Appendix A

Linear Algebra from a Geometric Point of View

Whoever thinks algebra is a trick in obtaining unknowns has thought it in vain. No attention should be paid to the fact that algebra and geometry are different in appearance. Algebras (al-jabbre and maqabeleh) are geometric facts which are proved by Propositions Five and Six of Book Two of [Euclid's] Elements.
— Omar Khayyam, a paper [**A:** Khayyam (1963)]

A.0. Where Do We Start?

Usual treatments of linear and affine algebra start with a vector space as a set of "vectors" and the operations of vector addition and scalar multiplication that satisfy the axioms for a vector space. In a vector space all vectors emanate from the origin. This works well algebraically; but it ignores our geometric images and experiences of vectors.

Geometrically, we start with our experiences of Euclidean geometry where there is no point that has been singled out as the origin and where there are no numerical distances (until after a unit distance is chosen).

The linear structure of Euclidean space is carried by the translations of the space. We picture vectors as directed line segments from one point to another, and the translations serve to define when one vector is parallel to another. Any space whose translations satisfy the same properties as translations in Euclidean space is called a ***geometric affine space***, which is the subject of Section **A.1**.

The collection of all the vectors emanating from the same point is called the tangent space at that point. Using translations we may define addition and scalar multiplication of vectors. The properties of these two operations on vectors are the defining properties of a ***vector space***, which is the subject of Section **A.2**.

A.1. Geometric Affine Spaces

A ***geometric affine space over the field*** K is a space, S, together with bijections, $T_{ba}\colon S \to S$, $T_{ba}(a) = b$, which, for every pair of points a, b in S, satisfy the Properties (0)-(8), below. In this text K will always be the field of real numbers $\mathbf{R}$. We call T_{ba} ***the translation from*** a ***to*** b. We call the ordered pair (a,b) ***the (bound) vector from*** a ***to*** b. The most basic property of translations is:

(0) T_{ba} is unique in the sense that, if $T_{dc}(a) = b$, then $T_{ba} = T_{dc}$; and translations are closed under composition in the sense that

$$T_{ba}T_{dc} = T_{ec}\text{, where } e = T_{ba}(d)\text{ , where } (AB)(x) = A(B(x)).$$

Further, we assume that T_{aa} = identity [that is, $T_{aa}(x) = x$, for all x in S.]

[Note the implication that, if a is distinct from b, then T_{ba} has no fixed points.]

This property allows us to define when two bound vectors are equivalent. We say that (a,b) ***is parallel to*** (c,d) if there is a translation that takes (a,b) to (c,d), in symbols $(a,b)\approx(c,d)$. Property (0) assures us that this translation is unique, and, in this case, $T_{ca}(b) = d$, and thus, by (0), $T_{ca} = T_{db}$. Then we can also conclude, using (0) again, that

$$T_{ba} = T_{bd}T_{dc}T_{ca} = (T_{bd}T_{dc})T_{db} = (T_{dc}T_{bd})T_{db} = T_{dc}(T_{bd}T_{db}) = T_{dc}.$$

We can easily check that relation of being parallel is an equivalence relation: that is,

- $(a,b) \approx (a,b)$,
- $(a,b) \approx (c,d)$ if and only if $(c,d) \approx (a,b)$, and
- $(a,b) \approx (c,d)$ and $(a,b) \approx (e,f)$ implies that $(e,f) \approx (c,d)$.

In addition, parallel bound vectors are unique in the sense that:

- $(a,b) \approx (c,d)$ and $(a,b) \approx (c,e)$ implies that $d = e$.

This is the main property that distinguishes an affine space form other spaces. This equivalence is the same as parallel transport in Euclidean space. Parallel transport of vectors is definable in very general settings but is unique only when the space is locally isometric to Euclidean space.

We define the ***free vectors*** in S to be the equivalence classes of bound vectors. We write the equivalence class of the bound vector (a,b) to be the free vector $\mathbf{v} = [a,b]$. Note that Property (0) implies that:

- $[a,b] \leftrightarrow T_{ba}$ is a one-to-one correspondence between translations and free vectors. Thus, it follows from Property (0) that, for any point c in S,

$$(a,b) \approx (c, T_{ba}(c)).$$

Thus, every free vector $\mathbf{v}$ has a representative bound to every point c. We denote this bound vector by $\mathbf{v}_c$. We can define the addition of free vectors by

$$[c,d] + [a,b] = [c,d] + [d, T_{ba}(d)] \equiv [c, T_{ba}(d)].$$

Since $d = T_{dc}(c)$, we see that

$$[c,d] + [a,b] = [c, T_{ba}T_{dc}(c)].$$

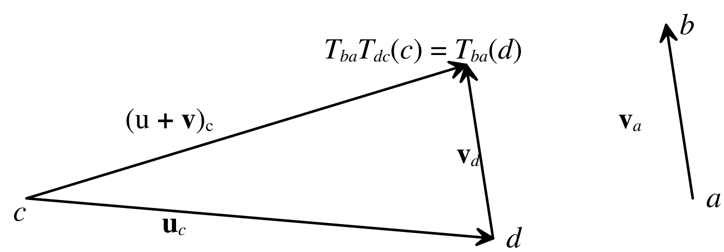

Figure A.1.1. Adding free vectors.

We can now write the further properties of a geometric affine space in terms of either free vectors or translations. For all free vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in S, and r, s in K:

(1) there is a free vector $\mathbf{0}$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$;

(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{w}$, $[T_{ba}T_{dc} = T_{dc}T_{ba}]$;

(3) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$, $[T_{fe}(T_{dc}T_{ba}) = (T_{fe}T_{dc})T_{ba}]$.

We now assume further that we have defined, for each r in K, a scalar multiplication of vectors, $r(a,b)$, [or an exponentiation of translations $(T_{ba})^r$, where $r(a,b) \equiv (a,(T_{ba})^r(a))$], with the following additional properties:

(4) $0\mathbf{v} = \mathbf{0}$, $[T_{ba}{}^0 = T_{aa}]$;

(5) $1\mathbf{v} = \mathbf{v}$, $[T_{ba}{}^1 = T_{ba}]$;

(6) $(r+s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$, $[\,(T_{ba})^{(r+s)} = (T_{ba})^r(T_{ba})^s\,]$;

{In particular, for n a positive integer,

$$n\mathbf{v} = \mathbf{v}+\mathbf{v}+...+\mathbf{v} \ (n \text{ times})$$
$$[T_{ba}{}^n = T_{ba}T_{ba}...T_{ba} \ (n \text{ times}) \text{ and } T_{ba}{}^{-1} = T_{ab}\,].\}$$

(7) $(rs)\mathbf{v} = r(s\mathbf{v})$, $[\,(T_{ba})^{(rs)} = ((T_{ba})^r)^s\,]$;

(8) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$, $[\,(T_{dc}T_{ba})^r = T_{dc}{}^rT_{ba}{}^r\,]$.

The collection of all (bound) vectors bound to a point a in S is called the ***tangent space at*** a, written

$$S_a = \{ \text{ vectors } (a,b) \mid b \text{ is in } S \ \}.$$

Since each free vector $\mathbf{v} = [a,b]$ is represented by the translation T_{ba}, we can define, for each point c in S, another point:

$$\mathbf{v}(c) = [a,b](c) = T_{ba}(c); \text{ in drawings this is:}$$

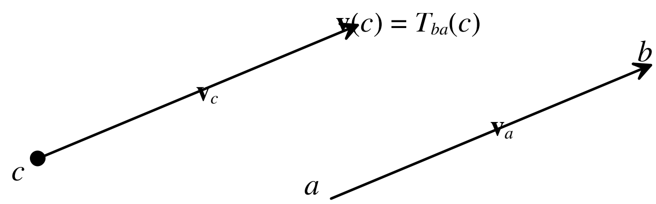

Figure A.1.2. Representing free vectors by translations.

A subset $R \subset S$ is called an ***affine subspace*** if

$$r[a,b](R) = T_{ba}^r(R) = R,$$

for every pair of points a, b in R and every r in K. If $\{a_0, a_1, ..., a_n\}$ is a finite collection of points from S, then we call the ***affine span of*** $\{a_0, a_1, ..., a_n\}$, denoted by $asp\{a_0, a_1, ..., a_n\}$, the smallest affine subspace containing each of $a_0, a_1, ..., a_n$. For two points, $a \neq b$, we call $asp\{a,b\}$ the ***line determined by*** a and b. We say that $\{a_0, a_1, ..., a_n\}$ are ***affinely independent*** if, for each i,

$$a_i \text{ is not in } asp\{a_0, a_1, ...a_{i-1}, a_{i+1}, ..., a_n\}.$$

If $\{a_0, a_1, ..., a_n\}$ are affinely independent, then we say that

$$asp\{a_0, a_1, ..., a_n\}$$

is an ***n-dimensional (affine) subspace***.

Theorem A.1.1. *The vector b is in $asp\{a_0, a_1, ..., a_n\}$ if and only if*

$$b = (r_n[a_n,a_0] + ... + r_2[a_2,a_0] + r_1[a_1,a_0])(a_0) \ \ [\text{or, } b = (T_{a_na_0}^{r_n} \cdots T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0)\ \ ],$$

for some $r_1\ r_2\ ...\ r_n$ in K.

Proof: Let R denote $asp\{a_0, a_1, ..., a_n\}$. By the definition of affine subspace,

$$T_{a_1a_0}^{r_1}(a_0) \in R \ ,$$

and thus,

$$(T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0) = T_{a_2a_0}^{r_2}(T_{a_1a_0}^{r_1}(a_0)) \ \in R \quad .$$

In this way we see that

$$b = (T_{a_na_0}^{r_n} \cdots T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0) = T_{a_na_0}^{r_n}(\cdots(T_{a_2a_0}^{r_2}(T_{a_1a_0}^{r_1}(a_0)))\cdots) \in R \quad .$$

Denote by R^* the collection of all points in R of the form:

$$(T_{a_na_0}^{r_n} \cdots T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0) \ \ .$$

Then, clearly, R^* contains each of $a_0, a_1, ..., a_n$. We now check that R^* is an affine subspace (in which case it easily follows that $R^* = R$):

Let

$$\begin{aligned} a &= (T_{a_na_0}^{s_n} \cdots T_{a_2a_0}^{s_2} T_{a_1a_0}^{s_1})(a_0) \\ b &= (T_{a_na_0}^{r_n} \cdots T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0) \\ c &= (T_{a_na_0}^{q_n} \cdots T_{a_2a_0}^{q_2} T_{a_1a_0}^{q_1})(a_0) \end{aligned}$$

be any three points in R^*. By the properties of a geometric affine space:

$$(T_{a_na_0}^{r_n-s_n} \cdots T_{a_2a_0}^{r_2-s_2} T_{a_1a_0}^{r_1-s_1})(a) = b \ \ ,$$

and thus, by Property (0),

$$(T_{a_na_0}^{r_n-s_n} \cdots T_{a_2a_0}^{r_2-s_2} T_{a_1a_0}^{r_1-s_1}) = T_{ba} \ \ .$$

Then

$$T_{ba}(c) = T_{ba}((T_{a_na_0}^{q_n} \cdots T_{a_2a_0}^{q_2} T_{a_1a_0}^{q_1})(a_0))$$

$$= (T_{a_na_0}^{r_n-s_n} \cdots T_{a_2a_0}^{r_2-s_2} T_{a_1a_0}^{r_1-s_1})((T_{a_na_0}^{q_n} \cdots T_{a_2a_0}^{q_2} T_{a_1a_0}^{q_1})(a_0))$$

$$= (T_{a_na_0}^{r_n-s_n+q_n} \cdots T_{a_2a_0}^{r_2-s_2+q_2} T_{a_1a_0}^{r_1-s_1+q_1})(a_0) \in \ R^*.$$

Corollary A.1.2. *If* $\{a_0, a_1, ..., a_n\}$ *are affinely independent, then the field elements* $r_1\ r_2\ ...\ r_n$ *in Theorem* **A.1.1** *are unique.*

Proof: If the $r_1\ r_2\ ...\ r_n$ were not unique, then

$$b = (T_{a_na_0}^{r_n} \cdots T_{a_2a_0}^{r_2} T_{a_1a_0}^{r_1})(a_0) = (T_{a_na_0}^{s_n} \cdots T_{a_2a_0}^{s_2} T_{a_1a_0}^{s_1})(a_0) \quad ,$$

and, if i were the first index for which $r_i \neq s_i$, then

$$T_{a_ia_0}^{s_i-r_i}(a_0) = (T_{a_na_0}^{r_n-s_n} \cdots T_{a_{i-2}a_0}^{r_{i-2}})(a_0)$$

and

$$a_i = T_{a_ia_0}(a_0) = (T_{a_ia_0}^{s_i-r_i})^{1/(s_i-r_i)}(a_0) = (T_{a_na_0}^{r_n-s_n} \cdots T_{a_{i-2}a_0}^{r_{i-2}})^{1/(s_i-r_i)}(a_0) \quad .$$

Thus a_i would be in $asp\{a_n, ..., a_{1+1}\}$, which contradicts the hypothesis that $\{a_0, a_1, ..., a_n\}$ are affinely independent.

Theorem A.1.3. *Let* $\{a_0, a_1, ..., a_n\}$ *be any collection of affinely independent points, and* T_{ba} *be any translation, and denote* $b_i = T_{ab}(a_i)$. *Then:*

a. $\{b_0, b_1, ..., b_n\}$ *are affinely independent, and*

b. $asp\{b_0, b_1, ..., b_n\} = T_{ba}(asp\{a_0, a_1, ..., a_n\})$.

Proof. The reader can check easily that this theorem is true because

$$T_{a_ia_j}(b_j) = T_{a_ia_j}(T_{ab}(a_j)) = T_{ba}(T_{a_ia_j}(a_j)) = T_{ba}(a_i) = b_i \quad ,$$

and thus, $T_{a_ia_j} = T_{b_ib_j}$.

A.2. Vector Spaces

We can define ***vector addition*** and ***scalar multiplication*** on the tangent space S_a at the point a in S as follows:

$$(a, b) + (a, c) \equiv (a, T_{ba}(c)) = (a, T_{ba}T_{ca}(a)),$$

and

$$r(a,b) \equiv (a, T^r_{ab}(a)).$$

These operations satisfy the following properties, which follow from the same-numbered properties of an affine space:

(1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$;

(2) there is a vector $\mathbf{0}$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$;

(3) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$;

(4) $0\mathbf{u} = \mathbf{0}$;

(5) $1\mathbf{u} = \mathbf{u}$;

(6) $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$;

(7) $(rs)\mathbf{u} = r(s\mathbf{u})$; and

(8) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$.

Any set V with two binary operations satisfying these properties is called a ***vector space over the field K***.

A subset $R \subset V$ is called a (***linear) subspace*** if $r\mathbf{u} + s\mathbf{v}$ is in R, for every pair of vectors $\mathbf{u}$, $\mathbf{v}$ in R and every r in K. If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a finite collection of vectors from V, then we call the (***linear***) ***span of*** $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$, denoted by $sp\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$, the smallest (linear) subspace containing each of $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$. We say that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ are (***linearly***) ***independent*** if, for each i, $\mathbf{u}_i$ is not in

$$sp\{\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_{i-1}, \mathbf{u}_{i+1}, ..., \mathbf{u}_n\}.$$

If $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ are linearly independent, then we say that

$$R = sp\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$$

is an n-***dimensional (linear) subspace*** and that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a ***basis*** for R.

The proofs of the two results below follow from the proofs to A.1.1. and A.1.2.

Theorem A.2.1. $\mathbf{v}$ *is in* $sp\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *if and only if*

$$\mathbf{v} = r^1\mathbf{u}_1 + r^2\mathbf{u}_2 + ... + r^n\mathbf{u}_n ,$$

for some $r^1, r^2, ..., r^n$ *in K*.

Corollary A.2.2. *If* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *are linearly independent, then, the field elements* $r^1, r^2, ...,$ r^n *in Theorem* **A.2.1** *are unique.*

If $\{\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_n\}$ is a ***basis*** for the subspace R, then, with respect to this basis, every **vector** v in R has a unique representation in terms of the $r_0, r_1, ..., r_n$. We write variously,

$$\mathbf{v} = \Sigma_{i=1}^{n} r^i \mathbf{u}_i = \Sigma\, r^i \mathbf{u}_i = r^i \mathbf{u}_i = (r^1, r^2, \cdots, r^n) \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{bmatrix}.$$

We will usually write $\mathbf{v} = \Sigma\, r^i \mathbf{u}_i$.

We can easily check that the properties of a vector space imply that if $\mathbf{v} = \Sigma\, v^i \mathbf{u}_i$ and $\mathbf{w} = \Sigma\, w^i \mathbf{u}_i$, then

$$\mathbf{v}+\mathbf{w} = \Sigma(v^i + w^i)\mathbf{u}_i \text{ and } r\mathbf{v} = \Sigma(rv^i)\mathbf{u}_i.$$

A.3. Inner Product — Lengths and Angles

In our usual experience of Euclidean space, the notion of angle is fundamental (See Chapter 3 of [**Tx**: Henderson]), and once we have chosen a unit length, then we know how to determine the length $|\mathbf{v}|$ of any vector **v**. Assuming we know what lengths and angles are, we can define the ***Euclidean inner product*** (variously called ***the standard inner product*** or ***the dot product***) of two vectors to be:

$$\langle \mathbf{v}, \mathbf{w} \rangle = |\mathbf{v}||\mathbf{w}| \cos\theta, \text{ where } \theta \text{ is the angle between } \mathbf{v} \text{ and } \mathbf{w}.$$

We can check that this inner product satisfies the following properties:

1. ***Symmetric***, $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$;
2. ***Bilinear***, $r\langle \mathbf{v}, \mathbf{w} \rangle = \langle r\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, r\mathbf{w} \rangle$, for all $r \in \mathbf{R}$,
   $\langle \mathbf{v} + \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$,
   $\langle \mathbf{v}, \mathbf{u} + \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$,
3. ***Positive definite,*** $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$.

In an abstract vector space we can start by asserting the existence of ***an inner product***, which satisfies these three properties. Then we define:

$$|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad \text{and} \quad \cos\theta = \tfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{|\mathbf{v}||\mathbf{w}|}.$$

Note that **v** and **w** are ***perpendicular*** (or ***orthogonal***) if and only if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. The collection of all vectors orthogonal to **v** is called the ***orthogonal complement of*** **v**.

We can use the inner product to express projections:

The (***orthogonal***) ***projection onto the vector*** **a**, $P_{\mathbf{a}}(\mathbf{v})$, is defined by the following picture in $sp\{\mathbf{a},\mathbf{v}\}$.

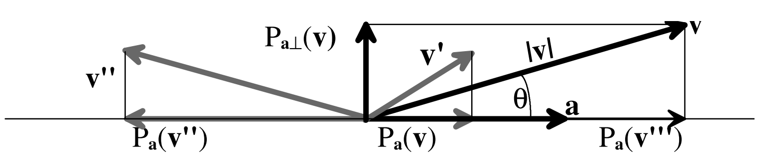

Figure A.3.1. Projection onto *a*.

Note that

$$P_{\mathbf{a}}(\mathbf{v}) = |\mathbf{v}| \cos\theta\left(\tfrac{1}{|a|}\right)\boldsymbol{a},$$

where $|\mathbf{x}|$ is the length of the vector $\mathbf{x}$ and θ is the angle between $\mathbf{v}$ and $\mathbf{a}$. Now using inner products we have:

$$\mathrm{P}_a(\mathbf{v}) = |\mathbf{v}|\cos\theta\,\frac{\mathbf{a}}{|\mathbf{a}|} = |\mathbf{v}|\frac{\langle \mathbf{v},\mathbf{a}\rangle}{|\mathbf{v}||\mathbf{a}|}\,\frac{\mathbf{a}}{|\mathbf{a}|} = \frac{\langle \mathbf{v},\mathbf{a}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle}\,\mathbf{a}\,.$$

- ***projection onto the orthogonal complement of a vector*** **a**:

$$\mathrm{P}_{\mathbf{a}\perp}(\mathbf{v}) = \mathbf{v} - \mathrm{P}_{\mathbf{a}}(\mathbf{v}) = \mathbf{v} - \frac{\langle \mathbf{v},\mathbf{a}\rangle}{\langle \mathbf{a},\mathbf{a}\rangle}\,\mathbf{a}.$$

{If the reader has not seen this before, then the reader should check carefully that this is true both formally (using the properties of the inner product) and geometrically (using Figure A.3.1).} We now apply this representation of projection to prove a famous result:

Theorem A.3.1. (***Gram-Schmidt Orthonormalization***)
If $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ *is any basis for V, then there is another basis* $\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$ *for V, such that*

a. *for each i,* $sp\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_i\} = sp\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_i\}$,

b. $\langle \mathbf{e}_i,\mathbf{e}_j\rangle = 0$, *for* $i \neq j$ (that is, $\mathbf{e}_i$ and $\mathbf{e}_j$ are *orthogonal*)*, and*

c. $\langle \mathbf{e}_i,\mathbf{e}_i\rangle = |\mathbf{e}_i|^2 = 1$ (that is, the $\mathbf{e}_i$ are *normalized*).

Proof: Let

$$\mathbf{e}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}\ ,\ \mathbf{e}_2 = \frac{\mathrm{P}_{\mathbf{e}_1\perp}(\mathbf{v}_2)}{|\mathrm{P}_{\mathbf{e}_1\perp}(\mathbf{v}_2)|} = \frac{\mathbf{v}_2 - \langle \mathbf{v}_2,\mathbf{e}_1\rangle\,\mathbf{e}_1}{|\mathbf{v}_2 - \langle \mathbf{v}_2,\mathbf{e}_1\rangle\,\mathbf{e}_1|}\,.$$

The reader should check that

$$sp\{\mathbf{v}_1\} = sp\{\mathbf{e}_1\},\ \ sp\{\mathbf{v}_1,\mathbf{v}_2\} = sp\{\mathbf{e}_1,\mathbf{e}_2\},$$

$$\langle \mathbf{e}_1,\mathbf{e}_2\rangle = 0,\ \ \langle \mathbf{e}_1,\mathbf{e}_1\rangle = 1, \text{ and } \langle \mathbf{e}_2,\mathbf{e}_2\rangle = 1.$$

In general, if $\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_k$ have been defined so that **A.3.1.**a, b, and c hold, then we can define

$$\mathbf{e}_{k+1} = \frac{\mathrm{P}_{\mathbf{e}_k\perp}(\cdots(\mathrm{P}_{\mathbf{e}_2\perp}(\mathrm{P}_{\mathbf{e}_1\perp}(\mathbf{v}_{k+1})))\cdots)}{|\mathrm{P}_{\mathbf{e}_k\perp}(\cdots(\mathrm{P}_{\mathbf{e}_2\perp}(\mathrm{P}_{\mathbf{e}_1\perp}(\mathbf{v}_{k+1})))\cdots)|} =$$

$$= \frac{\mathbf{v}_{k+1} - \langle \mathbf{v}_{k+1},\mathbf{e}_1\rangle\,\mathbf{e}_1 - \langle \mathbf{v}_{k+1},\mathbf{e}_2\rangle\,\mathbf{e}_2 - \cdots - \langle \mathbf{v}_{k+1},\mathbf{e}_k\rangle\,\mathbf{e}_k}{|\mathbf{v}_{k+1} - \langle \mathbf{v}_{k+1},\mathbf{e}_1\rangle\,\mathbf{e}_1 - \langle \mathbf{v}_{k+1},\mathbf{e}_2\rangle\,\mathbf{e}_2 - \cdots - \langle \mathbf{v}_{k+1},\mathbf{e}_k\rangle\,\mathbf{e}_k|}\,.$$

This last expression holds because the $\mathbf{e}_i$ are orthogonal and normal, and thus,

$$\begin{aligned}\mathrm{P}_{\mathbf{e}_2\perp}(\mathrm{P}_{\mathbf{e}_1\perp}(\mathbf{v})) &= \mathrm{P}_{\mathbf{e}_2\perp}(\mathbf{v} - \tfrac{\langle \mathbf{v},\mathbf{e}_1\rangle}{\langle \mathbf{e}_1,\mathbf{e}_1\rangle}\mathbf{e}_1) = \mathrm{P}_{\mathbf{e}_2\perp}(\mathbf{v} - \langle \mathbf{v},\mathbf{e}_1\rangle\mathbf{e}_1) = \\ &= (\mathbf{v} - \langle \mathbf{v},\mathbf{e}_1\rangle\mathbf{e}_1) - \langle(\mathbf{v} - \langle \mathbf{v},\mathbf{e}_1\rangle\mathbf{e}_1),\mathbf{e}_2\rangle\mathbf{e}_2 = \\ &= (\mathbf{v} - \langle \mathbf{v},\mathbf{e}_1\rangle\mathbf{e}_1) - (\langle \mathbf{v},\mathbf{e}_2\rangle - \langle \mathbf{v},\mathbf{e}_1\rangle\langle \mathbf{e}_1,\mathbf{e}_2\rangle)\mathbf{e}_2 = \\ &= \mathbf{v} - \langle \mathbf{v},\mathbf{e}_1\rangle\mathbf{e}_1 - \langle \mathbf{v},\mathbf{e}_2\rangle\mathbf{e}_2.\end{aligned}$$

The reader can easily check that a., b., and c. hold.

Lemma A.3.2. (***The Cauchy-Schwarz Inequality***) *If V has an inner product, and* $\mathbf{v}, \mathbf{w} \in V$*, then* $|\langle \mathbf{v},\mathbf{w}\rangle| \le |\mathbf{v}||\mathbf{w}|$. *Furthermore, we have equality if and only if* **v** *and* **w** *are linearly dependent.*

Proof: If we have defined the inner product geometrically as $|\mathbf{v}||\mathbf{w}|\cos\theta$, then this lemma is trivially true from the properties of the cosine. However, if we define the inner product abstractly, then the inequality in the lemma is exactly what we need in order that the definition

$$\cos\theta = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{|\mathbf{v}||\mathbf{w}|}$$

is well defined. A proof based only on the formal properties of the inner product can be found in most linear algebra books.

A.4. Linear Transformations and Operators

A ***linear transformation*** is a function T: $V \to W$ from a vector space V to a vector space W that preserves vector addition and scalar multiplication; that is,

$$\mathrm{T}(\mathbf{v} + \mathbf{u}) = \mathrm{T}(\mathbf{v}) + \mathrm{T}(\mathbf{u}) \text{ and } \mathrm{T}(r\mathbf{v}) = r\mathrm{T}(\mathbf{v}), \text{ for all } \mathbf{v}, \mathbf{u} \in V \text{ and } r \in K.$$

The proof of the following theorem is straightforward and can be found in most linear algebra books.

Theorem A.4.1. *For any linear transformation* T: $V \to W$:

a. T(V) *is a linear subspace of W* called ***the image of*** T, im(T), *and the dimension of* im(T) *called the **rank of*** T, rank(T);

b. *the kernel*

$$\ker(\mathrm{T}) \equiv \{\mathbf{v} \in V \mid \mathrm{T}(\mathbf{v}) = \mathbf{0}\}$$

*is a linear subspace of V and its dimension is called the **nullity of*** T, null(T)*;*

c. rank(T) + null(T) = dim(V)*;*

d. *if*

(i) $\{a_1,a_2,\ldots,a_m\}$ (m = null(T)) *is a basis for* null(T)*,*

(ii) $\{b_1,b_2,\ldots,b_r\}$ (r = rank(T)) *is a basis for* im(T)*, and*

(iii) *for each* $i = 1,2,\ldots,r$, *we pick* $c_i \in V$ *so that* $\mathrm{T}(c_i) = b_i$,

then

$$\{c_1,c_2,\ldots,c_r,a_1,a_2,\ldots,a_m\}$$

is a basis for V.

If $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is a basis for V and $\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$ is a basis for W, then $\mathrm{T}(\mathbf{v}_i) = \Sigma\, T_i^j \mathbf{w}_j$, where the T_i^j are numbers, and then

$$\mathrm{T}(\boldsymbol{a}) = \mathrm{T}(\Sigma\, a^i \mathbf{v}_i) = \Sigma\, a^i \mathrm{T}(\mathbf{v}_i) = \Sigma\, a^i (\Sigma\, T_i^j \mathbf{w}_j) = \Sigma\, \Sigma (a^i T_i^j) \mathbf{w}_j .$$

The numbers T_i^j form an n by m matrix, called the ***matrix of*** T ***with respect to the bases*** $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ ***and*** $\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$. For different bases there would be different matrices. And, conversely, with respect to these two bases any n by m matrix (M_i^j) will determine a linear transformation, M, by setting

$$M(\Sigma\, a^i \mathbf{v}_i) = \Sigma\, \Sigma (a^i M_i^j) \mathbf{w}_j .$$

The reader can easily check that, in Theorem **A.4.1**, with respect to the basis

$$\{c_1,c_2,\ldots,c_r,a_1,a_2,\ldots,a_n\} \text{ for } V$$

and any basis

$$\{b_1,b_2,\ldots,b_r,d_1,d_2,\ldots,d_s\} \text{ for } W,$$

we have the following corollary:

Corollary A.4.2. *For any linear transformation* T: $V \to W$, *there are bases for V and W such that with respect to these bases,* T *is represented by a matrix with* r (= rank(T)) 1*'s on the diagonal and all the other entries zero.*

If $V = W$, then a linear transformation T: $V \rightarrow V$ is called a ***linear operator***.
Examples of linear operators from R^n to R^n are:

- ***Dilation by*** λ: $\mathrm{M}_\lambda(\mathbf{v}) = \lambda\mathbf{v}$.
- ***Projection onto a vector*** **a**: (See Section **A.3**.)

$$\mathrm{P}_{\mathbf{a}}(\mathbf{v}) = |\mathbf{v}|\cos\theta\,\frac{\mathbf{a}}{|\mathbf{a}|} = \frac{\langle\mathbf{v},\mathbf{a}\rangle}{\langle\mathbf{a},\mathbf{a}\rangle}\,\mathbf{a}.$$

The fact that

$$\mathrm{P}_{\mathbf{a}}(\mathbf{v} + \mathbf{w}) = \mathrm{P}_{\mathbf{a}}(\mathbf{v}) + \mathrm{P}_{\mathbf{a}}(\mathbf{w})$$

can be seen geometrically by making a model (in your mind) of the subspace $sp\{\mathbf{a},\mathbf{v},\mathbf{w}\}$.

- ***Reflection through the orthogonal complement of a vector*** **a**:

$$\mathrm{F}_{\mathbf{a}\perp}(\mathbf{v}) = \mathbf{v} - 2\mathrm{P}_{\mathbf{a}}(\mathbf{v}) = \mathrm{P}_{\mathbf{a}\perp}(\mathbf{v}) - \mathrm{P}_{\mathbf{a}}(\mathbf{v}).$$

- ***Projection onto the orthogonal complement of a vector*** **a**:

$$\mathrm{P}_{\mathbf{a}\perp}(\mathbf{v}) = \mathbf{v} - \mathrm{P}_{\mathbf{a}}(\mathbf{v}).$$

- ***Dilation by*** λ ***in the direction of*** **a**:

$$\mathrm{M}_{\lambda\mathbf{a}}(\mathbf{v}) = \mathrm{P}_{\mathbf{a}\perp}(\mathbf{v}) + \lambda\mathrm{P}_{\mathbf{a}}(\mathbf{v}).$$

Note that if $\lambda = 0$, then this dilation is just a projection onto the orthogonal complement of **a**; and if $\lambda = -1$, then it is a reflection through the orthogonal complement of **a**.

- For two orthogonal vectors, $\langle\mathbf{w},\mathbf{a}\rangle = 0$, a (**w**,**a**)-***shear*** is the linear operator

$$\mathrm{S}_{\mathbf{w},\mathbf{a}}(\mathbf{v}) = \mathbf{v} + \langle\mathbf{v},\mathbf{w}\rangle\mathbf{a}.$$

The reader should check that a (**w**,**a**)-shear takes **a** to **a** and preserves planes that are orthogonal to **w**. [Hint: Check that $\langle\mathrm{S}_{\mathbf{w},\mathbf{a}}(\mathbf{v}),\mathbf{w}\rangle = \langle\mathbf{v},\mathbf{w}\rangle$.]

- ***Rotation in the plane of*** **a** ***and*** **b** ***through an angle*** θ:

$$\mathrm{R}_{\theta,\mathbf{ab}}(\mathbf{v}) = \mathbf{v} - [\mathrm{P}_{\mathbf{a}}(\mathbf{v}) + \mathrm{P}_{\mathbf{b}}(\mathbf{v})] + \mathrm{R}_\theta[\mathrm{P}_{\mathbf{a}}(\mathbf{v}) + \mathrm{P}_{\mathbf{b}}(\mathbf{v})],$$

where R_θ is the ordinary rotation through angle θ about **0** in plane $as\{\mathbf{a},\mathbf{b}\}$.

- ***The sum or product of linear operators is a linear operator***: that is, if T and S are linear operators, then T + S and TS are also linear operators, where

$$(\mathrm{T} + \mathrm{S})(\mathbf{v}) = \mathrm{T}(\mathbf{v}) + \mathrm{S}(\mathbf{v}) \text{ and } (\mathrm{TS})(\mathbf{v}) = \mathrm{T}(\mathrm{S}(\mathbf{v})).$$

Theorem A.4.3. *Every linear operator from* $\mathbf{R}^n$ *to* $\mathbf{R}^n$ *over the field of reals* **R** *is the composition of a finite number of shears, reflections, and dilations.*

Outline of a geometric proof—an algebraic proof will be given at the end of this section: Let T be any linear operator and use Theorem **A.4.1** followed by the Gramm-Schmidt Orthonormalization (**A.3.1**) to find an orthonormal basis $\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$ for $\mathbf{R}^n$, such that

$$\mathscr{B} = \{T(\mathbf{e}_1),T(\mathbf{e}_2),...,T(\mathbf{e}_r)\}$$

is a basis for $T(\mathbf{R}^n)$ and $\{\mathbf{e}_{r+1},\mathbf{e}_{r+2},...,\mathbf{e}_n\}$ is a basis for null(T), where $r = n\text{-}m$ is the rank of T (that is, the dimension of $T(\mathbf{R}^n)$). Since T is determined by the n vectors $\mathscr{B}$, we will have proved the theorem if we show that there is a composition of a finite number of projections, shears, reflections, dilations, and rotations that takes $\{\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_n\}$ to $\mathscr{B}$. Since $\mathscr{B}$ is a basis for $T(\mathbf{R}^n)$, we can use the proof of **A.3.1** to find an orthonormal basis $\{\mathbf{b}_1,\mathbf{b}_2,...,\mathbf{b}_r\}$ for $T(\mathbf{R}^n)$ satisfying the conclusions (a), (b), and (c) of the theorem. We will use this basis $\{\mathbf{b}_1,\mathbf{b}_2,...,\mathbf{b}_r\}$ in our proof.

First, if $\mathbf{e}_1 \neq \mathbf{b}_1$ then we can reflect (through the orthogonal complement of $\mathbf{e}_1 - \mathbf{b}_1$) to take

$$\mathbf{e}_1 \text{ to } \mathbf{b}_1 = T(\mathbf{e}_1)/|T(\mathbf{e}_1)|$$

and then dilate by $|T(\mathbf{e}_1)|$ in the direction of $T(\mathbf{e}_1)$. The result will be an operator A_1 that is the composition of a reflection and a dilation, such that $A_1(\mathbf{e}_1) = T(\mathbf{e}_1)$ and such that

$$\{A_1(\mathbf{e}_1),A_1(\mathbf{e}_2),...,A_1(\mathbf{e}_n)\}$$

is an orthogonal basis with $|A_1(\mathbf{e}_i)| = 1$, for $i > 1$.

Second, if $A_1(\mathbf{e}_2) \neq \mathbf{b}_2$, then reflect through the orthogonal complement of $A_1(\mathbf{e}_2) - \mathbf{b}_2$ to take

$$A(\mathbf{e}_2) \text{ to } \mathbf{b}_2 = \frac{P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))}{|P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))|},$$

then dilate by $|P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))|$ in the direction of $P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))$, and finally perform a $(\lambda\mathbf{b}_2,\mathbf{b}_1)$-shear to take $P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))$ to $T(\mathbf{e}_1)$. (See Figure A.4.2.)

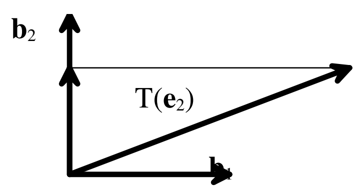

Figure A.4.2. Second step in the proof of A.4.3.

The reader can check that

$$\lambda = \frac{|P_{\mathbf{b}_1}(T(\mathbf{e}_2))|}{|P_{\mathbf{b}_1\perp}(T(\mathbf{e}_2))|}$$

and thus is positive. Note that this rotation is a rigid motion, and that the dilation and shear do not change any of the images of $\{\mathbf{e}_1,\mathbf{e}_3,...,\mathbf{e}_n\}$.

Thus, after the second stage, we have a composition, A_2, of rotations, shears, and dilations such that

$$A_2(\mathbf{e}_1) = T(\mathbf{e}_1),\ A_2(\mathbf{e}_2) = T(\mathbf{e}_2),\ \{A_2(\mathbf{e}_1),A_2(\mathbf{e}_2),...,A_2(\mathbf{e}_n)\}$$

is an orthogonal basis, and $A_2(\mathbf{e}_i) = 1$, for $i > 2$. The interested reader can now see how to continue.

The reader can check that with respect to the orthonormal basis $\{\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_n\}$, we have the following matrices:

- ***Scaling*** (***magnifying***) ***by*** λ:

$$\operatorname{matrix}(\mathrm{M}_\lambda) = \begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 0 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} = \lambda\mathrm{I}\,.$$

♦ ***Projection onto the vector*** $\mathbf{e}_k$:

$$\operatorname{matrix}(\mathrm{P}_{\mathbf{e}_k}) = \text{ 0-matrix except } \begin{array}{c} \\ \\ k\rightarrow \\ \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

♦ ***Reflection through the orthogonal complement of*** $\mathbf{e}_k$:

$$\operatorname{matrix}(\mathrm{F}_{\mathbf{e}_k\perp}) = \text{I-matrix except } \begin{array}{c} \\ \\ k\rightarrow \\ \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \cdots & \vdots & \ddots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & -1 & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

♦ ***Reflection through the orthogonal complement of*** $\mathbf{e}_k$**-**$\mathbf{e}_l$**:**

$$\operatorname{matrix}(\mathrm{F}_{(\mathbf{e}_k-\mathbf{e}_l)\perp}) = \text{I-matrix except } \begin{array}{c} \\ k\rightarrow \\ \\ l\rightarrow \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \cdots & \vdots & \ddots \\ \cdots & 0 & \cdots & 1 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdots & 1 & \cdots & 0 & \cdots \\ \ddots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}.$$

Note that multiplying by this matrix (on the left) is the same as the *elementary row operation* of interchanging the k-th and l-th rows.

♦ ***Projection onto the orthogonal complement of*** $\mathbf{e}_k$:

$$\operatorname{matrix}(\mathrm{P}_{\mathbf{e}_k\perp}) = \text{I-matrix except } \begin{array}{c} \\ \\ k\rightarrow \\ \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

♦ ***Dilation by*** **λ** ***in the direction of*** $\mathbf{e}_k$:

$$\operatorname{matrix}(\mathrm{M}_{\lambda\mathbf{e}_k}) = \text{I-matrix except } \begin{array}{c} \\ \\ k\rightarrow \\ \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & \lambda & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Note that multiplying by this matrix (on the left) is the same as multiplying the k-th row by λ.

- $(\mathbf{e}_k, \lambda\mathbf{e}_l)$***-shear***:

$$\text{matrix}(\mathrm{S}_{\mathbf{e}_k,\lambda\mathbf{e}_l}) = \text{I-matrix except } \begin{array}{c} \\ k\to \\ \\ l\to \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \cdots & \vdots & \ddots \\ \cdots & 1 & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdots & \lambda & \cdots & 1 & \cdots \\ \ddots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}.$$

Note that multiplying by this matrix (on the left) is the same as the *elementary row operation* of adding to the k-th row λ times the l-th row.

- ***Rotation in the plane of*** $\mathbf{e}_k$ ***and*** $\mathbf{e}_l$ ***through an angle*** θ:

$$\text{matrix}(\mathrm{R}_{\theta,\mathbf{e}_k\mathbf{e}_l}) = \text{I-matrix except } \begin{array}{c} \\ k\to \\ \\ l\to \\ \\ \end{array} \begin{pmatrix} \ddots & \vdots & \cdots & \vdots & \ddots \\ \cdots & \cos\theta & \cdots & \sin\theta & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdots & -\sin\theta & \cdots & \cos\theta & \cdots \\ \ddots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}.$$

- ***The sum or product of linear transformations is linear***: that is, if T and S are linear transformations with matrices T_i^j and S_i^j, then

$$(\mathrm{T} + \mathrm{S})_i^j = (\mathrm{T}_i^j + \mathrm{S}_i^j) \text{ and } (\mathrm{T}(\mathrm{S}))_i^j = \Sigma \mathrm{T}_k^j \mathrm{S}_i^k.$$

Alternate proof of Theorem **A.4.3**: If (T) is the matrix of the linear operator with respect to some orthonormal basis, the matrix (T) can be reduced to a diagonal matrix by a finite number of the types of row (or column) operations:

- interchanging the k-th and l-th rows (columns), which is the same as multiplying on the left (right) by $\text{matrix}(\mathrm{F}_{(\mathbf{e}_k-\mathbf{e}_l)\perp})$,
- adding to the k-th row (column) λ times the l-th row (column), which is the same as multiplying on the left (right) by $\text{matrix}(\mathrm{S}_{\mathbf{e}_k,\lambda\mathbf{e}_l})$.

Thus, we have $E(\mathrm{T}) = D$ (or, $(\mathrm{T})E = D$), where E is a finite product of reflection or shear matrices and D is a diagonal matrix. But each of the reflection and shear matrices has an inverse, which is of the same type, and any diagonal matrix is the product of matrices that

- multiply one row by a scalar λ, which is the same as multiplying by $\text{matrix}(\mathrm{M}_{\lambda\mathbf{e}_k})$.

Thus, we can write

$$(\mathrm{T}) = E^{-1}M \text{ (or, } M E^{-1}),$$

where now the right hand side is a finite product of reflections, shears, and dilations.

A.5. Areas, Cross Products, and Triple Products

Definition: The ***cross product*** $\mathbf{v} \times \mathbf{w}$ of two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbf{R}^3$ is the vector

1. whose magnitude is the area of the parallelogram formed by $\mathbf{v}$ and $\mathbf{w}$,
2. which is perpendicular to both $\mathbf{v}$ and $\mathbf{w}$, and

3. whose direction is such that $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{v} \times \mathbf{w}$ (in this order) form a right-hand system. (If you curl the fingers of your right hand from $\mathbf{v}$ to $\mathbf{w}$, then your thumb points in the direction of $\mathbf{v} \times \mathbf{w}$.)

Note that if the magnitude of $\mathbf{v} \times \mathbf{w}$ is 0, then the direction of $\mathbf{v} \times \mathbf{w}$ is not defined, which is correct in this case, because then $\mathbf{v} \times \mathbf{w} = \mathbf{0}$.

Theorem A.5.1. *For any two vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in* $\mathbf{R}^3$, *we have:*

a. $\mathbf{v} \parallel \mathbf{w}$ *if and only if* $\mathbf{v} \times \mathbf{w} = \mathbf{0}$,

b. $|\mathbf{v} \times \mathbf{w}| = |\mathbf{v}|\,|\mathbf{w}| \sin \theta_{\mathbf{v},\mathbf{w}} = \sqrt{|\mathbf{v}|^2|\mathbf{w}|^2 - \langle \mathbf{v}, \mathbf{w}\rangle^2}$,

c. $\mathbf{v} \times \mathbf{w}$ *is bilinear* $[\lambda(\mathbf{v} \times \mathbf{w}) = (\lambda\mathbf{v}) \times \mathbf{v} = \mathbf{v} \times (\lambda\mathbf{w})$
$(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{w})$
$\mathbf{v} \times (\mathbf{w} + \mathbf{u}) = (\mathbf{v} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{u})]$, *and*

d. $\mathbf{v} \times \mathbf{w}$ *is anticommutative* $[\mathbf{v} \times \mathbf{w} = -(\mathbf{w} \times \mathbf{v})]$.

Proof: Parts **a**, **b,** and **d** follow immediately from the definition and (for **b**) the geometric definition of $\langle \mathbf{v}, \mathbf{w}\rangle$. To see part **c**, check that the function

$$\mathbf{x} \to \mathbf{v} \times \mathbf{x} \text{ is equal to } \mathrm{R} \circ \mathrm{P}_{\mathbf{v}\perp},$$

where $\mathrm{P}_{\mathbf{v}\perp}$ is the projection onto the plane orthogonal to $\mathbf{v}$, and R is the $\pi/2$-rotation in the plane orthogonal to $\mathbf{v}$ in the direction counterclockwise from the point-of-view of $\mathbf{v}$. The linearity of the cross product now follows because both R and $\mathrm{P}_{\mathbf{v}\perp}$ are linear.

Definition: The ***triple product*** (or ***box product***) $[\mathbf{uvw}]$ of three vectors in $\mathbf{R}^3$ is defined as a number whose

- absolute value is the volume of the parallelepiped determined by the three vectors, and
- sign is positive (negative) if $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ forms a right (left) hand system.

Theorem A.5.2. *For any three vectors* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in* $\mathbf{R}^3$, *we have*:

a. $[\mathbf{uvw}] = [\mathbf{vwu}] = [\mathbf{wuv}] = -[\mathbf{uvw}] = -[\mathbf{vwu}] = -[\mathbf{wuv}]$.

b. $[\mathbf{uvw}] = \langle \mathbf{u} \times \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{v} \times \mathbf{w}\rangle = \langle \mathbf{w} \times \mathbf{u}, \mathbf{v}\rangle$.

c. $[\mathbf{uvw}] = 0$ *if and only if the three vectors are linearly dependent*.

d. *If* $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ *is a right-handed orthonormal basis for* $\mathbf{R}^3$, *then with respect to this basis,*

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = \langle \mathbf{u} \times \mathbf{v}, \mathbf{w}\rangle = \det\begin{pmatrix} u^1 & v^1 & w^1 \\ u^2 & v^2 & w^2 \\ u^3 & v^3 & w^3 \end{pmatrix},$$

where

$$\mathbf{u} = \Sigma u^i\mathbf{e}_i,\ \mathbf{v} = \Sigma v^i\mathbf{e}_i,\ \textit{and}\ \mathbf{w} = \Sigma w^i\mathbf{e}_i.$$

Proof: The reader should be able to check parts **a**, **b**, **c** directly from the definition of $[\mathbf{uvw}]$ and the geometric definition of the inner product. Part **d** follows directly from the definition of determinant in **A.6** or can at this point be taken as a definition of the determinant in $\mathbf{R}^3$. Or more directly, it is easy to check that elementary column operations will not change the volume.

Theorem A.5.3. (Double Cross Formula) *For any three vectors* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in* $\mathbf{R}^3$ *we have*:

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = \langle \mathbf{u}, \mathbf{w}\rangle\mathbf{v} - \langle \mathbf{v}, \mathbf{w}\rangle\mathbf{u}$$

and thus

$$\mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \langle \mathbf{v}, \mathbf{w} \rangle \mathbf{u} - \langle \mathbf{u}, \mathbf{w} \rangle \mathbf{v}.$$

Proof: In the special case that $\mathbf{u}$ equals $\mathbf{w}$, $(\mathbf{u} \times \mathbf{v}) \times \mathbf{u}$ is in the same direction as $P_{\mathbf{u}\perp}(\mathbf{v})$–to see this, use your right hand–and its magnitude is $|\mathbf{u}|^2|P_{\mathbf{u}\perp}(\mathbf{v})|$. Thus,

$$(\mathbf{u}\times\mathbf{v}) \times\mathbf{u} = |\mathbf{u}|^2 P_{\mathbf{u}\perp}(\mathbf{v}) = |\mathbf{u}|^2\left[\mathbf{v} - \tfrac{\langle \mathbf{u},\mathbf{v}\rangle}{|\mathbf{u}|^2}\mathbf{u}\right] = \langle \mathbf{u},\mathbf{u}\rangle\mathbf{v} - \langle \mathbf{u},\mathbf{v}\rangle\mathbf{u}.$$

If $\mathbf{u}$ and $\mathbf{v}$ are parallel, then both sides are zero. If $\mathbf{u}$ and $\mathbf{v}$ are not parallel, then $\{\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v}\}$ is a basis for $\mathbf{R}^3$, and thus, $\mathbf{w}$ is a linear combination:

$$\mathbf{w} = r\mathbf{u} + s\mathbf{v} + t(\mathbf{u} \times \mathbf{v}).$$

Therefore,

$$\begin{gathered}
(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{v}) \times (r\mathbf{u} + s\mathbf{v} + t(\mathbf{u} \times \mathbf{v})) = \\
= r((\mathbf{u} \times \mathbf{v}) \times \mathbf{u}) + s((\mathbf{u} \times \mathbf{v}) \times \mathbf{v}) + t((\mathbf{u} \times \mathbf{v}) \times (\mathbf{u} \times \mathbf{v})) = \\
= r((\mathbf{u} \times \mathbf{v}) \times \mathbf{u}) - s((\mathbf{v} \times \mathbf{u}) \times \mathbf{v}) + 0 = \\
= r\langle \mathbf{u}, \mathbf{u}\rangle\mathbf{v} - r\langle \mathbf{v}, \mathbf{u}\rangle\mathbf{u} - s\langle \mathbf{v}, \mathbf{v}\rangle\mathbf{u} + \langle \mathbf{v}, \mathbf{u}\rangle\mathbf{v} = \\
= \langle r\mathbf{u}+s\mathbf{v}, \mathbf{u}\rangle\mathbf{v} - \langle r\mathbf{u}+s\mathbf{v}, \mathbf{v}\rangle\mathbf{v} = \\
= \langle r\mathbf{u}+s\mathbf{v}+t(\mathbf{u} \times \mathbf{v}), \mathbf{u}\rangle\mathbf{v} - \langle r\mathbf{u}+s\mathbf{v}+t(\mathbf{u} \times \mathbf{v}), \mathbf{v}\rangle\mathbf{v} = \\
= \langle \mathbf{u}, \mathbf{w}\rangle\mathbf{v} - \langle \mathbf{v}, \mathbf{w}\rangle\mathbf{u}.
\end{gathered}$$

A.6. Volumes, Orientation, and Determinants

We now examine the effect of a linear operator on volumes in higher dimensions. The usual approach in linear algebra books is to start with the algebra and then go to the geometry. They define the determinant of a square matrix, and then show that all the matrices that represent the same linear operator have the same determinant, which can then be called the determinant of the linear operator. It is then shown that if T is a linear operator from $\mathbf{R}^n$ to $\mathbf{R}^n$, then this determinant is equal to

$$\frac{n\text{-volume of } \mathrm{T}(C)}{n\text{-volume of } C}$$

for any n-cube C in $\mathbf{R}^n$. We will proceed in the reverse direction. We will start with the geometry and proceed to the algebra.

Theorem A.6.1. *For every linear operator* T *from* $\mathbf{R}^n$ *to* $\mathbf{R}^n$ *and every n-cube C in* $\mathbf{R}^n$, *the ratio*

$$\frac{|n\text{-volume of } \mathrm{T}(C)|}{|n\text{-volume of } C|}$$

is a constant independent of C.

Proof: Since T is the product of reflections, shears, and dilations (Theorem **A.4.3**), we need only prove the theorem for these 3 types of operators. The reader can check that reflections and shears do not change volumes and that a dilation of λ in the direction of a vector $\mathbf{V}$ changes volumes by the ratio of λ. This completes the proof.

Definition: We denote the constant from **A.6.1** by $|\det(\mathrm{T})|$. If you are in a setting (such as $\mathbf{R}^2$ or $\mathbf{R}^3$) where there is a clear understanding of orientation then you can give $\det(\mathrm{T})$ a sign by declaring that

$$\det(\mathrm{T}) = |\det(\mathrm{T})|, \text{ if } C \text{ and } \mathrm{T}(C) \text{ have the same orientation,}$$

and

$$\det(\mathrm{T}) = -|\det(\mathrm{T})|, \text{ if } C \text{ and } \mathrm{T}(C) \text{ have different orientations.}$$

We will return to the issue of orientation later (**A.6.4** and **A.6.5**).

Theorem A.6.2. *If* T *and* S *are two linear operators from* $\mathbf{R}^n$ *to* $\mathbf{R}^n$, *then*

$$|\det(\mathrm{TS})| = |\det(\mathrm{T})|\,|\det(\mathrm{S})|.$$

Proof. We see that

$$|\det(\mathrm{TS})| = \frac{|n\text{-volume of } \mathrm{TS}(C)|}{|n\text{-volume of } C|} = \frac{|n\text{-volume of } \mathrm{T}(\mathrm{S}(C))|}{|n\text{-volume of } \mathrm{S}(C)|}\frac{|n\text{-volume of } \mathrm{S}(C)|}{|n\text{-volume of } C|} =$$

$$= \frac{|n\text{-volume of } \mathrm{T}(\mathrm{S}(C))|}{|n\text{-volume of } \mathrm{S}(C)|}|\det(\mathrm{S})|.$$

So the proof will be completed if we show that

$$\det(\mathrm{T}) = \frac{|n\text{-volume of } \mathrm{T}(\mathrm{S}(C))|}{|n\text{-volume of } \mathrm{S}(C)|}.$$

But n-volume can be calculated by filling the region with smaller and smaller cubes and then taking a limit. Thus, the ratio of the areas of the little cubes and their images will be $|\det(\mathrm{T})|$ and so also the limit.

Corollary A.6.3. *The following are equivalent:*

a. $|\det(\mathrm{T})| \neq 0$.

b. T *takes any basis to another basis.*

c. T *has an inverse* T^{-1} *such that* $\mathrm{TT}^{-1} = \mathrm{T}^{-1}\mathrm{T} =$ identity, *and*

$$|\det(\mathrm{T}^{-1})| = \frac{1}{|\det(\mathrm{T})|}.$$

Definition: If M is a matrix that represents the linear operator T with respect to a basis $\mathscr{B}$, then we define $|\det(M)| = |\det(\mathrm{T})|$. This definition is well defined because the *Alternate proof of Theorem* **A.4.3** shows that $|\det(\mathrm{S})|$ and $|\det(\mathrm{T})|$ are both determined in the same way by their common matrix M, using either row or column operations.

Theorem A.6.4. *If* T *is a linear operator* T *with* $|\det(\mathrm{T})| \neq 0$, *then there is a one-parameter family of linear operators* T_t $(0 \le t \le 1)$ *such that*

a. $\mathrm{T}_0 = \mathrm{T}$,

b. *for all* t, $|\det(\mathrm{T}_t)| \neq 0$,

c. *the function* $t \to M_t$ *is continuous where* M_t *is the matrix for* T_t *with respect to a fixed basis.*

d. T_1 *is either the identity operator or a reflection through the orthogonal complement of a fixed vector* **V**.

Any one-parameter family satisfying **a-c** is called an ***isotopy of*** T.

Proof: First note that the theorem is true if T is either a shear, dilation, or reflection, and in the case of shears and dilations (with $\lambda = 1$), T_1 is the identity. If the theorem is true for operators T and S, then $\mathrm{T}_t\mathrm{S}_t$ is easily seen to be an isotopy of TS with $(\mathrm{TS})_1 = \mathrm{T}_1\mathrm{S}_1$ equal either to the identity or to a reflection or to the product of two reflections. The product of two reflections is a rotation, which is clearly isotopic to the identity. The theorem now follows because every linear operator is the product of dilations, shears, and reflections.

Definition: If T_1 is the identity we say that T is ***orientation preserving***. If T_1 is a reflection we say that T is ***orientation reversing***.

Theorem A.6.5. *If* T *is a linear operator with* |det(T)| ≠ 0, *then* T *cannot be both orientation preserving and orientation reversing. Thus, we can define*

$$\det(T) = |\det(T)|, \textit{ if } T \textit{ is orientation preserving}$$

$$\det(T) = -|\det(T)|, \textit{ if } T \textit{ is orientation reversing.}$$

Proof: In dimensions 2 and 3 this is geometrically obvious because there is a clear meaning of orientation, and we know that a right-hand system cannot be isotoped to a left-hand system. Dimensions 2 and 3 are the only dimensions that we need in this book. In higher dimensions, the geometry is not so obvious, and thus, one may wish to resort to the algebraic proof that there is a unique function defined on n-tuples of n-vectors, which is

1. ***multilinear***—linear in each (vector) variable,
2. ***alternating***—if you interchange two vectors, you change the sign (note that this corresponds to a reflection), and
3. ***normed***—equal to 1 on the identity matrix.

See [**LA**: Damiano] for a proof of this.

A.7. Eigenvalues and Eigenvectors

A nonzero vector **v** is called an ***eigenvector*** for the linear operator T if

$$T(\mathbf{v}) = \lambda\mathbf{v}, \text{ for some scalar } \lambda.$$

The scalar is called the ***eigenvalue*** associated with the eigenvector **v**. Note that if **v** is an eigenvector, then so is a**v**, for and $a \neq 0$.

Theorem A.7.1. *The following statements about the linear operator* T *are equivalent:*

a. T *has an inverse*,

b. ker(T) = 0,

c. det(T) ≠ 0,

d. 0 *is* ***not*** *an eigenvalue for* T.

Proof: T has an inverse if and only if T is one-to-one and onto, which is true (by **A.4.1**) if and only if ker(T) = 0. T has an inverse if and only if (by **A.6.3**) det(T) ≠ 0. But, if T has 0 as an eigenvalue, then the associated eigenvector is in the kernel ker(T), and if **v** is a nonzero vector in ker(T), then $T(\mathbf{v}) = \mathbf{0} = 0\mathbf{v}$, and thus, 0 is an eigenvalue.

Thus, if λ is an eigenvalue for T, and **v** is its associated eigenvector, then $T(\mathbf{v})-\lambda\mathbf{v} = \mathbf{0}$ and **v** is in the kernel of T−λI, and thus, we have:

Theorem A.7.2. *For a linear operator* T, λ *is an eigenvalue if and only if*

$$\det(T-\lambda I) = 0.$$

Once we have found the eigenvalues, then the eigenvectors can be found by solving the linear equation $T(\mathbf{v})-\lambda\mathbf{v} = \mathbf{0}$ for **v**.

Theorem A.7.3. (The Principal Axis Theorem). *If* T *is a* ***symmetric*** *linear operator on* $\mathbf{R}^n$, *that is, for every* **v**, **w**,

$$\langle T(\mathbf{v}),\mathbf{w}\rangle = \langle \mathbf{v},T(\mathbf{w})\rangle,$$

then there is an orthonormal basis, $\{\mathbf{e}_1,...,\mathbf{e}_n\}$, *for* $\mathbf{R}^n$, *consisting entirely of eigenvectors of* T *with real eigenvalues.*

Proofs can be found in many linear algebra texts.

A.8. Introduction to Tensors

We do not use the language of tensors significantly in this book, but many of the notions in this book can be described in the tensor language. In this section we will introduce the terminology of tensors that is used in many treatments of differential geometry, and then we will use this terminology to describe many of the notions used in this book.

If V is a vector space over the field K, then a linear function from V to K is called a ***linear functional***.

Examples of linear functionals:

1. The i-th coordinate functional: $\mathbf{v} \to v^i$, which assigns to each vector $\mathbf{v}$ its i-th coordinate with respect some fixed basis.
2. The directional derivative of f with respect to a tangent vector: $\mathbf{X}_p \to \mathbf{X}_p f$, which assigns to each vector $\mathbf{X}_p$ in the tangent space T_pM the number that is the rate of change at p of f along a curve with velocity vector $\mathbf{X}_p$. This was shown to be linear in Problem **4.8**.

We can define addition and scalar multiplication for linear functionals on V as follows: If α and β are linear functionals on V, and k is an element in K, then

$$(\alpha + \beta)(\mathbf{v}) \equiv \alpha(\mathbf{v}) + \beta(\mathbf{v}) \text{ and } (k\alpha)(\mathbf{v}) \equiv k(\alpha(\mathbf{v})).$$

With these operations the space of all linear functionals on V forms a vector space called the ***conjugate space*** (or ***dual space***) to V and is often denoted by V'. If $\{\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_n\}$ is a basis for V, then define $\mathbf{e}^i$ to be the linear functional that assigns to each vector $\mathbf{v}$ its i-th coordinate with respect to the basis. Then for any linear functional α, we have

$$\begin{aligned}\alpha(\mathbf{v}) &= \alpha(v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + ... + v^n\mathbf{e}_n) = \\ &= \alpha(v^1\mathbf{e}_1) + \alpha(v^2\mathbf{e}_2) + ... + \alpha(v^n\mathbf{e}_n) = \\ &= v^1\alpha(\mathbf{e}_1) + v^2\alpha(\mathbf{e}_2) + ... + v^n\alpha(\mathbf{e}_n).\end{aligned}$$

Thus, if we define $\alpha(\mathbf{e}_i) \equiv \alpha_i$, then

$$\alpha(\mathbf{v}) = \Sigma v^i\alpha_i = \Sigma\alpha_i\mathbf{e}^i(\mathbf{v}) = \Sigma(\alpha_i\mathbf{e}^i)(\mathbf{v}),$$

and we can write

$$\alpha = \alpha_1\mathbf{e}^1 + \alpha_2\mathbf{e}^2 + ... + \alpha_n\mathbf{e}^n.$$

Thus $\{\mathbf{e}^1,\mathbf{e}^2,...,\mathbf{e}^n\}$ is a basis for V', and hence, the conjugate space has the same dimension as V.

Linear functionals are often called ***covectors*** because they are dual in the sense that if you apply a covector (linear functional) to a vector, you get a number (element of the field K) and, conversely, if you apply a vector to a covector, you also get a number. In fact, vectors in V can be expressed as linear functionals on the conjugate space V' by the identification:

$$\mathbf{v}(\alpha) \equiv \alpha(\mathbf{v}).$$

We express this by saying that $V = V''$.

A ***tensor of type*** (p,q) on the vector space V is a real-valued function F of p vectors variables and q covector (linear functional) variables:

$$F(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_p;\alpha_1,\alpha_2,\ldots,\alpha_q) \in \mathbf{R},$$

which is linear in each variable separately.

Examples of tensors:

1. Any linear functional is a tensor of type (1,0). Thus, if f is any differentiable real-valued function defined on a smooth surface M, then $\mathbf{X}_p \to \mathbf{X}_p f$ is a tensor of type (1,0) that is determined by the two numbers $\mathbf{x}_1 f$, $\mathbf{x}_2 f$.
2. Any vector $\mathbf{v}$ in V determines a unique tensor $\hat{\mathbf{v}}$ of type (0,1) by the identification: $\hat{\mathbf{v}}(\alpha)\equiv\alpha(\mathbf{v})$. Thus, it is possible to think of a vector as a tensor of type (0,1).
3. The Riemannian metric is a tensor of type (2,0) on the tangent space at a point.
4. A linear operator T determines a unique tensor $\mathrm{T}^\wedge$ of type (1,1) by the identification: $\mathrm{T}^\wedge(\mathbf{v},\alpha) = \alpha(\mathrm{T}(\mathbf{v}))$.
5. By Problem **8.5**, the Riemann curvature tensors are tensors of types (1,3) and (0,4).

But not everything is a tensor. For example, if we have a C^2 coordinate patch on a manifold M, then

$$\mathbf{X},\mathbf{Y} \to \nabla_{\mathbf{X}}\mathbf{Y}$$

is a vector-valued function, and thus its i-th coordinate $\mathbf{e}^i(\nabla_{\mathbf{X}}\mathbf{Y})$ is a number and thus,

$$\mathbf{X},\mathbf{Y},\alpha \to \alpha(\nabla_{\mathbf{X}}\mathbf{Y}) = \Sigma\alpha_i\mathbf{e}^i(\nabla_{\mathbf{X}}\mathbf{Y})$$

looks like it might be a tensor of type (2,1) but fails to be a tensor, because it depends on the values of the vector *field* $\mathbf{Y}$ near p and not just on $\mathbf{Y}_p$.