APPENDIX A

REVIEW OF LINEAR ALGEBRA

The following is a summary (almost entirely without proofs) of the linear algebra needed in the projective geometry notes. All the material is standard, but there are some relatively minor differences in viewpoint compared to standard undergraduate courses and texts.

1. There is more emphasis on certain results which are particularly important in the use of linear algebra to study geometric questions.
2. In most cases, the system of scalars is taken to be as general as possible. — This has both advantages and disadvantages. Maximum generality is needed to show that systems which satisfy the short list of axioms for an affine 3-space always have somewhat reasonable coordinate systems. However, especially at the beginning, the need to consider abstract mathematical systems of scalars and their specific properties (for example, whether or not multiplication of scalars satisfies $ab = ba$) may be distracting or even worse. A reader who is only interested in scalars for which $ab = ba$ may substitute "field" for "skew-field" throughout and ignore the adjectives "left" and "right" in discussions of vector spaces. If a reader wishes to go further and only consider scalars that are real or complex numbers, this can be done by assuming that "field" refers to one or both of these systems.

More complete treatments of linear algebra appear in the appropriate books listed in the Bibliography.

1. Vector spaces

The basic operations involving ordinary 2- and 3-dimensional vectors are vector addition and scalar multiplication (in this context *scalar* means real number). We begin by listing the relevant properties of the real numbers below. For each statement, it is assumed that a, b, c are arbitrary elements of the real numbers $\mathbb{R}$.

(F-1) (*Commutative Law of Addition*). $a + b = b + a$.
(F-2) (*Associative Law of Addition*). $a + (b + c) = (a + b) + c$.
(F-3) (*Existence of Zero Element*). There is a unique $0 \in \mathbb{R}$ such that $a + 0 = a$.
(F-4) (*Existence of Negative Elements*). If $a \in \mathbb{R}$, then there is a unique $(-a) \in \mathbb{R}$ such that $a + (-a) = 0$.
(F-5) (*Left Distributive Law*). $a \cdot (b + c) = (ab) + (ac)$.
(F-6) (*Right Distributive Law*). $(a + b) \cdot c = (ac) + (bc)$.
(F-7) (*Existence of Unit Element*). There is a unique $1 \in \mathbb{R}$ such that $a \cdot 1 = 1 \cdot a = a$.
(F-8) (*Existence of Nonzero Reciprocals*). If $a \in \mathbb{R}$ is not equal to zero, then there is a unique $a^{-1} \in \mathbb{R}$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.
(F-9) (*Associative Law of Multiplication*). $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
(F-10) (*Commutative Law of Multiplication*). $a \cdot b = b \cdot a$.

We call any triple $(\mathbb{F},\ +,\ \cdot\ )$ consisting of a set with binary operations "+" and "$\cdot$" satisfying **(F-1)**–**(F-10)** a (commutative) **field**. Many undergraduate texts on linear algebra begin by allowing the scalars to belong to an arbitrary field. In fact, a great deal of the theory does not depend upon property **(F-10)**, and for the purposes of projective geometry it is especially useful to drop this assumption. Thus we shall call a triple which only satisfies the first nine assumptions a **skew-field** (such a system is also called a **division ring** in the literature). If the final property is also satisfied, then frequently we shall say that the system is **commutative**.

THE MAIN EXAMPLES. Three basic number systems for arithmetic and algebra are fields: The *real numbers*, which we denote by $\mathbb{R}$, the *complex numbers*, which we denote by $\mathbb{C}$, and the *rational numbers*, which we denote by $\mathbb{Q}$. In each case the binary operations are given by ordinary addition and multiplication. — Another important class of examples are given by the finite fields of *integers mod(ulo) p*, where p is a prime number; these fields are denoted by $\mathbb{Z}_p$.

An example of a skew-field which is **not** commutative is given by the *quaternions*, which were first described by W. R. Hamilton (1805–1865) in the 19$^{\text{th}}$ century and will be denoted by $\mathbb{K}$ in these notes.[1] These consist of all formal expressions

$$\mathbf{x} \ = \ x_0\mathbf{1} \ + \ x_1\mathbf{i} \ + \ x_2\mathbf{j} \ + \ x_3\mathbf{k}$$

with real coefficients x_m, and two such expressions are taken to be equal if and only if all four coefficients are equal. The *sum* of two quaternions $\mathbf{x} + \mathbf{y}$ is the expression whose coeffients are $x_m + y_m$ for $m = 0, 1, 2, 3$. The general formula for the *product* is much less concise, and it is probably easier to think of the product as uniquely defined by the basic assumptions — particularly the left and right distributive laws — and the following multiplication formulas:

$$\mathbf{1}\cdot\mathbf{x} \ = \ \mathbf{x}\cdot\mathbf{1} \ = \ \mathbf{x} \quad (\text{all } \mathbf{x}), \qquad \mathbf{i}^2 \ = \ \mathbf{j}^2 \ = \ \mathbf{k}^2 \ = \ -1$$

$$\mathbf{i}\cdot\mathbf{j} \ = \ -\mathbf{j}\cdot\mathbf{i} \ = \ \mathbf{k}, \quad \mathbf{j}\cdot\mathbf{k} \ = \ -\mathbf{k}\cdot\mathbf{j} \ = \ \mathbf{i}, \quad \mathbf{k}\cdot\mathbf{i} \ = \ -\mathbf{i}\cdot\mathbf{k} \ = \ \mathbf{j}$$

The verification that $\mathbb{K}$ satisfies the first seven properties is essentially a simple exercise in bookkeeping, and the verification of **(F-9)** is similar but requires more work. To prove **(F-8)**, one first defines the *conjugate* of $\mathbf{x}$ by

$$\overline{\mathbf{x}} \ = \ x_0\mathbf{1} \ - \ x_1\mathbf{i} \ - \ x_2\mathbf{j} \ - \ x_3\mathbf{k}$$

and then verifies that

$$\left(x_0^2 + x_1^2 + x_2^2 + x_3^2 \right)^{-1} \cdot \overline{\mathbf{x}}$$

satisfies the properties one wants for $\mathbf{x}^{-1}$ (more elementary but somewhat lengthy bookkeeping — note the similarity between this formula and a standard formula for the reciprocal of a nonzero complex number). Of course, the relation $\mathbf{i}\cdot\mathbf{j} \ = \ -\mathbf{j}\cdot\mathbf{i}$ shows that **(F-10)** does not hold in $\mathbb{K}$. In all the examples we have discussed, we have $\mathbf{x}\cdot\mathbf{y} \ = \ \pm\,\mathbf{y}\cdot\mathbf{x}$, but it is also possible (and not really difficult) to construct examples for which we have $\mathbf{x}\cdot\mathbf{y}$ and $\mathbf{y}\cdot\mathbf{x}$ are not real multiples of each other.

VECTOR OPERATIONS. Having described the basic properties of addition and multiplication that we shall need, the next step is to list the properties of vector addition and scalar multiplication that will be important for our purposes. In each case, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$ are arbitrary elements of the space of vectors V, and a and b denote arbitrary scalars.

(V-1) (*Commutative Law of Addition*). $\mathbf{v} \ + \ \mathbf{w} \ = \ \mathbf{w} \ + \ \mathbf{v}$.
(V-2) (*Associative Law of Addition*). $\mathbf{v} \ + \ (\mathbf{w} + \mathbf{x}) \ = \ (\mathbf{v} + \mathbf{w}) \ + \ \mathbf{x}$.

[1]The material in this paragraph may be skipped if the reader is only interested in commutative fields or the basic number systems described above.

(V-3) (*Existence of Zero Element*). There is some $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
(V-4) (*Existence of Negative Elements*). If $\mathbf{v} \in \mathbb{R}$, then there is some $(-\mathbf{v}) \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
(V-5) (*Left Distributive Law*). $a \cdot (\mathbf{v} + \mathbf{w}) = (a\mathbf{v}) + (a\mathbf{w})$.
(V-6) (*Right Distributive Law*). $(a + b) \cdot \mathbf{v} = (a\mathbf{v}) + (b\mathbf{v})$.
(V-7) (*Unit Multiplication Property*). $1 \cdot \mathbf{v} = \mathbf{v}$.
(V-8) (*Mixed Associative Law*). $a \cdot (b\mathbf{v}) = (ab) \cdot \mathbf{v}$.

We call any triple $(V, +, \cdot)$ consisting of a set with a binary operation "+" and *left scalar multiplication "$\cdot$" by elements of a skew-field* $\mathbb{F}$ satisfying **(V-1)**–**(V-8)** a **left vector space** over $\mathbb{F}$.

STANDARD EXAMPLE. Let $\mathbb{F}^n$ be the set of all ordered n-tuples of elements in $\mathbb{F}$ (if $n = 2$ or 3 and $\mathbb{F} = \mathbb{R}$ then these are ordinary vectors). Define addition and scalar multiplication by

$$(\mathbf{x}_1, \cdots, \mathbf{x}_n) + (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (\mathbf{x}_1 + \mathbf{y}_1, \cdots, \mathbf{x}_n + \mathbf{y}_n)$$

$$c \cdot (\mathbf{x}_1, \cdots, \mathbf{x}_n) = (c\,\mathbf{x}_1, \cdots, c\,\mathbf{x}_n) .$$

Verification of **(V-1)** – **(V-8)** is yet another exercise in bookkeeping.

The above definition has a "right handed" counterpart in which scalars are multiplied on the **right**. As one might expect, such systems are called right vector spaces. They all satisfy **(V-1)** – **(V-4)** and analogs of **(V-5)** – **(V-8)** in which the order of vectors and scalars in all multiplications is reversed. The correct analog of the Mixed Associative Law is $\mathbf{v} \cdot (ab) = (\mathbf{v} \cdot a) \cdot b$. We may make $\mathbb{F}^n$ into a right vector space by defining addition as before and setting

$$(\mathbf{x}_1, \cdots, \mathbf{x}_n) \cdot c = (\mathbf{x}_1 c, \cdots, \mathbf{x}_n c) .$$

If $\mathbb{F}$ is commutative, this right multiplication differs from the left multiplication only in the order of writing vectors and scalars, with both operations yielding the same vector. However, if $\mathbb{F}$ is not commutative then the two multiplications may differ considerably.

Henceforth, we shall do everything for left vector spaces almost exclusively. In all cases the corresponding right-handed results may be obtained by making a few obvious transpositions.

Vector subspaces

Definition. A nonempty subset W of a (left) vector space V is called a (left) *vector subspace* provided the following hold:

(*i*) If $\mathbf{x}$ and $\mathbf{y}$ are in W, then so is $\mathbf{x} + \mathbf{y}$.

(*ii*) If $\mathbf{x} \in W$ and $c \in \mathbb{F}$, then $c\,\mathbf{x} \in W$.

The vector space identity $0 \cdot \mathbf{v} = \mathbf{0}$ is a direct consequence of **(V-1)**–**(V-8)**, and thus it follows that $\mathbf{0} \in W$. Furthermore, it follows immediately that W also satisfies **(V-1)**–**(V-8)**, and therefore W is a vector space in its own right with respect to the addition and scalar multiplication operations it inherits from V.

Theorem A.1. *Let V be a vector space, and let W_α be a family of subspaces of V, where α runs over all elements of some indexing set A. Then the intersection*

$$\bigcap_{\alpha\in A} W_\alpha$$

is also a vector subspace of V.

Theorem A.2. *Let V be a vector space, and let W_1 and W_2 be vector subspaces of V. Then the set $W_1 + W_2$ given by*

$$\{\ \mathbf{x}\in V \mid \mathbf{x} = \mathbf{v}_1 + \mathbf{v}_2,\ \text{where } \mathbf{x}_1\in W_1 \text{ and } \mathbf{x}_2\in W_2\ \}\ .$$

is a subspace of V.

Besides addition and multiplication, the real numbers have a concept of *order*; given two unequal real numbers a and b, we know that one is greater than the other. Since $a < b$ if and only if $b - a$ is *positive*, it follows that the ordering is determined by the set of positive elements $\mathbb{R}^+$. This set has the following properties:

(O-1) If $x,\ y\in\mathbb{R}^+$, then $x+y,\ xy\in\mathbb{R}^+$.

(O-2) If $x\in\mathbb{R}$, then **exactly one** of the statements $x\in\mathbb{R}^+$, $x=0$, and $-x\in\mathbb{R}^+$ is true.

Given a skew-field $\mathbb{F}$ and a nonempty set of elements $\mathbb{F}^+$ satisfying these two properties, we shall call the quadruple $(\mathbb{F},+,\cdot,\mathbb{F}^+)$ an *ordered skew-field*, and we call $\mathbb{F}^+$ its set of *positive elements*. The basic properties of ordered fields in the commutative case are presented in Sections I.3 and II.4 of Birkhoff and MacLane. One additional property suffices to give a complete characterization of the real numbers. Since its presentation is a bit lengthy, the reader is referred to Chapter IV of Birkhoff and MacLane for further information.

2. Dimension

Definition. Let V be a vector space over $\mathbb{F}$, and let $A\subset V$ be a nonempty subset. A vector $\mathbf{x}\in V$ is said to be a *linear combination of the vectors in A* if there exist $\mathbf{v}_1,\ \cdots\ ,\mathbf{x}_n\in A$ and $c_1,\ \cdots\ ,c_n\in\mathbb{F}$ such that $\mathbf{x}=\sum_i\ c_i\mathbf{x}_i$. We say that A is a *spanning set for V* (or the vectors in A *span the vector space V*) if every vector in V is a linear combination of the vectors in A. In general, if $A\subset V$ then $\mathsf{Span}(A)$ is defined to be the set of all linear combinations of vectors in A. It is immediate that $\mathsf{Span}(A)$ is a subspace of V.

EXAMPLE. If $\mathbf{e}_i\in\mathbb{F}^n$ is the n-tuple whose i^{th} entry is 1 and whose other entries are zero, then the vectors in $\{\,\mathbf{e}_1,\ \cdots\ ,\mathbf{e}_n\,\}$ span $\mathbb{F}^n$ because $\mathbf{x}=(x_1,\ \cdots\ ,x_n)\in\mathbb{F}^n$ implies

$$\mathbf{x}\ =\ \sum_{i=1}^{n}\ x_i\mathbf{e}_i\ .$$

Definition. A nonempty set of vectors A with $A\neq\{\,\mathbf{0}\,\}$ is *linearly independent* if for every $\mathbf{x}\in A$ the vectors $\mathbf{x}$ is not a linear combination of vectors in $A-\{\,\mathbf{x}\,\}$. A set is *linearly dependent* if it is not linearly independent. Thus the set $\{\,\mathbf{0}\,\}$ is linearly dependent by convention.

EXAMPLE. The set $\{\,\mathbf{e}_1,\ \cdots\ ,\mathbf{e}_n\,\}$ is linearly independent; for each i, the i^{th} coordinate of $\mathbf{e}_i$ is equal to 1, but every linear combination the remaining vectors has an i^{th} coordinate equal to zero.

The next result is not always emphasized in linear algebra texts or courses, but the proof is fairly straightforward and the conclusion is important for geometrical purposes.

Theorem A.3. *Let $A \subset V$ is linearly independent and $\mathbf{y} \notin \mathsf{Span}(A)$. Then $A \cup \{\,\mathbf{y}\,\}$ is linearly independent.*

Theorem A.4. *The set $A \subset V$ is linearly indepnedent if and only if every nonzero vector in $\mathsf{Span}(A)$ has an essentially unique expression as a linear combination as a linear combination of the vectors in A.*

More precisely, the condition in Theorem A.4 means that if the expressions

$$\sum_{i=1}^{r} c_i\mathbf{v}_i \qquad\qquad \sum_{j=1}^{s} d_j\mathbf{w}_j$$

(with all $c_i,\ d_j \neq 0$) yield the same vector, then $r = s$, the ordered lists of vectors $\mathbf{v}_1,\ \cdots\ ,\mathbf{v}_r$ and $\mathbf{w}_1,\ \cdots\ ,\mathbf{w}_r$ are rearrangements of each other, and if $j(1),\ \cdots\ ,j(r)$ is the rearrangement such that $\mathbf{w}_{j(i)} = \mathbf{v}_i$ for all i, then the corresponding coefficients are equal, so that $d_{j(i)} = c_i$.[2]

Remark. Actually, a set of vectors is linearly independent if and only if the zero vector has only one expression as a linear combination of vectors in the set; namely, the expression in which all the coefficients are zero.

Definition. A *basis* for V is a subset A which is linearly independent and spans V. If $V = \mathbb{F}^n$, the preceding discussion shows that the set of unit vectors $\{\,\mathbf{e}_1,\ \cdots\ ,\mathbf{e}_n\,\}$ is a basis.

Theorem A.5. *Suppose that V is a vector space having a basis with n elements, where n is some positive integer. Then the following hold:*

(*i*) *Every other basis for V has exactly n elements.*

(*ii*) *Every subset of V containing $(n+1)$ vectors is linearly dependent.*

(*iii*) *Every linearly independent subset $A \subset V$ is contained in a subset $B \subset V$ which is a basis.*

Definition. If V has a basis with n (finitely many) elements, then V is said to be *finite-dimensional* and n is called the *dimension* of V, written $\dim V$. By the preceding theorem, the number n is independent of the choice of basis. If V has no finite basis, we shall say that V is *infinite-dimensional*.

The following theorems are a natural continuation of Theorem A.5.

---

[2]The reason for the qualifying term "essentially" reflects the lack of a finiteness assumption on the spanning set A. In our setting, only finite sums are meaningful, and if A is infinite then any finite linear combination $\sum_i\ c_i\mathbf{a}_i$ can be rewritten as $\sum_i\ c_i\mathbf{a}_i\ +\ \sum_j\ 0\,\mathbf{b}_j$, where $\mathbf{b}_1,\ \cdots\ ,\mathbf{b}_q$ is a subset of A which is disjoint from $\mathbf{a}_1,\ \cdots\ ,\mathbf{a}_n$. Of course, these two expressions are essentially the same even if they might not be completely identical.

Theorem A.6. *Suppose that V has a spanning set with finitely many (say n) elements. Then V is finite-dimensional and $\dim V \leq n$. In fact, given a finite spanning set A, there is a subset $B \subset A$ which is a basis for V.*

Theorem A.7. *If W is a vector subspace of V and V is finite-dimensional, then so is W and $\dim W \leq \dim V$. In fact, every basis for W is contained in a basis for V.*

Theorem A.8. *Suppose that W_1 and W_2 are subspaces of the finite-dimensional vector space V such that $W_1 \subset W_2$ and $\dim W_1 = \dim W_2$. Then $W_1 = W_2$.*

Finally, we have the following important formula relating the discussions of the dimensions of two subspaces with those of their intersection and sum (see Theorems A.1 – A.2).

Theorem A.9. *If V is a vector space and W_1 and W_2 are finite-dimensional subspaces, then $W_1 \cap W_2$ and $W_1 + W_2$ are also finite-dimensional and*

$$\dim W_1 \ + \ \dim W_2 \ = \ \dim(W_1 \cap W_2) \ + \ \dim(W_1 + W_2) \ .$$

3. Systems of linear equations

Since we are considering skew-fields in which multiplication need not be commutative, we shall need to distinguish between the left and right scalar multiplications in this section. We remark that the entire discussion remains valid if one interchanges right and left and reverses the order of factors occurring in all products.

Suppose we are given a system of m homogeneous linear equations in n unknowns:

$$\begin{aligned} a_{1,1}x_1 \ + \ \cdots \ + \ a_{1,n}x_n \ &= \ 0 \\ a_{2,1}x_1 \ + \ \cdots \ + \ a_{2,n}x_n \ &= \ 0 \\ \cdots & \\ \cdots & \\ a_{m,1}x_1 \ + \ \cdots \ + \ a_{m,n}x_n \ &= \ 0 \end{aligned}$$

The *solution space* of this system is defined to be the set of all $\mathbf{x} = (x_1, \ \cdots \ , x_n) \in \mathbb{F}^n$ for which the equations above are satisfied. It is easy to see that this is a *right* subspace of $\mathbb{F}^n$. We shall describe how the coefficients $a_{i,j}$ determine its dimension.

Recall that an $m \times n$ matrix with entries in a set Y (or over a set Y) is a rectangular array of indexed elements in Y with m rows and n columns. We may view the rows as elements of the n-fold product Y^n.

Clearly the coefficients of the given system of linear equations describe and $m \times n$ matrix over $\mathbb{F}$, and we may view the rows of the matrix as vectors in $\mathbb{F}^n$. Define the (left) *row space* to be the (left) subspace spanned by the rows, and set the (left) *row rank* equal to the dimension of this subspace. By the results of the previous section, the row rank does not exceed the smaller of m and n.

Theorem A.10. *Let $A = (\, a_{i,j}\,)$ be the $m \times n$ matrix defined by the system of equations displayed above, and assume that its row rank is r. Then the dimension of the solution space for the original system of linear equations is $n - r$.*

Since there is a basically mechanical method for finding the row rank (the *Gauss-Jordan elimination procedure*), this result gives an effective method for finding the dimension of the solution space. In fact, one can push the method a little further and use it to obtain an explicit basis for the solution space.

HYPOTHESIS FOR THE REST OF THIS SECTION. From this point until the end of the current section, we shall assume that $\mathbb{F}$ is a field (so that multiplication is commutative).

If $\mathbb{F}$ is a field, there is an alternate characterization of the row rank in terms of *determinants*. The determinant is a function assigning to each square matrix A an element of $\mathbb{F}$ denoted by $\det A$ (the determinant of A), and it has the following properties:

(D-1) If A is $n \times n$, then $\det A \neq 0$ if and only if the row rank of A is equal to n.

(D-2) (*Expansion by minors*) For each all $i,\ j$ such that $1 \leq i,\ j \leq n$, let $A_{i,j}$ be the matrix obtained by deleting the i^{th} row and j^{th} column. Then for all p and q between 1 and n we have

$$\det A \ = \ \sum_{i=1}^{n} (-1)^{i+q} a_{i,q} \cdot \det A_{i,q} \ = \ \sum_{j=1}^{n} (-1)^{p+j} a_{p,j} \cdot \det A_{p,j} \ .$$

The first expression for the determinant is called its *expansion by minors along the q^{th} column* and the second is called its *expansion by minors along the p^{th} row*.

(D-3) If $A = (\, a\,)$ is 1×1, then $\det A = a$.

In particular, from these rules one can derive the standard formulas for 2×2 and 3×3 determinants that appear in high school algebra (and likewise for the more complicated rules which hold for all n and are discussed in undergraduate linear algebra courses).

Definition. If A is a matrix, then a *square submatrix* is a square matrix obtained from a by deleting certain rows and columns; for example, if $k \leq m,\ n$ then we obtain a $k \times k$ submatrix by deleting exactly $(m - k)$ rows and $(n - k)$ columns. The *determinant rank* of a (not necessarily square) matrix A is the largest d such that A contains an $r \times r$ submatrix with nonvanishing determinant.

Theorem A.11. *The row rank and determinant rank of a matrix are equal.*

4. Linear transformations

Definition. Let $\mathbb{F}$ be a skew-field, and let V and W be (left) vector space over $\mathbb{F}$. A *linear transformation* $T : V \to W$ is a function from V to W satisfying the following conditions for all $\mathbf{x},\ \mathbf{y} \in V$ and $c \in \mathbb{F}$:

(1) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$.

(2) $T(c \cdot \mathbf{x}) = c \cdot T(\mathbf{x})$.

Theorem A.12. *Let V be an n-dimensional (left) vector space over the skew-field $\mathbb{F}$ with basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, and let W be a (left) vector space over $\mathbb{F}$ with $\mathbf{w}_1, \cdots, \mathbf{w}_n \in W$. Then there is a unique linear transformation $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$, for all i.*

Important special case of Theorem A.12. If we view $\mathbb{F}$ as a vector space over itself using the given operations of addition and multiplication, then every linear transformation from $\mathbb{F}$ to itself is multiplication by an element of $\mathbb{F}$.

Definition. Let $T : V \to W$ be a linear transformation as above. The *kernel* of T, written Kernel(T), is the set of all $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{0}$, and the *image* of T, written Image(T), is the set of all $\mathbf{w} \in W$ such that $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v} \in V$. These sets are (left) vector subspaces of V and W respectively. If T_1 and T_2 are linear transformations from V to W, their *sum* is the linear transformation defined by

$$\big[T_1 + T_2\big](\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$$

for all $\mathbf{v} \in V$. If $T : V \to W$ and $U : W \to X$ are linear transformations, their *product* is the composite linear transformations $U \circ T : V \to X$ defined by

$$\big[U \circ T\big](\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$$

for all $\mathbf{v} \in V$.

Theorem A.13. *With respect to the above definitions of addition and multiplication, linear transformations from V satisfy properties* **(F-1)** – **(F-7)** *and* **(F-9)** *for skew-fields* PROVIDED *the relevant sums and products are defined.*

Remarks. **1.** The zero linear transformation $V \to W$ (where V and W are arbitrary) is the one sending every vector of V to zero. If $T : V \to W$ is linear, then so is its negative $-T : V \to W$ defined by $[-T](\mathbf{v}) = -\big[T(\mathbf{v})\big]$. If V is a vector space, then the identity map $I_V : V \to V$ defined by $I_V(\mathbf{v}) = \mathbf{v}$ is also a linear transformation.

2. Both **(F-8)** and **(F-10)** fail completely for sums and products of linear transformations. A linear transformation $T : V \to W$ for which there is a transformation $S : W \to V$ such that $S \circ T = I_V$ and $T \circ S = I_W$ is said to be *invertible*. If there is a linear transformation S with these properties, then there is only one such linear transformation, and it satisfies the inverse function relationship $\mathbf{v} = S(\mathbf{w})$ if and only if $\mathbf{w} = T(\mathbf{v})$. Thus S is the inverse mapping T^{-1}.

Theorem A.14. (*i*) *A linear transformation $T : V \to W$ is invertible if and only if it is one-to-one (equivalently, its kernel is zero) and onto (equivalently, its image is W).*

(*ii*) *If V and W are finite-dimensional, an invertible linear transformation $T : V \to W$ exists if and only if* $\dim V = \dim W$.

(*iii*) *If $T : V \to W$ and $U : W \to X$ are invertible, then so is $U \circ T$ and $[U \circ T]^{-1} = T^{-1} \circ U^{-1}$.*

(*iv*) *If $T : V \to W$ is an invertible linear transformation and V_0 is a k-dimensional vector subspace of V, then $T[V_0]$ is a k-dimensional vector subspace of W. Conversely if $T[V_0]$ is a k-dimensional vector subspace of W, then V_0 is a k-dimensional vector subspace of V.*

(*v*) *Let $\{ \mathbf{v}_1, \cdots, \mathbf{v}_n \}$ and $\{ \mathbf{w}_1, \cdots, \mathbf{w}_n \}$ be ordered bases for V. Then there is a unique invertible linear transformation $T : V \to V$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$, for all i.*

There are several basic criteria for recognizing invertible linear transformations of finite-dimensional vector spaces.

Theorem A.15. *Suppose that V and W are finite-dimensional vector spaces such that* $\dim V = \dim W$ *and $T : V \to W$ is linear. Then the following are equivalent:*

(*i*) *The linear transformation T is invertible.*

(*ii*) *If $\{ \mathbf{v}_1, \cdots, \mathbf{v}_n \}$ is a basis for V, then $\{ T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n) \}$ is a basis for W.*

(*iii*) *The linear transformation T is one-to-one.*

(*iv*) *The linear transformation T is onto.*

The reader should compare the equivalences $(i) \Leftrightarrow (iii)$ and $(i) \Leftrightarrow (iv)$ with Theorem A.13(i).

If $\mathbb{F}$ is *commutative* (hence a field), then one can also define the scalar product of a linear transformation by the rule $[cT](\mathbf{v}) = c \cdot T(\mathbf{v})$. If we let $\mathbf{Lin}_{\mathbb{F}}(V, W)$ denote the set of $\mathbb{F}$-linear transformations from one vector space V into another vector space W, this scalar product and the previously defined addition make $\mathbf{Lin}_{\mathbb{F}}(V, W)$ into a vector space over $\mathbb{F}$ in its own right. Scalar multiplication also has the following other properties:

$$\begin{aligned} c\,(U \circ T) \;&=\; (cU) \circ T \;=\; U \circ (cT) \\ (c\,I) \circ T \;&=\; c \cdot T \;=\; T \circ (c\,I) \end{aligned}$$

HYPOTHESIS. For the remainder of this section we shall assume that $\mathbb{F}$ is a (commutative) field.

Linear transformations and matrices

Linear transformations from an n-dimensional vector space to an m-dimensional vector space are closely related to $m \times n$ matrices over $\mathbb{F}$. Let $\mathcal{A} = \{ \mathbf{v}_1, \cdots, \mathbf{v}_n \}$ be an ordered basis for V, and let $\mathcal{B} = \{ \mathbf{w}_1, \cdots, \mathbf{w}_m \}$ be an ordered basis for W. Then we may define an $m \times n$ matrix whose entries $a_{i,j}$ are given by the formulas

$$T(\mathbf{v}_j) \;=\; \sum_{i=1}^{m} a_{i,j}\, \mathbf{w}_i \ .$$

We shall denote this matrix by $[T]^{\mathcal{B}}_{\mathcal{A}}$. The following relations are immediate consequences of this definition:

$$[T+S]^{\mathcal{B}}_{\mathcal{A}} \;=\; [T]^{\mathcal{B}}_{\mathcal{A}} \;+\; [S]^{\mathcal{B}}_{\mathcal{A}}$$
$$[c\,T]^{\mathcal{B}}_{\mathcal{A}} \;=\; c\,[T]^{\mathcal{B}}_{\mathcal{A}}$$

As usual, addition and scalar multiplication of matrices are defined entry by entry.

We define product matrices by the usual formula

$$C \;= A \cdot B \;\;\Longrightarrow\;\; c_{r,s} \;=\; \sum_{t=1}^{n} a_{r,t}\, b_{t,s}\ .$$

Such products are defined only if the number of columns in A equals the number of rows in B; matrix products satisfy analogs of the associative and distributive laws for scalars (and also analogs of the scalar multiplication identities displayed above), but multiplication is not commutative, even if both AB and BA are defined and have the same numbers of rows and columns.

The definition of matrix product is set up so that the associated matrices satisfy a basic compatibility relation with respect to composites of linear transformations:

$$[U \circ T]^{\mathcal{C}}_{\mathcal{A}} \;=\; [U]^{\mathcal{C}}_{\mathcal{B}} \cdot [T]^{\mathcal{B}}_{\mathcal{A}}$$

Theorem A.16. *If* $\dim V = n$ *and* $\dim W = m$*, then the mapping sending a linear transformation* T *to the matrix* $[T]^{\mathcal{B}}_{\mathcal{A}}$ *defines an invertible linear transformation of vector spaces.*

The following **Change of Basis Theorem** is useful in several contexts.

Theorem A.17. *Let* $T : V \to V$ *be a linear transformation, where* V *is finite-dimensional. Given two ordered bases* $\mathcal{A} = \{\, \mathbf{v}_1,\ \cdots\ , \mathbf{v}_n \,\}$ *and* $\mathcal{B} = \{\, \mathbf{w}_1,\ \cdots\ , \mathbf{w}_m \,\}$ *for* V*, define an* $n \times n$ *matrix* P *by the equations*

$$\mathbf{w}_j \;=\; \sum_{i=1}^{n} p_{i,j}\, \mathbf{v}_i\ .$$

Then $[T]^{\mathcal{B}}_{\mathcal{B}} \;=\; P \cdot [T]^{\mathcal{A}}_{\mathcal{A}} \cdot P^{-1}$.

Remark. If I_V is the identity transformation, then $[I_V]^{\mathcal{A}}_{\mathcal{A}} = (\,\delta_{i,j}\,)$, where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq nj$ (the *Kronecker delta symbol*). This holds for every ordered basis $\mathcal{A}$. As usual, we denote this matrix by I_n or (more simply) I.

The transposition operation on matrices

Definition. If A is an $m \times n$ matrix over $\mathbb{F}$, its **transpose** is the $n \times m$ matrix ${}^{\mathbf{T}}A = (\,[{}^{\mathbf{T}}(a)]_{i,j}\,)$ defined by $[{}^{\mathbf{T}}(a)]_{i,j} = a_{j,i}$. In other words, the rows and columns A and ${}^{\mathbf{T}}A$ are the columns and rows of A. The construction assigning to a matrix its transpose has several important properties:

Theorem A.18. *If the relevant sums and products are defined, then the following hold:*

(*i*) ${}^{\mathbf{T}}(A + B) = {}^{\mathbf{T}}A + {}^{\mathbf{T}}B$.

(*ii*) ${}^{\mathbf{T}}(c\,A) = c\,{}^{\mathbf{T}}A$.

(*iii*) ${}^{\mathbf{T}}\big({}^{\mathbf{T}}A\big) = A$.

(*iv*) *If A is a square matrix, then* $\det {}^{\mathbf{T}}A = \det A$.

(*v*) ${}^{\mathbf{T}}(AB) = {}^{\mathbf{T}}B \cdot {}^{\mathbf{T}}A$.

(*vi*) ${}^{\mathbf{T}}I_n = I_n$.

Furthermore, the (row and determinant) ranks of A and ${}^{\mathbf{T}}A$ are equal.

Invertible matrices

We say that an $n \times n$ matrix A is *invertible* if there is another $n \times n$ matrix A^{-1} for which $A \cdot A^{-1} = A^{-1} \cdot A = I$. This matrix is unique, for if B and C satisfy $BA = AC = I$, then we have

$$B = B \cdot I = B \cdot (AC) = (BA) \cdot C = I \cdot C = C\ .$$

Inverse matrices satisfy analogs of the product formula for inverse linear transformations in Theorem A.14(*ii*), and as in the case of inverse functions one has the identity

$$\big(A^{-1}\big)^{-1} = A\ .$$

Finally, we note that $I^{-1} = I$ and if A has an inverse and $c \neq 0$, then (cA) also has an inverse, which is equal to $c^{-1}\,A^{-1}$.

Theorem A.19. (*i*) *Let V be an n-dimensional vector space over $\mathbb{F}$, and let $\mathcal{A}$ be an ordered basis for V. Then the linear transformation $T : V \to V$ is invertible if and only if $[T]^{\mathcal{A}}_{\mathcal{A}}$ is, in which case we have*

$$\big[T^{-1}\big]^{\mathcal{A}}_{\mathcal{A}} = \big([T]^{\mathcal{A}}_{\mathcal{A}}\big)^{-1}\ .$$

(*ii*) *If A is an invertible $n \times n$ matrix over $\mathbb{F}$, then so is ${}^{\mathbf{T}}A$ and*

$${}^{\mathbf{T}}\big(A^{-1}\big) = \big({}^{\mathbf{T}}A\big)^{-1}\ .$$

EXERCISES

1. Suppose that V is an n-dimensional vector space over $\mathbb{F}$ and $S \subset V$ is a subset. Prove that S is an $(n-1)$-dimensional vector subspace if and only if there is a nonzero linear transformation $g : V \to \mathbb{F}$ such that S is the kernel of g. [*Hints:* If S is a vector subspace of the type described, take a basis $\mathbf{v}_1, \cdots, \mathbf{v}_{n-1}$ for S, extend it to a basis for V by adding a suitable vector $\mathbf{v}_n$, and define a linear transformation g by $g(\mathbf{v}_i) = 0$ if $i < n$ and $g(\mathbf{v}_n) = 1$. Verify that V has the required properties. — Conversely, if we are given an arbitrary linear transformation $T : V \to W$, then we know that its kernel is a vector subspace. In particular, this applies to S. It remains to show that $\dim V = n$ implies $\dim S = (n-1)$; since g is nonzero, all we can say a *priori* is that $r = \dim S < n$. As before, choose a basis take a basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$ for V such that

the first r vectors form a basis for S; we need to show that $r = (n-1)$. If $\mathbf{x} \in V$ is arbitrary, explain why the assumptions on S and g imply

$$\mathbf{x} \;-\; \left(\frac{g(\mathbf{x})}{g(\mathbf{v}_{r+1})}\right) \cdot \mathbf{v}_{r+1} \;\in\; S$$

and using this show that $\dim V$ must be equal to $r+1$, so that $r = \dim S = (n-1)$.]

2. Let V and W be vector spaces over $\mathbb{F}$, and define a **direct sum** vector space structure $V \oplus W$ on the Cartesian product $V \times W$ coordinatewise. In other words, we have the following:

$$(\mathbf{v}_1,\, \mathbf{w}_1) \;+\; (\mathbf{v}_2,\, \mathbf{w}_2) \;=\; (\mathbf{v}_1 + \mathbf{v}_2,\, \mathbf{w}_1 + \mathbf{w}_2), \qquad\qquad c\,(\mathbf{v}, \mathbf{w}) \;=\; (c\mathbf{v},\, c\mathbf{w})$$

Verify that these operations make $V \oplus W$ into a vector space. If V and W are finite-dimensional, show that the same is true for $V \oplus W$ and $\dim(V \oplus W) = \dim V + \dim W$. [*Hint:* If $\mathbf{v}_1, \cdots, \mathbf{v}_n$ is a basis for V and $\mathbf{w}_1, \cdots, \mathbf{w}_m$ is a basis for W, show that the vectors $(\mathbf{v}_i, \mathbf{0})$ and $(\mathbf{0}, \mathbf{w}_j)$ combine to form a basis for $V \oplus W$.]

3. Prove that every linear transformation $T : V \oplus W \to X$ has the form

$$T(\mathbf{v},\, \mathbf{w}) \;=\; T_1(\mathbf{v}) \;+\; T_2(\mathbf{w})$$

for some linear transformations $T_1 : V \to X$ and $T_2 : W \to X$. Conversely, show that every function T satisfying the displayed identity is a linear transformation (assuming that T_1 and T_2 are linear transformations).

5. Dot and cross products

Much of the theory of dot products in $\mathbb{R}^n$ and cross products in $\mathbb{R}^3$ can be extended to $\mathbb{F}^n$ and $\mathbb{F}^3$ for an arbitrary (commutative) field $\mathbb{F}$. We shall develop as much as possible in this degree of generality; additional properties which require the use of real numbers will be discussed at the end of this section.

The standard *dot product* (or *inner product*) in $\mathbb{R}^n$ is the scalar defined by the formula

$$\langle\, \mathbf{x},\, \mathbf{y} \,\rangle \;=\; \mathbf{x} \cdot \mathbf{y} \;=\; \sum_{i=1}^{n} x_i y_i$$

where the coordinates of $\mathbf{v}$ are given by $v_1, \cdot, v_n$. One can define a *standard bilinear pairing* on $\mathbb{F}^n$ by the same formula, and it is elementary to verify that this pairing, which is a mapping from $\mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$, has the following properties (here $\mathbf{x}\ \mathbf{y},\ \mathbf{z} \in \mathbb{F}^n$ and $c \in \mathbb{F}$ are arbitrary):

(B-1) (*Bilinearity in the right hand variable*) $\langle\, \mathbf{x},\, \mathbf{y} + \mathbf{z} \,\rangle \;=\; \langle\, \mathbf{x},\, \mathbf{y} \,\rangle \;+\; \langle\, \mathbf{x},\, \mathbf{z} \,\rangle$.

(B-2) (*Homogeneity*) $\langle\, c\,\mathbf{x},\, \mathbf{y} \,\rangle \;=\; c\,\langle\, \mathbf{x},\, \mathbf{y} \,\rangle \;=\; \langle\, \mathbf{x},\, c\,\mathbf{z} \,\rangle$.

(B-3) (*Symmetry in the variables*) $\langle\, \mathbf{x},\, \mathbf{y} \,\rangle \;=\; \langle\, \mathbf{y},\, \mathbf{x} \,\rangle$.

(B-4) (*Nondegeneracy*) For each nonzero $\mathbf{x} \in \mathbb{F}^n$ there is some $\mathbf{y} \in \mathbb{F}^n$ such that $\langle\, \mathbf{x},\, \mathbf{y} \,\rangle \;\neq\; 0$.

The verifications of the first three properties is once again a sequence of routine computations. Several further properties are immediate consequences of the given ones; one particular example is *bilinearity in the left hand variable*:

$$\langle \mathbf{x} + \mathbf{y}\, \mathbf{z} \rangle \;=\; \langle \mathbf{x},\, \mathbf{z} \rangle \;+\; \langle \mathbf{y},\, \mathbf{z} \rangle$$

Over the real numbers, one has a property which is stronger than **(B-4)**; namely, if $\mathbf{x} \neq 0$, then $\langle \mathbf{x},\, \mathbf{y} \rangle$ is the square of the length of the vector $\mathbf{x}$ (see the final portion of this section). If $\mathbb{F}$ is an arbitrary field, one can verify **(B-4)** as follows: If $\mathbf{x} \neq \mathbf{0}$, then some coordinate x_i is nonzero, and hence if $\mathbf{e}_i$ is the i^{th} unit vector then it follows that

$$\langle \mathbf{x},\, \mathbf{e}_i \rangle \;=\; x_i \;\neq\; 0\ .$$

Similarly, the *cross product* in $\mathbb{F}^3$, where $\mathbb{F}$ is an arbitrary field, can be defined by the following standard formula:

$$\mathbf{x} \times \mathbf{y} \;=\; \left(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} ,\ \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} ,\ \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right)$$

If we adopt the standard unit vector notation

$$\begin{aligned} \mathbf{i} &= (1,0,0) \\ \mathbf{j} &= (0,1,0) \\ \mathbf{k} &= (0,0,1) \end{aligned}$$

then the formula above may be restated more concisely as the following "formal3×3 determinant," which is expanded by minors along the first row:

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

The cross product has the following properties, all of which are relatively easy to verify:

(B-1) (*Bilinearity in the right hand variable*) $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) \;=\; (\mathbf{x} \times \mathbf{y}) \;+\; (\mathbf{x} \times \mathbf{z})$.

(B-2) (*Homogeneity*) $(c\,\mathbf{x} \times \mathbf{y}) \;=\; c\,(\mathbf{x} \times \mathbf{y}) \;=\; (\mathbf{x} \times c\,\mathbf{z})$.

(B-3) (*Anticommutativity*) $\mathbf{x} \times \mathbf{y} \;=\; -(\mathbf{y} \times \mathbf{x})$.

(B-4) (*Triple Product Formula*) $\langle \mathbf{x},\, \mathbf{y} \times \mathbf{z} \rangle \;=\; \langle \mathbf{z},\, \mathbf{x} \times \mathbf{y} \rangle \;=\; \langle \mathbf{y},\, \mathbf{z} \times \mathbf{x} \rangle \;=$

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}\ .$$

Since $\{\, \mathbf{i},\ \mathbf{j},\ \mathbf{k}\,\}$ is a basis for $\mathbb{F}^3$, the cross product is completely determined by the first three properties and the standard multiplication formulas

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

Several further properties are immediate consequences of the given ones; one particular example is *bilinearity in the left hand variable*:

$$(\mathbf{x} + \mathbf{y}) \times \mathbf{z} \;=\; (\mathbf{x} \times \mathbf{z}) \;+\; (\mathbf{y} \times \mathbf{z})$$

Another is the the following *Class Two nilpotence condition*:

$$\mathbf{y} \times \mathbf{y} \;=\; \mathbf{0} \qquad \text{for all } \mathbf{y}$$

NOTE. The similarity between the cross product formulas and the multiplication rules for quaternions when $\mathbb{F} = \mathbb{R}$ is not coincidental. If we identify $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ with its natural quaternionic counterpart, then we have

$$\mathbf{xy} \text{ (\textit{quaternion product})} \;=\; \mathbf{x} \times \mathbf{y} \;-\; \langle\, \mathbf{x}\,,\, \mathbf{y}\,\rangle \cdot \mathbf{1}\ .$$

In fact, for several decades during the 19^{th} century quaternions were used for many of the mathematical and physical formulas that are now expressed using 3-dimensional vector cross products and the language of vector analysis. The latter was developed by J. W. Gibbs (1839–1903) and O. Heaviside (1850–1925).

CAUTION. The cross product does **not** satisfy an associativity identity, and in general

$$\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) \;\neq\; (\mathbf{x} \times \mathbf{y}) \times \mathbf{z}\ .$$

To see this, note that $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) \;=\; \mathbf{i} \times -\mathbf{k} \;=\; -\mathbf{j}$ but $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} \;=\; \mathbf{0} \times \mathbf{j} \;=\; \mathbf{0}$. However, there is a formula which is helpful in manipulating large cross product expressions relatively well:

Theorem A.20. *For all* $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ *we have*

$$\mathbf{a} \;\times\; (\mathbf{b} \times \mathbf{c}) \;=\; \langle\, \mathbf{a},\, \mathbf{c}\,\rangle\, \mathbf{b} \;-\; \langle\, \mathbf{a},\, \mathbf{b}\,\rangle\, \mathbf{c}\ .$$

If we reverse the orders of scalars and vectors and use capital letters, the right hand side becomes

$$\mathbf{B}\,(\mathbf{A} \cdot \mathbf{C}) \;-\; \mathbf{C}\,(\mathbf{A} \cdot \mathbf{B})$$

which leads to the name **Back–Cab Rule** for the threefold product.

The following result is obtained by combining three cases of the Back–Cab Rule:

Theorem A.21. (Jacobi identity[3]) *For all* $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ *we have*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \;+\; \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) \;+\; \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) \;=\; \mathbf{0}\ .$$

For the sake of completeness, we mention one more fundamental relation between the dot and cross products:

Theorem A.22. *For all* $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ *we have*

$$\langle\, \mathbf{a} \times \mathbf{b},\, \mathbf{c} \times \mathbf{d}\,\rangle \;=\; \langle\, \mathbf{a},\, \mathbf{c}\,\rangle \cdot \langle\, \mathbf{b},\, \mathbf{d}\,\rangle \;-\; \langle\, \mathbf{a},\, \mathbf{d}\,\rangle \cdot \langle\, \mathbf{b},\, \mathbf{c}\,\rangle\ .$$

Geometric interpretations over $\mathbb{R}$

We have already mentioned that the dot product over the reals has the following property:

(B-5) (*Positive definiteness*) For each $\mathbf{x} \in \mathbb{R}^n$ we have $\langle\, \mathbf{x},\, \mathbf{x}\,\rangle \;\geq\; 0$, with equality if and only if $\mathbf{x} = \mathbf{0}$.

[3]**Carl Gustav Jacob Jacobi (1084–1851)** made many important contributions to a wide range of mathematical areas. The identity bearing his name arose in his work on differential equations and mathematical physics.

In fact, $\langle\, \mathbf{x},\, \mathbf{x}\, \rangle \,=\, |\mathbf{x}|^2$, where $|\mathbf{x}|$ is the length of $\mathbf{x}$ (equivalently, the distance between $\mathbf{x}$ and $\mathbf{0}$). More generally, we can interpret the dot product $\langle\, \mathbf{x},\, \mathbf{y}\, \rangle \,=\, \mathbf{x}\cdot\mathbf{y}$ geometrically as $|\mathbf{x}|\cdot|\mathbf{y}|\cdot\cos\alpha$, where α is the measurement of the angle $\angle\,\mathbf{x0y}$ determined by $\mathbf{x},\ \mathbf{0},\ \mathbf{y}$.

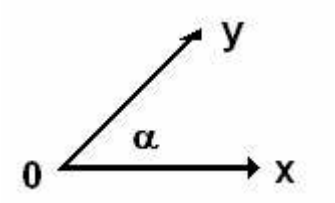

Figure A.1

It follows that the lines joining $\mathbf{0}$ to $\mathbf{x}$ and $\mathbf{0}$ to $\mathbf{y}$ are perpendicular if and only if $\mathbf{x}\cdot\mathbf{y} = 0$.

Similarly, the cross product may be viewed as a vector whose length is $|\mathbf{x}|\cdot|\mathbf{y}|\cdot\sin\alpha$. It is perpendicular to $\mathbf{x}$ and $\mathbf{y}$ by the triple product formula and basic properties of determinants (they vanish if two rows are the same), and its direction is specified by the *right hand rule*, which is illustrated below:

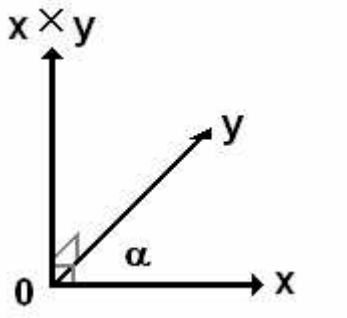

Figure A.2

Addendum. Rigid motions of $\mathbb{R}^n$

As noted at the end of Section II.6, a *rigid motion* of $\mathbb{R}^n$ is a 1–1 correspondence of $\mathbb{R}^n$ with itself that preserves distances. In that discussion we made two unproved assertions:

(A) Rigid motions belong to the affine group $\mathrm{Aff}(\mathbb{R}^n)$.

(B) Rigid motions preserve angle measurements; more precisely, if $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ are noncollinear points and T is a rigid motion, then the cosine of $\angle\,\mathbf{abc}$ is equal to the cosine of $\angle\, T(\mathbf{a})T(\mathbf{b})T(\mathbf{c})$.

Both are fairly easy to prove, and arguments appear in some undergraduate linear algebra or geometry texts, but the proofs are not always easy to locate so we shall prove both assertions here. The following result disposes of **(A)** and is the key step in proving **(B)**.

Theorem A.23. *Let T be an isometry of $\mathbb{R}^n$. Then T may be expressed in the form*

$$T(\mathbf{x}) \ = \ \mathbf{b} \ + \ A(\mathbf{x})$$

where $\mathbf{b} \in \mathbb{R}^n$ is some fixed vector and A is an orthogonal linear tranformation of $\mathbb{R}^n$ (i.e., in matrix form we have that ${}^{\mathbf{T}}A = A^{-1}$).

REMARKS. It is an elementary exercise to verify that the composite of two isometries is an isometry (and the inverse of an isometry is an isometry). If A is orthogonal, then it is elementary to prove that $T(\mathbf{x}) \ = \ \mathbf{b} \ + \ A(\mathbf{x})$ is an isometry, and in fact this is done in most if not all undergraduate linear algebra texts. On the other hand, if $A = I$ then the map above reduces to a **translation** of the form $T(\mathbf{x}) = \mathbf{b} + \mathbf{x}$, and such maps are isometries because they satisfy the even stronger identity

$$T\big(\mathbf{x} \ - \ \mathbf{y}\big) \ = \ \mathbf{x} \ - \ \mathbf{y} \ .$$

Therefore every map of the form $T(\mathbf{x}) = \mathbf{b} + A(\mathbf{x})$, where $\mathbf{b} \in \mathbb{R}^n$ is some fixed vector and A is an orthogonal linear transformation of $\mathbb{R}^n$, is an isometry of the latter. Therefore the proposition gives a complete characterization of all isometries of $\mathbb{R}^n$.

Proof. This argument is often given in linear algebra texts, and if this is not done then hints are frequently given in the exercises, so we shall merely indicate the basic steps.

First of all, the set of all isometries of $\mathbb{R}^n$ is a group (sometimes called the *Galilean group* of $\mathbb{R}^n$). It contains both the subgroups of orthogonal matrices and the subgroup of translations (maps of the form $G(\mathbf{x}) = \mathbf{x} + \mathbf{c}$ for some fixed vector $\mathbf{c}$), which is isomorphic as an additive group to $\mathbb{R}^n$ with the vector addition operation. Given $\mathbf{b} \in \mathbb{R}^n$ let $\mathbf{S_b}$ be translation by $\mathbf{b}$, so that $A = \mathbf{S}_{-T(\mathbf{0})} \circ T$ is an isometry from $\mathbb{R}^n$ to itself satisfying $A(\mathbf{0}) = \mathbf{0}$. If we can show that A is linear, then it will follow that A is given by an orthogonal matrix and the proof will be complete.

Since A is an isometry it follows that

$$\big|A(\mathbf{x}) - A(\mathbf{y})\big|^2 \ = \ \big|\mathbf{x} - \mathbf{y}\big|^2$$

and since $A(0) = 0$ it also follows that A is length preserving. If we combine these special cases with the general formula displayed above we conclude that $\langle A(\mathbf{x}), A(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \ \mathbf{y} \in \mathbb{R}^n$. In particular, it follows that A sends orthonormal bases to orthonormal bases. Let $\{\mathbf{u}_1, \ \cdots \ , \mathbf{u}_n\}$ be an orthonormal basis; then we have

$$\mathbf{x} \ = \ \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i\rangle \cdot \mathbf{u}_i$$

and likewise we have

$$A(\mathbf{x}) \ = \ \sum_{i=1}^{n} \langle A(\mathbf{x}), A(\mathbf{u}_i)\rangle \cdot A(\mathbf{u}_i) \ .$$

Since A preserves inner products we know that

$$\langle \mathbf{x}, \mathbf{u}_i\rangle \ = \ \langle A(\mathbf{x}), A(\mathbf{u}_i)\rangle$$

for all i, so that

$$A(\mathbf{x}) \ = \ \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{u}_i\rangle \cdot A(\mathbf{u}_i) \ .$$

and this implies that A is a linear transformation.■

Isometries and angle measurements

We shall use the preceding result to prove **(B)**. Specifically, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry, we need to prove that if $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ are noncollinear, then we have

$$\frac{(\mathbf{a}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{b})}{|\mathbf{a}-\mathbf{b}|\,|\mathbf{c}-\mathbf{b}|} \;=\; \frac{(\,T(\mathbf{a})-T(\mathbf{b})\,)\cdot(\,T(\mathbf{c})-T(\mathbf{b})\,)}{|T(\mathbf{a})-T(\mathbf{b})|\cdot|T(\mathbf{c})-T(\mathbf{b})|}\ .$$

Note that if T_1 and T_2 are 1–1 correspondences of $\mathbb{R}^n$ to itself and both satisfy the equation above, then their composite $T_2 \circ T_1$ also has this property.

Since an arbitrary isometry is the composite of an orthogonal linear transformation and a translation, by the preceding sentence it suffices to know that both orthogonal linear transformations and translations satisfy the displayed equation. If A is an orthogonal linear transformation, then the identity $A(\mathbf{x}-\mathbf{y}) = A(\mathbf{x}) - A(\mathbf{y})$ implies

$$\frac{A(\mathbf{a}-\mathbf{b})\cdot A(\mathbf{c}-\mathbf{b})}{|A(\mathbf{a}-\mathbf{b})|\cdot|A(\mathbf{c}-\mathbf{b})|} \;=\; \frac{(\,A(\mathbf{a})-A(\mathbf{b})\,)\cdot(\,A(\mathbf{c})-A(\mathbf{b})\,)}{|A(\mathbf{a})-A(\mathbf{b})|\cdot|A(\mathbf{c})-A(\mathbf{b})|}\ .$$

and the left hand side of this equation equals

$$\frac{(\mathbf{a}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{b})}{|\mathbf{a}-\mathbf{b}|\,|\mathbf{c}-\mathbf{b}|}$$

because orthogonal linear transformations preserve inner products and vector lengths. Thus orthogonal linear transformations preserve angle measurements. Similarly, if G is a translation, then the identity $G(\mathbf{x}-\mathbf{y}) = G(\mathbf{x}) - G(\mathbf{y})$ implies

$$\frac{(\mathbf{a}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{b})}{|\mathbf{a}-\mathbf{b}|\,|\mathbf{c}-\mathbf{b}|} \;=\; \frac{(\,G(\mathbf{a})-G(\mathbf{b})\,)\cdot(\,G(\mathbf{c})-G(\mathbf{b})\,)}{|G(\mathbf{a})-G(\mathbf{b})|\cdot|G(\mathbf{c})-G(\mathbf{b})|}$$

and hence translations also preserve angle measurements.■

Further topics

Additional information on isometries, classical geometric notions of congruence, and Euclidean similarity is contained in the following online document:

`http://math.ucr.edu/~res/math133/metgeom.pdf`

EXERCISES

Definition. An invertible square matrix P over the real numbers is said to be *conformal* if preserves angle measures in the sense described above; specifically, if $\mathbf{x},\ \mathbf{y},\ \mathbf{z}$ are noncollinear points (hence their images under P are noncollinear), then the cosine of $\angle\,\mathbf{xyz}$ is equal to the cosine of $\angle\,P\mathbf{x}P\mathbf{y}P\mathbf{z}$. — By the linearity of P, this condition holds for all noncollinear $\mathbf{x},\ \mathbf{y},\ \mathbf{z}$ if and only if it holds for triples such that $\mathbf{y} = \mathbf{0}$. Every orthogonal matrix is conformal, and nonzero scalar multiples of the identity matrix are non-orthogonal conformal matrices if the scalar is not ± 1.

1. Prove that the composite of two conformal matrices is conformal, and the inverse of a conformal matrix is conformal.

2. Prove that an $n \times n$ matrix is conformal if and only if if has the form $c\,A$, where c is a nonzero scalar and A is an orthogonal matrix. [*Hint:* Every matrix of the given form is conformal by the preceding exercise and the comments following the definition. Let $\mathbf{e}_1, \cdots, \mathbf{e}_n$ be the usual orthonormal basis of unit vectors for the space of 3×1 column vectors, and explain why the image vectors $P\mathbf{e}_1, \cdots, P\mathbf{e}_n$ are pairwise orthogonal. Let $c_i > 0$ be the length of $P\mathbf{e}_i$, and consider the angles $\angle(\mathbf{e}_1 + \mathbf{e}_i)\mathbf{0}\,\mathbf{e}_1$ for $i > 1$. Show that the conformal property implies $c_i = c_1$ for all i by considering what will happen if $c_1 < c_i$ or $c_1 > c_i$, and using this show that ${}^{\mathbf{T}}P \cdot P = c_1^2\, I$. Why does this imply that $c_1^{-1}\, P$ is orthogonal?]

3. Prove that a 2×2 matrix P is conformal if and only if there are real numbers a and b which are not both zero such that

$$P = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} .$$

[*Hints:* Writing out all 2×2 orthogonal matrices is fairly straightforward; use this to check that the given conditions hold for conformal matrices. Conversely, if P is conformal, what is ${}^{\mathbf{T}}P \cdot P$, and how can one use this to show P is a multiple of an orthogonal matrix?]

> **Remark.** (*For readers who are somewhat familiar with the theory of functions of a complex variable.*) The conclusion of this exercise may be viewed as an abstract version of a basic principle in the theory of functions of a complex variable; namely, complex analytic functions define conformal mappings of the complex plane $\mathbb{C}$. — In other words, if α and β are two parametrized curves with the same value at zero and nonzero tangent vectors there, and f is an analytic function of a complex variable which is defined near that point and has a nonzero derivative there, then the angle at which α and β intersect equals the angle at which there images $f \circ \alpha$ and $f \circ \beta$ intersect. The condition on the matrix coefficients is essentially an abstract version of the *Cauchy-Riemann equations* for analytic functions.