Chapter 12

Euclidean Spaces

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

12.1 Inner Products, Euclidean Spaces

So far the framework of vector spaces allows us to deal with ratios of vectors and linear combinations, but there is no way to express the notion of angle or to talk about orthogonality of vectors. A Euclidean structure allows us to deal with *metric notions* such as angles, orthogonality, and length (or distance).

This chapter covers the bare bones of Euclidean geometry. Deeper aspects of Euclidean geometry are investigated in Chapter 13. 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. 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.

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 prove that every finite-dimensional Euclidean space has orthonormal bases. 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 23.

For a more detailed treatment of Euclidean geometry see Berger [11, 12], Snapper and Troyer [162], or any other book on geometry, such as Pedoe [136], Coxeter [44], Fresnel [65], Tisseron [175], or Cagnac, Ramis, and Commeau [32]. Serious readers should consult Emil

Artin's famous book [6], 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 [84, 85] and Rouché and de Comberousse [139]. The first edition of [84] 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 12.1. A *Euclidean space* is a real vector space E equipped with a symmetric bilinear form $\varphi\colon E\times E \to \mathbb{R}$ that is *positive definite*. More explicitly, $\varphi\colon E\times E \to \mathbb{R}$ 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 $\Phi\colon E \to \mathbb{R}_+$ 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 $\varphi\colon E \times E \to \mathbb{R}$ 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 12.1. The standard example of a Euclidean space is $\mathbb{R}^n$, under the inner product $\cdot$ defined such that

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

This Euclidean space is denoted by $\mathbb{E}^n$.

There are other examples.

Example 12.2. 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

$$\varphi(x_1 e_1 + y_1 e_2,\, x_2 e_1 + y_2 e_2) = a x_1 x_2 + b(x_1 y_2 + x_2 y_1) + c y_1 y_2$$

yields a Euclidean structure on E. In this case,

$$\Phi(x e_1 + y e_2) = a x^2 + 2bxy + c y^2.$$

Example 12.3. Let $\mathcal{C}[a,b]$ denote the set of continuous functions $f\colon [a,b] \to \mathbb{R}$. 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 it 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!

Example 12.4. Let $E = \mathrm{M}_n(\mathbb{R})$ be the vector space of real $n \times n$ matrices. If we view a matrix $A \in \mathrm{M}_n(\mathbb{R})$ as a "long" column vector obtained by concatenating together its columns, we can define the inner product of two matrices $A, B \in \mathrm{M}_n(\mathbb{R})$ as

$$\langle A, B \rangle = \sum_{i,j=1}^{n} a_{ij} b_{ij},$$

which can be conveniently written as

$$\langle A, B \rangle = \mathrm{tr}(A^\top B) = \mathrm{tr}(B^\top A).$$

Since this can be viewed as the Euclidean product on $\mathbb{R}^{n^2}$, it is an inner product on $\mathrm{M}_n(\mathbb{R})$. The corresponding norm

$$\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$$

is the Frobenius norm (see Section 9.2).

Let us observe that φ can be recovered from Φ.

Proposition 12.1. *We have*

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

for all $u, v \in E$. We say that φ is the **polar form of** Φ.

Proof. By bilinearity and symmetry, we have

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

□

If E is finite-dimensional and if $\varphi\colon E \times E \to \mathbb{R}$ is a bilinear form on E, given any basis $(e_1, \ldots, e_n)$ of E, we can write $x = \sum_{i=1}^n x_i e_i$ and $y = \sum_{j=1}^n y_j e_j$, and we have

$$\varphi(x, y) = \varphi\left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\right) = \sum_{i,j=1}^n x_i y_j \varphi(e_i, e_j).$$

If we let G be the matrix $G = (\varphi(e_i, e_j))$, and if x and y are the column vectors associated with $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, then we can write

$$\varphi(x, y) = x^\top G y = y^\top G^\top x.$$

Note that we are committing an abuse of notation since $x = \sum_{i=1}^n x_i e_i$ is a vector in E, but the column vector associated with $(x_1, \ldots, x_n)$ belongs to $\mathbb{R}^n$. To avoid this minor abuse, we could denote the column vector associated with $(x_1, \ldots, x_n)$ by $\mathbf{x}$ (and similarly $\mathbf{y}$ for the column vector associated with $(y_1, \ldots, y_n)$), in wich case the "correct" expression for $\varphi(x, y)$ is

$$\varphi(x, y) = \mathbf{x}^\top G \mathbf{y}.$$

However, in view of the isomorphism between E and $\mathbb{R}^n$, to keep notation as simple as possible, we will use x and y instead of $\mathbf{x}$ and $\mathbf{y}$.

Also observe that φ is symmetric iff $G = G^\top$, and φ is positive definite iff the matrix G is positive definite, that is,

$$x^\top G x > 0 \quad \text{for all } x \in \mathbb{R}^n,\ x \neq 0.$$

The matrix G associated with an inner product is called the *Gram matrix* of the inner product with respect to the basis $(e_1, \ldots, e_n)$.

Conversely, if A is a symmetric positive definite $n \times n$ matrix, it is easy to check that the bilinear form

$$\langle x, y\rangle = x^\top A y$$

is an inner product. If we make a change of basis from the basis $(e_1, \ldots, e_n)$ to the basis $(f_1, \ldots, f_n)$, and if the change of basis matrix is P (where the jth column of P consists of the coordinates of f_j over the basis $(e_1, \ldots, e_n)$), then with respect to coordinates x' and y' over the basis $(f_1, \ldots, f_n)$, we have

$$x^\top G y = x'^\top P^\top G P y',$$

so the matrix of our inner product over the basis $(f_1, \ldots, f_n)$ is $P^\top G P$. We summarize these facts in the following proposition.

Proposition 12.2. *Let E be a finite-dimensional vector space, and let $(e_1, \ldots, e_n)$ be a basis of E.*

1. *For any inner product $\langle -, - \rangle$ on E, if $G = (\langle e_i, e_j \rangle)$ is the Gram matrix of the inner product $\langle -, - \rangle$ w.r.t. the basis $(e_1, \ldots, e_n)$, then G is symmetric positive definite.*

2. *For any change of basis matrix P, the Gram matrix of $\langle -, - \rangle$ with respect to the new basis is $P^\top G P$.*

3. *If A is any $n \times n$ symmetric positive definite matrix, then*

$$\langle x, y \rangle = x^\top A y$$

is an inner product on E.

We will see later that a symmetric matrix is positive definite iff its eigenvalues are all positive.

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

Proposition 12.3. *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

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

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

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

Let us now consider the case where we have the equality

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

There are two cases. If $\Phi(v) = 0$, then $v = 0$ and u and v are linearly dependent. If $\Phi(v) \neq 0$, then the above quadratic equation has a double root λ_0, and we have $\Phi(u + \lambda_0 v) = 0$. Since φ is positive definite, $\Phi(u + \lambda_0 v) = 0$ implies that $u + \lambda_0 v = 0$, which shows that 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

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

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 9.1, and since φ is bilinear and positive definite, it also satisfies conditions (N1) and (N2) of Definition 9.1, and thus it is a *norm* on E. The norm induced by φ is called the *Euclidean norm induced by* φ.

The Cauchy–Schwarz inequality can be written as

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

and the Minkowski inequality as

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

If u and v are nonzero vectors then the Cauchy–Schwarz inequality implies that

$$-1 \leq \frac{u \cdot v}{\|u\| \, \|v\|} \leq +1.$$

Then there is a unique $\theta \in [0, \pi]$ such that

$$\cos \theta = \frac{u \cdot v}{\|u\| \, \|v\|}.$$

We have $u = v$ iff $\theta = 0$ and $u = -v$ iff $\theta = \pi$. For $0 < \theta < \pi$, the vectors u and v are linearly independent and there is an orientation of the plane spanned by u and v such that θ is the angle between u and v. See Problem 12.8 for the precise notion of orientation. If u is a unit vector (which means that $\|u\| = 1$), then the vector

$$(\|v\| \cos \theta) u = (u \cdot v) u = (v \cdot u) u$$

is called the *orthogonal projection* of v onto the space spanned by u.

Remark: One might wonder if every norm on a vector space is induced by some Euclidean inner product. In general this is false, but remarkably, there is a simple necessary and sufficient condition, which is that the norm must satisfy the *parallelogram law*:

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

See Figure 12.1.

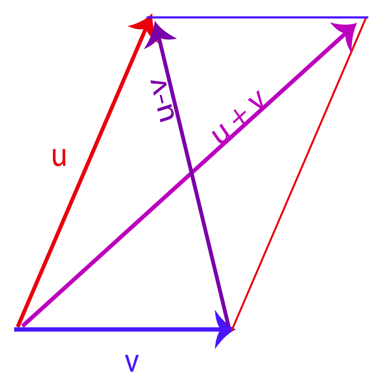

Figure 12.1: The parallelogram law states that the sum of the lengths of the diagonals of the parallelogram determined by vectors u and v equals the sum of all the sides.

If $\langle -, - \rangle$ is an inner product, then we have

$$\begin{aligned} \|u + v\|^2 &= \|u\|^2 + \|v\|^2 + 2\langle u, v \rangle \\ \|u - v\|^2 &= \|u\|^2 + \|v\|^2 - 2\langle u, v \rangle, \end{aligned}$$

and by adding and subtracting these identities, we get the parallelogram law and the equation

$$\langle u, v\rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2),$$

which allows us to recover $\langle -, -\rangle$ from the norm.

Conversely, if $\| \ \|$ is a norm satisfying the parallelogram law, and if it comes from an inner product, then this inner product must be given by

$$\langle u, v\rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2).$$

We need to prove that the above form is indeed symmetric and bilinear.

Symmetry holds because $\|u-v\| = \|-(u-v)\| = \|v-u\|$. Let us prove additivity in the variable u. By the parallelogram law, we have

$$2(\|x+z\|^2 + \|y\|^2) = \|x+y+z\|^2 + \|x-y+z\|^2$$

which yields

$$\begin{aligned}
\|x+y+z\|^2 &= 2(\|x+z\|^2 + \|y\|^2) - \|x-y+z\|^2\\
\|x+y+z\|^2 &= 2(\|y+z\|^2 + \|x\|^2) - \|y-x+z\|^2,
\end{aligned}$$

where the second formula is obtained by swapping x and y. Then by adding up these equations, we get

$$\|x+y+z\|^2 = \|x\|^2 + \|y\|^2 + \|x+z\|^2 + \|y+z\|^2 - \frac{1}{2}\|x-y+z\|^2 - \frac{1}{2}\|y-x+z\|^2.$$

Replacing z by $-z$ in the above equation, we get

$$\|x+y-z\|^2 = \|x\|^2 + \|y\|^2 + \|x-z\|^2 + \|y-z\|^2 - \frac{1}{2}\|x-y-z\|^2 - \frac{1}{2}\|y-x-z\|^2,$$

Since $\|x-y+z\| = \|-(x-y+z)\| = \|y-x-z\|$ and $\|y-x+z\| = \|-(y-x+z)\| = \|x-y-z\|$, by subtracting the last two equations, we get

$$\begin{aligned}
\langle x+y, z\rangle &= \frac{1}{4}(\|x+y+z\|^2 - \|x+y-z\|^2)\\
&= \frac{1}{4}(\|x+z\|^2 - \|x-z\|^2) + \frac{1}{4}(\|y+z\|^2 - \|y-z\|^2)\\
&= \langle x, z\rangle + \langle y, z\rangle,
\end{aligned}$$

as desired.

Proving that

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle \quad \text{for all } \lambda \in \mathbb{R}$$

is a little tricky. The strategy is to prove the identity for $\lambda \in \mathbb{Z}$, then to promote it to $\mathbb{Q}$, and then to $\mathbb{R}$ by continuity.

Since

$$\begin{aligned}\langle -u, v\rangle &= \frac{1}{4}(\|-u+v\|^2 - \|-u-v\|^2)\\ &= \frac{1}{4}(\|u-v\|^2 - \|u+v\|^2)\\ &= -\langle u, v\rangle,\end{aligned}$$

the property holds for $\lambda = -1$. By linearity and by induction, for any $n \in \mathbb{N}$ with $n \geq 1$, writing $n = n - 1 + 1$, we get

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle \quad \text{for all } \lambda \in \mathbb{N},$$

and since the above also holds for $\lambda = -1$, it holds for all $\lambda \in \mathbb{Z}$. For $\lambda = p/q$ with $p, q \in \mathbb{Z}$ and $q \neq 0$, we have

$$q\langle (p/q)u, v\rangle = \langle pu, v\rangle = p\langle u, v\rangle,$$

which shows that

$$\langle (p/q)u, v\rangle = (p/q)\langle u, v\rangle,$$

and thus

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle \quad \text{for all } \lambda \in \mathbb{Q}.$$

To finish the proof, we use the fact that a norm is a continuous map $x \mapsto \|x\|$. Then, the continuous function $t \mapsto \frac{1}{t}\langle tu, v\rangle$ defined on $\mathbb{R} - \{0\}$ agrees with $\langle u, v\rangle$ on $\mathbb{Q} - \{0\}$, so it is equal to $\langle u, v\rangle$ on $\mathbb{R} - \{0\}$. The case $\lambda = 0$ is trivial, so we are done.

We now define orthogonality.

12.2 Orthogonality and Duality in Euclidean Spaces

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 12.2. 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 12.5. Going back to Example 12.3 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 $\sqrt{2\pi}$.

Proposition 12.4. *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.*

Proof. Assume there is a linear dependence

$$\sum_{j \in J} \lambda_j u_j = 0$$

for some $\lambda_j \in \mathbb{R}$ and some finite subset J of I. By taking the inner product with u_i for any $i \in J$, and using the the bilinearity of the inner product and the fact that $u_i \cdot u_j = 0$ whenever $i \neq j$, we get

$$\begin{aligned} 0 = u_i \cdot 0 = u_i \cdot \left(\sum_{j \in J} \lambda_j u_j\right) \\ = \sum_{j \in J} \lambda_j (u_i \cdot u_j) = \lambda_i (u_i \cdot u_i), \end{aligned}$$

so

$$\lambda_i (u_i \cdot u_i) = 0, \qquad \text{for all } i \in J,$$

and since $u_i \neq 0$ and an inner product is positive definite, $u_i \cdot u_i \neq 0$, so we obtain

$$\lambda_i = 0, \qquad \text{for all } i \in J,$$

which shows that the family $(u_i)_{i \in I}$ is linearly independent. $\square$

We leave the following simple result as an exercise.

Proposition 12.5. *Given a Euclidean space E, any two vectors $u, v \in E$ are orthogonal iff*

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

See Figure 12.2 for a geometrical interpretation.

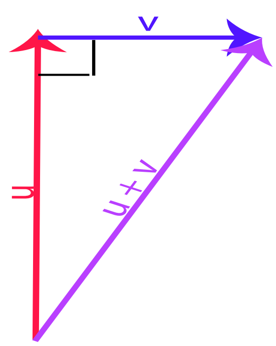

Figure 12.2: The sum of the lengths of the two sides of a right triangle is equal to the length of the hypotenuse; i.e. the Pythagorean theorem.

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. See Figure 12.3. If the basis is orthogonal but not necessarily

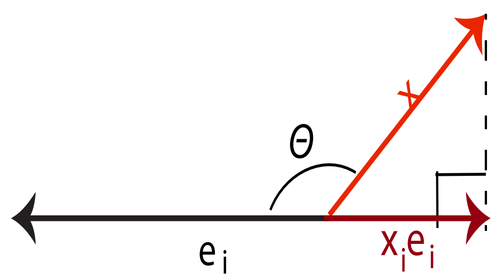

Figure 12.3: The orthogonal projection of the red vector x onto the black basis vector e_i is the maroon vector $x_i e_i$. Observe that $x \cdot e_i = \|x\| \cos \theta$.

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.

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^*. The reason is that an inner product $\cdot\colon E \times E \to \mathbb{R}$ defines a nondegenerate pairing, as defined in Definition 11.4. Indeed, if $u \cdot v = 0$ for all $v \in E$ then $u = 0$, and similarly if $u \cdot v = 0$ for all $u \in E$ then $v = 0$ (since an inner product is positive definite and symmetric). By Proposition 11.6, there is a canonical isomorphism between E and E^*. We feel that the reader will appreciate if we exhibit this mapping explicitly and reprove that it is an isomorphism.

The mapping from E to E^* is defined as follows.

Definition 12.3. For any vector $u \in E$, let $\varphi_u\colon E \to \mathbb{R}$ be the map defined such that

$$\varphi_u(v) = u \cdot v, \quad \text{for all } v \in E.$$

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

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

Theorem 12.6. *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 Theorem 12.6 we have the following corollary.

Corollary 12.7. *If E is a Euclidean space of finite dimension, every linear form $f \in E^*$ corresponds to a unique $u \in E$ such that*

$$f(v) = u \cdot v, \quad \text{for every } v \in E.$$

In particular, if f is not the zero 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) Theorem 12.6 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 $\mathbb{R}^4$ (the Lorentz form) 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*.

Example 12.6. Consider $\mathbb{R}^n$ with its usual Euclidean inner product. Given any differentiable function $f\colon U \to \mathbb{R}$, where U is some open subset of $\mathbb{R}^n$, by definition, for any $x \in U$, the *total derivative* df_x of f at x is the linear form defined so that for all $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$,

$$df_x(u) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)\, u_i.$$

The unique vector $v \in \mathbb{R}^n$ such that

$$v \cdot u = df_x(u) \quad \text{for all } u \in \mathbb{R}^n$$

is the transpose of the *Jacobian matrix* of f at x, the $1 \times n$ matrix

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

This is the *gradient* $\mathrm{grad}(f)_x$ of f at x, given by

$$\mathrm{grad}(f)_x = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Example 12.7. Given any two vectors $u, v \in \mathbb{R}^3$, let $c(u, v)$ be the linear form given by

$$c(u, v)(w) = \det(u, v, w) \quad \text{for all } w \in \mathbb{R}^3.$$

Since

$$\begin{aligned} \det(u, v, w) = \begin{vmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{vmatrix} &= w_1 \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} - w_2 \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} + w_3 \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \\ &= w_1(u_2v_3 - u_3v_2) + w_2(u_3v_1 - u_1v_3) + w_3(u_1v_2 - u_2v_1), \end{aligned}$$

we see that the unique vector $z \in \mathbb{R}^3$ such that

$$z \cdot w = c(u, v)(w) = \det(u, v, w) \quad \text{for all } w \in \mathbb{R}^3$$

is the vector

$$z = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

This is just the *cross-product* $u \times v$ of u and v. Since $\det(u, v, u) = \det(u, v, v) = 0$, we see that $u \times v$ is orthogonal to both u and v. The above allows us to generalize the cross-product to $\mathbb{R}^n$. Given any $n - 1$ vectors $u_1, \ldots, u_{n-1} \in \mathbb{R}^n$, the cross-product $u_1 \times \cdots \times u_{n-1}$ is the unique vector in $\mathbb{R}^n$ such that

$$(u_1 \times \cdots \times u_{n-1}) \cdot w = \det(u_1, \ldots, u_{n-1}, w) \quad \text{for all } w \in \mathbb{R}^n.$$

Example 12.8. Consider the vector space $\mathrm{M}_n(\mathbb{R})$ of real $n \times n$ matrices with the inner product

$$\langle A, B\rangle = \mathrm{tr}(A^\top B).$$

Let $s\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ be the function given by

$$s(A) = \sum_{i,j=1}^{n} a_{ij},$$

where $A = (a_{ij})$. It is immediately verified that s is a linear form. It is easy to check that the unique matrix Z such that

$$\langle Z, A\rangle = s(A) \quad \text{for all } A \in \mathrm{M}_n(\mathbb{R})$$

is the matrix $Z = \mathbf{ones}(n, n)$ whose entries are all equal to 1.

12.3 Adjoint of a Linear Map

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 [169]).

Let E be a Euclidean space of finite dimension n, and let $f\colon E \to E$ 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 Theorem 12.6, 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 proposition shows that the map f^* is linear.

Proposition 12.8. *Given a Euclidean space E of finite dimension, for every linear map $f\colon E \to E$, there is a unique linear map $f^*\colon E \to E$ such that*

$$f^*(u) \cdot v = u \cdot f(v), \quad \textit{for all } u, v \in E.$$

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,

$$f^*(u_1) \cdot v = u_1 \cdot f(v) \quad \text{and} \quad f^*(u_2) \cdot v = u_2 \cdot f(v),$$

for all $v \in E$. Thus we get

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

for all $v \in E$. Since our inner product is positive definite, 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

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

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. $\square$

Definition 12.4. Given a Euclidean space E of finite dimension, for every linear map $f\colon E \to E$, the unique linear map $f^*\colon E \to E$ such that

$$f^*(u) \cdot v = u \cdot f(v), \quad \text{for all } u, v \in E$$

given by Proposition 12.8 is called the *adjoint of f (w.r.t. to the inner product)*. Linear maps $f\colon E \to E$ such that $f = f^*$ are called *self-adjoint* maps.

Self-adjoint linear maps 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: Proposition 12.8 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 $f\colon E \to F$, it is immediately verified that the proof of Proposition 12.8 can be adapted to show that there is a unique linear map $f^*\colon F \to E$ 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.

The following properties immediately follow from the definition of the adjoint map:

(1) For any linear map $f\colon E \to F$, we have

$$f^{**} = f.$$

(2) For any two linear maps $f, g\colon E \to F$ and any scalar $\lambda \in \mathbb{R}$:

$$\begin{aligned}(f+g)^* &= f^* + g^*\\ (\lambda f)^* &= \lambda f^*.\end{aligned}$$

(3) If E, F, G are Euclidean spaces with respective inner products $\langle -,-\rangle_1, \langle -,-\rangle_2$, and $\langle -,-\rangle_3$, and if $f\colon E \to F$ and $g\colon F \to G$ are two linear maps, then

$$(g \circ f)^* = f^* \circ g^*.$$

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 Gram matrices defining the inner products; see Problem 12.5. We will do so with respect to orthonormal bases in Proposition 12.14(2). 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 $u, v \in E$.

12.4 Existence and Construction of Orthonormal Bases

We can also use Theorem 12.6 to show that any Euclidean space of finite dimension has an orthonormal basis.

Proposition 12.9. *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 Theorem 12.6, 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. $\square$

As a consequence of Proposition 12.9, 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 $u = u_1e_1 + \cdots + u_ne_n$ and $v = v_1e_1 + \cdots + v_ne_n$, the inner product $u \cdot v$ is expressed as

$$u \cdot v = (u_1e_1 + \cdots + u_ne_n) \cdot (v_1e_1 + \cdots + v_ne_n) = \sum_{i=1}^{n} u_iv_i,$$

and the norm $\|u\|$ as

$$\|u\| = \|u_1e_1 + \cdots + u_ne_n\| = \left(\sum_{i=1}^{n} u_i^2\right)^{1/2}.$$

The fact that a Euclidean space always has an orthonormal basis implies that any Gram matrix G can be written as

$$G = Q^\top Q,$$

for some invertible matrix Q. Indeed, we know that in a change of basis matrix, a Gram matrix G becomes $G' = P^{\top}GP$. If the basis corresponding to G' is orthonormal, then $G' = I$, so $G = (P^{-1})^{\top}P^{-1}$.

There is a more constructive way of proving Proposition 12.9, using a procedure known as the *Gram–Schmidt orthonormalization procedure*. Among other things, the Gram–Schmidt orthonormalization procedure yields the *QR-decomposition for matrices*, an important tool in numerical methods.

Proposition 12.10. *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 proposition. $\square$

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.

Example 12.9. For a specific example of this procedure, let $E = \mathbb{R}^3$ with the standard Euclidean norm. Take the basis

$$e_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad e_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Then

$$u_1 = 1/\sqrt{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

and

$$u'_2 = e_2 - (e_2 \cdot u_1)u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - 2/3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1/3 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

This implies that

$$u_2 = 1/\sqrt{6} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

and that

$$u'_3 = e_3 - (e_3 \cdot u_1)u_1 - (e_3 \cdot u_2)u_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - 2/3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1/6 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 1/2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

To complete the orthonormal basis, normalize u'_3 to obtain

$$u_3 = 1/\sqrt{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

An illustration of this example is provided by Figure 12.4.

Remarks:

(1) The QR-decomposition can now be obtained very easily, but we postpone this until Section 12.8.

(2) The proof of Proposition 12.10 also works for a countably infinite basis for E, producing a countably infinite orthonormal basis.

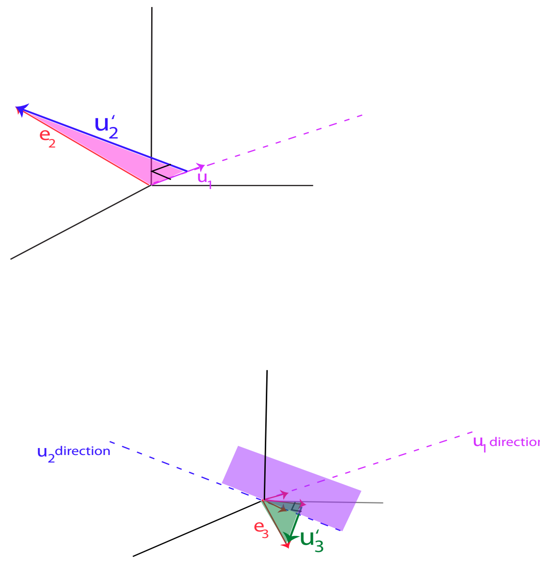

Figure 12.4: The top figure shows the construction of the blue u_2' as perpendicular to the orthogonal projection of e_2 onto u_1, while the bottom figure shows the construction of the green u_3' as normal to the plane determined by u_1 and u_2.

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 u_{k+1}', instead of projecting e_{k+1} onto $u_1, \ldots, u_k$ in a single step, it is better to perform k projections. We compute $u_1^{k+1}, u_2^{k+1}, \ldots, u_k^{k+1}$ as follows:

$$\begin{aligned} u_1^{k+1} &= e_{k+1} - (e_{k+1} \cdot u_1)\, u_1, \\ u_{i+1}^{k+1} &= u_i^{k+1} - (u_i^{k+1} \cdot u_{i+1})\, u_{i+1}, \end{aligned}$$

where $1 \leq i \leq k-1$. It is easily shown that $u_{k+1}' = u_k^{k+1}$.

Example 12.10. Let us apply the modified Gram–Schmidt method to the (e_1, e_2, e_3) basis of Example 12.9. The only change is the computation of u_3'. For the modified Gram–Schmidt procedure, we first calculate

$$u_1^3 = e_3 - (e_3 \cdot u_1)u_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - 2/3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1/3 \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

Then

$$u_2^3 = u_1^3 - (u_1^3 \cdot u_2)u_2 = 1/3 \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} + 1/6 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 1/2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

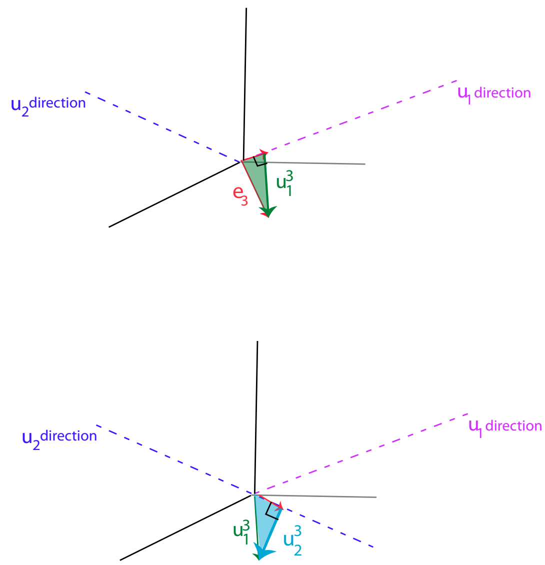

Figure 12.5: The top figure shows the construction of the blue u_1^3 as perpendicular to the orthogonal projection of e_3 onto u_1, while the bottom figure shows the construction of the sky blue u_2^3 as perpendicular to the orthogonal projection of u_1^3 onto u_2.

and observe that $u_2^3 = u'_3$. See Figure 12.5.

The following Matlab program implements the modified Gram–Schmidt procedure.

```
function q = gramschmidt4(e)
n = size(e,1);
for i = 1:n
    q(:,i) = e(:,i);
    for j = 1:i-1
        r = q(:,j)'*q(:,i);
        q(:,i) = q(:,i) - r*q(:,j);
    end
    r = sqrt(q(:,i)'*q(:,i));
    q(:,i) = q(:,i)/r;
end
end
```

If we apply the above function to the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

the ouput is the matrix

$$\begin{pmatrix} 0.5774 & 0.4082 & 0.7071 \\ 0.5774 & -0.8165 & -0.0000 \\ 0.5774 & 0.4082 & -0.7071 \end{pmatrix},$$

which matches the result of Example 12.9.

Example 12.11. 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:

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

Here is an explicit summation for $P_n(x)$:

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \binom{n}{k} \binom{2n-2k}{n} x^{n-2k}.$$

The polynomials Q_n are related to the Legendre polynomials P_n as follows:

$$Q_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x).$$

Example 12.12. Consider polynomials over $[-1, 1]$, with the symmetric bilinear form

$$\langle f, g\rangle = \int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} f(t)g(t)dt.$$

We leave it as an exercise to prove that the above defines an inner product. It can be shown that the polynomials $T_n(x)$ given by

$$T_n(x) = \cos(n \arccos x), \quad n \geq 0,$$

(equivalently, with $x = \cos\theta$, we have $T_n(\cos\theta) = \cos(n\theta)$) are orthogonal with respect to the above inner product. These polynomials are the *Chebyshev polynomials*. Their norm is not equal to 1. Instead, we have

$$\langle T_n, T_n\rangle = \begin{cases} \frac{\pi}{2} & \text{if } n > 0, \\ \pi & \text{if } n = 0. \end{cases}$$

Using the identity $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$ and the binomial formula, we obtain the following expression for $T_n(x)$:

$$T_n(x) = \sum_{k=0}^{\lfloor n/2\rfloor} \binom{n}{2k} (x^2-1)^k x^{n-2k}.$$

The Chebyshev polynomials are defined inductively as follows:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x), \quad n \geq 1. \end{aligned}$$

Using these recurrence equations, we can show that

$$T_n(x) = \frac{(x - \sqrt{x^2-1})^n + (x + \sqrt{x^2-1})^n}{2}.$$

The polynomial T_n has n distinct roots in the interval $[-1, 1]$. The Chebyshev polynomials play an important role in approximation theory. They are used as an approximation to a best polynomial approximation of a continuous function under the sup-norm (∞-norm).

The inner products of the last two examples are special cases of an inner product of the form

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

where $W(t)$ is a *weight function*. If W is a continuous function such that $W(x) > 0$ on $(-1, 1)$, then the above bilinear form is indeed positive definite. Families of orthogonal polynomials used in approximation theory and in physics arise by a suitable choice of the weight function W. Besides the previous two examples, the *Hermite polynomials* correspond to $W(x) = e^{-x^2}$, the *Laguerre polynomials* to $W(x) = e^{-x}$, and the *Jacobi polynomials* to $W(x) = (1-x)^\alpha(1+x)^\beta$, with $\alpha, \beta > -1$. Comprehensive treatments of orthogonal polynomials can be found in Lebedev [114], Sansone [144], and Andrews, Askey and Roy [3].

We can also prove the following proposition regarding orthogonal spaces.

Proposition 12.11. *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 Proposition 12.9, 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 Proposition 12.10, 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 $w = w_1u_1 + \cdots + w_nu_n \in E$ 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 $w = w_1u_1 + \cdots + w_nu_n \in E$ 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 $F^{\perp\perp} = F$. $\square$

12.5 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 12.5. Given any two nontrivial Euclidean spaces E and F of the same finite dimension n, a function $f\colon E \to F$ is *an orthogonal transformation, or a linear isometry*, if it is linear and

$$\|f(u)\| = \|u\|, \quad \text{for all } u \in E.$$

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 12.5 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 $f\colon E \to F$ (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 proposition shows that all these definitions are equivalent when E and F are of finite dimension, and for functions such that $f(0) = 0$.

Proposition 12.12. *Given any two nontrivial Euclidean spaces E and F of the same finite dimension n, for every function $f\colon E \to F$, 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 $g\colon E \to F$ defined such that

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

for all $u \in E$ is a map satisfying Condition (2), and that (2) implies (3). 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 since $(e_1, \ldots, e_n)$ and $(f(e_1), \ldots, f(e_n))$ are orthonormal bases, for any $u \in E$ we have

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

and

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

and since f preserves inner products, this shows that

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

which proves 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. $\square$

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) The implication that (3) implies (1) holds if we also assume that f is surjective, even if E has infinite dimension.

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

$$f(\tau + u) = f(\tau) + g(u) \quad \text{for all } u \in E.$$

By Proposition 24.7, this shows that f is affine with associated linear map g.

This fact is worth recording as the following proposition.

Proposition 12.13. *Given any two nontrivial Euclidean spaces E and F of the same finite dimension n, for every function $f\colon E \to F$, if*

$$\|f(v) - f(u)\| = \|v - u\| \quad \textit{for all } u, v \in E,$$

then f is an affine map, and its associated linear map g is an isometry.

In view of Proposition 12.12, we usually abbreviate "linear isometry" as "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.

12.6 The Orthogonal Group, Orthogonal Matrices

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

Proposition 12.14. *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

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

where I_n denotes the identity matrix of order n, iff the columns of A form an orthonormal basis of $\mathbb{R}^n$, iff the rows of A form an orthonormal basis of $\mathbb{R}^n$.

Proof. (1) The linear map $f\colon E \to E$ 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,

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

But an endomorphism f of a finite-dimensional vector space that has a left inverse is an isomorphism, so $f \circ f^* = \mathrm{id}$. The converse is established by doing the above steps backward.

(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 $A^\top$). $\square$

Proposition 12.14 shows that the inverse of an isometry f is its adjoint f^*. Recall that the set of all real $n \times n$ matrices is denoted by $\mathrm{M}_n(\mathbb{R})$. Proposition 12.14 also motivates the following definition.

Definition 12.6. 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 $(v_1, \ldots, v_n)$, since the columns of P are the coordinates of the vectors v_j with respect to the basis $(u_1, \ldots, u_n)$, if $v_{j_1} = \sum_{i_1=1}^n p_{i_1 j_1} u_{i_1}$ and $v_{j_2} = \sum_{i_2=1}^n p_{i_2 j_2} u_{i_2}$, since $(u_1, \ldots, u_n)$ is orthonormal,

$$v_{j_1} \cdot v_{j_2} = \sum_{i_1=1}^n \sum_{i_2=1}^n p_{i_1 j_1} p_{i_2 j_2} (u_{i_1} \cdot u_{i_2}) = \sum_{i=1}^n p_{i j_1} p_{i j_2},$$

and since $(v_1, \ldots, v_n)$ is orthonormal, $v_{j_1} \cdot v_{j_2} = \delta_{j_1\, j_2}$, so the columns of P are orthonormal, and by Proposition 12.14 (2), the matrix P is orthogonal.

The proof of Proposition 12.12 (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 $f \colon E \to E$ 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 $\det(AB) = \det(A)\det(B)$. 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 12.7. Given a Euclidean space E of dimension n, the set of isometries $f \colon E \to E$ 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*.

12.7 The Rodrigues Formula

When $n = 3$ and A is a skew symmetric matrix, it is possible to work out an explicit formula for e^A. For any 3×3 real skew symmetric matrix

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

if we let $\theta = \sqrt{a^2 + b^2 + c^2}$ and

$$B = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix},$$

then we have the following result known as *Rodrigues' formula* (1840). The (real) vector space of $n \times n$ skew symmetric matrices is denoted by $\mathfrak{so}(n)$.

Proposition 12.15. *The exponential map* $\exp \colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ *is given by*

$$e^A = \cos\theta\, I_3 + \frac{\sin\theta}{\theta} A + \frac{(1 - \cos\theta)}{\theta^2} B,$$

or, equivalently, by

$$e^A = I_3 + \frac{\sin\theta}{\theta} A + \frac{(1 - \cos\theta)}{\theta^2} A^2$$

if $\theta \neq 0$, *with* $e^{0_3} = I_3$.

Proof sketch. First observe that

$$A^2 = -\theta^2 I_3 + B,$$

since

$$\begin{aligned} A^2 &= \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} = \begin{pmatrix} -c^2 - b^2 & ba & ca \\ ab & -c^2 - a^2 & cb \\ ac & cb & -b^2 - a^2 \end{pmatrix} \\ &= \begin{pmatrix} -a^2 - b^2 - c^2 & 0 & 0 \\ 0 & -a^2 - b^2 - c^2 & 0 \\ 0 & 0 & -a^2 - b^2 - c^2 \end{pmatrix} + \begin{pmatrix} a^2 & ba & ca \\ ab & b^2 & cb \\ ac & cb & c^2 \end{pmatrix} \\ &= -\theta^2 I_3 + B, \end{aligned}$$

and that

$$AB = BA = 0.$$

From the above, deduce that

$$A^3 = -\theta^2 A,$$

and for any $k \geq 0$,

$$\begin{aligned} A^{4k+1} &= \theta^{4k} A, \\ A^{4k+2} &= \theta^{4k} A^2, \\ A^{4k+3} &= -\theta^{4k+2} A, \\ A^{4k+4} &= -\theta^{4k+2} A^2. \end{aligned}$$

Then prove the desired result by writing the power series for e^A and regrouping terms so that the power series for $\cos\theta$ and $\sin\theta$ show up. In particular

$$\begin{aligned}
e^A = I_3 + \sum_{p\geq 1}\frac{A^p}{p!} &= I_3 + \sum_{p\geq 0}\frac{A^{2p+1}}{(2p+1)!} + \sum_{p\geq 1}\frac{A^{2p}}{(2p)!} \\
&= I_3 + \sum_{p\geq 0}\frac{(-1)^p\theta^{2p}}{(2p+1)!}A + \sum_{p\geq 1}\frac{(-1)^{p-1}\theta^{2(p-1)}}{(2p)!}A^2 \\
&= I_3 + \frac{A}{\theta}\sum_{p\geq 0}\frac{(-1)^p\theta^{2p+1}}{(2p+1)!} - \frac{A^2}{\theta^2}\sum_{p\geq 1}\frac{(-1)^p\theta^{2p}}{(2p)!} \\
&= I_3 + \frac{\sin\theta}{\theta}A - \frac{A^2}{\theta^2}\sum_{p\geq 0}\frac{(-1)^p\theta^{2p}}{(2p)!} + \frac{A^2}{\theta^2} \\
&= I_3 + \frac{\sin\theta}{\theta}A + \frac{(1-\cos\theta)}{\theta^2}A^2,
\end{aligned}$$

as claimed. □

The above formulae are the well-known formulae expressing a rotation of axis specified by the vector (a, b, c) and angle θ.

The Rodrigues formula can used to show that the exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is surjective.

Given any rotation matrix $R \in \mathbf{SO}(3)$, we have the following cases:

(1) The case $R = I$ is trivial.

(2) If $R \neq I$ and $\mathrm{tr}(R) \neq -1$, then

$$\exp^{-1}(R) = \left\{ \frac{\theta}{2\sin\theta}(R - R^T) \;\middle|\; 1 + 2\cos\theta = \mathrm{tr}(R) \right\}.$$

(Recall that $\mathrm{tr}(R) = r_{1\,1} + r_{2\,2} + r_{3\,3}$, the *trace* of the matrix R).

Then there is a unique skew-symmetric B with corresponding θ satisfying $0 < \theta < \pi$ such that $e^B = R$.

(3) If $R \neq I$ and $\mathrm{tr}(R) = -1$, then R is a rotation by the angle π and things are more complicated, but a matrix B can be found. We leave this part as a good exercise: see Problem 17.8.

The computation of a logarithm of a rotation in $\mathbf{SO}(3)$ as sketched above has applications in kinematics, robotics, and motion interpolation.

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

12.8 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.

Definition 12.8. Given any real $n \times n$ matrix A, a *QR-decomposition* of A is any pair of $n \times n$ matrices (Q, R), where Q is an orthogonal matrix and R is an upper triangular matrix such that $A = QR$.

Note that if A is not invertible, then some diagonal entry in R must be zero.

Proposition 12.16. *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 Proposition 12.10 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

$$\begin{aligned}
A^1 &= \|Q'^1\| Q^1, \\
&\;\vdots \\
A^j &= (A^j \cdot Q^1)\, Q^1 + \cdots + (A^j \cdot Q^i)\, Q^i + \cdots + (A^j \cdot Q^{j-1})\, Q^{j-1} + \|Q'^j\| Q^j, \\
&\;\vdots \\
A^n &= (A^n \cdot Q^1)\, Q^1 + \cdots + (A^n \cdot Q^{n-1})\, Q^{n-1} + \|Q'^n\| Q^n.
\end{aligned}$$

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 diagonal entries $r_{k\,k} = \|Q'^k\| = A^k \cdot Q^k$ are indeed positive. $\square$

The reader should try the above procedure on some concrete examples for 2×2 and 3×3 matrices.

Remarks:

(1) Because the diagonal entries of R are positive, it can be shown that Q and R are unique. More generally, if A is invertible and if $A = Q_1 R_1 = Q_2 R_2$ are two QR-decompositions for A, then
$$R_1 R_2^{-1} = Q_1^\top Q_2.$$
The matrix $Q_1^\top Q_2$ is orthogonal and it is easy to see that $R_1 R_2^{-1}$ is upper triangular. But an upper triangular matrix which is orthogonal must be a diagonal matrix D with diagonal entries ± 1, so $Q_2 = Q_1 D$ and $R_1 = D R_2$.

(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 Proposition 13.4, and also Strang [169, 170], Golub and Van Loan [80], Trefethen and Bau [176], Demmel [48], Kincaid and Cheney [102], or Ciarlet [41]).

For better numerical stability, it is preferable to use the modified Gram–Schmidt method to implement the QR-factorization method. Here is a `Matlab` program implementing QR-factorization using modified Gram–Schmidt.

```
function [Q,R] = qrv4(A)
n = size(A,1);
for i = 1:n
    Q(:,i) = A(:,i);
    for j = 1:i-1
        R(j,i) = Q(:,j)'*Q(:,i);
        Q(:,i) = Q(:,i) - R(j,i)*Q(:,j);
    end
    R(i,i) = sqrt(Q(:,i)'*Q(:,i));
    Q(:,i) = Q(:,i)/R(i,i);
end
end
```

Example 12.13. Consider the matrix
$$A = \begin{pmatrix} 0 & 0 & 5 \\ 0 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

To determine the QR-decomposition of A, we first use the Gram-Schmidt orthonormalization procedure to calculate $Q = (Q^1 Q^2 Q^3)$. By definition

$$A^1 = Q'^1 = Q^1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and since $A^2 = \begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix}$, we discover that

$$Q'^2 = A^2 - (A^2 \cdot Q^1)Q^1 = \begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}.$$

Hence, $Q^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Finally,

$$Q'^3 = A_3 - (A^3 \cdot Q^1)Q^1 - (A^3 \cdot Q^2)Q^2 = \begin{pmatrix} 5 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix},$$

which implies that $Q^3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. According to Proposition 12.16, in order to determine R we need to calculate

$$\begin{aligned} r_{11} &= \left\| Q'^1 \right\| = 1 & r_{12} &= A^2 \cdot Q^1 = 1 & r_{13} &= A^3 \cdot Q^1 = 1 \\ & & r_{22} &= \left\| Q'^2 \right\| = 4 & r_{23} &= A_3 \cdot Q^2 = 1 \\ & & & & r_{33} &= \left\| Q'^3 \right\| = 5. \end{aligned}$$

In summary, we have found that the QR-decomposition of $A = \begin{pmatrix} 0 & 0 & 5 \\ 0 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is

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

Example 12.14. Another example of QR-decomposition is

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

Example 12.15. If we apply the above `Matlab` function to the matrix

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

we obtain

$$Q = \begin{pmatrix} 0.9701 & -0.2339 & 0.0619 & -0.0166 & 0.0046 \\ 0.2425 & 0.9354 & -0.2477 & 0.0663 & -0.0184 \\ 0 & 0.2650 & 0.9291 & -0.2486 & 0.0691 \\ 0 & 0 & 0.2677 & 0.9283 & -0.2581 \\ 0 & 0 & 0 & 0.2679 & 0.9634 \end{pmatrix}$$

and

$$R = \begin{pmatrix} 4.1231 & 1.9403 & 0.2425 & 0 & 0 \\ 0 & 3.7730 & 1.9956 & 0.2650 & 0 \\ 0 & 0 & 3.7361 & 1.9997 & 0.2677 \\ 0 & 0 & & 073.7324 & 2.0000 \\ 0 & 0 & 0 & 0 & 3.5956 \end{pmatrix}.$$

Remark: The `Matlab` function `qr`, called by `[Q, R] = qr(A)`, does not necessarily return an upper-triangular matrix whose diagonal entries are positive.

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 in Chapter 23), and for finding eigenvalues; see Chapter 18. 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 triangular matrix with positive diagonal entries. For more on QR, see Strang [169, 170], Golub and Van Loan [80], Demmel [48], Trefethen and Bau [176], or Serre [156].

A somewhat surprising consequence of the QR-decomposition is a famous determinantal inequality due to Hadamard.

Proposition 12.17. *(Hadamard) For any real $n \times n$ matrix $A = (a_{ij})$, we have*

$$|\det(A)| \leq \prod_{i=1}^{n} \Big(\sum_{j=1}^{n} a_{ij}^2 \Big)^{1/2} \quad \textit{and} \quad |\det(A)| \leq \prod_{j=1}^{n} \Big(\sum_{i=1}^{n} a_{ij}^2 \Big)^{1/2}.$$

Moreover, equality holds iff either A has orthogonal rows in the left inequality or orthogonal columns in the right inequality.

Proof. If $\det(A) = 0$, then the inequality is trivial. In addition, if the righthand side is also 0, then either some column or some row is zero. If $\det(A) \neq 0$, then we can factor A as $A = QR$, with Q is orthogonal and $R = (r_{ij})$ upper triangular with positive diagonal entries. Then since Q is orthogonal $\det(Q) = \pm 1$, so

$$|\det(A)| = |\det(Q)|\,|\det(R)| = \prod_{j=1} r_{jj}.$$

Now as Q is orthogonal, it preserves the Euclidean norm, so

$$\sum_{i=1}^{n} a_{ij}^2 = \left\| A^j \right\|_2^2 = \left\| QR^j \right\|_2^2 = \left\| R^j \right\|_2^2 = \sum_{i=1}^{n} r_{ij}^2 \geq r_{jj}^2,$$

which implies that

$$|\det(A)| = \prod_{j=1}^{n} r_{jj} \leq \prod_{j=1}^{n} \left\| R^j \right\|_2 = \prod_{j=1}^{n} \Big(\sum_{i=1}^{n} a_{ij}^2 \Big)^{1/2}.$$

The other inequality is obtained by replacing A by $A^\top$. Finally, if $\det(A) \neq 0$ and equality holds, then we must have

$$r_{jj} = \left\| A^j \right\|_2, \quad 1 \leq j \leq n,$$

which can only occur if A has orthogonal columns. ∎

Another version of Hadamard's inequality applies to symmetric positive semidefinite matrices.

Proposition 12.18. *(Hadamard) For any real $n \times n$ matrix $A = (a_{ij})$, if A is symmetric positive semidefinite, then we have*

$$\det(A) \leq \prod_{i=1}^{n} a_{ii}.$$

Moreover, if A is positive definite, then equality holds iff A is a diagonal matrix.

Proof. If $\det(A) = 0$, the inequality is trivial. Otherwise, A is positive definite, and by Theorem 8.10 (the Cholesky Factorization), there is a unique upper triangular matrix B with positive diagonal entries such that

$$A = B^\top B.$$

Thus, $\det(A) = \det(B^\top B) = \det(B^\top)\det(B) = \det(B)^2$. If we apply the Hadamard inequality (Proposition 12.17) to B, we obtain

$$\det(B) \leq \prod_{j=1}^{n}\left(\sum_{i=1}^{n} b_{ij}^2\right)^{1/2}. \tag{$*$}$$

However, the diagonal entries a_{jj} of $A = B^\top B$ are precisely the square norms $\|B^j\|_2^2 = \sum_{i=1}^n b_{ij}^2$, so by squaring $(*)$, we obtain

$$\det(A) = \det(B)^2 \leq \prod_{j=1}^{n}\left(\sum_{i=1}^{n} b_{ij}^2\right) = \prod_{j=1}^{n} a_{jj}.$$

If $\det(A) \neq 0$ and equality holds, then B must have orthogonal columns, which implies that B is a diagonal matrix, and so is A. $\square$

We derived the second Hadamard inequality (Proposition 12.18) from the first (Proposition 12.17). We leave it as an exercise to prove that the first Hadamard inequality can be deduced from the second Hadamard inequality.

12.9 Some Applications of Euclidean Geometry

Euclidean geometry has applications in computational geometry, in particular Voronoi diagrams and Delaunay triangulations. In turn, Voronoi diagrams have applications in motion planning (see O'Rourke [133]).

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 17). 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 12.8, or better, using Householder matrices, as shown in Section 13.2. 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). 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 22). 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 [170, 169], Golub and Van Loan [80], Demmel [48], Serre [156], and Trefethen and Bau [176].

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. More details on the method of least squares and pseudo-inverses can be found in Chapter 23.

12.10 Summary

The main concepts and results of this chapter are listed below:

- Bilinear forms; *positive definite* bilinear forms.
- *Inner products*, *scalar products*, *Euclidean spaces*.
- *Quadratic form* associated with a bilinear form.

- The Euclidean space $\mathbb{E}^n$.
- The *polar form* of a quadratic form.
- *Gram matrix* associated with an inner product.
- The *Cauchy–Schwarz inequality*; the *Minkowski inequality*.
- The *parallelogram law*.
- *Orthogonality*, *orthogonal complement* $F^{\perp}$; *orthonormal family*.
- The *musical isomorphisms* $\flat\colon E \to E^*$ and $\sharp\colon E^* \to E$ (when E is finite-dimensional); Theorem 12.6.
- The *adjoint* of a linear map (with respect to an inner product).
- Existence of an orthonormal basis in a finite-dimensional Euclidean space (Proposition 12.9).
- The *Gram–Schmidt orthonormalization procedure* (Proposition 12.10).
- The *Legendre* and the *Chebyshev* polynomials.
- *Linear isometries* (*orthogonal transformations*, *rigid motions*).
- The *orthogonal group*, *orthogonal matrices*.
- The matrix representing the adjoint f^* of a linear map f is the transpose of the matrix representing f.
- The *orthogonal group* $\mathbf{O}(n)$ and the *special orthogonal group* $\mathbf{SO}(n)$.
- *QR-decomposition* for invertible matrices.
- The *Hadamard inequality* for arbitrary real matrices.
- The *Hadamard inequality* for symmetric positive semidefinite matrices.
- The *Rodrigues formula* for rotations in $\mathbf{SO}(3)$.

12.11 Problems

Problem 12.1. E be a vector space of dimension 2, and let (e_1, e_2) be a basis of E. Prove that if $a > 0$ and $b^2 - ac < 0$, then the bilinear form defined such that

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

is a Euclidean inner product.

Problem 12.2. Let $\mathcal{C}[a, b]$ denote the set of continuous functions $f\colon [a, b] \to \mathbb{R}$. Given any two functions $f, g \in \mathcal{C}[a, b]$, let

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

Prove that the above bilinear form is indeed a Euclidean inner product.

Problem 12.3. Consider the inner product

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

of Problem 12.2 on the vector space $\mathcal{C}[-\pi, \pi]$. Prove that

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

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

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

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

Problem 12.4. Prove that the following matrix is orthogonal and skew-symmetric:

$$M = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 \\ -1 & 1 & 0 & -1 \\ -1 & -1 & 1 & 0 \end{pmatrix}.$$

Problem 12.5. Let E and F be two finite Euclidean spaces, let $(u_1, \ldots, u_n)$ be a basis of E, and let $(v_1, \ldots, v_m)$ be a basis of F. For any linear map $f\colon E \to F$, if A is the matrix of f w.r.t. the basis $(u_1, \ldots, u_n)$ and B is the matrix of f^* w.r.t. the basis $(v_1, \ldots, v_m)$, if G_1 is the Gram matrix of the inner product on E (w.r.t. $(u_1, \ldots, u_n)$) and if G_2 is the Gram matrix of the inner product on F (w.r.t. $(v_1, \ldots, v_m)$), then

$$B = G_1^{-1} A^\top G_2.$$