ELEN 162
A. I. Zecevic

Lecture Notes for Week 1

Linear Vector Spaces

Set S is referred to as a *linear vector space* if it satisfies a number of well defined properties, the most important of which is described below:

Property 1. If $x \in S$ and $y \in S$, and a and b are numbers, then $ax + by \in S$.

The following example shows how operations such as addition and multiplication by a number can be defined in a given set S. It also illustrates how Property 1 can be tested in practice.

Example 1. Consider the set of all vectors of the form

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{1}$$

where x_1, x_2 and x_3 are real numbers. Since each element in this set is a 3×1 vector, it is commonly denoted as R^3.

If we define the operation of addition as

$$x + y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix} \tag{2}$$

and multiplication by a number as

$$ax = \begin{bmatrix} ax_1 \\ ax_2 \\ ax_3 \end{bmatrix} \tag{3}$$

it is easily verified that

$$ax + by = \begin{bmatrix} ax_1 \\ ax_2 \\ ax_3 \end{bmatrix} + \begin{bmatrix} by_1 \\ by_2 \\ by_3 \end{bmatrix} = \begin{bmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ ax_3 + by_3 \end{bmatrix} \tag{4}$$

is a 3×1 vector for any choice of a and b. As a result, we can conclude that Property 1 is satisfied in this case. Since all the other properties hold as well, set R^3 can be classified as a linear vector space.

Linear Independence

An important concept related to linear spaces and their elements is the notion of *linear independence*. In space R^n (whose elements are real vectors of dimension $n \times 1$) linear independence can be defined in the following way.

Definition 1. A set of vectors $\{x_1, x_2, \ldots, x_p\}$ in R^n is said to be linearly independent if

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_p x_p = 0 \tag{5}$$

implies $\alpha_1 = \ldots = \alpha_p = 0$.

To get a sense for what this definition means, suppose that (5) holds but that some elements of set $\{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ are *not* zero. If one of them happens to be α_k, equation (5) would allow us to express x_k as

$$x_k = -\frac{1}{\alpha_k} \sum_{i \neq k} \alpha_i x_i \tag{6}$$

Since x_k obviously depends on vectors $\{x_i\}$ $(i \neq k)$, adding it to this set would not contribute any new information. Given that this is the case, we would have no grounds to treat vector x_k as an "independent" entity.

How does one determine whether a set of vectors is linearly independent? In order to explain that, we first need to say a few things about how systems of linear equations are solved using Gaussian elimination. The following two examples illustrate the main features of this method.

Example 2. Suppose that we want to solve system

$$Ax = b \tag{7}$$

where

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{8}$$

and

$$b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \tag{9}$$

The main idea behind Gaussian elimination is to transform matrix A and vector b in such a way that the matrix becomes *upper triangular*. This is desirable because it allows us to compute the variables one by one (using a procedure known as *backward substitution*).

Since our objective is to eliminate all nonzero elements below the diagonal, we will begin with the first column of A. If we multiply row 1 by -2 and add it to row 2, we obtain a new matrix

$$A_1 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -5 \\ 1 & 1 & 1 \end{bmatrix} \tag{10}$$

If we do the same to vector b, we get

$$b_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \tag{11}$$

Because these two operations are equivalent to multiplying both sides of equation (7) by matrix

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

it follows that the solution remains unchanged.

Multiplying the first row of matrix A_1 and vector b_1 by -1 and adding it to the third row transforms the system into

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -5 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \tag{13}$$

which brings us one step closer to an upper triangular structure. It is easily verified that such an operation is equivalent to multiplying both sides of the equation by matrix

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{14}$$

which produces

$$(P_2 P_1 A)x = P_2 P_1 b \tag{15}$$

where

$$P_2 P_1 A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -5 \\ 0 & 1 & -2 \end{bmatrix} \equiv A_2 \tag{16}$$

and

$$P_2 P_1 b = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \equiv b_2 \tag{17}$$

Since column 2 of matrix A_2 has a single nonzero element below the diagonal, we need just one more step to obtain an upper triangular matrix. This step amounts to multiplying row 2 of A_2 by -0.5 and adding it to row 3, after which the system becomes

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -5 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \tag{18}$$

Since such a transformation is equivalent to multiplying A_2 and b_2 by matrix

$$P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -0.5 & 1 \end{bmatrix} \tag{19}$$

vector x will not be affected in any way.

Observing that the matrix is now upper triangular, we can solve the resulting system

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & -5 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \tag{20}$$

in a step by step manner, beginning with variable x_3. If we do so, we obtain

$$\begin{aligned} &0.5x_3 = 1 \implies x_3 = 2 \\ &2x_2 - 5x_3 = -2 \implies x_2 = 4 \\ &x_1 + 3x_3 = 1 \implies x_1 = -5 \end{aligned} \tag{21}$$

This procedure is known as *backward substitution*, because it always starts from the *last* component of vector x.

Our next example demonstrates how this approach can be extended to *singular* matrices (i.e., matrices whose determinant is *zero*). We will see that in such cases system

$$Ax = 0 \tag{22}$$

has infinitely many solutions (as opposed to the scenario where A is non-singular). These solutions constitute what is known as the *nullspace of matrix* A, and our objective in the following will be to find a compact representation for this space.

Example 3. Suppose that we are given a system of the form (22) where

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 2 \\ 0 & 4 & 2 & 0 \end{bmatrix} \tag{23}$$

and are asked to find all of its solutions. If we multiply the first row by -2 and add it to row 3, we obtain

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{24}$$

After one more step, Gaussian elimination transforms the original system into

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{25}$$

where the matrix is upper triangular. What makes this matrix different from the one in (20) is the fact that rows 3 and 4 contain only zeros. As a result, the last two equations will be satisfied for *any* choice of vector x.

This "built in" redundancy reduces the system to two equations in four unknowns, which means that we can choose x_3 and x_4 *arbitrarily.* Setting $x_3 = s$ and $x_4 = t$ (where s and t are unspecified real numbers), we obtain

$$2x_2 + x_3 = 2x_2 + s = 0 \implies x_2 = -\frac{s}{2} \tag{26}$$

and

$$x_1 + x_2 + x_4 = x_1 - \frac{s}{2} + t = 0 \implies x_1 = \frac{s}{2} - t \tag{27}$$

respectively. These expressions allow us to describe the nullspace of matrix A in a very simple way, since the solutions of system (25) have the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} s/2 - t \\ -s/2 \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{28}$$

Such a representation is clearly convenient, because it requires only *two* vectors (although there are infinitely many possibilities).

An alternative approach for deriving expression (28) would be to rewrite the first two equations in the system as

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_4 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{29}$$

If we once again set $x_3 = s$ and $x_4 = t$, (29) becomes

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s \begin{bmatrix} 0 \\ -1 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \end{bmatrix} \tag{30}$$

and multiplying both sides by matrix

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \end{bmatrix} \tag{31}$$

produces

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \end{bmatrix} \tag{32}$$

It is not difficult to see that this matches expression (28) precisely.

The procedure that we used in Example 3 is very useful when it comes to determining whether a given set of vectors is linearly independent. The following example illustrates how this can be done, and shows how Gaussian elimination works in cases when matrix A is *rectangular.*

Example 4. Suppose that we are given three 5×1 vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 3 \\ 0 \end{bmatrix}; \quad v_3 = \begin{bmatrix} 2 \\ -1/2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{33}$$

and are asked to determine if they are linearly independent. If this were the case, we know that v_1, v_2 and v_3 would satisfy

$$a_1 v_1 + a_2 v_2 + a_3 v_3 = 0 \tag{34}$$

only when $a_1 = a_2 = a_3 = 0$. We must therefore check whether the vectors described in (33) conform to this condition.

In order to do that, let us rewrite expression (34) as

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \\ 1 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{35}$$

and consider whether this system of equations has a nonzero solution. If we apply Gaussian elimination to the matrix on the left hand side (which is rectangular in this case), we obtain the following pair of transformations.

STEP 1

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{36}$$

STEP 2

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{37}$$

Since the last three equations in (37) are satisfied for *any* choice of a_1, a_2 and a_3, we need to focus only on the first two. Following the same procedure as in the previous example, we can rewrite these equations as

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + a_3 \begin{bmatrix} 2 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{38}$$

and set $a_3 = t$, (where t is an arbitrary real number). If we do so, (38) becomes

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} -2 \\ -1/2 \end{bmatrix} \tag{39}$$

Multiplying both sides by

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \tag{40}$$

we obtain

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ -1/2 \end{bmatrix} = t \begin{bmatrix} -5/2 \\ 1/2 \end{bmatrix} \tag{41}$$

which indicates that the solutions of system (35) have the general form

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = t \begin{bmatrix} -5/2 \\ 1/2 \\ 1 \end{bmatrix} \tag{42}$$

Given that coefficients a_1, a_2 and a_3 have nonzero values when $t \neq 0$, we can conclude that vectors v_1, v_2 and v_3 are *not* linearly independent.

In cases such as this, it is often of interest to find a subset of vectors which *are* linearly independent. With that in mind, let us consider v_1 and v_2, and examine whether system

$$a_1 v_1 + a_2 v_2 = 0 \tag{43}$$

has a nontrivial solution. If we rewrite this system as

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{44}$$

and execute two steps of Gaussian elimination, we obtain

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{45}$$

Since the first two equations (which are the only ones that matter) have the form

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{46}$$

it is easily verified that $a_1 = a_2 = 0$ is the only possible solution. This implies that vectors v_1 and v_2 are linearly independent.

The Concept of a Basis

Linear independence is closely related to the notion of a *basis*. In order to explain what this means, let us once again consider space R^3. If addition and multiplication by a number are defined in the manner shown in (2) and (3), it is easily verified that every one of these elements can be expressed as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 E_1 + x_2 E_2 + x_3 E_3 \tag{47}$$

where

$$E_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad E_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \quad E_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{48}$$

Vectors $\{E_1, E_2, E_3\}$ are said to be a *basis* in this space, since any vector $x \in R^3$ can be represented as their linear combination.

Remark 1. It is important to recognize in this context that set $\{E_1, E_2, E_3\}$ is not the only possible basis, and that there are many other equally viable choices. What is common to all of them, however, is that they consist of three vectors (which is why this space is said to be *three dimensional*).

The following result establishes an important connection between bases and linear independence in space R^n.

Lemma 1.1. Any set of n linearly independent vectors in R^n constitutes a basis in this space.

Scalar Products and Orthogonality

Bases such as $\{E_1, E_2, E_3\}$ are interesting because they possess a property known as *orthonormality.* This property is associated with the notion of a *scalar product*, which represents a mapping that transforms the elements of a vector space into a number. In space R^3, the scalar product of vectors

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{49}$$

and

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{50}$$

is defined as

$$\langle x, y \rangle = x^T y = x_1 y_1 + x_2 y_2 + x_3 y_3 \tag{51}$$

It is not difficult to see that this operation always produces a real number, and that

$$\langle x, x \rangle = x_1^2 + x_2^2 + x_3^2 \tag{52}$$

is positive for all $x \neq 0$.

The fact that $\langle x, x \rangle > 0$ allows us to define the *norm* of x as

$$||x|| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{53}$$

Such a definition provides us with a convenient way to measure the "size" of this vector, and to evaluate the "distance" between vectors x and y (which can be done by computing $||x - y||$). In the special case when $||x|| = 1$, we say that vector x is *normalized.* This turns out to be a very useful property, both in linear algebra and in functional analysis.

Remark 2. Expression (53) is *not* the only way to define a vector norm - we will look at several alternative possibilities later.

Now that we know what a scalar product is, we can introduce the concept of orthonormality. In space R^n, elements x and y are considered to be *orthogonal* if their scalar product satisfies

$$\langle x, y \rangle = 0 \tag{54}$$

Because vectors E_1, E_2 and E_3 meet this condition (and are also *normalized*), we say that set $\{E_1, E_2, E_3\}$ represents an *orthonormal basis* in space R^3. We will see that bases of this sort play a crucial role in quantum mechanics, since they are associated with measurable states.

The following two examples illustrate how these ideas can be applied to other types of linear spaces.

Example 5. Consider the set of all real 2×2 matrices, whose elements have the general form

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{55}$$

If we define addition and multiplication by a number as

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \tag{56}$$

and

$$cA = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix} \tag{57}$$

respectively, it is easily verified that Property 1 of linear spaces is satisfied (as are all the other ones). It is also straightforward to show that matrices

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad E_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{58}$$

constitute a basis in this space, since any 2×2 matrix can be represented as

$$A = a_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{59}$$

The fact that this basis consists of four elements indicates that such a space is *four dimensional.*

A simple way to introduce a scalar product into this space involves the pairwise multiplication of matrix elements. If we adopt such an approach, the scalar product of matrices A and B can be defined as

$$\langle A, B \rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22} \tag{60}$$

(which is obviously a number). Given such a definition, it is not difficult to show that set $\{E_1, E_2, E_3, E_4\}$ represents an *orthonormal* basis, since

$$\langle E_i, E_j \rangle = 0 \tag{61}$$

when $i \neq j$ and

$$\langle E_i, E_i \rangle = 1 \tag{62}$$

for $i = 1, 2, 3, 4$.

Example 6. Let C^n be the set of all $n \times 1$ vectors whose components are *complex numbers*. In this space, the scalar product can be defined as

$$\langle x, y \rangle = x^T y^* = x_1 y_1^* + x_2 y_2^* + \ldots + x_n y_n^* \tag{63}$$

where y_i^* denotes the complex conjugate of y_i. It is not difficult to see that the scalar product defined in (63) satisfies

$$\langle y, x \rangle = \langle x, y \rangle^* \tag{64}$$

$$\langle \lambda x, y \rangle = \lambda \langle x, y \rangle \tag{65}$$

and

$$\langle x, \lambda y \rangle = \lambda^* \langle x, y \rangle \tag{66}$$

for any pair of complex vectors x, y and any complex number λ.

Since every vector $x \in C^n$ can be represented as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 E_1 + x_2 E_2 + \ldots + x_n E_n \tag{67}$$

where

$$E_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \quad E_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}; \quad \ldots \quad E_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{68}$$

and x_i are complex numbers, it follows that C^n is an n-dimensional space. Given that $\langle E_i, E_j \rangle = 0$ when $i \neq j$ and that $\langle E_i, E_i \rangle = 1$ $(i = 1, 2, \ldots, n)$, it is obvious that vectors $\{E_1, E_2, \ldots, E_n\}$ consitute an orthonormal basis in this space.

The Gram-Schmidt Method

What other orthonormal bases exist in R^n (apart from $\{E_1, \ldots, E_n\}$)? In order to answer this question, we should first recall that *any* set of linearly independent vectors $\{v_1, v_2, \ldots, v_n\}$ constitutes a basis in R^n (see Lemma 1.1). What we will now consider is how one can use such sets to construct a variety of different *orthonormal* bases in this space.

The Gram-Schmidt method is the simplest and most elegant way to achieve this objective. Before we explain how this process works, however, it is helpful to describe what is meant by the *span* of a collection of vectors. Given a set of $n \times 1$ vectors $S = \{v_1, v_2, \ldots, v_k\}$, we will define the span of S as

$$\text{span}\,(S) = \{x \in R^n : x = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k\} \tag{69}$$

What this means is that every element of span (S) can be represented as a linear combination of vectors $v_1, v_2, \ldots, v_k$.

Remark 3. Note that this definition doesn't require vectors $v_1, v_2, \ldots, v_k$ to be linearly independent. It is also important to recognize that if vectors $v_1, v_2, \ldots, v_n$ happen to be a *basis* in R^n, then

$$\text{span}\,(v_1, v_2, \ldots, v_n) = R^n \tag{70}$$

(since every element of R^n can be represented as a linear combination of vectors $\{v_1, v_2, \ldots, v_n\}$).

Given a basis $\{v_1, v_2, \ldots, v_n\}$ in R^n, the Gram-Schmidt method allows us to construct a set of vectors $\{\xi_1, \xi_2, \ldots, \xi_n\}$ such that

$$\text{span}\,(v_1, v_2, \ldots, v_n) = \text{span}\,(\xi_1, \xi_2, \ldots, \xi_n) \tag{71}$$

and $\langle \xi_i, \xi_j \rangle = 0$ when $i \neq j$. Equation (71) ensures that $\{\xi_1, \xi_2, \ldots, \xi_n\}$ is a basis in R^n, and the elements of this set can be computed recursively as

$$\xi_k = v_k - \frac{\langle v_k, \xi_1 \rangle}{\langle \xi_1, \xi_1 \rangle} \cdot \xi_1 - \frac{\langle v_k, \xi_2 \rangle}{\langle \xi_2, \xi_2 \rangle} \cdot \xi_2 - \ldots - \frac{\langle v_k, \xi_{k-1} \rangle}{\langle \xi_{k-1}, \xi_{k-1} \rangle} \cdot \xi_{k-1} \tag{72}$$

if we assume that $\xi_1 = v_1$.

The following example illustrates how this procedure works.

Example 7. Let us consider vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}; \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}; \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{73}$$

which are linearly independent but are *not* orthogonal, since

$$\langle v_1, v_2 \rangle = \langle v_1, v_3 \rangle = \langle v_2, v_3 \rangle = 1 \tag{74}$$

In order to orthogonalize them, we will first set

$$\xi_1 = v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \tag{75}$$

and compute ξ_2 as

$$\xi_2 = v_2 - \frac{\langle v_2, \xi_1 \rangle}{\langle \xi_1, \xi_1 \rangle} \cdot \xi_1 \tag{76}$$

Observing that

$$\langle v_2, \xi_1 \rangle = 1 \tag{77}$$

and

$$\langle \xi_1, \xi_1 \rangle = 2 \tag{78}$$

(76) becomes

$$\xi_2 = v_2 - \frac{1}{2} \cdot \xi_1 = \begin{bmatrix} -1/2 \\ 1 \\ 1/2 \end{bmatrix} \tag{79}$$

To find ξ_3, we now need to compute

$$\xi_3 = v_3 - \frac{\langle v_3, \xi_1 \rangle}{\langle \xi_1, \xi_1 \rangle} \cdot \xi_1 - \frac{\langle v_3, \xi_2 \rangle}{\langle \xi_2, \xi_2 \rangle} \cdot \xi_2 \tag{80}$$

This is not difficult to do, since we know that

$$\langle v_3, \xi_1 \rangle = 1 \tag{81}$$

$$\langle v_3, \xi_2 \rangle = 1/2 \tag{82}$$

and

$$\langle \xi_2, \xi_2 \rangle = 3/2 \tag{83}$$

Substituting these values into (80), we obtain

$$\xi_3 = v_3 - \frac{1}{2} \cdot \xi_1 - \frac{1}{3} \cdot \xi_2 = \begin{bmatrix} 2/3 \\ 2/3 \\ -2/3 \end{bmatrix} \tag{84}$$

At this point the procedure is complete, and we can claim that ξ_1, ξ_2 and ξ_3 are orthogonal vectors. If we divide each of these vectors by their norm, we obtain an *orthonormal basis* in R^3 whose elements are

$$\frac{\xi_1}{\|\xi_1\|} = \begin{bmatrix} 0.7071 \\ 0 \\ 0.7071 \end{bmatrix}; \quad \frac{\xi_2}{\|\xi_2\|} = \begin{bmatrix} -0.4082 \\ 0.8165 \\ 0.4082 \end{bmatrix}; \quad \frac{\xi_3}{\|\xi_3\|} = \begin{bmatrix} 0.5774 \\ 0.5774 \\ -0.5774 \end{bmatrix} \tag{85}$$

Eigenvalues and Eigenvectors

Before we move on to a more advanced discussion of scalar products and norms, it is important to say a few words about *eigenvalues* and *eigenvectors*. The following definition describes how these two concepts are understood in linear algebra.

Definition 2. A nonzero vector y_i is said to be an *eigenvector* of matrix A if there exists a number λ_i (which can be real or complex) such that

$$Ay_i = \lambda_i y_i \tag{86}$$

We refer to λ_i as the *eigenvalue* that corresponds to y_i.

The computation of eigenvalues and eigenvectors is illustrated by the following example.

Example 8. Consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix} \tag{87}$$

If λ is an eigenvalue of A, then equation

$$(A - \lambda I)y = 0 \tag{88}$$

must be satisfied for some nonzero vector y (in this expression, I denotes the identity matrix).

Since system (88) has a nontrivial solution if and only if

$$\Delta(\lambda) = \det(A - \lambda I) = 0 \tag{89}$$

it follows that the eigenvalues of A are actually the roots of polynomial $\Delta(\lambda)$ (which is known as the *characteristic polynomial* of matrix A). In our example, this polynomial has the form

$$\Delta(\lambda) = \lambda^2 + 5\lambda + 6 = (\lambda + 2)(\lambda + 3) \tag{90}$$

which implies that the eigenvalues of A are $\lambda_1 = -2$ and $\lambda_2 = -3$.

How can we determine the eigenvectors that correspond to λ_1 and λ_2? For $\lambda_1 = -2$, (88) becomes

$$\begin{bmatrix} 2 & 1 \\ -6 & -3 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{91}$$

which is a singular system of equations. This property becomes obvious after we multiply the first row by 3 and add it to row 2, which produces

$$\begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{92}$$

Observing that the second equation is satisfied for *any* choice of y_{11} and y_{12}, we can set $y_{12} = t$ (where t is an arbitrary real number). Component y_{11} can then be uniquely determined as

$$y_{11} = -\frac{t}{2} \tag{93}$$

Given this expression, it follows that the solutions of system (91) have the general form

$$y_1 = t \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} \tag{94}$$

The fact that t is an unspecified number indicates that we can associate *infinitely many* eigenvectors with λ_1. Note, however, that all of these eigenvectors are proportional to vector

$$\bar{y}_1 = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} \tag{95}$$

This allows us to choose $\bar{y}_1$ as a "representative" for the entire set.

We can proceed in a similar way to obtain the eigenvector that corresponds to $\lambda_2 = -3$. In that case we have

$$\begin{bmatrix} 3 & 1 \\ -6 & -2 \end{bmatrix} \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{96}$$

which becomes

$$\begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{97}$$

after mutliplying the first row by 2 and adding it to the second row. It is not difficult to verify that all the solutions of system (97) have the form

$$y_2 = t \begin{bmatrix} -1/3 \\ 1 \end{bmatrix} \tag{98}$$

which means that

$$\bar{y}_2 = \begin{bmatrix} -1/3 \\ 1 \end{bmatrix} \tag{99}$$

can be identified as the "representative" eigenvector that corresponds to λ_2.

When it comes to eigenvectors and eigenvalues, *symmetric matrices* (which satisfy $A^T = A$) have some particularly interesting properties. The following three results will be especially useful for our purposes (for the sake of consistency, I have numbered the theorems and lemmas in the same way as in the textbook).

Theorem 1.1. If A is a real, symmetric matrix, all of its eigenvalues must be *real*, and the corresponding eigenvectors must be *orthogonal*.

Corollary 1.1 If the eigenvalues of an $n \times n$ symmetric matrix are *distinct*, its normalized eigenvectors represent an *orthonormal basis* in R^n.

Lemma 1.6. Let A be a symmetric matrix with distinct eigenvalues, and suppose that its normalized eigenvectors constitute the columns of matrix T. The inverse of T will then satisfy

$$T^{-1} = T^T \tag{100}$$

Similarity Transformations and the Jordan Form

Since the eigenvalues of a matrix are one of its most important characteristics, it would be interesting to determine which types of transformations leave them unchanged. The following definition and the subsequent lemma will help us answer this question.

Definition 3. Matrices A and $\tilde{A}$ are said to be *similar* if there exists a nonsingular matrix T such that

$$\tilde{A} = T^{-1}AT \tag{101}$$

Lemma 1.3. If matrices A and $\tilde{A}$ are similar, they must have the *same* set of eigenvalues.

The fact that similar matrices have the same set of eigenvalues has several important implications. The following one will be particularly useful for our purposes.

Lemma 1.7. Let A be a symmetric matrix with distinct eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, and suppose that its eigenvectors constitute the columns of matrix T. Matrix A can then be transformed into a *diagonal matrix*

$$\Lambda = T^{-1}AT \tag{102}$$

which has the form

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{103}$$

This matrix is know as the *Jordan cannonical form* of A.

Remark 4. Note that Theorem 1.1 ensures that eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are *real*, since A is assumed to be symmetric. We also know that in this case $T^{-1} = T^T$ (by virtue of Lemma 1.6), so matrix Λ can be equivalently expressed as

$$\Lambda = T^T A T \tag{104}$$

Vector Norms Revisited

When we introduced the notion of a scalar product, we showed that the norm of vector $x \in R^n$ can be defined as

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \tag{105}$$

This particular norm is referred to as the *Euclidean norm*, and is typically denoted by $\|x\|_2$. The notion of a Euclidean norm can be extended to complex vectors as well, since the scalar product

$$\langle x, y\rangle = x^T y^* = x_1 y_1^* + x_2 y_2^* + \ldots + x_n y_n^* \tag{106}$$

(which was defined in Example 6) ensures that

$$\langle x, x\rangle = |x_1|^2 + |x_2|^2 + \ldots + |x_n|^2 \tag{107}$$

is a real number.

The following lemma establishes an important connection between scalar products and the Euclidean norm, which is known as the *Schwarz inequality.*

Lemma 1.4. If x and y are two vectors in R^n, their scalar product satisfies

$$|\langle x, y\rangle| \le \|x\|_2 \|y\|_2 \tag{108}$$

Remark 5. It is it not difficult to show that this inequality holds for any pair of *complex* vectors $x, y \in C^n$ as well.

The Infinity Norm and its Generalizations

The infinity norm of vector x is defined as

$$\|x\|_\infty = \max_i |x_i| \tag{109}$$

Given a vector ω whose components satisfy $\omega_i > 0$ $(i = 1, 2, \ldots, n)$, expression (109) can be generalized as

$$\|x\|_\infty^\omega = \max_i \frac{1}{\omega_i}|x_i| \tag{110}$$

This norm (which is known as the *weighted infinity norm*) provides us with some added flexibility, since there is obviously an unlimited number of ways to choose vector ω.

It is easily verified that these two norms satisfy the following four properties (the same is true for the Euclidean norm).

1. The norm of any vector $x \in R^n$ satisfies $\|x\| \geq 0$.
2. For any $x \in R^n$ and $a \in R$ we have that $\|ax\| = |a| \, \|x\|$.
3. The norm of x is zero if and only if $x = 0$.
4. The inequality
$$\|x + y\| \leq \|x\| + \|y\| \tag{111}$$
holds for any two vectors $x, y \in R^n$.