

Sections to be covered: 6.1, 6.2.

In elementary linear algebra courses, you learned the concept of *dot product* for vectors in  $\mathbb{R}^n$  and use the concept to define *orthogonal vectors*, *the norm (or length) of vectors*. In connection with the concept of dot product, you also learned

- Pythagorean theorem:  $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$  when  $\mathbf{u} \cdot \mathbf{v} = 0$ .
- Cauchy-Schwarz inequality:  $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\|$ .
- Triangle inequality:  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ .
- Orthogonal projections.
- Gram-Schmidt orthogonalization process.
- Orthogonal matrices.

These concepts are very useful. A natural question is: are there similar concepts for general vector spaces?

In a general vector space  $V$  (over a field  $\mathbb{F}$  such as  $\mathbb{R}$  or  $\mathbb{C}$ ), there are no pre-determined coordinates for vectors, although one can associate a vector to its coordinates once an ordered basis has been chosen. One possible approach is to choose an ordered basis, say,  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  for  $V$ , then for any vectors  $\mathbf{x}$  and  $\mathbf{y}$  in  $V$ , we can find scalars  $x_1, \dots, x_n, y_1, \dots, y_n$ , such that

$$\mathbf{x} = x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n, \quad \text{and} \quad \mathbf{y} = y_1\mathbf{v}_1 + \dots + y_n\mathbf{v}_n.$$

Namely,  $[x]_\beta = (x_1, \dots, x_n)^t$ , and  $[y]_\beta = (y_1, \dots, y_n)^t$ . One can attempt to define the dot product of  $\mathbf{x}$  and  $\mathbf{y}$  as  $\sum_{i=1}^n x_i y_i$ . However, when  $V$  does not have a preferred basis, how does one tell whether there will be a difference when using a different basis. Suppose  $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  is another ordered basis for  $V$ , and

$$\mathbf{x} = \hat{x}_1\mathbf{w}_1 + \dots + \hat{x}_n\mathbf{w}_n, \quad \mathbf{y} = \hat{y}_1\mathbf{w}_1 + \dots + \hat{y}_n\mathbf{w}_n.$$

Then one could have used  $\sum_{i=1}^n \hat{x}_i \hat{y}_i$  to define the dot product between  $\mathbf{x}$  and  $\mathbf{y}$ . Suppose the change of coordinate matrix from  $\beta$  to  $\gamma$  is given by

$$w_j = \sum_{i=1}^n q_{ij} \mathbf{v}_i, \text{ for } j = 1, \dots, n.$$

Then

$$x_i = \sum_{k=1}^n q_{ik} \hat{x}_k, \quad \text{and} \quad y_i = \sum_{l=1}^n q_{il} \hat{y}_l,$$

thus

$$\sum_{i=1}^n x_i y_i = \sum_{i=1}^n \sum_{k=1}^n \sum_{l=1}^n q_{ik} q_{il} \hat{x}_k \hat{y}_l = \sum_{k=1}^n \sum_{l=1}^n \hat{a}_{kl} \hat{x}_k \hat{y}_l,$$

where  $\hat{a}_{kl} = \sum_{i=1}^n q_{ik} q_{il}$ . But the result given by

$$\sum_{k=1}^n \sum_{l=1}^n \hat{a}_{kl} \hat{x}_k \hat{y}_l$$

may not agree with what is given by  $\sum_{i=1}^n \hat{x}_i \hat{y}_i$ , unless

$$\sum_{i=1}^n q_{ik} q_{il} = \delta_{kl}, \quad \text{namely, } Q^t Q = I_{n \times n}, \text{ where, } Q = (q_{ij}).$$

It turns out that a dot product can be defined naturally for many vector spaces without using a basis, and this product shares all the properties of the dot product defined on  $\mathbb{R}^n$ . Such dot products are also called inner products. What we do in **6.1** and **6.2** is to assume that a vector space  $V$  has an inner product, and develop its properties as we did for the dot product on  $\mathbb{R}^n$ . In the examples, we will learn how some of the natural inner products are defined on some vector spaces that we encounter often.

Much of what we will do in this chapter is to generalize the discussion of dot product for  $\mathbb{R}^n$  to an arbitrary vector space with an inner product. What is perhaps new to many students is the extensive use of complex numbers in this chapter. From our discussion on diagonalization, we already learned that complex valued eigenvalues may arise naturally even when we deal with linear operators defined on vector spaces over the field of real numbers. There are also natural reasons to discuss inner product in the context of vector spaces over the field of complex numbers, the most compelling of which is their natural occurrences in physical problems.

- For  $\mathbb{C}^n$ , we define the inner product of  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$  in  $\mathbb{C}^n$  by

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n.$$

When  $n = 1$ , and  $\mathbf{x} = a + bi$ ,  $\mathbf{y} = c + di$ , we find

$$\langle \mathbf{x}, \mathbf{y} \rangle = (a + bi)(\overline{c + di}) = (a + bi)(c - di) = ac + bd + (bc - ad)i \in \mathbb{C}.$$

When  $\mathbf{y} = \mathbf{x} = a + bi$ ,  $\langle \mathbf{x}, \mathbf{x} \rangle = a^2 + b^2$ , a non-negative real number, equal to zero only when  $a = b = 0$ . If we had used  $(a + bi)(c + di)$  as a candidate for inner product, then  $(a + bi)(a + bi) = a^2 - b^2 + 2abi$ , which may not be real number, so we could not use  $\sqrt{(a + bi)(a + bi)}$  to define length of  $a + bi$ .

- The conjugate transpose of a matrix is defined for similar reason. When

$$A = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11}i & a_{12} + b_{12}i \\ a_{21} + b_{21}i & a_{22} + b_{22}i \end{pmatrix}$$

with  $a_{ij}, b_{ij}$  real entries, then

$$A^* = \begin{pmatrix} \bar{z}_{11} & \bar{z}_{21} \\ \bar{z}_{12} & \bar{z}_{22} \end{pmatrix} = \begin{pmatrix} a_{11} - b_{11}i & a_{21} - b_{21}i \\ a_{12} - b_{12}i & a_{22} - b_{22}i \end{pmatrix}.$$

Note that

$$\begin{aligned} A^* A &= \begin{pmatrix} a_{11} - b_{11}i & a_{21} - b_{21}i \\ a_{12} - b_{12}i & a_{22} - b_{22}i \end{pmatrix} \begin{pmatrix} a_{11} + b_{11}i & a_{12} + b_{12}i \\ a_{21} + b_{21}i & a_{22} + b_{22}i \end{pmatrix} \\ &= \begin{pmatrix} |z_{11}|^2 + |z_{21}|^2 & (a_{11} - b_{11}i)(a_{12} + b_{12}i) + (a_{21} - b_{21}i)(a_{22} + b_{22}i) \\ (a_{12} - b_{12}i)(a_{11} + b_{11}i) + (a_{22} - b_{22}i)(a_{21} + b_{21}i) & |z_{12}|^2 + |z_{22}|^2 \end{pmatrix} \\ &= \begin{pmatrix} |z_{11}|^2 + |z_{21}|^2 & \bar{z}_{11}z_{12} + \bar{z}_{21}z_{22} \\ \bar{z}_{12}z_{11} + \bar{z}_{22}z_{21} & |z_{12}|^2 + |z_{22}|^2 \end{pmatrix}. \end{aligned}$$

The trace of this matrix is  $|z_{11}|^2 + |z_{21}|^2 + |z_{12}|^2 + |z_{22}|^2 \geq 0$ , and this matrix has the properties that its conjugate transpose is equal to itself. This example illustrates

that it pays to learn to do algebraic operations of complex numbers directly without using their real and imaginary parts, and that  $\langle A, B \rangle = \text{tr}(B^*A)$  is a reasonable choice for inner product on the space of complex valued matrices—at least in terms of fulfilling condition (d) in the definition.

- Note that the notion of orthogonal vectors are defined in the same (symbolic) way:  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ , and all the arguments involving the notion of orthogonality can be generalized in an analogous algebraic way. However, our visual notion of orthogonality may not always apply. For instance, for  $\mathbf{x}, \mathbf{y} \in \mathbb{C}$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$  iff  $ac + bd = 0$  and  $bc - ad = 0$ , which happens iff  $(a, b) \cdot (c, d) = 0$  and  $(a, b) \cdot (-d, c) = 0$ . These conditions hold iff  $(a, b) = (0, 0)$  or  $(c, d) = (0, 0)$ . In other words,  $\langle \mathbf{x}, \mathbf{y} \rangle \neq 0$  unless  $\mathbf{x} = 0$  or  $\mathbf{y} = 0$ , even though when we interpret  $\mathbf{x}$  and  $\mathbf{y}$  as vectors in  $\mathbb{R}^2$ ,  $\mathbf{x} \perp \mathbf{y}$  when  $(a, b) \cdot (c, d) = 0$ . The key to understand this point is that  $\mathbb{C}$  endowed with the inner product  $\langle \mathbf{x}, \mathbf{y} \rangle = x\bar{y}$  is a different inner product space from  $\mathbb{R}^2$  endowed with  $(a, b) \cdot (c, d) = ac + bd$ ! Another point that was mentioned earlier is that  $\mathbb{C}$  is a one dimensional vector space over  $\mathbb{C}$ , but two dimensional over  $\mathbb{R}$ . In fact, the inner product  $\langle \mathbf{x}, \mathbf{y} \rangle = x\bar{y}$  on  $\mathbb{C}$  as a one dimensional vector space over  $\mathbb{C}$  is related to the inner product  $(a, b) \cdot (c, d) = ac + bd$  on  $\mathbb{C}$  as a two dimensional vector space over  $\mathbb{R}$ .
- If  $V$  is an  $n$ -dimensional vector space over the field  $\mathbb{C}$ , then it can also be considered as a  $2n$ -dimensional vector space over the field  $\mathbb{R}$ . We denote the former by  $(V, \mathbb{C})$  and the latter by  $(V, \mathbb{R})$ . When  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a basis for  $(V, \mathbb{C})$ , then  $\{\mathbf{v}_1, i\mathbf{v}_1, \mathbf{v}_2, i\mathbf{v}_2, \dots, \mathbf{v}_n, i\mathbf{v}_n\}$  is a basis for  $(V, \mathbb{R})$ . If  $\langle \mathbf{x}, \mathbf{y} \rangle$  denotes an inner product on  $(V, \mathbb{C})$ , then we can split  $\langle \mathbf{x}, \mathbf{y} \rangle$  into its real part  $\langle \mathbf{x}, \mathbf{y} \rangle_1$  and imaginary part  $\langle \mathbf{x}, \mathbf{y} \rangle_2$ :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_1 + i \langle \mathbf{x}, \mathbf{y} \rangle_2 .$$

Then the properties (a)-(d) imply that  $\langle \mathbf{x}, \mathbf{y} \rangle_1$  is an inner product for  $(V, \mathbb{R})$ . In our example above,  $V = \mathbb{C}$ , then  $\langle \mathbf{x}, \mathbf{y} \rangle = x\bar{y}$ , and  $\langle \mathbf{x}, \mathbf{y} \rangle_1 = ac + bd$  is our usual inner product on  $\mathbb{C}$  considered to be the same as  $\mathbb{R}^2$ . The point that needs to be emphasized at this point is that one has to be clear about which vector space and which inner product one is referring to when using the notions of orthogonality, orthogonal projection, etc.