Chapter 9

Euclidean Spaces

9.1 Inner Products, Euclidean Spaces

The framework of vector spaces allows us deal with ratios of vectors and linear combinations, 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 will allow us to deal with *metric notions* such as orthogonality and length (or distance).

First, we define a Euclidean structure on a vector space.

Definition 9.1. A real vector space E is a *Euclidean space* iff it is equipped with a symmetric bilinear form $\varphi\colon E \times E \to \mathbb{R}$ which is also *positive definite*, which means that

$$\varphi(u, u) > 0, \quad \text{for every } u \neq 0.$$

More explicitly, $\varphi\colon E \times E \to \mathbb{R}$ satisfies the following axioms:

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

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

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

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

Example 2. 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,

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

Example 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 as an easy exercise that $\langle -, -\rangle$ is indeed an inner product on $\mathcal{C}[a, b]$.

When $[a, b] = [-\pi, \pi]$ (or $[a, b] = [0, 2\pi]$, 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 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 6.2).

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.

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

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

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 $\mathbf{y}$.

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

Proposition 9.1. *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.

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

Proposition 9.2. *Let E be a Euclidean space with inner product φ and quadratic form Φ. 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 Minkovski 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 $\lambda > 0$.

Sketch of proof. 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 can show that

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

Since φ is positive definite, we have $T(\lambda) \geq 0$ 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, If $\Phi(v) > 0$, then

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

can't have distinct roots, which means that its discriminant

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

is zero or negative, which is precisely the Cauchy-Schwarz inequality.

The Minkovski inequality can then be shown.

The Minkovski inequality

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

shows that the map $u \mapsto \sqrt{\Phi(u)}$ satisfies the *triangle inequality*, condition (N3) of definition 6.1, and since φ is bilinear and positive definite, it also satisfies conditions (N1) and (N2) of definition 6.1, 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 Minkovski inequality as

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

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*:

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

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

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

Proving that the above form is indeed symmetric and bilinear is quite tricky.

We now define orthogonality.

9.2 Orthogonality, Duality, Adjoint Maps

Definition 9.2. Given a Euclidean space E, any two vectors $u, v \in E$ are *orthogonal, or perpendicular* iff $u \cdot v = 0$. Given a family $(u_i)_{i \in I}$ of vectors in E, we say that $(u_i)_{i \in I}$ is *orthogonal* iff $u_i \cdot u_j = 0$ for all $i, j \in I$, where $i \neq j$. We say that the family $(u_i)_{i \in I}$ is *orthonormal* iff $u_i \cdot u_j = 0$ 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 5. Going back to example 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 9.3. *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.*

Proposition 9.4. *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 one 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 3.8.

By Proposition 3.16, 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 isomorphphism.

The mapping from E to E^* is defined as follows. 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 9.5. *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.

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

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).

(2) Theorem 9.5 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 9.1. 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 9.2. 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

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

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

$$z = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}.$$

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

$$(u_1 \times \cdots \times u_{n-1}) \cdot w = \det(u_1, \ldots, u_{n-1}, w)$$

for all $w \in \mathbb{R}^n$.

Example 9.3. 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.

As a consequence of Theorem 9.5, 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.

Theorem 9.5 allows us to define the adjoint of a linear map on a Euclidean space.

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 9.5, there is a unique vector in E denoted as $f^*(u)$, such that

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

for every $v \in E$.

Proposition 9.6. *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*

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

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

Remark: Proposition 9.6 still holds if the inner product on E is replaced by a nondegenerate symmetric bilinear form φ.

Linear maps $f\colon E \to E$ 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).

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

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

also play an important role. They are *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 as $\langle -, - \rangle_1$ and the inner product on F is denoted as $\langle -, - \rangle_2$, given any linear map $f \colon E \to F$, it is immediately verified that the proof of Proposition 9.6 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.

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.

We can also use Theorem 9.5 to show that any Euclidean space of finite dimension has an orthonormal basis.

Proposition 9.7. *Given any nontrivial Euclidean space E of finite dimension $n \geq 1$, there is an orthonormal basis $(u_1, \ldots, u_n)$ for E.*

There is a more constructive way of proving Proposition 9.7, 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.

Proposition 9.8. *Given any nontrivial Euclidean space E of 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 define the vectors u_k and u'_k as follows.

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

and for the inductive step

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

We need to show that u'_{k+1} is nonzero, and we conclude by induction.

Remarks:

(1) 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, we normalize u'_{k+1}.

The QR-decomposition can now be obtained very easily. We will do this in section 9.4.

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

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

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 Proposition 9.8 also works for a countably infinite basis for E, producing a countably infinite orthonormal basis.

Example 6. 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:

If we let 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:

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

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

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

As a consequence of Proposition 9.7 (or Proposition 9.8), 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

$$\|u\| = \|u_1e_1 + \cdots + u_ne_n\| = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

We can also prove the following proposition regarding orthogonal spaces.

Proposition 9.9. *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 $F^{\perp\perp} = F$.*

9.3 Linear Isometries (Orthogonal Transformations)

In this section, we consider linear maps between Euclidean spaces that preserve the Euclidean norm.

Definition 9.3. 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* iff it is linear and

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

for all $u \in E$.

Thus, a linear isometry is a linear map that preserves the norm.

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 9.3 is called a *metric map*, and a linear isometry is defined as a *bijective* metric map.

Also, 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 9.10. *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.

For (2), we shall prove a slightly stronger result. We prove that if

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

for all $u, v \in E$, 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 linear map such that $g(0) = 0$ and (3) holds.

Remarks:

(i) The dimension assumption is only needed 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 3.13, this shows that f is affine with associated linear map g.

This fact is worth recording as the following proposition.

Proposition 9.11. *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 9.10, we usually abbreviate "linear isometry" as "isometry," unless we wish to emphasize that we are dealing with a map between vector spaces.

9.4 The Orthogonal Group, Orthogonal Matrices

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

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

Proposition 9.12. *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 $\mathbb{R}^n$.

Proposition 9.12 shows that the inverse of an isometry f is its adjoint f^*. Proposition 9.12 also motivates the following definition:

Definition 9.4. A real $n \times n$ matrix is an *orthogonal matrix* iff

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

Remarks: 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 $(u_1, \ldots, u_n)$, and since $(v_1, \ldots, v_n)$ is orthonormal, the columns of P are orthonormal, and by Proposition 9.12 (2), the matrix P is orthogonal.

The proof of Proposition 9.10 (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.

Recall that the determinant $\det(f)$ of an endomorphism $f\colon E \to E$ 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.

Definition 9.5. Given a Euclidean space E of dimension n, the set of isometries $f\colon E \to E$ forms a group denoted as $\mathbf{O}(E)$, 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 as $\mathbf{SO}(E)$, 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*.

9.5 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.

Proposition 9.13. *Given any $n \times n$ real 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 9.8 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

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

If we let $r_{k\,k} = \|Q'^k\|$, $r_{i\,j} = A^j \cdot Q^i$ when $1 \leq i \leq j-1$, and then $Q = (Q^1, \ldots, Q^n)$ and $R = (r_{i\,j})$, then R is upper-triangular, Q is orthogonal, and

$$A = QR.$$

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 also Strang [31]).

Example 7. 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

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

Another example of QR-decomposition is

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

where

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

and

$$R = \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{pmatrix}$$

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 triangular matrix with positive diagonal entries. For more on QR, see Strang [31].

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.

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

Proposition 9.14. *(Hadamard) For any real $n \times n$ matrix $A = (a_{ij})$, we have*

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

Moreover, equality holds iff either A has a zero column in the left inequality or a zero row in the right inequality, or A is orthogonal.

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

Proposition 9.15. *(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.