2.
Vector Space

Vector spaces provide the fundamental mathematical framework for analysis and design of linear control systems. It is a common practice to express dynamic systems as a matrix differential equation, and the state of a system in terms of a vector in an appropriate Euclidean space. For example, the motion of a point mass is uniquely described by Newton's laws of motion, and in terms of a vector consisting of two elements, position and velocity, in the two dimensional Euclidean space. Once a system is expressed in an appropriate vector space, its properties, such as stability, transient response, etc., can be quickly found by analyzing certain characteristics of system matrices. In addition, the desired performance of the system is also described in terms of suitable vectors and matrices, which the designer uses to design a controller. This chapter introduces the fundamentals of vector space that will be frequently used in latter chapters, while detailed concepts can be found in standard texts, such as [, , 23, 3].

2.1 Vector Space

A vector space $\mathcal{X}$ over the field of real or complex numbers is a nonempty set of elements equipped with two operations, addition and scalar multiplication, satisfying the following axioms:

1) If $\mathbf{x}$ and $\mathbf{y}$ are elements in $\mathcal{X}$ then $\mathbf{x}+\mathbf{y}$ is in $\mathcal{X}$ (Closure under addition)
2) $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$ (Commutativity of addition)
3) $\mathbf{x}+(\mathbf{y}+\mathbf{z})=(\mathbf{x}+\mathbf{y})+\mathbf{z}$ (Associativity of addition)
4) There is a null element 0 in $\mathcal{X}$ such that $\mathbf{x}+0=\mathbf{x}$ (Existence of a null element)
5) For each $\mathbf{x}$ there is a negative $\mathbf{x}$, denoted, $-\mathbf{x}$, such that $\mathbf{x}+(-\mathbf{x})=0$ (Existence if a negative element)
6) If $\mathbf{x}$ is an element in $\mathcal{X}$ then $c\,\mathbf{x}$ is also an element in $\mathbf{x}$, where c is a scalar (Closure under scalar multiplication)

7) For any scalar c, $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$ (Distributivity under scalar multiplication)
8) $(c_1 + c_2)\mathbf{x} = c_1\mathbf{x} + c_2\mathbf{x}$, where c_1 and c_2 are scalars (Distributivity under scalar multiplication)
9) $c_1(c_2\mathbf{x}) = c_1c_2\mathbf{x}$ (Associativity under scalar multiplication)
10) $1\mathbf{x} = \mathbf{x}$ (Existence of a unit element in the field)

The three dimensional Euclidean space $\mathcal{R}^3$ is a vector space. Vectors in $\mathcal{R}^3$ have the mathematical representation as $\mathbf{x} = [x_1\ x_2\ x_3]'$, and geometrical representation as a directed line segment joining the origin and the point $[x_1\ x_2\ x_3]'$. The reader can easily verify that all the axioms listed above are satisfied by vectors in $\mathcal{R}^3$. Note however that vectors in n-dimensional Euclidean space, for $n \geq 3$, don't have any geometric representation perceivable by human eyes.

The set of all three dimensional vectors whose entries are non-negative is not a vector space; the reader can easily verify that not all the axioms of a vector space described above are satisfied.

Consider the electrical circuit consisting of 2 meshes. Using the Kirchoff's voltage law, we can completely describe the circuit by writing two mesh equations, with the mesh currents $I = (I_1, I_2)$ that are unknown. Theoretically these two mesh currents can take any arbitrary values depending on the applied voltage and the resistances in the circuit. We can say that I is an element in the vector space $\mathcal{R}^2$.

The motion of a particle in space can be uniquely described in terms of its position, x, and velocity, v; the vector $[\,x \quad v\,]'$ is an element of the two dimensional Euclidean space $\mathcal{R}^2$.

The notion of vector space can be further generalized to include spaces that contain not just traditional vectors, but other kinds of elements. Consider, for example, $\mathcal{X}$ being the space of all 2×2 matrices with entries that are real numbers and take values from $-\infty$ to $+\infty$. An element in $\mathcal{X}$ can be represented by

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix},$$

where x_1, x_2, x_3, x_4 are arbitrary real numbers, and a null element has the form

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

For the space $\mathcal{X}$, define addition of vectors and scalar multiplication by

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}, \qquad c\mathbf{A} = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}.$$

Then it can be verified that all the axioms of a vector space given above are satisfied. Therefore the space $\mathcal{X}$ defined above is a vector space. But the elements of this vector space are not *vectors* in the usual sense. For this reason, a vector space is often called a *linear space*. For control system design and analysis we shall deal mostly with the n-dimensional Euclidean vector space. Therefore the rest of this chapter will be restricted to the real vector space $\mathcal{R}^n$.

Scalar Product

Suppose $\mathcal{X}$ is a linear vector space, and $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathcal{X}$. A scalar product (also known as inner product) is a function that maps vectors from the vector space $\mathcal{X}$ to real numbers satisfying the following properties:

1) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x}$,
2) $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = 0$,
3) $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$, where bar denotes conjugate,
4) $\langle c\mathbf{x}, \mathbf{y} \rangle = \bar{c}\langle \mathbf{x}, \mathbf{y} \rangle, \quad \langle \mathbf{x}, c\mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$, where c is a complex number.
5) $\langle \mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2 \rangle = \langle \mathbf{x}, \mathbf{y}_1 \rangle + \langle \mathbf{x}, \mathbf{y}_2 \rangle, \qquad \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{X}$.

For $\mathcal{R}^n$, let $\mathbf{x}$ and $\mathbf{y}$ be two arbitrary vectors defined as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Then for real vector space, the scalar product can be defined as

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\ &= \sum_{i=1}^{n} x_i y_i \\ &= \mathbf{x}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{x}. \end{aligned} \tag{2.1}$$

The reader should verify that all properties of scalar product are satisfied. In case the vectors are complex, one takes

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

where $*$ denotes conjugate transpose of the vector $\mathbf{x}$.

For example, suppose

$$\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ 4 \\ 0 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 2 \end{bmatrix} \tag{2.2}$$

then it is easy to compute the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle = 16$. Similarly, for

$$\mathbf{x} = \begin{bmatrix} 1+i \\ 2-i \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 1-2i \\ 1-i \end{bmatrix}$$

the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle = 2 - 4i$. It is also easily verified that $\langle \mathbf{y}, \mathbf{x} \rangle = 2 + 4i$, which is the conjugate of $\langle \mathbf{x}, \mathbf{y} \rangle$.

Scalar product is also frequently known as *dot product*, and is denoted by $\mathbf{x} \cdot \mathbf{y}$.

Norm

Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in a vector space $\mathcal{X}$, and k be a real scalar. Then the norm, denoted $\|\cdot\|$, is a function mapping the space $\mathcal{X}$ to real numbers satisfying properties

1) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x}$,
2) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$,
3) $\|k\mathbf{x}\| = |k|\|\mathbf{x}\|$,
4) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\ \| + \|\mathbf{y}\|$, (Triangle inequality).

A frequently used definition of norm for $\mathcal{R}^n$, is given by

$$\|\mathbf{x}\| = \left[\sum_{i=1}^{n} x_i^2\right]^{\frac{1}{2}}. \tag{2.3}$$

For $\mathcal{R}^n$, the first three properties of norm are easily verified; the fourth property will be proved after the introduction of Cauchy-Schwartz inequality.

Using (2.1), it is also clear that $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle$. For example, if

$$\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix},$$

the norm of $\mathbf{x}$ is $\|\mathbf{x}\| = \sqrt{2^2 + (-1)^2 + (-2)^2} = 3$. Norm for $\mathcal{R}^n$ can be defined in various other ways, see for example, exercise at the end of the chapter.

In case $\mathcal{X} = \mathcal{C}^n$, the complex space of dimension n, the norm could be defined as

$$\|\mathbf{x}\|^2 = \sum_{i=1}^{n} \bar{x}_i \, x_i = \mathbf{x}^* x$$

where $(\cdot)^*$ denotes the conjugate transpose.

In $\mathcal{R}^2$ and $\mathcal{R}^3$, the norm can be visualized as the actual length of a vector. The first three properties of norm can be easily proved using the definition (2.3). The fourth property is a generalization of the well-known fact of Euclidean geometry that sum of two sides of a triangle is always greater than the third side.

Norm on $\mathcal{R}^n$ can be defined in several ways, for example,

$$\begin{aligned}
\|\mathbf{x}\|_1 &= \sum_{i=1}^{n} |x_i|, \quad \text{taxi-cab norm} \\
\|\mathbf{x}\|_2 &= \Big(\sum_{i=1}^{n} |x_i|^2\Big)^{\frac{1}{2}}, \quad \text{Euclidean norm} \\
\|\mathbf{x}\|_p &= \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{\frac{1}{p}}, \; 1 \leq p < \infty, \quad p\text{-norm} \\
\|\mathbf{x}\|_\infty &= \max_i |x_i|, \quad \text{sup norm}
\end{aligned}$$

Note also that the various norms on $\mathcal{R}^n$ are equivalent. It can be shown that

$$\|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2 \le n\|\mathbf{x}\|_\infty$$
$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1$$

In fact, let a vector space $\mathcal{X}$ has two norms denoted as $\|\cdot\|_a$ and $\|\cdot\|_b$. Then these norms are said to be equivalent if

$$k_1 \le \frac{\|\mathbf{x}\|_a}{\|\mathbf{x}\|_b} \le k_2, \quad \text{for all } x \in \mathcal{X}, \quad k_1, k_2 > 0$$

Unit Vector: If the norm of a vector is 1, then it is called a *unit vector.* Mathematically a unit vector can be expressed as $\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$.

Orthogonal Vector: *The vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *are said to be orthogonal, if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. *If two unit vectors are orthogonal, they are called orthonormal vectors.* For example, the vectors $\mathbf{x} = [1 \quad 2 \quad 0 \quad -2]'$ and $\mathbf{y} = [0 \quad -1 \quad 1 \quad -1]'$ are orthogonal since $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Furthermore, the vectors $\hat{\mathbf{x}} = \frac{1}{3}[1 \quad 2 \quad 0 \quad -2]'$ and $\hat{\mathbf{y}} = \frac{1}{\sqrt{3}}[0 \quad -1 \quad 1 \quad -1]'$ are orthonormal.

In $\mathcal{R}^2$, the vectors $\mathbf{x}$ and $\mathbf{y}$ in Figure 3.1,

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

are orthogonal since $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. In $\mathcal{R}^2$ and $\mathcal{R}^3$, orthogonality can also be visualized as two vectors being perpendicular to each other. In fact, the angle between any two vectors can be computed using

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|\,\|\mathbf{y}\| \cos\theta,$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. For the above example, $\cos\theta = 0$ so that the angle between the two vectors is $90°$. This geometric interpretation holds for higher dimensional spaces only in a generalized sense, but not in the sense of two or three dimensional Euclidean geometry.

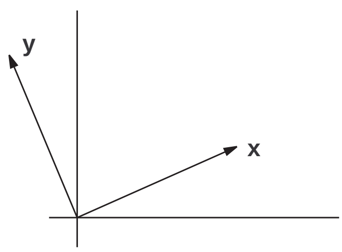


Figure 2.1 Orthogonal Vectors

Numerically, the norm of a vector can be computed using its basic definition. In the `MATLAB` environment, the macro `norm` can be used to compute the norm of a vector.

One of the important inequalities used in linear algebra and mathematical analyses of control systems is the **Cauchy-Schwarz inequality**: *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two vectors in* $\mathcal{R}^n$*. Then*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \; \|\mathbf{y}\|. \tag{2.4}$$

To prove this well-known result, consider the vector $\alpha\mathbf{x}+\mathbf{y}$, where α is a scalar. Then using the properties of scalar product and norm, we have

$$\begin{aligned} 0 &\leq \|\alpha\mathbf{x} + \mathbf{y}\|^2 \\ &= \langle \alpha\mathbf{x} + \mathbf{y}, \alpha\mathbf{x} + \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2\alpha^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle\alpha + \|\mathbf{y}\|^2. \end{aligned}$$

which is a quadratic polynomial of α. Since the value of this polynomial function is non-negative for all α, there can be no real root except possibly at 0. Thus its discriminant must be less than or equal to zero so that

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leq 0$$

which yields the inequality (2.3). Note that if $4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 > 0$, there will be two real roots of the polynomial, which means that part of the parabola will be below the zero line. ▷

Using the Cauchy-Schwarz inequality (2.4), we can prove the fourth property of norm, i.e., the triangle inequality. Indeed,

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2|\langle \mathbf{x}, \mathbf{y} \rangle| + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \; \|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \end{aligned}$$

which proves the triangle inequality.

Metric or Distance

Let $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ be vectors in a vector space $\mathcal{X}$. Then the distance (or metric) between the vectors $\mathbf{x}$ and $\mathbf{y}$, denoted $d(\mathbf{x}, \mathbf{y})$, is a scalar valued function satisfying

1) $d(\mathbf{x}, \mathbf{y}) \geq 0$,
2) $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$,
3) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$,
4) $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$, (Triangle inequality).

These properties are satisfied if one takes

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|. \tag{2.5}$$

For the vectors shown in equation (2.2), we obtain $d(\mathbf{x}, \mathbf{y}) = \sqrt{(2-1)^2 + (-1+2)^2 + (4-3)^2 + (0-2)^2} = \sqrt{7}$. In general, a vector $\mathbf{x}$ in $\mathcal{R}^n$ can be thought of as a point in $\mathcal{R}^n$. Then the metric represents the distance between the two given points.

The first three properties of metric are easily proved using the properties of norm. For the fourth property, i.e., the triangle inequality, note that

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{x} - \mathbf{z} + \mathbf{z} - \mathbf{y}\| \\ &\leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{y}\| \\ &= d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}). \end{aligned}$$

This fourth property of metric is a generalization of the fact that shortest distance between two points is a straight line. It can also be compared with the triangle inequality of norm if one notes that $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ represent the three vertices of a triangle.

2.2 Linear Dependence

Linear dependence is an important concept commonly used in the design of controllers for linear time invariant systems. As we shall see in subsequent chapters, feedback control design for multi-input systems requires selection of a set of linearly independent vectors from a given set of vectors. Specifically, for an n-dimensional dynamic system with m control inputs, one has to select a total of n linearly independent vectors from a set of nm vectors that form the controllability matrix. Also in Chapter 11 we shall see that the designer has to select a set of n linearly independent vectors representing the desired system modes or the desired transient response of the system. So what is linear dependence?

Let $\{\mathbf{x}_i\} = \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ be a set of vectors in $\mathcal{R}^n$. *The vectors $\{\mathbf{x}_i\}$ are said to be linearly dependent if there exists a set of n scalars, $\{a_i\}$, not all identically zero, such that*

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots a_m\mathbf{x}_m = 0. \tag{2.6}$$

This essentially means that any vector in the set $\{\mathbf{x}_i\}$ can be expressed as a linear combination of the rest of the vectors in the set, i.e., by summing up the rest of the vectors with appropriate multiplication factors.

If the vectors of the set $\{\mathbf{x}_i\}$ are not linearly dependent, then they are called linearly independent. More formally, if (2.6) implies $a_1 = a_2 = \cdots = a_m = 0$, then the vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ are linearly independent. This means that any vector in a linearly independent set cannot be expressed by linear combination of rest of the vectors in the set. The Kronecker vectors in $\mathcal{R}^n$ are linearly independent.

Example 2.1 In $\mathcal{R}^3$, consider the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

Then we observe that $\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_3 = 0$ so that $a_1 = 1$, $a_2 = 1$ and $a_3 = -1$, so that the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are linearly dependent. ▹

Example 2.2 For the set of vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

we observe that the equation (2.6) can hold only if $a_1 = a_2 = a_3 = 0$ so that the above vectors are linearly independent. ▹

Test for Linear Dependence

Linear dependence or independence of vectors cannot always be determined simply by observation, especially if the vectors do not consist of simple rational numbers. This section presents two formal mathematical methods that can be used to determine linear independence of vectors.

Equation (2.6) can be rewritten in the matrix form as

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = 0. \tag{2.7}$$

Since each vector $\mathbf{x}_i$ is a vector in $\mathcal{R}^n$ such that

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix}$$

$\mathbf{X} = [\,\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \mathbf{x}_m\,]'$ is a matrix of dimension $n \times m$. Thus equation (2.7) is equivalent to a system of n-linear equations with m-unknowns, $a_1, a_2, \cdots, a_m$. The question of linear dependence is, therefore, equivalent to the question of existence of a nontrivial solution of equation (2.7). If there is a nontrivial solution of equation (2.7) which implies that there is a set of constants, a_i's, not all zero, then the given vectors are linearly dependent. If the solution of the equation (2.7) is identically zero, i.e., each $a_i, i = 1, 2, \cdots, m$ is zero, then the vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ are linearly independent.

Test for Linear Dependence – 1

Let $\{\mathbf{x}_i\} = \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ be a set of n-vectors in $\mathcal{R}^n$. Then this set is linearly dependent if $|\mathbf{X}| = 0$, and linearly independent if $|\mathbf{X}| \neq 0$, where $|\mathbf{X}|$ denotes the determinant of the matrix $\mathbf{X} = [\,x_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n\,]$.

The proof follows easily from equation (2.7). Since each $\mathbf{x}_i$ is a vector in $\mathcal{R}^n$ and there are n vectors that are to be tested, the matrix $\mathbf{X}$ is of size $n \times n$. Define $\mathbf{a} = [\,a_1 \quad a_2 \quad \cdots \quad a_n\,]$. Then (2.7) represents a homogeneous equation with n-unknowns, and has the form $\mathbf{Xa} = 0$. If $\mathbf{X}$ is invertible, we have $\mathbf{a} = \mathbf{X}^{-1}0 = 0$ so that (2.7) has only a trivial solution, and hence, the set $\{\mathbf{x}_i\}$ is linearly independent. But a matrix is invertible only if its determinant is nonzero. Thus the set $\{\mathbf{x}_i\}$ is linearly independent if the determinant of the matrix $\mathbf{X}$ is nonzero.

Example 2.3 Considering the vectors of Example 2.1, we have

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

The determinant of the matrix $\mathbf{X}$ is zero so that the vectors are linearly dependent. Likewise the determinant of the matrix formed by the vectors of Example 2.2 is 1 so that the given vectors are linearly independent. ▷

Test for Linear Dependence – 2

The above method of linear dependency cannot be used if the matrix $\mathbf{X}$ formed by the given set of vectors is not square. This can arise if there are m vectors to be tested where each vector is in $\mathcal{R}^n$, and $m \neq n$. For this case the following modified method is helpful.

Consider the set of m-vectors $\{\mathbf{x}_i\}$, and let each $\mathbf{x}_i \in R^n$, where $n \neq m$. We premultiply the equation (2.6) sequentially by the transpose of the given vectors, i.e., by $\mathbf{x}_1', \mathbf{x}_2', \cdots, \mathbf{x}_m'$, and form the equations

$$\begin{aligned} & a_1\mathbf{x}_1'\mathbf{x}_1 + a_2\mathbf{x}_1'\mathbf{x}_2 + \cdots + a_n\mathbf{x}_1'\mathbf{x}_m = 0 \\ & a_1\mathbf{x}_2'\mathbf{x}_1 + a_2\mathbf{x}_2'\mathbf{x}_2 + \cdots + a_n\mathbf{x}_2'\mathbf{x}_m = 0 \\ & \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ & a_1\mathbf{x}_m'\mathbf{x}_1 + a_2\mathbf{x}_m'\mathbf{x}_2 + \cdots + a_m\mathbf{x}_m'\mathbf{x}_m = 0, \end{aligned} \tag{2.8}$$

which can be written as

$$\begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \cdots & \mathbf{x}_1'\mathbf{x}_m \\ \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 & \cdots & \mathbf{x}_2'\mathbf{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_m'\mathbf{x}_1 & \mathbf{x}_m'\mathbf{x}_2 & \cdots & \mathbf{x}_m'\mathbf{x}_m \end{bmatrix}_{m\times m} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.9}$$

Symbolically we express the above equation as

$$\mathbf{Ma} = 0 \tag{2.10}$$

where $\mathbf{M} = \mathbf{X}'\mathbf{X}$, with $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_m]$, and $\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_m]$. If the matrix $\mathbf{M}$ has a nonzero determinant, then from (2.10) we have $\mathbf{a} = \mathbf{M}^{-1}0 = 0$ so that the vectors $\mathbf{x}_1, x_2, \cdots x_m$ are linearly independent. On the other hand, if the determinant of the matrix $\mathbf{M}$ is zero, the (2.10) may have a nontrivial solution $\mathbf{a}$ satisfying (2.6), and the set $\{\mathbf{x}_i\}$ is linearly dependent. The result is summarized in the following:

A necessary and sufficient condition for the vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, x_m$ *to be linearly dependent is that* $|\mathbf{M}| = 0$*, where* $\mathbf{M} = \mathbf{X}'\mathbf{X}$*, and* $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_m]$*.* Note that the above method applies also to the case when there are n vectors and each vector is in R^n.

Example 2.4 Consider the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then we compute

$$\mathbf{M} = \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \mathbf{x}_1'\mathbf{x}_3 \\ \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_3 \\ \mathbf{x}_3'\mathbf{x}_1 & \mathbf{x}_3'\mathbf{x}_2 & \mathbf{x}_3'\mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 3 \\ 0 & 2 & 3 \\ 3 & 3 & 7 \end{bmatrix}$$

and $|\mathbf{M}| = 12$ so that the vectors $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ are linearly independent. ▷

Example 2.5 Considering the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 3 \\ -3 \\ -1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

we verify that $|\mathbf{M}| = 0$ so that the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are linearly dependent. Indeed, it is easy to see that $2\mathbf{x}_1 - \mathbf{x}_2 - \mathbf{x}_3 = 0$. ▷

Test for Linear Dependence – 3

A numerically efficient method for determining the vectors that are linearly independent is the use of elementary row operation. To determine linear dependence of a set of vectors, it suffices to solve equation (2.7) for the unknown coefficients, $a_1, a_2, \cdots, a_n$. If the solution is obtained as $a_1 = a_2 = \cdots = a_n = 0$, then the given vectors are linearly independent, and if some of these coefficients are non-zero, then the given vectors are linearly dependent. This can be explained through the following example:

Example 2.6 Consider the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1.4 \\ 1.6 \\ 0.6 \end{bmatrix},$$

Then we solve

$$\begin{bmatrix} 2 & 1 & 1.4 \\ 1 & 2 & 1.6 \\ 0 & 1 & 0.6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then we perform row operations as shown below to arrive at the row-reduced echelon form of $\mathbf{X}$.

$$\begin{bmatrix} 2 & 1 & 1.4 & 0 \\ 1 & 2 & 1.6 & 0 \\ 0 & 1 & 0.6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0.5 & 0.7 & 0 \\ 1 & 2 & 1.6 & 0 \\ 0 & 1 & 0.6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0.5 & 0.7 & 0 \\ 0 & 1.5 & 0.9 & 0 \\ 0 & 1 & 0.6 & 0 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & 0.5 & 0.7 & 0 \\ 0 & 1 & 0.6 & 0 \\ 0 & 1 & 0.6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0.4 & 0 \\ 0 & 1 & 0.6 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The row reduced matrix is interpreted as

$$\begin{bmatrix} 1 & 0 & 0.4 \\ 0 & 1 & 0.6 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which is equivalent to

$$a_1 + 0.4a_3 = 0$$
$$a_2 + 0.6a_3 = 0.$$

A possible solution of the above equation is taken as $a_1 = 0.4, a_2 = 0.6, a_3 = -1$ which gives

$$0.4\mathbf{x}_1 + 0.6\mathbf{x}_2 - \mathbf{x}_3 = 0.$$

This may be considered as equivalent to the statement that the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent, and $\mathbf{x}_3$ is linearly dependent on the first two vectors. One can also interpret the above equation as "$\mathbf{x}_1$ being linearly dependent on $\mathbf{x}_2, \mathbf{x}_3$", or as "$\mathbf{x}_2$ being dependent on $\mathbf{x}_1, \mathbf{x}_3$".

Note that the solution of the unknown coefficients, $\mathbf{a}$, is also quickly obtained if we simply take it as the last column of the row reduced coefficient matrix and replace its entry corresponding to the diagonal position by -1. Summarizing the above discussion we have the following:

- if the row-reduced matrix has a 1 in a diagonal position, the corresponding original vector can be taken as a linearly independent vector,
- if the row-reduced matrix has a 0 in a diagonal position with possibly nonzero numbers at off-diagonal positions, the corresponding original vector is linearly dependent. Furthermore, the entries of this column can be used to denote actual dependence of the dependent vector.
- There are multiple ways one can express linear dependence of vectors. $\triangleright$

As discussed at the beginning of Section 2.2, determination of linearly independent vectors is an important step of feedback control design for multi-input systems. We shall see an immediate application of the results presented here in Chapters 10 and 11.

2.3 Span and Basis

We know that the space $\mathcal{R}^3$ has three coordinate axes, and these three coordinates uniquely determine an arbitrary vector in $\mathcal{R}^3$. So a fundamental question is how many coordinate axes are there for a given vector space? First some definitions:

Let $\mathcal{V}$ be a vector space, and $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$, not all linearly independent, be a set of vectors in $\mathcal{V}$. Then the span of the set $\mathcal{S}$ is the subspace $\mathcal{W}$ of $\mathcal{V}$ consisting of vectors formed by all possible linear combination of the vectors in $\mathcal{S}$. Mathematically

$$\begin{aligned} \mathcal{W} &= \ \mathrm{Span}(\mathcal{S}) \\ &= \{v : v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n, \quad c_1, c_2, \cdots, c_n \quad \text{arbitrary}\}. \end{aligned}$$

If a vector in $\mathcal{S}$ is linearly dependent, then it can be removed from $\mathcal{S}$ without changing the space $\mathcal{W}$. For example, given $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, where $\mathbf{v}_3$ is linearly dependent, we have

$$\mathcal{W} = \mathrm{Span}(\mathcal{S}) = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}.$$

We say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$ span the space $\mathcal{W}$. This also means that any vector in $\mathcal{W}$ can be expressed as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_4$.

Basis

Consider the three dimensional Euclidean space $\mathcal{R}^3$. It has three coordinate axes, and any vector in $\mathcal{R}^3$ is expressed with respect to these three coordinate vectors. The concept of basis is a generalization of this coordinate axes so that elements in a vector space are uniquely expressed in terms of these *bases.*

Let $\mathcal{V}$ be a vector space. Then the minimal set of vectors in $\mathcal{V}$ that spans the entire space $\mathcal{V}$ is called its basis. This means that any vector in $\mathcal{V}$ can be constructed in terms of its basis, in particular, by taking the linear sum of vectors in the basis set. Note the phrase *minimal* in the definition of basis. The vectors in the basis set must also be linearly independent since these vectors must span the entire space. Addition of a linearly dependent vector does not constitute a minimal set since it is superfluous in the sense that it can be expressed as a linear combination of the rest.

An arbitrary vector in $\mathcal{V}$ can always be expressed as a linear combination of the basis vectors. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be the basis, and let $\mathbf{y} \in \mathcal{V}$ be an arbitrary vector. Then $\mathbf{y}$ can be expressed as

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n,$$

for some nonzero scalars $a_1, a_2, \cdots a_n$. *The dimension of the vector space $\mathcal{V}$ is the number of vectors in its basis set.* Thus for an n-dimensional space, there are exactly n-basis vectors. A vector space is called *finite dimensional space* if the number of vectors in the basis set is finite. $\mathcal{R}^n$ is a finite dimensional vector space. A linear space is *infinite dimensional space* if the basis set has infinitely many elements. This will be further discussed in Chapter 4.

Example 2.7 Consider the Euclidean space $\mathcal{R}^3$. Then Kronecker vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

is a basis for $\mathcal{R}^3$ since these vectors are linearly independent, and span the entire space $\mathcal{R}^3$, i.e., any vector in $\mathcal{R}^3$ can be expressed as a linear combination of these three vectors. ▷

For a given vector space, the basis set is not necessarily unique. Consider for example, in $\mathcal{R}^3$, the set of vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

First of all we verify that the above vectors are linearly independent. Secondly we form a space $\mathcal{W}$ as the span of these vectors, $\mathcal{W} = \text{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. The dimension of $\mathcal{W}$ is three since there are three linearly independent vectors. Finally we note that any vector in $\mathcal{R}^3$ can be expressed as a linear combination $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ so that $\mathcal{W}$ is same as $\mathcal{R}^3$ or the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ form a basis for $\mathcal{R}^3$. For example, if $\mathbf{y} = [1 \quad 2 \quad 3]'$, we have $\mathbf{y} = 0\mathbf{v}_1 + 2\mathbf{v}_2 + 1\mathbf{v}_3$. Note that the standard Kronecker vectors also form a basis for $\mathcal{R}^3$.

Consider the following set of vectors:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

It is easy to verify that these vectors are linearly dependent. Furthermore, one can take $\mathbf{v}_3$ as the dependent vector, and $\mathbf{v}_1$ and $\mathbf{v}_2$ as the linearly independent vectors. Can any vector in $\mathcal{R}^3$ be expressed as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$? Let $\mathbf{y} = [1 \quad 0 \quad 1]'$. Since the third entry of $\mathbf{y}$ is 1 and the third entries of both $\mathbf{v}_1$ and $\mathbf{v}_2$ are zero, it is impossible to express $\mathbf{y}$ in terms of a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Thus the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is not a basis of $\mathcal{R}^3$.

Consider the following example of four vectors in $\mathcal{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

We verify that the above set of vectors is not linearly independent, therefore cannot constitute a basis for $\mathcal{R}^3$. Next we prove that $\mathbf{v}_4$ can be considered as a linearly dependent vector, and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent, and span the entire $\mathcal{R}^3$. Thus a basis for $\mathcal{R}^3$ is the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

Expansion of a given vector in terms of Basis

By definition, a basis is a set of vectors that span the entire vector space, i.e., any vector in a vector space has a representation in terms of the basis of that vector space. Given an arbitrary vector, the following method can be used to express it as a linear combination of the basis. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be a basis set of $\mathcal{V}$, and $\mathbf{y}$ is a given vector in $\mathcal{V}$, such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{y}, \tag{2.11}$$

where $a_1, a_2, \cdots, a_n$ are unknown. Then scalar multiplying the above equation by the transpose of vectors of the basis set, we have

$$\begin{bmatrix} \mathbf{v}_1'\mathbf{v}_1 & \mathbf{v}_1'\mathbf{v}_2 & \cdots & \mathbf{v}_1'\mathbf{v}_n \\ \mathbf{v}_2'\mathbf{v}_1 & \mathbf{v}_2'\mathbf{v}_2 & \cdots & \mathbf{v}_2'\mathbf{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_n'\mathbf{v}_1 & \mathbf{v}_n'\mathbf{v}_2 & \cdots & \mathbf{v}_n'\mathbf{v}_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1'\mathbf{y} \\ \mathbf{v}_2'\mathbf{y} \\ \vdots \\ \mathbf{v}_n'\mathbf{y} \end{bmatrix}. \tag{2.12}$$

The above equation can then be solved for the unknowns $a_1, a_2, \cdots, a_n$. Defining the matrix $\mathbf{V} = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,]$, the above equation can be written in the compact form as

$$\mathbf{V}'\mathbf{V}\mathbf{a} = \mathbf{V}'\mathbf{y}$$

so that the unknown coefficients are obtained as

$$\mathbf{a} = [\mathbf{V}'\mathbf{V}]^{-1}\mathbf{V}'\mathbf{y}. \tag{2.13}$$

The following example shows the details:

Example 2.8 Consider the following basis for $\mathcal{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

and a vector $\mathbf{y} = [\,1 \quad 2 \quad 3\,]'$. Then from (2.12) we have

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix},$$

from which we obtain $a_1 = 0, a_2 = 2$ and $a_3 = 1$ so that $\mathbf{y} = 0\mathbf{v}_1 + 2\mathbf{v}_2 + 1\mathbf{v}_3$.

Another way of expressing the given vector in terms of a basis is to solve equation (2.11) directly. For example, rewriting (2.11) in a matrix form we have

$$[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = y$$

This equation can then be solved by elementary row reduction to obtain the expansion coefficients $a_1, a_2, \cdots, a_n$.

2.4 Change of Basis

Fundamentally basis vectors are the coordinate axes of a vector space. For example, the Kronecker vectors are the usual coordinate axes of the Euclidean space. However, system analysis often becomes relatively simple if one defines a new set of basis vectors rather than using the Kronecker vectors. We shall discuss engineering relevance of basis in Chapter 5, and in fact, we shall see that analysis of dynamic systems as well as design of controllers are relatively easy if one chooses an appropriate basis. As we have seen earlier, a vector space may have multiple bases, nevertheless for an engineering analysis, the choice of a particular basis may be more convenient than the other bases. This requires that the system description given in one set of basis be changed to a different set of basis. The following introduces the concept of change of basis:

Let's first consider an example in the two dimensional Euclidean space $\mathcal{R}^2$. The usual basis for this space is the Kronecker set $\{\mathbf{v}_1, \mathbf{v}_2\} = \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$. This means that an arbitrary vector, say, $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, has the representation

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.14}$$

Let's choose a new set of coordinate axes for $\mathcal{R}^2$ as $\{\mathbf{w}_1, \mathbf{w}_2\} = \{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\}$ as shown in the figure.

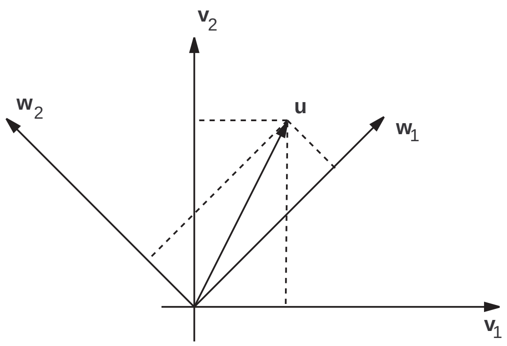


Figure 2.2 Change of Basis

Then it is easy to verify that the vector $\mathbf{u}$ can be represented as

$$\mathbf{u} = 1.5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Thus the vector $\mathbf{u}$ has two different representations in terms of two bases as

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\mathbf{v}\text{-basis}} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}_{\mathbf{w}\text{-basis}}.$$

Although the above transformation of the vector $\mathbf{u}$ has been done by observation or by trigonometric construction, indeed, as we shall see in the following that a change of basis essentially involves application of a certain transformation.

In fact, the transformation matrix for a change of basis can be found using the very definition of basis. Suppose $\mathbf{u}$ is a given vector in a vector space that has two basis sets, denoted $\mathbf{V}$ and $\mathbf{W}$. Then $\mathbf{u}$ has the representation both in $\mathbf{v}$ and $\mathbf{w}$ bases, so that

$$\mathbf{u} = b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + \cdots + b_n\mathbf{w}_n = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n,$$

or equivalently

$$\begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

so that

$$\mathbf{Wb} = \mathbf{Va},$$

or

$$\mathbf{b} = \mathbf{W}^{-1}\mathbf{Va}, \tag{2.15}$$

which gives the representation of a given vector $\mathbf{u}$ in $\mathbf{V}$-basis set in terms of a new $\mathbf{W}$-basis.

Example 2.9

Consider the following two bases for $\mathcal{R}^3$:

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \qquad \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} = \left\{ \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

and a vector $\mathbf{y} = [\,3 \quad 0 \quad 9\,]'$ expressed in terms of the Kronecker basis. Then using the method discussed in Section 2.3, equation (2.13), we obtain

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 = -2\mathbf{v}_1 + 5\mathbf{v}_2 + 2\mathbf{v}_3,$$

which can also be expressed in terms of $\mathbf{w}$-basis using (2.15) as

$$\begin{aligned} \mathbf{b} = \mathbf{W}^{-1}\mathbf{Va} &= \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 5 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -10.5 \\ 4.5 \\ -6.0 \end{bmatrix}. \end{aligned}$$

The transformation matrix for the change of basis is

$$\mathbf{W}^{-1}\mathbf{V} = \begin{bmatrix} 2 & -0.5 & -2 \\ 0 & 0.5 & 1 \\ 1 & 0 & -2 \end{bmatrix}.$$

$\triangleright$

Gram-Schmidt Orthogonalization Method

Mathematical analyses involving bases become significantly simple if the basis vectors are orthogonal. Consider, for example, the problem of finding the representation of an arbitrary vector in terms of a basis. The method has been presented above in Section 2.3, equation (2.12). If the basis vectors were orthogonal, one would have $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$, for $i \neq j$ so that equation (2.12) would simplify to

$$\begin{bmatrix} \mathbf{v}_1'\mathbf{v}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{v}_2'\mathbf{v}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{v}_n'\mathbf{v}_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1'\mathbf{y} \\ \mathbf{v}_2'\mathbf{y} \\ \vdots \\ \mathbf{v}_n'\mathbf{y} \end{bmatrix}, \tag{2.16}$$

which is easily solvable. In fact, if the basis vectors are orthonormal, then the coefficient matrix in the left hand side reduces even further to an identity matrix. Thus for simplification of mathematical analysis, a new set of orthonormal (or at least orthogonal) basis vectors be formed from the given set of basis. The Gram-Schmidt orthogonalization method is a commonly used tool for this purpose.

The problem is stated as follows: In $\mathcal{R}^n$, given a basis set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$, find a new orthonormal basis set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ satisfying

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$$

The Gram-Schmidt construction of the orthonormal vectors is presented in the following theorem:

Theorem 2.1. *Let $\{\mathbf{v}_i, i = 1, 2, \cdots, n\}$ be a basis set on $\mathcal{R}^n$. Then there exists an orthogonal basis, denoted $\{\mathbf{u}_i, i = 1, 2, \cdots, n\}$, such that*

$$\mathbf{u}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{v}_i \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j, \qquad i = 1, 2 \cdots n. \tag{2.17}$$

Furthermore, the set $\hat{\mathbf{u}}_i = \{\frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}\}$, $i = 1, 2, \cdots, n$, is orthonormal.

Proof. We prove the result by induction. Let $\mathbf{u}_1 = \mathbf{v}_1$, and $\mathbf{u}_2 = \mathbf{v}_2 - c_1\mathbf{u}_1$, where c_1 must be selected so that $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0$. Clearly, taking scalar product of $\mathbf{u}_2$ with $\mathbf{u}_1$, we have

$$0 = \langle \mathbf{u}_2, \mathbf{u}_1 \rangle = \langle \mathbf{v}_2, \mathbf{u}_1 \rangle - c_1 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle,$$

so that

$$c_1 = \frac{\langle \mathbf{v}_2, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}.$$

To find the next element of the basis set, i.e., $\mathbf{u}_3$, we take

$$\mathbf{u}_3 = \mathbf{v}_3 - d_1\mathbf{u}_1 - d_2\mathbf{u}_2, \tag{2.18}$$

where the unknown coefficients, d_1 and d_2 are chosen so that $\langle \mathbf{u}_3, \mathbf{u}_1 \rangle = 0$, and $\langle \mathbf{u}_3, \mathbf{u}_2 \rangle = 0$. Taking scalar product of (2.18) with respect to $\mathbf{u}_1$ and $\mathbf{u}_2$, and setting the results to zero, and noting that by construction $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0$, we obtain

$$d_1 = \frac{\langle \mathbf{v}_3, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}, \qquad d_2 = \frac{\langle \mathbf{v}_3, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}.$$

The process is extended inductively for the rest of the elements of the basis leading to (2.17). By construction, the sequence $\{\mathbf{u}_i\}$ is clearly orthogonal. To obtain the orthonormal set, what is needed is just normalization of the set $\{\mathbf{u}_i\}$ by dividing each vector by its norm. ▷

Example 2.10

Consider the following vectors in $\mathcal{R}^4$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}.$$

These vectors are linearly independent and form a basis for $\mathcal{R}^4$. Note however that the set $\{\mathbf{v}_i\}$ is not orthogonal. The new orthonormal basis set is constructed using the Gram-Schmidt method, equation (2.17), as

$$\mathbf{u}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_3, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{u}_4 = \mathbf{v}_4 - \frac{\langle \mathbf{v}_4, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_4, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2 - \frac{\langle \mathbf{v}_4, \mathbf{u}_3 \rangle}{\langle \mathbf{u}_3, \mathbf{u}_3 \rangle}\mathbf{u}_3 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 2 \end{bmatrix}.$$

It is also easily verified that $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for $i \neq j$. Thus $\{\mathbf{u}_i\}$ is an orthogonal set. Dividing each vector by its norm, one obtains the orthonormal set:

$$\{\hat{\mathbf{u}}_i\} = \{\frac{1}{\sqrt{3}}\mathbf{u}_1, \mathbf{u}_2, \frac{1}{\sqrt{2}}\mathbf{u}_3, \frac{1}{\sqrt{6}}\mathbf{u}_4\}.$$

To express the vector, $\mathbf{y} = [\,1 \quad 2 \quad -1 \quad 3\,]'$ in terms of the new basis set $\{\mathbf{u}_i\}$, we have

$$\mathbf{y} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + a_4\mathbf{u}_4.$$

Scalar multiplying the above equation by $\mathbf{u}_1$, and noting that
$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0, \langle \mathbf{u}_1, \mathbf{u}_3 \rangle = 0, \langle \mathbf{u}_1, \mathbf{u}_4 \rangle = 0,$, one obtains

$$\langle \mathbf{u}_1, \mathbf{y} \rangle = a_1 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle,$$

so that $a_1 = 1$. The coefficients a_2, a_3, a_4 are computed in a similar way giving $a_2 = 2$, $a_3 = -1$, $a_4 = 1$.

▷

Reciprocal Basis

Expansion of a given vector in $\mathcal{R}^n$ can also be simplified if it is possible to find a set of vectors that are orthonormal to the original basis. For example, suppose $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be a basis for $\mathcal{R}^n$. Consider the vector expansion of a given vector $\mathbf{y} \in \mathcal{R}^n$, i.e.,

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \tag{2.19}$$

where the expansion coefficients, $a_1, a_2, \cdots, a_n$, need to be determined. Suppose there exists a set of vectors $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n\}$ so that

$$\langle \mathbf{v}_i, \mathbf{r}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.20}$$

Then scalar multiplying equation (2.19) with $\mathbf{r}_i$, and using the property (2.20), one immediately obtains $a_i = \langle \mathbf{y}, \mathbf{r}_i \rangle$. Thus the vector expansion problem is greatly simplified. Indeed, the expansion (2.19) then takes the form

$$\mathbf{y} = \sum_{i=1}^{n} \langle \mathbf{y}, \mathbf{r}_i \rangle \mathbf{v}_i. \tag{2.21}$$

The set of vectors, $\{\mathbf{r}_i, i = 1, 2, \cdots n\}$, is called the reciprocal basis or dual basis corresponding to the set $\{\mathbf{v}_i, i = 1, 2, \cdots, n\}$. The reason for the nomenclature 'reciprocal' for the set $\{\mathbf{r}_i\}$ will be apparent after we introduce a method for its construction.

Using the definition (2.20), and writing all equations explicitly for all components of the set in the form of a matrix equation

$$\begin{bmatrix} \langle \mathbf{v}_1, \mathbf{r}_1 \rangle & \langle \mathbf{v}_1, \mathbf{r}_2 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{r}_n \rangle \\ \langle \mathbf{v}_2, \mathbf{r}_1 \rangle & \langle \mathbf{v}_2, \mathbf{r}_2 \rangle & \cdots & \langle \mathbf{v}_2, \mathbf{r}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_n, \mathbf{r}_1 \rangle & \langle \mathbf{v}_n, \mathbf{r}_2 \rangle & \cdots & \langle \mathbf{v}_n, \mathbf{r}_n \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

which can be expressed symbolically as

$$\mathbf{V}'\mathbf{R} = \mathbf{I}, \tag{2.22}$$

where $\mathbf{V} = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,]$, and $\mathbf{R} = [\,\mathbf{r}_1 \quad \mathbf{r}_2 \quad \cdots \quad \mathbf{r}_n\,]$. Thus the reciprocal basis set is obtained simply by inverting the transpose of the matrix of basis vectors as

$$\mathbf{R} = [\mathbf{V}']^{-1} \qquad \text{or equivalently} \qquad \mathbf{R}' = \mathbf{V}^{-1}. \tag{2.23}$$

that is, the rows of the inverse of the matrix of basis set are the vectors of the reciprocal basis. ▷

Although both the orthonormal and the reciprocal bases are computed for the same given basis, a fundamental difference between the two is that *the orthonormal basis vectors are orthonormal to themselves whereas the reciprocal basis vectors are normal to the original set.* In terms of their application for vector expansion, the use of the reciprocal basis set leads to an expansion in terms of the original set of basis vectors whereas for the other case the given vector is expanded in terms of the new orthonormal basis.

Example 2.11

For the set of vectors given in Example 2.10, the reciprocal basis set is constructed using (2.23) as:

$$\mathbf{r}_1 = \begin{bmatrix} -1/2 \\ -1 \\ 3/2 \\ 0 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} 3/2 \\ 1 \\ -3/2 \\ 0 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}, \quad \mathbf{r}_4 = \begin{bmatrix} -1/6 \\ 0 \\ -1/6 \\ 1/3 \end{bmatrix}$$

Given a vector, $\mathbf{y} = [\,1 \quad 2 \quad -1 \quad 3\,]'$, the coefficients for expansion in terms of the original basis set are easily obtained. Specifically, one has

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + a_4\mathbf{v}_4.$$

By virtue of orthogonality of the set $\{\mathbf{v}_i\}$ and the set $\{\mathbf{r}_i\}$, scalar multiplication of the above equation by $\mathbf{r}_1$ leads to

$$\langle \mathbf{y}, \mathbf{r}_1 \rangle = a_1 = -4.$$

The other coefficients of expansion are computed in a similar way as $a_2 = 5, a_3 = -1$, and $a_4 = 1$. ▷

The method easily extends to the case when the number of vectors in the basis set is not equal to n. Suppose $\mathcal{V}$ be a m-dimensional subspace of $\mathcal{R}^n$, with the basis set $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. It is desired to construct a reciprocal basis for $\mathcal{V}$. In light of above discussions, it is clear that one has to solve the equation

$$\mathbf{V}'\mathbf{R} = \mathbf{I}, \tag{2.24}$$

except that the matrix $\mathbf{V}$ is not invertible since it is not a square matrix. Solution of this type of equations will be the subject of further discussion later in this chapter. Nevertheless, equation (2.24) has a solution since the columns of $\mathbf{V}$ are linearly independent. Taking the reciprocal basis as

$$\mathbf{R} = \mathbf{V}[\mathbf{V}'\mathbf{V}]^{-1}, \tag{2.25}$$

it is observed that (2.24) is satisfied, which gives the m-vectors of the reciprocal basis set. ▷

Example 2.12

Let the space $\mathcal{V}$ be defined by the first three vectors, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ of Example 2.10. Then the reciprocal basis is computed using (2.25) as

$$\mathbf{r}_1 = \begin{bmatrix} -2/3 \\ -1 \\ 4/3 \\ 1/3 \end{bmatrix}, \qquad \mathbf{r}_2 = \begin{bmatrix} 3/2 \\ 1 \\ -3/2 \\ 0 \end{bmatrix}, \qquad \mathbf{r}_3 = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$$

Orthogonality relation (2.20) is easily proved using the above result. ▷

2.5 Linear Transformation

Consider two vector spaces $\mathcal{V}$ and $\mathcal{W}$. Then a transformation $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ is a mapping or rule that assigns to every elements $\mathbf{v}$ in $\mathcal{V}$ an elements $\mathbf{T}(\mathbf{v})$ in $\mathcal{W}$. The space $\mathcal{V}$ is called the domain of $\mathbf{T}$, and the space $\mathcal{W}$ is the range.

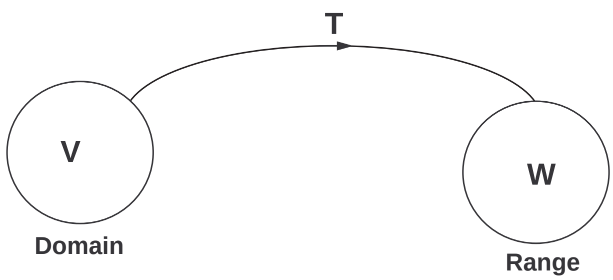


Figure 2.3 A Transformation

For example, define the operator $\mathbf{T}$ by

$$\mathbf{T}\mathbf{x} = \mathbf{A}\mathbf{x}$$

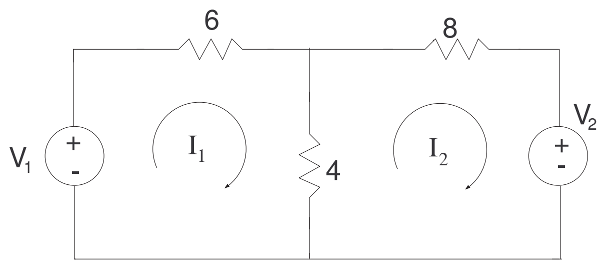
where $\mathbf{A}$ is a 3×2 matrix, and $\mathcal{V} = \mathcal{R}^{2\times 1}$, and $\mathcal{W} = \mathcal{R}^{3\times 1}$. Then for a vector $\mathbf{v} \in \mathcal{V}$, one has $\mathbf{w} = \mathbf{A}\mathbf{v}$, which is a vector in $\mathcal{W}$. Note also that the operator $\mathbf{T}$ defines an operation, rather that a matrix, i.e., $\mathbf{T}$ defines a *multiplication by a matrix* operation.

If a transformation satisfies the superposition property, then it is called a linear transformation, in other words, a linear transformation satisfies

$$\mathbf{T}(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1\mathbf{T}(\mathbf{v}_1) + a_2\mathbf{T}(\mathbf{v}_2),$$

where a_1 and a_2 are scalars and $\mathbf{v}_1, \mathbf{v}_2$ are vectors in $\mathcal{V}$. A matrix with constant entries is a linear transformation; in fact, linearity can be proved very easily. For the following circuit, using Kirchoff's voltage law, we obtain

$$\begin{bmatrix} 10 & -4 \\ -4 & 12 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} V_1 \\ -V_2 \end{bmatrix}$$



Define the transformation by

$$\mathbf{T}\mathbf{V} = \mathbf{M}^{-1}\mathbf{V}$$

where

$$\mathbf{M} = \begin{bmatrix} 10 & -4 \\ -4 & 12 \end{bmatrix}$$

where $\mathbf{M}$ is a 2×2 matrix. Here $\mathbf{T}$ is a more complex operator; more specifically, it the "multiplication by the inverse of a matrix". The elements of the domain space are vectors consisting of source voltages, and the elements of the range space are the mesh currents. One can verify the superposition property defined above are satisfied by $\mathbf{T}$. Since the operation between voltage and current is linear (i.e., the operator $\mathbf{T}$ is linear), in engineering terms, we simply say this circuit is linear.

A transformation is not always linear. For example, define the transformation $\mathbf{T} : \mathcal{R}^2 \to \mathcal{R}$ as

$$\mathbf{T}(x_1, x_2) = x_1^2 + x_2^2$$

Here the operator $\mathbf{T}$ defines the square of the norm of a vector, and it can be verified that $\mathbf{T}$ is not a linear operator. For a single loop circuit consisting of a diode and voltage source, the relation between voltage and current is not linear, since diode conducts in one direction only. This diode circuit defines a nonlinear transformation between the voltage source and the current.

From an engineering perspective, a transformation is a relation between an external input to intrinsic internal variables of the system. In the circuit given above, the transformation $\mathbf{T}$ describes the relation between the applied voltage and the mesh current. For a point mass moving in space, one can find a relation between the applied force and the position and velocity of the mass, which indeed is a differential equation. This is also a transformation, which is a differential operator.

Most physical systems exhibit nonlinearity in their operation, however, in engineering applications, we often restrict system operation within a certain bound within which the system behaves more or less as a linear system. In this context, the concept of linear transformation comes handy for their analysis. Mathematical concepts of nonlinear transformation is beyond the scope of this introductory book.

Domain and Range

The space on which the transformation operates is called its *domain*, and the space in which the transformed elements reside is called its *range*. Suppose $\mathbf{T}$ defines a matrix multiplication operator of dimension 2×3. The $\mathbf{T}$ operates on vectors $\mathbf{v}$ in $\mathcal{R}^2$ so that domain of $\mathbf{T}$, denoted $\mathcal{D}(\mathbf{T}) = \mathcal{R}^2$. After the matrix operation, the product $\mathbf{w} = \mathbf{T}\mathbf{v}$, is a vector in $\mathcal{R}^3$ so that the range of $\mathbf{T}$, denoted $\mathcal{R}(\mathbf{T}) \subset \mathcal{R}^3$. Note that the range of $\mathbf{T}$ need not be all of $\mathcal{R}^3$. This can be seen from the following example:

Consider the matrix

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

The domain of $\mathbf{A}$ is $\mathcal{R}^3$ since $\mathbf{A}$ operates on arbitrary vectors in $\mathcal{R}^3$. As for the range, we have

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x}, \quad \text{for some } \mathbf{x} \in \mathcal{R}^3\}$$

Clearly, the vector $\mathbf{y}$ has the representation

$$\mathbf{y} = x_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

where $\mathbf{x} = [\, x_1 \quad x_2 \quad x_3 \,]'$. But the last column of $\mathbf{A}$ is linearly dependent on its first two columns so that $\mathbf{y}$ can be represented as a linear combination of the first two columns of $\mathbf{A}$. Thus

$$\mathcal{R}(\mathbf{A}) = \text{ span } \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

This shows that for matrices, the range space is the span of linearly independent columns of the matrix. ▷

The range space of a transformation $\mathbf{T}$ signifies the existence of solution of the linear equation $\mathbf{Tx} = \mathbf{b}$. In order for the linear equation to have a solution, the vector $\mathbf{b}$ must belong to the range space of the transformation $\mathbf{T}$.

Rank of a Matrix

The rank of a matrix is defined as the number of linearly independent columns of the matrix. This is also known as the column rank of the matrix, which is same as the row rank or the number of linearly independent rows of the matrix. *The rank of a matrix* $\mathbf{M}$ *can also be defined as the size of the largest matrix with a nonzero determinant that can be formed from* $\mathbf{M}$. Computation of rank of a matrix by finding linearly independent rows or columns is numerically more efficient than finding a submatrix with nonzero determinant. Elementary row operations can be performed to determine linearly independent columns, and hence to determine the rank of a matrix. The MATLAB macro `rank` can also be used to compute the rank of a matrix.

Example 2.13 Consider the matrix given below, we perform elementary row operations to obtain row-reduced matrix as

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

which shows that there are two linearly independent columns in the matrix $\mathbf{M}$. Thus the rank of the matrix $\mathbf{M}$ is 2.

Alternately we find the rank by finding a submatrix of $\mathbf{M}$ with nonzero determinant. Since $|\mathbf{M}| = 0$ so that rank $\mathbf{M} \neq 3$. Now consider all 2×2 matrices that can be formed using the rows and columns of $\mathbf{M}$, for example, consider the following: $\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}$, $\begin{vmatrix} 1 & 3 \\ 2 & 3 \end{vmatrix}$, etc. Since the first submatrix has a nonzero determinant, rank of $\mathbf{M}$ is 2. ▷

Rank of a matrix can be computed even if the matrix $\mathbf{M}$ is not square. Suppose $\mathbf{M}$ is an $m \times n$ matrix. Then the maximum possible rank of $\mathbf{M}$ is $\min\{m, n\}$. In general, for any matrix $\text{rank}(\mathbf{M}) \leq \min\{m, n\}$. If $\mathbf{M}$ is a square matrix, say, $\mathbf{M}$ is an $n \times n$ matrix, and rank $\mathbf{M} = n$, then $\mathbf{M}$ is said to be full rank. ▷

The following property of rank of a matrix comes handy in control system analysis:

> If $\mathbf{C} = \mathbf{AB}$, then $\text{rank}(\mathbf{C}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$. Furthermore, if $\mathbf{A}$ is full rank, then $\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{B})$; likewise, if $\mathbf{B}$ is full rank, then $\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{A})$.

Null Space

The numm space of a matrix $\mathbf{M}$*, denoted* $\mathcal{N}(\mathbf{M})$*, is defined as the set of all vectors* $\mathbf{x}$ *such that* $\mathbf{Mx} = 0$. Formally, suppose $\mathbf{M} \in \mathcal{R}^{m\times n}$, then

$$\mathcal{N}\{\mathbf{M}\} = \{\mathbf{x} \in \mathcal{R}^n \;:\; \mathbf{Mx} = 0\}$$

Given the matrix

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}, \tag{2.26}$$

we verify that for

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

we have

$$\mathbf{Mx} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 0.$$

Thus the vector $\mathbf{x}$ given above is an element of the null space of $\mathbf{M}$. On the other hand, if the given matrix is

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}, \tag{2.27}$$

no nonzero $\mathbf{x}$ can be found for which $\mathbf{Mx} = 0$. In view of the above examples, we may ask the following questions:

a) Does there exist a nontrivial vector $\mathbf{x}$ so that $\mathbf{Mx} = 0$?
b) If there exists a nontrivial vector $\mathbf{x}$ for which $\mathbf{Mx} = 0$, is it unique? Are there more than one vector for which $\mathbf{Mx} = 0$?
c) If there exists a nontrivial vector for which $\mathbf{Mx} = 0$, how do we find it?

We look for answer to these questions in the sequel. For square matrices, if $|\mathbf{M}| \neq 0$, then we have $\mathbf{x} = \mathbf{M}^{-1}0 = 0$ so that the null space $\mathbf{M}$ is empty. *Thus for the null space of* $\mathbf{M}$ *to exist, the rank of* $\mathbf{M}$ *must be less than* n*, where* n *is the dimension of the matrix* $\mathbf{M}$. For the matrix given in (2.26), it is easily verified that rank $\mathbf{M} = 2$ so that there is a non-empty null space whereas the rank of the matrix given in (2.27) is three so that its null space is empty.

In general, *a nonempty null space exists if the column rank of the matrix is less than the number of columns of the matrix.* Suppose $\mathbf{A}$ is an $n \times m$ matrix, where $m > n$, i.e., there are more columns than the rows. Then there will be a nonempty null space if the number of linearly independent columns of $\mathbf{A}$ is less than m. For example, consider

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

This matrix has three columns, but only two of the columns are linearly independent. It is easy to see that for

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

the product $\mathbf{Ax} = 0$ so that $\mathbf{x}$ defined above is an element of the null space of $\mathbf{A}$.

The null space of a matrix may contain more than one element. For example, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Then we verify that for the following vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 0 \\ 2 \\ -1 \end{bmatrix},$$

one has $\mathbf{Ax}_1 = 0$ and $\mathbf{Ax}_2 = 0$, so that there are two vectors in the null space of $\mathbf{A}$. Observe that there are four columns in the matrix $\mathbf{A}$, and the column rank of $\mathbf{A}$ is 2, since only two of the columns are linearly independent. The dimension of the null space of a matrix is also known as *nullity* or *degeneracy* of a matrix. The following holds for any arbitrary matrix:

$$\begin{aligned}
&\text{Rank}(\mathbf{A}) = \text{number of independent columns of } \mathbf{A} \\
&\text{Rank}(\mathbf{A}) = \text{number of independent rows of } \mathbf{A} \\
&\text{Nullity}(\mathbf{A}) = \text{Number of columns of } \mathbf{A} - \text{rank}(\mathbf{A}) \\
&\text{Nullity}(\mathbf{A}') = \text{Number of rows of } \mathbf{A} - \text{rank}(\mathbf{A})
\end{aligned}$$

The following properties hold for any arbitrary matrix $\mathbf{A} \in \mathcal{R}^{m\times n}$:

a) $\text{rank }(\mathbf{A}) = \text{ rank }(\mathbf{A}')$
b) $\text{rank }(\mathbf{A}) \leq \min\{m, n\}$
c) $\text{rank }(\mathbf{A}) + \text{ nullity }(\mathbf{A}') = m$
d) $\text{rank }(\mathbf{A}') + \text{ nullity }(\mathbf{A}) = n$
e) $\text{rank }(\mathbf{A}) = \text{ rank }(\mathbf{AA}') = \text{ rank }(\mathbf{A}'\mathbf{A})$

Computation of the null space can be conveniently carried out using elementary row operations. This is described through an example. Given the matrix $\mathbf{M}$ as above, we have

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 5 \\ 1 & 2 & 4 \\ 0 & 1 & 1 \end{bmatrix} \quad \sim \quad \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The row-reduced matrix is deficient of a '1' in the leading position of the third column so that $\mathbf{M}$ is rank-deficient. This indicates the existence of a nonempty null space for $\mathbf{M}$. Furthermore, it is clearly observed that $\mathbf{m}_3 = 2\mathbf{m}_1 + 1\mathbf{m}_2$, i.e., the third column of $\mathbf{M}$ is 2 times the first column plus the first column. Rewriting the above statement, we have

$$[\,\mathbf{m}_1 \quad \mathbf{m}_2 \quad \mathbf{m}_3\,] \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 0$$

which shows that *the null space vectors are given by rank-deficient columns of the row-reduced matrix with the leading 0 replaced by* -1.

For nonsquare matrices, we augment the given matrix by zero rows to make it a square matrix. Then elementary row operations are carried out to reduce the matrix to the row-reduced echelon form, and null space vectors are obtained from the rank-deficient columns. This can be better explained by the following example. Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix},$$

which is augmented by adding two zero rows to make it a square matrix. Then we perform elementary row operations to obtain the equivalent matrix in row-echelon form:

$$\begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This shows that the rank of the given matrix is 2, and that there must be $4-2=2$ vectors in the null space of $\mathbf{A}$. Note that the leading entries in the second and the fourth rows are zero. The vectors in the null space of $\mathbf{A}$ are given by the second and the fourth columns of row-reduced matrix with their leading entries replaced by -1. Specifically

$$\mathcal{N}(\mathbf{A}) = \text{ Span } \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 0 \\ 2 \\ -1 \end{bmatrix} \right\}$$

How about if there are more rows than columns? How do you find the null space?

The following properties relate the null space and the range space of a matrix $\mathbf{M} \in \mathcal{R}^{m \times n}$:

a) $\mathcal{R}^{\perp}(\mathbf{M}) = \mathcal{N}(\mathbf{M}')$

b) $\mathcal{R}^n = \mathcal{R}(\mathbf{M}) \oplus \mathcal{N}(\mathbf{M}')$

c) $\mathcal{N}(\mathbf{M}'\mathbf{M}) = \mathcal{N}(\mathbf{M})$

d) $\mathcal{R}(\mathbf{M}\mathbf{M}') = \mathcal{R}(\mathbf{M})$

Before we end this section, some concluding remarks. The null space of a transformation is related to the nonuniqueness of solution of the linear equation $\mathbf{Tx} = \mathbf{b}$. If the transformation $\mathbf{T}$ has a nonempty null space, then the linear equation has multiple solutions. If the null space is empty, then the linear equation has a unique solution. We will explore these concepts in the next section.

2.6 General Solution of Linear Equation

An immediate application of the null space concept developed in the previous section is in finding the general solution of a linear equation. Consider the linear equation

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A}$ is an $m \times n$ given matrix, $\mathbf{b}$ is an $m \times 1$ given vector, and m is not necessarily equal to n. The problem is: find, if possible, the unknown n-vector $\mathbf{x}$ for which the above equality holds. Here we have m equations with n unknowns. It turns out that this equation may have a solution only under certain specific conditions. Define the augmented matrix $\mathbf{M} = [\mathbf{A} \mid \mathbf{b}]$. Then we have the following:

a) If rank $\mathbf{M} \neq$ rank $\mathbf{A}$, then there is no solution to the equation, and the equations are inconsistent.

b) If rank $\mathbf{M} =$ rank $\mathbf{A}$, then there is at least one solution. In particular, we have
   - if rank $\mathbf{M} =$ rank $\mathbf{A} = n$, then the solution is unique.
   - if rank $\mathbf{M} =$ rank $\mathbf{A} = p < n$, then there are infinitely many solutions which have $n - p$ free parameters.

This can be seen from the following: Let $\mathbf{A} = [\,\mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \mathbf{A}_n\,]$, where $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n$ are the columns of the matrix $\mathbf{A}$. Then we have

$$\mathbf{A}_1 x_1 + \mathbf{A}_2 x_2 + \cdots + \mathbf{A}_n x_n = \mathbf{b}. \tag{2.28}$$

This shows that if the equation $\mathbf{Ax} = \mathbf{b}$ has a solution, then the vector $\mathbf{b}$ must be linearly dependent on the vectors $\mathbf{A}_1, \mathbf{A}_2, \cdots \mathbf{A}_n$. Hence rank $\mathbf{A} =$ rank $\mathbf{M}$ must hold for a solution to exist.

If rank $\mathbf{A} = p$, it implies that only p-columns of the matrix $\mathbf{A}$ are linearly independent, and therefore all columns of $\mathbf{A}$ except the p independent columns can be removed from the above equation (2.28). Thus if rank $\mathbf{A} =$ rank $\mathbf{M} = p$, only p components of the unknown vector $\mathbf{x}$ can be computed using (2.28), and the remaining unknowns can be arbitrarily selected.

Example 2.14 Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Then we have rank $\mathbf{A} = 2$ and rank $\mathbf{M} = 3$ so that there is no solution to the equation $\mathbf{Ax} = \mathbf{b}$. ▷

Example 2.15

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

In this case we have rank $\mathbf{A} =$ rank $\mathbf{M} = 3$ so that there is a unique solution. ▷

Example 2.16 Consider the following example

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

For this example we have rank $\mathbf{A} =$ rank $\mathbf{M} = 2 < 3$ so that there are infinitely many solutions. We also see that there are $3 - 2 = 1$ free parameter that can be arbitrarily chosen to construct these infinitely many solutions. In fact the solution of the above equation is given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix},$$

where α is an arbitrary number. ▷

This last example requires special attention. In this problem we have two equations with three unknowns, and the equation has infinitely many solutions. In what follows we present a systematic method for computing the solution of systems of equations for which there are infinitely many solutions. The method is based on row reduced echelon form of the augmented matrix.

We augment the matrix $\mathbf{M} = [\mathbf{A} \mid \mathbf{b}]$ by adding zero rows so as to make it a square matrix, and carry out elementary row operations as usual. The complete solution of the system of equation is then constructed by adding the particular solution, (which is the last column of transformed $\mathbf{M}$) and the null solution of $\mathbf{M}$. The method is illustrated by the following example:

Considering the problem in Example 2.16, we form

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Now we add a zero row to $\mathbf{M}$ so as to make it a 3×4 matrix. The number of rows must be equal to the number of unknowns. Perform elementary row operations as shown below:

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

This gives us the particular solution as $[\,1 \quad 0 \quad 2\,]'$ and a null space $[\,0 \quad -1 \quad 0\,]'$ (obtained from the second column). The complete solution is then obtained by adding the two solutions as shown in Example 2.16.

Note that the row reduced matrix does not have the structure of"row echelon" form that is frequently used in many applications. A row echelon matrix has all zero rows at the bottom, which is clearly not the case above. Our objective is to get as many 1's as possible in the diagonal position. This approach easily identifies null space of a matrix, and hence the homogeneous solution, and give the the particular solution.

The approach given above is based on the fact that the solution of a linear system can always be expressed as the sum of two components: a particular solution, and a homogeneous or null solution. Consider the equation

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

and let $\mathbf{x}_p$ is the particular solution, and $\mathbf{x}_n$ is the homogeneous solution satisfying

$$\begin{aligned} \mathbf{A}\mathbf{x}_p &= \mathbf{b}, \\ \mathbf{A}\mathbf{x}_n &= 0. \end{aligned}$$

Then adding the two equations, we have

$$\mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) = \mathbf{b}.$$

so that $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$ is the general solution of the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. This gives us a way of finding the solution of linear equations in terms of its particular solution and the null space of the coefficient matrix. Consider the following example:

Example 2.17

$$\begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

We verify that the rank $\mathbf{A} = \text{rank}\,[\mathbf{A} \mid \mathbf{b}] = 2 < 4$ so that there are infinitely many solutions that have two $4 - 2 = 2$ free parameters. We perform the row operations on the augmented matrix as follows:

$$\begin{bmatrix} 1 & -2 & 0 & 1 & 5 \\ 0 & 0 & 1 & 2 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This shows that the particular solution of the equation is given by

$$\mathbf{x}_p = \begin{bmatrix} 5 \\ 0 \\ 6 \\ 0 \end{bmatrix}.$$

We also obtain the null space vectors from the second and the fourth columns as

$$\mathbf{x}_n = \text{ Span } \left\{ \begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix} \right\}.$$

Using the particular solution and the null space vectors, we have the complete solution given by

$$\mathbf{x} = \begin{bmatrix} 5 \\ 0 \\ 6 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix},$$

where α and β are arbitrary scalars. ▷

Minimum Norm Solution

As shown above, certain linear systems have infinitely many solutions. For example, consider the system $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A} \in R^{m \times n}$ and $\mathbf{x} \in R^n$, and $m < n$. This system has infinitely many solutions. In that case, the question is: out of the infinitely many solutions which one should be selected. Theoretically speaking all these infinitely many solutions are indeed valid solutions. However, in many engineering applications it is a common practice to choose the solution that has the minimum norm. In this section we discuss the method of computation of the minimum norm solution.

Fundamentally, what we have is a constrained minimization problem. Specifically, we minimize the norm of the solution subject to the condition $\mathbf{Ax} = \mathbf{b}$. In other words,

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|\mathbf{x}\|^2, \\ \text{subject to} \quad & \mathbf{Ax} - \mathbf{b} = 0. \end{aligned} \tag{2.29}$$

The solution of this minimization problem is obtained using the Kuhn-Tucker theorem [12, 36]. In this respect we have the following result:

Theorem 2.2. *If the continuously differentiable function $f(\mathbf{x})$ has a local minimum under the set of constraints $g(\mathbf{x}) = 0$ at the point $\mathbf{x}_0$, then there exists a vector Λ such that the Lagrangian function*

$$L(\mathbf{x}) = f(\mathbf{x}) + \langle \Lambda, g(\mathbf{x}) \rangle \tag{2.30}$$

is stationary at the point $\mathbf{x}_0$. *In other words, at the point* $\mathbf{x}_0$, *we have*

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \left[\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}\right]' \Lambda = 0. \tag{2.31}$$

$\triangleright$

Before we use this theorem for finding the minimum norm solution of linear equations, let's first consider a simple example.

Minimize

$$f(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 2)^2$$

subject to

$$x_1^2 + x_2^2 = 1$$

Clearly, $x_1 = 1$, $x_2 = 2$ is not the solution of this minimization problem since that will violate the equality condition. To apply the above theorem, let's form the Lagrangian

$$L(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 2)^2 + \lambda(x_1^2 + x_2^2 - 1)$$

and minimize it with respect to x_1, x_2 and λ. This is essentially an unconstrained minimization problem, and the solution is obtained by solving

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 2(x_1 - 1) + 2\lambda x_1 = 0 \\
\frac{\partial L}{\partial x_2} &= 2(x_2 - 2) + 2\lambda x_2 = 0 \\
\frac{\partial L}{\partial \lambda} &= x_1^2 + x_2^2 - 1 = 0
\end{aligned}$$

Solving these equations, we obtain $x_1 = \frac{1}{\sqrt{5}}$, $x_2 = \frac{2}{\sqrt{5}}$, and $\lambda = \sqrt{5} - 1$. This completes the solution of the above minimization problem.

Now we return to the minimum norm solution problem for the linear equation. The vector Λ is known as the Lagrangian multiplier, and the Lagrangian function for the minimization problem (2.29) is given by

$$L(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 + \langle \Lambda, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle \tag{2.32}$$

Differentiating the above equation with respect to $\mathbf{x}$ and Λ, and equating the results to zero, we obtain

$$\begin{aligned}
\frac{d}{d\mathbf{x}} L(\mathbf{x}) = 0 &\Longrightarrow \mathbf{x} + \mathbf{A}'\Lambda = 0 \\
\frac{d}{d\Lambda} L(\mathbf{x}) = 0 &\Longrightarrow \mathbf{A}\mathbf{x} - \mathbf{b} = 0.
\end{aligned}$$

Solving the first equation for $\mathbf{x}$, which is $\mathbf{x} = -\mathbf{A}'\Lambda$ and substituting it into the second equation, we obtain

$$-\mathbf{A}\mathbf{A}'\Lambda - \mathbf{b} = 0$$

which gives

$$\Lambda = -[\mathbf{A}\mathbf{A}']^{-1}\mathbf{b}.$$

Thus the solution of the minimization problem is

$$\begin{aligned} \mathbf{x} &= -\mathbf{A}'\Lambda \\ &= \mathbf{A}'[\mathbf{A}\mathbf{A}']^{-1}\mathbf{b}. \end{aligned} \tag{2.33}$$

The matrix $\mathbf{A}^{\dagger} = \mathbf{A}'[\mathbf{A}\mathbf{A}']^{-1}$ is known as the right pseudo inverse of the matrix $\mathbf{A}$ since multiplying the matrix $\mathbf{A}$ by $\mathbf{A}^{\dagger}$ on the right side results in the identity matrix. Note however that $\mathbf{A}^{\dagger}\mathbf{A} \neq \mathbf{I}$ so that $\mathbf{A}^{\dagger}$ does not have all the properties of an inverse.

As an example, consider the problem given in Example 2.17. Then we compute

$$\begin{aligned} \mathbf{A}^{\dagger} &= \mathbf{A}'[\mathbf{A}\mathbf{A}']^{-1} \\ &= \begin{bmatrix} 0.1923 & -0.0769 \\ -0.3846 & 0.1538 \\ -0.0769 & 0.2308 \\ 0.0385 & 0.3846 \end{bmatrix}, \end{aligned}$$

and the solution of the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ corresponding to the minimum norm is

$$\begin{aligned} \mathbf{x} &= \mathbf{A}^{\dagger}\mathbf{b} \\ &= \begin{bmatrix} 0.5 \\ -1.0 \\ 1.0 \\ 2.5 \end{bmatrix} \end{aligned}$$

It can be verified that $\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{I}$, but $\mathbf{A}^{\dagger}\mathbf{A} \neq \mathbf{I}$.

Certain linear equations with square matrices have infinitely many solutions. Is it possible to find the minimum norm solution for this type of problems? Why or why not?

Minimum Error Solution

As discussed earlier, given the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, we know that if rank$([\mathbf{A} \mid \mathbf{b}]) \neq$ rank$(\mathbf{A})$, then there is no solution to the equation. However, in certain applications it is necessary to find an $\mathbf{x}$ that comes closest to satisfying the given equation. This is discussed in the following:

Define the error vector

$$\mathbf{E} = \mathbf{A}\mathbf{x} - \mathbf{b}. \tag{2.34}$$

By assumption the system is inconsistent so that the error $\mathbf{E}$ will be a nonzero vector. The minimum error solution corresponds to an $\mathbf{x}$ for which this error is minimum.

To find this minimum error solution, we minimize the error, i.e.,

$$\begin{aligned}\text{Minimize} \quad \|\mathbf{E}\|^2 &= \|\mathbf{Ax} - \mathbf{b}\|^2 \\ &= \langle \mathbf{Ax} - \mathbf{b}, \mathbf{Ax} - \mathbf{b} \rangle.\end{aligned}$$

Differentiating the above equation with respect to $\mathbf{x}$ and setting it to zero, we obtain

$$2\mathbf{A}'\mathbf{Ax} - 2\mathbf{A}'\mathbf{b} = 0,$$

from which we obtain

$$\mathbf{x} = [\mathbf{A}'\mathbf{A}]^{-1}\mathbf{A}'\mathbf{b}. \tag{2.35}$$

Since the product $\mathbf{A}^{\ddagger}\mathbf{A} = \mathbf{I}$, the matrix $\mathbf{A}^{\ddagger}$ is known as the left pseudo inverse of $\mathbf{A}$. As an example, consider the problem given in Example 2.14. Then we compute

$$\mathbf{A}^{\ddagger} = [\mathbf{A}'\mathbf{A}]^{-1}\mathbf{A}' = \begin{bmatrix} 0.8333 & -0.3333 & 0.1667 \\ -0.3333 & 0.3333 & 0.3333 \end{bmatrix},$$

and the solution corresponding to the smallest (norm) error is

$$\mathbf{x} = \mathbf{A}^{\ddagger}\mathbf{b} = \begin{bmatrix} 0.6667 \\ 1.3333 \end{bmatrix}.$$

It can be easily seen that for this example, the equation $\mathbf{Ax} = \mathbf{b}$ is not satisfied for the computed solution, but the vector given above is the best as far as minimizing the error is concerned.

Least Square Curve Fit

One of the frequent applications of Minimum error solution or the least square solution can be found in curve fitting of a polynomial through observation or experimental data. Suppose experimental data collected from a scientific experiment be denoted by $\{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$. It is also assumed that these data points can be described by a polynomial

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{2.36}$$

where $n < m$, i.e., the number of data points is greater than the degree of the polynomial. In this polynomial the coefficients $a_0, a_1, a_2, \cdots, a_n$ must be determined so as to minimize the error between the experimental data and the data predicted by the polynomial. Since the polynomial (2.36) is expected to satisfy all observation data, we write

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_n x_1^n \\ y_2 &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_n x_2^n \\ &\vdots \qquad \vdots \\ y_m &= a_0 + a_1 x_m + a_2 x_m^2 + \cdots + a_n x_m^n, \end{aligned}$$

which can be expressed in the matrix form as

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}. \tag{2.37}$$

Clearly we have a matrix equation that has more equations than the number of unknowns. Furthermore, it is not expected that all data points will lie on the assumed curve so that not all equations are expected to be simultaneously satisfied. A minimum error solution can be computed using the result (2.35) discussed above. In fact using (2.37), (2.35) can be rewritten as

$$\begin{bmatrix} \sum_i 1 & \sum_i x_i & \sum_i x_i^2 & \cdots & \sum_i x_i^n \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 & \cdots & \sum_i x_i^{n+1} \\ \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^4 & \cdots & \sum_i x^{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_i x_i^n & \sum_i x_i^{n+1} & \sum_i x_i^{n+2} & \cdots & \sum_i x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \sum_i x_i^2 y_i \\ \vdots \\ \sum_i x_i^n y_i \end{bmatrix}$$

where the summation is carried over $i = 1, 2, \cdots m$. This equation can be easily solved for the unknown coefficients. Since the method entails minimizing the least squared error, it is known as the *least square curve fitting* in scientific and engineering literature. The MATLAB macro `polyfit` is available for least square curve fitting through observation data.

Example 2.18

The following data set has been collected in a scientific experiment, and the analyst decided to express the finding of the experiment in terms of a second degree polynomial

$$y = a_0 + a_1 x + a_2 x^2.$$

x	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
y	0.59	0.85	1.04	1.12	1.15	1.34	1.45	1.36	1.50	1.42	1.40

The coefficients of the unknown polynomial are calculated using (2.35) with the matrices obtained from (2.37). The best estimate of the coefficients of the polynomial that fits the experimental data points is obtained as

$$y = 0.6280 + 1.0113x - 0.3120x^2.$$

Figure 2.4 compares the experimental data set with its smooth polynomial approximation.

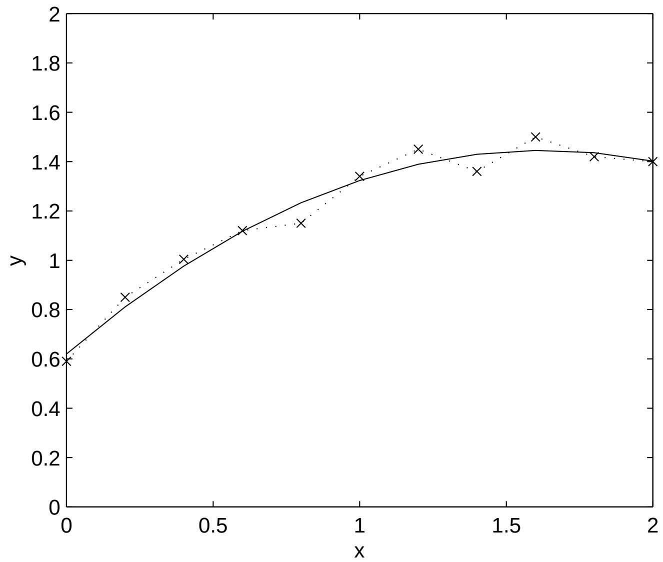


Fig 2.4 Least Square Curve Fit

2.7 Projection

Suppose $\mathcal{W}$ be a vector space, and $\mathcal{W} \subset \mathcal{V}$. Consider a vector $\mathbf{v} \in \mathcal{V}$ but $\mathbf{v} \notin \mathcal{W}$. Then one may ask the question: Which vector $\mathbf{w} \in \mathcal{W}$ is closest to $\mathbf{v}$ in the least square sense? The best approximation of the vector $\mathbf{v}$, denoted $\mathbf{w}$, $\mathbf{w} \in \mathcal{W}$, is called the projection of $\mathbf{v}$ onto $\mathcal{W}$.

The projection $\mathbf{w}$ can be easily calculated by minimizing the square of distance between the two vectors. We can express the projection $\mathbf{w}$ as

$$\mathbf{w} = \mathbf{W}\mathbf{a}, \tag{2.38}$$

where $\mathbf{W} = [\,\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_n\,]$ is the matrix formed by the basis vectors of $\mathcal{W}$, and $\mathbf{a}$ is the vector of expansion coefficients. These expansion coefficients are computed by minimizing the distance between $\mathbf{v}$ and $\mathbf{w}$ by setting

$$\frac{d}{d\mathbf{a}}\|\mathbf{W}\mathbf{a} - \mathbf{v}\|^2 = 0.$$

By direct differentiation with respect to $\mathbf{a}$, one obtains

$$\mathbf{W}'(\mathbf{W}\mathbf{a} - \mathbf{v}) = 0,$$

so that

$$\mathbf{a} = \left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\mathbf{v}. \tag{2.39}$$

The projection of $\mathbf{v}$ onto $\mathbf{W}$ is then obtained as

$$\mathbf{w} = \mathbf{W}\big(\mathbf{W}'\mathbf{W}\big)^{-1}\mathbf{W}'\mathbf{v}. \tag{2.40}$$

In case the basis vectors of $\mathcal{W}$ are orthogonal, the above equation simplifies to

$$\mathbf{w} = \langle \mathbf{v}, \mathbf{w}_1 \rangle \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|^2} + \langle \mathbf{v}, \mathbf{w}_2 \rangle \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|^2} + \cdots + \langle \mathbf{v}, \mathbf{w}_n \rangle \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|^2} \tag{2.41}$$

The proof follows from direct expansion of products in equation (2.40) and the orthogonality assumption of $\mathbf{W}$. ▷

The concept of projection can be better understood by considering the two dimensional space $\mathcal{R}^2$. In light of notations introduced above, we take $\mathcal{V} = \mathcal{R}^2$ and $\mathcal{W} = \{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$, the real line passing through the origin with a slope of 45°, as shown in Figure 2.5. Clearly, $\mathcal{W} \subset \mathcal{V}$. Now consider the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathcal{V}$, which is not in $\mathcal{W}$. Then we pose the following question: which vector in $\mathcal{W}$ is closest to $\mathbf{v}$? For this setup, we have

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

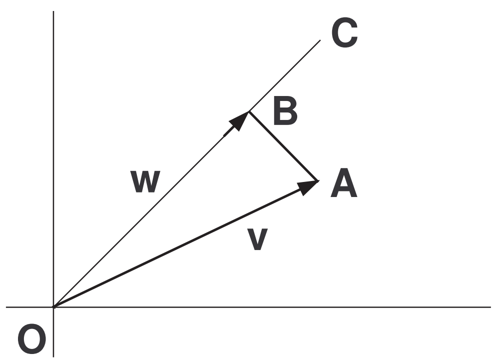


Figure 2.5 Projection

Substituting these matrices into (2.40), one obtains

$$\mathbf{w} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix},$$

which is the best approximation of $\mathbf{v}$ in $\mathbf{W}$ in the least square sense. Indeed, the result can also be verified by a simple trigonometric construction as shown in the figure. The projection of the vector $\vec{OA}$ onto the vector $\vec{OC}$ is the vector $\vec{OB}$. It is also noted that the vector $\vec{BA}$ is perpendicular to the vector $\vec{OB}$, i.e., the vector $\mathbf{v} - \mathbf{w}$ is orthogonal to all vectors of the space $\mathbf{W}$.

Example 2.19

In $\mathbf{V} = \mathcal{R}^4$, consider the subspace $\mathbf{W}$ spanned by the vectors,

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{w}_3 = \begin{bmatrix} 0 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

and the vector $\mathbf{v} = [\,2 \quad 1 \quad 3 \quad 4\,]'$. It is easily verified that $\mathbf{v}$ is linearly independent of the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$, so that $\mathbf{v} \notin \mathbf{W}$. The projection of $\mathbf{v}$ onto $\mathbf{W}$ is computed using (2.40) as

$$\mathbf{w} = \begin{bmatrix} 2.5 \\ 1.0 \\ 3.5 \\ 3.0 \end{bmatrix} \qquad \triangleright$$

A generalization of the orthogonal projection can be stated as follows: *For every vector $\mathbf{v} \in \mathcal{V}$, there exists an element $\mathbf{w} \in \mathcal{W}$, called the projection of $\mathcal{V}$ onto $\mathcal{W}$, such that*

$$\langle \mathbf{v} - \mathbf{w}, \mathbf{p} \rangle = 0 \tag{2.42}$$

where $\mathbf{p}$ is an arbitrary element of $\mathcal{W}$. The set of all vectors, $\mathbf{v} - \mathbf{w}$ is orthogonal to all elements of $\mathcal{W}$ (since it is orthogonal to each of $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$), and is called the **orthogonal complement** of $\mathcal{W}$, denoted $\mathcal{W}^\perp$. Then

$$\mathcal{W} \oplus \mathcal{W}^\perp = \mathcal{V}$$

In general, let $\mathcal{W}$ be a subspace of a vector space $\mathcal{V}$. Then its orthogonal complement, $\mathcal{W}^\perp$, is defined by

$$\mathcal{W}^\perp = \{x \in \mathcal{V} : \langle x, y \rangle = 0 \text{ for all } y \in \mathcal{W}\}$$

in other words, the orthogonal complement of $\mathcal{W} \subset \mathcal{V}$ is the set of all vectors in $\mathcal{V}$ that are orthogonal to every element in $\mathcal{W}$. Suppose $\mathcal{V} = \mathcal{R}^2$, and $\mathcal{W}$ as the line defined by $\mathcal{W} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Then if we take $\mathbf{x} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$, we have $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ so that the orthogonal complement of $\mathcal{W}$ is $\mathcal{W}^{\perp} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$, i.e., any multiple of this vector is an element of $\mathcal{W}^{\perp}$.

In addition, from the previous example, observe that the projection $\mathbf{w}$ of $\mathbf{v}$ onto $\mathbf{W}$ is the vector $\mathbf{w} = [\,2.5 \quad 1.0 \quad 3.5 \quad 3.0\,]'$. Furthermore, the vector $\mathbf{u} = \mathbf{v} - \mathbf{w} = [\,-0.5 \quad 0.0 \quad -0.5 \quad 1.0\,]'$ is orthogonal to arbitrary element in $\mathcal{W}$, and hence is an element of $\mathcal{W}^{\perp}$. It is also easily verified that the vectors $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{u}\}$ span the entire $\mathcal{R}^4$.

The Gram-Schmidt orthogonalization process could also be interpreted in the context of projection. For the Euclidean space $\mathcal{R}^n$, suppose using the Gram-Schmidt procedure, we have already computed $n-1$ orthogonal vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-1}$. Then using (2.17), we compute the next orthogonal vector $\mathbf{u}_n$ as

$$\mathbf{u}_n = \mathbf{v}_n - \sum_{j=1}^{n-1} \frac{\langle \mathbf{u}_j, \mathbf{v}_n \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j$$

which we rewrite as

$$\mathbf{v}_n = (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \cdots + c_{n-1} \mathbf{u}_{n-1}) + \mathbf{u}_n \tag{2.43}$$

where

$$c_j = \frac{\langle \mathbf{u}_j, \mathbf{v}_n \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle}$$

Denote $\mathcal{V} = \mathcal{R}^n$, and a subset $\mathcal{W} = \text{ Span } \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-1}\}$. Then the vector $\mathbf{v}_n \in \mathcal{V}$ is decomposed into two components; the sum $\sum_j^{n-1} c_j \mathbf{u}_j$ represents the projection of $\mathbf{v}_n$ onto the subspace $\mathcal{W}$, and $\mathbf{u}_n$ is an element in the orthogonal complement of $\mathcal{W}$; note that $\mathbf{u}_n$ is orthogonal to each of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-1}$. Recall that elements of the orthogonal complement of $\mathcal{W}$ are orthogonal to each element of $\mathcal{W}$.

From an engineering perspective, projection is an important concept that is frequently used in control system design. The desired performance of control systems is often described in terms of *dynamic modes* of the system. Mathematically these dynamic modes are essentially certain vectors characteristic to the dynamic response of the system. In case a desired mode is not achievable, the designer wishes to find a controller that achieves the best approximation of the desired mode, or the projection of the desired mode onto the achievable subspace. This gives the closed loop performance closest to what the engineer wishes to achieve. This will be the subject of further discussion in Chapter 11.

2.8 Examples

Example 2.20 Area of a rectangle

In $\mathcal{R}^2$, a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ can be visualized as a line segment joining the origin to the point (u_1, u_2). Thus one can form a parallelogram with the vectors, $\mathbf{u}, \mathbf{v}$, as two of its sides. Then if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, one actually gets a rectangle. Since the area of the rectangle is the product of the length and height of the rectangle, we have

$$\begin{aligned}
\text{Area}^2 &= \|\mathbf{u}\|^2 \, \|\mathbf{v}\|^2 \\
&= \left| \begin{bmatrix} \mathbf{u}' \\ \mathbf{v}' \end{bmatrix} [\,\mathbf{u} \quad \mathbf{v}\,] \right|, \qquad \text{since} \quad \mathbf{u}'\mathbf{v} = \mathbf{v}'\mathbf{u} = 0 \\
&= \begin{vmatrix} \mathbf{u}' \\ \mathbf{v}' \end{vmatrix} |\,\mathbf{u} \quad \mathbf{v}\,|, \qquad \text{since} \quad |\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}| \\
&= |\,\mathbf{u} \quad \mathbf{v}\,|^2,
\end{aligned}$$

which shows that

$$\text{Area of a rectangle} = |\,\mathbf{u} \quad \mathbf{v}\,|\,.$$

The result also holds for parallelogram when the vectors $\mathbf{u}$ and $\mathbf{v}$ are not orthogonal, however, the proof is somewhat more complex.

Example 2.21 Distance between a vector and a vector space

The concept of metric presented in Section 2.1 can be extended to define the distance between a vector and a vector space. Let $\mathcal{V} \subset \mathcal{R}^n$ be a vector space defined by

$$\mathcal{V} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\},$$

and let $\mathbf{y}$ be a vector in $\mathcal{R}^n$ such that $\mathbf{y} \notin \mathcal{V}$. The distance between the vector space $\mathcal{V}$ and the vector $\mathbf{y}$ is defined as

$$d(\mathbf{y}, \mathcal{V}) = \min_{\mathbf{v} \in \mathcal{V}} d(\mathbf{y}, \mathbf{v}), \tag{2.44}$$

in other words, *the distance between a vector and a vector space is the minimum distance between the given vector and an arbitrary vector in the vector space.* The following is an example. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

An arbitrary vector in $\mathcal{V}$ is described in terms of its basis as

$$\begin{aligned}
\mathbf{v} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 \\
&= [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \\
&= \mathbf{Va},
\end{aligned}$$

where the unknown coefficients a_1, a_2, a_3 are computed by minimizing the distance between the vector $\mathbf{v}$ and $\mathbf{y}$. For the minimum, we clearly have

$$\frac{d}{d\mathbf{a}}\|\mathbf{Va} - \mathbf{y}\|^2 = 0.$$

Carrying the differentiation, we obtain

$$2\mathbf{V}'\mathbf{Va} - 2\mathbf{V}'\mathbf{y} = 0,$$

so that

$$\mathbf{a} = [\mathbf{V}'\mathbf{V}]^{-1}\mathbf{V}'\mathbf{y},$$

Using the vectors given above, we then obtain the unknown coefficients as

$$\mathbf{a} = \begin{bmatrix} -0.2857 \\ 0.5714 \\ 0.4286 \end{bmatrix},$$

and the corresponding vector $\mathbf{v} \in \mathcal{V}$ which is nearest to $\mathbf{y}$ is

$$\begin{aligned} \mathbf{v} &= \mathbf{Va} \\ &= \begin{bmatrix} 1.2857 \\ 1.0000 \\ 0.5714 \\ 0.8571 \end{bmatrix}. \end{aligned}$$

Then the distance $\mathbf{y}$ and $\mathcal{V}$ is

$$d(\mathbf{y}, \mathcal{V}) = \|\mathbf{y} - \mathbf{v}\| = 0.5345.$$

$\triangleright$

Similar approach could be used to define the distance between two vector spaces. Let $\mathcal{V}$ and $\mathcal{W}$ be two subspaces of $\mathcal{R}^n$. Then the distance between the $\mathcal{V}$ and $\mathcal{W}$ is defined as

$$d(\mathcal{V}, \mathcal{W}) = \min_{\mathbf{v}\in\mathcal{V}} \min_{\mathbf{w}\in\mathcal{W}} d(\mathbf{v}, \mathbf{w})$$

in other words, distance between two vector spaces is the smallest distance between two arbitrary vectors contained in them.

Example 2.22 Intersection between two vector spaces

Consider two vector spaces defined by

$$\mathcal{M} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}, \qquad \mathcal{N} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\}$$

First we verify that the distance between the two spaces is zero, that is, the two spaces overlap to a certain extent. We would like to find the intersection of the two spaces.

An arbitrary element in $\mathcal{M}$ is defined by

$$\mathbf{m} = a_1\mathbf{m}_1 + a_2\mathbf{m}_2 = \mathbf{Ma},$$

where $\mathbf{m}_1$ and $\mathbf{m}_2$ are the two vectors of $\mathcal{M}$, and a_1 and a_2 are arbitrary scalars. Likewise an arbitrary element in $\mathcal{N}$ is defined as

$$\mathbf{n} = b_1\mathbf{n}_1 + b_2\mathbf{n}_2 = \mathbf{Nb}.$$

If there is an intersection between $\mathcal{M}$ and $\mathcal{N}$, there would exist an element $\mathbf{z}$ for which the above two equations will hold so that

$$\mathbf{z} = \mathbf{Ma} = \mathbf{Nb}$$

Thus

$$[\,\mathbf{M} \quad -\mathbf{N}\,] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = 0$$

so that

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \in \text{null}([\,\mathbf{M} \quad -\mathbf{N}\,]).$$

Finding the null space of $[\,\mathbf{M} \quad -\mathbf{N}\,]$, we obtain

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ \cdots \\ 1 \\ 1 \end{bmatrix}$$

so that

$$\mathbf{z} = \mathbf{Ma} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

As a final note, it is easily verified that $\mathbf{z} \in \mathcal{M}$ and $\mathbf{z} \in \mathcal{N}$ as expected so that

$$\mathcal{M} \cap \mathcal{N} = \{\mathbf{z}\}.$$

Furthermore, for this example $\mathcal{M} \cap \mathcal{N}$ happens to contain only a single element.

▷

Example 2.23

Given the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ as in the Example 2.21, find a vector which is orthonormal to all three given vectors.

Let $\mathbf{y}$ be the vector that is orthogonal to $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$. Then by definition of orthogonality, we have

$$\begin{aligned}\langle \mathbf{v}_1, \mathbf{y}\rangle &= 0\\ \langle \mathbf{v}_2, \mathbf{y}\rangle &= 0\\ \langle \mathbf{v}_3, \mathbf{y}\rangle &= 0,\end{aligned}$$

or equivalently

$$\mathbf{V}'\mathbf{y} = 0,$$

where $\mathbf{V} = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,]$. This shows that $\mathbf{y}$ is an element of the null space of $\mathbf{V}'$. Computing the null space of $\mathbf{V}'$, we obtain

$$\mathbf{y} = \begin{bmatrix} 2 \\ 0 \\ -3 \\ -1 \end{bmatrix}.$$

Orthogonality of the vector $\mathbf{y}$ with respect to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is easily verified. To find the orthonormal vector, we just normalize the vector $\mathbf{y}$ as

$$\hat{\mathbf{y}} = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 0 \\ -3 \\ -1 \end{bmatrix}.$$

$\triangleright$

Example 2.24

Consider the matrix equation

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

In this problem we have four equations with three unknowns. We first verify the rank condition for the existence of a solution. In this case we have $\text{rank}([\mathbf{A} \mid \mathbf{b}]) = \text{rank}(\mathbf{A}) = 2 < 3$(number of unknowns) so that there is a solution, in fact, there are infinitely many solutions. To find the solution, we row-reduce the augmented matrix, and find the homogeneous solution and the particular solution. Specifically,

$$\text{Augmented matrix} = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 0 & 3 \\ 2 & 1 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the last two rows are identically zero, the equations are consistent. The particular solution and the homogeneous solutions are obtained as

$$\mathbf{x}_p = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_h = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

The homogeneous solution is obtained from the third column of the row-reduced augmented matrix. Note that the fourth row of the row-reduced augmented matrix is discarded altogether since this equation is identically zero and does not convey any information, and that there are three unknown variables. The complete solution is then obtained as

$$\mathbf{x} = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

where α is arbitrary.

Example 2.25

Given the system

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$$

it is easy to verify that there are infinitely many solutions. Simple analysis then gives the complete solution as

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

If one is interested to find the minimum norm solution, a simple approach would be to set

$$\frac{d}{d\alpha}(\|\mathbf{x}_p + \alpha \mathbf{x}_h\|^2) = 0$$

which gives

$$\alpha = -\frac{4}{3}$$

and the corresponding minimum norm solution as

$$\mathbf{x} = \begin{bmatrix} \frac{5}{3} \\ -\frac{1}{3} \\ \frac{4}{3} \end{bmatrix}$$

2.9 References

1. Gantmacher, F.R., *Theory of Matrices*, Chelsea, 1959.
2. Gloub, Gene and C. F. van Loan, *Matrix Computation*, Johns Hopkins Press, 1989.
3. Strang, G., *Linear Algebra and its Applications*, Academic Press, 1980.
4. Halmos, P. R., *Finite Dimensional Vector Spaces*, van Nostrand, 1958.
5. Kailath T., *Linear Systems*, Prentice Hall, Englewood Cliffs, NJ, 1980.
6. Brogan, W., *Modern Control Theory*, Prentice Hall, 1991.
7. Bellman, R., *Introduction to Matrix Analysis*, McGraw Hill, New York, 1960.

2.10 Exercise

2.1 Verify that the set of all 2×2 matrices with standard addition and scalar multiplication definitions is a vector space.

2.2 Show that if $\mathbf{x}, \mathbf{y}, \mathbf{p}, \mathbf{q}$ be four vectors of $\mathcal{R}^n$, then

$$\langle \mathbf{x} + \mathbf{y}, \mathbf{p} + \mathbf{q} \rangle = \langle \mathbf{x}, \mathbf{p} \rangle + \langle \mathbf{x}, \mathbf{q} \rangle + \langle \mathbf{y}, \mathbf{p} \rangle + \langle \mathbf{y}, \mathbf{q} \rangle$$

2.3 Suppose $\mathbf{x}, \mathbf{y}$ be two vectors in $\mathcal{R}^n$ and $\mathbf{A}, \mathbf{B}$ be two matrices of dimension $m \times n$. Using the property of the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle$, show that $\langle \mathbf{Ax}, \mathbf{By} \rangle = \langle \mathbf{x}, \mathbf{A}'\mathbf{By} \rangle = \langle \mathbf{B}'\mathbf{Ax}, \mathbf{y} \rangle$

2.4 In $\mathcal{R}^2$, Pythagorean theorem states that *the square of the hypotenuse of a right angle triangle is equal to the sum of the squares of the other two sides.* Using the concepts of linear algebra, prove this statement. Let $\mathbf{u}$ and $\mathbf{v}$ be the two sides of the triangle that are perpendicular to each other. Then the hypotenuse is defined by $\mathbf{u} + \mathbf{v}$. Show that $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

2.5 In $\mathcal{R}^4$, two vectors be given by

$$\mathbf{x} = [\,2 \quad -1 \quad 1 \quad 0\,]', \qquad \mathbf{y} = [\,1 \quad 2 \quad 0 \quad 3\,]'$$

Compute the scalar product of the two vectors and show that the vectors are orthogonal. Find two orthonormal vectors corresponding to the given vectors. Find the norm of each vector. Find a vector that is normal to both $\mathbf{x}$ and $\mathbf{y}$.

2.6 For $\mathcal{R}^n$, the norm of a vector can be defined as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

Show that all the axioms of norm are satisfied by this definition of norm. This is known as the 1-norm of $\mathcal{R}^n$.

2.7 For $\mathcal{R}^n$, the norm of a vector can be defined as

$$\|\mathbf{x}\|_\infty = \max_i \{|x_i|, i = 1, 2, \cdots n\}$$

Show that all the axioms of norm are satisfied by this definition of norm. This is known as the infinity norm for $\mathcal{R}^n$

2.8 Let $\mathcal{X}$ be the vector space of all square matrices. Show that a norm for $\mathcal{X}$ be defined as

$$\|\mathbf{A}\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 = \text{Trace}(\mathbf{A}'\mathbf{A})$$

This norm is called the Frobenious norm of matrices. Verify that all properties of norm are satisfied.

For the space $\mathcal{X}$, we can also define scalar product by

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

Show that all axioms of scalar product are satisfied by this definition of scalar product.

2.9 The space ℓ_2 of all square summable infinite sequence of real numbers forms a vector space. Let x be an element of ℓ_2. Then $\mathbf{x}$ has the representation $\mathbf{x} = (x_1, x_2, x_3, \cdots)$ with $\sum_{i=1}^{\infty} x_i^2 < \infty$. Addition and scalar multiplication for ℓ_2 can be defined as

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1 + y_1, x_2 + y_2, x_3 + y_3, \cdots) \\ a\mathbf{x} &= (ax_1, ax_2, ax_3, \cdots) \end{aligned}$$

Verify that all axioms of vector space are satisfied for ℓ_2.

2.10 Suppose x and y be two arbitrary vectors in $\mathcal{R}^n$, $\mathbf{Q}$ is a positive definite matrix of dimension $n \times n$. Show that the following equation satisfies all properties of a scalar product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{Q}} = \mathbf{x}'\mathbf{Q}\mathbf{y}$$

2.11 Consider the vector space $\mathcal{M} = \mathcal{R}^{2\times 2}$ of all 2×2 real matrices which is closed under standard matrix addition and scalar multiplication.

a) Find a basis for $\mathcal{M}$.

b) What is the dimension of $\mathcal{M}$?

c) Consider an arbitrary matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Express the matrix $\mathbf{A}$ in terms of the elements of the basis set determined in question a).

2.12 What is the distance between the vectors given in Problem 2.5?

2.13 Show that all the properties of metric are satisfied for

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\|_1 \\ d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\|_\infty \end{aligned}$$

2.14 A vector space, $\mathcal{V}$, be described by the vectors

$$\mathbf{v}_1 = [\,1 \quad -1 \quad 1 \quad 2\,]', \qquad \mathbf{v}_2 = [\,0 \quad 2 \quad 1 \quad 1\,]', \qquad \mathbf{v}_3 = [\,2 \quad -1 \quad 2 \quad 0\,]'$$

Does the vector $\mathbf{y} = [\,-1 \quad 2 \quad 0 \quad 3\,]'$ belong to the vector space $\mathcal{V}$?

2.15 A linear space $\mathcal{W}$ is described by the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0.5 \\ 0 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0.5 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 0.6 \\ 1.5 \\ 0.6 \\ -0.6 \end{bmatrix}$$

a) Are the above vectors linearly independent? If there is any dependent vector express it in terms of the independent vectors.

b) Find a basis for $\mathcal{W}$. What is the dimension of $\mathcal{W}$?

2.16 A linear space $\mathcal{W}$ is described by the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -2 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 5 \\ 2 \\ -2 \end{bmatrix}$$

a) Are the above vectors linearly independent? If there is any dependent vector express it in terms of the independent vectors.

b) Find a basis for $\mathcal{W}$. What is the dimension of $\mathcal{W}$?

2.17 A linear space V is described by the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} 4 \\ 5 \\ -2 \end{bmatrix}$$

a) What is the dimension of $\mathcal{V}$? Find a basis for $\mathcal{V}$.

b) Find the norm of the vector $\mathbf{v}_1$, and the distance between the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

2.18 Given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Are these vectors linearly independent? What is the dimension of the vector space spanned by these vectors? Find $\|\mathbf{v}_1\|$, $\langle \mathbf{v}_1, \mathbf{v}_3 \rangle$, and the distance $d(\mathbf{v}_1, \mathbf{v}_2)$

2.19 Given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 2 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Are these vectors linearly independent? What is the dimension of the vector space spanned by these vectors?

2.20 Given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 2 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{v}_4 = \begin{bmatrix} 1 \\ 3 \\ -2 \\ 1 \end{bmatrix}$$

Are these vectors linearly independent? What is the dimension of the vector space spanned by these vectors? Construct a basis for the vector space spanned by these vectors.

2.21 A vector space be defined by the span of the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ given in Problem 2.20, i.e.,

$$\mathcal{V} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

Find the distance between $\mathbf{v}_4$ given in Problem 2.20 and the vector space $\mathcal{V}$.

2.22 Find the null space of the following matrices:

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & 1 & 1 \\ 0 & 5 & 7 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 1 & 2 \\ 1 & 3 & -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & -1 \\ 2 & 0 & 2 \end{bmatrix}$$

2.23 For each of the linear equation given below,

a) determine if there is a solution. If there exists a solution, is it unique?
b) If there is a solution, find the general solution as the sum of the particular solution, and the homogeneous solution.
c) If there are multiple solutions, find the solution that has the minimum norm.
d) If there is no solution, find the approximate solution for the minimum error.

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

$$\begin{bmatrix} 2 & 1 & 0 & -2 \\ 1 & 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

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

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 0 \\ 3 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

2.24 Find the general solution(s) (if one exists) of the following homogeneous equations:

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

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

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

$$\begin{bmatrix} 4 & 2 & 1 \\ 5 & 4 & 0 \\ 8 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 2 & -1 \\ -1 & 2 & 0 \\ 1 & 6 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

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

2.25 Find the rank of the following matrices:

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 3 \\ -1 & 2 & -2 & 2 \\ 0 & 1 & 1 & -2 \end{bmatrix}, \qquad \begin{bmatrix} 2 & -1 & 0 & 1 \\ 2 & 0 & 1 & 3 \\ 0 & -1 & 1 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 3 \\ -1 & 2 & -2 & 2 \\ 4 & -1 & 1 & 2 \end{bmatrix}, \qquad \begin{bmatrix} 4 & -2 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ -1 & 1 & -2 & -2 \\ 0 & 1 & 1 & -2 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

2.26 A vector space is described by the following basis:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 2 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

Find an orthonormal basis for this vector space. Suppose $\mathbf{y} = [1 \quad 2 \quad 3 \quad 4]'$.

a) Express $\mathbf{y}$ in terms of the new basis computed above.

b) Express $\mathbf{y}$ in terms of the original basis set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

2.27 Find the reciprocal basis for the vector space described in Problem 2.26. Using the reciprocal basis, express $\mathbf{y}$ given in Question 2.26 in terms of the basis set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$. Compare your answer with that of Question 2.26(b).

2.28 A subspace $\mathcal{W}$ of $\mathcal{R}^4$ is described by the basis:

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

and consider a vector $\mathbf{v} = [1 \quad 1 \quad 1 \quad 1]'$. Is the vector $\mathbf{v}$ a member of the vector space $\mathbf{W}$? If not, find the projection of $\mathbf{v}$ onto $\mathcal{W}$.

2.29 Consider the Euclidean space $\mathcal{R}^3$ with the Kronecker vectors as the basis. Also consider the x-y-plane of $\mathcal{R}^3$ as a subspace $\mathcal{W}$ of $\mathcal{R}^3$. Now assume a vector $\mathbf{y} = [1 \quad 1 \quad 0.2]'$. Show that $\mathbf{y}$ is not a member of the space $\mathbf{W}$? Using trigonometric construction, find the projection of $\mathbf{y}$ onto $\mathbf{W}$. Now find the projection using mathematical method described in Section 2.7. Show that the two answers are same.

2.30 Consider a vector space consisting of vectors $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in Question 2.20. Find the projection of the vector $\mathbf{v}_4$ onto $\mathcal{V}_4$.

2.31 Given the matrix $\mathbf{A}$, find the null space, X, and its orthogonal complement, Y, so that

$$X \oplus Y = R^4$$

where

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & -2 \\ 1 & 3 & -1 & 2 \end{bmatrix}$$

2.31 Find the domain space and range space of the following matrices:

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 2 \\ 0 & 1 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 5 \\ -1 & 2 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 2 & 2 \\ -1 & 1 & 1 & -3 \\ 0 & 1 & 3 & -1 \\ 1 & 1 & 5 & 1 \end{bmatrix}$$

2.32 Verify that for the matrix $\mathbf{A}$ give in problem (2.31),

$$\mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}') = \mathcal{R}^4$$