 \dots q+r+s}^{IKL} \alpha(v_I) \beta(v_K) \gamma(v_L) . \end{aligned}$$

It is easy to see that had we started with $(\alpha \wedge \beta) \wedge \gamma$, we would have arrived at the same sum.

As was the case with the tensor product, we simply write $\alpha \wedge \beta \wedge \gamma$ from now on. Note also that a similar calculation can be done for the wedge product of any number of terms. //

We now wish to prove the basic algebraic properties of the wedge product. This will be facilitated by a preliminary result on the alternation mapping.

Theorem 11.6 If $S \in \mathcal{T}_r$ and $T \in \mathcal{T}_s$, then

$$\mathcal{A}((\mathcal{A}S) \otimes T) = \mathcal{A}(S \otimes T) = \mathcal{A}(S \otimes (\mathcal{A}T)) .$$

Proof Using the bilinearity of the tensor product and the definition of $\mathcal{A}S$ we may write

$$(\mathcal{A}S) \otimes T = (1/r!) \sum_{\sigma \in S_r} (\text{sgn } \sigma) [(\sigma S) \otimes T] .$$

For each $\sigma \in S_r$, let $G \subset S_{r+s}$ be the set of permutations ϕ defined by

$$(\phi 1, \dots, \phi(r+s)) = (\sigma 1, \dots, \sigma r, r+1, \dots, r+s) .$$

In other words, G consists of all permutations $\phi \in S_{r+s}$ that have the same effect on $1, \dots, r$ as $\sigma \in S_r$, but leave the remaining terms $r+1, \dots, r+s$ unchanged. This means that $\text{sgn } \phi = \text{sgn } \sigma$, and $\phi(S \otimes T) = (\sigma S) \otimes T$ (see Exercise 11.3.1). We then have

$$(\mathcal{A}S) \otimes T = (1/r!) \sum_{\phi \in G} (\text{sgn } \phi) \phi(S \otimes T)$$

and therefore

$$\begin{aligned} \mathcal{A}((\mathcal{A}S) \otimes T) &= [1/(r+s)!] \sum_{\tau \in S_{r+s}} (\text{sgn } \tau) \tau((1/r!) \sum_{\phi \in G} (\text{sgn } \phi) \phi(S \otimes T)) \\ &= [1/(r+s)!] (1/r!) \sum_{\phi \in G} \sum_{\tau \in S_{r+s}} (\text{sgn } \tau) (\text{sgn } \phi) \tau \phi(S \otimes T) . \end{aligned}$$

But for each $\phi \in G$, we note that $S_{r+s} = \{\theta = \tau\phi : \tau \in S_{r+s}\}$, and hence

$$\begin{aligned} [1/(r+s)!] \sum_{\tau \in S_{r+s}} (\text{sgn } \tau) (\text{sgn } \phi) \tau \phi(S \otimes T) \\ &= [1/(r+s)!] \sum_{\tau \in S_{r+s}} (\text{sgn } \tau \phi) \tau \phi(S \otimes T) \\ &= [1/(r+s)!] \sum_{\theta \in S_{r+s}} (\text{sgn } \theta) \theta(S \otimes T) \\ &= \mathcal{A}(S \otimes T) . \end{aligned}$$

Since this is independent of the particular $\phi \in G$ and there are $r!$ elements in G , we then have

$$\begin{aligned} \mathcal{A}((\mathcal{A}S) \otimes T) &= (1/r!) \sum_{\phi \in G} \mathcal{A}(S \otimes T) \\ &= \mathcal{A}(S \otimes T) (1/r!) \sum_{\phi \in G} 1 \\ &= \mathcal{A}(S \otimes T) . \end{aligned}$$

The proof that $\mathcal{A}(S \otimes T) = \mathcal{A}(S \otimes (\mathcal{A}T))$ is similar, and we leave it to the reader (see Exercise 11.3.2). ■

Note that in defining the wedge product $\alpha \wedge \beta$, there is really nothing that requires us to have $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$. We could just as well be more general and let $\alpha \in \mathcal{T}_r(V)$ and $\beta \in \mathcal{T}_s(V)$. However, if this is the case, then the formula given in Theorem 11.4 most certainly is *not* valid. However, we do have the following corollary to Theorem 11.6.

Corollary For any $S \in \mathcal{T}_r$ and $T \in \mathcal{T}_s$ we have $\mathcal{A}S \wedge T = S \wedge T = S \wedge \mathcal{A}T$.

Proof This follows directly from Theorem 11.6 and the wedge product definition $S \wedge T = [(r+s)!/r!s!] \mathcal{A}(S \otimes T)$. ■

We are now in a position to prove some of the most important properties of the wedge product. Note that this next theorem is stated in terms of the

more general definition of the wedge product.

Theorem 11.7 Suppose $\alpha, \alpha_1, \alpha_2 \in \mathcal{T}_q(V)$, $\beta, \beta_1, \beta_2 \in \mathcal{T}_r(V)$, $\gamma \in \mathcal{T}_s(V)$ and $a \in \mathcal{F}$. Then

(a) The wedge product is bilinear. That is,

$$\begin{aligned}(\alpha_1 + \alpha_2) \wedge \beta &= \alpha_1 \wedge \beta + \alpha_2 \wedge \beta \\ \alpha \wedge (\beta_1 + \beta_2) &= \alpha \wedge \beta_1 + \alpha \wedge \beta_2 \\ (a\alpha) \wedge \beta &= \alpha \wedge (a\beta) = a(\alpha \wedge \beta)\end{aligned}$$

(b) $\alpha \wedge \beta = (-1)^{qr} \beta \wedge \alpha$.

(c) The wedge product is associative. That is,

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma = [(q+r+s)!/q!r!s!]\mathcal{A}(\alpha \otimes \beta \otimes \gamma) .$$

Proof (a) This follows from the definition of wedge product, the fact that $\otimes$ is bilinear and $\mathcal{A}$ is linear. This result may also be shown directly in the case that $\alpha, \alpha_1, \alpha_2 \in \bigwedge^q(V)$ and $\beta, \beta_1, \beta_2 \in \bigwedge^r(V)$ by using the corollary to Theorem 11.4 (see Exercise 11.3.3).

(b) This can also be shown directly from the corollary to Theorem 11.4 (see Exercise 11.3.4). Alternatively, we may proceed as follows. First note that for $\sigma \in S_r$, we see that (since for any other $\tau \in S_r$ we have $\tau(\sigma\alpha) = (\tau \circ \sigma)\alpha$, and hence $\tau(\sigma\alpha)(v_1, \dots, v_r) = \alpha(v_{\tau\sigma 1}, \dots, v_{\tau\sigma r})$)

$$\begin{aligned}\mathcal{A}(\sigma\alpha)(v_1, \dots, v_r) &= (1/r!) \sum_{\tau \in S_r} (\text{sgn } \tau) \tau(\sigma\alpha)(v_1, \dots, v_r) \\ &= (1/r!) \sum_{\tau \in S_r} (\text{sgn } \tau) \alpha(v_{\tau\sigma 1}, \dots, v_{\tau\sigma r}) \\ &= (1/r!) \sum_{\tau \in S_r} (\text{sgn } \tau\sigma) (\text{sgn } \sigma) \alpha(v_{\tau\sigma 1}, \dots, v_{\tau\sigma r}) \\ &= (\text{sgn } \sigma) (1/r!) \sum_{\theta \in S_r} (\text{sgn } \theta) \alpha(v_{\theta 1}, \dots, v_{\theta r}) \\ &= (\text{sgn } \sigma) \mathcal{A}\sigma(v_1, \dots, v_r) .\end{aligned}$$

Hence $\mathcal{A}(\sigma\alpha) = (\text{sgn } \sigma) \mathcal{A}\alpha$.

Now define $\sigma_0 \in S_{q+r}$ by

$$\sigma_0(1, \dots, q+r) = (q+1, \dots, q+r, 1, \dots, q) .$$

Since σ_0 is just a product of qr transpositions, it follows that $\text{sgn } \sigma_0 = (-1)^{qr}$. We then see that

$$\begin{aligned}\alpha \otimes \beta(v_1, \dots, v_{q+r}) &= (\beta \otimes \alpha)(v_{\sigma_0 1}, \dots, v_{\sigma_0(q+r)}) \\ &= \sigma_0(\beta \otimes \alpha)(v_1, \dots, v_{q+r}) .\end{aligned}$$

Therefore (ignoring the factorial multiplier which will cancel out from both sides of this equation),

$$\begin{aligned}\alpha \wedge \beta &= \mathcal{A}(\alpha \otimes \beta) = \mathcal{A}(\sigma_0(\beta \otimes \alpha)) = (\text{sgn } \sigma_0) \mathcal{A}(\beta \otimes \alpha) \\ &= (-1)^{qr} \beta \wedge \alpha .\end{aligned}$$

(c) Using Theorem 11.6, we simply calculate

$$\begin{aligned}(\alpha \wedge \beta) \wedge \gamma &= [(q+r+s)!/(q+r)!s!] \mathcal{A}((\alpha \wedge \beta) \otimes \gamma) \\ &= [(q+r+s)!/(q+r)!s!][(q+r)!/q!r!] \mathcal{A}(\mathcal{A}(\alpha \otimes \beta) \otimes \gamma) \\ &= [(q+r+s)!/(q!r!s!)] \mathcal{A}(\alpha \otimes \beta \otimes \gamma) .\end{aligned}$$

Similarly, we find that $\alpha \wedge (\beta \wedge \gamma)$ yields the same result. We are therefore justified (as we also saw in Example 11.5) in writing simply $\alpha \wedge \beta \wedge \gamma$. Furthermore, it is clear that this result can be extended to any finite number of products. ■

Example 11.6 Suppose $\alpha \in \mathcal{T}_r$ and $\beta \in \mathcal{T}_s$. Since $\alpha \wedge \beta = (-1)^{rs} \beta \wedge \alpha$, we see that if either r or s is even, then $\alpha \wedge \beta = \beta \wedge \alpha$, but if both r and s are odd, then $\alpha \wedge \beta = -\beta \wedge \alpha$. Therefore if r is odd we have $\alpha \wedge \alpha = 0$, but if r is even, then $\alpha \wedge \alpha$ is not necessarily zero. In particular, any 1-form α always has the property that $\alpha \wedge \alpha = 0$. //

Example 11.7 If $\alpha_1, \dots, \alpha_5$ are 1-forms on $\mathbb{R}^5$, let us define

$$\beta = \alpha_1 \wedge \alpha_3 + \alpha_3 \wedge \alpha_5$$

and

$$\gamma = 2\alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_2 \wedge \alpha_4 .$$

Using the properties of the wedge product given in Theorem 11.7 we then have

$$\begin{aligned}\beta \wedge \gamma &= (\alpha_1 \wedge \alpha_3 + \alpha_3 \wedge \alpha_5) \wedge (2\alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_2 \wedge \alpha_4) \\ &= 2\alpha_1 \wedge \alpha_3 \wedge \alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_3 \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_4 \\ &\quad + 2\alpha_3 \wedge \alpha_5 \wedge \alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_3 \wedge \alpha_5 \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_4 \\ &= -2\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5 - 0 + 0 - \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5 \\ &= -3\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5 .\end{aligned}$$

Example 11.8 Suppose $\alpha_1, \dots, \alpha_r \in \bigwedge^1(V)$ and $v_1, \dots, v_r \in V$. Using Theorem 11.5, it is easy to generalize the corollary to Theorem 11.4 to obtain (see Exercise 11.3.5)

$$\begin{aligned}\alpha_1 \wedge \cdots \wedge \alpha_r(v_1, \dots, v_r) &= \sum_{i_1, \dots, i_r} \varepsilon_{i_1 \dots i_r}^{i_1 \dots i_r} \alpha_1(v_{i_1}) \cdots \alpha_r(v_{i_r}) \\ &= \det(\alpha_i(v_j)) .\end{aligned}$$

(Note that the sum is not over any increasing indices because each α_i is only a 1-form.) As a special case, suppose $\{e_i\}$ is a basis for V and $\{\omega^j\}$ is the corresponding dual basis. Then $\omega^j(e_i) = \delta_i^j$ and hence

$$\begin{aligned}\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}(e_{j_1}, \dots, e_{j_r}) &= \sum_{k_1, \dots, k_r} \varepsilon_{j_1 \dots j_r}^{k_1 \dots k_r} \omega^{i_1}(e_{k_1}) \cdots \omega^{i_r}(e_{k_r}) \\ &= \varepsilon_{j_1 \dots j_r}^{i_1 \dots i_r} .\end{aligned}$$

In particular, if $\dim V = n$, choosing the indices $(i_1, \dots, i_n) = (1, \dots, n) = (j_1, \dots, j_n)$, we see that

$$\omega^1 \wedge \cdots \wedge \omega^n(e_1, \dots, e_n) = 1 . \quad //$$

Exercises

1. Show that $\phi(S \otimes T) = (\sigma S) \otimes T$ in the proof of Theorem 11.6.
2. Finish the proof of Theorem 11.6 by showing that

$$\mathcal{A}(S \otimes T) = \mathcal{A}(S \otimes (\mathcal{A}T)).$$

3. Using $\alpha_i \in \bigwedge^q(V)$ and $\beta_i \in \bigwedge^r(V)$, prove Theorem 11.7(a) directly from the corollary to Theorem 11.4.
4. Use the corollary to Theorem 11.4 to prove Theorem 11.7(b).
5. Suppose $\alpha_1, \dots, \alpha_r \in \bigwedge^1(V)$ and $v_1, \dots, v_r \in V$. Show that

$$\alpha_1 \wedge \cdots \wedge \alpha_r(v_1, \dots, v_r) = \det(\alpha_i(v_j)) .$$

6. Suppose $\{e_1, \dots, e_n\}$ is a basis for V and $\{\omega^1, \dots, \omega^n\}$ is the corresponding dual basis. If $\alpha \in \bigwedge^r(V)$ (where $r \leq n$), show that

$$\alpha = \sum_I \alpha(e_I) \omega^I = \sum_{i_1 < \cdots < i_r} \alpha(e_{i_1}, \dots, e_{i_r}) \omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$$

by applying both sides to $(e_{j_1}, \dots, e_{j_r})$. (See also Theorem 11.8.)

7. (**Interior Product**) Suppose $\alpha \in \bigwedge^r(V)$ and $v, v_2, \dots, v_r \in V$. We define the $(r-1)$ -form $i_v \alpha$ by

$$\begin{aligned} i_v \alpha &= 0 & \text{if } r = 0. \\ i_v \alpha &= \alpha(v) & \text{if } r = 1. \\ i_v \alpha(v_2, \dots, v_r) &= \alpha(v, v_2, \dots, v_r) & \text{if } r > 1. \end{aligned}$$

- (a) Prove that $i_{u+v} = i_u + i_v$.
 (b) If $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$, prove that $i_v: \bigwedge^{r+s}(V) \rightarrow \bigwedge^{r+s-1}(V)$ is an **anti-derivation**, i.e.,

$$i_v(\alpha \wedge \beta) = (i_v \alpha) \wedge \beta + (-1)^r \alpha \wedge (i_v \beta) .$$

- (c) If $v = v^i e_i$ and $\alpha = \sum_{i_1} a_{i_1} \dots i_r \omega^{i_1} \wedge \dots \wedge \omega^{i_r}$ where $\{\omega^i\}$ is the basis dual to $\{e_i\}$, show that

$$i_v \alpha = \sum_{i_2 < \dots < i_r} b_{i_2 \dots i_r} \omega^{i_2} \wedge \dots \wedge \omega^{i_r}$$

where

$$b_{i_2 \dots i_r} = \sum_j v^j a_{j i_2 \dots i_r} .$$

- (d) If $\alpha = f^1 \wedge \dots \wedge f^r$, show that

$$\begin{aligned} i_v \alpha &= \sum_{k=1}^r (-1)^{k-1} f^k(v) f^1 \wedge \dots \wedge f^{k-1} \wedge f^{k+1} \wedge \dots \wedge f^r \\ &= \sum_{k=1}^r (-1)^{k-1} f^k(v) f^1 \wedge \dots \wedge \widehat{f^k} \wedge \dots \wedge f^r \end{aligned}$$

where the $\widehat{}$ means that the term f^k is to be deleted from the expression.

8. Let $V = \mathbb{R}^n$ have the standard basis $\{e_i\}$, and let the corresponding dual basis for V^* be $\{\omega^i\}$.

- (a) If $u, v \in V$, show that

$$\omega^i \wedge \omega^j(u, v) = \begin{vmatrix} u^i & v^i \\ u^j & v^j \end{vmatrix}$$

and that this is $\pm$ the area of the parallelogram spanned by the projection of u and v onto the $x^i x^j$ -plane. What do you think is the significance of the different signs?

(b) Generalize this to $\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$ where $r \leq n$.

9. Suppose $V = \mathcal{F}^n$, and let $v_1, \dots, v_n \in V$ have components relative to the standard basis $\{e_i\}$ defined by $v_i = e_j v_i^j$. For any $1 \leq r < n$, let $s = n - r$ and define the r -form α by

$$\alpha(v_1, \dots, v_r) = \begin{vmatrix} v_1^1 & \cdots & v_r^1 \\ \vdots & & \vdots \\ v_1^r & \cdots & v_r^r \end{vmatrix}$$

and the s -form β by

$$\beta(v_1, \dots, v_s) = \begin{vmatrix} v_{r+1}^1 & \cdots & v_n^1 \\ \vdots & & \vdots \\ v_{r+1}^s & \cdots & v_n^s \end{vmatrix}.$$

- (a) Use Theorem 4.9 to show that $\alpha \wedge \beta$ is the determinant function D on $\mathcal{F}^n$.
 (b) Show that the sign of an (r, s) -shuffle is given by

$$\varepsilon_{1 \dots r \ r+1 \dots r+s}^{i_1 \dots i_r \ j_1 \dots j_s} = (-1)^{i_1 + \cdots + i_r + r(r+1)/2}$$

where $i_1, \dots, i_r$ and $j_1, \dots, j_s$ are listed in increasing order.

- (c) If $A = (a^i_j) \in M_n(\mathcal{F})$, prove the **Laplace expansion** formula

$$\det A = \sum_I (-1)^{i_1 + \cdots + i_r + r(r+1)/2} \begin{vmatrix} a^{i_1}_1 & \cdots & a^{i_1}_r \\ \vdots & & \vdots \\ a^{i_r}_1 & \cdots & a^{i_r}_r \end{vmatrix} \begin{vmatrix} a^{j_1}_{r+1} & \cdots & a^{j_1}_n \\ \vdots & & \vdots \\ a^{j_r}_{r+1} & \cdots & a^{j_r}_n \end{vmatrix}$$

where $I = \{i_1, \dots, i_r\}$ and $J = \{j_1, \dots, j_s\}$ are “complementary” sets of indices, i.e., $I \cap J = \emptyset$ and $I \cup J = \{1, 2, \dots, n\}$.

10. Let $\mathcal{B} = r! \mathcal{A}$ where $\mathcal{A}: \mathcal{T}_r \rightarrow \mathcal{T}_r$ is the alternation mapping. Define $\alpha \wedge \beta$ in terms of $\mathcal{B}$. What is $\mathcal{B}(f^1 \otimes \cdots \otimes f^r)$ where $f^i \in V^*$?
 11. Let $I = (i_1, \dots, i_q)$, $J = (j_1, \dots, j_p)$, and $K = (k_1, \dots, k_q)$. Prove the following generalization of Example 11.3:

$$\sum_J \varepsilon_{1 \dots p+q}^{IJ} \varepsilon_{JK}^{1 \dots p+q} = \varepsilon_K^I = n! \delta_{k_1}^{i_1} \cdots \delta_{k_q}^{i_q}$$

11.4 TENSOR ALGEBRAS

We define the direct sum of all tensor spaces $\mathcal{T}_r(V)$ to be the (infinite-dimensional) space $\mathcal{T}_0(V) \oplus \mathcal{T}_1(V) \oplus \cdots \oplus \mathcal{T}_r(V) \oplus \cdots$, and $\mathcal{T}(V)$ to be all elements of this space with finitely many nonzero components. This means that every element $T \in \mathcal{T}(V)$ has a unique expression of the form (ignoring zero summands)

$$T = T^{(1)}_{i_1} + \cdots + T^{(r)}_{i_r}$$

where each $T^{(k)}_{i_k} \in \mathcal{T}_{i_k}(V)$ and $i_1 < \cdots < i_r$. The tensors $T^{(k)}_{i_k}$ are called the **graded components** of T . In the special case that $T \in \mathcal{T}_r(V)$ for some r , then T is said to be of **order** r . We define addition in $\mathcal{T}(V)$ componentwise, and we also define multiplication in $\mathcal{T}(V)$ by defining $\otimes$ to be distributive on all of $\mathcal{T}(V)$. We have therefore made $\mathcal{T}(V)$ into an associative algebra over $\mathcal{F}$ which is called the **tensor algebra**.

We have seen that $\bigwedge^r(V)$ is a subspace of $\mathcal{T}_r(V)$ since $\bigwedge^r(V)$ is just the image of $\mathcal{T}_r(V)$ under $\mathcal{A}$. Recall also that $\bigwedge^0(V) = \mathcal{T}_0(V)$ is defined to be the scalar field $\mathcal{F}$. As might therefore be expected, we define $\bigwedge(V)$ to be the direct sum

$$\bigwedge(V) = \bigwedge^0(V) \oplus \bigwedge^1(V) \oplus \cdots \oplus \bigwedge^r(V) \oplus \cdots \subset \mathcal{T}(V) .$$

Note that $\bigwedge^r(V) = 0$ if $r > \dim V$.

It is important to realize that if $\alpha \in \bigwedge^r(V) \subset \mathcal{T}_r(V)$ and $\beta \in \bigwedge^s(V) \subset \mathcal{T}_s(V)$, then even though $\alpha \otimes \beta \in \mathcal{T}_{r+s}(V)$, it is not generally true that $\alpha \otimes \beta \in \bigwedge^{r+s}(V)$. Therefore $\bigwedge(V)$ is not a subalgebra of $\mathcal{T}(V)$. However, the wedge product is a mapping from $\bigwedge^r(V) \times \bigwedge^s(V) \rightarrow \bigwedge^{r+s}(V)$, and hence if we extend this product in the obvious manner to a bilinear mapping $\bigwedge(V) \times \bigwedge(V) \rightarrow \bigwedge(V)$, then $\bigwedge(V)$ becomes an algebra over $\mathcal{F} = \bigwedge^0(V)$. In other words, if $\alpha = \alpha_1 + \cdots + \alpha_r$ with each $\alpha_i \in \bigwedge^{r_i}(V)$, and $\beta = \beta_1 + \cdots + \beta_s$ with each $\beta_i \in \bigwedge^{s_i}(V)$, then we define

$$\alpha \wedge \beta = \sum_{i=1}^r \sum_{j=1}^s \alpha_i \wedge \beta_j .$$

This algebra is called the **Grassmann** (or **exterior**) **algebra**.

The astute reader may be wondering exactly how we add together the elements $\alpha_1 \in \bigwedge^{r_1}(V)$ and $\alpha_2 \in \bigwedge^{r_2}(V)$ (with $r_1 \neq r_2$) when none of the operations $(\alpha_1 + \alpha_2)(v_1, \dots, v_{r_1})$, $(\alpha_1 + \alpha_2)(v_1, \dots, v_{r_2})$ nor $(\alpha_1 + \alpha_2)(v_1, \dots, v_{r_1+r_2})$ makes any sense. The answer is that for purposes of the Grassmann algebra, we consider both α_1 and α_2 to be elements of $\bigwedge(V)$. For example, if

α_1 is a 1-form and α_2 is a 2-form, then we write $\alpha_1 = 0 + \alpha_1 + 0 + 0 + \cdots$ and $\alpha_2 = 0 + 0 + \alpha_2 + 0 + \cdots$, and hence $a_1\alpha_1 + a_2\alpha_2$ (where $a_i \in \mathcal{F}$) makes sense in $\bigwedge(V)$. In this way, every element of $\bigwedge(V)$ has a degree (recall that an r -form is said to be of degree r), and we say that $\bigwedge(V)$ is a **graded associative algebra**.

Unlike the infinite-dimensional algebra $\mathcal{T}(V)$, the algebra $\bigwedge(V)$ is finite-dimensional. This should be clear from the discussion following Example 11.2 where we showed that $\dim \bigwedge^r(V) = \binom{n}{r}$ (where $n = \dim V$), and hence that $\dim \bigwedge^r(V) = 0$ if $r > n$. Let us now prove this result again by constructing a specific basis for $\bigwedge^r(V)$.

Theorem 11.8 Suppose $\dim V = n$. Then for $r > n$ we have $\bigwedge^r(V) = \{0\}$, and if $0 \leq r \leq n$, then $\dim \bigwedge^r(V) = \binom{n}{r}$. Therefore $\dim \bigwedge(V) = 2^n$. Moreover, if $\{\omega^1, \dots, \omega^n\}$ is a basis for $V^* = \bigwedge^1(V)$, then a basis for $\bigwedge^r(V)$ is given by the set

$$\{\omega^{i_1} \wedge \cdots \wedge \omega^{i_r} : 1 \leq i_1 < \cdots < i_r \leq n\}.$$

Proof Suppose $\alpha \in \bigwedge^r(V)$ where $r > \dim V = n$. By multilinearity, α is determined by its values on a basis $\{e_1, \dots, e_n\}$ for V (see Example 11.2). But then we must have $\alpha(e_{i_1}, \dots, e_{i_r}) = 0$ because at least two of the e_{i_k} are necessarily the same and α is antisymmetric. This means that $\alpha(v_1, \dots, v_r) = 0$ for all $v_i \in V$, and hence $\alpha = 0$. Thus $\bigwedge^r(V) = \{0\}$ if $r > n$.

Now suppose that $\{\omega^1, \dots, \omega^n\}$ is the basis for V^* dual to $\{e_i\}$. From Theorem 11.2, we know that $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r} : 1 \leq i_1, \dots, i_r \leq n\}$ forms a basis for $\mathcal{T}_r(V)$, and since the alternation mapping $\mathcal{A}$ maps $\mathcal{T}_r(V)$ onto $\bigwedge^r(V)$ (Theorem 11.3(b)), it follows that the image of the basis $\{\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}\}$ for $\mathcal{T}_r(V)$ must span $\bigwedge^r(V)$. If $\alpha \in \bigwedge^r(V)$, then $\alpha \in \mathcal{T}_r(V)$, and hence

$$\alpha = \sum \alpha_{i_1 \dots i_r} \omega^{i_1} \otimes \cdots \otimes \omega^{i_r}$$

where the sum is over all $1 \leq i_1, \dots, i_r \leq n$ and $\alpha_{i_1 \dots i_r} = \alpha(e_{i_1}, \dots, e_{i_r})$. Using Theorems 11.3(a) and 11.7(c) we have

$$\begin{aligned} \alpha &= \mathcal{A}\alpha = \sum \alpha_{i_1 \dots i_r} \mathcal{A}(\omega^{i_1} \otimes \cdots \otimes \omega^{i_r}) \\ &= \sum \alpha_{i_1 \dots i_r} (1/r!) \omega^{i_1} \wedge \cdots \wedge \omega^{i_r} \end{aligned}$$

where the sum is still over all $1 \leq i_1, \dots, i_r \leq n$. However, by the antisymmetry of the wedge product, the collection $\{i_1, \dots, i_r\}$ must all be different, and hence the sum can only be over the $\binom{n}{r}$ distinct such combinations. For each

such combination there will be $r!$ permutations $\sigma \in S_r$ of the basis vectors. If we write each of these permutations in increasing order $i_1 < \dots < i_r$, then the wedge product changes by a factor $\text{sgn } \sigma$, as does $\alpha_{i_1} \dots \alpha_{i_r} = \alpha(e_{i_1}, \dots, e_{i_r})$. Therefore the signs cancel and we are left with

$$\alpha = \alpha_{|i_1 \dots i_r|} \omega^{i_1} \wedge \dots \wedge \omega^{i_r}$$

where, as mentioned previously, we use the notation $\alpha_{|i_1 \dots i_r|}$ to mean that the sum is over increasing sets $i_1 < \dots < i_r$. Thus we have shown that the $\binom{n}{r}$ elements $\omega^{i_1} \wedge \dots \wedge \omega^{i_r}$ with $1 \leq i_1 < \dots < i_r \leq n$ span $\bigwedge^r(V)$. We must still show that they are linearly independent.

Suppose $\alpha_{|i_1 \dots i_r|} \omega^{i_1} \wedge \dots \wedge \omega^{i_r} = 0$. Then for any set $\{e_{j_1}, \dots, e_{j_r}\}$ with $1 \leq j_1 < \dots < j_r \leq n$ we have (using Example 11.8)

$$\begin{aligned} 0 &= \alpha_{|i_1 \dots i_r|} \omega^{i_1} \wedge \dots \wedge \omega^{i_r} (e_{j_1}, \dots, e_{j_r}) \\ &= \alpha_{|i_1 \dots i_r|} \varepsilon_{j_1 \dots j_r}^{i_1 \dots i_r} \\ &= \alpha_{j_1 \dots j_r} \end{aligned}$$

since the only nonvanishing term occurs when $\{i_1, \dots, i_r\}$ is a permutation of $\{j_1, \dots, j_r\}$ and both are increasing sets. This proves linear independence.

Finally, using the binomial theorem, we now see that

$$\dim \bigwedge(V) = \sum_{r=0}^n \dim \bigwedge^r(V) = \sum_{r=0}^n \binom{n}{r} = (1+1)^n = 2^n \quad \blacksquare$$

Example 11.9 Another useful result is the following. Suppose $\dim V = n$, and let $\{\omega^1, \dots, \omega^n\}$ be a basis for V^* . If $\alpha^1, \dots, \alpha^n$ are any other 1-forms in $\bigwedge^1(V) = V^*$, then we may expand each α^i in terms of the ω^j as $\alpha^i = a^i_j \omega^j$. We then have

$$\begin{aligned} \alpha^1 \wedge \dots \wedge \alpha^n &= a^1_{i_1} \dots a^n_{i_n} \omega^{i_1} \wedge \dots \wedge \omega^{i_n} \\ &= a^1_{i_1} \dots a^n_{i_n} \varepsilon_{i_1 \dots i_n}^{1 \dots n} \omega^1 \wedge \dots \wedge \omega^n \\ &= \det(a^i_j) \omega^1 \wedge \dots \wedge \omega^n \end{aligned}$$

Recalling Example 11.1, if $\{\omega^i = dx^i\}$ is a local basis for a cotangent space V^* and $\{\alpha^i = dy^i\}$ is any other local basis, then $dy^i = (\partial y^i / \partial x^j) dx^j$ and

$$\det(a^i_j) = \frac{\partial(y^1 \cdots y^n)}{\partial(x^1 \cdots x^n)}$$

is just the usual Jacobian of the transformation. We then have

$$dy^1 \wedge \cdots \wedge dy^n = \frac{\partial(y^1 \cdots y^n)}{\partial(x^1 \cdots x^n)} dx^1 \wedge \cdots \wedge dx^n.$$

The reader may recognize $dx^1 \wedge \cdots \wedge dx^n$ as the volume element on $\mathbb{R}^n$, and hence differential forms are a natural way to describe the change of variables in multiple integrals. //

Theorem 11.9 If $\alpha^1, \dots, \alpha^r \in \bigwedge^1(V)$, then $\{\alpha^1, \dots, \alpha^r\}$ is a linearly dependent set if and only if $\alpha^1 \wedge \cdots \wedge \alpha^r = 0$.

Proof If $\{\alpha^1, \dots, \alpha^r\}$ is linearly dependent, then there exists at least one vector, say α^1 , such that $\alpha^1 = \sum_{j \neq 1} a_j \alpha^j$. But then

$$\begin{aligned} \alpha^1 \wedge \cdots \wedge \alpha^r &= (\sum_{j \neq 1} a_j \alpha^j) \wedge \alpha^2 \wedge \cdots \wedge \alpha^r \\ &= \sum_{j \neq 1} a_j (\alpha^j \wedge \alpha^2 \wedge \cdots \wedge \alpha^r) \\ &= 0 \end{aligned}$$

since every term in the sum contains a repeated 1-form and hence vanishes.

Conversely, suppose that $\alpha^1, \dots, \alpha^r$ are linearly *independent*. We can then extend them to a basis $\{\alpha^1, \dots, \alpha^n\}$ for V^* (Theorem 2.10). If $\{e_i\}$ is the corresponding dual basis for V , then $\alpha^1 \wedge \cdots \wedge \alpha^n(e_1, \dots, e_n) = 1$ which implies that $\alpha^1 \wedge \cdots \wedge \alpha^r \neq 0$. Therefore $\{\alpha^1, \dots, \alpha^r\}$ must be linearly *dependent* if $\alpha^1 \wedge \cdots \wedge \alpha^r = 0$. ■

11.5 THE TENSOR PRODUCT OF VECTOR SPACES

We now discuss the notion of the tensor product of vector spaces. Our reason for presenting this discussion is that it provides the basis for defining the Kronecker (or direct) product of two matrices, a concept which is very useful in the theory of group representations.

It should be remarked that there are many ways of defining the tensor product of vector spaces. While we will follow the simplest approach, there is another (somewhat complicated) method involving quotient spaces that is also

frequently used. This other method has the advantage that it includes infinite-dimensional spaces. The reader can find a treatment of this alternative method in, e.g., in the book by Curtis (1984).

By way of nomenclature, we say that a mapping $f: U \times V \rightarrow W$ of vector spaces U and V to a vector space W is **bilinear** if f is linear in each variable. This is exactly the same as we defined in Section 9.4 except that now f takes its values in W rather than the field $\mathcal{F}$. In addition, we will need the concept of a vector space generated by a set. In other words, suppose $S = \{s_1, \dots, s_n\}$ is some finite set of objects, and $\mathcal{F}$ is a field. While we may have an intuitive sense of what it should mean to write formal linear combinations of the form $a_1s_1 + \dots + a_ns_n$, we should realize that the $+$ sign as used here has no meaning for an arbitrary set S . We now go through the formalities involved in defining such terms, and hence make the set S into a vector space T over $\mathcal{F}$.

The basic idea is that we want to recast each element of S into the form of a function from S to $\mathcal{F}$. This is because we already know how to add functions as well as multiply them by a scalar. With these ideas in mind, for each $s_i \in S$ we define a function $s_i: S \rightarrow \mathcal{F}$ by

$$s_i(s_j) = 1\delta_{ij}$$

where 1 is the multiplicative identity of $\mathcal{F}$. Since addition in $\mathcal{F}$ is well-defined as is the addition of functions and multiplication of functions by elements of $\mathcal{F}$, we see that for any $a, b \in \mathcal{F}$ and $s_i, s_j \in S$ we have

$$\begin{aligned} (a+b)s_i(s_j) &= (a+b)\delta_{ij} = a\delta_{ij} + b\delta_{ij} = as_i(s_j) + bs_i(s_j) \\ &= (as_i + bs_i)(s_j) \end{aligned}$$

and therefore $(a+b)s_i = as_i + bs_i$. Similarly, it is easy to see that $a(bs_i) = (ab)s_i$.

We now define T to be the set of all functions from S to $\mathcal{F}$. These functions can be written in the form $a_1s_1 + \dots + a_ns_n$ with $a_i \in \mathcal{F}$. It should be clear that with our definition of the terms a_is_i , T forms a vector space over $\mathcal{F}$. In fact, it is easy to see that the functions $1s_1, \dots, 1s_n$ are linearly independent. Indeed, if 0 denotes the zero function, suppose $a_1s_1 + \dots + a_ns_n = 0$ for some set of scalars a_i . Applying this function to s_i (where $1 \leq i \leq n$) we obtain $a_i = 0$. As a matter of course, we simply write s_i rather than $1s_i$.

The linear combinations just defined are called **formal** linear combinations of the elements of S , and T is the vector space **generated** by the set S . T is therefore the vector space of all such formal linear combinations, and is sometimes called the **free** vector space of S over $\mathcal{F}$.

Theorem 11.10 Let U , V and W be finite-dimensional vector spaces over $\mathcal{F}$. Then there exists a finite-dimensional vector space over $\mathcal{F}$ denoted by T and a bilinear mapping $t: U \times V \rightarrow T$ denoted by $t(u, v) = u \otimes v$ satisfying the following properties:

(a) For every bilinear mapping $f: U \times V \rightarrow W$, there exists a unique linear transformation $\tilde{f}: T \rightarrow W$ such that $f = \tilde{f} \circ t$. In other words, for all $u \in U$ and $v \in V$ we have

$$f(u, v) = \tilde{f}(t(u, v)) = \tilde{f}(u \otimes v) .$$

(b) If $\{u_1, \dots, u_m\}$ is a basis for U and $\{v_1, \dots, v_n\}$ is a basis for V , then $\{u_i \otimes v_j\}$ is a basis for T and therefore

$$\dim T = mn = (\dim U)(\dim V) .$$

Proof Let $\{u_1, \dots, u_m\}$ be a basis for U and let $\{v_1, \dots, v_n\}$ be a basis for V . For each pair of integers (i, j) with $1 \leq i \leq m$ and $1 \leq j \leq n$ we let t_{ij} be a letter (i.e., an element of some set). We now define T to be the vector space over $\mathcal{F}$ consisting of all formal linear combinations of the elements t_{ij} . In other words, every element of T is of the form $a^{ij} t_{ij}$ where $a^{ij} \in \mathcal{F}$.

Define the bilinear map $t: U \times V \rightarrow T$ by

$$u_i \otimes v_j \equiv t(u_i, v_j) = t_{ij}$$

and hence to all of $U \times V$ by “bilinear extension.” In particular, if $u = x^i u_i \in U$ and $v = y^j v_j \in V$, let us define $u \otimes v$ to be that element of T given by

$$u \otimes v = t(u, v) = x^i y^j t_{ij} .$$

It should be obvious that this does indeed define a bilinear map.

Now suppose that $f: U \times V \rightarrow W$ is any bilinear map, and remember that every element of T is a linear combination of the t_{ij} . According to Theorem 5.1, we may define a unique linear transformation $\tilde{f}: T \rightarrow W$ by

$$\tilde{f}(t_{ij}) = f(u_i, v_j) .$$

Using the bilinearity of f and the linearity of $\tilde{f}$ we then have

$$\begin{aligned} f(u, v) &= f(x^i u_i, y^j v_j) = x^i y^j f(u_i, v_j) = x^i y^j \tilde{f}(t_{ij}) = \tilde{f}(x^i y^j t_{ij}) \\ &= \tilde{f}(u \otimes v) = \tilde{f}(t(u, v)) . \end{aligned}$$

This proves the existence and uniqueness of the mapping $\tilde{f}$ such that $f = \tilde{f} \circ t$ as specified in (a).

We have defined T to be the vector space generated by the mn elements $t_{ij} = u_i \otimes v_j$ where $\{u_1, \dots, u_m\}$ and $\{v_1, \dots, v_n\}$ were particular bases for U and V respectively. We now want to show that in fact $\{u'_i \otimes v'_j\}$ forms a basis for T where $\{u'_i\}$ and $\{v'_j\}$ are arbitrary bases for U and V . For any $u = x'^i u'_i \in U$ and $v = y'^j v'_j \in V$, we have (using the bilinearity of $\otimes$)

$$u \otimes v = x'^i y'^j (u'_i \otimes v'_j)$$

which shows that the mn elements $u'_i \otimes v'_j$ span T . If these mn elements are linearly dependent, then $\dim T < mn$ which contradicts the fact that the mn elements t_{ij} form a basis for T . Hence $\{u'_i \otimes v'_j\}$ is a basis for T . ■

The space T defined in this theorem is denoted by $U \otimes V$ and called the **tensor product** of U and V . Note that T can be any mn dimensional vector space. For example, if $m = n$ we could take $T = \mathcal{T}_2(V)$ with basis $t_{ij} = \omega^i \otimes \omega^j$, $1 \leq i, j \leq n$. The map $t(u_i, v_j) = u_i \otimes v_j$ then defines $u_i \otimes v_j = \omega^i \otimes \omega^j$.

Example 11.10 To show how this formalism relates to our previous treatment of tensors, consider the following example of the mapping $\tilde{f}$ defined in Theorem 11.10. Let $\{e_i\}$ be a basis for a real inner product space U , and let us define the real numbers $g_{ij} = \langle e_i, e_j \rangle$. If $\bar{e}_i = e_j p^j_i$ is another basis for U , then

$$\bar{g}_{ij} = \langle \bar{e}_i, \bar{e}_j \rangle = p^r_i p^s_j \langle e_r, e_s \rangle = p^r_i p^s_j g_{rs}$$

so that the g_{ij} transform like the components of a covariant tensor of order 2. This means that we may define the tensor $g \in \mathcal{T}_2(U)$ by $g(u, v) = \langle u, v \rangle$. This tensor is called the **metric tensor** on U (see Section 11.10).

Now suppose that we are given a positive definite symmetric bilinear form (i.e., an inner product) $g = \langle \ , \ \rangle: U \times U \rightarrow \mathcal{F}$. Then the mapping $\tilde{g}$ is just the metric because

$$\tilde{g}(e_i \otimes e_j) = g(e_i, e_j) = \langle e_i, e_j \rangle = g_{ij}.$$

Therefore, if $u = u^i e_i$ and $v = v^j e_j$ are vectors in U , we see that

$$\tilde{g}(u \otimes v) = g(u, v) = \langle u, v \rangle = u^i v^j \langle e_i, e_j \rangle = g_{ij} u^i v^j.$$

If $\{\omega^i\}$ is the basis for U^* dual to $\{e_i\}$, then according to our earlier formalism, we would write this as $\tilde{g} = g_{ij} \omega^i \otimes \omega^j$. //

Some of the main applications in mathematics and physics (e.g., in the theory of group representations) of the tensor product of two vector spaces are contained in the next two results. While the ideas are simple enough, the notation becomes somewhat awkward because of the double indices.

Theorem 11.11 Let U and V have the respective bases $\{u_1, \dots, u_m\}$ and $\{v_1, \dots, v_n\}$, and suppose the linear operators $S \in L(U)$ and $T \in L(V)$ have matrix representations $A = (a^i_j)$ and $B = (b^i_j)$ respectively. Then there exists a linear transformation $S \otimes T: U \otimes V \rightarrow U \otimes V$ such that for all $u \in U$ and $v \in V$ we have $(S \otimes T)(u \otimes v) = S(u) \otimes T(v)$.

Furthermore, the matrix C of $S \otimes T$ relative to the ordered basis

$$\{u_1 \otimes v_1, \dots, u_1 \otimes v_n, u_2 \otimes v_1, \dots, u_2 \otimes v_n, \dots, u_m \otimes v_1, \dots, u_m \otimes v_n\}$$

for $U \otimes V$ is the $mn \times mn$ matrix given in block matrix form as

$$C = \begin{pmatrix} a^1_1 B & a^1_2 B & \cdots & a^1_m B \\ \vdots & \vdots & & \vdots \\ a^m_1 B & a^m_2 B & \cdots & a^m_m B \end{pmatrix}.$$

The matrix C is called the **Kronecker** (or **direct** or **tensor**) **product** of the matrices A and B , and will also be written as $C = A \otimes B$.

Proof Since S and T are linear and $\otimes$ is bilinear, it is easy to see that the mapping $f: U \times V \rightarrow U \otimes V$ defined by $f(u, v) = S(u) \otimes T(v)$ is bilinear. Therefore, according to Theorem 11.10, there exists a unique linear transformation $\tilde{f} \in L(U \otimes V)$ such that $\tilde{f}(u \otimes v) = S(u) \otimes T(v)$. We denote the mapping $\tilde{f}$ by $S \otimes T$. Thus, $(S \otimes T)(u \otimes v) = S(u) \otimes T(v)$.

To find the matrix C of $S \otimes T$ is straightforward enough. We have $S(u_i) = u_j a^j_i$ and $T(v_j) = v_s b^s_j$, and hence

$$(S \otimes T)(u_i \otimes v_j) = S(u_i) \otimes T(v_j) = a^r_i b^s_j (u_r \otimes v_s).$$

Now recall that the i th column of the matrix representation of an operator is just the image of the i th basis vector under the transformation (see Theorem 5.11). In the present case, we will have to use double pairs of subscripts to label the matrix elements. Relative to the ordered basis

$$\{u_1 \otimes v_1, \dots, u_1 \otimes v_n, u_2 \otimes v_1, \dots, u_2 \otimes v_n, \dots, u_m \otimes v_1, \dots, u_m \otimes v_n\}$$

for $U \otimes V$, we then see that, for example, the $(1, 1)$ th column of C is the vector $(S \otimes T)(u_1 \otimes v_1) = a^r_1 b^s_1 (u_r \otimes v_s)$ given by

$$(a^1_1 b^1_1, \dots, a^1_1 b^n_1, a^2_1 b^1_1, \dots, a^2_1 b^n_1, \dots, a^m_1 b^1_1, \dots, a^m_1 b^n_1)$$

and in general, the (i, j) th column is given by

$$(a^1_i b^1_j, \dots, a^1_i b^n_j, a^2_i b^1_j, \dots, a^2_i b^n_j, \dots, a^m_i b^1_j, \dots, a^m_i b^n_j) .$$

This shows that the matrix C has the desired form. ■

Theorem 11.12 Let U and V be finite-dimensional vector spaces over $\mathcal{F}$.

(a) If $S_1, S_2 \in L(U)$ and $T_1, T_2 \in L(V)$, then

$$(S_1 \otimes T_1)(S_2 \otimes T_2) = S_1 S_2 \otimes T_1 T_2 .$$

Moreover, if A_i and B_i are the matrix representations of S_i and T_i respectively (relative to some basis for $U \otimes V$), then $(A_1 \otimes B_1)(A_2 \otimes B_2) = A_1 A_2 \otimes B_1 B_2$.

(b) If $S \in L(U)$ and $T \in L(V)$, then $\text{Tr}(S \otimes T) = (\text{Tr } S)(\text{Tr } T)$.

(c) If $S \in L(U)$ and $T \in L(V)$, and if S^{-1} and T^{-1} exist, then

$$(S \otimes T)^{-1} = S^{-1} \otimes T^{-1} .$$

Conversely, if $(S \otimes T)^{-1}$ exists, then S^{-1} and T^{-1} also exist, and $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$.

Proof (a) For any $u \in U$ and $v \in V$ we have

$$\begin{aligned} (S_1 \otimes T_1)(S_2 \otimes T_2)(u \otimes v) &= (S_1 \otimes T_1)(S_2(u) \otimes T_2(v)) \\ &= S_1 S_2(u) \otimes T_1 T_2(v) \\ &= (S_1 S_2 \otimes T_1 T_2)(u \otimes v) . \end{aligned}$$

As to the matrix representations, simply note that $A_i \otimes B_i$ is the representation of $S_i \otimes T_i$, and $A_1 A_2 \otimes B_1 B_2$ is the representation of $S_1 S_2 \otimes T_1 T_2$ (since the representation of a product of linear operators is the product of their matrices).

(b) Recall that the trace of a linear operator is defined to be the trace of any matrix representation of the operator (see Theorem 5.19). Therefore, if $A = (a^i_j)$ is the matrix of S and $B = (b^i_j)$ is the matrix of T , we see from Theorem 11.11 that the diagonal blocks of $A \otimes B$ are the matrices $a^1_1 B, \dots, a^m_m B$ and hence the diagonal elements of $A \otimes B$ are $a^1_1 b^1_1, \dots, a^1_1 b^n_1, \dots, a^m_m b^1_1, \dots, a^m_m b^n_1$. Therefore the sum of these diagonal elements is just

$$\begin{aligned}
\text{Tr}(A \otimes B) &= a^1_1(\sum_i b^i_i) + \cdots + a^m_m(\sum_i b^i_i) \\
&= (\sum_j a^j_j)(\sum_i b^i_i) \\
&= (\text{Tr } A)(\text{Tr } B) .
\end{aligned}$$

(c) We first note that if 1 denotes the identity transformation, then

$$(1 \otimes 1)(u \otimes v) = u \otimes v$$

and hence $1 \otimes 1 = 1$. Next note that $u \otimes v = (u + 0) \otimes v = u \otimes v + 0 \otimes v$, and hence $0 \otimes v = 0$. Similarly, it is clear that $u \otimes 0 = 0$. This then shows that

$$(S \otimes 0)(u \otimes v) = S(u) \otimes 0 = 0$$

so that $S \otimes 0 = 0$, and similarly $0 \otimes T = 0$.

Now, if S and T are invertible, then by part (a) we see that

$$(S^{-1} \otimes T^{-1})(S \otimes T) = SS^{-1} \otimes T^{-1}T = 1 \otimes 1 = 1$$

and similarly for $(S \otimes T)(S^{-1} \otimes T^{-1})$. Therefore $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$.

Conversely, suppose that $S \otimes T$ is invertible. To prove that S and T are also invertible we use Theorem 5.9. In other words, a surjective linear operator is invertible if and only if its kernel is zero. Since $S \otimes T$ is invertible we must have $T \neq 0$, and hence there exists $v \in V$ such that $T(v) \neq 0$. Suppose $u \in U$ is such that $S(u) = 0$. Then

$$0 = S(u) \otimes T(v) = (S \otimes T)(u \otimes v)$$

which implies that $u \otimes v = 0$ (since $S \otimes T$ is invertible). But $v \neq 0$, and hence we must have $u = 0$. This shows that S is invertible. Similarly, had we started with $S \neq 0$, we would have found that T is invertible. ■

Exercises

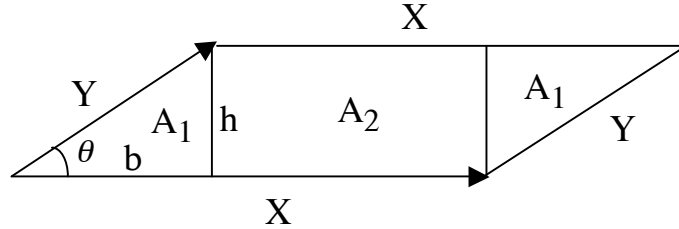
1. Give a direct proof of the matrix part of Theorem 11.12(a) using the definition of the Kronecker product of two matrices.
2. Suppose $A \in L(U)$ and $B \in L(V)$ where $\dim U = n$ and $\dim V = m$. Show that

$$\det(A \otimes B) = (\det A)^m (\det B)^n .$$

11.6 VOLUMES IN $\mathbb{R}^3$

Instead of starting out with an abstract presentation of volumes, we shall first go through an intuitive elementary discussion beginning with $\mathbb{R}^2$, then going to $\mathbb{R}^3$, and finally generalizing to $\mathbb{R}^n$ in the next section.

First consider a parallelogram in $\mathbb{R}^2$ (with the usual norm) defined by the vectors X and Y as shown.



Note that $h = \|Y\| \sin \theta$ and $b = \|Y\| \cos \theta$, and also that the area of each triangle is given by $A_1 = (1/2)bh$. Then the area of the rectangle is given by $A_2 = (\|X\| - b)h$, and the area of the entire parallelogram is given by

$$A = 2A_1 + A_2 = bh + (\|X\| - b)h = \|X\|h = \|X\|\|Y\|\sin \theta . \quad (1)$$

The reader should recognize this as the magnitude of the elementary “vector cross product” $X \times Y$ of the ordered pair of vectors (X, Y) that is *defined* to have a direction normal to the plane spanned by X and Y , and given by the “right hand rule” (i.e., *out* of the plane in this case).

If we define the usual orthogonal coordinate system with the x -axis parallel to the vector X , then

$$X = (x^1, x^2) = (\|X\|, 0)$$

and

$$Y = (y^1, y^2) = (\|Y\| \cos \theta, \|Y\| \sin \theta)$$

and hence we see that the determinant with columns formed from the vectors X and Y is just

$$\begin{vmatrix} x^1 & y^1 \\ x^2 & y^2 \end{vmatrix} = \begin{vmatrix} \|X\| & \|Y\| \cos \theta \\ 0 & \|Y\| \sin \theta \end{vmatrix} = \|X\|\|Y\|\sin \theta = A . \quad (2)$$

Notice that if we interchanged the vectors X and Y in the diagram, then the determinant would change sign and the vector $X \times Y$ (which by definition has a direction dependent on the *ordered* pair (X, Y)) would point *into* the page.

Thus the area of a parallelogram (which is always positive by definition) defined by two vectors in $\mathbb{R}^2$ is in general given by the absolute value of the determinant (2).

In terms of the usual inner product (or “dot product”) $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^2$, we have $\langle X, X \rangle = \|X\|^2$ and $\langle X, Y \rangle = \langle Y, X \rangle = \|X\|\|Y\| \cos \theta$, and hence

$$\begin{aligned} A^2 &= \|X\|^2 \|Y\|^2 \sin^2 \theta \\ &= \|X\|^2 \|Y\|^2 (1 - \cos^2 \theta) \\ &= \|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2. \end{aligned}$$

Therefore we see that the area is also given by the positive square root of the determinant

$$A^2 = \begin{vmatrix} \langle X, X \rangle & \langle X, Y \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle \end{vmatrix}. \quad (3)$$

It is also worth noting that the inner product may be written in the form $\langle X, Y \rangle = x^1 y^1 + x^2 y^2$, and thus in terms of matrices we may write

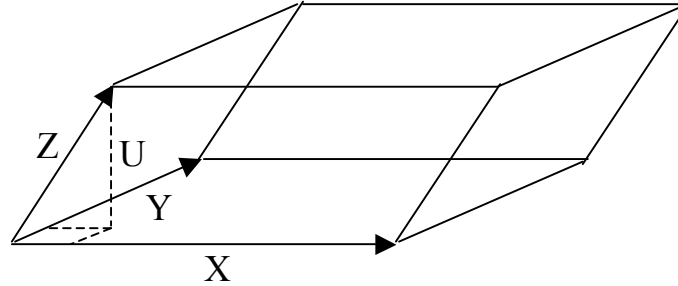
$$\begin{pmatrix} \langle X, X \rangle & \langle X, Y \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle \end{pmatrix} = \begin{pmatrix} x^1 & x^2 \\ y^1 & y^2 \end{pmatrix} \begin{pmatrix} x^1 & y^1 \\ x^2 & y^2 \end{pmatrix}.$$

Hence taking the determinant of this equation (using Theorems 4.8 and 4.1), we find (at least in $\mathbb{R}^2$) that the determinant (3) also implies that the area is given by the absolute value of the determinant in equation (2).

It is now easy to extend this discussion to a parallelogram in $\mathbb{R}^3$. Indeed, if $X = (x^1, x^2, x^3)$ and $Y = (y^1, y^2, y^3)$ are vectors in $\mathbb{R}^3$, then equation (1) is unchanged because any two vectors in $\mathbb{R}^3$ define the plane $\mathbb{R}^2$ spanned by the two vectors. Equation (3) also remains unchanged since its derivation did not depend on the specific coordinates of X and Y in $\mathbb{R}^2$. However, the left hand part of equation (2) does not apply (although we will see below that the three-dimensional version determines a volume in $\mathbb{R}^3$).

As a final remark on parallelograms, note that if X and Y are linearly dependent, then $aX + bY = 0$ so that $Y = -(a/b)X$, and hence X and Y are co-linear. Therefore θ equals 0 or π so that all equations for the area in terms of $\sin \theta$ are equal to zero. Since X and Y are dependent, this also means that the determinant in equation (2) equals zero, and everything is consistent.

We now take a look at volumes in $\mathbb{R}^3$. Consider three linearly independent vectors $X = (x^1, x^2, x^3)$, $Y = (y^1, y^2, y^3)$ and $Z = (z^1, z^2, z^3)$, and consider the parallelepiped with edges defined by these three vectors (in the given order (X, Y, Z)).



We claim that the volume of this parallelepiped is given by both the positive square root of the determinant

$$\begin{vmatrix} \langle X, X \rangle & \langle X, Y \rangle & \langle X, Z \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle & \langle Y, Z \rangle \\ \langle Z, X \rangle & \langle Z, Y \rangle & \langle Z, Z \rangle \end{vmatrix} \quad (4)$$

and the absolute value of the determinant

$$\begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix}. \quad (5)$$

To see this, first note that the volume of the parallelepiped is given by the product of the area of the base times the height, where the area A of the base is given by equation (3) and the height $\|U\|$ is just the projection of Z onto the orthogonal complement in $\mathbb{R}^3$ of the space spanned by X and Y . In other words, if W is the subspace of $V = \mathbb{R}^3$ spanned by X and Y , then (by Theorem 2.22) $V = W^\perp \oplus W$, and hence by Theorem 2.12 we may write

$$Z = U + aX + bY$$

where $U \in W^\perp$ and $a, b \in \mathbb{R}$ are uniquely determined (the uniqueness of a and b actually follows from Theorem 2.3 together with Theorem 2.12).

By definition we have $\langle X, U \rangle = \langle Y, U \rangle = 0$, and therefore

$$\begin{aligned}
\langle X, Z \rangle &= a\|X\|^2 + b\langle X, Y \rangle \\
\langle Y, Z \rangle &= a\langle Y, X \rangle + b\|Y\|^2 \\
\langle U, Z \rangle &= \|U\|^2.
\end{aligned} \tag{6}$$

We now wish to solve the first two of these equations for a and b by Cramer's rule (Theorem 4.13). Note that the determinant of the matrix of coefficients is just equation (3), and hence is just the square of the area A of the base of the parallelepiped. Applying Cramer's rule we have

$$\begin{aligned}
aA^2 &= \begin{vmatrix} \langle X, Z \rangle & \langle X, Y \rangle \\ \langle Y, Z \rangle & \langle Y, Y \rangle \end{vmatrix} = - \begin{vmatrix} \langle X, Y \rangle & \langle X, Z \rangle \\ \langle Y, Y \rangle & \langle Y, Z \rangle \end{vmatrix} \\
bA^2 &= \begin{vmatrix} \langle X, X \rangle & \langle X, Z \rangle \\ \langle Y, X \rangle & \langle Y, Z \rangle \end{vmatrix}.
\end{aligned}$$

Denoting the volume by $\text{Vol}(X, Y, Z)$, we now have (using the last of equations (6) together with $U = Z - aX - bY$)

$$\text{Vol}^2(X, Y, Z) = A^2\|U\|^2 = A^2\langle U, Z \rangle = A^2(\langle Z, Z \rangle - a\langle X, Z \rangle - b\langle Y, Z \rangle)$$

so that substituting the expressions for A^2 , aA^2 and bA^2 , we find

$$\begin{aligned}
\text{Vol}^2(X, Y, Z) &= \langle Z, Z \rangle \begin{vmatrix} \langle X, X \rangle & \langle X, Y \rangle \\ \langle Y, X \rangle & \langle Y, Y \rangle \end{vmatrix} + \langle X, Z \rangle \begin{vmatrix} \langle X, Y \rangle & \langle X, Z \rangle \\ \langle Y, Y \rangle & \langle Y, Z \rangle \end{vmatrix} \\
&\quad - \langle Y, Z \rangle \begin{vmatrix} \langle X, X \rangle & \langle X, Z \rangle \\ \langle Y, X \rangle & \langle Y, Z \rangle \end{vmatrix}.
\end{aligned}$$

Using $\langle X, Y \rangle = \langle Y, X \rangle$ etc., we see that this is just the expansion of a determinant by minors of the third row, and hence (using $\det A^T = \det A$)

$$\begin{aligned}
\text{Vol}^2(X, Y, Z) &= \begin{vmatrix} \langle X, X \rangle & \langle Y, X \rangle & \langle Z, X \rangle \\ \langle X, Y \rangle & \langle Y, Y \rangle & \langle Z, Y \rangle \\ \langle X, Z \rangle & \langle Y, Z \rangle & \langle Z, Z \rangle \end{vmatrix} \\
&= \begin{vmatrix} x^1 & x^2 & x^3 \\ y^1 & y^2 & y^3 \\ z^1 & z^2 & z^3 \end{vmatrix} \begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix} = \begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix}^2.
\end{aligned}$$

We remark that if the collection $\{X, Y, Z\}$ is linearly dependent, then the volume of the parallelepiped degenerates to zero (since at least one of the parallelograms that form the sides will have zero area). This agrees with the fact that the determinant (5) will vanish if two rows are linearly dependent. We also note that the area of the base is given by

$$|X \times Y| = \|X\| \|Y\| \sin \angle(X, Y)$$

where the direction of the vector $X \times Y$ is up (in this case). Therefore the projection of Z in the direction of $X \times Y$ is just Z dotted into a unit vector in the direction of $X \times Y$, and hence the volume of the parallelepiped is given by the number $Z \cdot (X \times Y)$. This is the so-called **scalar triple product** that should be familiar from elementary courses. We leave it to the reader to show that the scalar triple product is given by the determinant (5) (see Exercise 11.6.1).

Finally, note that if any two of the vectors X, Y, Z in equation (5) are interchanged, then the determinant changes sign even though the volume is unaffected (since it must be positive). This observation will form the basis for the concept of “orientation” to be defined later.

Exercises

1. Show that $Z \cdot (X \times Y)$ is given by the determinant in equation (5).
2. Find the area of the parallelogram whose vertices are:
 - (a) $(0, 0)$, $(1, 3)$, $(-2, 1)$, and $(-1, 4)$.
 - (b) $(2, 4)$, $(4, 5)$, $(5, 2)$, and $(7, 3)$.
 - (c) $(-1, 3)$, $(1, 5)$, $(3, 2)$, and $(5, 4)$.
 - (d) $(0, 0, 0)$, $(1, -2, 2)$, $(3, 4, 2)$, and $(4, 2, 4)$.
 - (e) $(2, 2, 1)$, $(3, 0, 6)$, $(4, 1, 5)$, and $(1, 1, 2)$.
3. Find the volume of the parallelepipeds whose adjacent edges are the vectors:
 - (a) $(1, 1, 2)$, $(3, -1, 0)$, and $(5, 2, -1)$.
 - (b) $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$.
4. Prove both algebraically and geometrically that the parallelogram with edges X and Y has the same area as the parallelogram with edges X and $Y + aX$ for any scalar a .

5. Prove both algebraically and geometrically that the volume of the parallelepiped in $\mathbb{R}^3$ with edges X , Y and Z is equal to the volume of the parallelepiped with edges X , Y and $Z + aX + bY$ for any scalars a and b .
6. Show that the parallelepiped in $\mathbb{R}^3$ defined by the three vectors $(2, 2, 1)$, $(1, -2, 2)$ and $(-2, 1, 2)$ is a cube. Find the volume of this cube.

11.7 VOLUMES IN $\mathbb{R}^n$

Now that we have a feeling for volumes in $\mathbb{R}^3$ expressed as determinants, let us prove the analogous results in $\mathbb{R}^n$. To begin with, we note that parallelograms defined by the vectors X and Y in either $\mathbb{R}^2$ or $\mathbb{R}^3$ contain all points (i.e., vectors) of the form $aX + bY$ for any $a, b \in [0, 1]$. Similarly, given three linearly independent vectors $X, Y, Z \in \mathbb{R}^3$, we may define the parallelepiped with these vectors as edges to be that subset of $\mathbb{R}^3$ containing all vectors of the form $aX + bY + cZ$ where $0 \leq a, b, c \leq 1$. The corners of the parallelepiped are the points $\delta_1 X + \delta_2 Y + \delta_3 Z$ where each δ_i is either 0 or 1.

Generalizing these observations, given any r linearly independent vectors $X_1, \dots, X_r \in \mathbb{R}^n$, we define an **r -dimensional parallelepiped** as the set of all vectors of the form $a_1 X_1 + \dots + a_r X_r$ where $0 \leq a_i \leq 1$ for each $i = 1, \dots, r$. In $\mathbb{R}^3$, by a **1-volume** we mean a length, a **2-volume** means an area, and a **3-volume** is just the usual volume. To define the volume of an r -dimensional parallelepiped we proceed by induction on r . In particular, if X is a nonzero vector (i.e., a 1-dimensional parallelepiped) in $\mathbb{R}^n$, we define its 1-volume to be its length $\langle X, X \rangle^{1/2}$. Proceeding, suppose the $(r - 1)$ -dimensional volume of an $(r - 1)$ -dimensional parallelepiped has been defined. If we let P_r denote the r -dimensional parallelepiped defined by the r linearly independent vectors $X_1, \dots, X_r$, then we say that the **base** of P_r is the $(r - 1)$ -dimensional parallelepiped defined by the $r - 1$ vectors $X_1, \dots, X_{r-1}$, and the **height** of P_r is the length of the projection of X_r onto the orthogonal complement in $\mathbb{R}^n$ of the space spanned by $X_1, \dots, X_{r-1}$. According to our induction hypothesis, the volume of an $(r - 1)$ -dimensional parallelepiped has already been defined. Therefore we define the **r -volume** of P_r to be the product of its height times the $(r - 1)$ -dimensional volume of its base.

The reader may wonder whether or not the r -volume of an r -dimensional parallelepiped in any way depends on which of the r vectors is singled out for projection. We proceed as if it does not and then, after the next theorem, we shall show that this is indeed the case.

Theorem 11.13 Let P_r be the r -dimensional parallelepiped defined by the r linearly independent vectors $X_1, \dots, X_r \in \mathbb{R}^n$. Then the r -volume of P_r is the positive square root of the determinant

$$\begin{vmatrix} \langle X_1, X_1 \rangle & \langle X_1, X_2 \rangle & \cdots & \langle X_1, X_r \rangle \\ \langle X_2, X_1 \rangle & \langle X_2, X_2 \rangle & \cdots & \langle X_2, X_r \rangle \\ \vdots & \vdots & & \vdots \\ \langle X_r, X_1 \rangle & \langle X_r, X_2 \rangle & \cdots & \langle X_r, X_r \rangle \end{vmatrix}. \quad (7)$$

Proof For the case of $r = 1$, we see that the theorem is true by the definition of length (or 1-volume) of a vector. Proceeding by induction, we assume the theorem is true for an $(r - 1)$ -dimensional parallelepiped, and we show that it is also true for an r -dimensional parallelepiped. Hence, let us write

$$A^2 = \text{Vol}^2(P_{r-1}) = \begin{vmatrix} \langle X_1, X_1 \rangle & \langle X_1, X_2 \rangle & \cdots & \langle X_1, X_{r-1} \rangle \\ \langle X_2, X_1 \rangle & \langle X_2, X_2 \rangle & \cdots & \langle X_2, X_{r-1} \rangle \\ \vdots & \vdots & & \vdots \\ \langle X_{r-1}, X_1 \rangle & \langle X_{r-1}, X_2 \rangle & \cdots & \langle X_{r-1}, X_{r-1} \rangle \end{vmatrix}$$

for the volume of the $(r - 1)$ -dimensional base of P_r . Just as we did in our discussion of volumes in $\mathbb{R}^3$, we write X_r in terms of its projection U onto the orthogonal complement of the space spanned by the $r - 1$ vectors $X_1, \dots, X_{r-1}$. This means that we can write

$$X_r = U + a_1 X_1 + \cdots + a_{r-1} X_{r-1}$$

where $\langle U, X_i \rangle = 0$ for $i = 1, \dots, r - 1$, and $\langle U, X_r \rangle = \langle U, U \rangle$. We thus have the system of equations

$$\begin{aligned} a_1 \langle X_1, X_1 \rangle + a_2 \langle X_1, X_2 \rangle + \cdots + a_{r-1} \langle X_1, X_{r-1} \rangle &= \langle X_1, X_r \rangle \\ a_1 \langle X_2, X_1 \rangle + a_2 \langle X_2, X_2 \rangle + \cdots + a_{r-1} \langle X_2, X_{r-1} \rangle &= \langle X_2, X_r \rangle \\ \vdots & \\ a_1 \langle X_{r-1}, X_1 \rangle + a_2 \langle X_{r-1}, X_2 \rangle + \cdots + a_{r-1} \langle X_{r-1}, X_{r-1} \rangle &= \langle X_{r-1}, X_r \rangle \end{aligned}$$

We write $M_1, \dots, M_{r-1}$ for the minors of the first $r - 1$ elements of the last row in (7). Solving the above system for the a_i using Cramer's rule, we obtain

$$\begin{aligned}
A^2 a_1 &= (-1)^{r-2} M_1 \\
A^2 a_2 &= (-1)^{r-3} M_2 \\
&\vdots \\
A^2 a_{r-1} &= M_{r-1}
\end{aligned}$$

where the factors of $(-1)^{r-k-1}$ in $A^2 a_k$ result from moving the last column of (7) over to become the k th column of the k th minor matrix.

Using this result, we now have

$$\begin{aligned}
A^2 U &= A^2(-a_1 X_1 - a_2 X_2 - \cdots - a_{r-1} X_{r-1} + X_r) \\
&= (-1)^{r-1} M_1 X_1 + (-1)^{r-2} M_2 X_2 + \cdots + (-1) M_{r-1} X_{r-1} + A^2 X_r
\end{aligned}$$

and hence, using $\|U\|^2 = \langle U, U \rangle = \langle U, X_r \rangle$, we find that (since $(-1)^{-k} = (-1)^k$)

$$\begin{aligned}
A^2 \|U\|^2 &= A^2 \langle U, X_r \rangle \\
&= (-1)^{r-1} M_1 \langle X_r, X_1 \rangle + (-1)(-1)^{r-1} M_2 \langle X_r, X_2 \rangle \\
&\quad + \cdots + A^2 \langle X_r, X_r \rangle \\
&= (-1)^{r-1} [M_1 \langle X_r, X_1 \rangle - M_2 \langle X_r, X_2 \rangle \\
&\quad + \cdots + (-1)^{r-1} A^2 \langle X_r, X_r \rangle] .
\end{aligned}$$

Now note that the right hand side of this equation is precisely the expansion of (7) by minors of the last row, and the left hand side is by definition the square of the r -volume of the r -dimensional parallelepiped P_r . This also shows that the determinant (7) is positive. ■

This result may also be expressed in terms of the matrix $(\langle X_i, X_j \rangle)$ as

$$\text{Vol}(P_r) = [\det(\langle X_i, X_j \rangle)]^{1/2} .$$

The most useful form of this theorem is given in the following corollary.

Corollary The n -volume of the n -dimensional parallelepiped in $\mathbb{R}^n$ defined by the vectors $X_1, \dots, X_n$ where each X_i has coordinates $(x^1_i, \dots, x^n_i)$ is the absolute value of the determinant of the matrix X given by

$$X = \begin{pmatrix} x^1_1 & x^1_2 & \cdots & x^1_n \\ x^2_1 & x^2_2 & \cdots & x^2_n \\ \vdots & \vdots & & \vdots \\ x^n_1 & x^n_2 & \cdots & x^n_n \end{pmatrix}.$$

Proof Note that $(\det X)^2 = (\det X)(\det X^T) = \det XX^T$ is just the determinant (7) in Theorem 11.13, which is the square of the volume. In other words, $\text{Vol}(P_n) = |\det X|$. ■

Prior to this theorem, we asked whether or not the r -volume depended on which of the r vectors is singled out for projection. We can now easily show that it does not. Suppose that we have an r -dimensional parallelepiped defined by r linearly independent vectors, and let us label these vectors $X_1, \dots, X_r$. According to Theorem 11.13, we project X_r onto the space orthogonal to the space spanned by $X_1, \dots, X_{r-1}$, and this leads to the determinant (7). If we wish to project any other vector instead, then we may simply relabel these r vectors to put a different one into position r . In other words, we have made some permutation of the indices in (7). However, remember that any permutation is a product of transpositions (Theorem 1.2), and hence we need only consider the effect of a single interchange of two indices.

Notice, for example, that the indices 1 and r only occur in rows 1 and r as well as in columns 1 and r . And in general, indices i and j only occur in the i th and j th rows and columns. But we also see that the matrix corresponding to (7) is symmetric about the main diagonal in these indices, and hence an interchange of the indices i and j has the effect of interchanging *both* rows i and j *as well as* columns i and j in exactly the same manner. Thus, because we have interchanged the same rows and columns there will be no sign change, and therefore the determinant (7) remains unchanged. In particular, it always remains positive. It now follows that the volume we have defined is indeed independent of which of the r vectors is singled out to be the height of the parallelepiped.

Now note that according to the above corollary, we know that $\text{Vol}(P_n) = \text{Vol}(X_1, \dots, X_n) = |\det X|$ which is always positive. While our discussion just showed that $\text{Vol}(X_1, \dots, X_n)$ is independent of any permutation of indices, the actual value of $\det X$ can change sign upon any such permutation. Because of this, we say that the vectors $(X_1, \dots, X_n)$ are **positively oriented** if $\det X > 0$, and **negatively oriented** if $\det X < 0$. Thus the orientation of a set of vectors depends on the order in which they are written. To take into account the sign of $\det X$, we define the **oriented volume** $\text{Vol}_o(X_1, \dots, X_n)$ to be $+\text{Vol}(X_1, \dots, X_n)$ if $\det X \geq 0$, and $-\text{Vol}(X_1, \dots, X_n)$ if $\det X < 0$. We will return to a careful discussion of orientation in a later section. We also

remark that $\det X$ is always nonzero as long as the vectors $(X_1, \dots, X_n)$ are linearly independent. Thus the above corollary may be expressed in the form

$$\text{Vol}_0(X_1, \dots, X_n) = \det(X_1, \dots, X_n)$$

where $\det(X_1, \dots, X_n)$ means the determinant as a function of the column vectors X_i .

Exercises

- Find the 3-volume of the three-dimensional parallelepipeds in $\mathbb{R}^4$ defined by the vectors:
 - $(2, 1, 0, -1)$, $(3, -1, 5, 2)$, and $(0, 4, -1, 2)$.
 - $(1, 1, 0, 0)$, $(0, 2, 2, 0)$, and $(0, 0, 3, 3)$.
- Find the 2-volume of the parallelogram in $\mathbb{R}^4$ two of whose edges are the vectors $(1, 3, -1, 6)$ and $(-1, 2, 4, 3)$.
- Prove that if the vectors $X_1, X_2, \dots, X_r$ are mutually orthogonal, the r -volume of the parallelepiped defined by them is equal to the product of their lengths.
- Prove that r vectors $X_1, X_2, \dots, X_r$ in $\mathbb{R}^n$ are linearly dependent if and only if the determinant (7) is equal to zero.

11.8 LINEAR TRANSFORMATIONS AND VOLUMES

One of the most useful applications of Theorem 11.13 and its corollary relates to linear mappings. In fact, this is the approach usually followed in deriving the change of variables formula for multiple integrals. Let $\{e_i\}$ be an orthonormal basis for $\mathbb{R}^n$, and let C_n denote the **unit cube** in $\mathbb{R}^n$. In other words,

$$C_n = \{t_1 e_1 + \dots + t_n e_n \in \mathbb{R}^n: 0 \leq t_i \leq 1\} .$$

This is similar to the definition of P_r given previously.

Now let $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation. Then the matrix of A relative to the basis $\{e_i\}$ is defined by $A(e_i) = e_j a_{ji}^j$. Let us write the image of e_i as X_i , so that $X_i = A(e_i) = e_j a_{ji}^j$. This means that the column vector X_i has

components $(a^1_i, \dots, a^n_i)$. Under the transformation A , the image of C_n becomes

$$A(C_n) = A(\sum t_i e_i) = \sum t_i A(e_i) = \sum t_i X_i$$

(where $0 \leq t_i \leq 1$) which is just the parallelepiped P_n spanned by the vectors $(X_1, \dots, X_n)$. Therefore the volume of $P_n = A(C_n)$ is given by

$$|\det(X_1, \dots, X_n)| = |\det(a^j_i)|.$$

Recalling that the determinant of a linear transformation is defined to be the determinant of its matrix representation, we have proved the next result.

Theorem 11.14 Let C_n be the unit cube in $\mathbb{R}^n$ spanned by the orthonormal basis vectors $\{e_i\}$. If $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear transformation and $P_n = A(C_n)$, then $\text{Vol}(P_n) = \text{Vol } A(C_n) = |\det A|$.

It is quite simple to generalize this result somewhat to include the image of an n -dimensional parallelepiped under a linear transformation A . First, we note that any parallelepiped P_n is just the image of C_n under some linear transformation B . Indeed, if $P_n = \{t_1 X_1 + \dots + t_n X_n: 0 \leq t_i \leq 1\}$ for some set of vectors X_i , then we may define the transformation B by $B(e_i) = X_i$, and hence $P_n = B(C_n)$. Thus

$$A(P_n) = A(B(C_n)) = (A \circ B)(C_n)$$

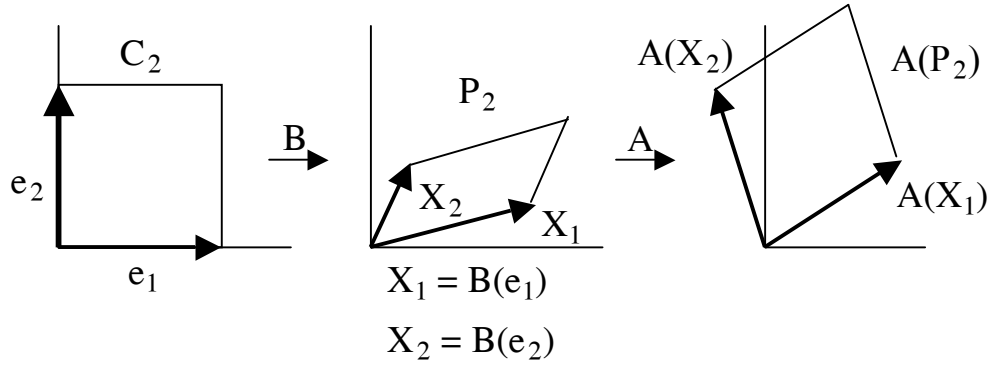
and therefore (using Theorem 11.14 along with the fact that the matrix of the composition of two transformations is the matrix product)

$$\begin{aligned} \text{Vol } A(P_n) &= \text{Vol}[(A \circ B)(C_n)] = |\det(A \circ B)| = |\det A| |\det B| \\ &= |\det A| \text{Vol}(P_n). \end{aligned}$$

In other words, $|\det A|$ is a measure of how much the volume of the parallelepiped changes under the linear transformation A . See the figure below for a picture of this in $\mathbb{R}^2$.

We summarize this discussion as a corollary to Theorem 11.14.

Corollary Suppose P_n is an n -dimensional parallelepiped in $\mathbb{R}^n$, and let $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation. Then $\text{Vol } A(P_n) = |\det A| \text{Vol}(P_n)$.



Now that we have an intuitive grasp of these concepts, let us look at this material from the point of view of exterior algebra. This more sophisticated approach is of great use in the theory of integration.

Let \$U\$ and \$V\$ be real vector spaces. Recall from Theorem 9.7 that given a linear transformation \$T \in L(U, V)\$, we defined the transpose mapping \$T^* \in L(V^*, U^*)\$ by

$$T^*\omega = \omega \circ T$$

for all \$\omega \in V^*\$. By this we mean if \$u \in U\$, then \$T^*\omega(u) = \omega(Tu)\$. As we then saw in Theorem 9.8, if \$U\$ and \$V\$ are finite-dimensional and \$A\$ is the matrix representation of \$T\$, then \$A^T\$ is the matrix representation of \$T^*\$, and hence certain properties of \$T^*\$ follow naturally. For example, if \$T_1 \in L(V, W)\$ and \$T_2 \in L(U, V)\$, then \$(T_1 \circ T_2)^* = T_2^* \circ T_1^*\$ (Theorem 3.18), and if \$T\$ is nonsingular, then \$(T^{-1})^* = (T^*)^{-1}\$ (Corollary 4 of Theorem 3.21).

Now suppose that \$\{e_i\}\$ is a basis for \$U\$ and \$\{f_j\}\$ is a basis for \$V\$. To keep the notation simple and understandable, let us write the corresponding dual bases as \$\{e^i\}\$ and \$\{f^j\}\$. We define the matrix \$A = (a^i_j)\$ of \$T\$ by \$Te_i = f_j a^j_i\$. Then (just as in the proof of Theorem 9.8)

$$\begin{aligned} (T^*f^i)e_j &= f^i(Te_j) = f^i(f_k a^k_j) = a^k_j f^i(f_k) = a^k_j \delta^i_k = a^i_j = a^i_k \delta^k_j \\ &= a^i_k e^k(e_j) \end{aligned}$$

which shows that

$$T^*f^i = a^i_k e^k. \quad (8)$$

We will use this result frequently below.

We now generalize our definition of the transpose. If \$\phi \in L(U, V)\$ and \$T \in \mathcal{T}_r(V)\$, we define the **pull-back** \$\phi^* \in L(\mathcal{T}_r(V), \mathcal{T}_r(U))\$ by

$$(\phi^*T)(u_1, \dots, u_r) = T(\phi(u_1), \dots, \phi(u_r))$$

where $u_1, \dots, u_r \in U$. Note that in the particular case of $r = 1$, the mapping ϕ^* is just the transpose of ϕ . It should also be clear from the definition that ϕ^* is indeed a linear transformation, and hence

$$\phi^*(aT_1 + bT_2) = a\phi^*T_1 + b\phi^*T_2 .$$

We also emphasize that ϕ need not be an isomorphism for us to define ϕ^* .

The main properties of the pull-back are given in the next theorem.

Theorem 11.15 If $\phi \in L(U, V)$ and $\psi \in L(V, W)$, then

- (a) $(\psi \circ \phi)^* = \phi^* \circ \psi^*$.
- (b) If $I \in L(U)$ is the identity map, then I^* is the identity in $L(\mathcal{T}_r(U))$.
- (c) If ϕ is an isomorphism, then so is ϕ^* , and $(\phi^*)^{-1} = (\phi^{-1})^*$.
- (d) If $T_1 \in \mathcal{T}_{r_1}(V)$ and $T_2 \in \mathcal{T}_{r_2}(V)$, then

$$\phi^*(T_1 \otimes T_2) = (\phi^*T_1) \otimes (\phi^*T_2) .$$

(e) Let U have basis $\{e_1, \dots, e_m\}$, V have basis $\{f_1, \dots, f_n\}$ and suppose that $\phi(e_i) = f_j a_{ji}^1$. If $T \in \mathcal{T}_r(V)$ has components $T_{i_1 \dots i_r} = T(f_{i_1}, \dots, f_{i_r})$, then the components of ϕ^*T relative to the basis $\{e_i\}$ are given by

$$(\phi^*T)_{j_1 \dots j_r} = T_{i_1 \dots i_r} a_{j_1 i_1}^1 \dots a_{j_r i_r}^1 .$$

Proof (a) Note that $\psi \circ \phi: U \rightarrow W$, and hence $(\psi \circ \phi)^*: \mathcal{T}_r(W) \rightarrow \mathcal{T}_r(U)$. Thus for any $T \in \mathcal{T}_r(W)$ and $u_1, \dots, u_r \in U$ we have

$$\begin{aligned} ((\psi \circ \phi)^*T)(u_1, \dots, u_r) &= T(\psi(\phi(u_1)), \dots, \psi(\phi(u_r))) \\ &= (\psi^*T)(\phi(u_1), \dots, \phi(u_r)) \\ &= ((\phi^* \circ \psi^*)T)(u_1, \dots, u_r) . \end{aligned}$$

- (b) Obvious from the definition of I^* .
- (c) If ϕ is an isomorphism, then ϕ^{-1} exists and we have (using (a) and (b))

$$\phi^* \circ (\phi^{-1})^* = (\phi^{-1} \circ \phi)^* = I^* .$$

Similarly $(\phi^{-1})^* \circ \phi^* = I^*$. Hence $(\phi^*)^{-1}$ exists and is equal to $(\phi^{-1})^*$.

- (d) This follows directly from the definitions (see Exercise 11.8.1).
- (e) Using the definitions, we have

$$\begin{aligned}
(\phi^*T)_{j_1 \dots j_r} &= (\phi^*T)(e_{j_1}, \dots, e_{j_r}) \\
&= T(\phi(e_{j_1}), \dots, \phi(e_{j_r})) \\
&= T(f_{i_1} a_{j_1}^{i_1}, \dots, f_{i_r} a_{j_r}^{i_r}) \\
&= T(f_{i_1}, \dots, f_{i_r}) a_{j_1}^{i_1} \dots a_{j_r}^{i_r} \\
&= T_{i_1 \dots i_r} a_{j_1}^{i_1} \dots a_{j_r}^{i_r} .
\end{aligned}$$

Alternatively, if $\{e^i\}$ and $\{f^j\}$ are the bases dual to $\{e_i\}$ and $\{f_j\}$ respectively, then $T = T_{i_1 \dots i_r} e^{i_1} \otimes \dots \otimes e^{i_r}$ and consequently (using the linearity of ϕ^* , part (d) and equation (8)),

$$\begin{aligned}
\phi^*T &= T_{i_1 \dots i_r} \phi^*e^{i_1} \otimes \dots \otimes \phi^*e^{i_r} \\
&= T_{i_1 \dots i_r} a_{j_1}^{i_1} \dots a_{j_r}^{i_r} f^{j_1} \otimes \dots \otimes f^{j_r}
\end{aligned}$$

which therefore yields the same result. ■

For our present purposes, we will only need to consider the pull-back as defined on the space $\bigwedge^r(V)$ rather than on $\mathcal{T}_r(V)$. Therefore, if $\phi \in L(U, V)$ then $\phi^* \in L(\mathcal{T}_r(V), \mathcal{T}_r(U))$, and hence we see that for $\omega \in \bigwedge^r(V)$ we have $(\phi^*\omega)(u_1, \dots, u_r) = \omega(\phi(u_1), \dots, \phi(u_r))$. This shows that $\phi^*(\bigwedge^r(V)) \subset \bigwedge^r(U)$. Parts (d) and (e) of Theorem 11.15 applied to the space $\bigwedge^r(V)$ yield the following special cases. (Recall that $|i_1 \dots i_r|$ means the sum is over increasing indices $i_1 < \dots < i_r$.)

Theorem 11.16 Suppose $\phi \in L(U, V)$, $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$. Then

(a) $\phi^*(\alpha \wedge \beta) = (\phi^*\alpha) \wedge (\phi^*\beta)$.

(b) Let U and V have bases $\{e_i\}$ and $\{f_j\}$ respectively, and let U^* and V^* have bases $\{e^i\}$ and $\{f^j\}$. If we write $\phi(e_i) = f_j a_i^j$ and $\phi^*(f^j) = a_i^j e^i$, and if $\alpha = a_{|i_1 \dots i_r|} f^{i_1} \wedge \dots \wedge f^{i_r} \in \bigwedge^r(V)$, then

$$\phi^*\alpha = \hat{a}_{|k_1 \dots k_r|} e^{k_1} \wedge \dots \wedge e^{k_r}$$

where

$$\hat{a}_{|k_1 \dots k_r|} = a_{|i_1 \dots i_r|} \varepsilon_{k_1 \dots k_r}^{j_1 \dots j_r} a_{j_1}^{i_1} \dots a_{j_r}^{i_r} .$$

Thus we may write

$$\hat{a}_{k_1 \dots k_r} = a_{|i_1 \dots i_r|} \det(a^I_K)$$

where

$$\det(a^I_K) = \begin{vmatrix} a^{i_1}_{k_1} & \cdots & a^{i_1}_{k_r} \\ \vdots & & \vdots \\ a^{i_r}_{k_1} & \cdots & a^{i_r}_{k_r} \end{vmatrix}.$$

Proof (a) For simplicity, let us write $(\phi^*\alpha)(u_J) = \alpha(\phi(u_J))$ instead of $(\phi^*\alpha)(u_1, \dots, u_r) = \alpha(\phi(u_1), \dots, \phi(u_r))$ (see the discussion following Theorem 11.4). Then, in an obvious notation, we have

$$\begin{aligned} [\phi^*(\alpha \wedge \beta)](u_I) &= (\alpha \wedge \beta)(\phi(u_I)) \\ &= \sum_{J, K} \varepsilon_I^{JK} (\alpha(\phi(u_J)) \beta(\phi(u_K))) \\ &= \sum_{J, K} \varepsilon_I^{JK} (\phi^*\alpha)(u_J) (\phi^*\beta)(u_K) \\ &= [(\phi^*\alpha) \wedge (\phi^*\beta)](u_I). \end{aligned}$$

By induction, this also obviously applies to the wedge product of a finite number of forms.

(b) From $\alpha = a_{i_1 \dots i_r} f^{i_1} \wedge \dots \wedge f^{i_r}$ and $\phi^*(f^i) = a^i_j e^j$, we have (using part (a) and the linearity of ϕ^*)

$$\begin{aligned} \phi^*\alpha &= a_{i_1 \dots i_r} \phi^*(f^{i_1}) \wedge \dots \wedge \phi^*(f^{i_r}) \\ &= a_{i_1 \dots i_r} a^{i_1}_{j_1} \cdots a^{i_r}_{j_r} e^{j_1} \wedge \dots \wedge e^{j_r}. \end{aligned}$$

But

$$e^{j_1} \wedge \dots \wedge e^{j_r} = \sum_K \varepsilon_{k_1 \dots k_r}^{j_1 \dots j_r} e^{k_1} \wedge \dots \wedge e^{k_r}$$

and hence we have

$$\begin{aligned} \phi^*\alpha &= a_{i_1 \dots i_r} \sum_K \varepsilon_{k_1 \dots k_r}^{j_1 \dots j_r} a^{i_1}_{j_1} \cdots a^{i_r}_{j_r} e^{k_1} \wedge \dots \wedge e^{k_r} \\ &= \hat{a}_{k_1 \dots k_r} e^{k_1} \wedge \dots \wedge e^{k_r} \end{aligned}$$

where

$$\hat{a}_{k_1 \dots k_r} = a_{i_1 \dots i_r} \varepsilon_{k_1 \dots k_r}^{j_1 \dots j_r} a^{i_1}_{j_1} \cdots a^{i_r}_{j_r}.$$

Finally, from the definition of determinant we see that

$$\varepsilon_{k_1 \dots k_r}^{j_1 \dots j_r} a^{i_1}_{j_1} \cdots a^{i_r}_{j_r} = \begin{vmatrix} a^{i_1}_{k_1} & \cdots & a^{i_1}_{k_r} \\ \vdots & & \vdots \\ a^{i_r}_{k_1} & \cdots & a^{i_r}_{k_r} \end{vmatrix}. \blacksquare$$

Example 11.11 (This is a continuation of Example 11.1.) An important example of $\phi^*\alpha$ is related to the change of variables formula in multiple integrals. While we are not in any position to present this material in detail, the idea is this. Suppose we consider the spaces $U = \mathbb{R}^3(u, v, w)$ and $V = \mathbb{R}^3(x, y, z)$ where the letters in parentheses tell us the coordinate system used for that particular copy of $\mathbb{R}^3$. Note that if we write $(x, y, z) = (x^1, x^2, x^3)$ and $(u, v, w) = (u^1, u^2, u^3)$, then from elementary calculus we know that $dx^i = (\partial x^i / \partial u^j) du^j$ and $\partial / \partial u^i = (\partial x^j / \partial u^i) (\partial / \partial x^j)$.

Now recall from Example 11.1 that at each point of $\mathbb{R}^3(u, v, w)$, the tangent space has the basis $\{e_i\} = \{\partial / \partial u^i\}$ and the cotangent space has the corresponding dual basis $\{e^i\} = \{du^i\}$, with a similar result for $\mathbb{R}^3(x, y, z)$. Let us define $\phi: \mathbb{R}^3(u, v, w) \rightarrow \mathbb{R}^3(x, y, z)$ by

$$\phi(\partial / \partial u^i) = (\partial x^j / \partial u^i) (\partial / \partial x^j) = a_j^i (\partial / \partial x^j) .$$

It is then apparent that (see equation (8))

$$\phi^*(dx^i) = a_j^i du^j = (\partial x^i / \partial u^j) du^j$$

as we should have expected. We now apply this to the 3-form

$$\alpha = a_{123} dx^1 \wedge dx^2 \wedge dx^3 = dx \wedge dy \wedge dz \in \Lambda^3(V) .$$

Since we are dealing with a 3-form in a 3-dimensional space, we must have

$$\phi^*\alpha = \hat{a} du \wedge dv \wedge dw$$

where $\hat{a} = \hat{a}_{123}$ consists of the single term given by the determinant

$$\begin{vmatrix} a_1^1 & a_1^2 & a_1^3 \\ a_2^1 & a_2^2 & a_2^3 \\ a_3^1 & a_3^2 & a_3^3 \end{vmatrix} = \begin{vmatrix} \partial x^1 / \partial u^1 & \partial x^1 / \partial u^2 & \partial x^1 / \partial u^3 \\ \partial x^2 / \partial u^1 & \partial x^2 / \partial u^2 & \partial x^2 / \partial u^3 \\ \partial x^3 / \partial u^1 & \partial x^3 / \partial u^2 & \partial x^3 / \partial u^3 \end{vmatrix}$$

which the reader may recognize as the so-called Jacobian of the transformation. This determinant is usually written as $\partial(x, y, z) / \partial(u, v, w)$, and hence we see that

$$\phi^*(dx \wedge dy \wedge dz) = \frac{\partial(x, y, z)}{\partial(u, v, w)} du \wedge dv \wedge dw .$$

This is precisely how volume elements transform (at least locally), and hence we have formulated the change of variables formula in quite general terms. //

This formalism allows us to define the determinant of a linear transformation in an interesting abstract manner. To see this, suppose $\phi \in L(V)$ where $\dim V = n$. Since $\dim \bigwedge^n(V) = 1$, we may choose any nonzero $\omega_0 \in \bigwedge^n(V)$ as a basis. Then $\phi^*: \bigwedge^n(V) \rightarrow \bigwedge^n(V)$ is linear, and hence for any $\omega = c_0\omega_0 \in \bigwedge^n(V)$ we have

$$\phi^*\omega = \phi^*(c_0\omega_0) = c_0\phi^*\omega_0 = c_0c\omega_0 = c(c_0\omega_0) = c\omega$$

for some scalar c (since $\phi^*\omega_0 \in \bigwedge^n(V)$ is necessarily of the form $c\omega_0$). Noting that this result did not depend on the scalar c_0 and hence is independent of $\omega = c_0\omega_0$, we see that the scalar c must be unique. We therefore define the **determinant** of ϕ to be the unique scalar, denoted by $\det \phi$, such that

$$\phi^*\omega = (\det \phi)\omega .$$

It is important to realize that this definition of the determinant does not depend on any choice of basis for V . However, let $\{e_i\}$ be a basis for V , and define the matrix (a^i_j) of ϕ by $\phi(e_i) = e_j a^j_i$. Then for any nonzero $\omega \in \bigwedge^n(V)$ we have

$$(\phi^*\omega)(e_1, \dots, e_n) = (\det \phi)\omega(e_1, \dots, e_n) .$$

On the other hand, Example 11.2 shows us that

$$\begin{aligned} (\phi^*\omega)(e_1, \dots, e_n) &= \omega(\phi(e_1), \dots, \phi(e_n)) \\ &= a^{i_1}_1 \cdots a^{i_n}_n \omega(e_{i_1}, \dots, e_{i_n}) \\ &= (\det(a^i_j))\omega(e_1, \dots, e_n) . \end{aligned}$$

Since $\omega \neq 0$, we have therefore proved the next result.

Theorem 11.17 If V has basis $\{e_1, \dots, e_n\}$ and $\phi \in L(V)$ has the matrix representation (a^i_j) defined by $\phi(e_i) = e_j a^j_i$, then $\det \phi = \det(a^i_j)$.

In other words, our abstract definition of the determinant is exactly the same as our earlier classical definition. In fact, it is now easy to derive some of the properties of the determinant that were not exactly simple to prove in the more traditional manner.

Theorem 11.18 If V is finite-dimensional and $\phi, \psi \in L(V, V)$, then

- (a) $\det(\phi \circ \psi) = (\det \phi)(\det \psi)$.
- (b) If ϕ is the identity transformation, then $\det \phi = 1$.
- (c) ϕ is an isomorphism if and only if $\det \phi \neq 0$, and if this is the case, then $\det \phi^{-1} = (\det \phi)^{-1}$.

Proof (a) By definition we have $(\phi \circ \psi)^*\omega = \det(\phi \circ \psi)\omega$. On the other hand, by Theorem 11.15(a) we know that $(\phi \circ \psi)^* = \psi^* \circ \phi^*$, and hence

$$\begin{aligned} (\phi \circ \psi)^*\omega &= \psi^*(\phi^*\omega) = \psi^*[(\det \phi)\omega] = (\det \phi)\psi^*\omega \\ &= (\det \phi)(\det \psi)\omega. \end{aligned}$$

(b) If $\phi = 1$ then $\phi^* = 1$ also (by Theorem 11.15(b)), and thus $\omega = \phi^*\omega = (\det \phi)\omega$ implies $\det \phi = 1$.

(c) First assume that ϕ is an isomorphism so that ϕ^{-1} exists. Then by parts (a) and (b) we see that

$$1 = \det(\phi\phi^{-1}) = (\det \phi)(\det \phi^{-1})$$

which implies $\det \phi \neq 0$ and $\det \phi^{-1} = (\det \phi)^{-1}$. Conversely, suppose that ϕ is not an isomorphism. Then $\text{Ker } \phi \neq 0$ and there exists a nonzero $e_1 \in V$ such that $\phi(e_1) = 0$. By Theorem 2.10, we can extend this to a basis $\{e_1, \dots, e_n\}$ for V . But then for any nonzero $\omega \in \bigwedge^n(V)$ we have

$$\begin{aligned} (\det \phi)\omega(e_1, \dots, e_n) &= (\phi^*\omega)(e_1, \dots, e_n) \\ &= \omega(\phi(e_1), \dots, \phi(e_n)) \\ &= \omega(0, \phi(e_2), \dots, \phi(e_n)) \\ &= 0 \end{aligned}$$

and hence we must have $\det \phi = 0$. ■

Exercises

1. Prove Theorem 11.15(d).
2. Show that the matrix ϕ^*T defined in Theorem 11.15(e) is just the r -fold Kronecker product $A \otimes \dots \otimes A$ where $A = (a^i_j)$.

The next three exercises are related.

3. Let $\phi \in L(U, V)$ be an isomorphism, and suppose $T \in \mathcal{T}_r^s(U)$. Define the **push-forward** $\phi_* \in L(\mathcal{T}_r^s(U), \mathcal{T}_r^s(V))$ by

$$\phi_* T(\alpha^1, \dots, \alpha^s, u_1, \dots, u_r) = T(\phi^* \alpha^1, \dots, \phi^* \alpha^s, \phi^{-1} u_1, \dots, \phi^{-1} u_r)$$

where $\alpha^1, \dots, \alpha^s \in U^*$ and $u_1, \dots, u_r \in U$. If $\psi \in L(V, W)$ is also an isomorphism, prove the following:

- (a) $(\psi \circ \phi)_* = \psi_* \circ \phi_*$.
- (b) If $I \in L(U)$ is the identity map, then so is $I_* \in L(\mathcal{T}_r^s(U))$.
- (c) ϕ_* is an isomorphism, and $(\phi_*)^{-1} = (\phi^{-1})_*$.
- (d) If $T_1 \in \mathcal{T}_{r_1}^{s_1}(U)$ and $T_2 \in \mathcal{T}_{r_2}^{s_2}(U)$, then

$$\phi_*(T_1 \otimes T_2) = (\phi_* T_1) \otimes (\phi_* T_2) .$$

4. Let $\phi \in L(U, V)$ be an isomorphism, and let U and V have bases $\{e_i\}$ and $\{f_i\}$ respectively. Define the matrices (a^i_j) and (b^i_j) by $\phi(e_i) = f_j a^j_i$ and $\phi^{-1}(f_i) = e_j b^j_i$. Suppose $T \in \mathcal{T}_r^s(U)$ has components $T^{i_1 \dots i_s}_{j_1 \dots j_r}$ relative to $\{e_i\}$, and $S \in \mathcal{T}_r^s(V)$ has components $S^{i_1 \dots i_s}_{j_1 \dots j_r}$ relative to $\{f_i\}$. Show that the components of $\phi_* T$ and $\phi_* S$ are given by

$$\begin{aligned} (\phi_* T)^{i_1 \dots i_s}_{j_1 \dots j_r} &= a^{i_1}_{p_1} \dots a^{i_s}_{p_s} T^{p_1 \dots p_s}_{q_1 \dots q_r} b^{q_1}_{j_1} \dots b^{q_r}_{j_r} \\ (\phi_* S)^{i_1 \dots i_s}_{j_1 \dots j_r} &= b^{i_1}_{p_1} \dots b^{i_s}_{p_s} S^{p_1 \dots p_s}_{q_1 \dots q_r} a^{q_1}_{j_1} \dots a^{q_r}_{j_r} . \end{aligned}$$

5. Let $\{\omega^i\}$ be the basis dual to $\{e_i\}$ for $\mathbb{R}^2$. Let

$$T = 2e_1 \otimes \omega^1 - e_2 \otimes \omega^1 + 3e_1 \otimes \omega^2$$

and suppose $\phi \in L(\mathbb{R}^2)$ and $\psi \in L(\mathbb{R}^3, \mathbb{R}^2)$ have the matrix representations

$$\phi = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad \psi = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 2 \end{pmatrix} .$$

Compute $\text{Tr } T$, $\phi^* T$, $\psi^* T$, $\text{Tr}(\psi^* T)$, and $\phi_* T$.

11.9 ORIENTATIONS AND VOLUMES

Suppose $\dim V = n$ and consider the space $\bigwedge^n(V)$. Since this space is 1-dimensional, we consider the n -form

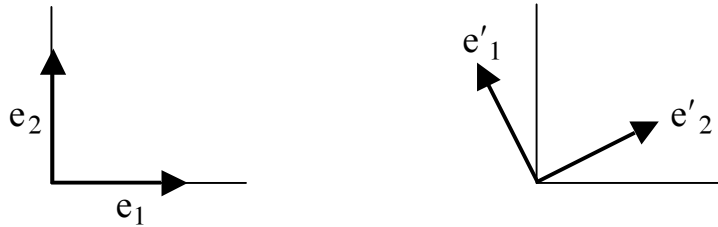
$$\omega = e^1 \wedge \cdots \wedge e^n \in \bigwedge^n(V)$$

where the basis $\{e^i\}$ for V^* is dual to the basis $\{e_i\}$ for V . If $\{v_i = e_j v^j_i\}$ is any set of n linearly independent vectors in V then, according to Examples 11.2 and 11.8, we have

$$\omega(v_1, \dots, v_n) = \det(v^j_i) \omega(e_1, \dots, e_n) = \det(v^j_i) .$$

However, from the corollary to Theorem 11.13, this is just the oriented n -volume of the n -dimensional parallelepiped in $\mathbb{R}^n$ spanned by the vectors $\{v_i\}$. Therefore, we see that an n -form in some sense represents volumes in an n -dimensional space. We now proceed to make this definition precise, beginning with a careful definition of the notion of orientation on a vector space.

In order to try and make the basic idea clear, let us first consider the space $\mathbb{R}^2$ with all possible orthogonal coordinate systems. For example, we may consider the usual “right-handed” coordinate system $\{e_1, e_2\}$ shown below, or we may consider the alternative “left-handed” system $\{e'_1, e'_2\}$ also shown.



In the first case, we see that rotating e_1 into e_2 through the smallest angle between them involves a counterclockwise rotation, while in the second case, rotating e'_1 into e'_2 entails a clockwise rotation. This effect is shown in the elementary vector cross product, where the direction of $e_1 \times e_2$ is defined by the “right-hand rule” to point out of the page, while $e'_1 \times e'_2$ points into the page.

We now ask whether or not it is possible to continuously rotate e'_1 into e_1 and e'_2 into e_2 while maintaining a basis at all times. In other words, we ask if these two bases are in some sense equivalent. Without being rigorous, it should be clear that this can not be done because there will always be one point where the vectors e'_1 and e'_2 will be co-linear, and hence linearly dependent. This observation suggests that we consider the determinant of the matrix representing this change of basis.

In order to formulate this idea precisely, let us take a look at the matrix relating our two bases $\{e_i\}$ and $\{e'_i\}$ for $\mathbb{R}^2$. We thus write $e'_i = e_j a^j_i$ and investigate the determinant $\det(a^j_i)$. From the above figure, we see that

$$\begin{aligned} e'_1 &= e_1 a^1_1 + e_2 a^2_1 & \text{where } a^1_1 < 0 \text{ and } a^2_1 > 0 \\ e'_2 &= e_1 a^1_2 + e_2 a^2_2 & \text{where } a^1_2 < 0 \text{ and } a^2_2 > 0 \end{aligned}$$

and hence $\det(a^j_i) = a^1_1 a^2_2 - a^1_2 a^2_1 < 0$.

Now suppose that we view this transformation as a continuous modification of the identity transformation. This means we consider the basis vectors e'_i to be continuous functions $e'_i(t)$ of the matrix $a^j_i(t)$ for $0 \leq t \leq 1$ where $a^j_i(0) = \delta^j_i$ and $a^j_i(1) = a^j_i$, so that $e'_i(0) = e_i$ and $e'_i(1) = e'_i$. In other words, we write $e'_i(t) = e_j a^j_i(t)$ for $0 \leq t \leq 1$. Now note that $\det(a^j_i(0)) = \det(\delta^j_i) = 1 > 0$, while $\det(a^j_i(1)) = \det(a^j_i) < 0$. Therefore, since the determinant is a continuous function of its entries, there must be some value $t_0 \in (0, 1)$ where $\det(a^j_i(t_0)) = 0$. It then follows that the vectors $e'_i(t_0)$ will be linearly dependent.

What we have just shown is that if we start with any pair of linearly independent vectors, and then transform this pair into another pair of linearly independent vectors by moving along any continuous path of linear transformations that always maintains the linear independence of the pair, then every linear transformation along this path must have positive determinant. Another way of saying this is that if we have two bases that are related by a transformation with negative determinant, then it is impossible to continuously transform one into the other while maintaining their independence. This argument clearly applies to $\mathbb{R}^n$ and is not restricted to $\mathbb{R}^2$.

Conversely, suppose we had assumed that $e'_i = e_j a^j_i$, but this time with $\det(a^j_i) > 0$. We want to show that $\{e_i\}$ may be continuously transformed into $\{e'_i\}$ while maintaining linear independence all the way. We first assume that both $\{e_i\}$ and $\{e'_i\}$ are orthonormal bases. After treating this special case, we will show how to take care of arbitrary bases.

(Unfortunately, the argument we are about to give relies on the topological concept of path connectedness. Since a complete discussion of this topic would lead us much too far astray, we shall be content to present only the fundamental concepts in Appendix C. Besides, this discussion is only motivation, and the reader should not get too bogged down in the details of this argument. Those readers who know some topology should have no trouble filling in the necessary details if desired.)

Since $\{e_i\}$ and $\{e'_i\}$ are orthonormal, it follows from Theorem 10.6 (applied to $\mathbb{R}$ rather than $\mathbb{C}$) that the transformation matrix $A = (a^j_i)$ defined by $e'_i = e_j a^j_i$ must be orthogonal, and hence $\det A = +1$ (by Theorem 10.8(a) and

the fact that we are assuming $\{e_i\}$ and $\{e'_i\}$ are related by a transformation with positive determinant). By Theorem 10.19, there exists a nonsingular matrix S such that $S^{-1}AS = M_\theta$ where M_θ is the block diagonal canonical form consisting of $+1$'s, -1 's, and 2×2 rotation matrices $R(\theta_i)$ given by

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

It is important to realize that if there are more than two $+1$'s or more than two -1 's, then each pair may be combined into one of the $R(\theta_i)$ by choosing either $\theta_i = \pi$ (for each pair of -1 's) or $\theta_i = 0$ (for each pair of $+1$'s). In this manner, we view M_θ as consisting entirely of 2×2 rotation matrices, and at most a single $+1$ and/or -1 . Since $\det R(\theta_i) = +1$ for any θ_i , we see that (using Theorem 4.14) $\det M_\theta = +1$ if there is no -1 , and $\det M_\theta = -1$ if there is a single -1 . From $A = SM_\theta S^{-1}$, we see that $\det A = \det M_\theta$, and since we are requiring that $\det A > 0$, we must have the case where there is no -1 in M_θ .

Since $\cos \theta_i$ and $\sin \theta_i$ are continuous functions of $\theta_i \in [0, 2\pi)$ (where the interval $[0, 2\pi)$ is a path connected set), we note that by parametrizing each θ_i by $\theta_i(t) = (1 - t)\theta_i$, the matrix M_θ may be continuously connected to the identity matrix I (i.e., at $t = 1$). In other words, we consider the matrix $M_{\theta(t)}$ where $M_{\theta(0)} = M_\theta$ and $M_{\theta(1)} = I$. Hence every such M_θ (i.e., any matrix of the same form as our particular M_θ , but with a different set of θ_i 's) may be continuously connected to the identity matrix. (For those readers who know some topology, note all we have said is that the torus $[0, 2\pi) \times \cdots \times [0, 2\pi)$ is path connected, and hence so is its continuous image which is the set of all such M_θ .)

We may write the (infinite) collection of all such M_θ as $M = \{M_\theta\}$. Clearly M is a path connected set. Since $A = SM_\theta S^{-1}$ and $I = SIS^{-1}$, we see that both A and I are contained in the collection $SMS^{-1} = \{SM_\theta S^{-1}\}$. But SMS^{-1} is also path connected since it is just the continuous image of a path connected set (matrix multiplication is obviously continuous). Thus we have shown that both A and I lie in the path connected set SMS^{-1} , and hence A may be continuously connected to I . Note also that every transformation along this path has positive determinant since $\det SM_\theta S^{-1} = \det M_\theta = 1 > 0$ for every $M_\theta \in M$.

If we now take any path in SMS^{-1} that starts at I and goes to A , then applying this path to the basis $\{e_i\}$ we obtain a continuous transformation from $\{e_i\}$ to $\{e'_i\}$ with everywhere positive determinant. This completes the proof for the special case of orthonormal bases.

Now suppose that $\{v_i\}$ and $\{v'_i\}$ are arbitrary bases related by a transformation with positive determinant. Starting with the basis $\{v_i\}$, we first apply

the Gram-Schmidt process (Theorem 2.21) to $\{v_i\}$ to obtain an orthonormal basis $\{e_i\} = \{v_j b_i^j\}$. This orthonormalization process may be visualized as a sequence $v_i(t) = v_j b_i^j(t)$ (for $0 \leq t \leq 1$) of continuous scalings and rotations that always maintain linear independence such that $v_i(0) = v_i$ (i.e., $b_i^j(0) = \delta_i^j$) and $v_i(1) = e_i$ (i.e., $b_i^j(1) = b_i^j$). Hence we have a continuous transformation $b_i^j(t)$ taking $\{v_i\}$ into $\{e_i\}$ with $\det(b_i^j(t)) > 0$ (the transformation starts with $\det(b_i^j(0)) = \det I > 0$, and since the vectors are always independent, it must maintain $\det(b_i^j(t)) \neq 0$). Similarly, we may transform $\{v'_i\}$ into an orthonormal basis $\{e'_i\}$ by a continuous transformation with positive determinant. (Alternatively, it was shown in Exercise 5.4.14 that the Gram-Schmidt process is represented by an upper-triangular matrix with all positive diagonal elements, and hence its determinant is positive.) Now $\{e_i\}$ and $\{e'_i\}$ are related by an orthogonal transformation that must also have determinant equal to $+1$ because $\{v_i\}$ and $\{v'_i\}$ are related by a transformation with positive determinant, and both of the Gram-Schmidt transformations have positive determinant. This reduces the general case to the special case treated above.

With this discussion as motivation, we make the following definition. Let $\{v_1, \dots, v_n\}$ and $\{v'_1, \dots, v'_n\}$ be two ordered bases for a real vector space V , and assume that $v'_i = v_j a_i^j$. These two bases are said to be **similarly oriented** if $\det(a_i^j) > 0$, and we write this as $\{v_i\} \approx \{v'_i\}$. In other words, $\{v_i\} \approx \{v'_i\}$ if $v'_i = \phi(v_i)$ with $\det \phi > 0$. We leave it to the reader to show that this defines an equivalence relation on the set of all ordered bases for V (see Exercise 11.9.1). We denote the equivalence class of the basis $\{v_i\}$ by $[v_i]$.

It is worth pointing out that had we instead required $\det(a_i^j) < 0$, then this would not have defined an equivalence relation. This is because if (b_i^j) is another such transformation with $\det(b_i^j) < 0$, then

$$\det(a_i^j b_j^k) = \det(a_i^j) \det(b_j^k) > 0.$$

Intuitively this is quite reasonable since a combination of two reflections (each of which has negative determinant) is not another reflection.

We now define an **orientation** of V to be an equivalence class of ordered bases. The space V together with an orientation $[v_i]$ is called an **oriented vector space** $(V, [v_i])$. Since the determinant of a linear transformation that relates any two bases must be either positive or negative, we see that V has exactly two orientations. In particular, if $\{v_i\}$ is any given basis, then every other basis belonging to the equivalence class $[v_i]$ of $\{v_i\}$ will be related to $\{v_i\}$ by a transformation with positive determinant, while those bases related to $\{v_i\}$ by a transformation with negative determinant will be related to each other by a transformation with positive determinant (see Exercise 11.9.1).

Now recall we have seen that n -forms seem to be related to n -volumes in an n -dimensional space V . To precisely define this relationship, we formulate orientations in terms of n -forms. To begin with, the nonzero elements of the 1-dimensional space $\bigwedge^n(V)$ are called **volume forms** (or sometimes **volume elements**) on V . If ω_1 and ω_2 are volume forms, then ω_1 is said to be **equivalent** to ω_2 if $\omega_1 = c\omega_2$ for some real $c > 0$, and in this case we also write $\omega_1 \approx \omega_2$. Since every element of $\bigwedge^n(V)$ is related to every other element by a relationship of the form $\omega_1 = a\omega_2$ for some real a (i.e., $-\infty < a < \infty$), it is clear that this equivalence relation divides the set of all nonzero volume forms into two distinct groups (i.e., equivalence classes). We can relate any ordered basis $\{v_i\}$ for V to a specific volume form by defining

$$\omega = v^1 \wedge \cdots \wedge v^n$$

where $\{v^i\}$ is the basis dual to $\{v_i\}$. That this association is meaningful is shown in the next result.

Theorem 11.19 Let $\{v_i\}$ and $\{\bar{v}_i\}$ be bases for V , and let $\{v^i\}$ and $\{\bar{v}^i\}$ be the corresponding dual bases. Define the volume forms

$$\omega = v^1 \wedge \cdots \wedge v^n$$

and

$$\bar{\omega} = \bar{v}^1 \wedge \cdots \wedge \bar{v}^n .$$

Then $\{v_i\} \approx \{\bar{v}_i\}$ if and only if $\omega \approx \bar{\omega}$.

Proof First suppose that $\{v_i\} \approx \{\bar{v}_i\}$. Then $\bar{v}_i = \phi(v_i)$ where $\det \phi > 0$, and hence (using

$$\omega(v_1, \dots, v_n) = v^1 \wedge \cdots \wedge v^n(v_1, \dots, v_n) = 1$$

as shown in Example 11.8)

$$\begin{aligned} \omega(\bar{v}_1, \dots, \bar{v}_n) &= \omega(\phi(v_1), \dots, \phi(v_n)) \\ &= (\phi^* \omega)(v_1, \dots, v_n) \\ &= (\det \phi) \omega(v_1, \dots, v_n) \\ &= \det \phi . \end{aligned}$$

If we assume that $\omega = c\bar{\omega}$ for some $-\infty < c < \infty$, then using $\bar{\omega}(\bar{v}_1, \dots, \bar{v}_n) = 1$ we see that our result implies $c = \det \phi > 0$ and thus $\omega \approx \bar{\omega}$.

Conversely, if $\omega = c \bar{\omega}$ where $c > 0$, then assuming that $\bar{v}_i = \phi(v_i)$, the above calculation shows that $\det \phi = c > 0$, and hence $\{v_i\} \approx \{\bar{v}_i\}$. ■

What this theorem shows us is that an equivalence class of bases uniquely determines an equivalence class of volume forms and conversely. Therefore it is consistent with our earlier definitions to say that an equivalence class $[\omega]$ of volume forms on V defines an **orientation** on V , and the space V together with an orientation $[\omega]$ is called an **oriented vector space** $(V, [\omega])$. A basis $\{v_i\}$ for $(V, [\omega])$ is now said to be **positively oriented** if $\omega(v_1, \dots, v_n) > 0$. Not surprisingly, the equivalence class $[-\omega]$ is called the **reverse orientation**, and the basis $\{v_i\}$ is said to be **negatively oriented** if $\omega(v_1, \dots, v_n) < 0$. Note that if the ordered basis $\{v_1, v_2, \dots, v_n\}$ is negatively oriented, then the basis $\{v_2, v_1, \dots, v_n\}$ will be positively oriented because $\omega(v_2, v_1, \dots, v_n) = -\omega(v_1, v_2, \dots, v_n) > 0$. By way of additional terminology, the **standard orientation** on $\mathbb{R}^n$ is that orientation defined by either the standard ordered basis $\{e_1, \dots, e_n\}$, or the corresponding volume form $e^1 \wedge \dots \wedge e^n$.

In order to proceed any further, we must introduce the notion of a metric on V . This is the subject of the next section.

Exercises

- Show that the collection of all similarly oriented bases for V defines an equivalence relation on the set of all ordered bases for V .
 - Let $\{v_i\}$ be a basis for V . Show that all other bases related to $\{v_i\}$ by a transformation with negative determinant will be related to each other by a transformation with positive determinant.
- Let (U, ω) and (V, μ) be oriented vector spaces with chosen volume elements. We say that $\phi \in L(U, V)$ is **volume preserving** if $\phi^*\mu = \omega$. If $\dim U = \dim V$ is finite, show that ϕ is an isomorphism.
- Let $(U, [\omega])$ and $(V, [\mu])$ be oriented vector spaces. We say that $\phi \in L(U, V)$ is **orientation preserving** if $\phi^*\mu \in [\omega]$. If $\dim U = \dim V$ is finite, show that ϕ is an isomorphism. If $U = V = \mathbb{R}^3$, give an example of a linear transformation that is orientation preserving but not volume preserving.

11.10 THE METRIC TENSOR AND VOLUME FORMS

We now generalize slightly our definition of inner products on V . In particular, recall from Section 2.4 (and the beginning of Section 9.2) that property (IP3) of an inner product requires that $\langle u, u \rangle \geq 0$ for all $u \in V$ and $\langle u, u \rangle = 0$ if and only if $u = 0$. If we drop this condition entirely, then we obtain an **indefinite inner product** on V . (In fact, some authors define an inner product as obeying only (IP1) and (IP2), and then refer to what we have called an inner product as a “positive definite inner product.”) If we replace (IP3) by the weaker requirement

$$(IP3') \quad \langle u, v \rangle = 0 \text{ for all } v \in V \text{ if and only if } u = 0$$

then our inner product is said to be **nondegenerate**. (Note that every example of an inner product given in this book up to now has been nondegenerate.) Thus a real nondegenerate indefinite inner product is just a real nondegenerate symmetric bilinear map. We will soon see an example of an inner product with the property that $\langle u, u \rangle = 0$ for some $u \neq 0$ (see Example 11.13 below).

Throughout the remainder of this chapter, we will assume that our inner products are indefinite and nondegenerate unless otherwise noted. We furthermore assume that we are dealing exclusively with real vector spaces.

Let $\{e_i\}$ be a basis for an inner product space V . Since in general we will not have $\langle e_i, e_j \rangle = \delta_{ij}$, we define the scalars g_{ij} by

$$g_{ij} = \langle e_i, e_j \rangle .$$

In terms of the g_{ij} , we have for any $X, Y \in V$

$$\langle X, Y \rangle = \langle x^i e_i, y^j e_j \rangle = x^i y^j \langle e_i, e_j \rangle = g_{ij} x^i y^j .$$

If $\{\bar{e}_i\}$ is another basis for V , then we will have $\bar{e}_i = e_j a_i^j$ for some nonsingular transition matrix $A = (a_i^j)$. Hence, writing $\bar{g}_{ij} = \langle \bar{e}_i, \bar{e}_j \rangle$ we see that

$$\bar{g}_{ij} = \langle \bar{e}_i, \bar{e}_j \rangle = \langle e_r a_i^r, e_s a_j^s \rangle = a_i^r a_j^s \langle e_r, e_s \rangle = a_i^r a_j^s g_{rs}$$

which shows that the g_{ij} transform like the components of a second-rank covariant tensor. Indeed, defining the tensor $g \in \mathcal{T}_2(V)$ by

$$g(X, Y) = \langle X, Y \rangle$$

results in

$$g(e_i, e_j) = \langle e_i, e_j \rangle = g_{ij}$$

as it should. We are therefore justified in defining the **(covariant) metric tensor**

$$g = g_{ij}\omega^i \otimes \omega^j \in \mathcal{T}_2(V)$$

(where $\{\omega^i\}$ is the basis dual to $\{e_i\}$) by $g(X, Y) = \langle X, Y \rangle$. In fact, since the inner product is nondegenerate and symmetric (i.e., $\langle X, Y \rangle = \langle Y, X \rangle$), we see that g is a nondegenerate symmetric tensor (i.e., $g_{ij} = g_{ji}$).

Next, we notice that given any vector $A \in V$, we may define a linear functional $\langle A, \cdot \rangle$ on V by the assignment $B \mapsto \langle A, B \rangle$. In other words, for any $A \in V$, we associate the 1-form α defined by $\alpha(B) = \langle A, B \rangle$ for every $B \in V$. Note that the kernel of the mapping $A \mapsto \langle A, \cdot \rangle$ (which is easily seen to be a vector space homomorphism) consists of only the zero vector (since $\langle A, B \rangle = 0$ for every $B \in V$ implies that $A = 0$), and hence this association is an isomorphism. Given any basis $\{e_i\}$ for V , the components a_i of $\alpha \in V^*$ are given in terms of those of $A = a^i e_i \in V$ by

$$a_i = \alpha(e_i) = \langle A, e_i \rangle = \langle a^j e_j, e_i \rangle = a^j \langle e_j, e_i \rangle = a^j g_{ji}$$

Thus, to any contravariant vector $A = a^i e_i \in V$, we can associate a unique covariant vector $\alpha \in V^*$ by

$$\alpha = a_i \omega^i = (a^j g_{ji}) \omega^i$$

where $\{\omega^i\}$ is the basis for V^* dual to the basis $\{e_i\}$ for V . In other words, we write

$$a_i = a^j g_{ji}$$

and we say that a_i arises by **lowering the index** j of a^j .

Example 11.12 If we consider the space $\mathbb{R}^n$ with a Cartesian coordinate system $\{e_i\}$, then we have $g_{ij} = \langle e_i, e_j \rangle = \delta_{ij}$, and hence $a_i = \delta_{ij} a^j = a^i$. Therefore, *in a Cartesian coordinate system*, there is no distinction between the components of covariant and contravariant vectors. This explains why 1-forms never arise in elementary treatments of vector analysis. //

Since the metric tensor is nondegenerate, the matrix (g_{ij}) must be nonsingular (or else the mapping $a^j \mapsto a_i$ would not be an isomorphism). We can therefore define the inverse matrix (g^{ij}) by

$$g^{ij} g_{jk} = g_{kj} g^{ji} = \delta^i_k .$$

Using (g^{ij}) , we see that the inverse of the mapping $a^j \mapsto a_i$ is given by

$$g^{ij}a_j = a^i .$$

This is called, naturally enough, **raising an index**. We will show below that the g^{ij} do indeed form the components of a tensor.

It is worth remarking that the “tensor” $g^i_j = g^{ik}g_{kj} = \delta^i_j (= \delta_j^i)$ is unique in that it has the same components in any coordinate system. Indeed, if $\{e_i\}$ and $\{\bar{e}_i\}$ are two bases for a space V with corresponding dual bases $\{\omega^i\}$ and $\{\bar{\omega}^i\}$, then $\bar{e}_i = e_j a^j_i$ and $\bar{\omega}^j = b^j_i \omega^i = (a^{-1})^j_i \omega^i$ (see the discussion following Theorem 11.2). Therefore, if we define the tensor δ to have the same values in the first coordinate system as the Kronecker delta, then $\delta^i_j = \delta(\omega^i, e_j)$. If we now define the symbol $\bar{\delta}^i_j$ by $\bar{\delta}^i_j = \delta(\bar{\omega}^i, \bar{e}_j)$, then we see that

$$\begin{aligned} \bar{\delta}^i_j &= \delta(\bar{\omega}^i, \bar{e}_j) = \delta((a^{-1})^i_k \omega^k, e_r a^r_j) = (a^{-1})^i_k a^r_j \delta(\omega^k, e_r) \\ &= (a^{-1})^i_k a^r_j \delta^k_r = (a^{-1})^i_k a^k_j = \delta^i_j . \end{aligned}$$

This shows that the δ^i_j are in fact the components of a tensor, and that these components are the same in any coordinate system.

We would now like to show that the scalars g^{ij} are indeed the components of a tensor. There are several ways that this can be done. First, let us write $g_{ij}g^{jk} = \delta^k_i$ where we know that both g_{ij} and δ^k_i are tensors. Multiplying both sides of this equation by $(a^{-1})^r_k a^i_s$ and using $(a^{-1})^r_k a^i_s \delta^k_i = \delta^r_s$ we find

$$g_{ij}g^{jk}(a^{-1})^r_k a^i_s = \delta^r_s .$$

Now substitute $g_{ij} = g_{it}\delta^t_j = g_{it}a^t_q(a^{-1})^q_j$ to obtain

$$[a^i_s a^t_q g_{it}][(a^{-1})^q_j (a^{-1})^r_k g^{jk}] = \delta^r_s .$$

Since g_{it} is a tensor, we know that $a^i_s a^t_q g_{it} = \bar{g}_{sq}$. If we write

$$\bar{g}^{qr} = (a^{-1})^q_j (a^{-1})^r_k g^{jk}$$

then we will have defined the g^{jk} to transform as the components of a tensor, and furthermore, they have the requisite property that $\bar{g}_{sq} \bar{g}^{qr} = \delta^r_s$. Therefore we have defined the (**contravariant**) **metric tensor** $G \in \mathcal{T}_0^2(V)$ by

$$G = g^{ij}e_i \otimes e_j$$

where $g^{ij}g_{jk} = \delta^i_k$.

There is another interesting way for us to define the tensor G . We have already seen that a vector $A = a^i e_i \in V$ defines a unique linear form $\alpha = a_j \omega^j \in V^*$ by the association $\alpha = g_{ij} a^i \omega^j$. If we denote the inverse of the matrix (g_{ij}) by (g^{ij}) so that $g^{ij}g_{jk} = \delta^i_k$, then to any linear form $\alpha = a_i \omega^i \in V^*$ there corresponds a unique vector $A = a^i e_i \in V$ defined by $A = g^{ij} a_j e_i$. We can now use this isomorphism to define an inner product on V^* . In other words, if $\langle , \rangle$ is an inner product on V , we define an inner product $\langle , \rangle$ on V^* by

$$\langle \alpha, \beta \rangle = \langle A, B \rangle$$

where $A, B \in V$ are the vectors corresponding to the 1-forms $\alpha, \beta \in V^*$.

Let us write an arbitrary basis vector e_i in terms of its components relative to the basis $\{\omega^j\}$ as $e_i = \delta^j_i \omega^j$. Therefore, in the above isomorphism, we may define the linear form $\hat{e}_i \in V^*$ corresponding to the basis vector e_i by

$$\hat{e}_i = g_{jk} \delta^j_i \omega^k = g_{ik} \omega^k$$

and hence using the inverse matrix, we find that

$$\omega^k = g^{ki} \hat{e}_i .$$

Applying our definition of the inner product in V^* we have $\langle \hat{e}_i, \hat{e}_j \rangle = \langle e_i, e_j \rangle = g_{ij}$, and therefore we obtain

$$\langle \omega^i, \omega^j \rangle = \langle g^{ir} \hat{e}_r, g^{js} \hat{e}_s \rangle = g^{ir} g^{js} \langle \hat{e}_r, \hat{e}_s \rangle = g^{ir} g^{js} g_{rs} = g^{ir} \delta^j_r = g^{ij}$$

which is the analogue in V^* of the definition $g_{ij} = \langle e_i, e_j \rangle$ in V .

Lastly, since $\bar{\omega}^j = (a^{-1})^j_i \omega^i$, we see that

$$\begin{aligned} \bar{g}^{ij} &= \langle \bar{\omega}^i, \bar{\omega}^j \rangle = \langle (a^{-1})^i_r \omega^r, (a^{-1})^j_s \omega^s \rangle = (a^{-1})^i_r (a^{-1})^j_s \langle \omega^r, \omega^s \rangle \\ &= (a^{-1})^i_r (a^{-1})^j_s g^{rs} \end{aligned}$$

so the scalars g^{ij} may be considered to be the components of a symmetric tensor $G \in \mathcal{T}_2^0(V)$ defined as above by $G = g^{ij}e_i \otimes e_j$.

Now let $g = \langle , \rangle$ be an arbitrary (i.e., possibly degenerate) real symmetric bilinear form on the inner product space V . It follows from the corollary to Theorem 9.14 that there exists a basis $\{e_i\}$ for V in which the matrix (g_{ij}) of g takes the unique diagonal form

$$g_{ij} = \begin{pmatrix} I_r & & \\ & -I_s & \\ & & 0_t \end{pmatrix}$$

where $r + s + t = \dim V = n$. Thus

$$g(e_i, e_i) = \begin{cases} 1 & \text{for } 1 \leq i \leq r \\ -1 & \text{for } r+1 \leq i \leq r+s \\ 0 & \text{for } r+s+1 \leq i \leq n \end{cases}.$$

If $r + s < n$, the inner product is degenerate and we say that the space V is **singular** (with respect to the given inner product). If $r + s = n$, then the inner product is nondegenerate, and the basis $\{e_i\}$ is orthonormal. In the orthonormal case, if either $r = 0$ or $r = n$, the space is said to be **ordinary Euclidean**, and if $0 < r < n$, then the space is called **pseudo-Euclidean**. Recall that the number $r - s = r - (n - r) = 2r - n$ is called the **signature** of g (which is therefore just the trace of (g_{ij})). Moreover, the number of -1 's is called the **index** of g , and is denoted by $\text{Ind}(g)$. If $g = \langle \cdot, \cdot \rangle$ is to be a metric on V , then by definition, we must have $r + s = n$ so that the inner product is nondegenerate. In this case, the basis $\{e_i\}$ is called **g-orthonormal**.

Example 11.13 If the metric g represents a positive definite inner product on V , then we must have $\text{Ind}(g) = 0$, and such a metric is said to be **Riemannian**. Alternatively, another well-known metric is the Lorentz metric used in the theory of special relativity. By definition, a **Lorentz** metric η has $\text{Ind}(\eta) = 1$. Therefore, if η is a Lorentz metric, an η -orthonormal basis $\{e_1, \dots, e_n\}$ ordered in such a way that $\eta(e_i, e_i) = +1$ for $i = 1, \dots, n-1$ and $\eta(e_n, e_n) = -1$ is called a **Lorentz frame**.

Thus, in terms of a g -orthonormal basis, a Riemannian metric has the form

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

while in a Lorentz frame, a Lorentz metric takes the form

$$(\eta_{ij}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}.$$

It is worth remarking that a Lorentz metric is also frequently defined as having $\text{Ind}(\eta) = n - 1$. In this case we have $\eta(e_1, e_1) = 1$ and $\eta(e_i, e_i) = -1$ for each $i = 2, \dots, n$. We also point out that a vector $v \in V$ is called **timelike** if $\eta(v, v) < 0$, **lightlike** (or **null**) if $\eta(v, v) = 0$, and **spacelike** if $\eta(v, v) > 0$. Note that a Lorentz inner product is clearly indefinite since, for example, the nonzero vector v with components $v = (0, 0, 1, 1)$ has the property that $\langle v, v \rangle = \eta(v, v) = 0$. //

We now show that introducing a metric on V leads to a unique volume form on V .

Theorem 11.20 Let g be a metric on an n -dimensional oriented vector space $(V, [\omega])$. Then, corresponding to the metric g , there exists a unique volume form $\mu = \mu(g) \in [\omega]$ such that $\mu(e_1, \dots, e_n) = 1$ for every positively oriented g -orthonormal basis $\{e_i\}$ for V . Moreover, if $\{v_i\}$ is any (not necessarily g -orthonormal) positively oriented basis for V with dual basis $\{v^i\}$, then

$$\mu = |\det(g(v_i, v_j))|^{1/2} v^1 \wedge \dots \wedge v^n .$$

In particular, if $\{v_i\} = \{e_i\}$ is a g -orthonormal basis, then $\mu = e^1 \wedge \dots \wedge e^n$.

Proof Since $\omega \neq 0$, there exists a positively oriented g -orthonormal basis $\{e_i\}$ such that $\omega(e_1, \dots, e_n) > 0$ (we can multiply e_1 by -1 if necessary in order that $\{e_i\}$ be positively oriented). We now define $\mu \in [\omega]$ by

$$\mu(e_1, \dots, e_n) = 1 .$$

That this defines a unique μ follows by multilinearity. We claim that if $\{f_i\}$ is any other positively oriented g -orthonormal basis, then $\mu(f_1, \dots, f_n) = 1$ also. To show this, we first prove a simple general result.

Suppose $\{v_i\}$ is any other basis for V related to the g -orthonormal basis $\{e_i\}$ by $v_i = \phi(e_i) = e_j a^j_i$ where, by Theorem 11.17, we have $\det \phi = \det(a^j_i)$. We then have $g(v_i, v_j) = a^r_i a^s_j g(e_r, e_s)$ which in matrix notation is $[g]_v = A^T [g]_e A$, and hence

$$\det(g(v_i, v_j)) = (\det \phi)^2 \det(g(e_r, e_s)) . \quad (9)$$

However, since $\{e_i\}$ is g -orthonormal we have $g(e_r, e_s) = \pm \delta_{rs}$, and therefore $|\det(g(e_r, e_s))| = 1$. In other words

$$|\det(g(v_i, v_j))|^{1/2} = |\det \phi| . \quad (10)$$

Returning to our problem, we have $\det(g(f_i, f_j)) = \pm 1$ also since $\{f_i\} = \{\phi(e_i)\}$ is g -orthonormal. Thus (10) implies that $|\det \phi| = 1$. But $\{f_i\}$ is positively oriented so that $\mu(f_1, \dots, f_n) > 0$ by definition. Therefore

$$\begin{aligned} 0 < \mu(f_1, \dots, f_n) &= \mu(\phi(e_1), \dots, \mu(\phi(e_n))) = (\phi^* \mu)(e_1, \dots, e_n) \\ &= (\det \phi) \mu(e_1, \dots, e_n) = \det \phi \end{aligned}$$

so that we must in fact have $\det \phi = +1$. In other words, $\mu(f_1, \dots, f_n) = 1$ as claimed.

Now suppose that $\{v_i\}$ is an arbitrary positively oriented basis for V such that $v_i = \phi(e_i)$. Then, analogously to what we have just shown, we see that $\mu(v_1, \dots, v_n) = \det \phi > 0$. Hence (10) shows that (using Example 11.8)

$$\begin{aligned} \mu(v_1, \dots, v_n) &= \det \phi \\ &= |\det(g(v_i, v_j))|^{1/2} \\ &= |\det(g(v_i, v_j))|^{1/2} v^1 \wedge \dots \wedge v^n(v_1, \dots, v_n) \end{aligned}$$

which implies

$$\mu = |\det(g(v_i, v_j))|^{1/2} v^1 \wedge \dots \wedge v^n . \blacksquare$$

The unique volume form μ defined in Theorem 11.20 is called the **g-volume**, or sometimes the **metric volume form**. A common (although rather careless) notation is to write $|\det(g(v_i, v_j))|^{1/2} = \sqrt{|g|}$. In this notation, the g -volume is written as

$$\sqrt{|g|} v^1 \wedge \dots \wedge v^n$$

where $\{v_1, \dots, v_n\}$ must be positively oriented. If the basis $\{v_1, v_2, \dots, v_n\}$ is negatively oriented, then clearly $\{v_2, v_1, \dots, v_n\}$ will be positively oriented. Furthermore, even though the matrix of g relative to each of these oriented bases will be different, the determinant actually remains unchanged (see the discussion following the corollary to Theorem 11.13). Therefore, for this negatively oriented basis, the g -volume is

$$\sqrt{|g|} v^2 \wedge v^1 \wedge \dots \wedge v^n = -\sqrt{|g|} v^1 \wedge v^2 \wedge \dots \wedge v^n .$$

We thus have the following corollary to Theorem 11.20.

Corollary Let $\{v_i\}$ be any basis for the n -dimensional oriented vector space $(V, [\omega])$ with metric g . Then the g -volume form on V is given by

$$\pm \sqrt{|g|} v^1 \wedge \dots \wedge v^n$$

where the “+” sign is for $\{v_i\}$ positively oriented, and the “−” sign is for $\{v_i\}$ negatively oriented.

Example 11.14 From Example 11.13, we see that for a Riemannian metric g and g -orthonormal basis $\{e_i\}$ we have $\det(g(e_i, e_j)) = +1$. Hence, from equation (9), we see that $\det(g(v_i, v_j)) > 0$ for any basis $\{v_i = \phi(e_i)\}$. Thus the g -volume form on a Riemannian space is given by $\pm\sqrt{g} v^1 \wedge \cdots \wedge v^n$.

For a Lorentz metric we have $\det(g(e_i, e_j)) = -1$ in a Lorentz frame, and therefore $\det(g(v_i, v_j)) < 0$ in an arbitrary frame. Thus the g -volume in a Lorentz space is given by $\pm\sqrt{-g} v^1 \wedge \cdots \wedge v^n$.

Let us point out that had we defined $\text{Ind}(\eta) = n - 1$ instead of $\text{Ind}(\eta) = 1$, then $\det(\eta(e_i, e_j)) < 0$ only in an even dimensional space. In this case, we would have to write the g -volume as in the above corollary. //

Example 11.15 (This example is a continuation of Example 11.11.) We remark that these volume elements are of great practical use in the theory of integration on manifolds. To see an example of how this is done, let us use Examples 11.1 and 11.11 to write the volume element as (remember that this applies only locally, and hence the metric depends on the coordinates)

$$d\tau = \sqrt{|g|} dx^1 \wedge \cdots \wedge dx^n .$$

If we go to a new coordinate system $\{\bar{x}^i\}$, then

$$\bar{g}_{ij} = \frac{\partial x^r}{\partial \bar{x}^i} \frac{\partial x^s}{\partial \bar{x}^j} g_{rs}$$

so that $|\bar{g}| = (J^{-1})^2 |g|$ where $J^{-1} = \det(\partial x^r / \partial \bar{x}^i)$ is the determinant of the inverse Jacobian matrix of the transformation. But using $d\bar{x}^i = (\partial \bar{x}^i / \partial x^j) dx^j$ and the properties of the wedge product, it is easy to see that

$$\begin{aligned} d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n &= \frac{\partial \bar{x}^1}{\partial x^{i_1}} \cdots \frac{\partial \bar{x}^n}{\partial x^{i_n}} dx^{i_1} \wedge \cdots \wedge dx^{i_n} \\ &= \det \left(\frac{\partial \bar{x}^i}{\partial x^j} \right) dx^1 \wedge \cdots \wedge dx^n \end{aligned}$$

and hence

$$d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n = J dx^1 \wedge \cdots \wedge dx^n$$

where J is the determinant of the Jacobian matrix. (Note that the proper transformation formula for the volume element in multiple integrals arises naturally in the algebra of exterior forms.) We now have

$$\begin{aligned} d\bar{\tau} &= \sqrt{|g|} d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n = J^{-1} \sqrt{|g|} J dx^1 \wedge \cdots \wedge dx^n \\ &= \sqrt{|g|} dx^1 \wedge \cdots \wedge dx^n = d\tau \end{aligned}$$

and hence $d\tau$ is a scalar called the **invariant volume element**. In the case of $\mathbb{R}^4$ as a Lorentz space, this result is used in the theory of relativity. //

Exercises

1. Suppose V has a metric g_{ij} defined on it. Show that for any $A, B \in V$ we have $\langle A, B \rangle = a_i b^i = a^i b_i$.
2. According to the special theory of relativity, the speed of light is the same for all unaccelerated observers regardless of the motion of the source of light relative to the observer. Consider two observers moving at a constant velocity β with respect to each other, and assume that the origins of their respective coordinate systems coincide at $t = 0$. If a spherical pulse of light is emitted from the origin at $t = 0$, then (in units where the speed of light is equal to 1) this pulse satisfies the equation $x^2 + y^2 + z^2 - t^2 = 0$ for the first observer, and $\bar{x}^2 + \bar{y}^2 + \bar{z}^2 - \bar{t}^2 = 0$ for the second observer. We shall use the common notation $(t, x, y, z) = (x^0, x^1, x^2, x^3)$ for our coordinates, and hence the Lorentz metric takes the form

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

where $0 \leq \mu, \nu \leq 3$.

- (a) Let the Lorentz transformation matrix be Λ so that $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$. Show that the Lorentz transformation must satisfy $\Lambda^T \eta \Lambda = \eta$.
- (b) If the $\{\bar{x}^\mu\}$ system moves along the x^1 -axis with velocity β , then it turns out that the Lorentz transformation is given by

$$\begin{aligned} \bar{x}^0 &= \gamma(x^0 - \beta x^1) \\ \bar{x}^1 &= \gamma(x^1 - \beta x^0) \end{aligned}$$

$$\begin{aligned}\bar{x}^2 &= x^2 \\ \bar{x}^3 &= x^3\end{aligned}$$

where $\gamma^2 = 1/(1 - \beta^2)$. Using $\Lambda^\mu{}_\nu = \partial\bar{x}^\mu/\partial x^\nu$, write out the matrix $(\Lambda^\mu{}_\nu)$, and verify explicitly that $\Lambda^T\eta\Lambda = \eta$.

(c) The electromagnetic field tensor is given by

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}.$$

Using this, find the components of the electric field $\vec{E}$ and magnetic field $\vec{B}$ in the $\{\bar{x}^\mu\}$ coordinate system. In other words, find $\bar{F}_{\mu\nu}$. (The actual definition of $F_{\mu\nu}$ is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ where $\partial_\mu = \partial/\partial x^\mu$ and $A_\mu = (\phi, A_1, A_2, A_3)$ is related to $\vec{E}$ and $\vec{B}$ through the classical equations $\vec{E} = -\nabla\phi - \partial\vec{A}/\partial t$ and $\vec{B} = \nabla \times \vec{A}$. See also Exercise 11.1.6.)

3. Let V be an n -dimensional vector space with a Lorentz metric η , and let W be an $(n - 1)$ -dimensional subspace of V . Note that

$$W^\perp = \{v \in V: \eta(v, w) = 0 \text{ for all } w \in W\}$$

is the 1-dimensional subspace of all normal vectors for W . We say that W is **timelike** if every normal vector is spacelike, **null** if every normal vector is null, and **spacelike** if every normal vector is timelike. Prove that η restricted to W is

- (a) Positive definite if W is spacelike.
 - (b) A Lorentz metric if W is timelike.
 - (c) Degenerate if W is null.
4. (a) Let D be a 3×3 determinant considered as a function of three contravariant vectors $A^i_{(1)}$, $A^i_{(2)}$, and $A^i_{(3)}$. Show that under a change of coordinates, D does not transform as a scalar, but that $D\sqrt{|g|}$ does transform as a proper scalar. [Hint: Use Exercise 11.2.8.]
- (b) Show that $\varepsilon_{ijk}\sqrt{|g|}$ transforms like a tensor. (This is the Levi-Civita tensor in general coordinates. Note that in a g -orthonormal coordinate system this reduces to the Levi-Civita symbol.)
- (c) What is the contravariant version of the tensor in part (b)?