Chapter 3
Vector Spaces and Linear Transformations

3.1 Vector Spaces

Recall that a vector in the xy plane is a line segment directed from the origin to a point in the plane as shown in Figure 3.1. Recall also that the sum of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is a vector obtained by the parallelogram rule. Also, for a real number c, $c\mathbf{v}$ is a vector whose magnitude is $|c|$ times the magnitude of $\mathbf{v}$ and whose direction is the same as the direction of $\mathbf{v}$ if $c > 0$, and opposite to the direction of $\mathbf{v}$ if $c < 0$.

A convenient way to represent a vector in the xy plane is to consider it as an ordered pair of two real numbers as $\mathbf{v} = (\alpha, \beta)$ where α and β are the components of $\mathbf{v}$ along the x and y axes. This representation allows us to define the sum of two vectors $\mathbf{v}_1 = (\alpha_1, \beta_1)$ and $\mathbf{v}_2 = (\alpha_2, \beta_2)$ in terms of their components as

$$\mathbf{v}_1 + \mathbf{v}_2 = (\alpha_1 + \alpha_2, \beta_1 + \beta_2)$$

and a scalar multiple of a vector $\mathbf{v} = (\alpha, \beta)$ as

$$c\mathbf{v} = (c\alpha, c\beta)$$

The representation of a vector in the xy plane by a pair also allows us to derive some desirable properties of vector addition and scalar multiplication. For example,

$$\mathbf{v}_1 + \mathbf{v}_2 = (\alpha_1 + \alpha_2, \beta_1 + \beta_2) = (\alpha_2 + \alpha_1, \beta_2 + \beta_1) = \mathbf{v}_2 + \mathbf{v}_1$$

and

$$(c + d)\mathbf{v} = ((c + d)\alpha, (c + d)\beta) = (c\alpha, c\beta) + (d\alpha, d\beta) = c\mathbf{v} + d\mathbf{v}$$

Finally, such a representation is useful in expressing a given vector in terms of some special vectors. For example, defining $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ to be the unit vectors along the x and y axes, we have

$$\mathbf{v} = (\alpha, \beta) = \alpha\,(1, 0) + \beta\,(0, 1) = \alpha\,\mathbf{i} + \beta\,\mathbf{j}$$

The idea of representing a vector in a plane by an ordered pair can be generalized to vectors in three dimensional xyz space, where we represent a vector $\mathbf{v}$ by a triple (α, β, γ), with α, β, and γ corresponding to the components of $\mathbf{v}$ along the x, y, and z axes. What about a quadruple $(\alpha, \beta, \gamma, \delta)$? Although we cannot visualize it as an arrow in a four dimensional space, we can still define the sum of two such

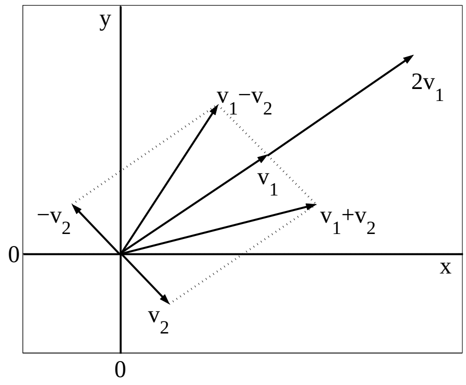

Figure 3.1: Representation of vectors in a plane

quadruples as well as a scalar multiple of a quadruple in terms of their components. This motivates the need for a more general and abstract definition of a vector.[1]

3.1.1 Definitions

A **vector space X** over a field **F** is a non-empty set, elements of which are called vectors, together with two operations called **addition** and **scalar multiplication** that have the following properties.

Addition operation associates with any two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ a unique vector denoted $\mathbf{x} + \mathbf{y} \in \mathbf{X}$, and satisfies the following conditions.

A1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$.

A2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{X}$.

A3. There exists an element of $\mathbf{X}$, denoted $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{X}$. $\mathbf{0}$ is called the **zero vector** or the **null vector**.

A4. For any $\mathbf{x} \in \mathbf{X}$ there is a vector $-\mathbf{x} \in \mathbf{X}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

Scalar multiplication operation associates with any vector $\mathbf{x} \in \mathbf{X}$ and any scalar $c \in \mathsf{F}$ a unique vector denoted $c\mathbf{x} \in \mathbf{X}$, and satisfies the following conditions.

S1. $(cd)\mathbf{x} = c(d\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$ and $c, d \in \mathsf{F}$.

S2. $1\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{X}$, where 1 is the multiplicative identity of F.

S3. $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$ for all $\mathbf{x} \in \mathbf{X}$ and $c, d \in \mathsf{F}$.

S4. $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ and $c \in \mathsf{F}$.

If the field over which a vector space is defined is clear from the context, we omit the phrase “over **F**” when referring to a vector space.

The following properties of a vector space follow directly from the definition.

[1] The reader may be accustomed to defining any directed line segment in the plane, such as an arrow directed from the tip of $\mathbf{v}_1$ to the tip of $\mathbf{v}_1 + \mathbf{v}_2$ in Figure 3.1, as a vector. However, this is just a visual aid, and as far as the addition and scalar multiplication operations just defined are concerned, that arrow is no different from $\mathbf{v}_2$. In this sense, $\mathbf{v}_2$ represents all arrows that have the same orientation and the same length as $\mathbf{v}_2$, which form an equivalence class.

a) $\mathbf{0}$ is unique

b) $-\mathbf{0} = \mathbf{0}$

c) $0\,\mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in \mathbf{X}$

d) $c\,\mathbf{0} = \mathbf{0}$ for all $c \in \mathsf{F}$

e) $-\mathbf{x}$ is unique for any $\mathbf{x} \in \mathbf{X}$

f) $(-1)\,\mathbf{x} = -\mathbf{x}$ for any $\mathbf{x} \in \mathbf{X}$

To prove (a), assume that there are two different vectors $\mathbf{0}_1 \neq \mathbf{0}_2$ satisfying condition A3. Then $\mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2$ (A3 with $\mathbf{x} = \mathbf{0}_2$ and $\mathbf{0} = \mathbf{0}_1$), and also $\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_1$ (A3 with $\mathbf{x} = \mathbf{0}_1$ and $\mathbf{0} = \mathbf{0}_2$). Then, by A1 we have $\mathbf{0}_2 = \mathbf{0}_1$, contradicting the assumption. Therefore, there can be no two distinct $\mathbf{0}$'s. Other properties can be proved similarly, and are left to the reader as an exercise.

Example 3.1

Consider the set of all ordered n-tuples[2] of the form

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

where $x_1, x_2, \ldots, x_n \in \mathsf{F}$. Defining addition and scalar multiplication operations element-by-element as

$$\begin{aligned} (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\ c\,(x_1, x_2, \ldots, x_n) &= (cx_1, cx_2, \ldots, cx_n) \end{aligned}$$

and letting

$$\begin{aligned} \mathbf{0} &= (0, 0, \ldots, 0) \\ -(x_1, x_2, \ldots, x_n) &= (-x_1, -x_2, \ldots, -x_n) \end{aligned}$$

all the properties of the vector addition and scalar multiplication are satisfied. Thus the set of all n-tuples of F is a vector space over F, called the ***n*-space** and denoted F^n. A real n-tuple is $(x_1, x_2, \ldots, x_n)$ is an obvious generalization of the familiar concept of a vector in the plane.

In particular, R^1, R^2 and R^3 can be identified with the real line, the xy plane and the xyz space, respectively.[3]

Example 3.2

The set of $m \times n$ matrices, $\mathsf{F}^{m\times n}$, together with the matrix addition and scalar multiplication operations defined in Section 1.2 is a vector space.[4]

In particular, $\mathsf{F}^{1\times n}$ and $\mathsf{F}^{n\times 1}$ are vector spaces. This is why we call a row matrix also a row vector, and a column matrix a column vector. In fact, both $\mathsf{F}^{1\times n}$ and $\mathsf{F}^{n\times 1}$ can be identified with F^n in Example 3.1. In other words, an n-tuple can be viewed as an

[2] From now on, we will distinguish an ordered set from an unordered set by enclosing its elements with parantheses rather than curly brackets

[3] Note that the set of real numbers is both a field and also a vector space. We distinguish the two by denoting the real field by R and the vector space of real numbers by R^1.

[4] The reader might ask: "When we multiply two $n \times n$ matrices, are we multiplying two vectors in $\mathsf{F}^{m\times n}$? Can we similarly multiply two vectors in R^n"? The answer is that when we multiply two matrices, we do not view them as vectors, but as something else that we will consider later. Multiplication of vectors is not defined, nor is it needed to construct a vector space.

element of either of the vector spaces $\mathbf{F}^n$, $\mathbf{F}^{1\times n}$ or $\mathbf{F}^{n\times 1}$, in which case it is represented respectively as

$$(x_1, x_2, \ldots, x_n)\,, \quad [\,x_1 \;\; x_2 \;\; \cdots \;\; x_n\,]\,, \quad \text{or} \quad \text{col}\,[\,x_1, x_2, \ldots, x_n\,]$$

* Example 3.3

An ordered real n-tuple $(x_1, x_2, \ldots, x_n)$ is a special case of a semi-infinite sequence

$$(x_k)_1^\infty = (x_1, x_2, \ldots)$$

of real numbers. Defining, by analogy to $\mathbf{R}^n$,

$$\begin{aligned} (x_1, x_2, \ldots) + (y_1, y_2, \ldots) &= (x_1 + y_1, x_2 + y_2, \ldots) \\ c\,(x_1, x_2, \ldots) &= (cx_1, cx_2, \ldots) \\ \mathbf{0} &= (0, 0, \ldots) \\ -(x_1, x_2, \ldots) &= (-x_1, -x_2, \ldots) \end{aligned}$$

we observe that the set of all such semi-infinite sequences is a vector space over $\mathbf{R}$.

Similarly, we can extend n-tuples in both directions and consider infinite sequences of the form

$$(x_k)_{-\infty}^\infty = (\ldots, x_{-1}, x_0, x_1, \ldots)$$

which form yet another vector space.

* Example 3.4

An ordered real n-tuple $(x_1, x_2, \ldots, x_n)$ in Example 3.1 can be viewed as a function $f : \mathbf{n} \to \mathbf{R}$, whose domain $\mathbf{n} = (\,1, 2, \ldots, n\,)$ is the ordered set of integers from 1 to n, and

$$f[k] = x_k\,, \quad k \in \mathbf{n}$$

Similarly, a semi-infinite sequence $(x_k)_1^\infty$ can be viewed as a function whose domain is the set of positive integers $\mathbf{N} = (\,1, 2, \ldots\,)$, and an infinite sequence $(x_k)_{-\infty}^\infty$ as a function whose domain is the set of all integers $\mathbf{Z} = (\,\ldots, -1, 0, 1, \ldots\,)$.

Consider the set $\mathcal{F}\,(\mathbf{D}, \mathbf{R})$ of all functions $f : \mathbf{D} \to \mathbf{R}$, where $\mathbf{D}$ is any finite or infinite discrete set like $\mathbf{n}$, or $\mathbf{N}$, or $\mathbf{Z}$. For $f, g \in \mathcal{F}\,(\mathbf{D}, \mathbf{R})$, we define their sum to be the function $f + g : \mathbf{D} \to \mathbf{R}$ such that

$$(f + g)[k] = f[k] + g[k]\,, \quad k \in \mathbf{D}$$

Likewise, the scalar multiple of f with a scalar c is defined to be the function $cf : \mathbf{D} \to \mathbf{R}$ such that

$$(cf)[k] = cf[k]\,, \quad k \in \mathbf{D}$$

Note that we do nothing new here, but just rephrase the definition of addition and scalar multiplication of n-tuples or sequences using an alternative formulation. We thus reach the conclusion that $\mathcal{F}\,(\mathbf{D}, \mathbf{R})$ is a vector space.

$\mathcal{F}\,(\mathbf{D}, \mathbf{R})$ in Example 3.4 is a typical example of a **function space**, a vector space whose elements are functions. Other examples of a function space are considered below.

* **Example 3.5**

Consider the set $\mathcal{F}(\mathcal{I}, \mathsf{R})$ of all real-valued functions $f : \mathcal{I} \rightarrow \mathsf{R}$ defined on a real interval $\mathcal{I}$. For $f, g \in \mathcal{F}(\mathcal{I}, \mathsf{R})$ and $c \in \mathsf{R}$, we define the functions $f + g$ and cf pointwise just like we did for $f, g \in \mathcal{F}(\mathsf{D}, \mathsf{R})$:

$$\begin{aligned}(f+g)(t) &= f(t)+g(t)\,, \quad t \in \mathcal{I}\\ (cf)(t) &= cf(t)\,, \quad t \in \mathcal{I}\end{aligned}$$

The zero function is one with

$$0(t) = 0\,, \quad t \in \mathcal{I}$$

and for any $f \in \mathcal{F}(\mathcal{I}, \mathsf{R})$, $-f$ is defined pointwise as

$$(-f)(t) = -f(t)\,, \quad t \in \mathcal{I}$$

With these definitions, $\mathcal{F}(\mathcal{I}, \mathsf{R})$ becomes a vector space over R.

The set of all real vector-valued functions $\mathbf{f} : \mathcal{I} \rightarrow \mathsf{R}^{n\times 1}$ is also a vector space over R, denoted $\mathcal{F}(\mathcal{I}, \mathsf{R}^{n\times 1})$. A vector-valued function $\mathbf{f}$ can also be viewed as a stack of scalar functions as

$$\mathbf{f} = \mathrm{col}\,[\, f_1, f_2, \ldots, f_n \,]$$

Note that a function f and its value $f(t)$ at a fixed t are different things. f is a vector, an element of $\mathcal{F}(\mathcal{I}, \mathsf{R})$, but $f(t)$ is a scalar, an element of R. This distinction is more apparent in the case of vector-valued functions: If $\mathbf{f} \in \mathcal{F}(\mathcal{I}, \mathsf{R}^{n\times 1})$ then $\mathbf{f}(t) \in \mathsf{R}^{n\times 1}$ for every $t \in \mathcal{I}$. Thus, although $\mathbf{f}$ and $\mathbf{f}(t)$ are both vectors, they are elements of different vector spaces.[5]

Similarly, the set $\mathcal{F}(\mathcal{I}, \mathsf{C})$ of all complex-valued functions $f : \mathcal{I} \rightarrow \mathsf{C}$ and the set $\mathcal{F}(\mathcal{I}, \mathsf{C}^{n\times 1})$ of all complex-vector-valued functions $\mathbf{f} : \mathcal{I} \rightarrow \mathsf{C}^{n\times 1}$ defined on a real interval $\mathcal{I}$ are vector spaces over C.

3.1.2 Subspaces

A subset $\mathbf{U} \subset \mathbf{X}$ of a vector space is called a **subspace** of $\mathbf{X}$ if it is itself a vector space with the same addition and scalar multiplication operations defined on $\mathbf{X}$. To check if a subset is a subspace we need not check all the conditions of a vector space. If $\mathbf{U}$ is a subspace then it must be closed under addition and scalar multiplication. That is, for all $\mathbf{u}, \mathbf{v} \in \mathbf{U}$, and $c \in \mathsf{F}$, we must have

$$\mathbf{u} + \mathbf{v} \in \mathbf{U}\,, \quad c\mathbf{u} \in \mathbf{U}$$

Usually these two conditions are combined into a single condition as

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \in \mathbf{U}$$

for all $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}$, and $c_1, c_2 \in \mathsf{F}$. Conversely, if $\mathbf{U}$ is closed under vector addition and scalar multiplication, then $-\mathbf{u} = (-1)\mathbf{u} \in \mathbf{U}$ for all $\mathbf{u} \in \mathbf{U}$, which in turn implies that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0} \in \mathbf{U}$. Since all other properties of vector addition and scalar multiplication are inherited from $\mathbf{X}$, we conclude that $\mathbf{U} \subset \mathbf{X}$ is a subspace if and only if it is closed under vector addition and scalar multiplication.

[5] Unfortunately, for the lack of an alternative we sometimes use the same notation to denote a function and its value. For example, e^t is used to denote both the exponential function and its value at t.

Example 3.6

Consider the following subset of $\mathbf{R}^3$.

$$\mathbf{U} = \{ (x, y, x-y) \mid x, y \in \mathbf{R} \}$$

For $\mathbf{x}_1 = (x_1, y_1, x_1 - y_1), \mathbf{x}_2 = (x_2, y_2, x_2 - y_2)$, and $c_1, c_2 \in \mathbf{R}$, we have

$$\begin{aligned} & c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \\ &= (c_1x_1, c_1y_1, c_1x_1 - c_1y_1) + (c_2x_2, c_2y_2, c_2x_2 - c_2y_2) \\ &= ((c_1x_1 + c_2x_2), (c_1y_1 + c_2y_2), (c_1x_1 + c_2x_2) - (c_1y_1 + c_2y_2)) \in \mathbf{U} \end{aligned}$$

Thus $\mathbf{U}$ is a subspace of $\mathbf{R}^3$. It is the set of all points $(x, y, z) \in \mathbf{R}^3$ that satisfy

$$x - y - z = 0$$

This is the equation of a plane through the origin $\mathbf{0} = (0, 0, 0)$. In $\mathbf{R}^3$, a plane through the origin is represented as the set of all points that satisfy

$$px + qy + rz = 0$$

for some p, q, r, not all zero. It is left to the reader to show that any such plane defines a subspace of $\mathbf{R}^3$. In particular, the equation $x = 0$ defines the yz plane, $y = 0$ the xz plane, and $z = 0$ the xy plane.

Now, consider the set of all points (x, y, z) that satisfy

$$\begin{bmatrix} p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since each equation above defines a plane through the origin, the points satisfying the above system are on the intersection of these two planes. If (p_1, q_1, r_1) and (p_2, q_2, r_2) are not proportional, then the two equations define distinct planes, and so their intersection is a straight line through the origin. Since the set of solutions of the above system is closed under addition and scalar multiplication, we conclude that any straight line through the origin is also a subspace of $\mathbf{R}^3$.

As an illustration, the first of the equations

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

describes the subspace $\mathbf{U}$ considered above, and the second describes the subspace

$$\mathbf{V} = \{ (x, y, -x) \mid x, y \in \mathbf{R} \}$$

Their intersection, which is the common solution of these equations, is the straight line described as

$$\mathbf{U} \cap \mathbf{V} = \{ (x, 2x, -x) \mid x \in \mathbf{R} \}$$

Clearly, a plane or a line not passing through the origin is not a subspace, simply because it does not include the zero vector of $\mathbf{R}^3$.

* **Example 3.7**

The set of polynomials of a complex variable s with complex coefficients is a vector space over $\mathbf{C}$, denoted $\mathbf{C}[s]$. The subset $\mathbf{C}_n[s]$, consisting of all polynomials with degree less than or equal to n, is a subspace of $\mathbf{C}[s]$. However, the set of polynomials with degree equal exactly to n is not a vector space. (Why?)

The set of polynomials in a real variable t with real coefficients is also a vector space, denoted $\mathbf{R}[t]$. Clearly, $\mathbf{R}[t]$ is a vector space over $\mathbf{R}$.

* **Example 3.8**

Let $\mathcal{C}_m(\mathcal{I}, \mathbf{R})$ denote the set of all real-valued functions defined on some real interval $\mathcal{I}$ such that $f, f', \ldots, f^{(m)}$ all exist and are continuous on $\mathcal{I}$. That is, $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$ is the set of continuous functions, $\mathcal{C}_1(\mathcal{I}, \mathbf{R})$ is the set of differentiable functions with a continuous derivative, etc. Also, let $\mathcal{C}_\infty(\mathcal{I}, \mathbf{R})$ denote the set of functions that have continuous derivatives of every order. By definition

$$\mathcal{F}(\mathcal{I}, \mathbf{R}) \supset \mathcal{C}_0(\mathcal{I}, \mathbf{R}) \supset \mathcal{C}_1(\mathcal{I}, \mathbf{R}) \supset \cdots \supset \mathcal{C}_\infty(\mathcal{I}, \mathbf{R})$$

Each of these sets is closed under the addition and scalar multiplication operations defined for the function space $\mathcal{F}(\mathcal{I}, \mathbf{R})$ in Example 3.4, and therefore, is a subspace of $\mathcal{F}(\mathcal{I}, \mathbf{R})$. The subspaces $\mathcal{C}_m(\mathcal{I}, \mathbf{R}^{n\times 1}) \subset \mathcal{F}(\mathcal{I}, \mathbf{R}^{n\times 1})$ can be defined similarly as

$$\mathcal{C}_m(\mathcal{I}, \mathbf{R}^{n\times 1}) = \{\, \mathbf{f} = \mathrm{col}\,[\, f_1, f_2, \ldots, f_n \,] \mid f_i \in \mathcal{C}_m(\mathcal{I}, \mathbf{R}),\ i = 1, \ldots, n \,\}$$

3.2 Span and Linear Independence

3.2.1 Span

Let $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k\}$ be a finite subset of a vector space $\mathbf{X}$. An expression of the form

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + \cdots + c_k\mathbf{r}_k$$

where $c_1, c_2, \ldots, c_k \in \mathbf{F}$, is called a **linear combination** of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$. Because of property A2 of vector addition, a linear combination unambiguously defines a vector in $\mathbf{X}$. The set of all linear combinations of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$ is called the **span** of $\mathbf{R}$, denoted $\mathrm{span}\,(\mathbf{R})$ or $\mathrm{span}\,(\mathbf{r}_1, \ldots, \mathbf{r}_k)$. Thus

$$\mathrm{span}\,(\mathbf{R}) = \{\, c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + \cdots + c_k\mathbf{r}_k \mid c_1, c_2, \ldots, c_k \in \mathbf{F} \,\}$$

If $\mathrm{span}\,(\mathbf{R}) = \mathbf{X}$, then $\mathbf{R}$ is called a **spanning set**.

The definition of span can be extended to infinite sets. The span of an infinite set of vectors is defined to be the set of all finite linear combinations of vectors of $\mathbf{R}$. More precisely,

$$\mathrm{span}\,(\mathbf{R}) = \{\, \sum_{i\in\mathbf{I}} c_i\mathbf{r}_i \mid \mathbf{I} \text{ is a finite index set}, c_i \in \mathbf{F}, \mathbf{r}_i \in \mathbf{R} \,\}$$

If $\mathbf{u}, \mathbf{v} \in \mathrm{span}\,(\mathbf{R})$, then $\mathbf{u} = \sum a_i\mathbf{r}_i$ and $\mathbf{v} = \sum b_i\mathbf{r}_i$ for some $a_i, b_i \in \mathbf{F}$. Then

$$c\mathbf{u} + d\mathbf{v} = \sum (ca_i + db_i)\mathbf{r}_i \in \mathrm{span}\,(\mathbf{R})$$

for any $c, d \in \mathbf{F}$. This shows that $\mathrm{span}\,(\mathbf{R})$ is a subspace of $\mathbf{X}$. In fact, it is the smallest subspace that contains all the vectors in $\mathbf{R}$.

Example 3.9

Let $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ denote the unit vectors along the x and y axes of the xy plane (R^2). Then

$$\text{span}\,(\,\mathbf{i}\,) = \{\,(\alpha, 0)\,|\,\alpha \in \mathsf{R}\}$$

and

$$\text{span}\,(\,\mathbf{j}\,) = \{\,(0, \beta)\,|\,\beta \in \mathsf{R}\}$$

are the x and y axes, and

$$\text{span}\,(\,\mathbf{i}, \mathbf{j}\,) = \{\,(\alpha, \beta)\,|\,\alpha, \beta \in \mathsf{R}\}$$

is the whole xy plane.

Example 3.10

In R^3, let

$$\mathbf{r}_1 = (0, 0, 1)\,, \quad \mathbf{r}_2 = (0, 1, -1)\,, \quad \mathbf{r}_3 = (1, 0, 1)\,, \quad \mathbf{r}_4 = (1, -1, 2)$$

Then

a) Span of each of the vectors is a straight line through the origin on which that vector lies. For example, $\text{span}\,(\mathbf{r}_1) = \{\,(0, 0, c)\mid c \in \mathsf{R}\,\}$, which is the z axis. Since the given vectors are different, each spans a different lines.

b) Any two of the given vectors span a plane through the origin that contain those two vectors. For example,

$$\text{span}\,(\mathbf{r}_2, \mathbf{r}_3) = \{\,(a, b, a-b)\,|\,a, b \in \mathsf{R}\,\}$$

which is the subspace $\mathbf{U}$ in Example 3.6.

c) $\text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \mathsf{R}^3$, because by definition $\text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \subset \mathsf{R}^3$, and for any $\mathbf{x} = (a, b, c) \in \mathsf{R}^3$

$$\mathbf{x} = (b + c - a)\,\mathbf{r}_1 + b\,\mathbf{r}_2 + a\,\mathbf{r}_3 \in \text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$$

so that $\mathsf{R}^3 \subset \text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ also. Similarly, $\text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4) = \text{span}\,(\mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_4) = \mathsf{R}^3$.

d) However, $\text{span}\,(\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \text{span}\,(\mathbf{r}_2, \mathbf{r}_3) = \text{span}\,(\mathbf{r}_2, \mathbf{r}_4) = \text{span}\,(\mathbf{r}_3, \mathbf{r}_4) = \mathbf{U}$.

e) Finally, $\text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \mathsf{R}^3$, simply because

$$\text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) \supset \text{span}\,(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$$

3.2.2 Linear Independence

A finite set of vectors $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k\}$ is said to be **linearly independent** if

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + \cdots + c_k\mathbf{r}_k = \mathbf{0} \tag{3.1}$$

holds only when $c_1 = c_2 = \cdots = c_k = 0$.

A set is said to be **linearly dependent** if it is not linearly independent. Alternatively, a finite set $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k\}$ is linearly dependent if there exist $c_1, \ldots, c_k$, not all 0, that satisfy (3.1).[6]

[6] We also say that the vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$ are linearly independent (dependent) to mean that the set consisting of these vectors is linearly independent (dependent).

By definition, a set containing only a single vector $\mathbf{r}$ is linearly independent if and only if $\mathbf{r} \neq \mathbf{0}$.

We have the following results concerning linear independence.

a) If $\mathbf{0} \in \mathbf{R}$ then $\mathbf{R}$ is linearly dependent.
b) If $\mathbf{R}$ is linearly independent and $\mathbf{S} \subset \mathbf{R}$, then $\mathbf{S}$ is also linearly independent. Equivalently, if $\mathbf{R}$ is linearly dependent and $\mathbf{S} \supset \mathbf{R}$, then $\mathbf{S}$ is also linearly dependent.
c) $\mathbf{R}$ is linearly dependent if and only if at least one vector in $\mathbf{R}$ can be written as a linear combination of some other vectors in $\mathbf{R}$ (assuming, of course, that $\mathbf{R}$ contains at least two vectors).

The rest being direct consequences of the definitions, only the necessity part of the last result requires a proof. If $\mathbf{R}$ is linearly dependent then there exist $c_1, \ldots, c_k$, not all 0, such that

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + \cdots + c_k\mathbf{r}_k = \mathbf{0}$$

Suppose $c_p \neq 0$. Then

$$\mathbf{r}_p = \sum_{q \neq p} (-c_q/c_p)\mathbf{r}_q$$

Property (b) above can be used to define linear dependence and independence of infinite sets. An infinite set is said to be linearly independent if every finite subset of it is linearly independent, and linearly dependent if it has a linearly dependent finite subset.

Example 3.11

The vectors $\mathbf{i}$ and $\mathbf{j}$ in Example 3.9 are linearly independent, because

$$\mathbf{0} = \alpha\mathbf{i} + \beta\mathbf{j} = (\alpha, \beta) \implies \alpha = \beta = 0$$

Example 3.12

Consider the vectors in Example 3.10. The set $\mathbf{R}_1 = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ is linearly independent, because

$$\mathbf{0} = c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + c_3\mathbf{r}_3 = (c_3, c_2, c_1 - c_2 + c_3)$$

implies

$$c_1 = c_2 = c_3 = 0$$

Similarly, the sets $\mathbf{R}_2 = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4\}$ and $\mathbf{R}_3 = \{\mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_4\}$ are linearly independent. Therefore, all subsets of these three sets, which include all singletons and pairs of $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ and $\mathbf{r}_4$, are also linearly independent.

However, the set $\mathbf{R}_4 = \{\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$ is linearly dependent, because

$$\mathbf{r}_2 - \mathbf{r}_3 + \mathbf{r}_4 = (0, 0, 0) = \mathbf{0}$$

Therefore, $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$ is also linearly dependent.

Example 3.13

In $\mathbf{C}^3$, let

$$\mathbf{x}_1 = (1, i, 0)\,, \quad \mathbf{x}_2 = (i, 0, 1)\,, \quad \mathbf{x}_3 = (0, 1, 1)$$

Then $\{\mathbf{x}_1, \mathbf{x}_2\}$ is linearly independent, because for $c_1 = a_1 + ib_1$ and $c_2 = a_2 + ib_2$,

$$\begin{aligned} \mathbf{0} &= c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \\ &= (a_1 + ib_1, -b_1 + ia_1, 0) + (-b_2 + ia_2, 0, a_2 + ib_2) \\ &= ((a_1 - b_2) + i(b_1 + a_2), -b_1 + ia_1, a_2 + ib_2)) \end{aligned}$$

implies $a_1 = b_1 = a_2 = b_2 = 0$, or equivalently, $c_1 = c_2 = 0$.

Similarly, the sets $\{\mathbf{x}_1, \mathbf{x}_3\}$ and $\{\mathbf{x}_2, \mathbf{x}_3\}$ are linearly independent. On the other hand, $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is linearly dependent, because

$$i\,\mathbf{x}_1 - \mathbf{x}_2 + \mathbf{x}_3 = \mathbf{0}$$

* **Example 3.14**

A set of functions $f_1, \ldots, f_k \in \mathcal{F}(\mathcal{I}, \mathbf{F})$ is linearly dependent if

$$c_1 f_1 + \cdots + c_k f_k = 0$$

for some scalars $c_1, \ldots, c_k \in \mathbf{F}$, not all 0. This is a functional equality, which is equivalent to

$$c_1 f_1(t) + \cdots + c_k f_k(t) = 0 \quad \text{for all} \;\; t \in \mathcal{I} \tag{3.2}$$

Consider the real-valued functions $\phi_1(t) = e^{\sigma_1 t}$ and $\phi_2(t) = e^{\sigma_2 t}$, where $\sigma_1 \neq \sigma_2 \in \mathbf{R}$. Unless $c_1 = c_2 = 0$, the equality

$$c_1 e^{\sigma_1 t} + c_2 e^{\sigma_2 t} = 0$$

can be satisfied for at most a single value of t (the graphs of $c_1 e^{\sigma_1 t}$ and $-c_2 e^{\sigma_2 t}$ either do not intersect, or intersect at a single point). Therefore, ϕ_1 and ϕ_2 are linearly independent on any interval $\mathcal{I}$.

Now consider two complex-valued functions $\psi_1(t) = e^{\lambda_1 t}$ and $\psi_2(t) = e^{\lambda_2 t}$, where $\lambda_1 \neq \lambda_2 \in \mathbf{C}$. The graphical argument above is of no use for we cannot plot graphs of complex-valued functions, and we need an algebraic method to test linear independence of $\psi_1(t)$ and $\psi_2(t)$. Such a method is based on the observation that if

$$c_1\psi_1(t) + c_2\psi_2(t) = 0 \quad \text{for all} \;\; t \in \mathcal{I}$$

then

$$c_1\psi_1'(t) + c_2\psi_2'(t) = 0 \quad \text{for all} \;\; t \in \mathcal{I}$$

provided ψ_1 and ψ_2 are differentiable on $\mathcal{I}$. For the given ψ_1 and ψ_2, which are differentiable everywhere, these two equations can be written in matrix form as

$$\begin{bmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ \lambda_1 e^{\lambda_1 t} & \lambda_2 e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

A simple elementary operation reduces the system to

$$\begin{bmatrix} e^{\lambda_1 t} & e^{\lambda_2 t} \\ 0 & (\lambda_2 - \lambda_1) e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since $\lambda_2 - \lambda_1 \neq 0$ and $e^{\lambda_2 t} \neq 0$ for all t, the second equation gives $c_2 = 0$. Similarly, since $e^{\lambda_1 t} \neq 0$ for all t, the first equation gives $c_1 = 0$. Hence ψ_1 and ψ_2 too are linearly independent on any interval $\mathcal{I}$.

Using the same technique we can show that the real-valued functions $\xi_1(t) = e^{\sigma t}$ and $\xi_2(t) = te^{\sigma t}$ are also linearly independent.

Note that the function pairs in the above three cases are solutions of a second order linear differential equation with constant coefficients whose characteristic polynomial has either the real roots $s_{1,2} = \sigma_{1,2}$ or the complex conjugate roots $s_{1,2} = \lambda_{1,2} = \sigma \mp i\omega$ or a double real root $s = \sigma$. In each case, the corresponding solutions are linearly independent either as elements of $\mathcal{F}(\mathcal{I}, \mathbf{R})$ or as elements of $\mathcal{F}(\mathcal{I}, \mathbf{C})$.

* **Example 3.15**

If $f_j = g_j + ih_j, f_j^* = g_j - ih_j, j = 1, \ldots, k$, are $2k$ linearly independent functions in $\mathcal{F}(\mathcal{I}, \mathbf{C})$, then their real and imaginary parts, $g_j, h_j, j = 1, \ldots, k$, are linearly independent in $\mathcal{F}(\mathcal{I}, \mathbf{R})$. To show this, suppose that

$$\sum_{j=1}^{k}(a_j g_j + b_j h_j) = 0$$

Noting that

$$g_j = \frac{1}{2}(f_j + f_j^*) \quad \text{and} \quad h_j = \frac{1}{2i}(f_j - f_j^*)$$

the above expression becomes

$$\frac{1}{2}\sum_{j=1}^{k}(c_j f_j + c_j^* f_j^*) = 0$$

where $c_j = a_j - ib_j$. Linear independence of $\{f_j, f_j^* \,|\, j = 1, \ldots, k\}$ implies $c_j = 0, j = 1, \ldots, k$, and therefore, $a_j = b_j = 0, j = 1, \ldots, k$.

Observe that this example explains why the real and imaginary parts of complex solutions of second order linear differential equation with constant coefficients, whose characteristic polynomial has a pair of complex-conjugate roots, are linearly independent real solutions.

3.2.3 Elementary Operations

Consider an $m \times n$ matrix A partitioned into its rows

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{bmatrix} \in \mathbf{F}^{m\times n}$$

Since the rows of A are vectors in $\mathbf{F}^{1\times n}$ (as noted in Example 3.2), the elementary row operations on A discussed in Section 1.4 can be viewed as operations involving the elements of the ordered set $\mathbf{R} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m) \subset \mathbf{F}^{1\times n}$. This observation suggests that similar operations can be defined for any ordered subset of a vector space.

The following operations on a finite ordered set of vectors $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k)$ are called **elementary operations**.

I: Interchange any two vectors

II: Multiply any vector by a nonzero scalar

III: Add a scalar multiple of a vector to another one

As we discussed in connection with elementary row operations, to every elementary operation there corresponds an inverse operation of the same type such that if $\mathbf{R}'$ is obtained from $\mathbf{R}$ by a single elementary operation, then $\mathbf{R}$ can be recovered from $\mathbf{R}'$ by performing the inverse operation.

Let $\mathbf{R}'$ be obtained from $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k)$ by a single elementary operation. If it is a Type I or Type II operation, then it is clear that $\text{span}\,(\mathbf{R}') = \text{span}\,(\mathbf{R})$. Suppose it is a Type III operation that consists of adding α times $\mathbf{r}_p$ to $\mathbf{r}_q$ for some $p \neq q$. That is,

$$\mathbf{r}'_i = \begin{cases} \mathbf{r}_i, & i \neq q \\ \mathbf{r}_q + \alpha\mathbf{r}_p, & i = q \end{cases} \tag{3.3}$$

For an arbitrary $\mathbf{x} \in \text{span}\,(\mathbf{R}')$

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{r}'_1 + \cdots + c_p\mathbf{r}'_p + \cdots + c_q\mathbf{r}'_q + \cdots + c_k\mathbf{r}'_k \\ &= c_1\mathbf{r}_1 + \cdots + c_p\mathbf{r}_p + \cdots + c_q(\mathbf{r}_q + \alpha\mathbf{r}_p) + \cdots + c_k\mathbf{r}_k \\ &= c_1\mathbf{r}_1 + \cdots + (c_p + \alpha c_q)\mathbf{r}_p + \cdots + c_q\mathbf{r}_q + \cdots + c_k\mathbf{r}_k \end{aligned} \tag{3.4}$$

so that $\mathbf{x} \in \text{span}\,(\mathbf{R})$. Hence, $\text{span}\,(\mathbf{R}') \subset \text{span}\,(\mathbf{R})$. Considering the inverse elementary operation, it can similarly be shown that $\text{span}\,(\mathbf{R}) \subset \text{span}\,(\mathbf{R}')$. Hence $\text{span}\,(\mathbf{R}') = \text{span}\,(\mathbf{R})$. Obviously, this property also holds if $\mathbf{R}'$ is obtained from $\mathbf{R}$ by a finite sequence of elementary operations.

Another property of elementary operations is the preservation of linear independence: If $\mathbf{R}'$ is obtained from $\mathbf{R}$ by a finite sequence of elementary operations, then $\mathbf{R}'$ is linearly independent if and only if $\mathbf{R}$ is linearly independent. Again, the proof is trivial if $\mathbf{R}'$ is obtained from $\mathbf{R}$ by a single Type I or Type II elementary operation. Suppose that $\mathbf{R}'$ is obtained from $\mathbf{R}$ by a single Type III elementary operation as described in (3.3), and consider a linear combination as in (3.4) with $\mathbf{x} = \mathbf{0}$. If $\mathbf{R}$ is linearly independent then all the coefficients in the last linear combination in (3.4) must be zero, which implies that all c_i's are zero, so that $\mathbf{R}'$ is also linearly independent. By considering the inverse elementary operation the converse can also be shown to be true.

These properties of elementary operations can be used to characterize the span of a set or to check its linear independence as illustrated by the following example.

Example 3.16

Consider the set $\mathbf{R}_1 = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ in Example 3.12. Identifying $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ with the rows of the matrix

$$R_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$

and performing elementary row operations on R_1, we observe that

$$R_1 \longrightarrow I$$

Thus

$$\mathbf{R}_1 \longrightarrow \mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$$

where

$$\mathbf{e}_1 = (1,0,0)\,, \quad \mathbf{e}_2 = (0,1,0)\,, \quad \mathbf{e}_3 = (0,0,1)$$

correspond to rows of I. Since $\mathbf{E}$ is linearly independent then so is $\mathbf{R}_1$. Also

$$\text{span}\,(\mathbf{R}_1) = \text{span}\,(\mathbf{E}) = \{\,(x,y,z)\,|\,x,y,z \in \mathsf{R}\,\} = \mathsf{R}^3$$

Now consider the set $\mathbf{R}_4 = (\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)$. Performing elementary operations on $\mathbf{R}_4$ as above, we obtain

$$\mathbf{R}_4 \longrightarrow \mathbf{S} = (\mathbf{r}_2, \mathbf{r}_3, \mathbf{0})$$

Then

$$\text{span}\,(\mathbf{R}_4) = \text{span}\,(\mathbf{S}) = \text{span}\,(\mathbf{r}_2, \mathbf{r}_3) = \mathbf{U}$$

Also, since $\mathbf{S}$ is linearly dependent then so is $\mathbf{R}_4$.

Note that these results have already been obtained in Examples 3.10 and 3.12.

3.3 Bases and Representations

If $\mathbf{R}$ is a spanning set then any $\mathbf{x} \in \mathbf{X}$ can be expressed as a linear combination of vectors in $\mathbf{R}$. A significant question is whether we need all the vectors in $\mathbf{R}$ to be able to do that for every $\mathbf{x} \in \mathbf{X}$. For example, the set $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$ in Example 3.10 spans R^3, but so also do $\mathbf{R}_1 = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$, $\mathbf{R}_2 = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4\}$, and $\mathbf{R}_3 = \{\mathbf{r}_1, \mathbf{r}_3, \mathbf{r}_4\}$. That is, any one of $\mathbf{r}_2$, $\mathbf{r}_3$, or $\mathbf{r}_4$ can be removed from $\mathbf{R}$ without losing the spanning property. On the other hand, if $\mathbf{r}_1$ is removed from $\mathbf{R}$, then the resulting set $\mathbf{R}_4 = \{\mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$ no longer spans R^3. Apparently, the sets $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ have a property that $\mathbf{R}_4$ does not have. Referring to Example 3.12 we find out that the first three sets are linearly independent while the last is not, and this may be a clue.

Lets take another look at one of those linearly independent spanning sets, say $\mathbf{R}_1$. If we remove one more vector from $\mathbf{R}_1$, then the resulting set of two vectors will only span a plane (a subspace), not the whole space. Thus $\mathbf{R}_1$ is a minimal spanning set. On the other hand, if we add any vector $\mathbf{r}$ different from $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ to $\mathbf{R}_1$, then the resulting set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}\}$ will no longer be linearly independent, because $\mathbf{r}$ can be expressed as a linear combination of the others. Hence $\mathbf{R}_1$ is also a maximal linearly independent set. The same are true also for $\mathbf{R}_2$ and $\mathbf{R}_3$.

These observations motivate a need to investigate the link between the concepts of span and linear independence.

3.3.1 Basis

The following theorem, which is one of the fundamental results of linear algebra, characterizes a linearly independent spanning set.

Theorem 3.1 *Let* $\mathbf{R}$ *be a subset of a vector space* $\mathbf{X}$*. Then the following are equivalent.*

a) $\mathbf{R}$ *is linearly independent and spans* $\mathbf{X}$*.*

b) $\mathbf{R}$ *spans* $\mathbf{X}$*, and no proper subset of* $\mathbf{R}$ *spans* $\mathbf{X}$*. (That is,* $\mathbf{R}$ *is a minimal spanning set.)*

c) $\mathbf{R}$ *is linearly independent, and no proper superset of* $\mathbf{R}$ *is linearly independent. (That is,* $\mathbf{R}$ *is a maximal linearly independent set.)*

d) Every vector $\mathbf{x} \in \mathbf{X}$ *can be expressed as a linear combination of the vectors of* $\mathbf{R}$ *in a unique way. That is,*

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{r}_i$$

for some $\mathbf{r}_1, \ldots, \mathbf{r}_k \in \mathbf{R}$ *and* $\alpha_1, \ldots, \alpha_k \in \mathsf{F}$*, all of which are uniquely determined by* $\mathbf{x}$*.*

Proof We will show that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b):

By the second part of the hypothesis $\mathbf{R}$ spans $\mathbf{X}$. If a proper subset $\mathbf{S} \subset \mathbf{R}$ spans $\mathbf{X}$, then there exists a vector $\mathbf{r} \in \mathbf{R} - \mathbf{S}$ that can be written as a linear combination of some vectors in $\mathbf{S}$. This implies that $\mathbf{R}$ is linearly dependent, contradicting the first part of the hypothesis. Hence no proper subset of $\mathbf{R}$ can span $\mathbf{X}$.

(b) $\Rightarrow$ (c):

If $\mathbf{R}$ is linearly dependent, then there exists $\mathbf{r} \in \mathbf{R}$ which can be written as a linear combination of some other vectors in $\mathbf{R}$. This implies that $\mathbf{R} - \{\mathbf{r}\}$ also spans $\mathbf{X}$, contradicting the second part of the hypothesis. Hence $\mathbf{R}$ is linearly independent. On the other hand, since every vector $\mathbf{x} \notin \mathbf{R}$ can be written as a linear combination of vectors of $\mathbf{R}$ (because $\mathbf{R}$ spans $\mathbf{X}$), no proper superset of $\mathbf{R}$ can be linearly independent.

(c) $\Rightarrow$ (d):

If there exists a nonzero vector $\mathbf{x}$ which cannot be expressed as a linear combination of vectors in $\mathbf{R}$, then $\mathbf{R} \cup \{\mathbf{x}\}$ is linearly independent (see Exercise 3.13), contradicting the second part of the hypothesis. Hence every vector can be expressed in terms of the vectors of $\mathbf{R}$. Now if a vector $\mathbf{x}$ can be expressed as two different linear combinations of the vectors of $\mathbf{R}$ as

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{r}_i = \sum_{i=1}^{k} \beta_i \mathbf{r}_i$$

then

$$\sum_{i=1}^{k} (\alpha_i - \beta_i) \mathbf{r}_i = \mathbf{0}$$

where at least one coefficient $\alpha_i - \beta_i$ is nonzero. This means that $\mathbf{R}$ is linearly dependent, contradicting the first part of the hypothesis. Hence the expression for $\mathbf{x}$ in terms of the vectors of $\mathbf{R}$ is unique.

(d) $\Rightarrow$ (a):

By hypothesis $\mathbf{R}$ spans $\mathbf{X}$. If $\mathbf{R}$ is linearly dependent, then there exists a vector $\mathbf{r} \in \mathbf{R}$ which can be expressed as

$$\mathbf{r} = \sum_{i=1}^{k} c_i \mathbf{r}_i$$

for some $\mathbf{r}_1, \ldots, \mathbf{r}_k \in \mathbf{R}$, which means that $\mathbf{r}$ has two different expressions in terms of $\mathbf{r}, \mathbf{r}_1, \ldots, \mathbf{r}_k \in \mathbf{R}$, contradicting the hypothesis. Hence $\mathbf{R}$ is also linearly independent.

A set $\mathbf{R}$ having the properties in Theorem 3.1 is called a **basis** for $\mathbf{X}$. A basis is a generalization of the concept of a coordinate system in a plane to abstract vector spaces. Consider the vectors $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ along the x and y axes of the xy plane (R^2). Since any vector $\mathbf{x} = (\alpha, \beta)$ has a unique representation as $\mathbf{x} = \alpha\mathbf{i} + \beta\mathbf{j}$, the vectors $\mathbf{i}$ and $\mathbf{j}$ form a basis for the xy plane. The basis vectors of a vector space play exactly the same role as do the vectors $\mathbf{i}$ and $\mathbf{j}$ in the xy plane.

Example 3.17

Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ denote columns of I_3. The set $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ spans $\mathsf{R}^{3\times 1}$, because any $\mathbf{x} = \operatorname{col}[\, x_1, x_2, x_3 \,]$ can be expressed as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \tag{3.5}$$

$\mathbf{E}$ is also linearly independent, because a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ as above is $\mathbf{0}$ only if all the coefficients x_1, x_2 and x_3 are zero. Hence $\mathbf{E}$ is a basis for $\mathsf{R}^{3\times 1}$, called the **canonical basis**. Canonical bases for $\mathsf{F}^n, \mathsf{F}^{1\times n}$ and $\mathsf{F}^{n\times 1}$ can be defined similarly. For example, the set $\mathbf{E}$ in Example 3.16 is the canonical basis for R^3.

We now claim that the set $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$, where

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

is also a basis for $\mathsf{R}^{3\times 1}$. To check if an arbitrary vector $\mathbf{x} = \operatorname{col}[\, x_1, x_2, x_3 \,]$ can be expressed in terms of the vectors in $\mathbf{R}$ we try to solve

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

for α_1, α_2 and α_3. Since the coefficient matrix of the above equation in already in a row echelon form, we obtain a unique solution by back substitution as

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \\ x_3 \end{bmatrix}$$

Thus

$$\mathbf{x} = (x_1 - x_2)\mathbf{r}_1 + (x_2 - x_3)\mathbf{r}_2 + x_3\mathbf{r}_3 \tag{3.6}$$

which shows that $\mathbf{R}$ spans $\mathsf{R}^{3\times 1}$. Moreover, since the coefficients of $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ in the above expression are uniquely determined by $\mathbf{x}$, $\mathbf{R}$ must also be linearly independent. Indeed,

$$c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + c_3\mathbf{r}_3 = \begin{bmatrix} c_1 + c_2 + c_3 \\ c_2 + c_3 \\ c_3 \end{bmatrix} = \mathbf{0}$$

implies $c_1 = c_2 = c_3 = 0$. This proves our claim that $\mathbf{R}$ is also a basis for $\mathsf{R}^{3\times 1}$.

* **Example 3.18**

In the vector space $\mathsf{C}[s]$ of polynomials, let $\mathcal{Q} = \{q_0, q_1, \ldots\}$, where the polynomials q_i are defined as

$$q_i(s) = s^i\,, \quad i = 0, 1, \ldots$$

Since any polynomial $p(s) = c_0 + c_1 s + \cdots + c_n s^n$ can be expressed as

$$p = c_0 q_0 + c_1 q_1 + \cdots + c_n q_n$$

$\mathcal{Q}$ spans $\mathsf{C}[s]$.

Consider the finite subset $\mathcal{Q}_n = \{q_0, q_1, \ldots, q_n\}$ of $\mathcal{Q}$, and let p be a linear combination of $q_0, q_1, \ldots, q_n$ expressed as above. If $p = 0$ then $p(s) = p'(s) = p''(s) = \cdots = 0$ for all s. Evaluating at $s = 0$, we get $c_1 = c_2 = \cdots = c_n = 0$, which shows that $\mathcal{Q}_n$ is linearly independent. Since any finite subset of $\mathcal{Q}$ is a subset of $\mathcal{Q}_n$ for some n, it follows that every finite subset of $\mathcal{Q}$ is linearly independent. Hence $\mathcal{Q}$ is linearly independent, and therefore, it is a basis for $\mathsf{C}[s]$.

The reader can show that the set $\mathcal{R} = \{r_0, r_1, \ldots\}$, where

$$r_i(s) = 1 + s + \cdots + s^i\,, \quad i = 0, 1, \ldots$$

is also a basis for $\mathsf{C}[s]$. In fact, $\mathcal{R}$ can be obtained from $\mathcal{Q}$ by a sequence of elementary operations.[7]

From the two examples above we observe that a vector space may have a finite or an infinite basis. A vector space with a finite basis is said to be **finite dimensional**, otherwise, **infinite dimensional**. Thus $\mathsf{R}^{3\times 1}$ in Example 3.17 is finite dimensional, and $\mathsf{C}[s]$ in Example 3.18 is infinite dimensional. These examples also illustrate that basis for a vector space is not unique.

The following corollary of Theorem 3.1 characterizes bases of a finite dimensional vector space.

Corollary 3.1.1 *Let* $\mathbf{X}$ *have a finite basis* $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$. *Then*

a) No subset of $\mathbf{X}$ *containing more than* n *vectors is linearly independent.*

b) No subset of $\mathbf{X}$ *containing less than* n *vectors spans* $\mathbf{X}$.

c) Any basis of $\mathbf{X}$ *contains exactly* n *vectors.*

d) Any linearly independent set that contains exactly n *vectors is a basis.*

e) Any spanning set that contains exactly n *vectors is a basis.*

[7] Although we defined elementary operations on a finite set only, the definition can easily be extended to infinite sets.

Proof

a) Consider a set of of vector $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m\}$, where $m > n$. Since $\mathbf{R}$ is a basis, each $\mathbf{s}_j$ can be expressed in terms of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ as

$$\mathbf{s}_j = \sum_{i=1}^{n} a_{ij}\mathbf{r}_i \,, \quad j = 1, 2, \ldots, m$$

Let $A = [\, a_{ij} \,]_{n\times m}$. Since $n < m$, the linear system $A\mathbf{c} = \mathbf{0}$ has a nontrivial solution, that is, there exist $c_1, c_2, \ldots, c_m$, not all zero, such that

$$\sum_{j=1}^{m} a_{ij}c_j = 0 \,, \quad i = 1, 2, \ldots, n$$

Then

$$\sum_{j=1}^{m} c_j\mathbf{s}_j = \sum_{j=1}^{m} c_j(\sum_{i=1}^{n} a_{ij}\mathbf{r}_i\,) = \sum_{i=1}^{n}(\sum_{j=1}^{m} a_{ij}c_j\,)\mathbf{r}_i = \mathbf{0}$$

which shows that $\mathbf{S}$ is linearly dependent.

b) Suppose that a set of vectors $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m\}$, where $m < n$, spans $\mathbf{X}$. Then each $\mathbf{r}_j$ can be expressed as

$$\mathbf{r}_j = \sum_{i=1}^{m} b_{ij}\mathbf{s}_i \,, \quad j = 1, 2, \ldots, n$$

Let $B = [\, b_{ij} \,]_{m\times n}$. Since $m < n$, we can show by following the same argument as in part (a) that there exist $c_1, c_2, \ldots, c_n$, not all zero, such that

$$\sum_{j=1}^{n} c_j\mathbf{r}_j = \sum_{j=1}^{n} c_j(\sum_{i=1}^{m} b_{ij}\mathbf{s}_i\,) = \sum_{i=1}^{m}(\sum_{j=1}^{n} b_{ij}c_j\,)\mathbf{s}_i = \mathbf{0}$$

Since this contradicts the assumption that $\mathbf{R}$ is linearly dependent $\mathbf{S}$ cannot span $\mathbf{X}$.

c) If $\mathbf{S}$ is a basis containing m vectors, then by (a) $m \le n$, and by (b) $m \ge n$, that is, $m = n$.

d) Let $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ be a linearly independent set. Then, by (a) no proper superset of $\mathbf{S}$ is linearly independent, and by Theorem 3.1, $\mathbf{S}$ is a basis.

e) Let $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ be a spanning set. Then, by (b) no proper subset of $\mathbf{S}$ spans $\mathbf{X}$, and by Theorem 3.1, $\mathbf{S}$ is a basis.

By Corollary 3.1(c), all bases of a finite dimensional vector space contain the same number of basis vectors. This fixed number is called the **dimension** of $\mathbf{X}$, denoted $\dim(\mathbf{X})$. If $\mathbf{X}$ is a trivial vector space containing only the zero vector, then it has no basis and $\dim(\mathbf{X}) = 0$.

From Example 3.17 we conclude that

$$\dim(\mathsf{F}^n) = \dim(\mathsf{F}^{1\times n}) = \dim(\mathsf{F}^{n\times 1}) = n$$

Example 3.19

From Examples 3.10 and 3.12 we conclude that the sets $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ are all bases for R^3. Since $\dim(\mathsf{R}^3) = 3$, the set $\mathbf{R}$, which contains four vectors, must be linearly dependent. Although the set $\mathbf{R}_4$ also contains three vectors, it is not a basis, because it is not linearly independent. Then it cannot span R^3. These results had already been obtained in Example 3.12.

Referring to the same example, we also observe that the set $\{\mathbf{r}_2, \mathbf{r}_3\}$ is a basis for the subspace $\mathbf{U}$ in Example 3.6. Hence $\dim(\mathbf{U}) = 2$. This is completely expected as $\mathbf{U}$ is essentially the same as the two-dimensional xy plane (or the yz or xz planes) except that it is tilted about the origin. The linearly independent vectors

$$\mathbf{s}_1 = (1,1,0) \quad \text{and} \quad \mathbf{s}_2 = (2,1,1)$$

form another basis for $\mathbf{U}$.

Example 3.20

Any two vectors not lying on the same straight line are linearly independent in $\mathbf{R}^2$. Since $\dim(\mathbf{R}^2) = 2$, any two such vectors form a basis for $\mathbf{R}^2$. For example, the vectors $\mathbf{u}_1 = (2.0, 1.0)$ and $\mathbf{u}_2 = (1.0, 2.0)$ shown in Figure 3.2 are linearly independent and form a basis for $\mathbf{R}^2$. The vectors $\mathbf{v}_1 = (1.1, 1.0)$ and $\mathbf{v}_2 = (1.0, 1.1)$ shown in the same figure are also linearly independent and form another basis for $\mathbf{R}^2$.

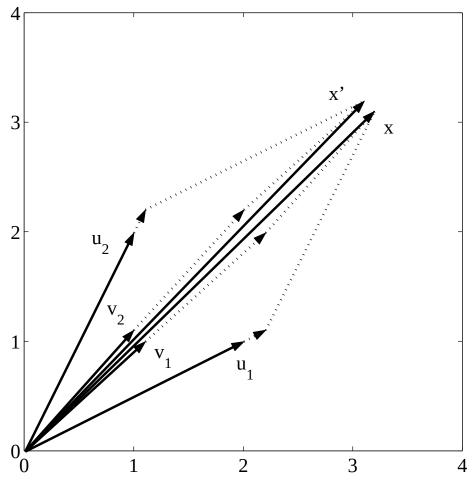

Figure 3.2: Two different bases for $\mathbf{R}^2$

Although $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2\}$ are both bases, they are quite different from a computational point of view. Consider two vectors $\mathbf{x} = (3.2, 3.1)$ and $\mathbf{x}' = (3.1, 3.2)$ which represent two close points in $\mathbf{R}^2$. We expect that when we express them in terms of a basis, then their corresponding coefficients multiplying the basis vectors should also be close. This is indeed the case for $\mathbf{U}$, where

$$\begin{aligned} \mathbf{x} &= 1.1\,\mathbf{u}_1 + 1.0\,\mathbf{u}_2 \\ \mathbf{x}' &= 1.0\,\mathbf{u}_1 + 1.1\,\mathbf{u}_2 \end{aligned}$$

On the other hand, when $\mathbf{x}$ and $\mathbf{x}'$ are expressed in terms of $\mathbf{V}$ as

$$\begin{aligned} \mathbf{x} &= 2.0\,\mathbf{v}_1 + 1.0\,\mathbf{v}_2 \\ \mathbf{x}' &= 1.0\,\mathbf{v}_1 + 2.0\,\mathbf{v}_2 \end{aligned}$$

their corresponding coefficients differ greatly.

To explain the situation we observe that finding the coefficients c_1 and c_2 of a given vector $\mathbf{x} = (x, y) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ is equivalent to solving the linear system

$$\begin{bmatrix} 1.1 & 1.0 \\ 1.0 & 1.1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Since this system is ill-conditioned, its solution (the coefficients c_1 and c_2) are very sensitive to small changes in x and y. The ill-conditioning of the system results from $\mathbf{v}_1$ and $\mathbf{v}_2$ being very much aligned with each other. We can say that they are closer to being linearly dependent than $\mathbf{u}_1$ and $\mathbf{u}_2$ are.[8]

Example 3.21

Any $A \in \mathsf{R}^{2\times 2}$ can be expressed as

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

Hence the set $\mathbf{M} = \{M_{11}, M_{12}, M_{21}, M_{22}\}$ spans $\mathsf{R}^{2\times 2}$. Since it is also linearly independent, it is a basis for $\mathsf{R}^{2\times 2}$. (The same conclusion can also be reached by observing that the coefficients of M_{ij} in the above expression are uniquely determined by A.) Therefore, $\dim(\mathsf{R}^{2\times 2}) = 4$.

The set

$$\mathsf{R}_s^{2\times 2} = \{\, S \in \mathsf{R}^{2\times 2} \mid S \text{ is symmetric} \,\}$$

is a subspace of $\mathsf{R}^{2\times 2}$. The matrices

$$S_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

form a basis for $\mathsf{R}_s^{2\times 2}$. Hence $\dim(\mathsf{R}_s^{2\times 2}) = 3$.

In general, $\dim(\mathsf{F}^{m\times n}) = mn$, and the set

$$\{\, M_{ij} \mid 1 \le i \le m, 1 \le j \le n \,\}$$

where $M_{ij} \in \mathsf{F}^{m\times n}$ consists of all 0's except a single 1 in the (i,j)th position, is a basis for $\mathsf{F}^{m\times n}$.

The following corollary is useful in constructing a basis for a vector space.

Corollary 3.1.2 *Let* $\dim(\mathbf{X}) = n$.

a) Any spanning set containing $m > n$ vectors can be reduced to a basis by deleting $m - n$ vectors from the set.

b) Any linearly independent set containing $k < n$ vectors can be completed to a basis by including $n - k$ more vectors into the set.

Proof

a) Let $\mathbf{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_m\}, m > n$, be a spanning set. By Corollary 3.1(a), it must be linearly dependent, and therefore, one of its vectors can be expressed in terms of the others. Deleting that vector from $\mathbf{R}$ reduces the number of vectors by one without destroying the spanning property. Continuing this process we finally obtain a subset of $\mathbf{R}$ which contains exactly n vectors and spans $\mathbf{X}$. By Corollary 3.1(e), it is a basis.

[8] We will mention about a measure of linear independence in Chapter 7.

The process of reducing $\mathbf{R}$ to a basis can be summarized by an algorithm:

$\mathbf{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_m\}$
For $i = m : 1$
  If span $(\mathbf{R} - \{\mathbf{r}_i\}) = \mathbf{X}$, $\mathbf{R} = \mathbf{R} - \{\mathbf{r}_i\}$
End

b) Let $\mathbf{R}_1 = \{\mathbf{r}_1, \ldots, \mathbf{r}_k\}, k < n$, be a linearly independent set, and let $\{\mathbf{r}_{k+1}, \ldots, \mathbf{r}_{k+n}\}$ be any basis for $\mathbf{X}$. Then $\mathbf{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_k, \mathbf{r}_{k+1}, \ldots, \mathbf{r}_{k+n}\}$ is a spanning set with $m = k + n$ elements. Application of the algorithm in part (a) to $\mathbf{R}$ reduces it to a basis which includes the first k vectors. Details are worked out in Exercise 3.14.

Example 3.22

Consider Example 3.10 again. The linearly independent set $\{\mathbf{r}_1, \mathbf{r}_2\}$ can be completed to a basis by adding $\mathbf{r}_3$ or $\mathbf{r}_4$. The spanning set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$ can be reduced to a basis by deleting $\mathbf{r}_2$, or $\mathbf{r}_3$, or $\mathbf{r}_4$.

3.3.2 Representation of Vectors With Respect to A Basis

Let $\dim(\mathbf{X}) = n$, and let $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)$ be an ordered basis for $\mathbf{X}$. Then any vector $\mathbf{x} \in \mathbf{X}$ can be expressed in terms of the basis vectors as

$$\mathbf{x} = \alpha_1 \mathbf{r}_1 + \cdots + \alpha_n \mathbf{r}_n$$

for some unique scalars $\alpha_i, i = 1, \ldots, n$. The column vector

$$\boldsymbol{\alpha} = \mathrm{col}\,[\,\alpha_1, \alpha_2, \ldots, \alpha_n\,] \in \mathsf{F}^{n\times 1}$$

is called the **representation** of $\mathbf{x}$ with respect to the basis $\mathbf{R}$. This way we establish a one-to-one correspondence between the vectors of $\mathbf{X}$ and the $n \times 1$ vectors of $\mathsf{F}^{n\times 1}$.[9]

Example 3.23

Consider the basis $\mathbf{E}$ of $\mathsf{R}^{3\times 1}$ in Example 3.17. From (3.5) we observe that the representation of a vector $\mathbf{x} = \mathrm{col}\,[\,x_1, x_2, x_3\,]$ with respect to $\mathbf{E}$ is itself. That is why $\mathbf{E}$ is called the canonical basis for $\mathsf{R}^{3\times 1}$.

Now consider the basis $\mathbf{R}$ in the same example. From (3.6), the representation of $\mathbf{x} = \mathrm{col}\,[\,x_1, x_2, x_3\,]$ with respect to $\mathbf{R}$ is obtained as

$$\boldsymbol{\alpha} = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \\ x_3 \end{bmatrix}$$

Example 3.24

The set $(\mathbf{r}_2, \mathbf{r}_3)$ in Example 3.10 is a basis for the subspace $\mathbf{U}$ in Example 3.6. The representation of $\mathbf{u} = (x, y, x-y) \in \mathbf{U}$ with respect to this basis is obtained by expressing $\mathbf{u}$ in terms of $\mathbf{r}_2$ and $\mathbf{r}_3$ as

$$\mathbf{u} = x\,\mathbf{r}_2 + y\,\mathbf{r}_3$$

[9]Note that although α_i are unique, their locations in $\boldsymbol{\alpha}$ depend on the ordering of the basis vectors. To guarantee that every vector has a unique column representation and that every colum represents a unique vector, it is necessary to associate an order with a basis. For this reson, from now on, whenever we deal with representations of vectors with respect to a basis, we will assume that the basis is ordered.

which gives

$$\boldsymbol{\alpha} = \begin{bmatrix} x \\ y \end{bmatrix}$$

The set $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2)$ in Example 3.19 is also a basis for $\mathbf{U}$. The representation of $\mathbf{u} = (x, y, x - y)$ with respect to $\mathbf{S}$ is obtained by expressing $\mathbf{u}$ in terms of $\mathbf{s}_1$ and $\mathbf{s}_2$ as

$$\mathbf{u} = 2y\,\mathbf{s}_1 + (x - y)\,\mathbf{s}_2$$

to be

$$\boldsymbol{\beta} = \begin{bmatrix} 2y \\ x - y \end{bmatrix}$$

Note that $\dim(\mathbf{U}) = 2$, and therefore, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbf{R}^{2\times 1}$.

* Example 3.25

Let $\mathbf{D}_N = (0, 1, \ldots, N-1)$, and consider the vector space $\mathcal{F}(\mathbf{D}_N, \mathbf{C})$ of complex-valued functions defined on $\mathbf{D}_N$.

Let the functions $e_p \in \mathcal{F}(\mathbf{D}_N, \mathbf{C})$ be defined as

$$e_p[k] = \begin{cases} 1, & k = p \\ 0, & k \neq p \end{cases}, \quad p = 0, 1, \ldots, N-1$$

Then any $f \in \mathcal{F}(\mathbf{D}_N, \mathbf{C})$ can be expressed in terms of e_p uniquely as

$$f = \sum_{p=0}^{N-1} a_p e_p\,, \quad a_p = f[p]\,, \quad p = 0, 1, \ldots, N-1$$

because

$$\Big(\sum_{p=0}^{N-1} a_p e_p\Big)[k] = \sum_{p=0}^{N-1} a_p e_p[k] = \sum_{p=0}^{N-1} f[p] e_p[k] = f[k]\,, \quad k \in \mathbf{D}_N$$

Hence $(e_0, e_1, \ldots, e_{N-1})$ is a basis for $\mathcal{F}(\mathbf{D}_N, \mathbf{C})$, and therefore $\dim(\mathcal{F}(\mathbf{D}_N, \mathbf{C})) = N$. From the expression above it also follows that the representation of f with respect to this basis is the column vector

$$\mathbf{f} = \mathrm{col}\,[a_0, a_1, \ldots, a_{N-1}] = \mathrm{col}\,[f[0], f[1], \ldots, f[N-1]]$$

There is nothing surprising about this result. $\mathcal{F}(\mathbf{D}_N, \mathbf{C})$ is essentially the same as $\mathbf{C}^{N\times 1}$, and a function $f \in \mathcal{F}(\mathbf{D}_N, \mathbf{C})$ is the same as the column vector $\mathbf{f}$. The functions $e_0, e_1, \ldots, e_{N-1}$ correspond to the canonical basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ of $\mathbf{C}^{N\times 1}$. That is why the representation of a function with respect to (e_p) is a column vector consisting of the values of f at $k = 0, 1, \ldots, N-1$.

Things become more interesting when we consider another basis for $\mathcal{F}(\mathbf{D}_N, \mathbf{C})$. Let

$$\phi_p[k] = e^{ip\frac{2\pi}{N}k}\,, \quad p = 0, 1, \ldots, N-1$$

It can be shown (see Exercise 3.20) that any function $f \in \mathcal{F}(\mathbf{D}_N, \mathbf{C})$ can be expressed in terms of ϕ_p uniquely as

$$f = \sum_{p=0}^{N-1} c_p \phi_p\,, \quad c_p = \frac{1}{N} \sum_{k=0}^{N-1} f[k] \phi_p^*[k]\,, \quad p = 0, 1, \ldots, N-1 \tag{3.7}$$

Hence $(\phi_0, \phi_1, \ldots, \phi_{N-1})$ is also a basis for $\mathcal{F}(\mathbf{D}_N, \mathbf{C})$, and the representation of f with respect to (ϕ_p) is

$$\mathbf{F} = \mathrm{col}\,[\,c_0, c_1, \ldots, c_{N-1}\,]$$

The representation of f as a linear combination of ϕ_p is known as the **discrete Fourier series** of f, and the coefficients c_p as the discrete Fourier coefficients of f.

As a specific example, suppose $N = 4$. Then the basis functions ϕ_p have the values tabulated below.

	$\phi_0[k]$	$\phi_1[k]$	$\phi_2[k]$	$\phi_3[k]$
$k=0$	1	1	1	1
$k=1$	1	i	-1	$-i$
$k=2$	1	-1	1	-1
$k=3$	1	$-i$	-1	i

Let

$$f[k] = \begin{cases} 2, & k = 0 \\ 4, & k = 1 \\ -2, & k = 2 \\ 0, & k = 3 \end{cases}$$

Then the discrete Fourier coefficients of f are computed as

$$\begin{aligned} c_0 &= \tfrac{1}{4}(2 + 4 - 2 + 0) &&= 1 \\ c_1 &= \tfrac{1}{4}(2 - 4i + 2 + 0) &&= 1 - i \\ c_2 &= \tfrac{1}{4}(2 - 4 - 2 + 0) &&= -1 \\ c_3 &= \tfrac{1}{4}(2 + 4i + 2 + 0) &&= 1 + i \end{aligned}$$

Hence the discrete Fourier series of f is

$$f = \phi_0 + (1 - i)\phi_1 - \phi_2 + (1 + i)\phi_3$$

and the representation of f with respect to (ϕ_p) is

$$\mathbf{F} = \begin{bmatrix} 1 \\ 1 - i \\ -1 \\ 1 + i \end{bmatrix}$$

From the examples above we observe that although the representation of a vector is unique with respect to a given basis, it has a different (but still unique) representation with respect to another basis. We now investigate how different representations of the same vector with respect to different bases are related.

Let $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$ and $\mathbf{R}' = (\mathbf{r}'_1, \mathbf{r}'_2, \ldots, \mathbf{r}'_n)$ be two ordered bases for $\mathbf{X}$.[10] Suppose that a vector $\mathbf{x}$ has the representations

$$\boldsymbol{\alpha} = \mathrm{col}\,[\,\alpha_1, \ldots, \alpha_n\,] \quad \text{and} \quad \boldsymbol{\alpha}' = \mathrm{col}\,[\,\alpha'_1, \ldots, \alpha'_n\,]$$

[10] Keep in mind that even when $\mathbf{R}$ and $\mathbf{R}'$ contain exactly the same vectors, if their orderings are different then $\mathbf{R}$ and $\mathbf{R}'$ are different.

with respect to $\mathbf{R}$ and $\mathbf{R}'$. That is,

$$\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{r}_j = \sum_{i=1}^{n} \alpha'_i \mathbf{r}'_i$$

Let the jth basis vector $\mathbf{r}_j$ be expressed in terms of the vectors of $\mathbf{R}'$ as

$$\mathbf{r}_j = \sum_{i=1}^{n} q_{ij} \mathbf{r}'_i \,, \quad j = 1, \ldots, n$$

so that it has a representation

$$\mathbf{q}_j = \mathrm{col}\,[\, q_{1j}, \ldots, q_{nj} \,] \,, \quad j = 1, \ldots, n$$

with respect to $\mathbf{R}'$. Then

$$\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{r}_j = \sum_{j=1}^{n} \alpha_j \,(\sum_{i=1}^{n} q_{ij} \mathbf{r}'_i \,) = \sum_{i=1}^{n} (\sum_{j=1}^{n} q_{ij} \alpha_j \,)\, \mathbf{r}'_i = \sum_{i=1}^{n} \alpha'_i \mathbf{r}'_i$$

By uniqueness of the representation of $\mathbf{x}$ with respect to $\mathbf{R}'$, we have

$$\alpha'_i = \sum_{j=1}^{n} q_{ij} \alpha_j \,, \quad i = 1, \ldots, n$$

By expressing these equalities in matrix form, we observe that the representations $\boldsymbol{\alpha}'$ and $\boldsymbol{\alpha}$ are related as

$$\boldsymbol{\alpha}' = Q\, \boldsymbol{\alpha}$$

The matrix

$$Q = [\, \mathbf{q}_1 \;\; \mathbf{q}_2 \;\; \cdots \;\; \mathbf{q}_n \,] = [\, q_{ij} \,]_{n \times n}$$

which is defined uniquely by $\mathbf{R}$ and $\mathbf{R}'$, is called the **matrix of change–of–basis** from $\mathbf{R}$ to $\mathbf{R}'$.

Now interchange the roles of the bases $\mathbf{R}$ and $\mathbf{R}'$. Let the jth basis vector $\mathbf{r}'_j$ be expressed in terms of the vectors of $\mathbf{R}$ as

$$\mathbf{r}'_j = \sum_{i=1}^{n} p_{ij} \mathbf{r}_i \,, \quad j = 1, \ldots, n$$

so that it has a representation

$$\mathbf{p}_j = \mathrm{col}\,[\, p_{1j}, \ldots, p_{nj} \,] \,, \quad j = 1, \ldots, n$$

with respect to $\mathbf{R}$. Defining

$$P = [\, \mathbf{p}_1 \;\; \mathbf{p}_2 \;\; \cdots \;\; \mathbf{p}_n \,] = [\, p_{ij} \,]_{n \times n}$$

to be the matrix of change–of–basis from $\mathbf{R}'$ to $\mathbf{R}$, we get

$$\boldsymbol{\alpha} = P\, \boldsymbol{\alpha}'$$

The reader might suspect that the matrices Q and P are related. Indeed, since $\boldsymbol{\alpha} = P\boldsymbol{\alpha}' = PQ\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}' = Q\boldsymbol{\alpha} = QP\boldsymbol{\alpha}'$ for all pairs $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathsf{R}^{n \times 1}$, we must have

$$PQ = QP = I_n$$

We will investigate such matrices in Chapter 4.

Example 3.26

Consider Example 3.17. Expressing the vectors of the canonical basis $\mathbf{E}$ in terms of $\mathbf{R}$ as

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{r}_1 \\ \mathbf{e}_2 &= -\mathbf{r}_1 + \mathbf{r}_2 \\ \mathbf{e}_3 &= -\mathbf{r}_2 + \mathbf{r}_3 \end{aligned}$$

we obtain the matrix of change–of–basis from $\mathbf{E}$ to $\mathbf{R}$ as

$$Q = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence the representation of a vector $\mathbf{x} = \text{col}\,[\,x_1, x_2, x_3\,]$ with respect to $\mathbf{R}$ is related to its canonical representation $\mathbf{x}$ as

$$\boldsymbol{\alpha} = Q\mathbf{x} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \\ x_3 \end{bmatrix}$$

which is the same as in Example 3.23.

Since the representations of $\mathbf{r}_j$ with respect to $\mathbf{E}$ are themselves, the matrix of change–of–basis from $\mathbf{R}$ to $\mathbf{E}$ is easily obtained as

$$P = [\,\mathbf{r}_1 \;\; \mathbf{r}_2 \;\; \mathbf{r}_3\,] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

If $\mathbf{x}$ has a representation

$$\boldsymbol{\alpha} = \text{col}\,[\,a, b, c\,]$$

with respect to $\mathbf{R}$, then

$$\mathbf{x} = a\mathbf{r}_1 + b\mathbf{r}_2 + c\mathbf{r}_3 = \begin{bmatrix} a+b+c \\ b+c \\ c \end{bmatrix} = P\boldsymbol{\alpha}$$

The reader should verify that $QP = PQ = I$.

3.4 Linear Transformations

Let $\mathbf{X}$ and $\mathbf{Y}$ be vector spaces over the same field F. A mapping $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ is called a **linear transformation** if for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}$ and for all $c_1, c_2 \in \mathsf{F}$

$$\mathcal{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathcal{A}(\mathbf{x}_1) + c_2\mathcal{A}(\mathbf{x}_2) \tag{3.8}$$

(3.8) is equivalent to

$$\mathcal{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathcal{A}(\mathbf{x}_1) + \mathcal{A}(\mathbf{x}_2) \tag{3.9}$$

and

$$\mathcal{A}(c\mathbf{x}) = c\mathcal{A}(\mathbf{x}) \tag{3.10}$$

which are known as **superposition** and **homogeneity**, respectively. (3.9) follows from (3.8) on choosing $c_1 = c_2 = 1$, and (3.10) on choosing $c_1 = 1, c_2 = 0$ and $\mathbf{x}_1 = \mathbf{x}$. Conversely, (3.9) and (3.10) imply that

$$\mathcal{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = \mathcal{A}(c_1\mathbf{x}_1) + \mathcal{A}(c_2\mathbf{x}_2) = c_1\mathcal{A}(\mathbf{x}_1) + c_2\mathcal{A}(\mathbf{x}_2)$$

$\mathbf{X}$ and $\mathbf{Y}$ are the **domain** and the **codomain** of $\mathcal{A}$. A linear transformation from a vector space $\mathbf{X}$ into itself is called a **linear operator** on $\mathbf{X}$.

If $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ is a linear transformation then

$$\mathcal{A}(\mathbf{0}_x) = \mathbf{0}_y$$

which follows from (3.8) on taking $c_1 = c_2 = 0$.

Example 3.27

The zero mapping $\mathcal{O} : \mathbf{X} \to \mathbf{Y}$ defined as

$$\mathcal{O}(\mathbf{x}) = \mathbf{0}_y \quad \text{for all } \mathbf{x} \in \mathbf{X}$$

is a linear transformation that satisfies (3.8) trivially.

The identity mapping $\mathcal{I} : \mathbf{X} \to \mathbf{X}$ defined as

$$\mathcal{I}(\mathbf{x}) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbf{X}$$

is also a linear transformation, because

$$\mathcal{I}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = c_1\mathcal{I}(\mathbf{x}_1) + c_2\mathcal{I}(\mathbf{x}_2)$$

Hence $\mathcal{I}$ is a linear operator on $\mathcal{X}$.

Example 3.28

The mapping $\mathcal{A} : \mathsf{R}^3 \to \mathsf{R}^2$ defined as

$$\mathcal{A}(x_1, x_2, x_3) = (x_1 + x_2, x_2 - x_3)$$

is a linear transformation. For $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$ and $c, d \in \mathsf{F}$

$$\begin{aligned} \mathcal{A}(c\mathbf{u} + d\mathbf{v}) &= \mathcal{A}(cu_1 + dv_1, cu_2 + dv_2, cu_3 + dv_3) \\ &= (cu_1 + dv_1 + cu_2 + dv_2, cu_2 + dv_2 - cu_3 - dv_3) \\ &= c(u_1 + u_2, u_2 - u_3) + d(v_1 + v_2, v_2 - v_3) \\ &= c\mathcal{A}(\mathbf{u}) + d\mathcal{A}(\mathbf{v}) \end{aligned}$$

However, none of the mappings

$$\begin{aligned} \mathcal{B}(x_1, x_2, x_3) &= (x_1 + x_3, x_2 + 1) \\ \mathcal{C}(x_1, x_2, x_3) &= (x_1 + x_3, x_2^2) \\ \mathcal{D}(x_1, x_2, x_3) &= (x_1 x_3, x_2) \end{aligned}$$

is a linear transformation. $\mathcal{B}$ is not linear simply because $\mathcal{B}(\mathbf{0}) \neq \mathbf{0}$. The reader is urged to explain why $\mathcal{C}$ and $\mathcal{D}$ are not linear.

Example 3.29

Let $\mathcal{A} : \mathbf{F}^{n\times 1} \rightarrow \mathbf{F}^{m\times 1}$ be defined as $\mathcal{A}(\mathbf{x}) = A\mathbf{x}$, where A is an $m \times n$ matrix with elements from $\mathbf{F}$. Since

$$A(a\mathbf{x} + b\mathbf{y}) = aA\mathbf{x} + bA\mathbf{y}$$

$\mathcal{A}$ is a linear transformation. This example shows that every matrix defines a linear transformation. Thus a linear transformation defined by an $n \times n$ matrix with elements from $\mathbf{F}$ is a linear operator on $\mathbf{F}^{n\times 1}$.

Example 3.30

Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be vector spaces over the same field, and let $\mathcal{A} : \mathbf{X} \rightarrow \mathbf{Y}$ and $\mathcal{B} : \mathbf{Y} \rightarrow \mathbf{Z}$ be linear transformations. Then the compound mapping $\mathcal{C} : \mathbf{X} \rightarrow \mathbf{Z}$ defined as

$$\mathcal{C}(\mathbf{x}) = (\mathcal{B} \circ \mathcal{A})(\mathbf{x}) = \mathcal{B}(\mathcal{A}(\mathbf{x}))$$

is also a linear transformation, because

$$\begin{aligned} \mathcal{C}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) &= \mathcal{B}(\mathcal{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2)) \\ &= \mathcal{B}(c_1\mathcal{A}(\mathbf{x}_1) + c_2\mathcal{A}(\mathbf{x}_2)) \\ &= c_1\mathcal{B}(\mathcal{A}(\mathbf{x}_1)) + c_2\mathcal{B}(\mathcal{A}(\mathbf{x}_2)) \\ &= c_1\mathcal{C}(\mathbf{x}_1) + c_2\mathcal{C}(\mathbf{x}_2) \end{aligned}$$

* Example 3.31

In Section 2.5 we defined the differential operator D as a mapping from a set of functions into itself such that $D(f) = f'$. We now take a closer look at D.

Recall from Example 3.8 that

$$\mathcal{C}_0(\mathcal{I}, \mathbf{R}) \supset \mathcal{C}_1(\mathcal{I}, \mathbf{R}) \supset \cdots \supset \mathcal{C}_\infty(\mathcal{I}, \mathbf{R})$$

are subspaces of $\mathcal{F}(\mathcal{I}, \mathbf{R})$. Also, if $f \in \mathcal{C}_m(\mathcal{I}, \mathbf{R})$ then

$$f' \in \mathcal{C}_{m-1}(\mathcal{I}, \mathbf{R}),\ f'' \in \mathcal{C}_{m-2}(\mathcal{I}, \mathbf{R}),\ \ldots,\ f^{(m)} \in \mathcal{C}_0(\mathcal{I}, \mathbf{R})$$

Hence, for any $m > 1$, the differential operator D is a mapping from $\mathcal{C}_m(\mathcal{I}, \mathbf{R})$ into $\mathcal{C}_{m-1}(\mathcal{I}, \mathbf{R})$, and therefore, into $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$. The property in (2.34) implies that D is a linear transformation. Then, as discussed in Example 3.30, the operator D^2 that is defined in terms of D as

$$D^2(f) = (D \circ D)(f) = D(D(f)) = D(f') = f''$$

is also a linear transformation. Consequently, each $D^k, k = 1, \ldots, n \leq m$, which can be defined recursively as

$$D^k(f) = (D \circ D^{k-1})(f) = D(D^{k-1}(f)) = D(f^{(k-1)}) = f^{(k)}$$

is a linear transformations from $\mathcal{C}_m(\mathcal{I}, \mathbf{R})$ into $\mathcal{C}_{m-k}(\mathcal{I}, \mathbf{R})$, and therefore, into $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$.

Finally, an nth order linear differential operator $L(D)$ is a linear transformation from $\mathcal{C}_n(\mathcal{I}, \mathbf{R})$ into $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$ (see Exercise 3.31). In fact, this is precisely the reason for calling $L(D)$ a linear operator.

* **Example 3.32**

Recall that a vector $\mathbf{x} \in \mathsf{R}^{n\times 1}$ can also be viewed as a function $f \in \mathcal{F}(\mathsf{n}, \mathsf{R})$. The linear operator $\mathcal{A} : \mathsf{R}^{n\times 1} \to \mathsf{R}^{n\times 1}$ defined by a matrix $A \in \mathsf{R}^{n\times n}$ can similarly be interpreted as a a linear operator $\mathcal{A} : \mathcal{F}(\mathsf{n}, \mathsf{R}) \to \mathcal{F}(\mathsf{n}, \mathsf{R})$ such that the image $g = \mathcal{A}(f)$ of a function f is defined pointwise as

$$g[p] = \sum_{q=1}^{n} a_{pq} f[q] \,, \quad p \in \mathsf{n}$$

Now consider the vector space $\mathcal{F}(\mathsf{Z}, \mathsf{R})$ of infinite sequences. We can define a linear operator $\mathcal{H}$ on $\mathcal{F}(\mathsf{Z}, \mathsf{R})$ such that if $g = \mathcal{H}(f)$ then

$$g[p] = \sum_{q=-\infty}^{\infty} h[p, q] f[q] \,, \quad p \in \mathsf{Z} \tag{3.11}$$

where $h[p, q] \in \mathsf{R}$.[11] We can think of $\mathcal{H}$ as defined by an infinitely large matrix H with elements $h[p, q]$ such that

$$\begin{bmatrix} \vdots \\ g[-1] \\ g[\ 0\ ] \\ g[\ 1\ ] \\ \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & \vdots & \vdots & \\ \cdots & h[-1,-1] & h[-1,0] & h[-1,1] & \cdots \\ \cdots & h[\ 0,-1] & h[\ 0,0] & h[\ 0,1] & \cdots \\ \cdots & h[\ 1,-1] & h[\ 1,0] & h[\ 1,1] & \cdots \\ & \vdots & \vdots & \vdots & \end{bmatrix} = \begin{bmatrix} \vdots \\ f[-1] \\ f[\ 0\ ] \\ f[\ 1\ ] \\ \vdots \end{bmatrix}$$

Let us go one step further, and consider the vector space $\mathcal{F}(\mathsf{R}, \mathsf{R})$ of real-valued functions defined on R. Now the domain of the functions is a continuum, and the infinite summation in (3.11) must be replaced with an integral: We can then define a linear operator $\mathcal{H}$ on $\mathcal{F}(\mathsf{R}, \mathsf{R})$ such that $g = \mathcal{H}(f)$ is characterized by

$$g(t) = \int_{-\infty}^{\infty} h(t, \tau) f(\tau)\, d\tau \,, \quad t \in \mathsf{R} \tag{3.12}$$

In the special cases when $h[p, q] = h[p - q]$ and $h(t, \tau) = h(t - \tau)$, the operators defined by (3.11) and (3.12) are known as **convolution**, and are widely used in system analysis.

3.4.1 Matrix Representation of Linear Transformations

Let $\dim(\mathbf{X}) = n$, let $\dim(\mathbf{Y}) = m$, let $\mathbf{R} = (\,\mathbf{r}_1, \ldots, \mathbf{r}_n\,)$ and $\mathbf{S} = (\,\mathbf{s}_1, \ldots, \mathbf{s}_m\,)$ be ordered bases for $\mathbf{X}$ and $\mathbf{Y}$, and let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ be a linear transformation.

Consider $\mathcal{A}(\mathbf{r}_j)$. Since it is a vector in $\mathbf{Y}$, it has a unique representation $\mathbf{a}_j \in \mathsf{F}^{m\times 1}$ with respect to the basis $\mathbf{S}$. That is,

$$\mathcal{A}(\mathbf{r}_j) = \sum_{i=1}^{m} a_{ij} \mathbf{s}_i \,, \quad j = 1, \ldots, n$$

[11] Of course, this definition requires that the infinite series in (3.11) converges for all $p \in \mathsf{Z}$, which puts some restrictions not only on $h[p, q]$ but also on f. These technical difficulties can be worked out by restricting f to a subspace of $\mathcal{F}(\mathsf{Z}, \mathsf{R})$ and by choosing $h[p, q]$ suitably.

and

$$\mathbf{a}_j = \mathrm{col}\,[\, a_{1j}, a_{2j}, \ldots, a_{mj} \,]\,, \quad j = 1, \ldots, n$$

The $m \times n$ matrix

$$A = [\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n \,] = [\, a_{ij} \,]$$

which is uniquely defined by $\mathcal{A}$, $\mathbf{R}$ and $\mathbf{S}$, is called the **matrix representation of a linear transformation** of $\mathcal{A}$ with respect to the basis pair $(\mathbf{R}, \mathbf{S})$.

The significance of the matrix representation of $\mathcal{A}$ is that if $\mathbf{x} \in \mathbf{X}$ has a representation $\boldsymbol{\alpha} \in \mathsf{F}^{n\times 1}$ with respect to $\mathbf{R}$ and $\mathbf{y} = \mathcal{A}(\mathbf{x}) \in \mathbf{Y}$ has a representation $\boldsymbol{\beta} \in \mathsf{F}^{m\times 1}$ with respect to $\mathbf{S}$, then the two representations are related as $\boldsymbol{\beta} = A\boldsymbol{\alpha}$. To show this, suppose

$$\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{r}_j\,, \quad \mathbf{y} = \mathcal{A}(\mathbf{x}) = \sum_{i=1}^{m} \beta_i \mathbf{s}_i$$

Then

$$\mathbf{y} = \sum_{j=1}^{n} \alpha_j \mathcal{A}(\mathbf{r}_j) = \sum_{j=1}^{n} \alpha_j \,(\sum_{i=1}^{m} a_{ij}\mathbf{s}_i \,) = \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij}\alpha_j \,)\mathbf{s}_i = \sum_{i=1}^{m} \beta_i \mathbf{s}_i$$

Since the representation of $\mathbf{y}$ in terms of $\mathbf{s}_i$ is unique, we must have

$$\beta_i = \sum_{j=1}^{n} a_{ij}\alpha_j\,, \quad i = 1, \ldots, m$$

or in matrix form

$$\boldsymbol{\beta} = A\boldsymbol{\alpha}$$

In conclusion, not only does an $m \times n$ matrix define a linear transformation from $\mathsf{F}^{n\times 1}$ into $\mathsf{F}^{m\times 1}$, but also a linear transformation from an n-dimensional vector space $\mathbf{X}$ into an m-dimensional vector space $\mathbf{Y}$ can be represented by an $m \times n$ matrix once a pair of bases for $\mathbf{X}$ and $\mathbf{Y}$ are fixed.[12]

Like the column representation of a vector with respect to basis, a linear transformation has different representations with respect to different bases. If $\mathcal{A}$ has a representation A with respect to $(\mathbf{R}, \mathbf{S})$ and a representation A' with respect to $(\mathbf{R}', \mathbf{S}')$, then

$$A' = Q_y A P_x$$

where Q_y is the matrix of change–of–basis from $\mathbf{S}$ to $\mathbf{S}'$ in $\mathbf{Y}$, and P_x is the matrix of change–of–basis from $\mathbf{R}'$ to $\mathbf{R}$ in $\mathbf{X}$. This follows from the fact that if $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ are representations of $\mathbf{x}$ with respect to $\mathbf{R}$ and $\mathbf{R}'$, and $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$ are representations of $\mathbf{y} = \mathcal{A}(\mathbf{x})$ with respect to $\mathbf{S}$ and $\mathbf{S}'$, then

$$A'\boldsymbol{\alpha}' = \boldsymbol{\beta}' = Q_y\boldsymbol{\beta} = Q_y A\boldsymbol{\alpha} = Q_y A P_x \boldsymbol{\alpha}'$$

[12] The question of whether a similar result can be derived for linear transformations between infinite dimensional vector spaces is beyond the scope of this book.

The relations between the vectors of $\mathbf{X}$ and $\mathbf{Y}$ and their representations are summarized by the diagram in Figure 3.3. From the diagram it is clear that A, Q_yA, AP_x and Q_yAP_x all represent the same linear transformation, each with respect to a different pair of bases. Q_yA represents $\mathcal{A}$ with respect to $(\mathbf{R}, \mathbf{S}')$, and AP_x represents $\mathcal{A}$ with respect to $(\mathbf{R}', \mathbf{S})$.

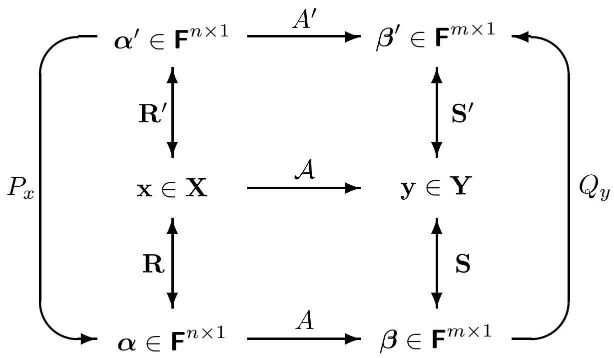

Figure 3.3: Matrix representation of a linear transformation

Example 3.33

Consider the linear transformation $\mathcal{A} : \mathsf{F}^{n\times 1} \to \mathsf{F}^{m\times 1}$ defined by a matrix $A^{m\times n} \in \mathsf{F}^{m\times n}$. Let $\mathbf{E}^n = (\, \mathbf{e}_1^n, \ldots, \mathbf{e}_n^n \,)$ and $\mathbf{E}^m = (\, \mathbf{e}_1^m, \ldots, \mathbf{e}_m^m \,)$ denote the canonical bases for $\mathsf{F}^{n\times 1}$ and $\mathsf{F}^{m\times 1}$, respectively. Then since $A\mathbf{e}_j^n = \mathbf{a}_j$ (the jth column of A), and since the column representation of $\mathbf{a}_j$ with respect to $\mathbf{E}^m$ is itself, it follows that the matrix representation of $\mathcal{A}$ with respect to the canonical bases of $\mathsf{F}^{n\times 1}$ and $\mathsf{F}^{m\times 1}$ is the matrix A itself.

Example 3.34

Consider the linear transformation in Example 3.28. If we choose the canonical bases $\mathbf{E}^3 = (\, \mathbf{e}_1^3, \mathbf{e}_2^3, \mathbf{e}_3^3 \,)$ and $\mathbf{E}^2 = (\, \mathbf{e}_1^2, \mathbf{e}_2^2 \,)$ for F^3 and F^2, then

$$\begin{array}{rclclcl} \mathcal{A}(\mathbf{e}_1^3) & = & \mathcal{A}(1,0,0) & = & (1,\ \ 0) & = & \mathbf{e}_1^2 \\ \mathcal{A}(\mathbf{e}_2^3) & = & \mathcal{A}(0,1,0) & = & (1,\ \ 1) & = & \mathbf{e}_1^2 + \mathbf{e}_2^2 \\ \mathcal{A}(\mathbf{e}_3^3) & = & \mathcal{A}(0,0,1) & = & (0,-1) & = & -\mathbf{e}_2^2 \end{array}$$

and the matrix representation of $\mathcal{A}$ is

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

Now suppose we choose

$$\mathbf{r}_1 = (1,0,0)\,, \quad \mathbf{r}_2 = (0,1,0)\,, \quad \mathbf{r}_3 = (-1,1,1)$$

as a basis for F^3, and

$$\mathbf{s}_1 = (1,0)\,, \quad \mathbf{s}_2 = (1,1)$$

as a basis for F^2. Then, since

$$\begin{array}{rclclcl} \mathcal{A}(\mathbf{r}_1) &=& \mathcal{A}(\ \ 1,0,0) &=& (1,0) &=& \mathbf{s}_1 \\ \mathcal{A}(\mathbf{r}_2) &=& \mathcal{A}(\ \ 0,1,0) &=& (1,1) &=& \mathbf{s}_2 \\ \mathcal{A}(\mathbf{r}_3) &=& \mathcal{A}(-1,1,1) &=& (0,0) &=& 0 \end{array}$$

$\mathcal{A}$ has the matrix representation

$$A' = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]$$

with respect to $(\mathbf{R},\mathbf{S})$.

Let us form the change–of–basis matrices Q_y and P_x. Since

$$\begin{array}{rcl} \mathbf{e}_1^2 &=& \mathbf{s}_1 \\ \mathbf{e}_2^2 &=& -\mathbf{s}_1+\mathbf{s}_2 \end{array}$$

the matrix of change–of–basis from $\mathbf{E}^2$ to $\mathbf{S}$ in F^2 is

$$Q_y = \left[\begin{array}{rr} 1 & -1 \\ 0 & 1 \end{array}\right]$$

The matrix of change–of–basis from $\mathbf{R}$ to $\mathbf{E}^3$ in F^3 is readily obtained as

$$P_x = \left[\begin{array}{rrr} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right]$$

The reader can easily verify that $A' = Q_y A P_x$.

In Example 3.30 we have seen that if $\mathcal{A}:\mathbf{X}\rightarrow\mathbf{Y}$ and $\mathcal{B}:\mathbf{Y}\rightarrow\mathbf{Z}$ are linear transformations, then the mapping $\mathcal{C}:\mathbf{X}\rightarrow\mathbf{Z}$ defined as

$$\mathcal{C}(\mathbf{x}) = (\mathcal{B}\circ\mathcal{A})(\mathbf{x}) = \mathcal{B}(\mathcal{A}(\mathbf{x}))$$

is also a linear transformation. In particular, if $\mathcal{A}:\mathsf{F}^{n\times 1}\rightarrow\mathsf{F}^{m\times 1}$ and $\mathcal{B}:\mathsf{F}^{m\times 1}\rightarrow\mathsf{F}^{p\times 1}$ are linear transformations defined as

$$\mathcal{A}(\mathbf{x}) = A\mathbf{x}\,,\quad \mathcal{B}(\mathbf{y}) = B\mathbf{y}$$

where A and B are $m\times n$ and $p\times m$ matrices then $\mathcal{C}$ is defined as

$$\mathcal{C}(\mathbf{x}) = \mathcal{B}(\mathcal{A}(\mathbf{x})) = \mathcal{B}(A\mathbf{x}) = BA\mathbf{x}$$

Conversely, if $\mathbf{X},\mathbf{Y}$ and $\mathbf{Z}$ are finite dimensional and $\mathcal{A}$ and $\mathcal{B}$ are represented by matrices A and B with respect to some fixed bases of $\mathbf{X},\mathbf{Y}$ and $\mathbf{Z}$, then $\mathcal{C}=\mathcal{B}\circ\mathcal{A}$ is represented by the matrix $C=BA$ with respect to the same bases (see Exercise 3.29). Thus a matrix product can be viewed as the representation of a linear transformation followed by another as illustrated in Figure 3.4.

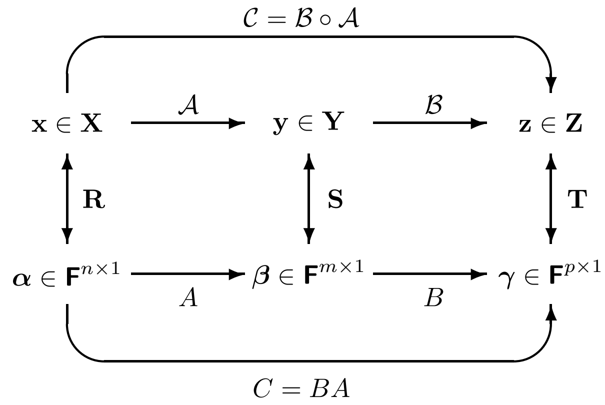

Figure 3.4: An interpretation matrix multiplication

3.4.2 Kernel and Image of a Linear Transformation

Let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ be a linear transformation. The set

$$\ker(\mathcal{A}) = \{\, \mathbf{x} \in \mathbf{X} \mid \mathcal{A}(\mathbf{x}) = \mathbf{0} \,\} \subset \mathbf{X}$$

is called the **kernel** of $\mathcal{A}$. Clearly, $\mathbf{0} \in \ker(\mathcal{A})$. Furthermore, if $\mathbf{x}_1, \mathbf{x}_2 \in \ker(\mathcal{A})$ then for any $c_1, c_2 \in \mathsf{F}$

$$\mathcal{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathcal{A}(\mathbf{x}_1) + c_2\mathcal{A}(\mathbf{x}_2) = c_1\mathbf{0} + c_2\mathbf{0} = \mathbf{0}$$

so that $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \in \ker(\mathcal{A})$. That is, $\ker(\mathcal{A})$ is closed under vector addition and scalar multiplication. Hence it is a subspace of $\mathbf{X}$, which is also called the **null space** of $\mathcal{A}$ and denoted by $\mathbf{N}(\mathcal{A})$. If it is finite dimensional, we define $\nu(\mathcal{A}) = \dim(\ker(\mathcal{A}))$ to be the **nullity** of $\mathcal{A}$.

The set

$$\operatorname{im}(\mathcal{A}) = \{\, \mathbf{y} \in \mathbf{Y} \mid y = \mathcal{A}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbf{X} \,\} \subset \mathbf{Y}$$

is called the **image** of $\mathcal{A}$. The reader can easily show that $\operatorname{im}(\mathcal{A})$ is a subspace of $\mathbf{Y}$, which is also called the **range space** of $\mathcal{A}$ and denoted as $\mathbf{R}(\mathcal{A})$. If it is finite dimensional, then we define $\rho(\mathcal{A}) = \dim(\operatorname{im}(\mathcal{A}))$ to be the **rank** of $\mathcal{A}$.

If $\mathcal{A} : \mathsf{F}^{n\times 1} \to \mathsf{F}^{m\times 1}$ is a linear transformation defined by an $m \times n$ matrix A, then we also use the notation $\ker(A)$ and $\operatorname{im}(A)$ to denote $\ker(\mathcal{A})$ and $\operatorname{im}(\mathcal{A})$.

Example 3.35

Consider the linear transformation $\mathcal{A} : \mathsf{R}^{4\times 1} \to \mathsf{R}^{3\times 1}$ defined by the matrix

$$A = \begin{bmatrix} 1 & 0 & -2 & -2 \\ 1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 3 \end{bmatrix} = [\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{a}_4 \,]$$

Then

$$\ker(\mathcal{A}) = \{\, \mathbf{x} \in \mathsf{R}^{4\times 1} \mid A\mathbf{x} = \mathbf{0} \,\}$$

that is, $\ker(\mathcal{A})$ is precisely the set of solutions of the homogeneous system $A\mathbf{x} = \mathbf{0}$. From the reduced row echelon form of A

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

we obtain two linearly independent solutions

$$\boldsymbol{\phi}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\phi}_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Hence $\ker(\mathcal{A}) = \operatorname{span}(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, and therefore, $\nu(\mathcal{A}) = 2$.

Clearly,

$$\operatorname{im}(\mathcal{A}) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$$

Performing elementary operations on the columns of A as

$$A \quad \begin{array}{c} 2\mathrm{C}_1 + \mathrm{C}_3 \rightarrow \mathrm{C}_3 \\ 2\mathrm{C}_1 + \mathrm{C}_4 \rightarrow \mathrm{R}_+ \\ \longrightarrow \end{array} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 3 \\ 0 & -1 & 1 & 3 \end{bmatrix}$$

$$\begin{array}{c} \mathrm{C}_2 + \mathrm{C}_3 \rightarrow \mathrm{C}_3 \\ 3\mathrm{C}_2 + \mathrm{C}_4 \rightarrow \mathrm{C}_4 \\ \longrightarrow \end{array} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{0} \;\; \mathbf{0}\,]$$

we determine that $\operatorname{im}(\mathcal{A}) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2)$, and hence $\rho(\mathcal{A}) = 2$.

Example 3.36

Let $\mathcal{A} : \mathbf{R}^{2\times 2} \rightarrow \mathbf{R}^{2\times 2}$ be defined as

$$\mathcal{A}(M) = M + M^t$$

It is easy to see that $\mathcal{A}$ is a linear transformation.

Since $M + M^t = O$ if and only if M is skew-symmetric,

$$\ker(\mathcal{A}) = \operatorname{span}\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right)$$

and hence $\nu(\mathcal{A}) = 1$.

Since $M + M^t$ is symmetric for any M,

$$\operatorname{im}(\mathcal{A}) = \mathbf{R}_s^{2\times 2}$$

where $\mathbf{R}_s^{2\times 2}$ is the subspace in Example 3.21. Hence $\rho(\mathcal{A}) = 3$.

* 3.4.3 Inverse Transformations

If $\ker(\mathcal{A}) = \{\,\mathbf{0}\,\}$ then to every $\mathbf{y} \in \operatorname{im}(\mathcal{A})$ there corresponds a unique $\mathbf{x} \in \mathbf{X}$ such that $\mathcal{A}(\mathbf{x}) = \mathbf{y}$, that is, $\mathcal{A}$ is one-to-one (see Exercise 3.38). It is then natural to expect that there exists a linear transformation $\hat{\mathcal{A}}_L : \mathbf{Y} \rightarrow \mathbf{X}$ that maps the image of every $\mathbf{x} \in \mathbf{X}$ back to $\mathbf{x}$ as illustrated in Figure 3.5. Such a linear transformation, if it exists, is called a **left inverse** of $\mathcal{A}$.[13]

In general, $\hat{\mathcal{A}}_L$ is not unique because of the arbitrariness in defining $\hat{\mathcal{A}}_L(\mathbf{y})$ when $\mathbf{y} \notin \operatorname{im}(\mathcal{A})$. However, by the very definition, it has the property that

$$\hat{\mathcal{A}}_L(\mathcal{A}(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbf{X}$$

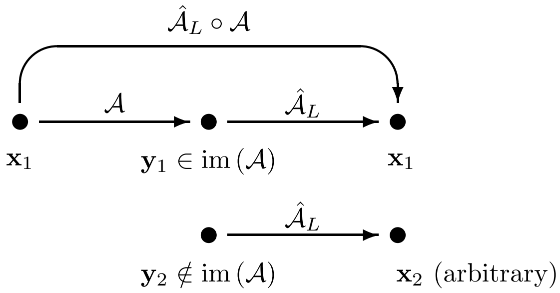

Figure 3.5: Left inverse of a linear transformation

Example 3.37

Let $\mathcal{A} : \mathbf{R}^{2\times 1} \to \mathbf{R}^{3\times 1}$ be defined by the matrix

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

It can easily be verified that the only solution of $A\mathbf{x} = \mathbf{0}$ is the trivial solution $\mathbf{x} = \mathbf{0}$, that is, $\ker(A) = \{\mathbf{0}\}$. Let

$$\hat{A}_L = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Then the mapping $\hat{\mathcal{A}}_L : \mathbf{R}^{3\times 1} \to \mathbf{R}^{2\times 1}$ defined by $\hat{A}_L$ is a left inverse of $\mathcal{A}$, because $\hat{A}_L A = I$ so that

$$\hat{\mathcal{A}}_L(\mathcal{A}(\mathbf{x})) = \hat{A}_L A\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbf{R}^{2\times 1}$$

The reader can verify that the matrix

$$\hat{A}'_L = \begin{bmatrix} -1 & 1 & -2 \\ 2 & -1 & 2 \end{bmatrix}$$

also defines a left inverse of $\mathcal{A}$.

[13]The proof of existence of $\hat{\mathcal{A}}_L$ in the general case is beyond the scope of this book. Left inverse of a linear transformation defined by a matrix is studied in Chepter 4.

Now suppose that $\text{im}\,(\mathcal{A}) = \mathbf{Y}$. Then for every $\mathbf{y} \in \mathbf{Y}$ there exists an $\mathbf{x} \in \mathbf{X}$, not necessarily unique, such that $\mathcal{A}(\mathbf{x}) = \mathbf{y}$, that is, $\mathcal{A}$ is onto. We can then define a linear transformation $\hat{\mathcal{A}}_R : \mathbf{Y} \to \mathbf{X}$ such that $\hat{\mathcal{A}}_R(\mathbf{y}) = \mathbf{x}$, where $\mathbf{x}$ is any fixed vector that satisfies $\mathcal{A}(\mathbf{x}) = \mathbf{y}$. Because of the arbitrariness in choosing $\mathbf{x}$ (if there are more than one $\mathbf{x}$ that satisfy $\mathcal{A}(\mathbf{x}) = \mathbf{y}$), $\hat{\mathcal{A}}_R$ is not unique either. However, it has the property that

$$\mathcal{A}(\hat{\mathcal{A}}_R(\mathbf{y})) = \mathbf{y}$$

for all $\mathbf{y} \in \mathbf{Y}$. Thus $\hat{\mathcal{A}}_R$ maps every $\mathbf{y} \in \mathbf{Y}$ to a vector whose image is $\mathbf{y}$ as illustrated in Figure 3.6, and is called a **right inverse** of $\mathcal{A}$.[14]

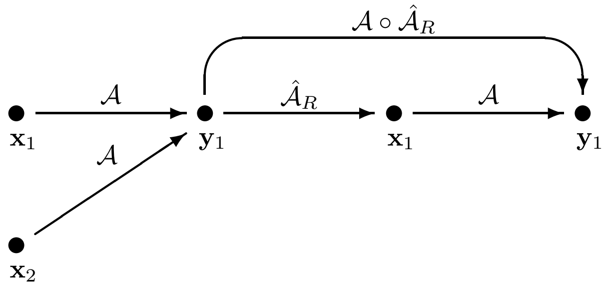

Figure 3.6: Right inverse of a linear transformation

Example 3.38

Let $\mathcal{B} : \mathbf{R}^{3\times 1} \to \mathbf{R}^{2\times 1}$ be defined by the matrix

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

Since $r(B) = 2$, the equation $B\mathbf{x} = \mathbf{y}$ is consistent for all $\mathbf{y}$, that is, $\text{im}\,(B) = \mathbf{R}^{2\times 1}$. Let

$$\hat{B}_R = \begin{bmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}$$

Then the mapping $\hat{\mathcal{B}}_R : \mathbf{R}^{2\times 1} \to \mathbf{R}^{3\times 1}$ defined by $\hat{B}_R$ is a right inverse of $\mathcal{B}$, because $B\hat{B}_R = I$, so that

$$\mathcal{B}(\hat{\mathcal{B}}_R(\mathbf{y})) = B\hat{B}_R\mathbf{y} = \mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbf{R}^{2\times 1}$$

As an exercise the reader may try to find a different right inverse of $\mathcal{B}$.

[14] Right inverse of a linear transformation defined by a matrix is also studied in Chepter 4.

Finally, suppose that $\ker(\mathcal{A}) = \{\mathbf{0}\}$ and also $\mathrm{im}(\mathcal{A}) = \mathbf{Y}$. Then $\mathcal{A}$ has both a left inverse $\hat{\mathcal{A}}_L$ and a right inverse $\hat{\mathcal{A}}_R$. Moreover, $\hat{\mathcal{A}}_L$ and $\hat{\mathcal{A}}_R$ are unique.[15] Since $\mathcal{A}(\hat{\mathcal{A}}_R(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in \mathbf{Y}$, from the definition of left inverse it follows that

$$\hat{\mathcal{A}}_L(\mathbf{y}) = \hat{\mathcal{A}}_L(\mathcal{A}(\hat{\mathcal{A}}_R(\mathbf{y}))) = \hat{\mathcal{A}}_R(\mathbf{y}) \quad \text{for all } \mathbf{y} \in \mathbf{Y}$$

that is, $\hat{\mathcal{A}}_L = \hat{\mathcal{A}}_R$. The unique common left and right inverse of $\mathcal{A}$ is simply called the **inverse** of $\mathcal{A}$, denoted $\mathcal{A}^{-1}$.

In summary, if $\ker(\mathcal{A}) = \{\mathbf{0}\}$ and $\mathrm{im}(\mathcal{A}) = \mathbf{Y}$ for a linear transformation $\mathcal{A} : \mathbf{X} \rightarrow \mathbf{Y}$, then there exists a unique inverse transformation $\mathcal{A}^{-1} : \mathbf{Y} \rightarrow \mathbf{X}$ such that

$$\mathcal{A}^{-1}(\mathcal{A}(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbf{X}$$

and

$$\mathcal{A}(\mathcal{A}^{-1}(\mathbf{y})) = \mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbf{Y}$$

This is somewhat an expected result, because if $\ker(\mathcal{A}) = \{\mathbf{0}\}$ and $\mathrm{im}(\mathcal{A}) = \mathbf{Y}$ then $\mathcal{A}$ establishes a one-to-one correspondence between the elements of $\mathbf{X}$ and $\mathbf{Y}$. Such a linear transformation is called an **isomorphism**, and any two vector spaces related by an isomorphism are called **isomorphic**. It is left to the reader to show that two finite dimensional vector spaces $\mathbf{X}$ and $\mathbf{Y}$ are isomorphic if and only if $\dim(\mathbf{X}) = \dim(\mathbf{Y})$ (Exercise 3.39).

Example 3.39

Let $\dim(\mathbf{X}) = n$, and let $\mathbf{R}$ be a basis for $\mathbf{X}$. Let the unique representation of a vector $\mathbf{x} \in \mathbf{X}$ with respect to the basis $\mathbf{R}$ be $\boldsymbol{\alpha}_x \in \mathsf{F}^{n\times 1}$. Then the mapping $\mathcal{A} : \mathbf{X} \rightarrow \mathsf{F}^{n\times 1}$ defined as

$$\mathcal{A}(\mathbf{x}) = \boldsymbol{\alpha}_x$$

is a linear transformation as can easily be shown using the definition of the column representation of a vector.

Since $\boldsymbol{\alpha}_x$ is uniquely defined by $\mathbf{x}$, $\mathbf{x} = \mathbf{0}$ is the only vector whose representation is $\mathbf{0}_{n\times 1}$. Hence $\ker(\mathcal{A}) = \{\mathbf{0}\}$. Also since every $\boldsymbol{\alpha} \in \mathsf{F}^{n\times 1}$ is the representation of some $\mathbf{x} \in \mathbf{X}$, $\mathrm{im}(\mathcal{A}) = \mathsf{F}^{n\times 1}$. Thus $\mathcal{A}$ is a one-to-one mapping from $\mathbf{X}$ onto $\mathsf{F}^{n\times 1}$ (an isomorphism). The inverse of $\mathcal{A}$ is defined as $\mathcal{A}^{-1}(\boldsymbol{\alpha}_x) = \mathbf{x}$.

3.5 Linear Equations

In Chapter 1 we considered linear systems of the form

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

where A is an $m \times n$ matrix. We have seen that a general solution is of the form

$$\mathbf{x} = \boldsymbol{\phi}_p + \boldsymbol{\phi}_c$$

[15] A rigorous proof of this statement is beyond the scope of this book. However, we can argue that since there exists no $y \notin \mathrm{im}(\mathcal{A})$, there is no arbitrariness in $\hat{\mathcal{A}}_L$. Also, since for any $\mathbf{y} \in \mathbf{Y}$, the vector $\mathbf{x}$ that satisfies $\mathcal{A}(\mathbf{x}) = \mathbf{y}$ is unique, there is no arbitrariness in $\hat{\mathcal{A}}_R$ either.

where $\mathbf{x} = \boldsymbol{\phi}_p$ is a particular solution, and $\mathbf{x} = \boldsymbol{\phi}_c$ is a complementary solution that contains arbitrary constants and satisfies the associated homogeneous equation.

In Chapter 2 we considered first and second order linear differential equations of the form

$$L(D)(y) = u(t)$$

where $L(D)$ is a linear differential operator with constant coefficients. Again, the solution is of the form

$$y = \phi_p(t) + \phi_c(t)$$

where $y = \phi_p$ is a particular solution, and $y = \phi_c$ is a complementary solution.

The similarities between the nature of solutions of linear differential equations and linear systems are striking but not surprising. Both a matrix and a linear differential operator are linear transformations, and both a linear system and a linear differential equation can be viewed as an equation

$$\mathcal{A}(\mathbf{x}) = \mathbf{y} \tag{3.13}$$

where $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ is a linear transformation and $\mathbf{y} \in \mathbf{Y}$ is a given vector. In the case of linear systems $\mathbf{X}$ and $\mathbf{Y}$ are $\mathsf{F}^{n\times 1}$ and $\mathsf{F}^{m\times 1}$, and in the case of linear differential equations, the set of real-valued piece-wise continuous functions defined on some interval. We now recall some definitions of Chapter 1 and Chapter 2.

An equation of the form (3.13), where $\mathcal{A}$ is a linear transformation, is called a **linear equation**. If $\mathbf{y} = \mathbf{0}$ then the equation is **homogeneous**. A vector $\mathbf{x} = \boldsymbol{\phi}$ is called a **solution** of (3.13) if $\mathcal{A}(\boldsymbol{\phi}) = \mathbf{y}$. If (3.13) has no solution, it is said to be **inconsistent**. Clearly, (3.13) is consistent if and only if $\mathbf{y} \in \operatorname{im}(\mathcal{A})$.

Consider the homogeneous linear equation

$$\mathcal{A}(\mathbf{x}) = \mathbf{0} \tag{3.14}$$

which is consistent as $\mathbf{x} = \mathbf{0}$ is a trivial solution. Clearly, the set of all solutions of (3.14) is $\ker(\mathcal{A})$. Suppose that $\dim(\ker(\mathcal{A})) = \nu(\mathcal{A}) = \nu$, and let $\{\, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu \,\}$ be a basis for $\ker(\mathcal{A})$. Then any solution of (3.14) can be expressed in terms of the basis vectors as

$$\mathbf{x} = c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu \tag{3.15}$$

for some choice of the constants $c_1, \ldots, c_\nu$.

Now consider the non-homogeneous linear equation (3.13). Assume that $\mathbf{y} \in \operatorname{im}(\mathcal{A})$, so that it has at least one particular solution $\mathbf{x} = \boldsymbol{\phi}_p$. Then for arbitrary $c_1, \ldots, c_\nu$

$$\mathbf{x} = \boldsymbol{\phi}_p + c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu \tag{3.16}$$

is also a solution, because

$$\mathcal{A}(\boldsymbol{\phi}_p + c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu) = \mathcal{A}(\boldsymbol{\phi}_p) + \sum_{i=1}^{\nu} c_i\mathcal{A}(\boldsymbol{\phi}_i) = \mathbf{y} + \sum_{i=1}^{\nu} c_i\mathbf{0} = \mathbf{y}$$

Conversely, if $\mathbf{x} = \boldsymbol{\phi}$ is any solution of (3.13), then since

$$\mathcal{A}(\boldsymbol{\phi} - \boldsymbol{\phi}_p) = \mathcal{A}(\boldsymbol{\phi}) - \mathcal{A}(\boldsymbol{\phi}_p) = \mathbf{y} - \mathbf{y} = \mathbf{0}$$

$\boldsymbol{\phi} - \boldsymbol{\phi}_p$ is a solution of the associated homogeneous equation (3.14), and therefore, can be expressed as in (3.15). This, in turn, implies that $\boldsymbol{\phi}$ is of the form (3.16). Thus (3.16) characterizes the solution set of (3.13), and is called a **general solution**.

Note that neither ϕ_p nor $\phi_i, i = 1, \ldots, \nu$, are unique. If ϕ'_p is another particular solution, and $\phi'_i, i = 1, \ldots, \nu$, form another basis for $\ker(\mathcal{A})$, then

$$\mathbf{x} = \phi'_p + c'_1\boldsymbol{\phi}'_1 + \cdots + c'_\nu\boldsymbol{\phi}'_\nu \tag{3.17}$$

is also a general solution. Although the expressions in (3.16) and (3.17) are different, they nevertheless define the same family of solutions (see Exercise 3.41).

Example 3.40

Consider the linear system

$$[1\ \ 2\ \ 3]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2$$

whose coefficient matrix is already in reduced row echelon form.

Following the standard procedure of Chapter 1, a general solution is obtained as

$$\mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + c_1\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + c_2\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

where $\boldsymbol{\phi}_p = \operatorname{col}[2, 0, 0]$ is a particular solution, and $\boldsymbol{\phi}_1 = \operatorname{col}[-2, 1, 0]$ and $\boldsymbol{\phi}_2 = \operatorname{col}[-3, 0, 1]$ form a basis for the kernel of the coefficient matrix.

On the other hand, $\boldsymbol{\phi}'_p = \operatorname{col}[0, 1, 0]$ is also a particular solution (obtained from the general solution above by choosing $c_1 = 1$ and $c_2 = 0$), and $\boldsymbol{\phi}'_1 = \operatorname{col}[-2, 1, 0]$ and $\boldsymbol{\phi}'_2 = \operatorname{col}[0, -3, 2]$ form another basis for the kernel of the coefficient matrix. Thus

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c'_1\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + c'_2\begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix}$$

is also a general solution.

The reader can verify that any solution obtained from the second expression by choosing arbitrary values for c'_1 and c'_2 can also be obtained from the first expression by choosing $c_1 = 1 + c'_1 - c'_2$ and $c_2 = 2c'_2$, and vice versa.

Linear equations of the form (3.13) are not limited to linear differential equations and linear systems. In the following two examples, we consider different types of linear equations.

Example 3.41

Consider the linear equation

$$\mathcal{A}(M) = M + M^t = N = \begin{bmatrix} 6 & 2 \\ 2 & -4 \end{bmatrix}$$

where $\mathcal{A}$ is the linear transformation considered in Example 3.36. Since the matrix on the right-hand side of the above equation is symmetric, it is in $\operatorname{im}(\mathcal{A})$, and hence, the equation is consistent.

A particular solution can be obtained by inspection to be

$$M_p = \frac{1}{2}N = \begin{bmatrix} 3 & 1 \\ 1 & -2 \end{bmatrix}$$

(Since N is symmetric, then so is M_p, and hence $M_p + M_p^t = 2M_p = N$.)

The general solution can then be obtained by complementing M_p with $\ker(\mathcal{A})$, which has already been characterized in Example 3.36, as

$$M = \begin{bmatrix} 3 & 1 \\ 1 & -2 \end{bmatrix} + c \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Thus

$$M = \begin{bmatrix} 3 & 0 \\ 2 & -2 \end{bmatrix}$$

is also a solution obtained from the general solution with $c = 1$. Indeed

$$M + M^t = \begin{bmatrix} 3 & 0 \\ 2 & -2 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & -4 \end{bmatrix} = N$$

Example 3.42

Suppose that we are interested in finding a sequence $f \in \mathcal{F}(\mathbf{N}, \mathbf{C})$ which satisfies an equation of the form

$$f[k+2] + a_1 f[k+1] + a_2 f[k] = u[k]\,, \quad k \geq 1 \tag{3.18}$$

where a_1, a_2 are fixed (complex) coefficients, and $u \in \mathcal{F}(\mathbf{N}, \mathbf{C})$ is a given sequence. Such an equation is called a (second order) **difference equation**.[16]

Obtaining a solution to a difference equation is easy: Choose $f[1]$ and $f[2]$ arbitrarily, and calculate $f[3], f[4]$, etc., recursively from (3.18). Thus

$$\begin{aligned} f[1] &= c_1 \\ f[2] &= c_2 \\ f[3] &= -a_2 f[1] - a_1 f[2] + u[1] = -a_2 c_1 - a_1 c_2 + u[1] \\ f[4] &= -a_2 f[2] - a_1 f[3] + u[2] = (a_1 a_2) c_1 + (a_1^2 - a_2) c_2 + u[2] - a_1 u[1] \end{aligned}$$

and so on. Certainly, any term of a solution sequence can be obtained after sufficient number of substitutions. However, it would be useful to have a formula for the kth term, which could be evaluated without working out all the intermediate terms.

Let us try to formulate the problem as a linear equation. For this purpose we define a shift operator $\Delta : \mathcal{F}(\mathbf{N}, \mathbf{C}) \to \mathcal{F}(\mathbf{N}, \mathbf{C})$ as

$$(\Delta f)[k] = f[k+1]\,, \quad k \in \mathbf{N}$$

Defining Δ^2, Δ^3, etc., similar to the powers of the differential operator D, the difference equation in (3.18) can be expressed as

$$L(\Delta)(f) = u[k] \tag{3.19}$$

where

$$L(\Delta) = \Delta^2 + a_1 \Delta + a_2 I$$

[16]Note that the recursion relations we considered in Section 2.8 in connection with numerical solution of differential equations are difference equations with specified initial conditions.

is a polynomial shift operator on $\mathcal{F}(\mathbf{N},\mathbf{C})$.

It is left to the reader to show that $L(\Delta)$ is a linear operator on $\mathcal{F}(\mathbf{N},\mathbf{C})$. Hence the linear difference equation (3.18) is a linear equation. Then it must have a general solution of the form

$$f = \phi_p[k] + \phi_c[k]$$

where ϕ_p is any particular solution sequence, and ϕ_c is a complementary solution sequence expressed as a linear combination of the basis vectors of $\ker(L(\Delta))$.

To be more specific, let us consider the difference equation

$$f[k+2] - \frac{5}{6}f[k+1] + \frac{1}{6}f[k] = 1\,, \quad k \in \mathbf{N}$$

or in operator notation

$$(\Delta^2 - \frac{5}{6}\Delta + \frac{1}{6}I)(f) = 1\,, \quad k \in \mathbf{N}$$

Since the right-hand side of the given equation is a constant, we suspect that a constant sequence $f[k] = C$ might be a solution. Substituting the assumed solution into the equation, and noting that $\Delta^2(C) = \Delta(C) = C$, we get

$$C - (5/6)C + (1/6)C = (1/3)C = 1$$

giving $C = 3$. A particular solution is thus obtained as $\phi_p[k] = 3$.

To find the complementary solution, we try a sequence of the form $f[k] = z^k$. Substituting the trial solution into the homogeneous equation, we get

$$L(\Delta)(z^k) = (z^{k+2} - (5/6)z^{k+1} + (1/6)z^k) = 0$$

or equating the corresponding terms

$$z^{k+2} - (5/6)z^{k+1} + (1/6)z^k = z^k(z^2 - (5/6)z + (1/6)) = 0$$

From the last equation, we observe that $\phi[k] = z^k$ is a solution if and only if z is a root of the characteristic equation

$$z^2 - (5/6)z + (1/6) = 0$$

Since the characteristic equation has two real roots, $\mu_1 = 1/2$ and $\mu_2 = 1/3$, each of the sequences

$$\phi_1[k] = (1/2^k) \quad \text{and} \quad \phi_2[k] = (1/3^k)$$

is a solution of the homogeneous equation. It can be shown that ϕ_1 and ϕ_2 are linearly independent sequences, and thus form a basis for $\ker(L(\Delta))$. The general solution of the non-homogeneous difference equation is thus obtained as

$$f = \phi[k] = 3 + c_1/2^k + c_2/3^k\,, \quad c_1, c_2 \in \mathbf{C}$$

* 3.6 Direct Sum and Projections

Consider an arbitrary vector $\mathbf{x} = (x_1, x_2, x_3) \in \mathbf{R}^3$. In terms of the canonical basis vectors $\mathbf{e}_1 = (1,0,0)$, $\mathbf{e}_2 = (0,1,0)$ and $\mathbf{e}_3 = (0,0,1)$, $\mathbf{x}$ can be expressed uniquely as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \tag{3.20}$$

Let $\mathbf{U}_i = \text{span}\,(\mathbf{e}_i), i = 1,2,3$, be the one-dimensional subspaces of $\mathbf{R}^3$ defined by the canonical basis vectors. (They represent the x, y and z axes in the the xyz space). Then we can interpret (3.20) as

$$\mathbf{x} = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$$

which is a decomposition of $\mathbf{x}$ into three components $\mathbf{u}_1 = x_1\mathbf{e}_1, \mathbf{u}_2 = x_2\mathbf{e}_2, \mathbf{u}_3 = x_3\mathbf{e}_3$ in the subspaces $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$.

Now let $\mathbf{V}_1 = \text{span}\,(\mathbf{e}_1, \mathbf{e}_2)$ and $\mathbf{V}_2 = \text{span}\,(\mathbf{e}_3)$, where $\mathbf{V}_1$ is a two-dimensional subspace (the xy plane) and $\mathbf{V}_2$ is a one-dimensional subspace (the z axis). Then (3.20) can also be written as

$$\mathbf{x} = \mathbf{v}_1 + \mathbf{v}_2$$

which gives a decomposition of $\mathbf{x}$ into two components $\mathbf{v}_1 = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ and $\mathbf{v}_2 = x_3\mathbf{e}_3$ in the subspaces $\mathbf{V}_1$ and $\mathbf{V}_2$. Obviously, these components are uniquely determined by $\mathbf{x}$ and the subspaces $\mathbf{V}_1$ and $\mathbf{V}_2$.

This example suggests that the idea of decomposing a vector into components along the one-dimensional subspaces defined by a given basis can be generalized to a decomposition into components in higher dimensional subspaces.

Let $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_k$ be subspaces of a vector space $\mathbf{X}$. Their **algebraic sum** is defined as

$$\sum_{i=1}^{k} \mathbf{U}_i = \{\, \mathbf{x} \,|\, \mathbf{x} = \sum_{i=1}^{k} \mathbf{u}_i, \mathbf{u}_i \in \mathbf{U}_i, i = 1, \ldots, k \,\}$$

It is easy to show that the algebraic sum is also a subspace of $\mathbf{X}$.

The subspaces $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_k$ are said to be linearly independent if for $\mathbf{u}_i \in \mathbf{U}_i, i = 1, \ldots, k$

$$\sum_{i=1}^{k} \mathbf{u}_i = \mathbf{0}$$

is satisfied only when $\mathbf{u}_i = \mathbf{0}, i = 1, \ldots, k$.

If $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_k$ are said to be linearly independent their algebraic sum is called a **direct sum**, denoted $\bigoplus_{i=1}^{k} \mathbf{U}_i$.

Note that algebraic sum of a family of subspaces is a generalization of the concept of span of a set of vectors. Similarly, linear independence of a family of subspaces is a generalization of the concept of linear independence of a set of vectors. The following theorem characterizes two linearly independent subspaces. The extension of the theorem to more than two subspaces is left to the reader as an exercise (see Exercise 3.43).

Theorem 3.2 *Let* $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots\}$ *and* $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots\}$ *be bases for the subspaces* $\mathbf{U}$ *and* $\mathbf{V}$*. Then the following are equivalent.*

a) $\mathbf{U}$ *and* $\mathbf{V}$ *are linearly independent.*

b) $\mathbf{U} \cap \mathbf{V} = \{\mathbf{0}\}$.

c) $\mathbf{T} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{s}_1, \mathbf{s}_2, \dots\}$ *is a basis for* $\mathbf{U} + \mathbf{V}$.

Proof We will show that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b):

By contradiction. Suppose that there exists $\mathbf{0} \neq \mathbf{x} \in \mathbf{U} \cap \mathbf{V}$. Let $\mathbf{u} = \mathbf{x} \in \mathbf{U}, \mathbf{v} = -\mathbf{x} \in \mathbf{V}$. Then $\mathbf{u} \neq \mathbf{0} \neq \mathbf{v}$ and $\mathbf{u} + \mathbf{v} = \mathbf{x} - \mathbf{x} = \mathbf{0}$.

(b) $\Rightarrow$ (c):

By contradiction. Since span $(\mathbf{T}) = \mathbf{U} + \mathbf{V}$, if $\mathbf{T}$ is not a basis for $\mathbf{U} + \mathbf{V}$ then it must be linearly dependent, in which case there exist scalars α_i and β_j, not all zero, such that

$$\sum_i \alpha_i \mathbf{r}_i + \sum_j \beta_j \mathbf{s}_j = \mathbf{0}$$

Let

$$\mathbf{x} = \sum_i \alpha_i \mathbf{r}_i = -\sum_j \beta_j \mathbf{s}_j$$

Then $\mathbf{x} \in \mathbf{U} \cap \mathbf{V}$, and $\mathbf{x} \neq \mathbf{0}$ as $\mathbf{R}$ and $\mathbf{S}$ are linearly independent.

(c) $\Rightarrow$ (a):

Suppose $\mathbf{u} + \mathbf{v} = \mathbf{0}$ for some

$$\mathbf{u} = \sum_i \alpha_i \mathbf{r}_i \quad \text{and} \quad \mathbf{v} = \sum_j \beta_j \mathbf{s}_j$$

that is,

$$\sum_i \alpha_i \mathbf{r}_i + \sum_j \beta_j \mathbf{s}_j = \mathbf{0}$$

Since $\mathbf{T}$ is a basis, we must have $\alpha_i = 0$ for all i implying $\mathbf{u} = \mathbf{0}$, and also $\beta_j = 0$ for all j implying $\mathbf{v} = \mathbf{0}$.

Example 3.43

In $\mathbf{R}^3$, let

$$\mathbf{U}_1 = \text{span}(\mathbf{e}_1, \mathbf{e}_2), \quad \mathbf{U}_2 = \text{span}(\mathbf{e}_2, \mathbf{e}_3), \quad \mathbf{U}_3 = \text{span}(\mathbf{e}_3)$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the canonical basis vectors. Then

$$\mathbf{U}_1 + \mathbf{U}_2 = \mathbf{U}_1 + \mathbf{U}_3 = \text{span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbf{R}^3$$

$\mathbf{U}_1$ and $\mathbf{U}_2$ are not linearly independent, because $\mathbf{e}_2 \in \mathbf{U}_1 \cap \mathbf{U}_2$. However, $\mathbf{U}_1$ and $\mathbf{U}_3$ are linearly independent, because $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a basis for $\mathbf{U}_1 + \mathbf{U}_3$. Hence

$$\mathbf{R}^3 = \mathbf{U}_1 \oplus \mathbf{U}_3$$

The reader can interpret these findings by identifying $\mathbf{R}^3$ with the xyz space, and $\mathbf{U}_1, \mathbf{U}_2$ and $\mathbf{U}_3$ with the xy plane, yz plane, and the z axis respectively.

From Theorem 3.2 it is clear that if $\mathbf{T} = \{ \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n \}$ is a basis for $\mathbf{X}$, partitioned arbitrarily into two disjoint sets, say

$$\mathbf{R} = \{ \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_k \} \quad \text{and} \quad \mathbf{S} = \{ \mathbf{t}_{k+1}, \mathbf{t}_{k+2}, \ldots, \mathbf{t}_n \}$$

then

$$\mathbf{X} = \operatorname{span}(\mathbf{R}) \oplus \operatorname{span}(\mathbf{S})$$

On the other hand, if $\mathbf{R} = \{ \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k \}$ is a basis for a k-dimensional subspace $\mathbf{U}$, then by Corollary 3.2, it can be completed to a basis by including $n-k$ more vectors. Let these additional vectors form a set $\mathbf{S}$, and let $\mathbf{V} = \operatorname{span}(\mathbf{S})$. Then $\mathbf{X} = \mathbf{U} \oplus \mathbf{V}$. The subspace $\mathbf{V}$ thus constructed is called a **complement** of $\mathbf{U}$. Since $\mathbf{S}$ can be chosen in many different ways, complement of $\mathbf{U}$ is not unique. For example, for any vector $\mathbf{v} = (a, b, c)$ with $c \neq 0$, $\mathbf{V} = \operatorname{span}(\mathbf{v})$ is a complement of $\mathbf{U}_1$ in Example 3.43, and in particular, so is $\mathbf{U}_3$. However, all complements of $\mathbf{U}$ must have the same dimension.

We also observe that if $\mathbf{X}$ is finite dimensional and is decomposed into a direct sum as $\mathbf{X} = \mathbf{U} \oplus \mathbf{V}$ then

$$\dim(\mathbf{X}) = \dim(\mathbf{U}) + \dim(\mathbf{V})$$

Let $\mathbf{X} = \mathbf{U} \oplus \mathbf{V}$, and let $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_k)$ and $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_{n-k})$ be ordered bases for $\mathbf{U}$ and $\mathbf{V}$. Since $\mathbf{T} = \mathbf{R} \cup \mathbf{S}$ is a basis for $\mathbf{X}$, any vector $\mathbf{x} \in \mathbf{X}$ can be expressed as

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{r}_i + \sum_{j=1}^{n-k} \beta_j \mathbf{s}_j = \mathbf{u} + \mathbf{v} \tag{3.21}$$

The vectors $\mathbf{u} \in \mathbf{U}$ and $\mathbf{v} \in \mathbf{V}$, which are uniquely defined by $\mathbf{x}$, are called the **components** of $\mathbf{x}$ in $\mathbf{U}$ and $\mathbf{V}$.

Let $\mathcal{P} : \mathbf{X} \to \mathbf{X}$ be a mapping which maps every vector in $\mathbf{X}$ to its component in $\mathbf{U}$, that is,

$$\mathcal{P}(\mathbf{x}) = \mathbf{u}$$

$\mathcal{P}$ is called a **projection** on $\mathbf{U}$ along $\mathbf{V}$. It is left to the reader to show that

a) $\mathcal{P}$ is a linear transformation

b) $\operatorname{im}(\mathcal{P}) = \mathbf{U}$

c) $\ker(\mathcal{P}) = \mathbf{V}$

The reader should note that the mapping $\mathcal{Q} : \mathbf{X} \to \mathbf{X}$ defined as $\mathcal{Q}(\mathbf{x}) = \mathbf{v}$ is also a projection (on $\mathbf{V}$ along $\mathbf{U}$).

The matrix P that represents $\mathcal{P}$ with respect to the basis $\mathbf{T}$ is called a **projection matrix**. Since

$$\mathcal{P}(\mathcal{P}(\mathbf{x})) = \mathcal{P}(\mathbf{u}) = \mathbf{u} = \mathcal{P}(\mathbf{x})$$

for any $\mathbf{x} \in \mathbf{X}$, it follows that $P^2 \boldsymbol{\alpha} = P \boldsymbol{\alpha}$ for any $\boldsymbol{\alpha} \in \mathsf{F}^{n \times 1}$ that stands for the column representation of some vector. Thus if P is a projection matrix then $P^2 = P$. Such a matrix is called **idempotent**. Conversely, if P is an $n \times n$ idempotent matrix, then

$$\mathsf{F}^{n \times 1} = \operatorname{im}(P) \oplus \ker(P)$$

and P defines a projection in $\mathsf{F}^{n \times 1}$ on $\operatorname{im}(P)$ along $\ker(P)$ (see Exercise 3.51).

Example 3.44

In Example 3.43, the projection on $\mathbf{U}_1$ along $\mathbf{U}_3$ is defined by the matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the projection on $\mathbf{U}_3$ along $\mathbf{U}_1$ is defined by the matrix

$$Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that $P^2 = P$ and $Q^2 = Q$.

Now consider the matrix

$$R = \frac{1}{3}\begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

Since R is idempotent, it defines a projection in $\mathbf{R}^{3\times 1}$ on

$$\operatorname{im}(R) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2)$$

along

$$\ker(R) = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right) = \operatorname{span}(\mathbf{u}_3)$$

The components of an arbitrary vector $\mathbf{x} = \operatorname{col}[\,a, b, c\,]$ in $\operatorname{im}(R)$ and $\ker(R)$ are

$$\mathbf{u} = R\mathbf{x} = \frac{1}{3}\begin{bmatrix} 2a+b-c \\ a+2b+c \\ -a+b+2c \end{bmatrix} = \frac{2a+b-c}{3}\,\mathbf{u}_1 + \frac{-a+b+2c}{3}\,\mathbf{u}_2$$

and

$$\mathbf{v} = (I-R)\mathbf{x} = \frac{1}{3}\begin{bmatrix} a-b+c \\ -a+b-c \\ a-b+c \end{bmatrix} = \frac{a-b+c}{3}\,\mathbf{u}_3$$

Example 3.45

In $\mathbf{C}[s]$, let

$$\begin{aligned} \mathbf{C}_e[s] &= \{\, p(s) = p_0 + p_1 s^2 + \cdots + p_m s^{2m} \mid m = 0, 1, \ldots \} \\ \mathbf{C}_o[s] &= \{\, q(s) = q_0 s + q_1 s^3 + \cdots + q_m s^{2m+1} \mid m = 0, 1, \ldots \} \end{aligned}$$

Then $\mathbf{C}[s] = \mathbf{C}_e[s] \oplus \mathbf{C}_o[s]$.

If

$$r(s) = r_0 + r_1 s + r_2 s^2 + \cdots + r_{2n+1} s^{2n+1}$$

then

$$r_e(s) = r_0 + r_2 s^2 + \cdots + r_{2n} s^{2n}$$

is the projection of r on $\mathbf{C}_e[s]$ along $\mathbf{C}_o[s]$.

This example illustrates that direct sum decomposition and projections are not limited to finite dimensional vector spaces.

3.7 Exercises

1. Prove properties (a-f) on p.87 of a vector space.
2. Show that a plane through the origin is a subspace of R^3.
3. Write down the equation of a line passing through two given points $\mathbf{p}, \mathbf{q} \in \mathsf{R}^n$. Under what conditions on $\mathbf{p}, \mathbf{q}$ does the line represent a subspace?
4. Discuss how an $m \times n$ real matrix A can be interpreted as a function $f : \mathsf{n} \to \mathsf{R}^{m\times 1}$.
5. Show that if $\mathbf{U}$ and $\mathbf{V}$ are subspaces of $\mathbf{X}$, then so is $\mathbf{U} \cap \mathbf{V}$. Is $\mathbf{U} \cup \mathbf{V}$ also a subspace?
6. Show that $\text{span}\,(\text{span}\,(\mathbf{R})) = \text{span}\,(\mathbf{R})$ for any subset $\mathbf{R} \subset \mathbf{X}$.
7. Prove facts (a-c) on p.93 concerning linear independence.
8. Let $\mathbf{R}$ be a finite set of vectors and let $\mathbf{R}'$ be obtained from $\mathbf{R}$ by a single Type I or Type II elementary operation.
   (a) Explain why $\text{span}\,(\mathbf{R}') = \text{span}\,(\mathbf{R})$.
   (b) Explain why $\mathbf{R}'$ is linearly independent if and only if $\mathbf{R}$ is.
9. Show that C is a vector space over R, and find a basis for it.
10. Show that the set of all 3×3 real skew-symmetric matrices is a subspace of $\mathsf{R}^{3\times 3}$, and find a basis for it.
11. Show that the set $\mathbf{R} = \{ r_0, r_1, \dots \}$ in Example 3.18 is a basis for $\mathsf{R}[s]$.
12. An $n \times n$ matrix N is said to be **nilpotent of index** k if $N^k = O$ but $N^{k-1} \neq O$.
   (a) Let $\mathbf{v}$ be such that $N^{k-1}\mathbf{v} \neq \mathbf{0}$. Show that the vectors $\mathbf{v}, N\mathbf{v}, \dots, N^{k-1}\mathbf{v}$ are linearly independent.
   (b) Show that the index of nilpotency cannot exceed n.
13. Suppose $\mathbf{R}$ is linearly independent, and $\mathbf{x} \notin \text{span}\,(\mathbf{R})$. Show that $\mathbf{R} \cup \{ \mathbf{x} \}$ is also linearly independent. Hint: Consider a finite subset $\mathbf{S} \subset \mathbf{R} \cup \{ \mathbf{x} \}$. If $\mathbf{x} \notin \mathbf{S}$ then $\mathbf{S} \subset \mathbf{R}$, and therefore $\mathbf{S}$ must be linearly independent. If $\mathbf{x} \in \mathbf{S}$ then $\mathbf{S} = \{ \mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{x} \}$ for some $\mathbf{r}_1, \dots, \mathbf{r}_k \in \mathbf{R}$.
14. Since the set $\mathbf{R} = \{ \mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_{k+1}, \dots, \mathbf{r}_{k+n} \}$ in the proof of Corollary 3.2(b) is linearly dependent, there exist some $c_1, \dots, c_{n+k}$, not all zero, such that

   $$\sum_{i=1}^{n+k} c_i \mathbf{r}_i = \mathbf{0}$$

   Show that at least one of $c_{k+1}, \dots, c_{n+k}$ must be nonzero. Explain why this implies that the set obtained by deleting the corresponding vector from $\mathbf{R}$ includes the first k vectors and still spans $\mathbf{X}$. On the basis of this reasoning, explain also why the algorithm in the proof of Corollary 3.2(a) reduces $\mathbf{R}$ to a basis that includes $\mathbf{R}_1$.
15. Let $\mathbf{U} = \text{span}\,(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p)$ and $\mathbf{V} = \text{span}\,(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_q)$ be subspaces of $\mathsf{R}^{n\times 1}$. Give an algorithm to obtain a basis for $\mathbf{U} \cap \mathbf{V}$.
16. In a two-dimensional vector space $\mathbf{X}$, a vector x has the representation $\boldsymbol{\alpha} = \text{col}\,[\,1, -1\,]$ with respect to some ordered basis $\mathbf{R} = (\,\mathbf{r}_1, \mathbf{r}_2\,)$. Let

   $$\mathbf{r}_1' = \mathbf{r}_1 + \mathbf{r}_2 \quad \text{and} \quad \mathbf{r}_2' = \mathbf{r}_1 + 2\mathbf{r}_2$$

   (a) Show that $\mathbf{R}' = (\,\mathbf{r}_1', \mathbf{r}_2'\,)$ is also a basis for $\mathbf{X}$.
   (b) Find the matrix of change–of–basis Q from $\mathbf{R}$ to $\mathbf{R}'$, and the matrix of change–of–basis P from $\mathbf{R}'$ to $\mathbf{R}$. Verify that $QP = PQ = I$.

(c) Obtain the representation $\boldsymbol{\alpha}'$ of x with respect to $\mathbf{R}'$.

17. Refer to Example 3.25. Let $N = 4$.
    (a) Find the matrix of change–of–basis Q from (e_p) to (ϕ_p).
    (b) Verify that $\mathbf{F} = Q\,\mathbf{f}$ for the sequence f considered in the example.
18. Let $f = (3, -1, -3, 5, 3, 5) \in \mathcal{F}(\mathsf{D}_6, \mathsf{C})$.
    (a) Compute the discrete Fourier coefficients of f.
    (b) Verify your result by using the MATLAB commands

```
f=[3 -1 -3 5 3 5];
c=fft(f)
```

19. Refer to Example 3.25.
    (a) Show that

$$\psi_p[k] = \begin{cases} 1, & p = 0 \\ \cos\frac{\pi}{2}k & p = 1 \\ \sin\frac{\pi}{2}k & p = 2 \\ \cos\pi k & p = 3 \end{cases}$$

    is also a basis for $\mathcal{F}(\mathsf{D}_4, \mathsf{C})$.
    (b) Find the representation of f with respect to (ψ_p).
20. Show (3.7) for $f \in \mathcal{F}(\mathsf{D}_N, \mathsf{C})$. Hint: Each of the complex numbers

$$s_p = e^{ip\frac{2\pi}{N}}, \quad p = 0, 1, \ldots, N-1$$

satisfies

$$s_p^N = 1$$

Use this fact to show that

$$\sum_{p=0}^{N-1} e^{ip\frac{2\pi}{N}q} = \begin{cases} N, & q = 0 \\ 0, & q \neq 0 \end{cases}$$

21. Consider $\mathcal{F}(\mathsf{D}, \mathsf{R})$ with $\mathsf{D} = \{1, 2, 3\}$.
    (a) Show that (g_1, g_2, g_3), where

$$\begin{aligned} g_1[k] &= 1, \quad k = 1, 2, 3 \\ g_2[k] &= \begin{cases} 1, & k = 1, 2 \\ 0, & k = 3 \end{cases} \\ g_3[k] &= \begin{cases} 0, & k = 1 \\ 1, & k = 2, 3 \end{cases} \end{aligned}$$

    is a basis for $\mathcal{F}(\mathsf{D}, \mathsf{R})$.
    (b) Find the column representation of

$$f[k] = k, \quad k = 1, 2, 3$$

    with respect to (g_1, g_2, g_3).

22. Prove that $\mathcal{F}(\mathbf{N}, \mathbf{R})$ is infinite dimensional. Hint: Assume that it has a finite dimension, say, M, and try to find a subspace of $\mathcal{F}(\mathbf{N}, \mathbf{R})$ with dimension larger than M.

23. Let $\mathcal{A} : \mathbf{R}^{2\times 1} \to \mathbf{R}^{2\times 1}$ be a linear transformation defined by the matrix

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

Can you find a basis for $\mathbf{R}^{2\times 1}$ with respect to which $\mathcal{A}$ is represented by a diagonal matrix?

24. Let $\mathcal{A} : \mathbf{R}^3 \to \mathbf{R}^2$ be defined as $\mathcal{A}(x, y, z) = (x + y, y - 2z)$. Choose arbitrary bases for $\mathbf{R}^3$ and $\mathbf{R}^2$ (other than the canonical bases) and obtain the matrix representation of $\mathcal{A}$ with respect to these bases.

25. Let

$$\mathbf{R}_3[s] = \{\, p(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 \,|\, p_0, p_1, p_2, p_3 \in \mathbf{R}\,\}$$

(a) Find a basis $\mathbf{R}$ for $\mathbf{R}_3[s]$.

(b) Find the column representation of $q(s) = 1 + s^2 - 2s^3$ with respect to $\mathbf{R}$.

(c) Let $\mathcal{A} : \mathbf{R}_3[s] \to \mathbf{R}_3[s]$ be defined as $\mathcal{A}(p) = p'$, where p' denotes the derivative of p. Show that $\mathcal{A}$ is a linear transformation, and find its matrix representation with respect to $\mathbf{R}$.

26. Let $\mathcal{A} : \mathbf{R}^{2\times 2} \to \mathbf{R}^{2\times 2}$ be defined as $\mathcal{A}(X) = CX$, where

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

Find the matrix representation of $\mathcal{A}$ with respect to the basis in Example 3.21.

27. (a) Show that the matrices

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ M_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},\ M_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ M_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

form an ordered basis $\mathbf{M} = (M_1, M_2, M_3, M_4)$ for $\mathbf{R}^{2\times 2}$.

(b) Let $\mathcal{A} : \mathbf{R}^{2\times 2} \to \mathbf{R}^{2\times 2}$ be defined as in Example 3.36. Find $Y = \mathcal{A}(X)$ for

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

(c) Find the column representations $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of X and Y with respect to $\mathbf{M}$.

(d) Find the matrix representation A of $\mathcal{A}$ with respect to $\mathbf{M}$, and show that $\boldsymbol{\beta} = A\boldsymbol{\alpha}$.

28. In special theory of relativity, the space and time coordinates of an object measured in two coordinate systems moving in the x direction at a constant relative speed are related by the **Lorentz transformation**

$$\begin{aligned} \mathcal{L}(v):\ x' &= k_v(x - vt) \\ t' &= k_v(-\frac{v}{c^2}x + t) \end{aligned}$$

where c is the speed of light, $v < c$ is the relative speed of the coordinate systems, and $k_v = 1/\sqrt{1 - (v/c)^2}$.

(a) Show that the Lorentz transformation is a linear transformation from $\mathbf{R}^2$ into itself, mapping (x, t) to (x', t').

(b) Find the matrix representation $L(v)$ of the Lorentz transformation with respect to the canonical basis in $\mathbf{R}^2$.

(c) Show that $L(u)L(v) = L(w)$ for some w, and find w in terms of u and v.

29. Let $\mathbf{R} = (\,\mathbf{r}_1, \ldots, \mathbf{r}_n\,), \mathbf{S} = (\,\mathbf{s}_1, \ldots, \mathbf{s}_m\,)$ and $\mathbf{T} = (\,\mathbf{t}_1, \ldots, \mathbf{t}_p\,)$ be bases for the vector spaces $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$, respectively, and let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ and $\mathcal{B} : \mathbf{Y} \to \mathbf{Z}$ be linear transformations. Show that if $\mathcal{A}$ and $\mathcal{B}$ are represented by the matrices A and B with respect to the given bases, then $\mathcal{C} : \mathbf{X} \to \mathbf{Z}$ defined as $\mathcal{C}(\mathbf{x}) = \mathcal{B}(\mathcal{A}(\mathbf{x}))$ is represented by the matrix $C = BA$. Hint: By definition

$$\mathcal{A}(\mathbf{r}_j) = \sum_{k=1}^{m} a_{kj}\mathbf{s}_k\,, \quad \mathcal{B}(\mathbf{s}_k) = \sum_{i=1}^{p} b_{ik}\mathbf{t}_i\,, \quad \mathcal{C}(\mathbf{r}_j) = \sum_{i=1}^{p} c_{ij}\mathbf{t}_i$$

Find an expression for c_{ij} in terms of b_{ik} and a_{kj}.

30. Let $\mathbf{X}$ and $\mathbf{Y}$ be vector spaces over the same field $\mathbf{F}$.

(a) Show that the set of all linear transformations from $\mathbf{X}$ into $\mathbf{Y}$ is also vector space over $\mathbf{F}$. This vector space is denoted by $\mathcal{L}(\mathbf{X}, \mathbf{Y})$. Define clearly the addition and scalar multiplication operations on $\mathcal{L}(\mathbf{X}, \mathbf{Y})$, as well as the null vector and the additive inverse.

(b) Find $\dim(\mathcal{L}(\mathbf{X}, \mathbf{Y}))$ if $\dim(\mathbf{X}) = n$ and $\dim(\mathbf{Y}) = m$.

31. In the light of the previous exercise, explain why $L(D)$ is a linear transformation from $\mathcal{C}_n(\mathcal{I}, \mathbf{R})$ into $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$.

32. Let

$$f[k] = \begin{cases} 1 + (-1)^k, & k \geq 0 \\ 0, & k < 0 \end{cases}$$

(a) Find the sequence g defined by (3.11) for

$$h[p, q] = \begin{cases} 1/2, & q = p \text{ or } q = p - 1 \\ 0, & q \neq p, p - 1 \end{cases}$$

Plot f and g pointwise for $-2 \leq k \leq 5$.

(b) Repeat part(a) for

$$h[p, q] = \begin{cases} 1/2, & q = p \\ -1/2, & q = p - 1 \\ 0, & q \neq p, p - 1 \end{cases}$$

33. Let

$$h(t, \tau) = \begin{cases} 1, & t - 1 < \tau < t \\ 0, & \tau < t - 1 \text{ or } \tau > t \end{cases}$$

and

$$f(t) = \begin{cases} e^{-t}, & t > 0 \\ 0, & t < 0 \end{cases}$$

Find the function g defined by (3.12). Plot f and g for $-1 \leq t \leq 5$.

34. Show that if $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ is a linear transformation, then $\mathrm{im}\,(\mathcal{A})$ is a subspace of $\mathbf{Y}$.

35. Let $\mathcal{A} : \mathbf{R}^3 \to \mathbf{R}^3$ be defined as $\mathcal{A}(x, y, z) = (x + y, x + 2z, 2x + y + 2z)$. Find bases for $\mathrm{im}\,(\mathcal{A})$ and $\ker(\mathcal{A})$.

36. Let $\mathcal{A}$ be the LT defined in Exercise 3.25. Find bases for for $\text{im}\,(\mathcal{A})$ and $\ker\,(\mathcal{A})$.

37. Find a 3×3 real matrix A such that $\mathbf{0} \neq \ker\,(A) \subset \text{im}\,(A) \neq \mathsf{R}^{3\times 1}$.

38. Let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ be a linear transformation.

    (a) Show that if $\{\, \mathcal{A}(\mathbf{x}_1), \ldots, \mathcal{A}(\mathbf{x}_k) \,\}$ is linearly independent, then so is $\{\, \mathbf{x}_1, \ldots, \mathbf{x}_k \,\}$.

    (b) Show that $\mathcal{A}$ is one-to-one if and only if $\ker\,(\mathcal{A}) = \{\mathbf{0}\}$. Hint: Suppose that corresponding to some $\mathbf{y} \in \text{im}\,(\mathcal{A})$ there exist $\mathbf{x}_1 \neq \mathbf{x}_2$ such that $\mathcal{A}(\mathbf{x}_1) = \mathcal{A}(\mathbf{x}_2) = \mathbf{y}$. Let $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2 \neq \mathbf{0}$, and consider $\mathcal{A}(\mathbf{x})$.

    (c) Show that if $\mathcal{A}$ is a one-to-one LT, and $\{\, \mathbf{x}_1, \ldots, \mathbf{x}_k \,\}$ is linearly independent, then $\{\, \mathcal{A}(\mathbf{x}_1), \ldots, \mathcal{A}(\mathbf{x}_k) \,\}$ is also linearly independent.

39. (a) Let $\dim\,(\mathbf{X}) = n$ and let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ be an isomorphism. Prove that $\dim\,(\mathbf{Y}) = n$. Hint: Let $(\, \mathbf{r}_1, \, \ldots \,, \mathbf{r}_n \,)$ be a basis for $\mathbf{X}$. Show that $(\, \mathcal{A}(\mathbf{r}_1), \, \ldots \,, \mathcal{A}(\mathbf{r}_n) \,)$ is a basis for $\mathbf{Y}$.

    (b) Let $\mathbf{X}$ and $\mathbf{Y}$ be vector spaces over F, and let $(\, \mathbf{r}_1, \, \ldots \,, \mathbf{r}_n \,)$ and $(\, \mathbf{s}_1, \, \ldots \,, \mathbf{s}_n \,)$ be bases for $\mathbf{X}$ and $\mathbf{Y}$. Define $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ such that

    $$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{r}_i \implies \mathcal{A}(\mathbf{x}) = \sum_{i=1}^{n} c_i \mathbf{s}_i$$

    Show that $\mathcal{A}$ is an isomorphism.

40. Let $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$, where $\mathbf{X}, \mathbf{Y}$ are finite dimensional. Show that there exist bases for $\mathbf{X}, \mathbf{Y}$ with respect to which $\mathcal{A}$ has a matrix representation

    $$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

41. Let ϕ_p and ϕ_p' be any two particular solutions of (3.13), and let $(\, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu \,)$ and $(\, \boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_\nu' \,)$ be any two bases for $\ker\,(A)$. Show that

    $$\mathbf{x} = \boldsymbol{\phi}_p + c_1\boldsymbol{\phi}_1 + c_2\boldsymbol{\phi}_2 + \cdots + c_\nu\boldsymbol{\phi}_\nu$$

    and

    $$\mathbf{x} = \boldsymbol{\phi}_p' + c_1'\boldsymbol{\phi}_1' + c_2'\boldsymbol{\phi}_2' + \cdots + c_\nu'\boldsymbol{\phi}_\nu'$$

    define the same family of solutions, and therefore, are both general solutions. Hint: Since $\boldsymbol{\phi}_p'$ is a solution, it is a member of the first family. Also, each $\boldsymbol{\phi}_j'$ can be expressed in terms of $\boldsymbol{\phi}_i\,, \quad i = 1, \ldots, \nu$.

42. Show that the polynomial shift operator $L(\Delta)$ in Example 3.42 is a linear operator on the vector space of $\mathcal{F}(\mathsf{N}, \mathsf{C})$.

43. State and prove Theorem 3.2 for more than two subspaces $\mathbf{U}_i, i = 1, \ldots, k$. Hint: Part (b) will be

    (b)$\mathbf{U}_i \cap \sum_{j\neq i} \mathbf{U}_j = \{\, \mathbf{0} \,\}, \quad i = 1, \ldots, k$

44. (a) Show that the set $\mathbf{U} = \{\, \text{col}\,[\, x, x, y \,] \mid x, y \in \mathsf{R} \,\}$ is a subspace of $\mathbf{R}^{3\times 1}$.

    (b) Find a basis for $\mathbf{U}$. What is the dimension of $\mathbf{U}$?

    (c) Obtain the representation of the vector $\mathbf{u} = \text{col}\,[\, 1, 1, 2 \,] \in \mathbf{U}$ with respect to the basis chosen in (b).

    (d) Characterize another subspace $\mathbf{V}$ such that $\mathbf{R}^{3\times 1} = \mathbf{U} \oplus \mathbf{V}$

45. Let $\mathbf{U} = \text{span}\,(\mathbf{u}_1, \mathbf{u}_2)$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

   (a) Characterize a subspace $\mathbf{V}$ such that $\mathbf{U} \oplus \mathbf{V} = \mathsf{R}^{3\times 1}$.

   (b) Find the projection of $\mathbf{x} = \text{col}\,[\,0, 2, 1\,]$ on $\mathbf{U}$ along $\mathbf{V}$.

   (c) Find a matrix P such that for any $\mathbf{x} \in \mathcal{R}^{3\times 1}$, $P\mathbf{x}$ is the projection of $\mathbf{x}$ on $\mathbf{U}$.

46. Let $\mathbf{U} = \text{span}\,(\mathbf{u}_1, \mathbf{u}_2)$, $\mathbf{V} = \text{span}\,(\mathbf{v}_1, \mathbf{v}_2)$, where

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

   (a) Show that $\mathbf{U} \oplus \mathbf{V} = \mathsf{R}^{4\times 1}$.

   (b) Find the projection of $\mathbf{x} = \text{col}\,[\,x_1, x_2, x_3, x_4\,]$ on $\mathbf{U}$ along $\mathbf{V}$.

   (c) Construct a matrix P such that for any $\mathbf{x} \in \mathsf{R}^{4\times 1}$, the projection of $\mathbf{x}$ on $\mathbf{U}$ along $\mathbf{V}$ is $P\mathbf{x}$.

47. Let $\mathbf{X}$ be the set of all semi-infinite sequences $f \in \mathcal{F}(\mathsf{N}, \mathsf{R})$ such that

$$f[k+2] = f[k] + f[k+1]\,, \quad k \in \mathsf{N}$$

   Such a sequence is known as a **Fibonacci sequence**.

   (a) Show that $\mathbf{X}$ is a vector space over R.

   (b) Show that the sequences

$$s_1 = (1, 1, 2, 3, 5, 8, 13, ...) \quad \text{and} \quad s_2 = (-1, 1, 0, 1, 1, 2, 3, ...)$$

   form a basis for $\mathbf{X}$.

   (c) Find the projection of the sequence $f = (1, 2, 3, 5, 8, 13, ...)$ on $\text{span}\,(s_2)$ along $\text{span}\,(s_1)$.

   (d) Let $\mathcal{A} : \mathbf{X} \to \mathsf{R}^2$ be defined as $\mathcal{A}(f) = (f[3], f[4])$. Find the matrix representation of $\mathcal{A}$ with respect to the basis of $\mathbf{X}$ in part (b) and the canonical basis of R^2.

48. Let $\mathbf{X} = \mathbf{U} \oplus \mathbf{V}$, and let $\mathbf{R} = (\,\mathbf{r}_1, \ldots, \mathbf{r}_k\,)$ and $\mathbf{S} = (\,\mathbf{s}_1, \ldots, \mathbf{s}_{n-k}\,)$ be bases for $\mathbf{U}$ and $\mathbf{V}$ so that $\mathbf{T} = (\,\mathbf{r}_1, \ldots, \mathbf{r}_k, \mathbf{s}_1, \ldots, \mathbf{s}_{n-k}\,)$ is a basis for $\mathbf{X}$. Let $\mathcal{P}$ be a projection on $\mathbf{U}$ along $\mathbf{V}$.

   (a) Find the matrix representation of $\mathcal{P}$ with respect to $\mathbf{T}$ if $\mathcal{P}$ is interpreted as a LT from $\mathbf{X}$ into itself.

   (b) Find the matrix representation of $\mathcal{P}$ with respect to $(\mathbf{T}, \mathbf{R})$ if $\mathcal{P}$ is interpreted as a LT from $\mathbf{X}$ into $\mathbf{U}$.

49. Show that $\mathcal{P} : \mathsf{R}^2 \to \mathsf{R}^2$ defined as

$$\mathcal{P}(\alpha, \beta) = \left(\frac{\alpha - \beta}{2}, \frac{\beta - \alpha}{2}\right)$$

   is a projection. Characterize $\text{im}\,(\mathcal{P})$ and $\ker\,(\mathcal{P})$. Illustrate the decomposition of a vector into components in $\text{im}\,(\mathcal{P})$ and $\ker\,(\mathcal{P})$ with the help of a picture.

50. Prove facts (a-c) on p.126 concerning a projection.

51. Show that if $P \in \mathsf{F}^{n \times n}$ is idempotent, then it defines a projection on $\mathrm{im}\,(P)$ along $\ker\,(P)$. Hint: For any $\mathbf{x} \in \mathsf{F}^{n \times 1}$, let $\mathbf{u} = P\mathbf{x} \in \mathrm{im}\,(P)$ and $\mathbf{v} = (I - P)\mathbf{x}$, and show that $\mathbf{v} \in \ker\,(P)$.

52. A projection in the xyz space projects every point onto the plane described by

$$x + y + z = 0$$

along its normal $\mathbf{n} = \mathrm{col}\,[\,1, 1, 1\,]$.

(a) Find the projection $\mathbf{u}$ of the point $\mathbf{x} = \mathrm{col}\,[\,a, b, c\,]$ on the plane.

(b) Find a matrix P such that $\mathbf{u} = P\mathbf{x}$.

(c) Verify that $P^2 = P$.