2.1 Introduction

Vector spaces are central to the development of state-space control theory. They provide a useful generalization of familiar geometric notions regarding lines and planes to more complicated, higher dimensional, settings. In addition, vector spaces provide the tools for developing the theory of linear equations and other aspects of matrix theory. In this chapter, we start with the basic definitions of linear algebra: vector spaces, linear independence, bases, and dimension, and proceed to derive matrix theory results which are used in this book, namely: solutions to linear equations including least-squares and minimum-norm solutions, eigenvalues and eigenvectors, and the singular value decomposition.

This chapter contains a number of useful results from linear algebra. We label these results as “facts” and enclose them in a box. The reason for using this terminology is that we view the elementary theorems from linear algebra as useful facts which will be applied in subsequent work. The proofs of some of the facts are given to convey a sense of the thought processes that go into the development of linear algebra. The reader who is already familiar with linear algebra may simply wish to read the “fact boxes” as a refresher. A summary of the notation which is introduced in this chapter is given in Table 2.1.

Vector Spaces

An abstract vector space consists of two sets $\mathcal{X}$ and $\mathcal{F}$ together with certain properties that the elements of those sets must obey. The elements of the set $\mathcal{X}$ are called vectors and the
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Page of First Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x \in \mathcal{X}$</td>
<td>$x$ is a member of the set $\mathcal{X}$</td>
<td>25</td>
</tr>
<tr>
<td>$\forall$</td>
<td>for all</td>
<td>25</td>
</tr>
<tr>
<td>$\implies$</td>
<td>implies</td>
<td>29</td>
</tr>
<tr>
<td>$\iff$</td>
<td>if and only if</td>
<td>25</td>
</tr>
<tr>
<td>$A^T$</td>
<td>transpose of the matrix $A$</td>
<td>35</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
<td>$</td>
</tr>
<tr>
<td>$A^{-1}$</td>
<td>inverse of the matrix $A$</td>
<td>45</td>
</tr>
<tr>
<td>$|x|$</td>
<td>norm of the vector $x$</td>
<td>49</td>
</tr>
<tr>
<td>$R^\perp$</td>
<td>orthogonal complement of the subspace $R$</td>
<td>54</td>
</tr>
<tr>
<td>$\text{row}(A)$</td>
<td>row-space of the matrix $A$</td>
<td>35</td>
</tr>
<tr>
<td>$\text{col}(A)$</td>
<td>column-space of the matrix $A$</td>
<td>34</td>
</tr>
<tr>
<td>$\rho(A)$</td>
<td>rank of the matrix $A$</td>
<td>56</td>
</tr>
<tr>
<td>$\nu(A)$</td>
<td>nullity of the matrix $A$</td>
<td>57</td>
</tr>
</tbody>
</table>

**Table 2.1** Summary of mathematical notation.
elements of the set $\mathcal{F}$ are called *scalars*. An addition operation must be defined for the vectors, multiplication and addition must be defined for the scalars, and a multiplication between scalars and vectors must be defined. While all these operations must obey many properties to define a vector space [42, 10], the most important properties are the following:

1. $\forall x, y \in \mathcal{X}, \ x + y \in \mathcal{X}$.

2. $\forall \alpha \in \mathcal{F}, \ \forall x \in \mathcal{X}, \ \alpha x \in \mathcal{X}$.

In words, the first property says that a vector space is closed under addition – the sum of any two vectors from the set $\mathcal{X}$ must also be in the set. The second property says that a vector space is closed under scalar multiplication.

In this chapter, linear algebra is used to develop matrix theory, and the matrices consist of real numbers. The rows and columns of an $m \times n$ matrix are, respectively, $n$-tuples and $m$-tuples of real numbers. Thus we deal mostly with the vector space consisting of real-valued $n$-dimensional vectors, which we now define.

### 2.2 The Vector Space $\mathbb{R}^n$

Let the set $\mathcal{X} = \mathbb{R}^n$, the set of $n$-tuples of real numbers, where $n$ is some finite integer. We assume that the $n$-tuples are arranged in a column, so that a typical element of $\mathcal{X}$ looks as follows

\[
\mathbf{x} = \begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}.
\] (2.1)

Let the set of scalars $\mathcal{F} = \mathbb{R}$. Then it can be shown that these two sets, together with the usual definitions of addition and multiplication, form a vector space [42, 10]. We will use lower-case boldface characters for an $n$-tuple (vector), and a subscripted normal font to indicate an element of an $n$-tuple. A subscripted boldface symbol is used to define a vector which is to be distinguished by other vectors with different subscripts. Also, we will refer to the vector space simply as $\mathbb{R}^n$. The associated set $\mathcal{F}$ of real numbers to be used as scalars will be understood.

Two vectors are said to be equal if they are equal element wise; that is

\[
\mathbf{x} = \mathbf{y} \iff x_i = y_i, \ i = 1, \ldots, n.
\] (2.2)

In a similar way, addition and subtraction of vectors is done element wise

\[
\mathbf{x} \pm \mathbf{y} = \begin{bmatrix}
x_1 \pm y_1 \\
\vdots \\
x_n \pm y_n
\end{bmatrix}.
\] (2.3)

The zero vector in $\mathbb{R}^n$ is denoted $\mathbf{0}$ and has the property that $\mathbf{x} \pm \mathbf{0} = \mathbf{x}, \ \forall \mathbf{x} \in \mathbb{R}^n$. Clearly we must have

\[
\mathbf{0} = \begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix}.
\] (2.4)
The multiplication of a vector by a scalar is also defined element-wise as follows, where \( x \) and \( y \) are vectors in \( \mathbb{R}^n \) and \( \alpha \) is a scalar (real number):

\[
y = \alpha x \iff y_i = \alpha x_i , \ i = 1, \cdots, n.
\] (2.5)

The above definition of multiplication of a scalar times a vector is easily shown to satisfy the following properties for any two real numbers \( \alpha \) and \( \beta \), and for any vectors \( x \) and \( y \):

1. \( (\alpha\beta)x = \alpha(\beta x) \).
2. \( (\alpha + \beta)x = \alpha x + \beta x \).
3. \( \alpha(x + y) = \alpha x + \alpha y \).
4. \( 1x = x \).

It is also easy to see that the two closure properties mentioned in the introduction to this chapter are satisfied.

### 2.3 Linear Independence, Bases, and Subspaces

Given a collection of \( m \) vectors \( x_1, \cdots, x_m \), consider the vector \( y \) formed from the \( x \)'s as follows

\[
y = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m.
\] (2.6)

The vector \( y \) is said to be a **linear combination** of the vectors \( x_1, \cdots, x_m \) and the numbers \( \alpha_1, \cdots, \alpha_m \) are called the **expansion coefficients of the linear combination**. If all the \( \alpha \)'s are zero, the linear combination is said to be **trivial**. If at least one \( \alpha \) is non-zero, the linear combination is said to be **non-trivial**.

**EXAMPLE 2.1**

Consider the following set of vectors in \( \mathbb{R}^n \)

\[
e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad e_n = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 1 \end{bmatrix}
\]

and consider an arbitrary vector \( x \) in \( \mathbb{R}^n \)

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\]

It is easy to see that

\[
x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.
\]

Since \( x \) is arbitrary, we see that any vector in \( \mathbb{R}^n \) can be written as a linear combination of \( e_1, \cdots, e_n \) and that the coefficients of the linear combination are just the elements of \( x \).
Definition 2.1 Let $X$ be a set of vectors in $\mathbb{R}^n$. The vectors in $X$ are said to be linearly independent if the only linear combination of them which equals the zero vector is the trivial linear combination (all coefficients equal zero). If a non-trivial linear combination of the vectors equals the zero vector, the vectors in $X$ are said to be linearly dependent.

In light of the above definition, a procedure to check if a given set of vectors is linearly dependent or independent must examine the following equation

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m = 0.$$  

(2.7)

To prove that the vectors are linearly independent, one must show that the above equation implies that all the $\alpha$'s equal zero. On the other hand, if it is possible to find a set of scalars $\alpha_1, \cdots, \alpha_m$ which are not all zero such that (2.7) is satisfied, then the given vectors are linearly dependent.

**Example 2.2**

Consider the vectors $e_1, \cdots, e_n$ introduced in Example 2.1. To test if these vectors are linearly independent or not, we use (2.7) which is written as follows

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_1 + \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \alpha_2 + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \alpha_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$  

or

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$  

The above equation says that all the coefficients equal zero, so the vectors $e_1 \cdots e_n$ are linearly independent.

**Fact 2.1** Given a set of vectors $x_1 \cdots x_m$, if one of the vectors, say $x_k$, is the zero vector, then the given vectors are linearly dependent.

**Proof.** Equation (2.7) is satisfied with the following choice of coefficients

$$\alpha_i = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}$$

which are not all zero. Thus the vectors are linearly dependent.
Fact 2.2 A set of vectors is linearly dependent if and only if one of the vectors in the set can be written as a linear combination of the others.

PROOF: If the vectors are linearly dependent, we know that there exists a set of coefficients which are not all zero such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m = 0.$$ 

Suppose that $\alpha_k$ is a non-zero coefficient. Then we can solve for $x_k$ from the above equation as follows

$$x_k = -\frac{1}{\alpha_k} [\alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \alpha_{k+1} x_{k+1} + \cdots + \alpha_m x_m].$$ \hfill (2.8) 

That is, one of the vectors can be written as a linear combination of the others. On the other hand, if we know that for a given set of vectors $\{x_1, \cdots, x_m\}$, one of them (say the $k$th vector) can be written as a linear combination of the others, then we have

$$x_k = \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \alpha_{k+1} x_{k+1} + \cdots + \alpha_m x_m$$

for some set of expansion coefficients. Rearranging the previous equation as follows:

$$\alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} - 1 \cdot x_k + \alpha_{k+1} x_{k+1} + \cdots + \alpha_m x_m = 0$$

shows that (2.7) is satisfied with at least one nonzero coefficient (the coefficient of $x_k$). Thus, from Definition 2.1 the set of vectors is linearly dependent.

The above fact says that the vector $x_k$ is redundant as far as linear combinations go. In other words, any linear combination of the vectors $x_1, \cdots, x_m$ can be written as linear combination of the vectors $x_1, \cdots, x_{k-1}, x_{k+1}, x_m$ simply by replacing $x_k$ by its expansion (2.8) in terms of the other vectors. Note that the above fact does not say that every vector in a linear dependent set of vectors can be written as a linear combination of the others. This is demonstrated in the following example.

EXAMPLE 2.3

Consider the following set of vectors

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ x_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ x_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$ 

These vectors are linearly dependent because (2.7) is satisfied with the coefficients $\alpha_1 = 1, \ alpha_2 = 1, \ alpha_3 = -1$, and $\alpha_4 = 0$. Note that either $x_1, x_2,$ or $x_3$ can be written in terms of the other vectors. But $x_4$ cannot be written as a linear combination of the other vectors.
Although Definition 2.1 is the standard definition of a set of linearly independent vectors, it is sometimes more helpful to characterize such set as follows:

**Fact N.1** A set of vectors is linearly independent if and only if none of the vectors in the set can be written as a linear combination of the others.

**PROOF:** Left as a exercise.

Linearly independent vectors have an important property of uniqueness of linear combinations. This property is expressed as

**Fact 2.3** Any vector that is a linear combination of linearly independent vectors has unique expansion coefficients.

**PROOF.** Suppose a vector \( y \) can be expressed as a linear combination of a linearly independent set of vectors \( x_1, \cdots, x_m \). Suppose that there are two sets of expansion coefficients for \( y \) as shown below

\[
y = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m
\]

and

\[
y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m.
\]

Subtracting the second equation from the first yields

\[
0 = (\alpha_1 - \beta_1) x_1 + (\alpha_2 - \beta_2) x_2 + \cdots + (\alpha_m - \beta_m) x_m.
\]

In the above equation, the linearly independent vectors \( x_i \) sum to the zero vector, so the expansion coefficients must all equal zero. Thus \( \alpha_i = \beta_i, \ i = 1, \cdots, m. \) In other words, the two expansions of the vector \( y \) must have identical expansion coefficients.

**Definition 2.2** If a vector \( x_{k+1} \) can be written as a linear combination of the vectors \( x_1, \cdots, x_k \), then we say that \( x_{k+1} \) is **linearly dependent on** \( x_1, \cdots, x_k \). In this case the set of vectors \( x_1, \cdots, x_{k+1} \) is linearly dependent. If \( x_{k+1} \) cannot be written as a linear combination of the vectors \( x_1, \cdots, x_k \), then we say that \( x_{k+1} \) is **linearly independent of** \( x_1, \cdots, x_k \).

The elements of the set \( \mathbb{R}^n \) satisfy certain properties mentioned in Section 2.2 which show that \( \mathbb{R}^n \) is a vector space. There are certain subsets of \( \mathbb{R}^n \) whose elements also satisfy the properties of a vector space. Thus these subsets of \( \mathbb{R}^n \) are themselves vector spaces, and are called **subspaces** of \( \mathbb{R}^n \). As shown in the following definition, there are only two conditions which must be checked to determine if a particular subset of \( \mathbb{R}^n \) is a subspace. All of the remaining vector space conditions are satisfied automatically by the fact that the elements of a subset of \( \mathbb{R}^n \) are elements of \( \mathbb{R}^n \) which is a vector space.

**Definition 2.3** Let \( S \) be a nonempty subset of \( \mathbb{R}^n \). Then \( S \) is a **subspace** of \( \mathbb{R}^n \) if

1. \( x, y \in S \implies x + y \in S \).

2. \( x \in S, \ alpha \in R \implies \alpha x \in S \).

In other words, a subspace is closed under addition and scalar multiplication. Note that \( S \) could equal \( \mathbb{R}^n \) in which case the above conditions are identical to the two conditions stated at the beginning of this chapter for \( \mathbb{R}^n \). If \( S \) is a proper subset of \( \mathbb{R}^n \), then the above two conditions say that \( S \) itself is a vector space. The two conditions also imply that the zero vector is an element of every subspace. The second condition says that if \( x \in S \) then \(-x \in S\) also. The first condition says that the sum of these vectors must be in \( S \), or \( x + (-x) = 0 \in S \). An example of subspaces of \( \mathbb{R}^n \) is given after the next fact and definition.

**Fact 2.4** Given \( x_1, \ldots, x_k \), let \( S \) be the set of all possible linear combinations of \( x_1, \ldots, x_k \). Then \( S \) is a subspace of \( \mathbb{R}^n \), and \( S \) is said to be the subspace generated by \( x_1, \ldots, x_k \).

**Proof.** Consider two vectors \( z \) and \( w \) both in \( S \). They can each be written as some linear combination of the vectors \( x_1 \cdots x_k \)

\[
\begin{align*}
z &= \alpha_1 x_1 + \cdots + \alpha_k x_k \\
w &= \beta_1 x_1 + \cdots + \beta_k x_k.
\end{align*}
\]

Summing the previous two equations yields

\[
z + w = (\alpha_1 + \beta_1)x_1 + \cdots + (\alpha_k + \beta_k)x_k \in S
\]

and the sum of \( z \) and \( w \) is in \( S \) because it can be written as a linear combination of \( x_1 \cdots x_k \). In a similar way, it can be shown that for any real number \( \gamma \), \( \gamma w \in S \) since it can be written as a linear combination of the vectors \( x_1, \ldots, x_k \) with expansion coefficients \( \gamma \beta_i \).

**Definition 2.4** A set of vectors \( x_1, \ldots, x_k \) is said to span a subspace \( S \) if every vector in \( S \) can be written as a linear combination of \( x_1, \cdots, x_k \).

**Example 2.4**

Consider the vector space \( \mathbb{R}^3 \) in which the elements of a vector \( x \) are \( x, y, z \) coordinates in Euclidean space

\[
x = \begin{bmatrix} x \\ y \\ z \end{bmatrix}
\]

The vector

\[
\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}
\]
generates a subspace of \( \mathbb{R}^3 \) consisting of the \( x \) axis. The vectors
\[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}, \quad \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\]
generate a subspace of \( \mathbb{R}^3 \) consisting of the \( x - y \) plane. It is not necessary for subspaces to be “parallel” to the coordinate axes. For example, the vector
\[
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}
\]
generates a subspace consisting of a line through the origin which is not parallel to a coordinate axis.

If the vectors which span a subspace are linearly dependent, then one of them can be removed from the set and the remaining vectors still span the subspace (see Fact 2.2 and the discussion following it). If the remaining set of vectors is still linearly dependent, then another vector can be removed. If all the redundant vectors were removed from the set, we would be left with a set of linearly independent vectors that would still span the subspace. Such a set is called a basis for the subspace. A formal definition of a basis is

**Definition 2.5** Let \( S \) be a subspace of \( \mathbb{R}^n \) and let \( \{ b_1, \ldots, b_k \} \) be a set of vectors in \( S \). Then \( B \) is said to be a basis for \( S \) if

1. The vectors \( b_1, \ldots, b_k \) are linearly independent, and
2. The vectors \( b_1, \ldots, b_k \) span the subspace \( S \).

From (2) of the above definition, we see that any vector \( x \in S \) can be written as a linear combination of the basis vectors
\[
x = \gamma_1 b_1 + \cdots + \gamma_m b_m.
\]
From (1) of the above definition and Fact 2.3, we know that the expansion coefficients for \( x \) are unique.

**Fact 2.5** The set of vectors \( e_1, \ldots, e_n \) introduced Example 2.1 is a basis for \( \mathbb{R}^n \). This set of vectors is called the standard basis for \( \mathbb{R}^n \).

**PROOF.** From Example 2.1 we see that any vector in \( \mathbb{R}^n \) can be written as a linear combination of \( e_1, \ldots, e_n \). From example 2.2 we know that the vectors \( e_1, \ldots, e_n \) are linearly independent. Thus the vectors satisfy both requirements for being a basis.

**Fact 2.6** Let \( S \) be a subspace with a given basis \( b_1 \cdots b_m \). Then every basis for \( S \) must contain exactly \( m \) vectors.

**PROOF.** The proof of this fact may be found in most linear algebra textbooks. For instance, see [42, 83, 84].
Since all bases for a subspace $S$ have the same number of elements, we have the following

**Definition 2.6** The dimension of a subspace $S$ is the number of elements in a basis for $S$.

Using the previous fact and Fact 2.5, we see that the dimension of $\mathbb{R}^n$ is $n$. Thus the name “$n$-dimensional Euclidean space” is appropriate for $\mathbb{R}^n$. A particular set of basis vectors called the standard basis was introduced in Fact 2.5. However the standard basis is not the only basis for $\mathbb{R}^n$. It turns out that any set of $n$ linearly independent vectors is a basis for $\mathbb{R}^n$. This fact is stated next for the general case of $m$-dimensional subspaces.

**Fact 2.7** Let $S$ be an $m$-dimensional subspace of $\mathbb{R}^n$. Then no set of linearly independent vectors in $S$ has more than $m$ elements. Furthermore, any set of $m$ linearly independent vectors in $S$ is a basis for $S$.

**PROOF.** The first part of this fact is a corollary to Fact 2.6. To prove the second part of this fact, let $b_1, \ldots, b_m$ be linearly independent vectors in $S$ and consider an arbitrary vector $x \in S$. Then the vectors $b_1, \ldots, b_m, x$ must be linearly dependent by the first part of this fact, since there are more than $m$ vectors, and $S$ has dimension $m$. Thus there exist coefficients $\alpha_0, \alpha_1, \ldots, \alpha_m$ not all zero such that

$$\alpha_0 x + \alpha_1 b_1 + \cdots + \alpha_m b_m = 0.$$  

We claim that $\alpha_0 \neq 0$, since if $\alpha_0 = 0$ then the above equation reduces to

$$\alpha_1 b_1 + \cdots + \alpha_m b_m = 0$$

and since $b_1, \ldots, b_m$ are linearly independent, this would imply that $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Thus all the $\alpha$’s (including $\alpha_0$) would be zero, which we know is not true. So we must have $\alpha_0 \neq 0$. Then we can solve for $x$ as follows

$$x = -\frac{1}{\alpha_0} [\alpha_1 b_1 + \cdots + \alpha_m b_m].$$

Since $x$ is arbitrary, we see that any vector in $S$ can be written as a linear combination of the vectors $b_1, \ldots, b_m$, i.e. these vectors span $S$. Since these vectors are also linearly independent, they satisfy Definition 2.5 and they are a basis for $S$.

Using the above Fact, we know that the vectors in Example 2.3 must be linearly dependent, because the number of vectors is greater than the dimension of the vector space. The following fact says that a set of linearly independent vectors in a subspace can always be extended to form a basis for the subspace.

**Fact 2.8** Let $S$ be a $m$-dimensional subspace of $\mathbb{R}^n$ and let $b_1, \ldots, b_k$ be linearly independent vectors in $S$ for some $k \leq m$. Then additional vectors $b_{k+1}, b_{k+2}, \ldots, b_m$ can be found such that $B = \{b_1, \ldots, b_m\}$ is a basis for $S$.

**PROOF.** If $k = m$, then we have $m$ linearly independent vectors in an $m$-dimensional subspace $S$, and by Fact 2.7, these vectors are a basis for $S$. Thus we don’t need to add
any vectors to the set to get a basis. Now assume that $k < m$. We claim that there exists a vector $b_{k+1} \in S$ such that the vectors $b_1, \ldots, b_{k+1}$ are linearly independent. If we could not choose such a linearly independent vector, then all vectors $x \in S$ would have the property that the vectors $x, b_1, b_2, \ldots, b_k$ are linearly dependent, and so $x$ could be written as a linear combination of $b_1, \ldots, b_k$ (see the proof of Fact 2.7). Then $b_1, \ldots, b_k$ would be $k$ linearly independent vectors which span $S$, with $k < m$. In other words, these vectors would be a basis for $S$ consisting of fewer than $m$ vectors. This is impossible by Fact 2.6. Thus our claim that we can find a vector $b_{k+1}$ such that the vectors $b_1, \ldots, b_{k+1}$ are linearly independent is indeed true. If $k + 1 < m$, the same argument shows that we can find a vector $b_{k+2}$ such that the vectors $b_1, \ldots, b_{k+2}$ are linearly independent. This argument can be repeated until finally we find a vector $b_m$ such that $b_1, \ldots, b_m$ are linearly independent. By Fact 2.7, these vectors are a basis for $S$, and they also contain the given vectors $b_1, \ldots, b_k$.

**Fact 2.9** Let $S$ be an $m$-dimensional subspace of $\mathbb{R}^n$ and let $B = \{b_1, \ldots, b_p\}$, $p \geq m$ be a set of vectors which spans $S$. Then $m$ vectors can be chosen from $B$ to be a basis for $S$.

**Proof.** Let $b_1 = b_i$ where $b_i \in B$ is the non-zero vector from $B$ with the smallest index $i$. Then let $b_2 = b_j$ where $b_j \in B$ with the smallest index $j$ such that the vectors $b_1, b_j$ are linearly independent. Continue in this way to extract a set of linearly independent vectors $B = \{b_1, \ldots, b_q\}$ from the given set of vectors $B$. We will now show that $B$ is a basis for $S$ and that $q = m$. To show this, let $b_k$ be any vector in $B$ that was not selected to be in $B$. Then the vectors $b_k, b_1, \ldots, b_q$ must be linearly independent. Since $b_1, \ldots, b_q$ are linearly independent, then $b_k$ can be written as a linear combination of them (see the proof of Fact 2.7). This is true for all vectors not selected to be in $B$. Thus each $b_1, \ldots, b_q$ can be written as a linear combination of vectors in $B$, so $B$ spans $S$. Since the vectors in $B$ are also linearly independent, $B$ is a basis for $S$. By Fact 2.6, $q$ must equal $m$, and the above procedure has extracted a basis for $S$.

The previous two facts show that a basis is a maximal independent set of vectors in a given subspace, and at the same time, a minimal spanning set of vectors. A basis cannot be made larger without losing independence, and it cannot be made smaller and still span the subspace.

**Fact 2.10** The set $S$ of all linear combinations of $m$ linearly independent vectors $b_1, \ldots, b_m$ is an $m$-dimensional subspace of $\mathbb{R}^n$.

**Proof.** From Fact 2.4, $S$ is a subspace spanned by the vectors $b_1, \ldots, b_m$. Since these vectors are also linearly independent, they form a basis for $S$ (Definition 2.5). Since there are $m$ vectors in this basis, the dimension of $S$ must be $m$ (Fact 2.6).
2.4 Matrices

The elements of the vector space $\mathbb{R}^n$ are columns of $n$ real numbers. While column vectors play an important role in state-space system theory, these vectors are often transformed by matrices. In addition, the theory of simultaneous linear equations, which is central to many results in state-space system theory, is developed in terms of subspaces defined by the rows or columns of a matrix. In this section, we give the basic definitions and operations of matrices.

**Definition 2.7** An $m \times n$ matrix is a rectangular array of numbers having $m$ rows and $n$ columns. The numbers $m$ and $n$ are called the dimensions of the matrix.

In this book, matrices are represented by upper-case boldface symbols. The elements of the matrix are written with double subscripts, the first of which indicates the row position of that element, and the second, the column position. For example, a $3 \times 4$ matrix $A$ has the following form.

$$A = \begin{bmatrix}
    a_{11} & a_{12} & a_{13} & a_{14} \\
    a_{21} & a_{22} & a_{23} & a_{24} \\
    a_{31} & a_{32} & a_{33} & a_{34}
\end{bmatrix}.$$

The number $a_{ij}$ is called the $(i, j)$-element of $A$. An $n \times 1$ matrix is a vector (sometimes called a column vector). A $1 \times n$ matrix is called a row vector.

Note that each column of an $m \times n$ matrix $A$ is an element of $\mathbb{R}^m$. For example, the first column of $A$, which we denote $a_1$ is

$$a_1 = \begin{bmatrix}
    a_{11} \\
    a_{21} \\
    \vdots \\
    a_{m1}
\end{bmatrix}.$$

Thus an $m \times n$ matrix may be thought of as a collection of vectors from $\mathbb{R}^m$ indexed according to their column position as follows

$$A = [a_1 \ a_2 \ \cdots \ a_n].$$

**Definition 2.8** Given an $m \times n$ matrix $A$, the vector space formed by all linear combinations of the columns of $A$ is called the column space of $A$, and is denoted $\text{col}(A)$. (In other treatments of matrices, the column space is sometimes called the range or image of $A$).

Given a matrix $A$, we can define a new matrix $A^T$ by setting the rows of $A^T$ equal to the columns of $A$. This leads to the following definition.

**Definition 2.9** Given an $m \times n$ matrix

$$A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \cdots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}$$
the $n \times m$ matrix obtained by interchanging the rows and columns of $A$ is called the transpose of the matrix $A$ and is denoted by $A^T$, where

$$A^T = \begin{bmatrix}
a_{11} & a_{21} & \cdots & a_{m1} \\
a_{12} & a_{22} & \cdots & a_{m2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1n} & a_{2n} & \cdots & a_{mn}
\end{bmatrix}$$

**EXAMPLE 2.5**

The transpose of a column vector

$$a = \begin{bmatrix}a_1 \\
a_2 \\
\vdots \\
a_n\end{bmatrix}$$

is a row vector

$$a^T = [a_1 \ a_2 \ \cdots \ a_n].$$

There are several useful identities involving the transpose operation which are shown in Table 2.2.

Previously we viewed a matrix $A$ as a collection of column vectors. We can also think of an $m \times n$ matrix $A$ as a collection of row vectors as follows.

$$A = \begin{bmatrix}
a_1^T \\
a_2^T \\
\vdots \\
a_m^T
\end{bmatrix}$$

where

$$a_i^T = [a_{i1} \ a_{i2} \ \cdots \ a_{im}], \ i = 1, \ldots, m.$$
1. \((\alpha \beta)A = \alpha(\beta A)\).
2. \((\alpha + \beta)A = \alpha A + \beta A\).
3. \(\alpha(A + B) = \alpha A + \alpha B\).
4. \(1 \cdot A = A\).

Table 2.3  Properties of scalar-matrix multiplication.

2.4.1 Operations with Matrices

Definition 2.11 Let \(A\) and \(B\) be \(m \times n\) matrices. The sum of \(A\) and \(B\) is the matrix \(C\) whose elements are given by

\[
c_{ij} = a_{ij} + b_{ij}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, n.
\]

We write

\[C = A + B.\]

Note that the sum of two matrices is defined only when they have the same dimensions, and that the addition of matrices is done element wise. Since the addition of real numbers is commutative and associative, it is easy to show that the addition of matrices is also commutative and associative so that the following fact is true.

Fact 2.11

1. \(A + B = B + A\).
2. \((A + B) + C = A + (B + C)\).

Definition 2.12 Let \(A\) be an \(m \times n\) matrix and let \(\alpha\) be a real number. The product of \(\alpha\) and \(A\), written \(\alpha A\) or \(A \alpha\) is the \(m \times n\) matrix whose elements are given by

\[
\alpha a_{ij}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, n.
\]

This definition says that the product of a scalar and a matrix is done element wise. A list of the properties of scalar-matrix multiplication is given in Table 2.3. Note that the properties are also true when the scalars appear to the right of the matrices.

Matrices can be multiplied by scalars as shown above. It is also possible to multiply two matrices together according to the following definition

Definition 2.13 Let \(A\) be a \(p \times m\) matrix and \(B\) be an \(m \times n\) matrix. The product of \(A\) and \(B\) is the \(p \times n\) matrix \(C\) whose elements are given by

\[
c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}, \quad i = 1, \cdots, p; \quad j = 1, \cdots, n.
\]

We write

\[C = AB.\]
1. \((AB)C = A(BC)\).
2. \(A(B + C) = AB + AC\).
3. \((A + B)C = AC + BC\).
4. \(\alpha(AB) = (\alpha A)B = A(\alpha B)\).
5. \(AB \neq BA\), in general.

Table 2.4 Properties of matrix multiplication.

An example of matrix multiplication is given on page 39 after several interpretations of matrix multiplication are given. Three important aspects of matrix multiplication can be inferred from the definition, and are listed below. Properties of matrix multiplication are given in Table 2.4.

1. A product of two matrices is only defined when the number of columns of the first factor equals the number of rows of the second factor.
2. The product matrix has the same number of rows as the first factor and the same number of columns as the second factor, as shown in the following equation

\[
C_{p \times n} = A_{p \times m}B_{m \times n}.
\]

3. Matrix multiplication does not commute, in general. The matrices \(AB\) and \(BA\) may not have the same dimensions. Even if the dimensions are the same, it is not necessarily true that \(AB = BA\).

We now investigate several different expressions of matrix multiplication. To do so, it is useful to first define two types of vector multiplication, and two types of matrix-vector multiplication.

Definition 2.14 Given two vectors \(x\) and \(y\) in \(\mathbb{R}^n\), their inner product \(\gamma\) is the real number given by

\[
\gamma = x^T y = y^T x = \sum_{i=1}^{n} x_i y_i.
\]

Definition 2.15 Given two vectors \(x \in \mathbb{R}^m\) and \(y \in \mathbb{R}^n\), their outer product results in a matrix. The outer product of \(x\) with \(y\) gives the \(m \times n\) matrix \(A\) defined by

\[
A = xy^T, \quad a_{ij} = x_i y_j.
\]

The outer product of \(y\) with \(x\) gives the \(n \times m\) matrix \(B\) defined by

\[
B = yx^T, \quad b_{ij} = y_i x_j.
\]

By the rules of matrix transposition (see Fact 2.2), \(B^T = (yx^T)^T = xy^T = A\) so that \(A = B^T\).

Before presenting different interpretations of matrix multiplication, we first give two interpretations of matrix-vector multiplication. Since a row or column vector can also be
thought of as a matrix with a single row or column, respectively, the next two facts follow from the definition (2.13) of matrix multiplication.

**Fact 2.12** The product of a matrix \( A \) and a column vector \( x \) can be written as a linear combination of the columns of \( A \) in which the expansion coefficients are the elements of the vector \( x \), i.e.

\[
Ax = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n.
\]

A similar fact is true about the product of a row vector and a matrix.

**Fact 2.13** The product of a row vector \( y^T \) and a matrix \( A \) can be written as a linear combination of the rows of \( A \) in which the expansion coefficients are the elements of the vector \( y \), i.e.

\[
y^T A = \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} = y_1a_1^T + y_2a_2^T + \cdots + y_ma_m^T.
\]

We can now state several different interpretations of matrix multiplication.

**Elements of a Product Matrix as Inner Products** If a \( p \times m \) matrix \( A \) is written in terms of its rows, and an \( m \times n \) matrix \( B \) is written in terms of its columns

\[
A = \begin{bmatrix} a_1^T \\ \vdots \\ a_p^T \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}
\]

then the definition of matrix multiplication (Definition 2.13) gives the following

**Fact 2.14** If the matrix \( C = AB \), then the \((i,j)\)-element of \( C \) is the inner product of the \( i \)th row of \( A \) with the \( j \)th column of \( B \), i.e.

\[c_{ij} = a_i^T b_j.\]

**A Product Matrix as a Sum of Outer Products** A less obvious interpretation of matrix multiplication than the one given in Fact 2.14 can also be obtained from Definition 2.13. To get this interpretation, we write \( A \) in terms of its columns and \( B \) in terms of its rows as follows

\[
A = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}, \quad B = \begin{bmatrix} b_1^T \\ \vdots \\ b_m^T \end{bmatrix}
\]
Then we have the following

**Fact 2.15** If the matrix $C = AB$, then $C$ can be written as the sum of outer products of columns of $A$ with rows of $B$, i.e.

$$C = a_1 b_1^T + a_2 b_2^T + \cdots + a_m b_m^T.$$  \hspace{1cm} (2.9)

**Matrix Multiplication as a Weighted Sum of Columns or Rows** There are two other interpretations of matrix multiplication that are related to the one just given. We can write the matrix $B$ in terms of its columns, and then multiply $A$ by each column of $B$ as shown below.

**Fact 2.16** The matrix product $AB$ can be written column-wise as $A$ times the columns of $B$

$$AB = A \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_n \end{bmatrix}.$$  

From Fact 2.12 the matrix-vector product $Ab_i$ is a linear combination of columns of $A$. Thus each column of the product $AB$ is a linear combination of the columns of $A$.

A different way of expressing matrix multiplication is to write the matrix $A$ in terms of its rows, and then multiply each row of $A$ by the matrix $B$.

**Fact 2.17** The matrix product $AB$ can be written row-wise as rows of $A$ times $B$

$$AB = \begin{bmatrix} a_1^T \\ \vdots \\ a_p^T \end{bmatrix} B = \begin{bmatrix} a_1^T B \\ \vdots \\ a_p^T B \end{bmatrix}.$$  

From Fact 2.13, each row of the product $AB$ is a linear combination of the rows of $B$.

**EXAMPLE 2.6**

This example illustrates the different interpretations of matrix multiplication mentioned above. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$
Applying Fact 2.14 we compute the elements of the product matrix $C = AB$ using

inner products

$$
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
1 \\
0
\end{bmatrix}
= 3
\begin{bmatrix}
1 & 2 & 3 \\
0 & 1 & 1
\end{bmatrix}
= 5
$$

$$
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
0 \\
1
\end{bmatrix}
= 9
\begin{bmatrix}
0 & 1 & 1
\end{bmatrix}
= 11
$$

resulting in

$$
C = \begin{bmatrix}
3 & 5 \\
9 & 11
\end{bmatrix}.
$$

The same matrix product $AB$ can be computed in a different way using Fact 2.15 by computing the following outer products

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

$$
\begin{bmatrix}
3 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 \\
6 & 0 & 0
\end{bmatrix}
= \begin{bmatrix}
0 & 3 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 \\
1 & 1 & 1
\end{bmatrix}
= \begin{bmatrix}
0 & 3 & 0
\end{bmatrix}
$$

The product matrix is obtained by adding the outer products computed above

$$
\begin{bmatrix}
1 & 0 \\
4 & 0
\end{bmatrix}
+ \begin{bmatrix}
2 & 2 & 2
\end{bmatrix}
+ \begin{bmatrix}
0 & 3 & 0
\end{bmatrix}
= \begin{bmatrix}
3 & 5 & 3
\end{bmatrix}.
$$

We can also compute the matrix product $AB$ column wise as in Fact 2.16. The columns of $C$ are computed as follows

$$
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
0
\end{bmatrix}
= \begin{bmatrix}
3 & 5 \\
9 & 11
\end{bmatrix},
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
0 \\
1 \\
1
\end{bmatrix}
= \begin{bmatrix}
5 & 11
\end{bmatrix}
$$

Notice that the calculations for this method are the same as the method of inner products except that here the rows of $A$ are grouped together.

Finally, we can compute the matrix product $AB$ row wise as in Fact 2.17. The rows of $C$ are computed as follows

$$
\begin{bmatrix}
1 & 0 \\
1 & 1 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix}
= \begin{bmatrix}
3 & 5 \\
4 & 5 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
1 \\
0
\end{bmatrix}
= \begin{bmatrix}
9 & 11
\end{bmatrix}.
$$

Notice that the calculations for this method are the same as the method of inner products except that here the columns of $B$ are grouped together.

We now define some special matrices which are useful in the sequel.
Definition 2.16 An $n \times n$ matrix $A$ is said to be lower triangular if all elements above the main diagonal are zero. A lower triangular matrix has the following form:

$$A = \begin{bmatrix} * & 0 & \cdots & 0 \\ * & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix}.$$  

An upper triangular matrix is defined in a similar way – all elements below the main diagonal are zero. A matrix which is both upper triangular and lower triangular has nonzero entries only on the main diagonal. Such a matrix has the following definition.

Definition 2.17 A matrix is said to be diagonal if all elements not on the main diagonal are zero as shown below:

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}.$$  

We write $A = \text{diag}(a_1, a_2, \cdots, a_n)$.

Definition 2.18 The $n \times n$ diagonal matrix $\text{diag}(1, 1, \cdots, 1)$ is called the $n \times n$ identity matrix and is denoted $I_n$. The subscript indicating the dimension will often be omitted.

2.4.2 Operations with Partitioned Matrices*

The matrix operations introduced in the previous section were defined in terms of the addition and multiplication of real numbers, i.e. the elements of the matrices. Suppose we now consider a matrix to consist of partitioned submatrices. A submatrix is a rectangular array of numbers within a given matrix. For instance, a $4 \times 4$ matrix $A$ can be partitioned into 4 submatrices as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (2.10)$$

where

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_{12} = \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}, \quad A_{22} = \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix}.$$  

It is clear that a matrix can be partitioned into submatrices in several different ways. For instance, the partition shown in (2.10) is also true with the following definitions:

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad A_{12} = \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} a_{41} & a_{42} \end{bmatrix}, \quad A_{22} = a_{44}.$$
This example shows that when dealing with partitioned submatrices, it is important to know the dimensions of each submatrix.

We now examine the possibility of performing matrix operations with partitioned matrices as if the submatrices were the “elements” of the matrix. Of course when the “elements” (submatrices) are added or multiplied, they must be combined according to the rules of matrix addition and multiplication. The results for operations with partitioned matrices are summarized in Fact 2.18 on page 42.

**Fact 2.18** Let the matrices $A$ and $B$ be partitioned as follows

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1l} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kl} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}$$

where $A_{ij}$ is a submatrix of $A$ with dimensions $\alpha_i \times \delta_j$ and $B_{ij}$ is a submatrix of $B$ with dimensions $\beta_i \times \gamma_j$. Then the addition, multiplication, and transpose operations can be performed as follows:

1. If $k = m$, $\alpha_i = \beta_i$, and $\delta_j = \gamma_j$, then

$$A + B = \begin{bmatrix} C_{11} & \cdots & C_{1l} \\ \vdots & \ddots & \vdots \\ C_{k1} & \cdots & C_{kl} \end{bmatrix}$$

where $C_{ij} = A_{ij} + B_{ij}$.

2. If $l = m$ and $\delta_i = \beta_i$ then

$$AB = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{k1} & \cdots & C_{kn} \end{bmatrix}$$

where $C_{ij} = \sum_{q=1}^{l} A_{iq}B_{qj}$.

3. $A^T = \begin{bmatrix} A^T_{11} & \cdots & A^T_{k1} \\ \vdots & \ddots & \vdots \\ A^T_{1l} & \cdots & A^T_{kl} \end{bmatrix}$.
The first two items in Fact 2.18 require that the dimensions of the submatrices of \( A \) and \( B \) are such that addition and multiplication of the submatrices can be performed. Two matrices whose submatrices satisfy this property are said to be partitioned conformally. The addition of conformally partitioned matrices is accomplished by simply adding the corresponding submatrices. Multiplication is accomplished by taking an “inner product” with the \( i \)-th block row and the \( j \)-th block column of the matrices \( A \) and \( B \), respectively. The \( i \)-th block row of \( A \) is shown in the rectangle below:

\[
\begin{bmatrix}
A_{11} & \cdots & A_{1l} \\
\vdots & \ddots & \vdots \\
A_{i1} & \cdots & A_{il} \\
\vdots & \ddots & \vdots \\
A_{k1} & \cdots & A_{kl}
\end{bmatrix}
\]

and the \( j \)-th block column of \( B \) is shown in the rectangle below:

\[
\begin{bmatrix}
B_{11} & \cdots & B_{1j} \\
\vdots & \ddots & \vdots \\
B_{m1} & \cdots & B_{mj} \\
\vdots & \ddots & \vdots \\
B_{mn}
\end{bmatrix}
\]

The “inner product” of the \( i \)-th block row of \( A \) and the \( j \)-th block column of \( B \) is defined by the summation shown in item 2 of Fact 2.18. Compare this summation with the one given in Definition 2.13 for ordinary matrix multiplication.

**EXAMPLE 2.7**

Consider the following block matrices:

\[
A = \begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix},
\]

and

\[
B = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
0 & 1
\end{bmatrix} = \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix}.
\]

These matrices are each \( 2 \times 2 \) block matrices, and so the product matrix \( C = AB \) can be computed by 4 block inner products as shown below.

\[
C_{11} = \begin{bmatrix}
1 & 2 \\
4 & 5
\end{bmatrix} \begin{bmatrix}
1 \\
1
\end{bmatrix} + \begin{bmatrix}
3 \\
6
\end{bmatrix} \begin{bmatrix}
0 \\
1
\end{bmatrix} = \begin{bmatrix}
3 \\
9
\end{bmatrix}
\]

\[
C_{12} = \begin{bmatrix}
1 & 2 \\
4 & 5
\end{bmatrix} \begin{bmatrix}
0 \\
1
\end{bmatrix} + \begin{bmatrix}
3 \\
6
\end{bmatrix} \begin{bmatrix}
1 \\
1
\end{bmatrix} = \begin{bmatrix}
5 \\
11
\end{bmatrix}
\]
The resulting product matrix is then

\[
C = \begin{bmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{bmatrix} = \begin{bmatrix}
3 & 5 \\
9 & 11 \\
15 & 17
\end{bmatrix}.
\]

We also note that the definition of multiplication for partitioned matrices given in Fact 2.18 can also be used to define the multiplication of a partitioned matrix and a partitioned vector simply by having the matrix \( B \) have one column. To more clearly illustrate partitioned matrix-vector multiplication, we use different notation for the partitioned vector as shown below. Let

\[
A = \begin{bmatrix}
A_{11} & \cdots & A_{1l} \\
\vdots & \ddots & \vdots \\
A_{k1} & \cdots & A_{kl}
\end{bmatrix}
\]

where \( A_{ij} \) is a submatrix of \( A \) with dimensions \( \alpha_i \times \delta_j \) and let

\[
x = \begin{bmatrix}
x_1 \\
\vdots \\
x_l
\end{bmatrix}
\]

where \( x_j \) is a subvector of \( x \) with dimensions \( \delta_j \times 1 \). Then the matrix-vector product can be written as

\[
Ax = \sum_{j=1}^{l} A_{ij} x_j.
\]

To illustrate the transpose operation on a partitioned matrix, we show a matrix \( A \) below with its \( i \)-th block row enclosed in a rectangle. When \( A \) is transposed, this block row becomes the \( i \)-th block column of \( A^T \) and the individual submatrices are each transposed as follows

\[
A = \begin{bmatrix}
A_{11} & \cdots & A_{1l} \\
\vdots & \ddots & \vdots \\
A_{i1} & \cdots & A_{il} \\
\vdots & \ddots & \vdots \\
A_{k1} & \cdots & A_{kl}
\end{bmatrix}, \quad A^T = \begin{bmatrix}
A_{11}^T & \cdots & A_{1l}^T \\
\vdots & \ddots & \vdots \\
A_{i1}^T & \cdots & A_{il}^T \\
\vdots & \ddots & \vdots \\
A_{k1}^T & \cdots & A_{kl}^T
\end{bmatrix}
\]

### 2.4.3 Determinants and Matrix Inverses

Given an \( n \times n \) matrix \( A \), the \( n \times n \) matrix \( B \) is called the inverse of \( A \) if it satisfies

\[
AB = BA = I.
\]
1. If $A$ and $B$ are both $n \times n$, then $|AB| = |A||B|$.
2. $|A| = |A^T|$.
3. If $A$ is nonsingular then $|A^{-1}| = 1/|A|$.
4. Multiplying all elements of any one row (or column) of a matrix $A$ by a scalar $\alpha$ yields a matrix whose determinant is $\alpha|A|$.
5. Any multiple of a row (column) can be added to any other row (column) without changing the value of the determinant.
6. The determinant of a triangular matrix equals the product of its diagonal elements.
7. The determinant of a block triangular matrix equals the product of the determinants of the matrices on the main diagonal.

Table 2.5 Some properties of determinants.

The inverse of the matrix $A$ is denoted $A^{-1}$. Note that only square matrices can have an inverse, because the product matrices in (2.11) ($AB$ and $BA$) will have the same dimensions only when $A$ and $B$ are square. Not all square matrices possess an inverse, however.

One way to test whether or not a given matrix $A$ has an inverse is to compute the determinant of $A$, denoted $\det(A)$ or $|A|$. The determinant of a $2 \times 2$ matrix is given by the following formula

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \quad (2.12)$$

The determinants of larger matrices may be computed by combining determinants of smaller submatrices [10, 84, 83]. In this book we will not be concerned with the hand calculation of determinants.

The formula for the inverse of a $2 \times 2$ matrix is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.13)$$

This formula can be verified by substituting it into (2.11). Note that the determinant of the matrix appears in the denominator of the expression for the inverse. This means that the inverse of a $2 \times 2$ matrix is defined only when its determinant is nonzero. This turns out to be true in general, and an $n \times n$ matrix $A$ has an inverse if and only if $|A|$ is not equal to zero. If $|A| \neq 0$ we say that $A$ is nonsingular or invertible. If $|A| = 0$ we say that $A$ is singular. Some useful properties of determinants are given in Table 2.5. We conclude this
section with two important facts involving matrix inverses.

**Fact 2.19** Let \( A \) and \( B \) be nonsingular matrices having the same dimensions. Then
\[
(AB)^{-1} = B^{-1}A^{-1}.
\]

The following fact says that the inverse and transpose operations can be performed in either order on a nonsingular matrix with the same result.

**Fact 2.20** If \( A \) is a nonsingular matrix then
\[
(A^{-1})^T = (A^T)^{-1} \overset{\text{def}}{=} A^{-T}.
\]

### 2.4.4 The Gram Matrix Test for Linear Independence

It is often important to know when a given set of vectors is linearly independent. The definition of linear independence (Definition 2.1) does not provide a convenient way to test a given set of vectors for linear independence. In this section, we state without proof a computational test based on the determinant of a Gram matrix. A proof may be found in [10].

**Definition 2.19** The **Gram Matrix** \( G \) associated with the \( k \) vectors \( x_1, \ldots, x_k \in \mathbb{R}^n \) is a \( k \times k \) matrix whose \((i,j)\) element is given by \( g_{ij} = x_i^T x_j \). If the given vectors are arranged as columns in a matrix as follows
\[
X = \begin{bmatrix} x_1 & x_2 & \cdots & x_k \end{bmatrix}
\]
then the Gram matrix can be written as
\[
G = X^T X.
\]

The Gram matrix can be used to test whether or not a given set of vectors is linearly independent, as shown by the following

**Fact 2.21** A set of vectors \( x_1, \ldots, x_k \in \mathbb{R}^n \) is linearly independent if and only if the determinant of the Gram matrix associated with these vectors is not equal to zero.

**Example 2.8**

Let us use the Gram determinant to test if the following vectors are linearly independent
\[
x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}.
\]
The Gram matrix is
\[ G = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \]

The determinant of \( G \) is \( |G| = 3 \), and so the vectors are linearly independent.

The Gram matrix can also be used to test if a given vector is in a certain subspace. For example, if \( x_1 \) and \( x_2 \) are linearly independent vectors, then Fact 2.10 says that the set of all linear combinations of \( x_1 \) and \( x_2 \) is a 2-dimensional subspace of \( \mathbb{R}^3 \). Let us denote this subspace as
\[ R = \{ \text{all linear combinations of } x_1 \text{ and } x_2 \}. \] (2.14)

Suppose we are given a vector \( y \). How could we check if \( y \in R \)? The answer is as follows. If \( y \in R \), it can be expressed as a linear combination of \( x_1 \) and \( x_2 \). In this case the vectors \( x_1, x_2, y \) are linearly dependent (see Definition 2.2 on page 29). On the other hand, if the vectors \( x_1, x_2, y \) are linearly independent, then \( y \) cannot be expressed as a linear combination of \( x_1, x_2 \) and so \( y \notin R \). This procedure to test if a given vector belongs to a subspace is summarized in the following fact.

\begin{fact}
Let \( R \) be an \( m \)-dimensional subspace of \( \mathbb{R}^n \) and let \( x_1, x_2, \ldots, x_m \) be a basis for \( R \). Let \( y \) be a vector in \( \mathbb{R}^n \) and define \( W \) to be the following matrix
\[ W \overset{\text{def}}{=} \begin{bmatrix} x_1 & x_2 & \cdots & x_m & y \end{bmatrix}. \]
If the columns of \( W \) are linearly independent then \( y \notin R \). If the columns of \( W \) are linearly dependent the \( y \in R \).
\end{fact}

The test for linear independence can be done by computing the determinant of the Gram matrix \( G = W^T W \).

\begin{example}
Let \( R \) be the subspace defined by
\[ R = \{ \text{all linear combinations of } x_1 \text{ and } x_2 \} \]
where \( x_1 \) and \( x_2 \) are defined in the previous example. Check whether the following vectors are in \( R \)
\[ y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad z = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}. \]

We first form the Gram matrix for \( x_1, x_2, y \).
\[ G_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 3 \end{bmatrix}. \]

The determinant of \( G_1 \) can be calculated to be 1. Thus \( y \notin R \).
The Gram matrix for \( x_1, x_2, z \) is
\[
G_2 = \begin{bmatrix}
    1 & 1 & 0 \\
    0 & 1 & 1 \\
    1 & 3 & 2
\end{bmatrix} \begin{bmatrix}
    1 & 0 & 1 \\
    1 & 1 & 3 \\
    0 & 1 & 2
\end{bmatrix} = \begin{bmatrix}
    2 & 1 & 4 \\
    1 & 2 & 5 \\
    4 & 5 & 14
\end{bmatrix}.
\]

The determinant of \( G_2 \) can be calculated to be zero, and so \( z \in R \). Since \( z \in R \), it should be possible to express \( z \) as a linear combination of \( x_1 \) and \( x_2 \). It is easy to see that
\[
z = x_1 + 2x_2.
\]

The previous example talked about a subspace \( R \) as the set of all linear combinations of two vectors. We now show a more convenient representation for such a subspace. Let the \( m \)-dimensional subspace \( R \) be defined as follows
\[
R = \{ \text{all linear combinations of } x_1, x_2, \ldots, x_m \}.
\]
(2.15)

Using Fact 2.12 we can write any linear combination of \( x_1, \ldots, x_m \) as \( X\alpha \), where
\[
X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}
\]
and
\[
\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}
\]
is a vector of expansion coefficients. Thus every vector in the subspace \( R \) can be represented as \( X\alpha \) for some \( \alpha \), and a definition of \( R \) which is equivalent to (2.15) is
\[
R = \{ x = X\alpha \text{ for some } \alpha \}.
\]
(2.16)

We conclude this section with a corollary to Fact 2.21. Consider the case when \( k = n \) so that the matrix \( X \) is \( n \times n \). The Gram matrix is \( G = X^T X \) and its determinant is
\[
|G| = |X^T X|
\]
\[
= |X^T||X|
\]
\[
= |X|^2
\]
where we have used some of the properties of determinants given in Table 2.5 on page 45. Since \( |G| = |X|^2 \), then \( |G| = 0 \) if and only if \( |X| = 0 \). Thus to test the linear independence of \( n \) vectors in \( \mathbb{R}^n \) we do not need to form the Gram matrix. The determinant of the matrix formed from the given vectors can be used. In this discussion, the \( n \) vectors were the columns of the matrix \( X \). A similar development is possible when \( n \) vectors in \( \mathbb{R}^n \) are placed as rows of a matrix. This gives the following

**Fact 2.23** The columns (or rows) of a \( n \times n \) matrix \( A \) are linearly independent if and only if \( \det(A) \neq 0 \).
2.5 Matrix Subspaces and Projections

Many of the design formulas used in state-space control theory reduce to the solution of linear equations. The theory of linear equations is in turn rooted in the foundations of linear algebra. Previous sections of this chapter have reviewed basic facts from linear algebra and matrices. Before dealing specifically with linear equations, we first develop the four fundamental subspaces associated with a matrix. We also develop in this section the notions of orthogonality and projections. With these tools, the next section then develops the theory of linear equations.

2.5.1 Orthogonality

In the previous section, we defined the inner product of two vectors in $\mathbb{R}^n$ (Definition 2.14). The inner product has a simple geometric interpretation in $\mathbb{R}^2$. To develop this interpretation, we first define the length (norm) of a vector.

**Definition 2.20** The norm $^1$ of a vector $x \in \mathbb{R}^n$, denoted by $\|x\|$ is given by the positive square root

$$\|x\| = \sqrt{x^T x}.$$ 

The above definition of a norm generalizes the notion of Euclidean distance in $\mathbb{R}^3$ to arbitrary dimensions. Note in Fig. 2.1 that the length of the vector $x \in \mathbb{R}^2$ is equal to 5 (by the Pythagorean theorem), and this result is also obtained using the definition of the norm of $x$.

![Figure 2.1](image.png)  

**Figure 2.1** The length of the vector $x$ is given by $\|x\| = \sqrt{3^2 + 4^2} = 5$.

We can expand the norm squared of $x$ in terms of the components of $x$ as follows

$$\|x\|^2 = x^T x = x_1^2 + x_2^2 + \cdots + x_n^2.$$  

$^1$This norm is called the Euclidean norm or 2-norm. There are many other norms for vectors in $\mathbb{R}^n$, but only the 2-norm will be used in this book.
Because the right-hand side of the above equation consists of a sum of squares, it can only equal zero if each \( x_i \) is zero. This establishes the following fact.

**Fact 2.24** 
\[ ||x|| = 0 \iff x = 0. \]

Suppose now that \( x \) and \( y \) are two vectors in \( \mathbb{R}^2 \) as shown in Fig. 2.2, and let \( c = x - y \). By the law of cosines for the triangle shown in Fig. 2.2, we have

\[ \|c\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\| \cos \theta. \] \hspace{1cm} (2.17)

The square of the norm of \( c = x - y \) can be expanded as follows

\[ \|x - y\|^2 = (x - y)^T(x - y) \]
\[ = x^T x - x^T y - y^T x + y^T y \]
\[ = \|x\|^2 - 2x^T y + \|y\|^2. \] \hspace{1cm} (2.18)

Substituting the above equation into (2.17) and solving for \( \cos \theta \) yields

\[ \cos \theta = \frac{x^T y}{\|x\|\|y\|}. \] \hspace{1cm} (2.19)

In words, the above equation says that the cosine of the (acute) angle between the vectors \( x \) and \( y \) is given by the inner product \( x^T y \) normalized by the lengths of \( x \) and \( y \). The development of (2.19) assumed the vector space was \( \mathbb{R}^2 \), but (2.19) can be evaluated for vectors in \( \mathbb{R}^n \) for arbitrary \( n \). For \( n \geq 3 \), we take (2.19) as the definition of the angle between vectors. This definition is then used to define the notion of orthogonal vectors.

The vectors \( x \) and \( y \) in Fig. 2.2 are orthogonal if the angle \( \theta = 90^\circ \). In this case, \( \cos \theta = 0 \), and from (2.19) this means that \( x^T y = 0 \). The inner product is used to define orthogonality for two vectors as follows.

**Definition 2.21** Two vectors \( x, y \in \mathbb{R}^n \) are said to be orthogonal if their inner product \( x^T y \) is zero.

By defining orthogonality in terms of the inner product, the definition applies to arbitrary dimensions. The definition of orthogonality between two vectors can be used to define a notion of orthogonality for sets of vectors, and for subspaces.
Definition 2.22 A set of nonzero vectors \( \{x_1, x_2, \cdots, x_k\} \) is said to be an orthogonal set of vectors if \( x_i \) is orthogonal to \( x_j \) for all \( i \neq j \). If, in addition, \( \|x_i\| = 1, i = 1, \cdots, k \) then the set of vectors is said to be orthonormal.

Fact 2.25 An orthogonal (or orthonormal) set of nonzero vectors is linearly independent.

\[
\text{PROOF.} \text{ To establish this result with a proof by contradiction, assume that the given set of vectors is linearly dependent. Then by Fact 2.2, it is possible to express one vector, say } x_i, \text{ as a linear combination of the others. That is}
\]
\[
x_i = \alpha_1 x_1 + \cdots + \alpha_{i-1} x_{i-1} + \alpha_{i+1} x_{i+1} + \cdots + \alpha_k x_k.
\]
If we multiply both sides of this equation by \( x_i^T \) and use the fact that all the inner products on the right-hand side of the equation are zero because of orthogonality, the result is
\[
\|x_i\|^2 = 0.
\]
From Fact 2.24 this implies that \( x_i = 0 \) which contradicts the fact that the given vectors are not zero.

Definition 2.23 Two subspaces \( S_1 \) and \( S_2 \) of \( \mathbb{R}^n \) are said to be orthogonal subspaces if, for any \( x \in S_1 \) and any \( y \in S_2 \), the vectors \( x \) and \( y \) are orthogonal. If \( x_1, \cdots, x_p \) is a basis for \( S_1 \) and \( y_1, \cdots, y_q \) is a basis for \( S_2 \) then \( S_1 \) and \( S_2 \) are orthogonal if \( x_i^T y_j = 0, i = 1, \cdots, p; j = 1, \cdots, q \).

2.5.2 Orthogonal Projections

The definition of orthogonality given in the previous section is a generalization of the familiar geometric concept of perpendicular lines. The importance of orthogonality comes from its use in deriving optimal approximations. From geometry, it is known that the shortest distance from a point to a line occurs along the line segment which is perpendicular to the given line. Similarly, the shortest distance from a point to a plane occurs along the line segment which is perpendicular to the plane.

The optimal solutions mentioned above can be expressed in the language of linear algebra as follows: the shortest distance between a vector and a subspace occurs along a vector which is orthogonal to the subspace. Fig. 2.3 shows a vector \( y \) (represented by an arrow) and a subspace \( R \) (represented by a plane). Of all vectors which connect the plane and the vector \( y \), the vector \( y_e \) is the one with the smallest norm. The vector \( y_e \) is orthogonal to the subspace \( R \). The vector \( y_p \) is the vector in \( R \) which is the best approximation to \( y \). The vector \( y_p \) is called the projection of \( y \) onto \( R \).

Assume that the subspace \( R \) is defined as the subspace spanned by a given set of linearly independent vectors \( x_1, \cdots, x_r \). From Fact 2.10, the given vectors are a basis for \( R \). The question we would like to answer now is: given a vector \( y \) and a basis for a subspace \( R \), how can we compute \( y_p \), the projection of \( y \) onto \( R \)?

The answer is quite simple, and it comes from the fact that the “error vector”

\[
y_e = y - y_p \tag{2.20}
\]
Figure 2.3 \( y_p \) is the best approximation to \( y \) of any vector in \( R \). \( y_p \) is obtained by projecting \( y \) onto \( R \). is orthogonal to the subspace \( R \), which means that the error vector must be orthogonal to every vector in a basis for \( R \) (see Definition 2.23), or

\[
y_e^T x_i = 0, \quad i = 1, \ldots, r.
\]  

(2.21)

However these orthogonality equations cannot be evaluated directly because \( y_e \) is a function of \( y_p \) (see (2.20)), which is the vector we are trying to find! We do know that \( y_p \) is in the subspace \( R \), and from (2.16) on page 48 we know that any vector in \( R \) can be represented as \( X\alpha \) for some vector \( \alpha \) of expansion coefficients, i.e.

\[
y_p = X\alpha \text{ for some } \alpha \text{ where } X = \begin{bmatrix} x_1 & \cdots & x_r \end{bmatrix}.
\]  

(2.22)

We can substitute this expression into the orthogonality equations and solve for the expansion coefficients \( \alpha \). Once we find the coefficients \( \alpha \) which satisfy the orthogonality conditions, (2.22) is used to find \( y_p \).

To be specific, we substitute (2.22) and (2.20) into (2.21) to obtain the following equations, which are called called the normal equations

\[
(y - X\alpha^*)^T x_i = 0, \quad i = 1, \cdots, r,
\]  

(2.23) or

\[
(y - X\alpha^*)^T X = 0
\]

or

\[
y^T X = \alpha^{*T}X^T X.
\]

We put a * on \( \alpha \) to show that this vector of coefficients corresponds to the optimal approximating vector \( y_p \). We can take the transpose of the above equation and solve for the coefficient vector \( \alpha^* \) which satisfies the normal equations as follows

\[
\alpha^* = (X^T X)^{-1}X^T y.
\]  

(2.24)

Note that the \( r \times r \) matrix \( X^T X \) is the Gram matrix for the vectors \( x_1, \cdots, x_r \). Since these vectors are linearly independent, this Gram matrix has a non-zero determinant and hence is invertible. Recall that the vector \( y_p \) that we are looking for is expressed as \( y_p = X\alpha \), and we now know that \( \alpha = \alpha^* \) satisfies the normal equations (orthogonality conditions). Thus using (2.24) we have

\[
y_p = X\alpha^* \\
= X(X^T X)^{-1}X^T y \\
= P_R y,
\]  

(2.25)
where the matrix $P_R$ is called the orthogonal projection matrix onto the subspace $R$. To project any vector $y$ onto the subspace $R$, simply multiply $y$ by $P_R$. From (2.25) we see that the projection matrix is built from the matrix $X$, whose columns are a basis for the subspace $R$, as follows

$$P_R = X(X^T X)^{-1} X^T.$$  

(2.26)

We note in passing that all orthogonal projection matrices share two properties, each of which is easy to verify using (2.26)

1. A projection matrix is symmetric: $P_R = P_R^T$.

2. A projection matrix is idempotent: $P_R P_R = P_R$.

**EXAMPLE 2.10**

In Example 2.9 we showed that the vector $y$ is not in the subspace $R$. In this example we compute the projection matrix $P_R$ and the vector $y_p$, the projection of $y$ onto $R$.

Recall that the basis vectors for $R$, arranged in the matrix $X$ are

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$  

Using (2.26) we compute $P_R$ as

$$P_R = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}.$$  

We can now project $y$ onto $R$ as follows

$$y_p = P_R y = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 4/3 \\ 2/3 \end{bmatrix}.$$  

The error between $y$ and its projection onto $R$ is

$$y_e = y - y_p = \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix}.$$  

We can show that the error vector is in fact orthogonal to the subspace $R$ by showing that it is orthogonal to the basis vectors $x_1$ and $x_2$.

$$x_1^T y_e = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix} = 0.$$
and
\[
x_2^T y = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix} = 0.
\]

2.5.3 Orthogonal Complements and the Four Fundamental Subspaces

Suppose we are given \( r \) linearly independent vectors in \( \mathbb{R}^n \), \( x_1, \ldots, x_r \), where \( r < n \). Then from Fact 2.10, we have that
\[
R = \{ \text{all linear combinations of } x_1 \cdots x_r \}
\]
is a subspace of \( \mathbb{R}^n \). Given any subspace of \( \mathbb{R}^n \), we can always define its **orthogonal complement**. For example, the orthogonal complement of the subspace \( R \), which we denote by \( R^\perp \), is defined as
\[
R^\perp = \{ y \in \mathbb{R}^n \text{ such that } y^T x = 0 \text{ for all } x \in R \}.
\]
That is, elements of \( R^\perp \) are orthogonal to every vector in \( R \). This definition of \( R^\perp \) is somewhat cumbersome, because it requires that we check for orthogonality with every vector in \( R \). Thus we now develop an equivalent definition of \( R^\perp \).

Recall that if a basis is defined for some vector space, then every vector in that space can be (uniquely) expressed as a linear combination of basis vectors. For the space \( R \) defined above, the linearly independent vectors \( x_1 \cdots x_r \) constitute a basis. If a vector \( y \) were orthogonal to each of the basis vectors in \( R \), then \( y \) would be orthogonal to every vector in \( R \) since every vector in \( R \) is expressible in terms of the basis. We can write an equation to express this fact as follows. Assume \( y^T x = 0 \), \( i = 1, \ldots, r \), and let \( x \) be an arbitrary vector in \( R \) which is expressed in terms of the basis vectors as \( x = \sum_{i=1}^{r} \alpha_i x_i \). Then
\[
y^T x = y^T \sum_{i=1}^{r} \alpha_i x_i \\
= \sum_{i=1}^{r} \alpha_i y^T x_i \\
= 0.
\]
Thus an equivalent definition for the orthogonal complement of a subspace \( R \) with basis \( x_1, \ldots, x_r \) is
\[
R^\perp = \{ y \in \mathbb{R}^n \text{ such that } y^T x = 0, i = 1, \ldots, r \}.
\]
If all the vectors in \( R^\perp \) are orthogonal to basis vectors for \( R \), then it is clear that basis vectors for \( R^\perp \) are orthogonal to basis vectors for \( R \). In other words, if \( \{x_1, \ldots, x_r\} \) is a basis for \( R \) and \( \{y_1, \ldots, y_s\} \) is a basis for \( R^\perp \) then
\[
x_i^T y_j = 0, \quad i = 1, \ldots, r \text{ and } j = 1, \ldots, s.
\]
Using (2.27) it can be shown that the vectors \( x_1, \ldots, x_r, y_1, \ldots, y_2 \) are linearly independent (compare Fact 2.25). Since these vectors belong to \( \mathbb{R}^n \), an \( n \)-dimensional vector
space, we must have $r + s \leq n$ (Fact 2.7). With a little more work it is possible to show that $r + s = n$ [84].

A useful property of $\mathbb{R}^n$ is that it can be decomposed into a **direct sum of orthogonal subspaces**. In particular, $\mathbb{R}^n$ can be written as a direct sum of the subspaces $R$ and $R^\perp$. This direct sum of subspaces is written as follows

$$
\mathbb{R}^n = R \oplus R^\perp
$$

and is explained by the following definition: any vector $v \in \mathbb{R}^n$ can be decomposed **uniquely** as the sum of two orthogonal vectors $v_R$ and $v_{R^\perp}$ where $v_R \in R$ and $v_{R^\perp} \in R^\perp$.

Notice that $v_R$ is the projection of $v$ onto the subspace $R$, and $v_{R^\perp}$ is the projection of $v$ onto the subspace $R^\perp$.

**Example 2.11**

Recall from Example 2.10 that

$$
y = y_p + y_e
$$

where $y_p \in R$ and $y_e$ is orthogonal to $R$ so that $y_e \in R^\perp$. We have seen that $y_p$ is the projection of $y$ onto $R$ and we now want to show that $y_e$ is the projection of $y$ onto $R^\perp$. In order to do this, we need a basis for $R^\perp$. In this example, the vector space is $\mathbb{R}^3$ ($n = 3$) and the subspace $R$ has dimension 2 ($r = 2$). Thus the dimension of the orthogonal complement $R^\perp$ must be $n - r = s = 1$. A basis for $R^\perp$ consists of a single vector which is orthogonal to the basis vectors for $R$. Such a vector is\(^2\)

$$
y_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.
$$

We use $y_1$ to form the projection matrix onto $R^\perp$ as follows.

$$
P_{R^\perp} = y_1(y_1^T y_1)^{-1} y_1 = \begin{bmatrix} 1/3 & -1/3 & 1/3 \\ -1/3 & 1/3 & -1/3 \\ 1/3 & -1/3 & 1/3 \end{bmatrix}.
$$

Then

$$
P_{R^\perp} y = \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix}
$$

which is the vector $y_e$ calculated in Example 2.10.

Now that orthogonal complements have been defined, we proceed to define the four fundamental subspaces associated with a matrix.

Given a matrix $A$, there are four subspaces that can be defined in terms of the rows and columns of $A$. Two of these subspaces have already been defined in Section 2.4, namely the row space and column space of a matrix (Definitions 2.8 and 2.10, respectively). The orthogonal complements of the row and column spaces are the other two fundamental\(^2\)

---

\(^2\)A tool for computing basis vectors for the orthogonal complement of a subspace is given in Section 2.7.
subspaces. We deal first with the orthogonal complement of the row space, which is called the null-space.

Definition 2.24 The null space of a matrix $A$, denoted $\text{null}(A)$, is the set of all vectors $x$ satisfying $Ax = 0$. The dimension of the null-space of $A$ is called the nullity of $A$ and is denoted $\nu(A)$.

Notice that any vector $x$ which satisfies $Ax = 0$ is orthogonal to all the rows of $A$. Thus

$$x \in \text{null}(A) \implies x \in \text{row}(A)^\perp.$$  

It can be shown that the reverse implication is also true, namely

$$x \in \text{row}(A)^\perp \implies x \in \text{null}(A).$$

Thus we have

Fact 2.26 $\text{null}(A) = \text{row}(A)^\perp$.

The fourth fundamental subspace is the orthogonal complement of the column space of $A$. There is no special notation for this subspace.

2.5.4 The Rank of a Matrix

An important property of a matrix which is used to characterize solutions to linear equations is its rank. The rank of a matrix is also used in Chapters 6 and 7 to test for system theoretic properties such as controllability and observability.

The row rank of a matrix $A$ is defined to be the maximum number of linearly independent rows in $A$. The column rank of a matrix $A$ is defined to be the maximum number of linearly independent columns in $A$.

Fact N.2 Given any $m \times n$ matrix $A$, the row rank of $A$ is always equal to the column rank of $A$.

PROOF: See [84].

The surprising thing about the previous fact is that it is true even when $m$ is not equal to $n$. Thus we can define the rank of a matrix as follows

Definition 2.25 The rank of an $m \times n$ matrix $A$, denoted $\rho(A)$, is equal to the number of linearly independent rows of $A$, or equivalently, the number of linearly independent columns of $A$. 

It is clear that an \( m \times n \) matrix \( A \) cannot have more than \( m \) linearly independent rows – it only has \( m \) rows! Similarly, \( A \) cannot have more than \( n \) linearly independent columns. Thus we obtain the simple but useful fact

**Fact 2.27** The rank of an \( m \times n \) matrix \( A \) is less than or equal to \( \min(m, n) \).

There is a useful fact relating the rank of a matrix and the dimension of its null space. Recall that the rank of a matrix equals the dimension of its row space, while the nullity is the dimension of the orthogonal complement of the row space. In the previous subsection we mentioned that the dimensions of a subspace and its orthogonal complement add up to the dimension of the vector space. Since the rows of an \( m \times n \) matrix are elements of \( \mathbb{R}^n \), we have the following fact.

**Fact 2.28** Given an \( m \times n \) matrix \( A \), \( \text{rank}(A) + \text{nullity}(A) = n \).

A proof of this fact can be found in [84].

### 2.6 Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are defined only for square matrices. Throughout this section we assume that \( A \) is an \( n \times n \) matrix.

**Definition 2.26** Given a matrix \( A \), a nonzero vector \( x \) is said to be an eigenvector of \( A \) corresponding to eigenvalue \( \lambda \) if

\[
Ax = \lambda x.
\]  

(2.28)

It is easy to see from this definition that eigenvectors are not unique. Indeed if \( x \) is an eigenvector of \( A \) corresponding to eigenvalue \( \lambda \), then so is \( \alpha x \) for any \( \alpha \neq 0 \). We can derive three useful facts about eigenvalues directly from the definition. The first fact shows how eigenvalues are affected when a matrix \( A \) is multiplied by a scalar \( \alpha \). The second fact shows how eigenvalues are affected when a scalar multiple of the identity matrix is added to \( A \). The third fact says that the eigenvalues are unchanged when a matrix is transformed in a certain way.

**Fact 2.29** If \( \lambda \) is an eigenvalue of \( A \) then \( \alpha \lambda \) is an eigenvalue of \( \alpha A \).

**PROOF**. From Definition 2.26 we know that there is an eigenvector \( x \) such that

\[
Ax = \lambda x.
\]

If we let \( M = \alpha A \) then

\[
Mx = (\alpha A)x
\]

\[
= \alpha (Ax)
\]

\[
= \alpha \lambda x
\]
which shows that $\alpha \lambda$ is an eigenvalue of $M = \alpha A$.

The proof of the next fact is left as an exercise.

**Fact 2.30** If $\lambda$ is an eigenvalue of $A$, then $\alpha + \lambda$ is an eigenvalue of $\alpha I + A$.

Before presenting the next fact we must first have the following definition.

**Definition 2.27** Let $A$ be an $n \times n$ matrix, let $T$ be a nonsingular $n \times n$ matrix, and let $\bar{A} \overset{\text{def}}{=} TAT^{-1}$. Then $A$ and $\bar{A}$ are said to be **similar** matrices. They are related by the **similarity transformation** matrix $T$.

**Fact 2.31** Similar matrices have the same eigenvalues.

**Proof.** We first show that if $\lambda$ is an eigenvalue of $A$, then it must also be an eigenvalue of $\bar{A}$. If $\lambda$ is an eigenvalue of $A$, then we have

\[
Ax = \lambda x
\]

\[
TAX = \lambda Tx
\]

\[
TAT^{-1}T^{-1}x = \lambda Tx
\]

\[
\bar{A}(Tx) = \lambda(Tx).
\]

The last line shows that $\lambda$ is an eigenvalue of $\bar{A}$, and we also see that the corresponding eigenvector is $Tx$.

We now show that if $\lambda$ is an eigenvalue of $\bar{A}$, then it must also be an eigenvalue of $A$.

We have

\[
\bar{A}y = \lambda y
\]

\[
TAT^{-1}y = \lambda y
\]

\[
A(T^{-1}y) = \lambda(T^{-1}y).
\]

The last line shows that $\lambda$ is an eigenvalue of $A$, corresponding to eigenvector $T^{-1}y$.

We now derive three more facts about eigenvalues which involve the use of the determinant of a special matrix.

**Fact 2.32** The scalar $\lambda$ is an eigenvalue of $A$ if and only if

\[
\det(\lambda I - A) = 0.
\]

**Proof.** If $\lambda$ is an eigenvalue of $A$, then there is an eigenvector $x$ such that

\[
Ax = \lambda x.
\]
We can rewrite $\lambda x$ as $\lambda Ix$ and rearrange the above equation to yield

$$(\lambda I - A)x = 0.$$  

Notice that the eigenvector $x$ is in the null space of the matrix $\lambda I - A$. Since $(\lambda I - A)$ has at least one nonzero vector in its null space, the nullity of $(\lambda I - A)$ is at least one, and so the rank of $(\lambda I - A)$ must be less than or equal to $n - 1$ (see Fact 2.28). Since the rank of $(\lambda I - A)$ is less than $n$, this means that its rows (and columns) are linearly dependent. From Fact 2.23 this means that $\det(\lambda I - A) = 0$.

To show the reverse implication, assume that $\det(\lambda I - A) = 0$. Then by Fact 2.23 the columns of $(\lambda I - A)$ are linearly dependent. That means that there exists a nonzero vector $x$ such that $(\lambda I - A)x = 0$ or $Ax = \lambda x$. Thus $\lambda$ is an eigenvalue of $A$.

The following fact is given without proof.

**Fact 2.33** Let $A$ be an $n \times n$ matrix. Then $\det(\lambda I - A)$ is a monic polynomial in $\lambda$ of degree $n$. It is denoted $a(\lambda)$ and called the **characteristic polynomial** of $A$. We have

$$a(\lambda) \overset{\text{def}}{=} \det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n.$$  

In Table 3.5 on page 114 we give an algorithm to compute the characteristic polynomial of a matrix. Putting the two previous facts together we get

**Fact 2.34** An $n \times n$ matrix $A$ has $n$ eigenvalues which are the roots of the characteristic polynomial $a(\lambda)$.

The above facts suggest a way to calculate the eigenvalues and eigenvectors of a matrix. The first step is to form the characteristic polynomial and find its roots $\lambda_1, \cdots, \lambda_n$. These are the eigenvalues of $A$. For each eigenvalue, form the matrix $(\lambda_i I - A)$ and find a nonzero vector in its null space. This vector will be an eigenvector of $A$ (see the proof of Fact 2.32). If the eigenvalues of $A$ are distinct it turns out that the nullity of $(\lambda I - A)$ is equal to one, and each eigenvector is picked as any nonzero vector from a one-dimensional null space. The situation is more complicated for repeated eigenvalues. Algorithms for computing eigenvectors can be found in [22, 37]. The case of repeated eigenvalues can be found in [10, 42].

**EXAMPLE 2.12**

Consider the following matrix

$$A = \begin{bmatrix} 1.3 & -0.4 \\ -0.2 & 1.1 \end{bmatrix}.$$
We can compute the characteristic polynomial as follows
\[ a(\lambda) = |\lambda I - A| \]
\[ = \begin{vmatrix} \lambda - 1.3 & 0.4 \\ 0.2 & \lambda - 1.1 \end{vmatrix} \]
\[ = (\lambda - 1.3)(\lambda - 1.1) - 0.8 \]
\[ = \lambda^2 - 2.4\lambda + 1.35. \]

The eigenvalues of \( A \) are the roots of the \( a(\lambda) \) can be calculated using the quadratic formula to be \( \lambda_1 = 1.5 \) and \( \lambda_2 = 0.9 \).

To compute the eigenvector corresponding to \( \lambda_1 \), we look for a vector in the null space of \( \lambda_1 I - A \).
\[ 1.5 I - A = \begin{bmatrix} 0.2 & 0.4 \\ 0.2 & 0.4 \end{bmatrix}. \]

A null-space vector for this matrix is found (by inspection) to be
\[ x_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}. \]

The second eigenvector is computed by finding a vector in the null space of \( \lambda_2 I - A \).
\[ 0.9 I - A = \begin{bmatrix} -0.4 & 0.4 \\ 0.2 & -0.2 \end{bmatrix}. \]

A null-space vector for this matrix is found (by inspection) to be
\[ x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \]

In the above example, the null-space vectors were found by inspection. There is a systematic procedure which can be carried out by hand to compute null-space vectors [10, 84]. However we do not stress hand computations in this book. We assume that the reader has access to a computer program such as MATLAB to perform the necessary calculations.

We now give another fact about eigenvalues without proof. The proof of the following fact may be found in [10, 42]

**Fact 2.35** Eigenvectors corresponding to distinct eigenvalues are linearly independent.

This fact can be used to derive a useful representation of a matrix. Suppose that the eigenvalues of a matrix \( A \) are distinct. Then we have
\[ A x_i = \lambda_i x_i, \quad i = 1, \ldots, n. \]

The above equations may be written as a single matrix equation as follows
\[ A \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \] (2.29)
Now let
\[ X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \]

From Fact 2.35 the \( n \) columns of \( X \) are linearly independent, and so by Fact 2.23 on page 48 \( X \) is nonsingular. Using the definitions of \( X \) and \( D \) we can rewrite (2.29) as

\[ AX = XD \]
or
\[ A = XDX^{-1}. \quad (2.30) \]

The above equation says that the matrix \( A \) is similar to a diagonal matrix consisting of its eigenvalues. The columns of the similarity transformation matrix are the eigenvectors of \( A \).

Equation (2.30) is useful for computing powers of a matrix. A matrix \( A \) raised to an integer power \( k \) is just the matrix \( A \) multiplied by itself \( k \) times as shown below

\[ A^k \overset{\text{def}}{=} A \cdot A \cdot \cdots \cdot A. \]

Substituting (2.30) into the above equation yields

\[ A^k = (XDX^{-1})(XDX^{-1}) \cdots (XDX^{-1}) = XD^kX^{-1}. \quad (2.31) \]

Notice that all the intermediate \( XX^{-1} \) terms cancel out in the expansion of \( A^k \). For example, consider the matrix \( A \) from Example 2.12. Using (2.31) we have

\[ \begin{bmatrix} 1.3 & -0.4 \\ -0.2 & 1.1 \end{bmatrix}^k = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1.5^k & 0 \\ 0 & 0.9^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}^{-1}. \]

The above equation is a useful computational tool to compute \( A^k \) for large values of \( k \), but it also gives useful information about \( A^k \) for large \( k \). For instance, because \( 1.5^k \) increases without bound as \( k \) increases, we see that the entries in the matrix \( A^k \) will increase without bound. However, if both of the eigenvalues had magnitudes less than one, then \( A^k \) would approach the zero matrix as \( k \) became large. Furthermore, the rate of increase or decrease of the elements in \( A^k \) depends on the magnitudes of the eigenvalues.

We conclude this section with four additional facts about eigenvalues.

**Fact 2.36** The determinant of a matrix equals the product of its eigenvalues.

**PROOF.** We give the proof only for the case in which the matrix has distinct eigenvalues. However, the fact is true for any square matrix.

Let \( A \) be a matrix with distinct eigenvalues, let \( X \) be a matrix whose columns are the eigenvectors of \( A \) and let the corresponding eigenvalues be placed in a diagonal matrix.
D. Let the determinant of the matrix $X$ be denoted $\alpha$. Then using the properties of determinants shown in Table 2.5 on page 45, we have

\[
|A| = |XDX^{-1}|
= |X||D||X^{-1}|
= \alpha|D|\frac{1}{\alpha}
= |D|
= \lambda_1 \lambda_2 \cdots \lambda_n.
\]

Since a matrix is nonsingular if and only if its determinant is nonzero, the above fact gives

**Fact 2.37** A matrix is nonsingular if and only if all its eigenvalues are nonzero.

From Fact 2.32 we know that eigenvalues are the roots of $\det(\lambda I - A)$ and from item 2 in Table 2.5 on page 45, we know that a matrix and its transpose have the same determinant. Thus the following fact is true.

**Fact 2.38** A matrix and its transpose have the same eigenvalues.

The derivation of the following result is left as an exercise.

**Fact 2.39** The eigenvalues of a triangular matrix are equal to the diagonal elements of the matrix.

An extension of the above fact to block-triangular matrices is

**Fact 2.40** The eigenvalues of a block-triangular matrix $A$ are the union of the eigenvalues of the matrices on the main diagonal of $A$.

### 2.7 The Singular Value Decomposition

Any $m \times n$ matrix $A$ of rank $r$ has a singular value decomposition (SVD) which is a product of three matrices as follows

\[
A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V^T_{n \times n}, \quad (2.32)
\]
where $U^T U = I_m$, and $V^T V = I_n$. The matrix $\Sigma$ takes the following form:

$$\Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \cdots & 0 \\ 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}. $$

The numbers $\sigma_i$ are called the *singular values* of $A$ and they are non-negative real numbers ordered in decreasing order, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Note that the matrix $A$ may also possess a number of zero singular values. It can be shown that the squares of the nonzero singular values of $A$ are the nonzero eigenvalues of $AA^T$ or $A^T A$ (both of these matrices have the same nonzero eigenvalues). It can also be shown that the columns of $U$ are eigenvectors of $AA^T$, and columns of $V$ are eigenvectors of $A^T A$.

We now partition the $U$ and $V$ matrices given by the SVD in (2.32) according to the rank $r$ of $A$:

$$U = [u_1 \cdots u_r | u_{r+1} \cdots u_m] = [U_1 | U_2], \quad (2.33)$$

and

$$V = [v_1 \cdots v_r | v_{r+1} \cdots v_n] = [V_1 | V_2]. \quad (2.34)$$

The vectors $u_i$ are called the *left singular vectors* of $A$, and it can be shown that they are eigenvectors of $AA^T$. The vectors $v_i$ are called the *right singular vectors* of $A$, and it can be shown that they are eigenvectors of $A^T A$.

Note that zero rows in $\Sigma$ multiply columns of $U$ and zero columns of $\Sigma$ multiply rows of $V^T$. These zero rows and columns of $\Sigma$ can be deleted, along with the corresponding columns and rows of $U$ and $V^T$ to produce the “economy size” SVD of $A$:

$$A = U_1 \Sigma_1 V_1^T,$$

where

$$U_1^T U_1 = I_r, \quad V_1^T V_1 = I_r,$$

and $U_1$ and $V_1$ contain singular vectors corresponding to nonzero singular values as shown in (2.33). Recall that $r$ in (2.33) is the number of nonzero singular values. Thus the rank of $A$ is the number of nonzero singular values.

**SVD and the Four Fundamental Subspaces**

The SVD shows the rank of a matrix as the number $r$ of non-zero singular values. When the matrices $U$ and $V$ are partitioned as shown in (2.33) and (2.34), the partitions provide orthonormal bases for the four fundamental subspaces of a matrix:

1. $U_1$ is an orthonormal basis for the column space of $A$.
2. $U_2$ is an orthonormal basis for the orthogonal complement of the column space of $A$.
3. $V_1$ is an orthonormal basis for the row space of $A$.
4. $V_2$ is an orthonormal basis for the null-space of $A$ (recall that the null space of a matrix is the same as the orthogonal complement of the row space).
EXAMPLE 2.13

In Examples 2.10 and 2.11 we considered a matrix \( \mathbf{X} \) which we show below as the matrix \( \mathbf{A} \).

\[
\mathbf{A} = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
0 & 1 
\end{bmatrix}.
\]

MATLAB gives the following SVD of \( \mathbf{A} \) (to 4 decimal places)

\[
\mathbf{U} = \begin{bmatrix}
0.4082 & -0.7071 & 0.5774 \\
0.8165 & 0.0000 & -0.5774 \\
0.4082 & 0.7071 & 0.5774
\end{bmatrix},
\mathbf{\Sigma} = \begin{bmatrix}
1.7321 & 0 \\
0 & 1.0000 \\
0 & 0
\end{bmatrix},
\]

and

\[
\mathbf{V} = \begin{bmatrix}
0.7071 & -0.7071 \\
0.7071 & 0.7071
\end{bmatrix}.
\]

Since there are two non-zero singular values, the rank of this matrix is 2. The first two columns of \( \mathbf{U} \) are a basis for the column space of \( \mathbf{A} \) and the third column on \( \mathbf{U} \) is a basis for the orthogonal complement of the column space of \( \mathbf{A} \). We can compute projection matrices for these subspaces using (2.26). Because the bases from the SVD are orthonormal, the term that is inverted in (2.26) equals an identity matrix, and can be omitted. The projection matrix for the column space of \( \mathbf{A} \) is

\[
P_R = \mathbf{U}_1 \mathbf{U}_1^T
\]

\[
= \begin{bmatrix}
0.4082 & -0.7071 \\
0.8165 & 0.0000 \\
0.4082 & 0.7071
\end{bmatrix} \begin{bmatrix}
0.4082 & 0.8165 & 0.4082 \\
-0.7071 & 0.0000 & 0.7071
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0.6667 & 0.3333 & -0.3333 \\
0.3333 & 0.6667 & 0.3333 \\
-0.3333 & 0.3333 & 0.6667
\end{bmatrix}.
\]

This is the same projection matrix that was obtained in Example 2.10.

In a similar way we can use the SVD to calculate the projection matrix onto the orthogonal complement of the column space of \( \mathbf{A} \). The result is

\[
P_{R^\perp} = \mathbf{U}_2 \mathbf{U}_2^T
\]

\[
= \begin{bmatrix}
0.5774 \\
-0.5774 \\
0.5774
\end{bmatrix} \begin{bmatrix}
0.5774 & -0.5774 & 0.5774
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0.3333 & -0.3333 & 0.3333 \\
-0.3333 & 0.3333 & -0.3333 \\
0.3333 & -0.3333 & 0.3333
\end{bmatrix}.
\]

which is the same projection matrix that was obtained in Example 2.11.
SVD and Rank

Another important use of the SVD has to do with determining the rank of a matrix. The SVD shows if a small perturbation of a matrix $A$ results in a matrix whose rank is less than that of $A$. This can occur if the smallest nonzero singular value of $A$ is much smaller that the largest singular value, as illustrated by the following example. Suppose $A$ is a $2 \times 2$ matrix given by

$$A = \begin{bmatrix} 1 & 0 \\ 1/\epsilon & 1 \end{bmatrix}$$

where $\epsilon$ is a small positive number. We note that the determinant of $A$ is 1, so $A$ is nonsingular (rank = 2 in this case). Now we add a very small matrix to $A$ to get

$$\tilde{A} = A + \begin{bmatrix} 0 & \epsilon \\ 0 & 0 \end{bmatrix}.$$ 

Note that the determinant of $\tilde{A} = 0$ so $\tilde{A}$ is singular (rank = 1), even though it is just a small perturbation away from a matrix of rank 2. The singular values of $A$ show that $A$ is close to a matrix of rank 1. The singular values of $A$ can be computed to give

$$\sigma_1^2 = 1 + \frac{1 + \sqrt{1 + 4\epsilon^2}}{2\epsilon^2}, \quad \sigma_2^2 = 1 + \frac{1 - \sqrt{1 + 4\epsilon^2}}{2\epsilon^2}. $$

As $\epsilon$ becomes small, the first singular value becomes large, while the second singular value goes to zero. For instance, when $\epsilon$ equals 0.001, $\sigma_1 = 1,000$ and $\sigma_2 = 10^{-6}$. Thus the matrix $A$ has only one singular value significantly different from zero, and so $A$ is very close to a matrix of rank 1.

2.8 Linear Equations

Consider a set of $m$ linear equations in $n$ unknowns written in matrix form as

$$A \underbrace{x}_{m \times n} = \underbrace{y}_{n \times 1}. \quad (2.35)$$

In this equation, $A$ and $y$ are known and $x$ is unknown. It is useful to define a matrix $W$ which consists of the matrix $A$ augmented with the vector $y$; that is, $W$ is the $m \times (n + 1)$ matrix given by

$$W \overset{\text{def}}{=} [A \ y]. \quad (2.36)$$

In the next two sections, we characterize solutions to (2.35).

2.8.1 Characterization by Rank

If $y \in \text{col}(A)$ then there exists a vector of expansion coefficients $x$ such that $y = Ax$. In other words, $y \in \text{col}(A)$ is a sufficient condition for (2.35) to have a solution. It is also a necessary condition because if $y \notin \text{col}(A)$, this means that $y$ cannot be expressed as a linear combination of the columns of $A$. In this case there does not exist a vector $x$ such that $y = Ax$. 

A test for whether or not \( y \in \text{col}(A) \) can be expressed in terms of the ranks of the matrices \( A \) and \( W \) (\( W \) is defined in (2.36)). The rank of \( A \), denoted \( \rho(A) \), is the number of linearly independent columns of \( A \). The matrix \( W \) contains the columns of \( A \) plus one additional column, the vector \( y \). There are only two ways that \( \rho(W) \) and \( \rho(A) \) can be related. Either \( \rho(W) = \rho(A) + 1 \), which means that \( y \) is linearly independent of the columns of \( A \), or \( \rho(W) = \rho(A) \), which means that \( y \) is linearly dependent on the columns of \( A \).

The discussion in the previous two paragraphs can be summarized as follows

\[
\rho(W) = \rho(A) \implies y \in \text{col}(A) \implies (2.35) \text{ has a solution.}
\]

\[
\rho(W) \neq \rho(A) \implies y \notin \text{col}(A) \implies (2.35) \text{ does not have a solution.}
\]

These results characterize the existence of solutions for linear equations. A complete characterization of the existence and uniqueness of solutions is given in the following fact.

**Fact 2.41**

1. If \( \rho(W) \neq \rho(A) \) then no solutions exist.
2. If \( \rho(W) = \rho(A) \) then at least one solution exists.
   
   (a) If \( \rho(W) = \rho(A) = n \), then there is a unique solution for \( x \).
   
   (b) If \( \rho(W) = \rho(A) = r < n \), then there is an infinite set of solution vectors parameterized by \( n - r \) free variables.
3. If \( \rho(A) = m \), a solution exists for every possible right hand side vector \( y \). This case can only occur if \( m \leq n \). From 2., the solution is unique if \( m = n \) and non-unique if \( m < n \).

While Fact 2.41 is theoretically complete, it falls short in two areas: first, it doesn’t tell you how to find solutions, or how to parameterize an infinite set of solutions, and second, it does not give any information about obtaining approximate solutions in cases when an exact solution does not exist. We address these two areas in the following sections. We begin by characterizing solutions in terms of the dimensions of the coefficient matrix.

### 2.8.2 Characterization by Matrix Dimension

We consider three cases: when \( m = n \) and the matrix \( A \) is square, when \( m > n \) so that the matrix \( A \) has more rows than columns, and finally the case when \( m < n \) so that the matrix \( A \) has more columns than rows. We first characterize the types of solutions that can occur with these three cases using the two conditions in Fact 2.41. Then we show how solutions can be calculated.

1. If \( m = n \) (square system of equations), then:
   
   (a) If \( \det(A) \neq 0 \) then a unique solution exists and is given by
   
   \[
   x = A^{-1}y.
   \]

   The reasoning for this case goes as follows: \( \det(A)=n \) implies by the Gram test that the columns of \( A \) are linearly independent, and so \( \rho(A) = n \). Then \( \rho(W) \)
can’t be larger than \( n \) because \( W \) has only \( n \) rows, and \( \rho(W) \) can’t be smaller than \( n \) because the \( n \) columns of \( A \) (which are known to be independent) are also columns of \( W \). Thus this case fits into case 2(a) of the theoretical characterization given in the previous subsection.

(b) If \( \det(A)=0 \) then solutions do not exist for an arbitrary vector \( y \). However, if \( y \) is such that \( \rho(W) = \rho(A) = r < n \), then an infinite number of solutions exist. (Note that \( r \) will be less than \( n \) in this case because \( \det(A) = 0 \).)

2. If \( m > n \) (overdetermined system of equations)

(a) The most common case of overdetermined equations occurs when \( \rho(A) = n \). In this case the unique least-squares solution (the one that minimizes the norm of \( y - Ax \) over all possible \( x \)) is given by projecting \( y \) onto the column space of \( A \), and the solution will shown in the next section to be given by

\[
x = (A^T A)^{-1} A^T y.
\]

(b) If \( \rho(A) < n \), then solutions exist if and only if \( \rho(A) = \rho(W) = r \) (where \( r \) is some number less than \( n \)). In this case, there are an infinite number of solutions parameterized by \( n - r \) variables. This case is examined later using the SVD.

3. If \( m < n \) (underdetermined system of equations)

(a) The most common case of underdetermined equations occurs when \( \rho(A) = m \). In this case an infinite number of solutions exists, but we can obtain the unique solution which has the smallest norm of any solution vector by projecting any solution onto the row space of \( A \). The solution will be shown to be given by

\[
x_{\min-norm} = A^T (AA^T)^{-1} y.
\]

(b) If \( \rho(A) < m \), then solutions may or may not exist, depending on whether or not \( \rho(W) \) equals \( \rho(A) \). This case is examined using the SVD later in this section.

2.8.3 Least Squares and Min-Norm Solutions – Full Rank Case

2.8.3.1 Least-Squares Solutions

Consider the case when \( \rho(W) \neq \rho(A) \) so that no solutions exist to the equation \( Ax = y \). In this case, we would like to find the vector \( x_{LS} \) which minimizes the squared norm between \( y \) and \( Ax \) for any vector \( x \). In other words, we would like to solve the following minimization problem

\[
\min_x \| y - Ax \|^2.
\]  (2.37)

Since the vector \( Ax \) is an element of \( \text{col}(A) \) then an equivalent problem is to find the vector in \( \text{col}(A) \) which is the best approximation to a given vector \( y \). We know the answer to this second problem: the best approximation is obtained by projecting \( y \) onto the subspace \( \text{col}(A) \). See Fig. 2.3 on page 52. The result of the projection is the vector \( y_p \).

Once \( y_p \) is known then any vector \( x_{LS} \) which satisfies

\[
Ax_{LS} = y_p
\]  (2.38)

will minimize the squared error in (2.37) We know that the above equation has a solution because \( y_p \in \text{col}(A) \) which implies that \( \rho(W) = \rho(A) \). If we make the further
assumption that \( \rho(A) = n \) (this is the most common case), then by 2(a) of section 4.1, the least-squares solution will be unique, and the matrix which projects onto \( \text{col}(A) \) can be constructed using (2.26). In other words,

\[
y_p = A(A^T A)^{-1} A^T y
\]

and we now solve for the least-squares solution from

\[
Ax_{LS} = y_p = A(A^T A)^{-1} A^T y.
\]

The above equation can be solved by multiplying both sides on the left by \( A^T \) and then by \( (A^T A)^{-1} \) to obtain

\[
x_{LS} = (A^T A)^{-1} A^T y. \tag{2.39}
\]

This is the unique least-squares solution when the columns of \( A \) are linearly independent (so that the rank of \( A \) is \( n \)). In order for \( A \) to have rank \( n \), it is necessary that \( m \geq n \). If \( m = n \) then the right-hand side of (2.39) reduces to \( A^{-1} y \) and the least-squares solution gives an error norm of zero in (2.37).

When the columns of \( A \) are dependent (an uncommon case), the least squares solution is not unique, and solutions can be calculated using the SVD as shown later in this section.

**EXAMPLE 2.14**

Consider the linear equations \( Ax = y \) where

\[
A = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
0 & 1
\end{bmatrix}, \quad y = \begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}.
\]

Using (2.39) we compute the least-squares solution (to 4 decimal places) to be

\[
x_{LS} = \begin{bmatrix}
0.6667 \\
0.6667
\end{bmatrix}.
\]

Equation (2.38) says that \( Ax_{LS} \) should equal \( y_p \), the projection of \( y \) onto the column space of \( A \). We have previously computed this projection in Example 2.10, and the result is

\[
y_p = \begin{bmatrix}
2/3 \\
4/3 \\
2/3
\end{bmatrix}.
\]

We get the same result by multiplying \( Ax_{LS} \) as shown below (to 4 decimal places)

\[
A x_{LS} = \begin{bmatrix}
1 & 1 & 0 \\
0 & 1 & 1
\end{bmatrix} \begin{bmatrix}
0.6667 \\
0.6667
\end{bmatrix} = \begin{bmatrix}
0.6667 \\
1.3333 \\
0.6667
\end{bmatrix}.
\]

**2.8.3.2 Min-Norm Solutions** We now turn our attention to underdetermined equations \( (m < n) \) and min-norm solutions in the case when the matrix \( A \) has rank \( m \) (that is, its rows are linearly independent). From 3 of Fact 2.41 we know that a solution exists for every possible right hand side vector \( y \). In addition, we know that there will be an infinite
number of solutions, since $\rho(A) = m < n$. We would like to find the solution which has the minimum norm. A method for parameterizing all solutions using the SVD will be given later in this section.

In order to find the min-norm solution, it is useful to consider the row-space of $A$, denoted $\text{row}(A)$, and its orthogonal complement, denoted $\text{row}(A)^\perp$. Recall that a subspace and its orthogonal complement form a direct sum decomposition of $\mathbb{R}^n$, so that

$$\mathbb{R}^n = \text{row}(A) \oplus \text{row}(A)^\perp.$$ 

Now a solution to $Ax = y$ will be a vector $x \in \mathbb{R}^n$, and so $x$ can be decomposed uniquely into its projection onto $\text{row}(A)$ and its projection onto $\text{row}(A)^\perp$ as follows

$$x = x_A + x_{A^\perp}.$$ \hspace{1cm} (2.40)

Thus for any solution $x$ we can write

$$Ax = y$$

$$A(x_A + x_{A^\perp}) = y$$

$$Ax_A = y.$$ 

We used the fact that $Ax_{A^\perp} = 0$ since vectors in $\text{row}(A)^\perp$ are orthogonal to rows of $A$. We now use the fact that $x_A$ is in $\text{row}(A)$, and so it has a representation as a linear combination of the rows of $A$. Since the rows of $A$ are the columns of $A^T$, we can write this representation as

$$x_A = A^T \alpha.$$ \hspace{1cm} (2.41)

If we multiply the above equation by $A$ and use the fact that $Ax_A = y$ we get

$$Ax_A = AA^T \alpha = y.$$ 

Now we notice that $AA^T$ is the Gram matrix for the *rows* of $A$, which are assumed to be linearly independent, so we can multiply through by $(AA^T)^{-1}$ to obtain the unique solution for $\alpha$

$$\alpha = (AA^T)^{-1}y.$$ 

and using (2.41) we get that

$$x_A = A^T (AA^T)^{-1}y$$

where $x_A$ is the unique vector in $\text{row}(A)$ which satisfies $Ax_A = y$. Notice that $y = Ax$ so that the above equation can be written as

$$x_A = A^T (AA^T)^{-1}Ax$$

where the matrix $A^T (AA^T)^{-1}A$ is *the orthogonal projection matrix onto row$(A)$, the row-space of the matrix $A$* (compare with the formulas for projection onto the space formed from column vectors in Section 2.5.2). We can interpret the above equation as saying that $x_A$ is obtained by *projecting* any solution $x$ onto $\text{row}(A)$.

Recall from (2.40) that any solution can be expressed as $x = x_A + x_{A^\perp}$. However, since $x_A$ is orthogonal to $x_{A^\perp}$ it is easy to show that

$$\|x\|^2 = \|x_A\|^2 + \|x_{A^\perp}\|^2 \geq \|x_A\|^2$$
and that the solution of minimum norm occurs when we choose $x = x_A$. In other words, 

$$x_{\text{min-norm}} = x_A = A^T(AA^T)^{-1}y. \quad (2.42)$$

Finally, notice that $\text{row}(A)^\perp$ is just the null-space of $A$, which is of dimension $n - m$ when $A$ has rank $m$. Thus we can find $n - m$ vectors which are a basis for $\text{row}(A)^\perp$. Any linear combination of these vectors can be added to $x_{\text{min-norm}}$ to generate other solutions to $Ax = y$.

**EXAMPLE 2.15**

Consider the equation $Ax = y$ where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$ 

The min-norm solution is computed using $(2.42)$ as follows (results shown to 4 decimal places)

$$x_{\text{min-norm}} = A^T(AA^T)^{-1}y = \begin{bmatrix} 0.6667 & -0.3333 \\ 0.3333 & 0.3333 \\ -0.3333 & 0.6667 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.6667 \\ 2.3333 \\ 1.6667 \end{bmatrix}.$$ 

It can be verified that $Ax_{\text{min-norm}}$ does in fact equal $y$, but there are other solutions to the equation. From Examples 2.10 and 2.11 we know that a basis for the null-space of $A$ is 

$$v = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$ 

Thus all solutions to $Ax = y$ can be written as 

$$x = x_{\text{min-norm}} + \alpha v$$

for some value of $\alpha$. Notice that $x_{\text{min-norm}} \perp v$ so that

$$\|x\|^2 = \|x_{\text{min-norm}} + \alpha v\|^2 = \|x_{\text{min-norm}}\|^2 + \alpha^2 \|v\|^2 \geq \|x_{\text{min-norm}}\|^2.$$ 

2.8.4 Using SVD to Solve Linear Equations

The SVD was introduced in Section 2.7. The SVD provides orthonormal bases for the four fundamental subspaces of a matrix, and these subspaces contain the information necessary to solve linear equations in any situation: unique solution, multiple solutions, least-squares solutions, min-norm solutions, etc. Thus the SVD can be used to solve all of the types of linear equations mentioned in the previous section! A description of how this is done is given in Table 2.6.
Consider the equation $Ax = y$ where the matrix $A$ is $m \times n$. Let the SVD of $A$ be partitioned as follows

$$A = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}.$$  

The $r \times r$ matrix $\Sigma_1$ contains the $r$ non-zero singular values of $A$. The matrices $U_1$ and $V_1$ each contain $r$ columns. If $r = m$ then $U_2$ is not present (all the columns of $U$ appear in $U_1$), and if $r = n$ then $V_2$ is not present.

If $y = 0$ the equation of interest reduces to $Ax = 0$ and we are interested in the null space of $A$. Since $V_2$ provides a basis for the null-space of $A$ all solutions can be written as

$$x = V_2 \alpha$$

where $\alpha$ is a vector containing $n - r$ free parameters.

If $y \neq 0$ we compute the following vector

$$z = V_1 \Sigma_1^{-1} U_1^T y.$$  

The vector $z$ has the following properties.

1. If $Ax = y$ has a unique solution, $z$ is the solution.
2. If $Ax = y$ does not have a solution, $z$ gives a least-squares solution. If the least-squares solution is not unique, $z$ is the least-squares solution with minimum norm. Other least-squares solutions can be generated by adding vectors from the null-space of $A$ to $z$. All solutions can be written as

$$x = z + V_2 \alpha$$

where $\alpha$ is a vector containing $n - r$ free parameters.

3. If $Ax = y$ has an infinite number of solutions, $z$ is the min-norm solution. Other solutions can be generated by adding vectors from the null-space of $A$ to $z$ as shown in the equation in item 2.

Table 2.6  Using SVD to solve linear equations.
EXAMPLE 2.16

In this example we use the SVD to compute all least-squares solutions to a rank-deficient least-squares problem. Consider the following matrix:

\[
A = \begin{bmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
1 & -1 & 1 & 1 \\
0 & 0 & -1 & -1
\end{bmatrix}.
\]

MATLAB gives the singular value decomposition of \(A\) as follows

```matlab
>> [u,s,v]=svd(A)
```

\(u =\)

\[
\begin{align*}
u & =  \\
0.2764 & -0.7236 & 0.5499 & -0.3124 \\
0.4472 & 0.4472 & 0.6243 & 0.4586 \\
0.7236 & -0.2764 & -0.5499 & 0.3124 \\
-0.4472 & -0.4472 & 0.0744 & 0.7710
\end{align*}
\]

\(s =\)

\[
\begin{align*} 
s & =  \\
2.6900 & 0 & 0 & 0 \\
0 & 1.6625 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{align*}
\]

\(v =\)

\[
\begin{align*}
v & =  \\
0.3717 & -0.6015 & 0.7071 & 0.0000 \\
-0.3717 & 0.6015 & 0.7071 & -0.0000 \\
0.6015 & 0.3717 & -0.0000 & -0.7071 \\
0.6015 & 0.3717 & -0.0000 & 0.7071
\end{align*}
\]

Since there are 2 non-zero singular values, the rank of \(A\) is 2. Furthermore, since \(A\) has 4 columns, the nullity of \(A\) is also 2 (see Fact 2.28). Suppose we want to parameterize all least-squares solutions to \(Ax = y\) where

\[
y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.
\]
Since the columns of $A$ are not linearly independent, the full-rank solution in (2.39) cannot be used. However using the formula in Table 2.6 we can compute

$$z = V_1 \Sigma_1^{-1} U_T^T y$$

$$= \begin{bmatrix} 0.3717 & -0.6015 \\ -0.3717 & 0.6015 \\ 0.6015 & 0.3717 \end{bmatrix} \begin{bmatrix} 2.6900 & 0 \\ 0 & 1.6625 \end{bmatrix}^{-1} \begin{bmatrix} 0.2764 & 0.4472 & 0.7236 & -0.4472 \\ -0.7236 & 0.4472 & -0.2764 & -0.4472 \end{bmatrix} y$$

$$= \begin{bmatrix} 1.1 \\ -1.1 \\ -0.2 \\ -0.2 \end{bmatrix}.$$  

The vector $z$ is the least-squares solution of minimum norm. All least-squares solutions to $Ax = y$ can be written as

$$x_{LS} = z + V_2 \alpha$$

where $\alpha$ is a vector of two coefficients and

$$V_2 = \begin{bmatrix} 0.7071 & 0.0000 \\ 0.7071 & -0.0000 \\ -0.0000 & -0.7071 \\ -0.0000 & 0.7071 \end{bmatrix}.$$  

Notice that

$$||Az - y||^2 = ||A(z + V_2 \alpha) - y||^2$$

since $AV_2 = 0.$

2.9 Chapter Summary

In this chapter we presented results from linear algebra and matrix theory which are used in the design of state-space control systems. We developed the concept of linear independence of a set of vectors and used it to define the rank of a matrix as the number of linearly independent rows or columns. Simple, but useful, interpretations of matrix multiplication were shown.

We introduced the four fundamental subspaces of a matrix: the row space, the column space, and their orthogonal complements. Important facts about the eigenvalues and eigenvectors of a square matrix were presented. We characterized the solution of simultaneous linear equations. Conditions for existence and uniqueness of solutions were established. We showed how to compute min-norm solutions when solutions are not unique, and least-squares approximations when no solutions exist. We also showed how the singular value decomposition could be used to solve any type of linear equation.

2.10 Problems

1. Indicate whether the following statements are true or false. If a statement is true, give the reason why it is true. (Use any Fact from Chapter 2.) If a statement is false, give an example in $\mathbb{R}^2$ showing that it is false.
(a) If two vectors $x_1$ and $x_2$ are linearly independent then they must be orthogonal.

(b) Let $A$ be a matrix whose columns are $x_1, x_2, x_3$, i.e.
\[
A = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}.
\]
If the rank of $A$ is 2, then $x_1$ and $x_2$ must be linearly independent.

(c) For the matrix $A$ given in part (b), if the rank of $A$ is 2, then any two linearly independent columns are a basis for the column space of $A$.

2. Given $p$ linearly independent vectors $x_1, \ldots, x_p$ and $p$ other vectors $y_1, \ldots, y_p$, form vectors $z_i$ by concatenating the vectors $x_i$ and $y_i$ as follows:
\[
z_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad i = 1, \ldots, p.
\]
Show that the vectors $z_1, \ldots, z_p$ are linearly independent regardless of the values of $y_1, \ldots, y_p$.

3. Suppose you are given a matrix $A_{m \times n}$ with $m > n$ and you are told that the rank of $A$ is $n$.

   (a) Does there exist a matrix $B$ such that $AB = I_{m \times m}$? If yes, give a formula for $B$ in terms of $A$; if not, state why not.

   (b) Does there exist a matrix $B$ such that $BA = I_{n \times n}$? If yes, give a formula for $B$ in terms of $A$; if not, state why not.

4. Given $X$ and $y$ below, show that $y \in \text{col}(X)$. Find the expansion coefficients that express $y$ as a linear combination of columns of $X$.
\[
X = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 1 & 2 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 6 \\ 5 \end{bmatrix}
\]

5. Suppose it is true that $Ax = Ay$. Under what conditions can you conclude that $x = y$?

6. Under what conditions could it be that $x \neq y$ but $Ax = Ay$?

7. Given $x \in \text{row}(A)$ and $y \in \text{null}(A)$, show that $x^T y = 0$.

8. Let $x_1, \ldots, x_m$ be a basis for an $m$-dimensional subspace $X$ of $\mathbb{R}^n$. Let $y_1, \ldots, y_{n-m}$ be a basis for the orthogonal complement of $X$, $X^\perp = \{ y \text{ s.t. } y^T x_i = 0, \quad i = 1, \ldots, m \}$. Let $P_X$ be the $n \times n$ orthogonal projection matrix onto $X$.

   (a) What does $P_X x_i$ equal for $i = 1, \ldots, m$?

   (b) What does $P_X y_i$ equal for $i = 1, \ldots, n - m$?

   (c) $P_X$ has $n$ eigenvalues. Show using (a) and (b) that the eigenvalues will either equal zero or one. How many eigenvalues will be 1 and how many will be 0?

9. Let $P_X$ be the orthogonal projection matrix defined in the previous problem. Use (2.26) to show that $P_X$ is a symmetric matrix, and also that $P_X$ is idempotent ($P_X P_X = P_X$).
10. Let \( x_1, \ldots, x_m \) be linearly independent vectors in \( \mathbb{R}^n \) with \( m < n \), and let \( A = [x_1 \cdots x_m] \).
   (a) Is \( A^T A \) an invertible matrix? Why or why not.
   (b) Is \( A A^T \) an invertible matrix? Why or why not.

11. Given a \( p \times r \) matrix \( A \) and an \( r \times q \) matrix \( B \), show that
    \[
    \text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}.
    \]
    (Hint: Use Facts 2.16 and 2.17 and the sentences following them, as well as Definition 2.25.)

12. Let \( w_i \) be the \( i \)-th column of the matrix \( W \) shown below
    \[
    W = \begin{bmatrix}
    1 & 2 & 3 & 4 \\
    2 & 3 & 4 & 5 \\
    3 & 4 & 5 & 6 \\
    4 & 5 & 6 & 7 
    \end{bmatrix}.
    \]
    (a) What is the dimension of the subspace \( S = \text{col}(W) \)?
    (b) Find a basis for \( S \) whose elements are columns of \( W \).
    (c) Find a basis for \( S \) whose elements are columns of \( W \) and is different from the basis found in (b).

13. State whether the following statements are true or false and give an explanation for your answer.
   (a) Given \( A_{m \times n} \), if the rank of \( A = n \) then the only vector \( x \) such that \( Ax = 0 \) is \( x = 0 \).
   (b) Given 3 non-zero vectors \( x_1, x_2, x_3 \), if \( Ax_1 = 0, Ax_2 = 0, \) and \( Ax_3 = 0 \) then the nullity of \( A \) is greater than or equal to 3.
   (c) Let \( x \) and \( y \) be \( n \)-vectors. If \( A = xy^T \) then \( x \) is an eigenvector of \( A \).
   (d) If 3 linearly independent vectors \( x, y, \) and \( z \) are elements of a vector space \( S \) then the dimension of \( S \) is greater than or equal to 3.

14. Suppose that \( x_1, \ldots, x_k \) are basis vectors for a subspace \( S \) of \( \mathbb{R}^n \). These vectors can be placed as columns of a matrix \( X \). Let \( T \) be a \( k \times k \) nonsingular matrix and let
    \[ Y = XT. \]
    (a) Show that the columns of \( Y \) are linearly independent. (Hint: form the Gram matrix and use properties of determinants.)
    (b) Calculate the projection matrix for the column space of \( Y \). Show that this projection matrix is equal to the projection matrix for the column space of \( X \).

15. Consider the following vectors
    \[
    x = \begin{bmatrix}
    1 \\
    2 \\
    3 \\
    4 
    \end{bmatrix}, \quad v_1 = \begin{bmatrix}
    1 \\
    1 \\
    0 \\
    1 
    \end{bmatrix}, \quad v_2 = \begin{bmatrix}
    1 \\
    0 \\
    1 \\
    1 
    \end{bmatrix}.
    \]
Let $S$ be the subspace generated by $v_1$ and $v_2$.

(a) Show that $x$ is not in the subspace $S$.

(b) Compute $y = $ the projection of $x$ onto $S$.

(c) Show that $y$ is in $S$.

(d) Let $e = x - y$. Show that $e$ is orthogonal to $S$.

(e) Express $y$ as a linear combination of the vectors $v_1$ and $v_2$.

16. Given

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}, \quad y_1 = \begin{bmatrix} 6 \\ 9 \\ 12 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix}$$

(a) Show that $Ax = y_1$ has an infinite number of solutions parameterized by 1 free variable. Choose two different values of this variable, and generate two different solution vectors $x_1$ and $x_2$. Show by direct substitution that $Ax_i = y_1$, $i = 1, 2$.

(b) Show that $Ax = y_2$ does not have any solutions, but it does have a 1-parameter family of least-squares solutions. Calculate the min-norm solution $x_1$ and any other solution $x_2$ and show that they yield the same error: $\|Ax_1 - y_2\| = \|Ax_2 - y_2\|$.

17. Consider

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} x = \begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

(a) Calculate the min-norm solution.

(b) Generate another solution $x_1$ by adding a non-zero vector from the null-space of $A$ to $x_{\min - \text{norm}}$. Show that $\|x_1\| > \|x_{\min - \text{norm}}\|$.

18. Consider

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Show that a 2-parameter family of solutions exists for the above equation. Find 2 different solutions.

19. Let $A$ be an $m \times p$ matrix and $B$ be a $p \times m$ matrix. Suppose $x$ is an eigenvector of $AB$ corresponding to the non-zero eigenvalue $\lambda$. Show that $\lambda$ is also an eigenvalue of $BA$, and find the corresponding eigenvector. (Hint: Use the definition of an eigenvector given in (2.28).)

20. Show that if $\lambda$ is an eigenvalue of $A$, then $\alpha + \lambda$ is an eigenvalue of $\alpha I + A$.

21. (a) Show that the eigenvalues of a triangular matrix are equal to the elements on the main diagonal. (Hint: Use Fact 2.32 as well as a property of determinants given in Table 2.5 on page 45.)

(b) Show that the eigenvalues of a block triangular matrix are equal to the union of the eigenvalues of the matrices on the main diagonal. (Hint: Use the hint given for part (a).)
22. Test the following sets of vectors for linear dependence. If a set is linearly dependent, express one of the vectors in terms of the others in that set.

(a) 
\[
\begin{bmatrix}
1 \\
0 \\
0 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
1 \\
1 \\
1 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
3 \\
4 \\
1 \\
\end{bmatrix}
\]

(b) 
\[
\begin{bmatrix}
1 \\
0 \\
0 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
1 \\
1 \\
0 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
1 \\
1 \\
1 \\
\end{bmatrix}
\]

(c) 
\[
\begin{bmatrix}
1 \\
0 \\
1 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
0 \\
1 \\
1 \\
\end{bmatrix}, \quad 
\begin{bmatrix}
2 \\
1 \\
1 \\
\end{bmatrix}
\]

23. What is the dimension of the subspace generated by the sets of vectors in the previous problem? Give a basis for each subspace.

24. Show that, for any \( m \times n \) matrix \( A \), the rank of \( A \) plus the nullity of \( A^T \) equals \( m \). (Hint: use the definitions of rank and nullity.)

25. Using the matrix \( A \) and the vector \( y \) shown below, decompose \( y \) as \( y = y_A + y_{A^\perp} \) where \( y_A \in \text{col}(A) \) and \( y_{A^\perp} \in \text{orthogonal complement of col}(A) \).

\[
A = \begin{bmatrix}
1 & 1 \\
1 & 0 \\
1 & 1 \\
\end{bmatrix}, \quad y = \begin{bmatrix}
1 \\
2 \\
3 \\
\end{bmatrix}
\]

26. This problem shows how the singular value decomposition of a matrix \( A \) is related to an eigenvalue decomposition of the “squared” matrices \( A^T A \) and \( AA^T \).

(a) Show that if \( u \) is a left singular vector of \( A \) then it is an eigenvector of \( AA^T \).

(b) Show that if \( v \) is a right singular vector of \( A \) then it is an eigenvector of \( A^T A \).

(c) Show that if \( \sigma \) is a nonzero singular value of \( A \) then \( \lambda = \sigma^2 \) is a nonzero eigenvalue of \( AA^T \) or \( A^T A \).

27. This problem considers some special properties of eigenvectors and singular vectors of symmetric matrices. Let \( A \) be a (square) symmetric matrix so that \( A = A^T \).

(a) Let \( u_1 \) and \( u_2 \) be eigenvectors of \( A \) corresponding to distinct eigenvalues \( \lambda_1 \) and \( \lambda_2 \) \( (\lambda_1 \neq \lambda_2) \). Show that \( u_1 \) and \( u_2 \) are orthogonal. (Hint: show that \( \lambda_2 u_2^T u_1 = \lambda_1 u_2^T u_1 \).

(b) Let \( A \) be a symmetric matrix with distinct eigenvalues. The eigenvalue decomposition of \( A \) can always be written as \( A = U \Sigma U^{-1} \). However, because of the result in part (a), the eigenvalue decomposition of a symmetric matrix can be written as

\[
A = U \Sigma U^T, \quad U^T U = I.
\]
i. Show that the left singular vectors of $A$ can be chosen to be the columns of $U$.

ii. Show that the right singular vectors of $A$ can be chosen to be the columns of $U$.

(Hint: Use the fact that the left singular vectors of $A$ are eigenvectors of $AA^T$ and right singular vectors of $A$ are eigenvectors of $A^TA$.)

28. Given the vectors $x_1$ and $x_2$ shown below, construct the matrix $P$ which projects onto the subspace generated by $x_1, x_2$. First calculate $P$ by definition, then compare with the result obtained using the SVD.

\[
x_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

29. Let $S_1$ be the subspace generated by $x_1, x_2$, and let the subspace $S_2$ be the subspace generated by $y_1, y_2$.

(a) Develop a test to determine if $S_1 = S_2$. Use the test on the subspaces generated by the vectors in (b) and (c) below.

(b) \[
x_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad y_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}
\]

(c) \[
x_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad y_1 = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}
\]

30. Calculate the least-squares solution to the following problem using the “full-rank” formula. Show that the same result is obtained using the SVD.

\[
\begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}
\]

31. The least-squares solution for a full-rank least-squares problem is given by

\[
x_{LS} = (A^TA)^{-1}A^Ty.
\]

(a) Show that $\Sigma_1$ and $V_1$ from an SVD of $A$ are both square, invertible matrices. Also show that $V_1^{-1} = V_1^T$.

(b) Derive the alternate expression for $x_{LS}$ in terms of the SVD of $A$:

\[
x_{LS} = V_1 \Sigma_1^{-1} U_1^T y.
\]

32. Consider the linear equations $Ax = y$ where $A$ is an $m \times n$ matrix. Suppose $y \notin \text{col}(A)$ so that the equations do not have an exact solution. A least-squares solution
can be found using the SVD. Suppose further that the matrix $A$ has a non-trivial null
space. The SVD gives an orthonormal basis for this null space as the columns in $V_2$. The least-squares solution is not unique in this case, but all least-squares solutions are given by

$$x_{LS} = z + V_2\alpha$$

where $\alpha$ is a vector of $q$ free parameters and $V_2$ is an $n \times q$ matrix. Different choices for $\alpha$ will result in least-squares solutions with different properties. For example, if we choose $\alpha = 0$, then $x_{LS} = z$ is the least-squares solution of minimum norm.

Suppose we want to use the degrees of freedom given by $\alpha$ to satisfy additional con-
straints on $x_{LS}$ where the constraints are specified as the following linear equations

$$Mx_{LS} = w$$

and $M$ is a $q \times n$ matrix with $w \in \text{col}(M)$.

(a) Show that we can always choose $\alpha$ to obtain a least-squares solution which satis-

(b) Under what conditions will $\alpha$ be unique?