6
Basics of Euclidean Geometry

Rien n'est beau que le vrai.
—**Hermann Minkowski**

6.1 Inner Products, Euclidean Spaces

In affine geometry it is possible to deal with ratios of vectors and barycenters of points, but there is no way to express the notion of length of a line segment or to talk about orthogonality of vectors. A Euclidean structure allows us to deal with *metric notions* such as orthogonality and length (or distance).

This chapter and the next two cover the bare bones of Euclidean geometry. One of our main goals is to give the basic properties of the transformations that preserve the Euclidean structure, rotations and reflections, since they play an important role in practice. As affine geometry is the study of properties invariant under bijective affine maps and projective geometry is the study of properties invariant under bijective projective maps, Euclidean geometry is the study of properties invariant under certain affine maps called *rigid motions*. Rigid motions are the maps that preserve the distance between points. Such maps are, in fact, affine and bijective (at least in the finite–dimensional case; see Lemma 7.4.3). They form a group $\mathbf{Is}(n)$ of affine maps whose corresponding linear maps form the group $\mathbf{O}(n)$ of orthogonal transformations. The subgroup $\mathbf{SE}(n)$ of $\mathbf{Is}(n)$ corresponds to the orientation–preserving rigid motions, and there is a corresponding

subgroup $\mathbf{SO}(n)$ of $\mathbf{O}(n)$, the group of rotations. These groups play a very important role in geometry, and we will study their structure in some detail.

Before going any further, a potential confusion should be cleared up. Euclidean geometry deals with affine spaces $\left(E, \overrightarrow{E}\right)$ where the associated vector space $\overrightarrow{E}$ is equipped with an inner product. Such spaces are called *Euclidean affine spaces*. However, inner products are defined on vector spaces. Thus, we must first study the properties of vector spaces equipped with an inner product, and the linear maps preserving an inner product (the orthogonal group $\mathbf{SO}(n)$). Such spaces are called *Euclidean spaces* (omitting the word affine). It should be clear from the context whether we are dealing with a Euclidean vector space or a Euclidean affine space, but we will try to be clear about that. For instance, in this chapter, except for Definition 6.2.9, we are dealing with Euclidean vector spaces and linear maps.

We begin by defining inner products and Euclidean spaces. The Cauchy–Schwarz inequality and the Minkowski inequality are shown. We define orthogonality of vectors and of subspaces, orthogonal bases, and orthonormal bases. We offer a glimpse of Fourier series in terms of the orthogonal families $(\sin px)_{p\geq 1} \cup (\cos qx)_{q\geq 0}$ and $(e^{ikx})_{k\in\mathbb{Z}}$. We prove that every finite–dimensional Euclidean space has orthonormal bases. Orthonormal bases are the Euclidean analogue for affine frames. The first proof uses duality, and the second one the Gram–Schmidt orthogonalization procedure. The QR-decomposition for invertible matrices is shown as an application of the Gram–Schmidt procedure. Linear isometries (also called orthogonal transformations) are defined and studied briefly. We conclude with a short section in which some applications of Euclidean geometry are sketched. One of the most important applications, the method of least squares, is discussed in Chapter 13.

For a more detailed treatment of Euclidean geometry, see Berger [12, 13], Snapper and Troyer [160], or any other book on geometry, such as Pedoe [136], Coxeter [35], Fresnel [66], Tisseron [169], or Cagnac, Ramis, and Commeau [25]. Serious readers should consult Emil Artin's famous book [4], which contains an in-depth study of the orthogonal group, as well as other groups arising in geometry. It is still worth consulting some of the older classics, such as Hadamard [81, 82] and Rouché and de Comberousse [144]. The first edition of [81] was published in 1898, and finally reached its thirteenth edition in 1947! In this chapter it is assumed that all vector spaces are defined over the field $\mathbb{R}$ of real numbers unless specified otherwise (in a few cases, over the complex numbers $\mathbb{C}$).

First, we define a Euclidean structure on a vector space. Technically, a Euclidean structure over a vector space E is provided by a symmetric bilinear form on the vector space satisfying some extra properties. Recall that a bilinear form $\varphi\colon E \times E \to \mathbb{R}$ is *definite* if for every $u \in E$, $u \neq 0$ implies that $\varphi(u, u) \neq 0$, and *positive* if for every $u \in E$, $\varphi(u, u) \geq 0$.

Definition 6.1.1 A *Euclidean space* is a real vector space E equipped with a symmetric bilinear form $\varphi: E \times E \to \mathbb{R}$ that is *positive definite*. More explicitly, $\varphi: E \times E \to \mathbb{R}$ satisfies the following axioms:

$$\begin{aligned}
\varphi(u_1 + u_2, v) &= \varphi(u_1, v) + \varphi(u_2, v),\\
\varphi(u, v_1 + v_2) &= \varphi(u, v_1) + \varphi(u, v_2),\\
\varphi(\lambda u, v) &= \lambda\varphi(u, v),\\
\varphi(u, \lambda v) &= \lambda\varphi(u, v),\\
\varphi(u, v) &= \varphi(v, u),\\
u \neq 0 \text{ implies that } &\varphi(u, u) > 0.
\end{aligned}$$

The real number $\varphi(u, v)$ is also called the *inner product (or scalar product) of u and v*. We also define the *quadratic form associated with φ* as the function $\Phi: E \to \mathbb{R}_+$ such that

$$\Phi(u) = \varphi(u, u),$$

for all $u \in E$.

Since φ is bilinear, we have $\varphi(0, 0) = 0$, and since it is positive definite, we have the stronger fact that

$$\varphi(u, u) = 0 \quad \text{iff} \quad u = 0,$$

that is, $\Phi(u) = 0$ iff $u = 0$.

Given an inner product $\varphi: E \times E \to \mathbb{R}$ on a vector space E, we also denote $\varphi(u, v)$ by

$$u \cdot v \quad \text{or} \quad \langle u, v\rangle \quad \text{or} \quad (u|v),$$

and $\sqrt{\Phi(u)}$ by $\|u\|$.

Example 6.1 The standard example of a Euclidean space is $\mathbb{R}^n$, under the inner product $\cdot$ defined such that

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

There are other examples.

Example 6.2 For instance, let E be a vector space of dimension 2, and let (e_1, e_2) be a basis of E. If $a > 0$ and $b^2 - ac < 0$, the bilinear form defined such that

$$\varphi(x_1e_1 + y_1e_2,\, x_2e_1 + y_2e_2) = ax_1x_2 + b(x_1y_2 + x_2y_1) + cy_1y_2$$

yields a Euclidean structure on E. In this case,

$$\Phi(xe_1 + ye_2) = ax^2 + 2bxy + cy^2.$$

Example 6.3 Let $\mathcal{C}[a, b]$ denote the set of continuous functions $f: [a, b] \to \mathbb{R}$. It is easily checked that $\mathcal{C}[a, b]$ is a vector space of infinite dimension.

Given any two functions $f, g \in \mathcal{C}[a, b]$, let

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt.$$

We leave as an easy exercise that $\langle -, -\rangle$ is indeed an inner product on $\mathcal{C}[a, b]$. In the case where $a = -\pi$ and $b = \pi$ (or $a = 0$ and $b = 2\pi$, this makes basically no difference), one should compute

$$\langle \sin px, \sin qx\rangle, \quad \langle \sin px, \cos qx\rangle, \quad \text{and} \quad \langle \cos px, \cos qx\rangle,$$

for all natural numbers $p, q \geq 1$. The outcome of these calculations is what makes Fourier analysis possible!

Let us observe that φ can be recovered from Φ. Indeed, by bilinearity and symmetry, we have

$$\begin{aligned}\Phi(u + v) &= \varphi(u + v,\, u + v)\\ &= \varphi(u,\, u + v) + \varphi(v,\, u + v)\\ &= \varphi(u,\, u) + 2\varphi(u,\, v) + \varphi(v,\, v)\\ &= \Phi(u) + 2\varphi(u,\, v) + \Phi(v).\end{aligned}$$

Thus, we have

$$\varphi(u,\, v) = \frac{1}{2}[\Phi(u + v) - \Phi(u) - \Phi(v)].$$

We also say that φ is the *polar form of* Φ. We will generalize polar forms to polynomials, and we will see that they play a very important role.

One of the very important properties of an inner product φ is that the map $u \mapsto \sqrt{\Phi(u)}$ is a norm.

Lemma 6.1.2 *Let E be a Euclidean space with inner product φ, and let Φ be the corresponding quadratic form. For all $u, v \in E$, we have the* Cauchy–Schwarz inequality

$$\varphi(u, v)^2 \leq \Phi(u)\Phi(v),$$

the equality holding iff u and v are linearly dependent.

We also have the Minkowski inequality

$$\sqrt{\Phi(u + v)} \leq \sqrt{\Phi(u)} + \sqrt{\Phi(v)},$$

the equality holding iff u and v are linearly dependent, where in addition if $u \neq 0$ and $v \neq 0$, then $u = \lambda v$ for some $\lambda > 0$.

Proof. For any vectors $u, v \in E$, we define the function $T\colon \mathbb{R} \to \mathbb{R}$ such that

$$T(\lambda) = \Phi(u + \lambda v),$$

for all $\lambda \in \mathbb{R}$. Using bilinearity and symmetry, we have

$$\Phi(u + \lambda v) = \varphi(u + \lambda v,\, u + \lambda v)$$

$$\begin{aligned}&= \varphi(u,\, u+\lambda v) + \lambda\varphi(v,\, u+\lambda v)\\ &= \varphi(u,\, u) + 2\lambda\varphi(u,\, v) + \lambda^2\varphi(v,\, v)\\ &= \Phi(u) + 2\lambda\varphi(u,\, v) + \lambda^2\Phi(v).\end{aligned}$$

Since φ is positive definite, Φ is nonnegative, and thus $T(\lambda) \geq 0$ for all $\lambda \in \mathbb{R}$. If $\Phi(v) = 0$, then $v = 0$, and we also have $\varphi(u,\, v) = 0$. In this case, the Cauchy–Schwarz inequality is trivial, and $v = 0$ and u are linearly dependent.

Now, assume $\Phi(v) > 0$. Since $T(\lambda) \geq 0$, the quadratic equation

$$\lambda^2\Phi(v) + 2\lambda\varphi(u,\, v) + \Phi(u) = 0$$

cannot have distinct real roots, which means that its discriminant

$$\Delta = 4(\varphi(u,\, v)^2 - \Phi(u)\Phi(v))$$

is null or negative, which is precisely the Cauchy–Schwarz inequality

$$\varphi(u, v)^2 \leq \Phi(u)\Phi(v).$$

If

$$\varphi(u, v)^2 = \Phi(u)\Phi(v),$$

then the above quadratic equation has a double root λ_0, and we have $\Phi(u + \lambda_0 v) = 0$. If $\lambda_0 = 0$, then $\varphi(u,\, v) = 0$, and since $\Phi(v) > 0$, we must have $\Phi(u) = 0$, and thus $u = 0$. In this case, of course, $u = 0$ and v are linearly dependent. Finally, if $\lambda_0 \neq 0$, since $\Phi(u + \lambda_0 v) = 0$ implies that $u + \lambda_0 v = 0$, then u and v are linearly dependent. Conversely, it is easy to check that we have equality when u and v are linearly dependent.

The Minkowski inequality

$$\sqrt{\Phi(u+v)} \leq \sqrt{\Phi(u)} + \sqrt{\Phi(v)}$$

is equivalent to

$$\Phi(u+v) \leq \Phi(u) + \Phi(v) + 2\sqrt{\Phi(u)\Phi(v)}.$$

However, we have shown that

$$2\varphi(u,\, v) = \Phi(u+v) - \Phi(u) - \Phi(v),$$

and so the above inequality is equivalent to

$$\varphi(u,\, v) \leq \sqrt{\Phi(u)\Phi(v)},$$

which is trivial when $\varphi(u,\, v) \leq 0$, and follows from the Cauchy–Schwarz inequality when $\varphi(u,\, v) \geq 0$. Thus, the Minkowski inequality holds. Finally, assume that $u \neq 0$ and $v \neq 0$, and that

$$\sqrt{\Phi(u+v)} = \sqrt{\Phi(u)} + \sqrt{\Phi(v)}.$$

When this is the case, we have

$$\varphi(u,\, v) = \sqrt{\Phi(u)\Phi(v)},$$

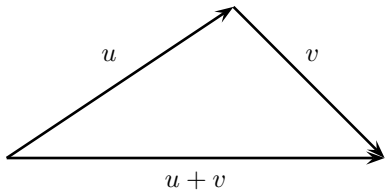

Figure 6.1. The triangle inequality

and we know from the discussion of the Cauchy–Schwarz inequality that the equality holds iff u and v are linearly dependent. The Minkowski inequality is an equality when u or v is null. Otherwise, if $u \neq 0$ and $v \neq 0$, then $u = \lambda v$ for some $\lambda \neq 0$, and since

$$\varphi(u,\, v) = \lambda\varphi(v,\, v) = \sqrt{\Phi(u)\Phi(v)},$$

by positivity, we must have $\lambda > 0$. □

Note that the Cauchy–Schwarz inequality can also be written as

$$|\varphi(u, v)| \leq \sqrt{\Phi(u)}\sqrt{\Phi(v)}.$$

Remark: It is easy to prove that the Cauchy–Schwarz and the Minkowski inequalities still hold for a symmetric bilinear form that is positive, but not necessarily definite (i.e., $\varphi(u, v) \geq 0$ for all $u, v \in E$). However, u and v need not be linearly dependent when the equality holds.

The Minkowski inequality

$$\sqrt{\Phi(u+v)} \leq \sqrt{\Phi(u)} + \sqrt{\Phi(v)}$$

shows that the map $u \mapsto \sqrt{\Phi(u)}$ satisfies the convexity inequality (also known as triangle inequality), condition (N3) of Definition 17.2.2, and since φ is bilinear and positive definite, it also satisfies conditions (N1) and (N2) of Definition 17.2.2, and thus it is a *norm* on E. The norm induced by φ is called the *Euclidean norm induced by* φ.

Note that the Cauchy–Schwarz inequality can be written as

$$|u \cdot v| \leq \|u\|\|v\|,$$

and the Minkowski inequality as

$$\|u + v\| \leq \|u\| + \|v\|.$$

Figure 6.1 illustrates the triangle inequality.

We now define orthogonality.

6.2 Orthogonality, Duality, Adjoint of a Linear Map

An inner product on a vector space gives the ability to define the notion of orthogonality. Families of nonnull pairwise orthogonal vectors must be linearly independent. They are called orthogonal families. In a vector space of finite dimension it is always possible to find orthogonal bases. This is very useful theoretically and practically. Indeed, in an orthogonal basis, finding the coordinates of a vector is very cheap: It takes an inner product. Fourier series make crucial use of this fact. When E has finite dimension, we prove that the inner product on E induces a natural isomorphism between E and its dual space E^*. This allows us to define the adjoint of a linear map in an intrinsic fashion (i.e., independently of bases). It is also possible to orthonormalize any basis (certainly when the dimension is finite). We give two proofs, one using duality, the other more constructive using the Gram–Schmidt orthonormalization procedure.

Definition 6.2.1 Given a Euclidean space E, any two vectors $u, v \in E$ are *orthogonal, or perpendicular*, if $u \cdot v = 0$. Given a family $(u_i)_{i \in I}$ of vectors in E, we say that $(u_i)_{i \in I}$ is *orthogonal* if $u_i \cdot u_j = 0$ for all $i, j \in I$, where $i \neq j$. We say that the family $(u_i)_{i \in I}$ is *orthonormal* if $u_i \cdot u_j = 0$ for all $i, j \in I$, where $i \neq j$, and $\|u_i\| = u_i \cdot u_i = 1$, for all $i \in I$. For any subset F of E, the set

$$F^\perp = \{v \in E \mid u \cdot v = 0, \text{ for all } u \in F\},$$

of all vectors orthogonal to all vectors in F, is called the *orthogonal complement of* F.

Since inner products are positive definite, observe that for any vector $u \in E$, we have

$$u \cdot v = 0 \quad \text{for all } v \in E \quad \text{iff} \quad u = 0.$$

It is immediately verified that the orthogonal complement $F^\perp$ of F is a subspace of E.

Example 6.4 Going back to Example 6.3 and to the inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)g(t)dt$$

on the vector space $\mathcal{C}[-\pi, \pi]$, it is easily checked that

$$\langle \sin px, \sin qx \rangle = \begin{cases} \pi & \text{if } p = q,\ p, q \geq 1, \\ 0 & \text{if } p \neq q,\ p, q \geq 1, \end{cases}$$

$$\langle \cos px, \cos qx \rangle = \begin{cases} \pi & \text{if } p = q,\ p, q \geq 1, \\ 0 & \text{if } p \neq q,\ p, q \geq 0, \end{cases}$$

and

$$\langle \sin px, \cos qx \rangle = 0,$$

for all $p \geq 1$ and $q \geq 0$, and of course, $\langle 1, 1 \rangle = \int_{-\pi}^{\pi} dx = 2\pi$.

As a consequence, the family $(\sin px)_{p \geq 1} \cup (\cos qx)_{q \geq 0}$ is orthogonal. It is not orthonormal, but becomes so if we divide every trigonometric function by $\sqrt{\pi}$, and 1 by $\sqrt{2\pi}$.

Remark: Observe that if we allow complex–valued functions, we obtain simpler proofs. For example, it is immediately checked that

$$\int_{-\pi}^{\pi} e^{ikx} dx = \begin{cases} 2\pi & \text{if } k = 0, \\ 0 & \text{if } k \neq 0, \end{cases}$$

because the derivative of e^{ikx} is ike^{ikx}.

However, beware that something strange is going on. Indeed, unless $k = 0$, we have

$$\langle e^{ikx}, e^{ikx} \rangle = 0,$$

since

$$\langle e^{ikx}, e^{ikx} \rangle = \int_{-\pi}^{\pi} (e^{ikx})^2 dx = \int_{-\pi}^{\pi} e^{i2kx} dx = 0.$$

The inner product $\langle e^{ikx}, e^{ikx} \rangle$ should be strictly positive. What went wrong?

The problem is that we are using the wrong inner product. When we use complex-valued functions, we must use the *Hermitian inner product*

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx,$$

where $\overline{g(x)}$ is the *conjugate* of $g(x)$. The Hermitian inner product is not symmetric. Instead,

$$\langle g, f \rangle = \overline{\langle f, g \rangle}.$$

(Recall that if $z = a + ib$, where $a, b \in \mathbb{R}$, then $\overline{z} = a - ib$. Also, $e^{i\theta} = \cos\theta + i \sin\theta$). With the Hermitian inner product, everything works out beautifully! In particular, the family $(e^{ikx})_{k \in \mathbb{Z}}$ is orthogonal. Hermitian spaces and some basics of Fourier series will be discussed more rigorously in Chapter 10.

We leave the following simple two results as exercises.

Lemma 6.2.2 *Given a Euclidean space E, for any family $(u_i)_{i \in I}$ of nonnull vectors in E, if $(u_i)_{i \in I}$ is orthogonal, then it is linearly independent.*

Lemma 6.2.3 *Given a Euclidean space E, any two vectors $u, v \in E$ are orthogonal iff*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

One of the most useful features of orthonormal bases is that they afford a very simple method for computing the coordinates of a vector over any basis vector. Indeed, assume that $(e_1, \ldots, e_m)$ is an orthonormal basis. For any vector

$$x = x_1 e_1 + \cdots + x_m e_m,$$

if we compute the inner product $x \cdot e_i$, we get

$$x \cdot e_i = x_1 e_1 \cdot e_i + \cdots + x_i e_i \cdot e_i + \cdots + x_m e_m \cdot e_i = x_i,$$

since

$$e_i \cdot e_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

is the property characterizing an orthonormal family. Thus,

$$x_i = x \cdot e_i,$$

which means that $x_i e_i = (x \cdot e_i)e_i$ is the orthogonal projection of x onto the subspace generated by the basis vector e_i. If the basis is orthogonal but not necessarily orthonormal, then

$$x_i = \frac{x \cdot e_i}{e_i \cdot e_i} = \frac{x \cdot e_i}{\|e_i\|^2}.$$

All this is true even for an infinite orthonormal (or orthogonal) basis $(e_i)_{i \in I}$.

However, remember that every vector x is expressed as a linear combination

$$x = \sum_{i \in I} x_i e_i$$

where the family of scalars $(x_i)_{i \in I}$ has **finite support**, which means that $x_i = 0$ for all $i \in I - J$, where J is a finite set. Thus, even though the family $(\sin px)_{p \geq 1} \cup (\cos qx)_{q \geq 0}$ is orthogonal (it is not orthonormal, but becomes so if we divide every trigonometric function by $\sqrt{\pi}$, and 1 by $\sqrt{2\pi}$; we won't because it looks messy!), the fact that a function $f \in \mathcal{C}^0[-\pi, \pi]$ can be written as a Fourier series as

$$f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

does not mean that $(\sin px)_{p \geq 1} \cup (\cos qx)_{q \geq 0}$ is a basis of this vector space of functions, because in general, the families (a_k) and (b_k) **do not** have

finite support! In order for this infinite linear combination to make sense, it is necessary to prove that the partial sums

$$a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

of the series converge to a limit when n goes to infinity. This requires a topology on the space.

Still, a small miracle happens. If $f \in \mathcal{C}[-\pi, \pi]$ can indeed be expressed as a Fourier series

$$f(x) = a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx),$$

the coefficients a_0 and a_k, b_k, $k \geq 1$, can be computed by projecting f over the basis functions, i.e., by taking inner products with the basis functions in $(\sin px)_{p\geq 1} \cup (\cos qx)_{q\geq 0}$. Indeed, for all $k \geq 1$, we have

$$a_0 = \frac{\langle f, 1\rangle}{\|1\|^2},$$

and

$$a_k = \frac{\langle f, \cos kx\rangle}{\|\cos kx\|^2}, \qquad b_k = \frac{\langle f, \sin kx\rangle}{\|\sin kx\|^2},$$

that is,

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx,$$

and

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\, dx, \qquad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\, dx.$$

If we allow f to be complex-valued and use the family $(e^{ikx})_{k\in\mathbb{Z}}$, which is is indeed orthogonal w.r.t. the Hermitian inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx,$$

we consider functions $f \in \mathcal{C}[-\pi, \pi]$ that can be expressed as the sum of a series

$$f(x) = \sum_{k\in\mathbb{Z}} c_k e^{ikx}.$$

Note that the index k is allowed to be a negative integer. Then, the formula giving the c_k is very nice:

$$c_k = \frac{\langle f, e^{ikx}\rangle}{\|e^{ikx}\|^2},$$

that is,

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx.$$

Note the presence of the negative sign in e^{-ikx}, which is due to the fact that the inner product is Hermitian. Of course, the real case can be recovered from the complex case. If f is a real-valued function, then we must have

$$a_k = c_k + c_{-k} \quad \text{and} \quad b_k = i(c_k - c_{-k}).$$

Also note that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx$$

is defined not only for all discrete values $k \in \mathbb{Z}$, but for all $k \in \mathbb{R}$, and that if f is continuous over $\mathbb{R}$, the integral makes sense. This suggests defining

$$\widehat{f}(k) = \int_{-\infty}^{\infty} f(x)e^{-ikx}dx,$$

called the *Fourier transform* of f. The Fourier transform analyzes the function f in the "frequency domain" in terms of its spectrum of harmonics. Amazingly, there is an inverse Fourier transform (change e^{-ikx} to e^{+ikx} and divide by the scale factor 2π) that reconstructs f (under certain assumptions on f).

Some basics of Fourier series will be discussed more rigorously in Chapter 10. For more on Fourier analysis, we highly recommend Strang [165] for a lucid introduction with lots of practical examples, and then move on to a good real analysis text, for instance Lang [109, 110], or [145].

A very important property of Euclidean spaces of finite dimension is that the inner product induces a canonical bijection (i.e., independent of the choice of bases) between the vector space E and its dual E^*.

Given a Euclidean space E, for any vector $u \in E$, let $\varphi_u \colon E \to \mathbb{R}$ be the map defined such that

$$\varphi_u(v) = u \cdot v,$$

for all $v \in E$.

Since the inner product is bilinear, the map φ_u is a linear form in E^*. Thus, we have a map $\flat \colon E \to E^*$, defined such that

$$\flat(u) = \varphi_u.$$

Lemma 6.2.4 *Given a Euclidean space E, the map $\flat \colon E \to E^*$ defined such that*

$$\flat(u) = \varphi_u$$

is linear and injective. When E is also of finite dimension, the map $\flat \colon E \to E^*$ is a canonical isomorphism.

Proof. That $\flat\colon E \to E^*$ is a linear map follows immediately from the fact that the inner product is bilinear. If $\varphi_u = \varphi_v$, then $\varphi_u(w) = \varphi_v(w)$ for all $w \in E$, which by definition of φ_u means that

$$u \cdot w = v \cdot w$$

for all $w \in E$, which by bilinearity is equivalent to

$$(v - u) \cdot w = 0$$

for all $w \in E$, which implies that $u = v$, since the inner product is positive definite. Thus, $\flat\colon E \to E^*$ is injective. Finally, when E is of finite dimension n, we know that E^* is also of dimension n, and then $\flat\colon E \to E^*$ is bijective. □

The inverse of the isomorphism $\flat\colon E \to E^*$ is denoted by $\sharp\colon E^* \to E$.

As a consequence of Lemma 6.2.4, if E is a Euclidean space of finite dimension, every linear form $f \in E^*$ corresponds to a unique $u \in E$ such that

$$f(v) = u \cdot v,$$

for every $v \in E$. In particular, if f is not the null form, the kernel of f, which is a hyperplane H, is precisely the set of vectors that are orthogonal to u.

Remarks:

(1) The "musical map" $\flat\colon E \to E^*$ is not surjective when E has infinite dimension. The result can be salvaged by restricting our attention to continuous linear maps, and by assuming that the vector space E is a *Hilbert space* (i.e., E is a complete normed vector space w.r.t. the Euclidean norm). This is the famous "little" Riesz theorem (or Riesz representation theorem).

(2) Lemma 6.2.4 still holds if the inner product on E is replaced by a nondegenerate symmetric bilinear form φ. We say that a symmetric bilinear form $\varphi\colon E \times E \to \mathbb{R}$ is *nondegenerate* if for every $u \in E$,

$$\text{if} \quad \varphi(u, v) = 0 \quad \text{for all } v \in E, \quad \text{then} \quad u = 0.$$

For example, the symmetric bilinear form on $\mathbb{R}^4$ defined such that

$$\varphi((x_1, x_2, x_3, x_4), (y_1, y_2, y_3, y_4)) = x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4$$

is nondegenerate. However, there are nonnull vectors $u \in \mathbb{R}^4$ such that $\varphi(u, u) = 0$, which is impossible in a Euclidean space. Such vectors are called *isotropic*.

The existence of the isomorphism $\flat\colon E \to E^*$ is crucial to the existence of adjoint maps. The importance of adjoint maps stems from the fact that

the linear maps arising in physical problems are often self-adjoint, which means that $f = f^*$. Moreover, self-adjoint maps can be diagonalized over orthonormal bases of eigenvectors. This is the key to the solution of many problems in mechanics, and engineering in general (see Strang [165]).

Let E be a Euclidean space of finite dimension n, and let $f: E \to E$ be a linear map. For every $u \in E$, the map

$$v \mapsto u \cdot f(v)$$

is clearly a linear form in E^*, and by Lemma 6.2.4, there is a unique vector in E denoted by $f^*(u)$ such that

$$f^*(u) \cdot v = u \cdot f(v),$$

for every $v \in E$. The following simple lemma shows that the map f^* is linear.

Lemma 6.2.5 *Given a Euclidean space E of finite dimension, for every linear map $f: E \to E$, there is a unique linear map $f^*: E \to E$ such that*

$$f^*(u) \cdot v = u \cdot f(v),$$

for all $u, v \in E$. The map f^* is called the adjoint of f (w.r.t. to the inner product).

Proof. Given $u_1, u_2 \in E$, since the inner product is bilinear, we have

$$(u_1 + u_2) \cdot f(v) = u_1 \cdot f(v) + u_2 \cdot f(v),$$

for all $v \in E$, and

$$(f^*(u_1) + f^*(u_2)) \cdot v = f^*(u_1) \cdot v + f^*(u_2) \cdot v,$$

for all $v \in E$, and since by assumption,

$$f^*(u_1) \cdot v = u_1 \cdot f(v)$$

and

$$f^*(u_2) \cdot v = u_2 \cdot f(v),$$

for all $v \in E$, we get

$$(f^*(u_1) + f^*(u_2)) \cdot v = (u_1 + u_2) \cdot f(v),$$

for all $v \in E$. Since $\flat$ is bijective, this implies that

$$f^*(u_1 + u_2) = f^*(u_1) + f^*(u_2).$$

Similarly,

$$(\lambda u) \cdot f(v) = \lambda(u \cdot f(v)),$$

for all $v \in E$, and

$$(\lambda f^*(u)) \cdot v = \lambda(f^*(u) \cdot v),$$

for all $v \in E$, and since by assumption,

$$f^*(u) \cdot v = u \cdot f(v),$$

for all $v \in E$, we get

$$(\lambda f^*(u)) \cdot v = (\lambda u) \cdot f(v),$$

for all $v \in E$. Since $\flat$ is bijective, this implies that

$$f^*(\lambda u) = \lambda f^*(u).$$

Thus, f^* is indeed a linear map, and it is unique, since $\flat$ is a bijection. □

Linear maps $f: E \to E$ such that $f = f^*$ are called *self-adjoint* maps. They play a very important role because they have real eigenvalues, and because orthonormal bases arise from their eigenvectors. Furthermore, many physical problems lead to self-adjoint linear maps (in the form of symmetric matrices).

Remark: Lemma 6.2.5 still holds if the inner product on E is replaced by a nondegenerate symmetric bilinear form φ.

Linear maps such that $f^{-1} = f^*$, or equivalently

$$f^* \circ f = f \circ f^* = \mathrm{id},$$

also play an important role. They are *linear isometries*, or *isometries*. Rotations are special kinds of isometries. Another important class of linear maps are the linear maps satisfying the property

$$f^* \circ f = f \circ f^*,$$

called *normal linear maps*. We will see later on that normal maps can always be diagonalized over orthonormal bases of eigenvectors, but this will require using a Hermitian inner product (over $\mathbb{C}$).

Given two Euclidean spaces E and F, where the inner product on E is denoted by $\langle -, - \rangle_1$ and the inner product on F is denoted by $\langle -, - \rangle_2$, given any linear map $f: E \to F$, it is immediately verified that the proof of Lemma 6.2.5 can be adapted to show that there is a unique linear map $f^*: F \to E$ such that

$$\langle f(u), v \rangle_2 = \langle u, f^*(v) \rangle_1$$

for all $u \in E$ and all $v \in F$. The linear map f^* is also called the *adjoint of* f.

Remark: Given any basis for E and any basis for F, it is possible to characterize the matrix of the adjoint f^* of f in terms of the matrix of f, and the symmetric matrices defining the inner products. We will do so with

respect to orthonormal bases. Also, since inner products are symmetric, the adjoint f^* of f is also characterized by

$$f(u) \cdot v = u \cdot f^*(v),$$

for all $u, v \in E$.

We can also use Lemma 6.2.4 to show that any Euclidean space of finite dimension has an orthonormal basis.

Lemma 6.2.6 *Given any nontrivial Euclidean space E of finite dimension $n \geq 1$, there is an orthonormal basis $(u_1, \ldots, u_n)$ for E.*

Proof. We proceed by induction on n. When $n = 1$, take any nonnull vector $v \in E$, which exists, since we assumed E nontrivial, and let

$$u = \frac{v}{\|v\|}.$$

If $n \geq 2$, again take any nonnull vector $v \in E$, and let

$$u_1 = \frac{v}{\|v\|}.$$

Consider the linear form φ_{u_1} associated with u_1. Since $u_1 \neq 0$, by Lemma 6.2.4, the linear form φ_{u_1} is nonnull, and its kernel is a hyperplane H. Since $\varphi_{u_1}(w) = 0$ iff $u_1 \cdot w = 0$, the hyperplane H is the orthogonal complement of $\{u_1\}$. Furthermore, since $u_1 \neq 0$ and the inner product is positive definite, $u_1 \cdot u_1 \neq 0$, and thus, $u_1 \notin H$, which implies that $E = H \oplus \mathbb{R}u_1$. However, since E is of finite dimension n, the hyperplane H has dimension $n-1$, and by the induction hypothesis, we can find an orthonormal basis $(u_2, \ldots, u_n)$ for H. Now, because H and the one dimensional space $\mathbb{R}u_1$ are orthogonal and $E = H \oplus \mathbb{R}u_1$, it is clear that $(u_1, \ldots, u_n)$ is an orthonormal basis for E. □

There is a more constructive way of proving Lemma 6.2.6, using a procedure known as the *Gram–Schmidt orthonormalization procedure*. Among other things, the Gram–Schmidt orthonormalization procedure yields the so-called *QR-decomposition for matrices*, an important tool in numerical methods.

Lemma 6.2.7 *Given any nontrivial Euclidean space E of finite dimension $n \geq 1$, from any basis $(e_1, \ldots, e_n)$ for E we can construct an orthonormal basis $(u_1, \ldots, u_n)$ for E, with the property that for every k, $1 \leq k \leq n$, the families $(e_1, \ldots, e_k)$ and $(u_1, \ldots, u_k)$ generate the same subspace.*

Proof. We proceed by induction on n. For $n = 1$, let

$$u_1 = \frac{e_1}{\|e_1\|}.$$

For $n \geq 2$, we also let

$$u_1 = \frac{e_1}{\|e_1\|},$$

and assuming that $(u_1, \ldots, u_k)$ is an orthonormal system that generates the same subspace as $(e_1, \ldots, e_k)$, for every k with $1 \leq k < n$, we note that the vector

$$u'_{k+1} = e_{k+1} - \sum_{i=1}^{k} (e_{k+1} \cdot u_i)\, u_i$$

is nonnull, since otherwise, because $(u_1, \ldots, u_k)$ and $(e_1, \ldots, e_k)$ generate the same subspace, $(e_1, \ldots, e_{k+1})$ would be linearly dependent, which is absurd, since $(e_1, \ldots, e_n)$ is a basis. Thus, the norm of the vector u'_{k+1} being nonzero, we use the following construction of the vectors u_k and u'_k:

$$u'_1 = e_1, \qquad u_1 = \frac{u'_1}{\|u'_1\|},$$

and for the inductive step

$$u'_{k+1} = e_{k+1} - \sum_{i=1}^{k} (e_{k+1} \cdot u_i)\, u_i, \qquad u_{k+1} = \frac{u'_{k+1}}{\|u'_{k+1}\|},$$

where $1 \leq k \leq n-1$. It is clear that $\|u_{k+1}\| = 1$, and since $(u_1, \ldots, u_k)$ is an orthonormal system, we have

$$u'_{k+1} \cdot u_i = e_{k+1} \cdot u_i - (e_{k+1} \cdot u_i) u_i \cdot u_i = e_{k+1} \cdot u_i - e_{k+1} \cdot u_i = 0,$$

for all i with $1 \leq i \leq k$. This shows that the family $(u_1, \ldots, u_{k+1})$ is orthonormal, and since $(u_1, \ldots, u_k)$ and $(e_1, \ldots, e_k)$ generates the same subspace, it is clear from the definition of u_{k+1} that $(u_1, \ldots, u_{k+1})$ and $(e_1, \ldots, e_{k+1})$ generate the same subspace. This completes the induction step and the proof of the lemma. □

Note that u'_{k+1} is obtained by subtracting from e_{k+1} the projection of e_{k+1} itself onto the orthonormal vectors $u_1, \ldots, u_k$ that have already been computed. Then, u'_{k+1} is normalized. The Gram–Schmidt orthonormalization procedure is illustrated in Figure 6.2.

Remarks:

(1) The QR-decomposition can now be obtained very easily, but we postpone this until Section 6.4.

(2) We could compute u'_{k+1} using the formula

$$u'_{k+1} = e_{k+1} - \sum_{i=1}^{k} \left(\frac{e_{k+1} \cdot u'_i}{\|u'_i\|^2} \right) u'_i,$$

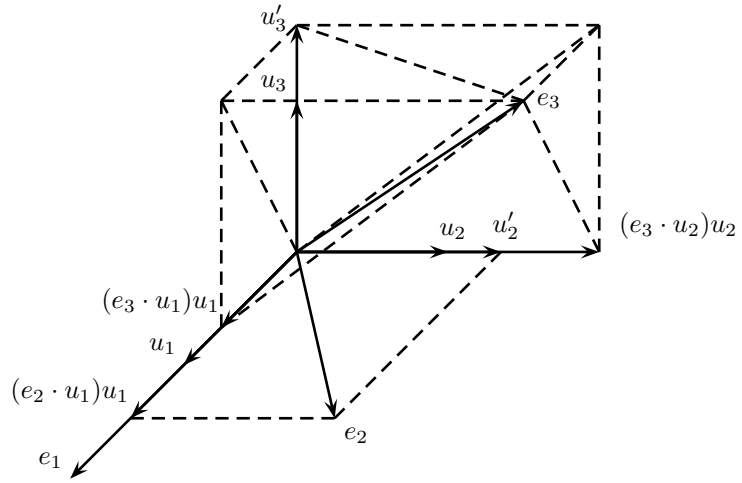

Figure 6.2. The Gram–Schmidt orthonormalization procedure

and normalize the vectors u'_k at the end. This time, we are subtracting from e_{k+1} the projection of e_{k+1} itself onto the orthogonal vectors $u'_1, \ldots, u'_k$. This might be preferable when writing a computer program.

(3) The proof of Lemma 6.2.7 also works for a countably infinite basis for E, producing a countably infinite orthonormal basis.

Example 6.5 If we consider polynomials and the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)dt,$$

applying the Gram–Schmidt orthonormalization procedure to the polynomials

$$1, x, x^2, \ldots, x^n, \ldots,$$

which form a basis of the polynomials in one variable with real coefficients, we get a family of orthonormal polynomials $Q_n(x)$ related to the *Legendre polynomials*.

The Legendre polynomials $P_n(x)$ have many nice properties. They are orthogonal, but their norm is not always 1. The Legendre polynomials $P_n(x)$ can be defined as follows. Letting f_n be the function

$$f_n(x) = (x^2 - 1)^n,$$

we define $P_n(x)$ as follows:

$$P_0(x) = 1, \quad \text{and} \quad P_n(x) = \frac{1}{2^n n!}\, f_n^{(n)}(x),$$

where $f_n^{(n)}$ is the nth derivative of f_n.

They can also be defined inductively as follows:

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ P_{n+1}(x) &= \frac{2n+1}{n+1}\, xP_n(x) - \frac{n}{n+1}\, P_{n-1}(x). \end{aligned}$$

It turns out that the polynomials Q_n are related to the Legendre polynomials P_n as follows:

$$Q_n(x) = \frac{2^n (n!)^2}{(2n)!} P_n(x).$$

As a consequence of Lemma 6.2.6 (or Lemma 6.2.7), given any Euclidean space of finite dimension n, if $(e_1, \ldots, e_n)$ is an orthonormal basis for E, then for any two vectors $u = u_1 e_1 + \cdots + u_n e_n$ and $v = v_1 e_1 + \cdots + v_n e_n$, the inner product $u \cdot v$ is expressed as

$$u \cdot v = (u_1 e_1 + \cdots + u_n e_n) \cdot (v_1 e_1 + \cdots + v_n e_n) = \sum_{i=1}^{n} u_i v_i,$$

and the norm $\|u\|$ as

$$\|u\| = \|u_1 e_1 + \cdots + u_n e_n\| = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

We can also prove the following lemma regarding orthogonal spaces.

Lemma 6.2.8 *Given any nontrivial Euclidean space E of finite dimension $n \geq 1$, for any subspace F of dimension k, the orthogonal complement $F^\perp$ of F has dimension $n - k$, and $E = F \oplus F^\perp$. Furthermore, we have $F^{\perp\perp} = F$.*

Proof. From Lemma 6.2.6, the subspace F has some orthonormal basis $(u_1, \ldots, u_k)$. This linearly independent family $(u_1, \ldots, u_k)$ can be extended to a basis $(u_1, \ldots, u_k, v_{k+1}, \ldots, v_n)$, and by Lemma 6.2.7, it can be converted to an orthonormal basis $(u_1, \ldots, u_n)$, which contains $(u_1, \ldots, u_k)$ as an orthonormal basis of F. Now, any vector $w = w_1 u_1 + \cdots + w_n u_n \in E$ is orthogonal to F iff $w \cdot u_i = 0$, for every i, where $1 \leq i \leq k$, iff $w_i = 0$ for every i, where $1 \leq i \leq k$. Clearly, this shows that $(u_{k+1}, \ldots, u_n)$ is a basis of $F^\perp$, and thus $E = F \oplus F^\perp$, and $F^\perp$ has dimension $n - k$. Similarly, any vector $w = w_1 u_1 + \cdots + w_n u_n \in E$ is orthogonal to $F^\perp$ iff $w \cdot u_i = 0$, for every i, where $k + 1 \leq i \leq n$, iff $w_i = 0$ for every i, where $k + 1 \leq i \leq n$. Thus, $(u_1, \ldots, u_k)$ is a basis of $F^{\perp\perp}$, and $F^{\perp\perp} = F$.

We now define Euclidean affine spaces.

Definition 6.2.9 An affine space $\left(E, \overrightarrow{E}\right)$ is a *Euclidean affine space* if its underlying vector space $\overrightarrow{E}$ is a Euclidean vector space. Given any two

points $a, b \in E$, we define the *distance between a and b, or length of the segment* (a, b), as $\|\mathbf{ab}\|$, the Euclidean norm of $\mathbf{ab}$. Given any two pairs of points (a, b) and (c, d), we define their inner product as $\mathbf{ab} \cdot \mathbf{cd}$. We say that (a, b) and (c, d) are *orthogonal, or perpendicular*, if $\mathbf{ab} \cdot \mathbf{cd} = 0$. We say that two affine subspaces F_1 and F_2 of E are *orthogonal* if their directions F_1 and F_2 are orthogonal.

The verification that the distance defined in Definition 6.2.9 satisfies the axioms of Definition 17.2.1 is immediate. Note that a Euclidean affine space is a normed affine space, in the sense of Definition 17.2.3. We denote by $\mathbb{E}^m$ the Euclidean affine space obtained from the affine space $\mathbb{A}^m$ by defining on the vector space $\mathbb{R}^m$ the standard inner product

$$(x_1, \ldots, x_m) \cdot (y_1, \ldots, y_m) = x_1 y_1 + \cdots + x_m y_m.$$

The corresponding Euclidean norm is

$$\|(x_1, \ldots, x_m)\| = \sqrt{x_1^2 + \cdots + x_m^2}.$$

6.3 Linear Isometries (Orthogonal Transformations)

In this section we consider linear maps between Euclidean spaces that preserve the Euclidean norm. These transformations, sometimes called *rigid motions*, play an important role in geometry.

Definition 6.3.1 Given any two nontrivial Euclidean spaces E and F of the same finite dimension n, a function $f: E \to F$ is *an orthogonal transformation, or a linear isometry*, if it is linear and

$$\|f(u)\| = \|u\|,$$

for all $u \in E$.

Remarks:

(1) A linear isometry is often defined as a linear map such that

$$\|f(v) - f(u)\| = \|v - u\|,$$

for all $u, v \in E$. Since the map f is linear, the two definitions are equivalent. The second definition just focuses on preserving the distance between vectors.

(2) Sometimes, a linear map satisfying the condition of Definition 6.3.1 is called a *metric map*, and a linear isometry is defined as a *bijective* metric map.

An isometry (without the word linear) is sometimes defined as a function $f: E \to F$ (not necessarily linear) such that

$$\|f(v) - f(u)\| = \|v - u\|,$$

for all $u, v \in E$, i.e., as a function that preserves the distance. This requirement turns out to be very strong. Indeed, the next lemma shows that all these definitions are equivalent when E and F are of finite dimension, and for functions such that $f(0) = 0$.

Lemma 6.3.2 *Given any two nontrivial Euclidean spaces E and F of the same finite dimension n, for every function $f: E \to F$, the following properties are equivalent:*

(1) f is a linear map and $\|f(u)\| = \|u\|$, for all $u \in E$;

(2) $\|f(v) - f(u)\| = \|v - u\|$, for all $u, v \in E$, and $f(0) = 0$;

(3) $f(u) \cdot f(v) = u \cdot v$, for all $u, v \in E$.

Furthermore, such a map is bijective.

Proof. Clearly, (1) implies (2), since in (1) it is assumed that f is linear.

Assume that (2) holds. In fact, we shall prove a slightly stronger result. We prove that if

$$\|f(v) - f(u)\| = \|v - u\|$$

for all $u, v \in E$, then for any vector $\tau \in E$, the function $g: E \to F$ defined such that

$$g(u) = f(\tau + u) - f(\tau)$$

for all $u \in E$ is a linear map such that $g(0) = 0$ and (3) holds. Clearly, $g(0) = f(\tau) - f(\tau) = 0$.

Note that from the hypothesis

$$\|f(v) - f(u)\| = \|v - u\|$$

for all $u, v \in E$, we conclude that

$$\begin{aligned} \|g(v) - g(u)\| &= \|f(\tau + v) - f(\tau) - (f(\tau + u) - f(\tau))\|, \\ &= \|f(\tau + v) - f(\tau + u)\|, \\ &= \|\tau + v - (\tau + u)\|, \\ &= \|v - u\|, \end{aligned}$$

for all $u, v \in E$. Since $g(0) = 0$, by setting $u = 0$ in

$$\|g(v) - g(u)\| = \|v - u\|,$$

we get

$$\|g(v)\| = \|v\|$$

for all $v \in E$. In other words, g preserves both the distance and the norm.

To prove that g preserves the inner product, we use the simple fact that

$$2u \cdot v = \|u\|^2 + \|v\|^2 - \|u - v\|^2$$

for all $u, v \in E$. Then, since g preserves distance and norm, we have

$$\begin{aligned} 2g(u) \cdot g(v) &= \|g(u)\|^2 + \|g(v)\|^2 - \|g(u) - g(v)\|^2 \\ &= \|u\|^2 + \|v\|^2 - \|u - v\|^2 \\ &= 2u \cdot v, \end{aligned}$$

and thus $g(u) \cdot g(v) = u \cdot v$, for all $u, v \in E$, which is (3).

In particular, if $f(0) = 0$, by letting $\tau = 0$, we have $g = f$, and f preserves the scalar product, i.e., (3) holds.

Now assume that (3) holds. Since E is of finite dimension, we can pick an orthonormal basis $(e_1, \ldots, e_n)$ for E. Since f preserves inner products, $(f(e_1), \ldots, f(e_n))$ is also orthonormal, and since F also has dimension n, it is a basis of F. Then note that for any $u = u_1 e_1 + \cdots + u_n e_n$, we have

$$u_i = u \cdot e_i,$$

for all i, $1 \leq i \leq n$. Thus, we have

$$f(u) = \sum_{i=1}^{n} (f(u) \cdot f(e_i)) f(e_i),$$

and since f preserves inner products, this shows that

$$f(u) = \sum_{i=1}^{n} (u \cdot e_i) f(e_i) = \sum_{i=1}^{n} u_i f(e_i),$$

which shows that f is linear. Obviously, f preserves the Euclidean norm, and (3) implies (1).

Finally, if $f(u) = f(v)$, then by linearity $f(v-u) = 0$, so that $\|f(v-u)\| = 0$, and since f preserves norms, we must have $\|v-u\| = 0$, and thus $u = v$. Thus, f is injective, and since E and F have the same finite dimension, f is bijective. □

Remarks:

(i) The dimension assumption is needed only to prove that (3) implies (1) when f is not known to be linear, and to prove that f is surjective, but the proof shows that (1) implies that f is injective.

(ii) In (2), when f does not satisfy the condition $f(0) = 0$, the proof shows that f is an affine map. Indeed, taking any vector τ as an origin, the map g is linear, and

$$f(\tau + u) = f(\tau) + g(u)$$

for all $u \in E$, proving that f is affine with associated linear map g.

(iii) Paul Hughett showed me a nice proof of the following interesting fact: The implication that (3) implies (1) holds if we also assume that f is surjective, even if E has infinite dimension. Indeed, observe that

$$\begin{aligned}(f(\lambda u+\mu v)-\lambda f(u)-\mu f(v))\cdot f(w)&\\ = f(\lambda u+\mu v)\cdot f(w)-\lambda f(u)\cdot f(w)-\mu f(v)\cdot f(w)&\\ = (\lambda u+\mu v)\cdot w-\lambda u\cdot w-\mu v\cdot w = 0,&\end{aligned}$$

since f preserves the inner product. However, if f is surjective, every $z\in E$ is of the form $z=f(w)$ for some $w\in E$, and the above equation implies that

$$(f(\lambda u+\mu v)-\lambda f(u)-\mu f(v))\cdot z=0$$

for all $z\in E$, which implies that

$$f(\lambda u+\mu v)-\lambda f(u)-\mu f(v)=0,$$

proving that f is linear.

In view of Lemma 6.3.2, we will drop the word "linear" in "linear isometry," unless we wish to emphasize that we are dealing with a map between vector spaces.

We are now going to take a closer look at the isometries $f\colon E\to E$ of a Euclidean space of finite dimension.

6.4 The Orthogonal Group, Orthogonal Matrices

In this section we explore some of the basic properties of the orthogonal group and of orthogonal matrices.

Lemma 6.4.1 *Let E be any Euclidean space of finite dimension n, and let $f\colon E\to E$ be any linear map. The following properties hold:*

(1) The linear map $f\colon E\to E$ is an isometry iff

$$f\circ f^*=f^*\circ f=\mathrm{id}.$$

(2) For every orthonormal basis $(e_1,\ldots,e_n)$ of E, if the matrix of f is A, then the matrix of f^* is the transpose $A^\top$ of A, and f is an isometry iff A satisfies the identities

$$AA^\top=A^\top A=I_n,$$

where I_n denotes the identity matrix of order n, iff the columns of A form an orthonormal basis of E, iff the rows of A form an orthonormal basis of E.

Proof. (1) The linear map $f: E \to E$ is an isometry iff

$$f(u) \cdot f(v) = u \cdot v,$$

for all $u, v \in E$, iff

$$f^*(f(u)) \cdot v = f(u) \cdot f(v) = u \cdot v$$

for all $u, v \in E$, which implies

$$(f^*(f(u)) - u) \cdot v = 0$$

for all $u, v \in E$. Since the inner product is positive definite, we must have

$$f^*(f(u)) - u = 0$$

for all $u \in E$, that is,

$$f^* \circ f = f \circ f^* = \mathrm{id}.$$

(2) If $(e_1, \ldots, e_n)$ is an orthonormal basis for E, let $A = (a_{i,j})$ be the matrix of f, and let $B = (b_{i,j})$ be the matrix of f^*. Since f^* is characterized by

$$f^*(u) \cdot v = u \cdot f(v)$$

for all $u, v \in E$, using the fact that if $w = w_1 e_1 + \cdots + w_n e_n$ we have $w_k = w \cdot e_k$ for all k, $1 \leq k \leq n$, letting $u = e_i$ and $v = e_j$, we get

$$b_{j,i} = f^*(e_i) \cdot e_j = e_i \cdot f(e_j) = a_{i,j},$$

for all i, j, $1 \leq i, j \leq n$. Thus, $B = A^\top$. Now, if X and Y are arbitrary matrices over the basis $(e_1, \ldots, e_n)$, denoting as usual the jth column of X by X_j, and similarly for Y, a simple calculation shows that

$$X^\top Y = (X_i \cdot Y_j)_{1 \leq i,j \leq n}.$$

Then it is immediately verified that if $X = Y = A$, then

$$A^\top A = A\, A^\top = I_n$$

iff the column vectors $(A_1, \ldots, A_n)$ form an orthonormal basis. Thus, from (1), we see that (2) is clear (also because the rows of A are the columns of $A^\top$). □

Lemma 6.4.1 shows that the inverse of an isometry f is its adjoint f^*. Lemma 6.4.1 also motivates the following definition. The set of all real $n \times n$ matrices is denoted by $\mathrm{M}_n(\mathbb{R})$.

Definition 6.4.2 A real $n \times n$ matrix is an *orthogonal matrix* if

$$A\, A^\top = A^\top A = I_n.$$

Remark: It is easy to show that the conditions $A\, A^\top = I_n$, $A^\top A = I_n$, and $A^{-1} = A^\top$, are equivalent. Given any two orthonormal bases $(u_1, \ldots, u_n)$

and $(v_1, \ldots, v_n)$, if P is the change of basis matrix from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$ (i.e., the columns of P are the coordinates of the v_j w.r.t. $(u_1, \ldots, u_n)$), since the columns of P are the coordinates of the vectors v_j with respect to the basis $(u_1, \ldots, u_n)$, and since $(v_1, \ldots, v_n)$ is orthonormal, the columns of P are orthonormal, and by Lemma 6.4.1 (2), the matrix P is orthogonal.

The proof of Lemma 6.3.2 (3) also shows that if f is an isometry, then the image of an orthonormal basis $(u_1, \ldots, u_n)$ is an orthonormal basis. Students often ask why ortho*gon*al matrices are not called ortho*norm*al matrices, since their columns (and rows) are orthonormal bases! I have no good answer, but isometries do preserve orthogonality, and orthogonal matrices correspond to isometries.

Recall that the determinant $\det(f)$ of a linear map $f: E \to E$ is independent of the choice of a basis in E. Also, for every matrix $A \in \mathrm{M}_n(\mathbb{R})$, we have $\det(A) = \det(A^\top)$, and for any two $n \times n$ matrices A and B, we have $\det(AB) = \det(A)\det(B)$ (for all these basic results, see Lang [107]). Then, if f is an isometry, and A is its matrix with respect to any orthonormal basis, $A\,A^\top = A^\top A = I_n$ implies that $\det(A)^2 = 1$, that is, either $\det(A) = 1$, or $\det(A) = -1$. It is also clear that the isometries of a Euclidean space of dimension n form a group, and that the isometries of determinant $+1$ form a subgroup. This leads to the following definition.

Definition 6.4.3 Given a Euclidean space E of dimension n, the set of isometries $f: E \to E$ forms a subgroup of $\mathbf{GL}(E)$ denoted by $\mathbf{O}(E)$, or $\mathbf{O}(n)$ when $E = \mathbb{R}^n$, called the *orthogonal group (of E)*. For every isometry f, we have $\det(f) = \pm 1$, where $\det(f)$ denotes the determinant of f. The isometries such that $\det(f) = 1$ are called *rotations, or proper isometries, or proper orthogonal transformations*, and they form a subgroup of the special linear group $\mathbf{SL}(E)$ (and of $\mathbf{O}(E)$), denoted by $\mathbf{SO}(E)$, or $\mathbf{SO}(n)$ when $E = \mathbb{R}^n$, called the *special orthogonal group (of E)*. The isometries such that $\det(f) = -1$ are called *improper isometries, or improper orthogonal transformations, or flip transformations*.

As an immediate corollary of the Gram–Schmidt orthonormalization procedure, we obtain the QR-decomposition for invertible matrices.

6.5 QR-Decomposition for Invertible Matrices

Now that we have the definition of an orthogonal matrix, we can explain how the Gram–Schmidt orthonormalization procedure immediately yields the QR-decomposition for matrices.

Lemma 6.5.1 *Given any real $n \times n$ matrix A, if A is invertible, then there is an orthogonal matrix Q and an upper triangular matrix R with positive diagonal entries such that $A = QR$.*

Proof. We can view the columns of A as vectors $A_1, \ldots, A_n$ in $\mathbb{E}^n$. If A is invertible, then they are linearly independent, and we can apply Lemma 6.2.7 to produce an orthonormal basis using the Gram–Schmidt orthonormalization procedure. Recall that we construct vectors Q_k and Q'_k as follows:

$$Q'_1 = A_1, \qquad Q_1 = \frac{Q'_1}{\|Q'_1\|},$$

and for the inductive step

$$Q'_{k+1} = A_{k+1} - \sum_{i=1}^{k} (A_{k+1} \cdot Q_i)\, Q_i, \qquad Q_{k+1} = \frac{Q'_{k+1}}{\|Q'_{k+1}\|},$$

where $1 \leq k \leq n-1$. If we express the vectors A_k in terms of the Q_i and Q'_i, we get the triangular system

$$\begin{aligned}
A_1 &= \|Q'_1\| Q_1, \\
&\cdots \\
A_j &= (A_j \cdot Q_1)\, Q_1 + \cdots + (A_j \cdot Q_i)\, Q_i + \cdots + \|Q'_j\| Q_j, \\
&\cdots \\
A_n &= (A_n \cdot Q_1)\, Q_1 + \cdots + (A_n \cdot Q_{n-1})\, Q_{n-1} + \|Q'_n\| Q_n.
\end{aligned}$$

Letting $r_{k,k} = \|Q'_k\|$, and $r_{i,j} = A_j \cdot Q_i$ (the reversal of i and j on the right-hand side **is** intentional!), where $1 \leq k \leq n$, $2 \leq j \leq n$, and $1 \leq i \leq j-1$, and letting $q_{i,j}$ be the ith component of Q_j, we note that $a_{i,j}$, the ith component of A_j, is given by

$$a_{i,j} = r_{1,j} q_{i,1} + \cdots + r_{i,j} q_{i,i} + \cdots + r_{j,j} q_{i,j} = q_{i,1} r_{1,j} + \cdots + q_{i,i} r_{i,j} + \cdots + q_{i,j} r_{j,j}.$$

If we let $Q = (q_{i,j})$, the matrix whose columns are the components of the Q_j, and $R = (r_{i,j})$, the above equations show that $A = QR$, where R is upper triangular (the reader should work this out on some concrete examples for 2×2 and 3×3 matrices!). The diagonal entries $r_{k,k} = \|Q'_k\| = A_k \cdot Q_k$ are indeed positive. □

Remarks:

(1) Because the diagonal entries of R are positive, it can be shown that Q and R are unique.

(2) The QR-decomposition holds even when A is not invertible. In this case, R has some zero on the diagonal. However, a different proof is needed. We will give a nice proof using Householder matrices (see Lemma 7.3.2, and also Strang [165, 166], Golub and Van Loan [75], Trefethen and Bau [170], Kincaid and Cheney [100], or Ciarlet [33]).

Example 6.6 Consider the matrix

$$A = \begin{pmatrix} 0 & 0 & 5 \\ 0 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

We leave as an exercise to show that $A = QR$, with

$$Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 4 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

Example 6.7 Another example of QR-decomposition is

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

The QR-decomposition yields a rather efficient and numerically stable method for solving systems of linear equations. Indeed, given a system $Ax = b$, where A is an $n \times n$ invertible matrix, writing $A = QR$, since Q is orthogonal, we get

$$Rx = Q^\top b,$$

and since R is upper triangular, we can solve it by Gaussian elimination, by solving for the last variable x_n first, substituting its value into the system, then solving for x_{n-1}, etc. The QR-decomposition is also very useful in solving least squares problems (we will come back to this later on), and for finding eigenvalues. It can be easily adapted to the case where A is a rectangular $m \times n$ matrix with independent columns (thus, $n \leq m$). In this case, Q is not quite orthogonal. It is an $m \times n$ matrix whose columns are orthogonal, and R is an invertible $n \times n$ upper diagonal matrix with positive diagonal entries. For more on QR, see Strang [165, 166], Golub and Van Loan [75], or Trefethen and Bau [170].

It should also be said that the Gram–Schmidt orthonormalization procedure that we have presented is not very stable numerically, and instead, one should use the *modified Gram–Schmidt method*. To compute Q'_{k+1}, instead of projecting A_{k+1} onto $Q_1, \ldots, Q_k$ in a single step, it is better to perform k projections. We compute $Q^1_{k+1}, Q^2_{k+1}, \ldots, Q^k_{k+1}$ as follows:

$$\begin{aligned} Q^1_{k+1} &= A_{k+1} - (A_{k+1} \cdot Q_1)\, Q_1, \\ Q^{i+1}_{k+1} &= Q^i_{k+1} - (Q^i_{k+1} \cdot Q_{i+1})\, Q_{i+1}, \end{aligned}$$

where $1 \leq i \leq k-1$. It is easily shown that $Q'_{k+1} = Q^k_{k+1}$. The reader is urged to code this method.

6.6 Some Applications of Euclidean Geometry

Euclidean geometry has applications in computational geometry, in particular Voronoi diagrams and Delaunay triangulations, discussed in Chapter 9. In turn, Voronoi diagrams have applications in motion planning (see O'Rourke [132]).

Euclidean geometry also has applications to matrix analysis. Recall that a real $n \times n$ matrix A is *symmetric* if it is equal to its transpose $A^\top$. One of the most important properties of symmetric matrices is that they have real eigenvalues and that they can be diagonalized by an orthogonal matrix (see Chapter 11). This means that for every symmetric matrix A, there is a diagonal matrix D and an orthogonal matrix P such that

$$A = PDP^\top.$$

Even though it is not always possible to diagonalize an arbitrary matrix, there are various decompositions involving orthogonal matrices that are of great practical interest. For example, for every real matrix A, there is the *QR-decomposition*, which says that a real matrix A can be expressed as

$$A = QR,$$

where Q is orthogonal and R is an upper triangular matrix. This can be obtained from the Gram–Schmidt orthonormalization procedure, as we saw in Section 6.5, or better, using Householder matrices, as shown in Section 7.3. There is also the *polar decomposition*, which says that a real matrix A can be expressed as

$$A = QS,$$

where Q is orthogonal and S is symmetric positive semidefinite (which means that the eigenvalues of S are nonnegative; see Chapter 11). Such a decomposition is important in continuum mechanics and in robotics, since it separates stretching from rotation. Finally, there is the wonderful *singular value decomposition*, abbreviated as SVD, which says that a real matrix A can be expressed as

$$A = VDU^\top,$$

where U and V are orthogonal and D is a diagonal matrix with nonnegative entries (see Chapter 12). This decomposition leads to the notion of *pseudo-inverse*, which has many applications in engineering (least squares solutions, etc). For an excellent presentation of all these notions, we highly recommend Strang [166, 165], Golub and Van Loan [75], and Trefethen and Bau [170].

The method of least squares, invented by Gauss and Legendre around 1800, is another great application of Euclidean geometry. Roughly speaking, the method is used to solve inconsistent linear systems $Ax = b$, where the number of equations is greater than the number of variables. Since this

is generally impossible, the method of least squares consists in finding a solution x minimizing the Euclidean norm $\|Ax - b\|^2$, that is, the sum of the squares of the "errors." It turns out that there is always a unique solution x^+ of smallest norm minimizing $\|Ax-b\|^2$, and that it is a solution of the square system

$$A^\top Ax = A^\top b,$$

called the system of *normal equations*. The solution x^+ can be found either by using the QR-decomposition in terms of Householder transformations, or by using the notion of pseudo-inverse of a matrix. The pseudo-inverse can be computed using the SVD decomposition. Least squares methods are used extensively in computer vision; see Trucco and Verri [171], or Jain, Katsuri, and Schunck [93]. More details on the method of least squares and pseudo-inverses can be found in Section 13.1.

6.7 Problems

Problem 6.1 Prove Lemma 6.2.2.

Problem 6.2 Prove Lemma 6.2.3.

Problem 6.3 Let $(e_1, \ldots, e_n)$ be an orthonormal basis for E. If X and Y are arbitrary $n \times n$ matrices, denoting as usual the jth column of X by X_j, and similarly for Y, show that

$$X^\top Y = (X_i \cdot Y_j)_{1 \leq i,j \leq n}.$$

Use this to prove that

$$A^\top A = A\, A^\top = I_n$$

iff the column vectors $(A_1, \ldots, A_n)$ form an orthonormal basis. Show that the conditions $A\, A^\top = I_n$, $A^\top A = I_n$, and $A^{-1} = A^\top$ are equivalent.

Problem 6.4 Given any two linear maps $f\colon E \to F$ and $g\colon F \to E$, where $\dim(E) = n$ and $\dim(F) = m$, prove that

$$(-\lambda)^m \det(g \circ f - \lambda\, I_n) = (-\lambda)^n \det(f \circ g - \lambda\, I_m),$$

and thus that $g \circ f$ and $f \circ g$ have the same nonnull eigenvalues.
Hint. If A is an $m \times n$ matrix and B is an $n \times m$ matrix, observe that

$$\begin{vmatrix} AB - X\, I_m & 0_{m,n} \\ B & -X\, I_n \end{vmatrix} = \begin{vmatrix} A & X I_m \\ I_n & 0_{n,m} \end{vmatrix} \begin{vmatrix} B & -X I_n \\ -I_m & A \end{vmatrix}$$

and

$$\begin{vmatrix} A & X I_m \\ I_n & 0_{n,m} \end{vmatrix} \begin{vmatrix} B & -X I_n \\ -I_m & A \end{vmatrix} = \begin{vmatrix} BA - X\, I_n & XB \\ 0_{m,n} & -X\, I_m \end{vmatrix},$$

where X is a variable.

Problem 6.5 (a) Let $\mathcal{C}_1 = (C_1, R_1)$ and $\mathcal{C}_2 = (C_2, R_2)$ be two distinct circles in the plane $\mathbb{E}^2$ (where C_i is the center and R_i is the radius). What is the locus of the centers of all circles tangent to both $\mathcal{C}_1$ and $\mathcal{C}_2$?
Hint. When is it one conic, when is it two conics?

(b) Repeat question (a) in the case where $\mathcal{C}_2$ is a line.

(c) Given three pairwise distinct circles $\mathcal{C}_1 = (C_1, R_1)$, $\mathcal{C}_2 = (C_2, R_2)$, and $\mathcal{C}_3 = (C_3, R_3)$ in the plane $\mathbb{E}^2$, prove that there are at most eight circles simultaneously tangent to $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ (this is known as the *problem of Apollonius*). What happens if the centers C_1, C_2, C_3 of the circles are collinear? In the latter case, show that there are at most two circles exterior and tangent to $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$.
Hint. You may want to use a carefully chosen inversion (see the problems in Section 5.14, especially Problem 5.37).

(d) Prove that the problem of question (c) reduces to the problem of finding the circles passing through a fixed point and tangent to two given circles. In turn, by inversion, this problem reduces to finding all lines tangent to two circles.

(e) Given four pairwise distinct spheres $\mathcal{C}_1 = (C_1, R_1)$, $\mathcal{C}_2 = (C_2, R_2)$, $\mathcal{C}_3 = (C_3, R_3)$, and $\mathcal{C}_4 = (C_4, R_4)$, prove that there are at most sixteen spheres simultaneously tangent to $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_4$. Prove that this problem reduces to the problem of finding the spheres passing through a fixed point and tangent to three given spheres. In turn, by inversion, this problem reduces to finding all planes tangent to three spheres.

Problem 6.6 (a) Given any two circles $\mathcal{C}_1$ and $\mathcal{C}_2$ in $\mathbb{E}^2$ of equations

$$x^2 + y^2 - 2ax - 2by + c = 0 \quad \text{and} \quad x^2 + y^2 - 2a'x - 2'by + c' = 0,$$

we say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are *orthogonal* if they intersect and if the tangents at the intersection points are orthogonal. Prove that $\mathcal{C}_1$ and $\mathcal{C}_2$ are orthogonal iff

$$2(aa' + bb') = c + c'.$$

(b) For any given $c \in \mathbb{R}$ ($c \neq 0$), there is a pencil $\mathcal{F}$ of circles of equations

$$x^2 + y^2 - 2ux - c = 0,$$

where $u \in \mathbb{R}$ is arbitrary. Show that the set of circles orthogonal to all circles in the pencil $\mathcal{F}$ is the pencil $\mathcal{F}^\perp$ of circles of equations

$$x^2 + y^2 - 2vy + c = 0,$$

where $v \in \mathbb{R}$ is arbitrary.

Problem 6.7 Let $P = \{p_1, \ldots, p_n\}$ be a finite set of points in $\mathbb{E}^3$. Show that there is a unique point c such that the sum of the squares of the distances from c to each p_i is minimal. Find this point in terms of the p_i.

Problem 6.8 (1) Compute the real Fourier coefficients of the function $id(x) = x$ over $[-\pi, \pi]$ and prove that

$$x = 2\left(\frac{\sin x}{1} - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \cdots\right).$$

What is the value of the Fourier series at $\pm\pi$? What is the value of the Fourier near $\pm\pi$? Do you find this surprising?

(2) Plot the functions obtained by keeping $1, 2, 4, 5$, and 10 terms. What do you observe around $\pm\pi$?

Problem 6.9 The *Dirac delta function* (which is **not** a function!) is the spike function s.t. $\delta(k2\pi) = +\infty$ for all $k \in \mathbb{Z}$, and $\delta(x) = 0$ everywhere else. It has the property that for "well-behaved" functions f (including constant functions and trigonometric functions),

$$\int_{-\pi}^{+\pi} f(t)\delta(t)dt = f(0).$$

(1) Compute the real Fourier coefficients of δ over $[-\pi, \pi]$, and prove that

$$\delta(x) = \frac{1}{2\pi}\left(1 + 2\cos x + 2\cos 2x + 2\cos 3x + \cdots + 2\cos nx + \cdots\right).$$

Also compute the complex Fourier coefficients of δ over $[-\pi, \pi]$, and prove that

$$\delta(x) = \frac{1}{2\pi}\left(1 + e^{ix} + e^{-ix} + e^{i2x} + e^{-i2x} + \cdots + e^{inx} + e^{-inx} + \cdots\right).$$

(2) Prove that the partial sum of the first $2n + 1$ complex terms is

$$\delta_n(x) = \frac{\sin\left((2n+1)(x/2)\right)}{2\pi \sin\left(x/2\right)}.$$

What is $\delta_n(0)$?

(3) Plot $\delta_n(x)$ for $n = 10, 20$ (over $[-\pi, \pi]$). Prove that the area under the curve δ_n is independent of n. What is it?

Problem 6.10 (1) If an upper triangular $n \times n$ matrix R is invertible, prove that its inverse is also upper triangular.

(2) If an upper triangular matrix is orthogonal, prove that it must be a diagonal matrix.

If A is an invertible $n \times n$ matrix and if $A = Q_1R_1 = Q_2R_2$, where R_1 and R_2 are upper triangular with positive diagonal entries and Q_1, Q_2 are orthogonal, prove that $Q_1 = Q_2$ and $R_1 = R_2$.

Problem 6.11 (1) Review the modified Gram–Schmidt method. Recall that to compute Q'_{k+1}, instead of projecting A_{k+1} onto $Q_1, \ldots, Q_k$ in a single step, it is better to perform k projections. We compute $Q^1_{k+1}, Q^2_{k+1}, \ldots,$ Q^k_{k+1} as follows:

$$Q^1_{k+1} = A_{k+1} - (A_{k+1} \cdot Q_1)\, Q_1,$$

$$Q_{k+1}^{i+1} = Q_{k+1}^{i} - (Q_{k+1}^{i} \cdot Q_{i+1})\, Q_{i+1},$$

where $1 \leq i \leq k-1$.

Prove that $Q'_{k+1} = Q_{k+1}^{k}$.

(2) Write two computer programs to compute the QR-decomposition of an invertible matrix. The first one should use the standard Gram–Schmidt method, and the second one the modified Gram–Schmidt method. Run both on a number of matrices, up to dimension at least 10. Do you observe any difference in their performance in terms of numerical stability?

Run your programs on the Hilbert matrix $H_n = (1/(i+j-1))_{1 \leq i,j \leq n}$. What happens?

Extra Credit. Write a program to solve linear systems of equations $Ax = b$, using your version of the QR-decomposition program, where A is an $n \times n$ matrix.

Problem 6.12 Let E be a Euclidean space of finite dimension n, and let $(e_1, \ldots, e_n)$ be an orthonormal basis for E. For any two vectors $u, v \in E$, the linear map $u \otimes v$ is defined such that

$$u \otimes v(x) = (v \cdot x)\, u,$$

for all $x \in E$. If U and V are the column vectors of coordinates of u and v w.r.t. the basis $(e_1, \ldots, e_n)$, prove that $u \otimes v$ is represented by the matrix

$$U^\top V.$$

What sort of linear map is $u \otimes u$ when u is a unit vector?

Problem 6.13 Let $\varphi \colon E \times E \to \mathbb{R}$ be a bilinear form on a real vector space E of finite dimension n. Given any basis $(e_1, \ldots, e_n)$ of E, let $A = (\alpha_{i\,j})$ be the matrix defined such that

$$\alpha_{i\,j} = \varphi(e_i, e_j),$$

$1 \leq i, j \leq n$. We call A *the matrix of* φ *w.r.t. the basis* $(e_1, \ldots, e_n)$.

(a) For any two vectors x and y, if X and Y denote the column vectors of coordinates of x and y w.r.t. the basis $(e_1, \ldots, e_n)$, prove that

$$\varphi(x, y) = X^\top A Y.$$

(b) Recall that A is a *symmetric* matrix if $A = A^\top$. Prove that φ is symmetric if A is a symmetric matrix.

(c) If $(f_1, \ldots, f_n)$ is another basis of E and P is the change of basis matrix from $(e_1, \ldots, e_n)$ to $(f_1, \ldots, f_n)$, prove that the matrix of φ w.r.t. the basis $(f_1, \ldots, f_n)$ is

$$P^\top A P.$$

The common rank of all matrices representing φ is called the *rank* of φ.

Problem 6.14 Let $\varphi \colon E \times E \to \mathbb{R}$ be a symmetric bilinear form on a real vector space E of finite dimension n. Two vectors x and y are said to be

conjugate w.r.t. φ if $\varphi(x, y) = 0$. The main purpose of this problem is to prove that there is a basis of vectors that are pairwise conjugate w.r.t. φ.

(a) Prove that if $\varphi(x, x) = 0$ for all $x \in E$, then φ is identically null on E.

Otherwise, we can assume that there is some vector $x \in E$ such that $\varphi(x, x) \neq 0$. Use induction to prove that there is a basis of vectors that are pairwise conjugate w.r.t. φ.

For the induction step, proceed as follows. Let $(e_1, e_2, \ldots, e_n)$ be a basis of E, with $\varphi(e_1, e_1) \neq 0$. Prove that there are scalars $\lambda_2, \ldots, \lambda_n$ such that each of the vectors

$$v_i = e_i + \lambda_i e_1$$

is conjugate to e_1 w.r.t. φ, where $2 \leq i \leq n$, and that $(e_1, v_2, \ldots, v_n)$ is a basis.

(b) Let $(e_1, \ldots, e_n)$ be a basis of vectors that are pairwise conjugate w.r.t. φ, and assume that they are ordered such that

$$\varphi(e_i, e_i) = \begin{cases} \theta_i \neq 0 & \text{if } 1 \leq i \leq r, \\ 0 & \text{if } r+1 \leq i \leq n, \end{cases}$$

where r is the rank of φ. Show that the matrix of φ w.r.t. $(e_1, \ldots, e_n)$ is a diagonal matrix, and that

$$\varphi(x, y) = \sum_{i=1}^{r} \theta_i x_i y_i,$$

where $x = \sum_{i=1}^{n} x_i e_i$ and $y = \sum_{i=1}^{n} y_i e_i$.

Prove that for every symmetric matrix A, there is an invertible matrix P such that

$$P^\top A P = D,$$

where D is a diagonal matrix.

(c) Prove that there is an integer p, $0 \leq p \leq r$ (where r is the rank of φ), such that $\varphi(u_i, u_i) > 0$ for exactly p vectors of every basis $(u_1, \ldots, u_n)$ of vectors that are pairwise conjugate w.r.t. φ (*Sylvester's inertia theorem*).

Proceed as follows. Assume that in the basis $(u_1, \ldots, u_n)$, for any $x \in E$, we have

$$\varphi(x, x) = \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2 - \alpha_{p+1} x_{p+1}^2 - \cdots - \alpha_r x_r^2,$$

where $x = \sum_{i=1}^{n} x_i u_i$, and that in the basis $(v_1, \ldots, v_n)$, for any $x \in E$, we have

$$\varphi(x, x) = \beta_1 y_1^2 + \cdots + \beta_q y_q^2 - \beta_{q+1} y_{q+1}^2 - \cdots - \beta_r y_r^2,$$

where $x = \sum_{i=1}^{n} y_i v_i$, with $\alpha_i > 0$, $\beta_i > 0$, $1 \leq i \leq r$.

Assume that $p > q$ and derive a contradiction. First, consider x in the subspace F spanned by

$$(u_1, \ldots, u_p, u_{r+1}, \ldots, u_n),$$

and observe that $\varphi(x,x) \geq 0$ if $x \neq 0$. Next, consider x in the subspace G spanned by

$$(v_{q+1}, \ldots, v_r),$$

and observe that $\varphi(x,x) < 0$ if $x \neq 0$. Prove that $F \cap G$ is nontrivial (i.e., contains some nonnull vector), and derive a contradiction. This implies that $p \leq q$. Finish the proof.

The pair $(p, r-p)$ is called the *signature* of φ.

(d) A symmetric bilinear form φ is *definite* if for every $x \in E$, if $\varphi(x,x) = 0$, then $x = 0$.

Prove that a symmetric bilinear form is definite iff its signature is either $(n,0)$ or $(0,n)$. In other words, a symmetric definite bilinear form has rank n and is either positive or negative.

(e) The *kernel* of a symmetric bilinear form φ is the subspace consisting of the vectors that are conjugate to all vectors in E. We say that a symmetric bilinear form φ is *nondegenerate* if its kernel is trivial (i.e., equal to $\{0\}$).

Prove that a symmetric bilinear form φ is nondegenerate iff its rank is n, the dimension of E. Is a definite symmetric bilinear form φ nondegenerate? What about the converse?

Prove that if φ is nondegenerate, then there is a basis of vectors that are pairwise conjugate w.r.t. φ and such that φ is represented by the matrix

$$\begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}$$

where (p,q) is the signature of φ.

(f) Given a nondegenerate symmetric bilinear form φ on E, prove that for every linear map $f\colon E \to E$, there is a unique linear map $f^*\colon E \to E$ such that

$$\varphi(f(u),\, v) = \varphi(u,\, f^*(v)),$$

for all $u, v \in E$. The map f^* is called the *adjoint of f (w.r.t. to φ)*. Given any basis $(u_1, \ldots, u_n)$, if Ω is the matrix representing φ and A is the matrix representing f, prove that f^* is represented by $\Omega^{-1} A^\top \Omega$.

Prove that Lemma 6.2.4 also holds, i.e., the map $\flat\colon E \to E^*$ is a canonical isomorphism.

A linear map $f\colon E \to E$ is an *isometry w.r.t.* φ if

$$\varphi(f(x),\, f(y)) = \varphi(x,\, y)$$

for all $x, y \in E$. Prove that a linear map f is an isometry w.r.t. φ iff

$$f^* \circ f = f \circ f^* = \mathrm{id}.$$

Prove that the set of isometries w.r.t. φ is a group. This group is denoted by $\mathbf{O}(\varphi)$, and its subgroup consisting of isometries having determinant $+1$ by $\mathbf{SO}(\varphi)$. Given any basis of E, if Ω is the matrix representing φ and A

is the matrix representing f, prove that $f \in \mathbf{O}(\varphi)$ iff

$$A^\top \Omega A = \Omega.$$

Given another nondegenerate symmetric bilinear form ψ on E, we say that φ and ψ are *equivalent* if there is a bijective linear map $h\colon E \to E$ such that

$$\psi(x,\, y) = \varphi(h(x),\, h(y)),$$

for all $x, y \in E$. Prove that the groups of isometries $\mathbf{O}(\varphi)$ and $\mathbf{O}(\psi)$ are isomomorphic (use the map $f \mapsto h \circ f \circ h^{-1}$ from $\mathbf{O}(\psi)$ to $\mathbf{O}(\varphi)$).

If φ is a nondegenerate symmetric bilinear form of signature (p, q), prove that the group $\mathbf{O}(\varphi)$ is isomorphic to the group of $n \times n$ matrices A such that

$$A^\top \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix} A = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}.$$

Remark: In view of question (f), the groups $\mathbf{O}(\varphi)$ and $\mathbf{SO}(\varphi)$ are also denoted by $\mathbf{O}(p, q)$ and $\mathbf{SO}(p, q)$ when φ has signature (p, q). They are Lie groups. In particular, the group $\mathbf{SO}(3, 1)$, known as the *Lorentz group*, plays an important role in the theory of special relativity.

Problem 6.15 (a) Let C be a circle of radius R and center O, and let P be any point in the Euclidean plane $\mathbb{E}^2$. Consider the lines Δ through P that intersect the circle C, generally in two points A and B. Prove that for all such lines,

$$\mathbf{PA} \cdot \mathbf{PB} = \|\mathbf{PO}\|^2 - R^2.$$

Hint. If P is not on C, let B' be the antipodal of B (i.e., $\mathbf{OB'} = -\mathbf{OB}$). Then $\mathbf{AB} \cdot \mathbf{AB'} = 0$ and

$$\mathbf{PA} \cdot \mathbf{PB} = \mathbf{PB'} \cdot \mathbf{PB} = (\mathbf{PO} - \mathbf{OB}) \cdot (\mathbf{PO} + \mathbf{OB}) = \|\mathbf{PO}\|^2 - R^2.$$

The quantity $\|\mathbf{PO}\|^2 - R^2$ is called the *power of P w.r.t. C*, and it is denoted by $\mathcal{P}(P, C)$.

Show that if Δ is tangent to C, then $A = B$ and

$$\|\mathbf{PA}\|^2 = \|\mathbf{PO}\|^2 - R^2.$$

Show that P is inside C iff $\mathcal{P}(P, C) < 0$, on C iff $\mathcal{P}(P, C) = 0$, outside C if $\mathcal{P}(P, C) > 0$.

If the equation of C is

$$x^2 + y^2 - 2ax - 2by + c = 0,$$

prove that the power of $P = (x, y)$ w.r.t. C is given by

$$\mathcal{P}(P, C) = x^2 + y^2 - 2ax - 2by + c.$$

(b) Given two nonconcentric circles C and C', show that the set of points having equal power w.r.t. C and C' is a line orthogonal to the line through the centers of C and C'. If the equations of C and C' are

$$x^2 + y^2 - 2ax - 2by + c = 0 \quad \text{and} \quad x^2 + y^2 - 2a'x - 2b'y + c' = 0,$$

show that the equation of this line is

$$2(a - a')x + 2(b - b')y + c' - c = 0.$$

This line is called the *radical axis* of C and C'.

(c) Given three distinct nonconcentric circles C, C', and C'', prove that either the three pairwise radical axes of these circles are parallel or that they intersect in a single point ω that has equal power w.r.t. C, C', and C''. In the first case, the centers of C, C', and C'' are collinear. In the second case, if the power of ω is positive, prove that ω is the center of a circle Γ orthogonal to C, C', and C'', and if the power of ω is negative, ω is inside C, C', and C''.

(d) Given any $k \in \mathbb{R}$ with $k \neq 0$ and any point a, recall that an *inversion of pole a and power k* is a map $h\colon (\mathbb{E}^n - \{a\}) \to \mathbb{E}^n$ defined such that for every $x \in \mathbb{E}^n - \{a\}$,

$$h(x) = a + k\frac{\mathbf{ax}}{\|\mathbf{ax}\|^2}.$$

For example, when $n = 2$, chosing any orthonormal frame with origin a, h is defined by the map

$$(x, y) \mapsto \left(\frac{kx}{x^2 + y^2}, \frac{ky}{x^2 + y^2}\right).$$

When the centers of C, C' and C'' are not collinear and the power of ω is positive, prove that by a suitable inversion, C, C' and C'' are mapped to three circles whose centers are collinear.

Prove that if three distinct nonconcentric circles C, C', and C'' have collinear centers, then there are at most eight circles simultaneously tangent to C, C', and C'', and at most two for those exterior to C, C', and C''.

(e) Prove that an inversion in $\mathbb{E}^3$ maps a sphere to a sphere or to a plane. Prove that inversions preserve tangency and orthogonality of planes and spheres.