Arindama Singh

Introduction to Linear Algebra

Contents

Preface **1**

1 Vector Spaces **3**
- 1.1 Introduction . . . 3
- 1.2 What is a vector space? . . . 5
- 1.3 Subspaces . . . 10
- 1.4 Span . . . 14
- 1.5 Linear independence . . . 18
- 1.6 Basis . . . 23
- 1.7 Dimension . . . 28
- 1.8 Extracting a basis . . . 32

2 Inner Product Spaces **36**
- 2.1 Inner Products . . . 36
- 2.2 Orthonormal basis . . . 39
- 2.3 Gram-Schmidt orthogonalization . . . 43
- 2.4 Best approximation . . . 46

3 Linear Transformations **50**
- 3.1 Linear maps . . . 50
- 3.2 Action on a basis . . . 53
- 3.3 Range space and null space . . . 56
- 3.4 Isomorphisms . . . 61
- 3.5 Adjoint of a Linear Transformation . . . 66

4 Matrices **71**
- 4.1 Matrix of a linear transformation . . . 71
- 4.2 Matrix operations . . . 76
- 4.3 Special types of matrices . . . 85
- 4.4 Trace and determinant . . . 91
- 4.5 Change of basis . . . 94
- 4.6 Equivalence . . . 100

5 Systems of Linear Equations **103**
- 5.1 Solvability . . . 103
- 5.2 Elementary row operations . . . 106

5.3 Row reduced echelon form 110
5.4 Gaussian and Gauss-Jordan elimination 114
5.5 Least squares solution 120

6 Spectral Representation 125
6.1 Eigenvalues and eigenvectors 125
6.2 Characteristic polynomial 128
6.3 Schur triangularization 135
6.4 Diagonalizability 141
6.5 Jordan form 147
6.6 Singular value decomposition 160
6.7 Polar decomposition 166

Bibliography 172

Index 174

Preface

Over the years, Linear Algebra has been accepted as a core course in Science and Engineering curricula at Bachelor's level. Teaching Linear Algebra through matrices and restricting to subspaces of finite dimensional real or complex vector spaces are often recommended at this stage. Such an approach has been taken in [10]. This approach teaches the students all that is required in applications leaving apart the geometry and the abstraction, which are at the heart of a budding mathematician. The present book tries to fill that gap by introducing the method of abstraction to a student at an early stage. It is guided by the geometry of vector spaces, which form the intuitive core of Linear Algebra. A slightly more advanced text that takes a similar approach to the subject is [9]; and the present text serves as an introduction to that. You will find many examples and problems common to these three books.

This book introduces basic notions connected with a vector space in an abstract but enjoyable way. It discusses matrices as specific types of linear transformations, which are maps that preserve the vector space structure. The approach shows why some specific operations on matrices have been defined in that particular manner. Elementary operations on matrices are used as computational tools in extracting a basis for a subspace, and also in solving linear systems. Various matrix factorizations have been derived via elegant matrix representations of linear transformations.

As this is an introductory text, we do not deal with direct sums, quotient spaces, orthogonal complements, dual spaces, LU decomposition, Cholesky factorization, and minimal polynomials, which are usually found in other books meant for texts in masters level. The driving force is to keep the material as elementary as possible without sacrificing on rigour, and at the same time obtain results that are important to matrix representations such as triangularization, diagonalization, Jordan form, singular value decomposition, rank factorization, and polar decomposition. On the way, we discuss Gram-Schmidt orthogonalization, QR factorization, best approximation, and the least squares solutions of linear systems.

Keeping the modest goal of an introductory text book on linear algebra, which may be covered in one semester, it presents bare essentials. Perhaps, more conceptual problems will be added to it in future. Right now, one may look at problems in [1, 9, 10]. Attempt has been made to state theorems in plain English, avoiding heavy use of symbols, and when necessary, explain the rigorous and exact meaning of the statement in the proof. It is felt that this will introduce the students to getting used to mathematical slang.

I acknowledge the pains taken by my students (M.Sc. mathematics and B.Tech) in pointing out typographical errors. Their difficulties in grasping the notions have contributed a lot towards the contents and this particular sequencing of topics. I cheerfully thank my colleagues Sounak Mishra, A. V. Jayanthan, and R. Balaji for using the earlier drafts for teaching Linear Algebra to undergraduate engineering and science students at IIT Madras.

By no means, it implies that the book is perfect and error-free. Indeed, I do not claim perfection. If you are using the book, then please point out improvements. I welcome you to write to me at asingh@iitm.ac.in.

This book is intended for internal circulation and academic use. Do not print it for commercial purposes; and do not purchase it from anyone; rather ask me for a free copy.

Chennai *Arindama Singh*
July 2018

1

Vector Spaces

1.1 Introduction

Consider a vector $\vec{u}$ in the plane. As we know, two vectors are equal iff they have the same direction and same length. Thus, we may draw $\vec{u}$ anywhere on the plane. Let us fix the x and the y axes, and mark the origin as $(0,0)$. Next, we draw $\vec{u}$ with its initial point at the origin. Its endpoint is some point, say, (a,b). Thus, by making the convention that each plane vector has its initial point at the origin, we see that any plane vector is identified with a point $(a,b) \in \mathbb{R}^2$.

Take another vector $\vec{v}$ with its initial point at the origin and endpoint as (c,d). By using the parallelogram law of adding two vectors, we see that the vector $\vec{u}+\vec{v}$ has initial point $(0,0)$ and endpoint $(a+c, b+d)$. This gives rise to addition of two points in $\mathbb{R}^2$ as in the following:

$$(a,b)+(c,d) = (a+c, b+d).$$

We call it *component-wise* addition. Similarly, for any $\alpha \in \mathbb{R}$, we see that the endpoint of $\alpha\vec{u}$ is $(\alpha a, \alpha b)$. That is, the scalar multiplication may be defined on $\mathbb{R}^2$ by the rule

$$\alpha(a,b) = (\alpha a, \alpha b).$$

These two operations of addition of two points and left-multiplication of a point by a scalar have the same properties as the corresponding operations on plane vectors.

Moreover, by the technique of identification, some other entities such as functions can also be seen to have similar properties. For example, consider S as the set of all functions from $\{1,2\}$ to $\mathbb{R}$. Such a function $f \in S$ is completely specified if we know how f acts on 1 and how it acts on 2. That is, any element $f \in S$ is completely specified by the ordered pair $(f(1), f(2))$, which is, of course, a point in $\mathbb{R}^2$. This way, S is identified with $\mathbb{R}^2$.

Now, what about the addition and scalar multiplication on S? Suppose $f \in S$ and $\alpha \in \mathbb{R}$. As per the scalar multiplication in $\mathbb{R}^2$, we have

$$\alpha(f(1), f(2)) = (\alpha f(1), \alpha f(2)).$$

That is, we may define the scalar multiplication on S by the rule

$$(\alpha f)(x) = \alpha f(x) \quad \text{for each } x \in \{1, 2\}.$$

This defines the new function αf obtained from the function f and the scalar α. Similarly, if $f, g \in S$, the addition on $\mathbb{R}^2$ says that

$$(f(1), f(2)) + (g(1), g(2)) = (f(1) + g(1), f(2) + g(2)).$$

Thus, our identification dictates us to define the addition of two functions in S by the rule

$$(f + g)(x) = f(x) + g(x) \quad \text{for each } x \in \{1, 2\}.$$

This defines the new function $f + g \in S$ obtained from the functions f and g.

As another example, consider solving the following linear equations:

$$x + y + z + w = 0, \quad 2x - y + z - 2w = 0.$$

By adding the two equations, we can eliminate the unknown y. Further, from the second equation, this y can be expressed in terms of the others. It yields

$$3x + 2z - w = 0, \quad y = 2x + z - 2w.$$

We may express w in terms of x and z using the first equation here. And then replacing this in the second, we obtain:

$$w = 3x + 2z, \quad y = -4x - 3z.$$

We see that any solution of the original equations can be computed by assigning values to the unknowns x and z arbitrarily, and then computing the values of y and w using the above expressions. Assigning x to a, and z to b, we may write a solution as the quadruple

$$(a, -4a - 3b, b, 3a + 2b), \quad \text{for arbitrary } a, b.$$

If we want only real solutions, then the set of all solutions of the original equations is given by

$$\text{Sol} = \{(a, -4a - 3b, b, 3a + 2b) : a, b \in \mathbb{R}\}.$$

If $(a_1, -4a_1 - 3b_1, b_1, 3a_1 + 2b_1), (a_2, -4a_2 - 3b_2, b_2, 3a_2 + 2b_2) \in \text{Sol}$, then their component-wise sum

$$(a_1 + a_2,\ -4(a_1 + a_2) - 3(b_1 + b_2),\ b_1 + b_2,\ 3(a_1 + a_2) + 2(b_1 + b_2)) \in \text{Sol}.$$

Further, multiplying a solution (component-wise) with a real number α also lands up in Sol:

$$(\alpha a_1, -4\alpha a_1 - 3\alpha b_1, \alpha b_1, 3\alpha a_1 + 2\alpha b_1) \in \mathrm{Sol}\,.$$

Moreover, we may associate any solution $(a, -4a-3b, b, 3a+2b) \in \mathrm{Sol}$ with the point $(a,b) \in \mathbb{R}^2$.

We see that points in $\mathbb{R}^2$, the functions in S, and also the solutions in Sol may all be thought of as vectors in the plane so far as these two operations of addition and scalar multiplication are concerned. That is, if we say that the set of all plane vectors is a vector space, then so are $\mathbb{R}^2$, the set of functions S, and also the set Sol of solutions of the homogeneous linear equations given earlier.

The essential properties of vectors in the plane are the following:

Addition is commutative and associative.

There exists a vector, denoted by $\vec{0}$, such that for all vectors $\vec{x}$, $\vec{x}+\vec{0} = \vec{x} = \vec{0}+\vec{x}$.

For each vector $\vec{x}$, there exists (possibly) another vector, denoted by $-\vec{x}$, such that $\vec{x}+(-\vec{x}) = \vec{0} = (-\vec{x})+\vec{x}$.

Addition distributes over scalar multiplication.

For each vector $\vec{x}$, $1 \cdot \vec{x} = \vec{x}$.

The approach is to find out the essential properties of plane vectors with respect to these operations. If we use these properties as the defining criteria, then our familiar way of working with plane vectors will yield results in any unfamiliar domain that satisfies these criteria. This is the greedy principle of abstraction used in mathematics. Using this principle, we end up with the notion of a vector space.

1.2 What is a vector space?

Let $\mathbb{R}$ denote the set of all real numbers and let $\mathbb{C}$ denote the set of all complex numbers. To talk about them both, we let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. The familiar properties of addition and multiplication of numbers from $\mathbb{F}$ are the following:

For all $a, b, c \in \mathbb{F}$,

1. $a+b = b+a$, $(a+b)+c = a+(b+c)$, $0+a = a$, $a-a := a+(-a) = 0$,
2. $ab = ba$, $(ab)c = a(bc)$, $1\,a = a$, $a\,a^{-1} = 1$ for $a \neq 0$,
3. $a(b+c) = (ab)+(ac) := ab+ac$, $(a+b)c = (ac)+(bc) := ac+bc$.

In fact, any set having at least two distinct elements, written 0 and 1, where the operations of addition and multiplication are defined, and satisfy the above properties is called a *field*. And we need a field for defining a vector space. To make the matter simple, we take our field as $\mathbb{F}$, which is either $\mathbb{R}$ or $\mathbb{C}$.

In what follows, we will write the Cartesian product of $\mathbb{F}$ with itself, taken n times as $\mathbb{F}^n$. That is,

$$\mathbb{F}^n := \{(a_1, \ldots, a_n) : a_1, \ldots, a_n \in \mathbb{F}\}.$$

An n-tuple in $\mathbb{F}^n$ is also written in two other forms, such as

$$[a_1 \ \cdots \ a_n], \qquad \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

The first is called a **row vector**, and the second is called a **column vector**. We also write a column vector using the *transpose* notation as in the following:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{is written as} \quad [a_1 \ \cdots \ a_n]^{\mathrm{t}}.$$

To distinguish the set of row vectors and the set of column vectors, we write

$$\begin{aligned} \mathbb{F}^{1\times n} &:= \{[a_1 \ \cdots \ a_n] : a_1, \ldots, a_n \in \mathbb{F}\}, \\ \mathbb{F}^{n\times 1} &:= \{[a_1 \ \cdots \ a_n]^{\mathrm{t}} : a_1, \ldots, a_n \in \mathbb{F}\}. \end{aligned}$$

In fact, when we are not very specific, we write a row vector and also a column vector as an n-tuple. So, both $\mathbb{F}^{1\times n}$ and $\mathbb{F}^{n\times 1}$ are written as $\mathbb{F}^n$.

A nonempty set V with two operations: $+$ (addition) that associates any two elements u, v in V to a single element $u + v$ in V, and $\cdot$ (scalar multiplication) that associates a number $\alpha \in \mathbb{F}$ and an element v in V to an element $\alpha \cdot v$ in V, is said to be a **vector space** over $\mathbb{F}$ iff it satisfies the following conditions:

(1) For all $x, y \in V$, $x + y = y + x$.

(2) For all $x, y, z \in V$, $(x + y) + z = x + (y + z)$.

(3) There exists an element $0 \in V$ such that $x + 0 = x$ for all $x \in V$.

(4) For each $x \in V$, there exists $(-x) \in V$ such that $x + (-x) = 0$.

(5) For each $x \in V$, $1 \cdot x = x$.

(6) For all $\alpha, \beta \in \mathbb{F}$ and for all $x, y \in V$, $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$.

(7) For each $\alpha \in \mathbb{F}$ and for all $x, y \in V$, $\alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y)$.

(8) For all $\alpha, \beta \in \mathbb{F}$ and for each $x \in V$, $(\alpha + \beta) \cdot x = (\alpha \cdot x) + (\beta \cdot x)$.

If the underlying field $\mathbb{F}$ is $\mathbb{R}$ we say that V is a **real vector space**; and if $\mathbb{F} = \mathbb{C}$, we say that V is a **complex vector space**. Elements of a vector space are called **vectors**. We will use u, v, w, x, y, z with or without subscripts for vectors. Elements of the underlying field $\mathbb{F}$ are called **scalars**, and they will be denoted by the Roman letters $a, b, c, \ldots$ and also by the Greek letters $\alpha, \beta, \gamma, \ldots$ with or without subscripts.

The symbol 0 will stand for both 'zero vector' and 'zero scalar'; you should know which one it represents from its use in a specific context. We write $\alpha \cdot x$ as αx. We accept the usual precedence rule of arithmetic; that is, the expression $\alpha x + \beta y$ will be understood as $(\alpha \cdot x) + (\beta \cdot y)$. Also, we will shorten $y + (-x)$ to $y - x$, and $(-x) + y$ to $-x + y$.

Notice that the first property of commutativity of addition allows us to write $x + y$ as $y + x$ whenever it facilitates our understanding. Similarly, the second property of associativity of addition allows us to put more parentheses or remove some according to our convenience. Analogous comments hold for other properties.

In mentioning the properties, we have used a short hand. In the third property, when we say "there exists an element $0 \in V$ such that", what we mean is "there exists an element $y \in V$, which we write as 0, such that". Similarly, in the fourth property, "for each $x \in V$ there exists $(-x) \in V$ such that" means: for each $x \in V$, there exists an element $y \in V$, which we denote as $-x$, such that".

Example 1.1

1. $\{0\}$ is a vector space over $\mathbb{F}$ with $0 + 0 = 0$ and $\alpha \cdot 0 = 0$ for each $\alpha \in \mathbb{F}$.
2. $\mathbb{F}$ is a vector space over $\mathbb{F}$ with addition and multiplication as in $\mathbb{F}$.
3. $\mathbb{R}^n$, $\mathbb{R}^{1\times n}$ and $\mathbb{R}^{n\times 1}$ are real vector spaces with component-wise addition and scalar multiplication, for any $n \in \mathbb{N}$.
4. $\mathbb{C}^n$, $\mathbb{C}^{1\times n}$ and $\mathbb{C}^{n\times 1}$ are complex vector spaces with component-wise addition and scalar multiplication, for any $n \in \mathbb{N}$.
5. Consider $\mathbb{C}$ with usual addition of complex numbers. For any $\alpha \in \mathbb{R}$, consider the scalar multiplication αx as the real number α multiplied with the complex number x for any $x \in \mathbb{C}$. Then $\mathbb{C}$ is a real vector space. Similarly, $\mathbb{C}^n$, $\mathbb{C}^{1\times n}$ and $\mathbb{C}^{n\times 1}$ are also real vector spaces.
6. $V = \{(a, b) \in \mathbb{R}^2 : b = 0\}$ is a real vector space under component-wise addition and scalar multiplication.
7. $V = \{(a, b) \in \mathbb{R}^2 : 2a + b = 0\}$ is a real vector space under component-wise addition and scalar multiplication.
8. Let $V = \{(a, b) \in \mathbb{R}^2 : 3a + 5b = 1\}$. We see that $(1/3, 0)$, $(0, 1/5) \in V$. But their sum $(1/3, 1/5) \notin V$. [Also, $3(1/3, 0) \notin V$.] Thus V is not a vector space with component-wise addition and scalar multiplication.

9. $\mathbb{F}_n[t] := \{a_0 + a_1 t + \cdots + a_n t^n : a_i \in \mathbb{F}\}$ with addition as the usual addition of polynomials and scalar multiplication as multiplication of a polynomial by a number, is a vector space over $\mathbb{F}$. Here, $\mathbb{F}_n[t]$ contains all polynomials in the variable t of degree less than or equal to n.
10. $\mathbb{F}[t]$:= the set of all polynomials (of all degrees) with coefficients from $\mathbb{F}$ is a vector space over $\mathbb{F}$ with + as the addition of two polynomials and $\cdot$ as the multiplication of a polynomial by a number from $\mathbb{F}$.
11. Let $V = \mathbb{R}^2$. For (a, b), $(c, d) \in V$ and $\alpha \in R$, define addition as component-wise addition, and scalar multiplication as in the following:

$$\begin{cases} \alpha(a,b) = (0,0) & \text{if } \alpha = 0 \\ \alpha(a,b) = (\alpha a, b/\alpha) & \text{if } \alpha \neq 0. \end{cases}$$

Then $(1+1)(0,1) = 2(0,1) = (0, 1/2)$ but $1(0,1) + 1(0,1) = (0,2)$. Thus V is not a vector space over $\mathbb{R}$.
12. Let V be the set of all functions from a nonempty set S to $\mathbb{F}$. Define addition of two functions and scalar multiplication by

$$(f+g)(x) = f(x) + g(x), \quad (\alpha f)(x) = \alpha f(x), \quad \text{for } x \in S,\ \alpha \in \mathbb{F}.$$

Then V is a vector space over $\mathbb{F}$. Here, *zero vector* is the zero map 0 given by $0(x) = 0$ for all $x \in S$; and for each $f \in V$, its additive inverse $-f$ is the map given by $(-f)(x) = -f(x)$ for $x \in S$.
13. Let V be the set of all functions from the closed interval $[a, b]$ to $\mathbb{R}$. Define addition and scalar multiplication as in (12). Then V is a real vector space.
14. Let V be the set of all two times differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ such that $f'' + f = 0$, where f' denotes the derivative of f with respect to the independent real variable t. Define addition and scalar multiplication as in (12). For $f, g \in V$ and $\alpha \in \mathbb{R}$, we see that

$$(f+g)'' + (f+g) = (f''+f) + (g''+g) = 0$$
$$(\alpha f)'' + (\alpha f) = \alpha(f''+f) = 0.$$

Therefore, the operations of addition and scalar multiplication are well-defined on V. It is easy to verify the eight properties. Hence V is a real vector space.
15. Let $V = \mathbb{R}^\infty$, the set of all sequences of real numbers. For sequences (a_n), $(b_n) \in V$, define $(a_n) + (b_n) := (a_n + b_n)$; and for $\alpha \in \mathbb{R}$, define $\alpha(a_n) := (\alpha a_n)$. Then V is a real vector space.
16. Let $V = c_{00}$, the set of all sequences of real numbers having a finite number of nonzero terms. With the addition and scalar multiplication as in (15), V is a real vector space. $\square$

Any vector that behaves like the 0 in the third property, is called a *zero vector*. Similarly, any vector that behaves like $(-x)$ for a given vector x, is called an *additive inverse* of x. In fact, there cannot be more than one zero vector, and there cannot be more than one additive inverse of any vector. Along with this we show some other expected facts.

***Theorem* 1.2**
Let V be a vector space over $\mathbb{F}$. Then the following are true.

(1) *Zero vector is unique.*

(2) *Each vector has a unique additive inverse.*

(3) *For any $x, y, z \in V$, and any $\alpha \in \mathbb{F}$, the following hold:*

  (a) *If* $x + y = x + z$, *then* $y = z$.
  (b) $0 \cdot x = 0$.
  (c) $\alpha \cdot 0 = 0$.
  (d) $(-1) \cdot x = -x$.
  (e) *If* $\alpha \cdot x = 0$, *then* $\alpha = 0$ *or* $x = 0$.

Proof (1) Let θ_1 and θ_2 be zero vectors in V. Then $\theta_2 = \theta_2 + \theta_1 = \theta_1 + \theta_2 = \theta_1$.
(2) Let $x \in V$. Suppose $x_1, x_2 \in V$ satisfy $x + x_1 = 0 = x + x_2$. Then

$$x_1 = x_1 + 0 = x_1 + x + x_2 = x_2 + x + x_1 = x_2 + 0 = x_2.$$

(3) Let $x, y, z \in V$ and let $\alpha \in \mathbb{F}$.

(a) $x + y = x + z \Rightarrow -x + x + y = -x + x + z \Rightarrow 0 + y = 0 + z \Rightarrow y = z$.

(b) $0 \cdot x + 0 = 0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x \Rightarrow 0 \cdot x = 0$, by (a).

(c) $0 + \alpha \cdot 0 = \alpha \cdot 0 = \alpha \cdot (0 + 0) = \alpha \cdot 0 + \alpha \cdot 0 \Rightarrow \alpha \cdot 0 = 0$, by (a).

(d) $x + (-1)x = 1 \cdot x + (-1) \cdot x = (1 + (-1)) \cdot x = 0 \cdot x = 0 \Rightarrow (-1)x = -x$.

(e) Suppose $\alpha \cdot x = 0$ but $\alpha \neq 0$. Then α^{-1} exists in $\mathbb{F}$. Consequently,

$$x = 1 \cdot x = \alpha^{-1} \cdot \alpha \cdot x = \alpha^{-1} \cdot 0 = 0. \qquad \blacksquare$$

Theorem 1.2 shows that we may work with vectors the same way as we work with numbers. However, we cannot multiply two vectors, since no such operation is available in a vector space. This is the reason we used α^{-1} in the proof of Theorem 1.2(3e) instead of using x^{-1}. In fact, there is no such vector as x^{-1}.

In what follows, we write V as a vector space without mentioning the underlying field. We assume that the underlying field is $\mathbb{F}$ which may be $\mathbb{R}$ or $\mathbb{C}$. We consider $\mathbb{F}^n$ to be a vector space over $\mathbb{F}$. In particular, $\mathbb{R}^n$ is taken as a real vector space; and if nothing is specified, we take $\mathbb{C}^n$ as a complex vector space.

Exercises for § 1.2

1. In each of the following a nonempty set V is given and some operations are defined. Check whether V is a vector space with these operations. Whenever V is a vector space, write explicitly the zero vector and the additive inverse of any vector $v \in V$.
   (a) $V = \{(a,0) : a \in \mathbb{R}\}$ with $+$ and $\cdot$ as in $\mathbb{R}^2$.
   (b) $V = \{(a,b) \in \mathbb{R}^2 : 2a+3b=0\}$ with $+$ and $\cdot$ as in $\mathbb{R}^2$.
   (c) $V = \{(a,b) \in \mathbb{R}^2 : a+b=1\}$ with $+$ and $\cdot$ as in $\mathbb{R}^2$.
   (d) $V = \mathbb{R}^2$ with $+$ as in $\mathbb{R}^2$, and $\cdot$ defined by $0 \cdot (a,b) = (0,0)$, and for $\alpha \neq 0$, $\alpha \in \mathbb{R}$, $\alpha \cdot (a,b) = (a/\alpha, \alpha b)$.
   (e) $V = \mathbb{R}^2$ with $+$ as in $\mathbb{R}^2$, and $\cdot$ defined by $\alpha \cdot (a,b) = (a,0)$ for $\alpha \in \mathbb{R}$.
   (f) $V = \mathbb{C}^2$ with $+$ and $\cdot$ defined by $(a,b)+(c,d) = (a+2c, b+3d)$, and $\alpha \cdot (a,b) = (\alpha a, \alpha b)$ for (a,b), $(c,d) \in V$ and $\alpha \in \mathbb{C}$.
   (g) $V = [0,\infty)$ with addition $\oplus$ and scalar multiplication $\odot$ defined by $x \oplus y = xy$, and $\alpha \odot x = x^\alpha$ for $x, y \in V$ and $\alpha \in \mathbb{R}$.
   (h) $(\mathbb{R} \setminus \mathbb{Q}) \cup \{0,1\}$ with $+$ and $\cdot$ as in $\mathbb{R}$.
   (i) $V = \mathbb{R}$ with addition $\oplus$ as $a \oplus b = a+b-2$ and the scalar multiplication defined by $\alpha \odot b = \alpha b + 2(\alpha - 1)$ for $\alpha \in \mathbb{R}$.
   (j) $V = \mathbb{R}^n$ with addition $\oplus$ as $u \oplus v = u+v-w$ and scalar multiplication $\odot$ as $\alpha \odot u = \alpha(u-w)+w$, where w is a fixed given vector in V, and $\alpha \in \mathbb{R}$, $u, v \in V$.
   (k) $V = \{(a,b) \in \mathbb{R}^2 : a+b=1\}$ with addition $\oplus$ and scalar multiplication $\odot$ as $(a,b) \oplus (c,d) = (a+c-1, b+d)$, and $\alpha \odot (a,b) = (\alpha a - \alpha + 1, \alpha b)$.
2. Is the set of all polynomials of degree 5 with usual addition and scalar multiplication of polynomials a vector space?
3. Let V be a vector space over $\mathbb{F}$. Show the following:
   (a) For all $x, y, z \in V$, $x+y = z+y$ implies $x = z$.
   (b) Let $\alpha, \beta \in \mathbb{F}, x \in V, x \neq 0$. Then, $\alpha x \neq \beta x$ iff $\alpha \neq \beta$.
   (c) If V has two vectors, then V has an infinite number of vectors.

1.3 Subspaces

Consider the following two nonempty subsets of $\mathbb{R}^2$:

$$U = \{(a,b) \in \mathbb{R}^2 : 2a+b=0\}, \quad W = \{(a,b) \in \mathbb{R}^2 : 2a+b=1\}.$$

We have seen that U is a vector space with the same operations of addition and scalar multiplication as in $\mathbb{R}^2$. Of course, the operations are well defined on U.

That is, whenever $x, y \in U$ and $\alpha \in \mathbb{F}$, we have $x + y \in U$ and $\alpha x \in U$. But W is not a vector space with the same operations. In fact, the sum of two vectors from W does not necessarily result in a vector from W. For instance, $(0, 1) \in W$ and $(1, -1) \in W$ but $(0, 1) + (1, -1) = (1, 0) \notin W$. Similarly, multiplying a scalar with a vector from W may not result in a vector from W. We would like to separate out the first interesting case of U.

Let V be a vector space. A subset U of V is called a **subspace of** V iff the following conditions are satisfied:

(1) $U \neq \emptyset$.

(2) For all $x, y \in U$, $x + y \in U$.

(3) For each $x \in U$ and for each $\alpha \in \mathbb{F}$, $\alpha x \in U$.

A subspace of V which is not equal to V is called a **proper subspace** of V.

Notice that the second and the third conditions together, in the definition of a subspace, is equivalent to the following single condition:

$$\text{For all } x, y \in U \text{ and for each } \alpha \in U,\quad x + \alpha y \in U.$$

Example 1.3

1. If V is a vector space, then $\{0\}$ is a subspace of V. It is called the *zero subspace* of V.
2. Any vector space V is a subspace of itself.
3. The x-axis, that is, $\{(a, 0) : a \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$. We identify the x-axis with $\mathbb{R}$ and say that $\mathbb{R}$ is a proper subspace of $\mathbb{R}^2$.
4. Let $m < n$. Then $\{(a_1, \ldots, a_m, 0, \ldots, 0) \in \mathbb{F}^n\}$ is a subspace of $\mathbb{F}^n$. We identify this subspace as $\mathbb{F}^m$, and say that $\mathbb{F}^m$ is a proper subspace of $\mathbb{F}^n$.
5. All straight lines in $\mathbb{R}^2$ passing through the origin are proper subspaces of $\mathbb{R}^2$. What are the other subspaces?
6. $W = \{(a, b, c) \in \mathbb{R}^3 : a - 3b + 2c = 0\}$ is a proper subspace of $\mathbb{R}^3$.
7. All planes and all straight lines in $\mathbb{R}^3$ passing through the origin are proper subspaces of $\mathbb{R}^3$.
8. $\mathbb{F}_m[t]$ is a proper subspace of $\mathbb{F}_n[t]$ for $m < n$.
9. $C[a,\ b]$:= the set of all continuous functions from $[a,\ b]$ to $\mathbb{R}$, is a proper subspace of the set of all functions from the closed interval $[a,\ b]$ to $\mathbb{R}$, with usual operations of + and $\cdot$; see Example 1.1(12).
10. Let $V = C[-1,\ 1]$. Let $U = \{f \in V : f \text{ is an odd function }\}$. As a convention, the zero function is taken as an odd function. So, $U \neq \emptyset$. If $f,\ g \in U$ and $\alpha \in \mathbb{R}$, then

    $$(f + \alpha g)(-x) = f(-x) + \alpha g(-x) = -f(x) + \alpha(-g(x)) = -(f + \alpha g)(x).$$

    So, $f + \alpha g \in U$. Therefore, U is a subspace of V.

11. $\mathcal{R}[a, b] :=$ the set of all Riemann integrable functions from $[a, b]$ to $\mathbb{R}$ is a real vector space. And the vector space $C[a, b]$ is a proper subspace of $\mathcal{R}[a, b]$.
12. $C^k[a, b] :=$ the set of all k times continuously differentiable functions from $[a, b]$ to $\mathbb{R}$ is a proper subspace of $C[a, b]$.
13. Interpret each polynomial $p(t) \in \mathbb{R}_n[t]$ as a function, where $t \in [a, b]$. Then $\mathbb{R}_n[t]$ is a proper subspace of $C[a, b]$. □

In fact, a subspace is a vector space on its own right, with the operations inherited from the parent vector space.

***Theorem* 1.4**

Let V be a vector space over $\mathbb{F}$ with $+$ as the addition and $\cdot$ as the scalar multiplication. Let U be a subspace of V. Then U is a vector space over $\mathbb{F}$ with the addition as the restriction of $+$ to U, and scalar multiplication as the restriction of $\cdot$ to U.

Proof Since U is a subspace of V, the restriction of $+$ and $\cdot$ to U are well defined operations on U. The commutativity and associativity of addition, distributive properties, and scalar multiplication with 1 are satisfied in V; and hence, they are true in U too. Therefore, we only need to verify the existence of 'zero vector' and 'additive inverse' in U.

Since $U \neq \emptyset$, there exists an $x \in U$. In V, we know that $-x = (-1)x$. Since $(-1)x \in U$, we see that $-x \in U$. Now, since both x and $-x \in U$, we have $x + (-x) = 0 \in U$. Moreover, this 0 serves as the additive identity in U; and this $-x$ serves as the additive inverse of x in U.

Therefore, U is a vector space with the restricted operations. ∎

The proof of Theorem 1.4 reveals that the zero vector of a subspace is the same zero vector of the parent vector space. And the additive inverse of a vector x in U is the same additive inverse $-x$ in V.

Verify that all subspaces given in Example 1.3 are vector spaces on their own right.

Given two subspaces, can we use set operations for obtaining new subspaces?

***Theorem* 1.5**

The intersection of two subspaces of a vector space is also a subspace.

Proof Let U and W be subspaces of a vector space V. Since $0 \in U$ and $0 \in W$, $U \cap W \neq \emptyset$.

Let $x, y \in U \cap W$ and let $\alpha \in \mathbb{F}$. Then $x + \alpha y \in U$ and $x + \alpha y \in W$, since they are subspaces. So, $x + \alpha y \in U \cap W$. Therefore, $U \cap W$ is a subspace. ∎

For illustration, consider two distinct planes passing through the origin in $\mathbb{R}^3$. Their intersection is a straight line passing through the origin; it is a subspace of $\mathbb{R}^3$.

However, union of two subspaces need not be a subspace. For example, consider the x-axis $X = \{(a,0) : a \in \mathbb{R}\}$ and the y-axis $Y = \{(0,b) : b \in \mathbb{R}\}$. These are subspaces of $\mathbb{R}^2$. But their union is not a subspace of $\mathbb{R}^2$. Reason?

$$(1,0),\ (0,1) \in X \cup Y, \text{ but } (1,0) + (0,1) = (1,1) \notin X \cup Y.$$

***Theorem* 1.6**
Union of two subspaces is a subspace iff one of them is a subset of the other.

Proof Let U and W be subspaces of a vector space V with $U \subseteq W$ or $W \subseteq U$. If $U \subseteq W$, then $U \cup W = W$. If $W \subseteq U$, then $U \cup W = U$. In either case, $U \cup W$ is a subspace of V.

Conversely, assume that $U \cup W$ is a subspace of V. Suppose, on the contrary that $U \nsubseteq W$ and $W \nsubseteq U$. Then there exist vectors x and y such that $x \in U$, $x \notin W$, and $y \in W$, $y \notin U$. Now, both $x, y \in U \cup W$. Where is $x + y$?

If $x + y \in U$, then $y = (x + y) - x \in U$, which is wrong.

If $x + y \in W$, then $x = (x + y) - y \in W$; which is also wrong.

Therefore $x + y \notin U \cup W$. This contradicts the assumption that $U \cup W$ is a subspace of V. ∎

Exercises for § 1.3

1. In each of the following, a vector space V and a subset U are given. Check whether U is a subspace of V.
   (a) $V = \mathbb{R}^2$, $U = \{(a,b) \in V : b = 2a - \alpha\}$ for some $\alpha \neq 0$.
   (b) $V = \mathbb{R}^3$, $U = \{(a,b,c) \in V : 2a - b - c = 0\}$.
   (c) $V = \mathbb{F}_3[t]$, $U = \{a + bt + ct^2 + dt^3 \in V :\ a = 0\}$.
   (d) $V = \mathbb{R}_3[t]$, $U = \{at + bt^2 + ct^3 \in V : a, b, c \in \mathbb{R}\}$.
   (e) $V = \mathbb{C}_3[t]$, $U = \{a + bt + ct^2 + dt^3 \in V : a + b + c + d = 0\}$.
   (f) $V = \mathbb{C}_3[t]$, $U = \{a + bt + ct^2 + dt^3 \in V : a, b, c, d \text{ integers}\}$.
   (g) $V = C[-1,1]$, $U = \{f \in V : f \text{ is an even function}\}$.
   (h) $V = C[0,1]$, $U = \{f \in V : f(x) \geq 0 \text{ for all } x\}$.
2. Describe all subspaces of $\mathbb{R}^2$, and of $\mathbb{R}^3$.
3. Let $U = \{(a_1, a_2, \ldots, a_n) : a_1, a_2, \ldots, a_n \in \mathbb{R},\ a_1 + 2a_2 + \cdots + na_n = b\}$. For which real numbers b, is U a subspace of $\mathbb{R}^n$?
4. Let U be a subspace of V and let V be a subspace of W. Is U a subspace of W?
5. Why is the set of all polynomials of degree n not a vector space, with usual addition and scalar multiplication of polynomials?

6. Show that $\mathcal{R}[a,b]$, the set of all real valued Riemann integrable functions on $[a,b]$, is a vector space.
7. Let S be a nonempty set and let $s \in S$. Let V be the set of all functions $f : S \to \mathbb{R}$ with $f(s) = 0$. Is V a vector space over $\mathbb{R}$ with the usual addition and scalar multiplication of functions?
8. Determine whether the following are real vector spaces:
   (a) $\mathbb{R}^\infty :=$ the set of all sequences of real numbers.
   (b) $\ell^\infty :=$ the set of all bounded sequences of real numbers.
   (c) $\ell^1 :=$ the set of all absolutely convergent sequences of real numbers.
9. Show that the set $B(S)$ of all bounded functions from a nonempty set S to $\mathbb{R}$ is a real vector space.

1.4 Span

We have seen that the union of two subspaces may fail to be a subspace since the union need not be closed under the operations. It is quite possible that we enlarge the union so that it becomes a subspace. Of course, a trivial enlargement of the union is the whole vector space. A better option would be to enlarge the union in a minimal way; that is by including only those vectors that are required to obtain a subspace.

To see the requirement in a general way, let S be a nonempty subset of a vector space V. Suppose S is not a subspace of V. Let $u \in S$. In any minimal enlargement of S, all vectors of the form αu must be present. Similarly, if $v \in S$, then all vectors of the form βv must be present in the enlargement. Then the vectors of the form $\alpha u + \beta v$ must also be in the enlargement. In general, if $v_1, \ldots, v_n \in S$, then in this enlargement, we must have all vectors of the form

$$\alpha_1 v_1 + \cdots + \alpha_n v_n \quad \text{for } \alpha_1, \ldots, \alpha_n \in \mathbb{F}.$$

Moreover, trivially, if $S = \varnothing$, then the minimal way we can enlarge it to a subspace is the zero space $\{0\}$.

Let V be a vector space. Let $v_1, \ldots, v_n \in V$. A vector $v \in V$ is said to be a **linear combination** of $v_1, \ldots, v_n$ iff $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$.

Let S be a nonempty subset of V. Then the set of all linear combinations of elements of S is called the **span** of S, and is denoted by $\operatorname{span}(S)$. Further, span of the empty set is taken to be $\{0\}$.

Notice that a linear combination is always a finite sum. If S is a finite set, say $S = \{v_1, \ldots, v_m\}$, then

$$\mathrm{span}\,(S) = \{\alpha_1 v_1 + \cdots + \alpha_m v_m : \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}.$$

In general if S is a nonempty set, then we can write its span as

$$\mathrm{span}\,(S) = \{\alpha_1 v_1 + \cdots + \alpha_n v_n : \alpha_1, \ldots, \alpha_n \in \mathbb{F},\ v_1, \ldots, v_n \in S \text{ for some } n \in \mathbb{N}\}.$$

Example 1.7

1. $\mathrm{span}\,(\varnothing) = \mathrm{span}\,(\{0\}) = \{0\}$.
2. $\mathbb{C} = \mathrm{span}\,\{1, i\}$ with scalars from $\mathbb{R}$.
3. $\mathbb{C} = \mathrm{span}\,\{1\} = \mathrm{span}\,\{1, i\}$ with scalars from $\mathbb{C}$.
4. Let $e_1 = (1, 0),\ e_2 = (0, 1)$. Then $\mathbb{F}^2 = \mathrm{span}\,\{e_1, e_2\}$, where scalars are from $\mathbb{F}$.
5. Let e_i denote the vector in $\mathbb{F}^n$ having 1 at the ith place and 0 elsewhere. Then $\mathbb{F}^n = \mathrm{span}\,\{e_1, \ldots, e_n\}$, where scalars are from $\mathbb{F}$.
6. $\mathrm{span}\,\{(1, 2, 3)\}$ in $\mathbb{R}^3$ is the straight line passing through the origin and the point $(1, 2, 3)$.
7. $\mathrm{span}\,\{(1, 2, 3), (3, 2, 1)\}$ in $\mathbb{R}^3$ is the plane passing through the points $(0, 0, 0)$, $(1, 2, 3)$ and $(3, 2, 1)$.
8. $\mathbb{F}_3[t] = \mathrm{span}\,\{1, t, t^2, t^3\}$ with scalars from $\mathbb{F}$.
9. $\mathbb{F}[t] = \mathrm{span}\,\{1, t, t^2, \ldots\}$ with scalars from $\mathbb{F}$. □

Caution: The set S can have infinitely many elements, but a linear combination is a sum of finitely many elements from S, multiplied with some scalars.

We show that the notion of 'span' serves its purpose in enlarging a subset to a minimal subspace.

***Theorem* 1.8**

The span of a subset is the minimal subspace that contains the subset.

Proof Let S be a subset of a vector space V over $\mathbb{F}$. If $S = \varnothing$, then $\mathrm{span}\,(S) = \{0\}$; which is clearly the smallest subspace containing $\varnothing$. So, let $S \neq \varnothing$. Let $x, y \in \mathrm{span}\,(S)$. Then $x = a_1 x_1 + \cdots + a_n x_n$ and $y = b_1 y_1 + \cdots + b_m y_m$ for some $a_i, b_j \in \mathbb{F}$ and $x_i, y_j \in S$. Let $\alpha \in \mathbb{F}$. Consequently,

$$x + \alpha y = a_1 x_1 + \cdots + a_n x_n + \alpha b_1 y_1 + \cdots + \alpha b_m y_m \in \mathrm{span}\,(S).$$

Therefore, $\mathrm{span}\,(S)$ is a subspace. If U is a subspace containing S, then U contains all linear combinations of vectors from S. That is, U contains $\mathrm{span}\,(S)$. So, $\mathrm{span}\,(S)$ is the minimal subspace containing S. ∎

In particular, if U and W are subspaces of a vector space V, then any linear combination of vectors from U is a vector in U. Similarly, any linear combination of vectors from W is also a vector in W. We guess that any vector in span $(U \cup W)$ is a sum of vectors from U and W.

Let S_1 and S_2 be nonempty subsets of a vector space V. The **Sum of** these subsets is defined as

$$S_1 + S_2 := \{x + y : x \in S_1,\ y \in S_2\}.$$

For example, consider the sum of the x-axis and the y-axis in $\mathbb{R}^2$. We see that each vector $(a, b) \in \mathbb{R}^2$ can be written as $(a, b) = (a, 0) + (0, b)$. Hence the sum of the x-axis and the y-axis is the whole of $\mathbb{R}^2$.

In $\mathbb{R}^3$, the sum of x-axis and the y-axis is the xy-plane.

We show that the sum of two subspaces is the minimal enlargement of their union so that the enlarged set is a subspace.

***Theorem* 1.9**

The sum of two subspaces of a vector space is equal to the span of their union, the minimal subspace containing their union.

Proof Let U and W be subspaces of a vector space V. Since $U \subseteq U + W$, $U + W \neq \emptyset$. Let $z_1, z_2 \in U + W$. There exist vectors $x_1, x_2 \in U$ and $y_1, y_2 \in W$ such that $z_1 = x_1 + y_1$ and $z_2 = x_2 + y_2$. Let $\alpha \in \mathbb{F}$. Then

$$z_1 + \alpha z_2 = x_1 + y_1 + \alpha(x_2 + y_2) = (x_1 + \alpha x_2) + (y_1 + \alpha y_2) \in U + W.$$

Hence $U + W$ is a subspace of V that contains $U \cup W$. Further, $U + W \subseteq$ span $(U \cup W)$. Since span $(U \cup W)$ is the minimal subspace of V that contains $U \cup W$, it follows that $U + W \supseteq$ span $(U \cup W)$. Hence, $U + W =$ span $(U \cup W)$, the minimal subspace of V that contains $U \cup W$. ∎

Though a vector space is very large, there can be a small subset whose span is the vector space. For example, $\mathbb{R}^2 = \text{span}\{(1, 0), (0, 1)\}$.

A subset S of a vector space V is said to **span** V iff span $(S) = V$. In this case, we also say that S is a **spanning set** of V, and V is **spanned by** S.

Example 1.10

1. The subset $\{(1, 0), (0, 1)\}$ of $\mathbb{R}^2$ spans $\mathbb{R}^2$.
2. The subset $\{(1, 2), (2, 1), (2, 2)\}$ of $\mathbb{R}^2$ spans $\mathbb{R}^2$.
3. The subset $\{e_1, \ldots, e_n\}$ of $\mathbb{F}^n$ is a spanning set of $\mathbb{F}^n$.
4. $\mathbb{F}_n[t]$ is spanned by $\{1, t, \ldots, t^n\}$.
5. The subset $\{(1, 2)\}$ of $\mathbb{R}^2$ spans the vector space $\{\alpha(1, 2) : \alpha \in \mathbb{R}\}$. Here, the vector space is the straight line that passes through the origin and the point $(1, 2)$; it is a proper subspace of $\mathbb{R}^2$. For instance, $(1, 1) \notin \text{span}\{(1, 2)\}$. We see that $\{(1, 2)\}$ does not span $\mathbb{R}^2$.

6. The subset $\{(1,1,1),(0,1,1),(1,-1,-1),(1,3,3)\}$ of $\mathbb{R}^3$ spans the plane in $\mathbb{R}^3$ that contains the points $(0,0,0)$, $(1,1,1)$ and $(0,1,1)$. Reason?

$$\begin{aligned} &a(1,1,1)+b(0,1,1)+c(1,-1,-1)+d(1,3,3) \\ &\qquad = (a+c+d)(1,1,1)+(b-2c+2d)(0,1,1). \end{aligned} \qquad \square$$

If S is a spanning set of a vector space V, then any superset of S is also a spanning set of V. However, a subset of S may or may not be a spanning set. In Example 1.6 (6), the plane is spanned by $\{(1,1,1),\ (0,1,1)\}$; but it is not spanned by $\{(1,1,1)\}$.

The vector space V spans itself, though our interest is in finding smaller subsets of V that would span V.

Exercises for § 1.4

1. Do the polynomials t^3-2t^2+1, $4t^2-t+3$, and $3t-2$ span $\mathbb{F}_3[t]$?
2. What is $\operatorname{span}\{t^n : n=0,2,4,6,\ldots\}$ in $\mathbb{R}[t]$?
3. In $\mathbb{F}^3$, let $e_1=(1,0,0)$, $e_2=(0,1,0)$ and $e_3=(0,0,1)$. What is $\operatorname{span}\{e_1+e_2, e_2+e_3, e_3+e_1\}$?
4. Let $u, v_1, v_2, \ldots, v_n$ be $n+1$ distinct vectors in a vector space V. Take $S_1=\{v_1,v_2,\ldots,v_n\}$ and $S_2=\{u,v_1,v_2,\ldots,v_n\}$. Prove that $u\in\operatorname{span}(S_1)$ iff $\operatorname{span}(S_1)=\operatorname{span}(S_2)$.
5. Let S be a subset of a vector space V. Show that S is a subspace iff $S=\operatorname{span}(S)$.
6. Let $u_1(t)=1$, and for $n=2,3,\ldots$, let $u_n(t)=1+t+\ldots+t^{n-1}$. Show that $\{u_1,\ldots,u_n\}$ spans $\mathbb{F}_{n-1}[t]$. Is it true that $\{u_1,u_2,\ldots\}$ spans $\mathbb{F}[t]$?
7. Let S be a subset of a vector space V. Prove that $\operatorname{span}(S)$ is the intersection of all subspaces that contain S.
8. Show that each vector space has at least two spanning sets.
9. We know that $e^t = 1+t+\frac{1}{2!}t^2+\cdots$ for each $t\in\mathbb{R}$. Does it imply that $e^t\in\operatorname{span}\{1,t,t^2,\ldots\}$?
10. Let V be the real vector space of all functions from $\{1,2\}$ to $\mathbb{R}$. Construct a spanning set of V with two elements.
11. Let A and B be subsets of a vector space V. Prove or disprove:
    (a) $\operatorname{span}(\operatorname{span}(A))=\operatorname{span}(A)$.
    (b) If $A\subseteq B$, then $\operatorname{span}(A)\subseteq\operatorname{span}(B)$.
    (c) $\operatorname{span}(A\cap B)\subseteq\operatorname{span}(A)\cap\operatorname{span}(B)$.
    (d) $\operatorname{span}(A)\cap\operatorname{span}(B)\subseteq\operatorname{span}(A\cap B)$.
    (e) $\operatorname{span}(A)\setminus\operatorname{span}(B)\subseteq\operatorname{span}(A\setminus B)$.
    (f) $\operatorname{span}(A\setminus B)\subseteq(\operatorname{span}(A)\setminus\operatorname{span}(B))\cup\{0\}$.
12. Give suitable real vector spaces U,V,W so that $U+V=U+W$ but $V\neq W$.

13. Let U and W be subspaces of a vector space such that $U \cap W = \{0\}$. Prove that if $x \in U + W$, then there exist unique $u \in U, w \in W$ such that $x = u + w$.
14. Let U, V, and W be subspaces of a vector space X.
    (a) Prove that $(U \cap V) + (U \cap W) \subseteq U \cap (V + W)$.
    (b) Give suitable U, V, W, X so that $U \cap (V + W) \nsubseteq (U \cap V) + (U \cap W)$.
    (c) Prove that $U + (V \cap W) \subseteq (U + V) \cap (U + W)$.
    (d) Give suitable U, V, W, X so that $(U + V) \cap (U + W) \nsubseteq U + (V \cap W)$.
15. Let e_i be the sequence $(0, 0, \ldots, 0, 1, 0, 0, \ldots)$ where the ith term is 1 and the rest are all 0. What is $\operatorname{span}(\{e_1, e_2, \ldots\})$?

1.5 Linear independence

We would like to have a spanning set none of whose proper subsets is a spanning set. In that case, no vector in such a set is in the span of the rest.

Let S be a subset of a vector space V. S is said to be **linearly dependent** iff there exists a vector $v \in S$ such that $v \in \operatorname{span}(S \setminus \{v\})$.

A list of vectors $v_1, \ldots, v_n$ is said to be **linearly dependent** iff either a vector is repeated in the list or the set $\{v_1, \ldots, v_n\}$ is linearly dependent.

A set or a list is called **linearly independent** iff it is not linearly dependent.

When a list of vectors $v_1, \ldots, v_n$ is linearly dependent (or independent), we say that the vectors $v_1, \ldots, v_n$ are linearly dependent (or independent).

Observe that $\varnothing$ is linearly independent, and $\{0\}$ is linearly dependent. Moreover, a nonempty set of vectors is linearly dependent iff one of the vectors in the set is a linear combination of others.

Example 1.11

1. The set $\{(1, 0), (0, 1)\}$ is linearly independent in $\mathbb{R}^2$.
2. The set $\{(1, 1), (2, 1), (2, 3)\}$ is linearly dependent in $\mathbb{R}^2$.
3. The vectors $(1, 1), (2, 1), (2, 3)$ are linearly dependent in $\mathbb{R}^2$.
4. The vectors $1, i$ in the real vector space $\mathbb{C}$ are linearly independent.
5. The vectors $1, i$ in the complex vector space $\mathbb{C}$ are linearly dependent.
6. The set $\{1, 1 + t, t^2 - t^3, 2t + t^2 - t^3, 1 + t + t^2 + t^3\}$ is linearly dependent in $\mathbb{F}_3[t]$. Reason: $2t + t^2 - t^3 = -2(1) + 2(1 + t) + (t^2 - t^3)$.
7. The set of functions $\{t,\ \cos t\}$ is linearly independent in $C[0, \pi/2]$. For, if $\cos t = \alpha t$, then at $t = 0$, we have $1 = \alpha \times 0 = 0$, which is impossible. On the other hand, if $t = \beta \cos t$, then at $t = \frac{\pi}{2}$, we obtain $\frac{\pi}{2} = \beta \cos(\frac{\pi}{2}) = 0$, which is again impossible.

□

Given a set of vectors, how do we determine whether one of them is a linear combination of others? Or, how do we show that none of the vectors in the set is a linear combination of others?

***Theorem* 1.12**
Let $v_1, \ldots, v_n$ be vectors in a vector space V. Then

(1) $\{v_1, \ldots, v_n\}$ *is linearly dependent iff* $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ *where at least one of the scalars* $\alpha_1, \ldots, \alpha_n$ *is nonzero.*

(2) $\{v_1, \ldots, v_n\}$ *is linearly independent iff for scalars* $\alpha_1, \ldots, \alpha_n$,

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \;\; \textit{implies} \;\; \alpha_1 = \cdots = \alpha_n = 0.$$

Proof (1) Suppose that $\{v_1, \ldots, v_n\}$ is linearly dependent. Then we have at least one $j \in \{1, \ldots, n\}$ such that

$$v_j = \alpha_1 v_1 + \cdots + \alpha_{j-1} v_{j-1} + \alpha_{j+1} v_{j+1} + \cdots + \alpha_n v_n.$$

Then $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ where $\alpha_j = -1 \neq 0$.

Conversely, suppose that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$, where $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, and we have (some) $j \in \{1, \ldots, n\}$ such that $\alpha_j \neq 0$. Then

$$v_j = -(\alpha_j)^{-1} \alpha_1 v_1 - \cdots - (\alpha_j)^{-1} \alpha_n v_n.$$

Therefore, $\{v_1, \ldots, v_n\}$ is linearly dependent.

(2) If the given condition is satisfied, then there does not exist a nonzero α_j such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. That is, $\{v_1, \ldots, v_n\}$ is linearly independent.

Conversely, if the condition is not satisfied, then at least one $\alpha_j \neq 0$ and $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. That is, the set $\{v_1, \ldots, v_n\}$ is linearly dependent. ∎

Theorem 1.12 says that given vectors are linearly independent iff the only way the zero vector can be written as a linear combination of the vectors is the *trivial* linear combination. It thus gives us a method of determining whether a given set of vectors is linearly dependent or independent. We start with the equation

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$$

and try to determine the scalars $\alpha_1, \ldots, \alpha_n$. If it turns out that all of them must be zero, then the vectors are linearly independent. In case, we fail to show this, we would obtain a *nontrivial* linear combination of the zero vector. That would supply us with a vector that can be expressed as a linear combination of others.

Example 1.13

1. $\{(1,0),(1,1)\}$ is linearly independent in $\mathbb{R}^2$. To illustrate Theorem 1.12, we assume that $\alpha(1,0)+\beta(1,1)=(0,0)$. It gives $(\alpha+\beta,\beta)=(0,0)$. That is, $\alpha=\beta=0$.
2. Is $\{(1,0,0),(1,2,0),(1,1,1)\}$ linearly independent in $\mathbb{R}^3$? We start with the equation
$$\alpha(1,0,0)+\beta(1,2,0)+\gamma(1,1,1)=(0,0,0).$$
Comparing the components, we obtain: $\alpha+\beta+\gamma=0$, $2\beta+\gamma=0$, and $\gamma=0$. It implies $\alpha=\beta=\gamma=0$.
Thus, the vectors are linearly independent.
3. $\{1,t,t^2\}$ is a linearly independent set in $\mathbb{F}_2[t]$. For, if $a\,1+bt+ct^2=0$, the zero polynomial, then $a=b=c=0$.
4. Is $\{\sin t,\cos t\}$ linearly independent in $C[0,\ \pi]$? Assume that
$$\alpha\sin t+\beta\cos t=0.$$
Notice that the 0 on the right hand side is the zero function. Putting $t=0$ we get $\beta=0$. By taking $t=\pi/2$, we get $\alpha=0$. Hence the set $\{\sin t,\cos t\}$ is linearly independent.
5. If $\{u_1,\ldots,u_n\}$ is linearly dependent, then for any vector $v\in V$, the set $\{u_1,\ldots,u_n,v\}$ is linearly dependent. Why?
For, suppose $\alpha_1u_1+\cdots+\alpha_nu_n=0$ with $\alpha_i\neq0$ for at least one i. Then $\alpha_1u_1+\cdots+\alpha_nu_n+0\cdot v=0$ with that $\alpha_i\neq0$. That is, $\{u_1,\ldots,u_n,v\}$ is linearly dependent. $\square$

We observe the following:

1. Any set containing the zero vector is linearly dependent.
2. The vectors u,v are linearly dependent iff one of them is a scalar multiple of the other.
3. Each superset of a linearly dependent set is linearly dependent.
4. Each subset of a linearly independent set is linearly independent.
5. Moreover, the set $\{v_1,\ldots,v_n\}$ is linearly dependent *does not imply* that *each* vector is in the span of the remaining vectors.

Given a linearly dependent list of five vectors, suppose, we have discovered that the second vector is a linear combination of first, third, and fourth, where the coefficient of the fourth vector is nonzero. Then the fourth vector can also be expressed as a linear combination of the first, second and third.

***Theorem* 1.14**

A list of vectors $v_1,v_2,\ldots,v_n$ in a vector space V is linearly dependent iff there exists $k\in\{1,\ldots,n\}$ such that the list $v_1,\ldots,v_{k-1}$ is linearly independent, and $v_k\in\operatorname{span}\{v_1,\ldots,v_{k-1}\}$.

In the notation used, if $k = 1$, then the list of vectors $v_1, \ldots, v_{k-1}$ is taken as the empty list $\varnothing$.

Proof Write $S_0 := \varnothing$, and for $1 \le j \le n$, define the list S_j as the list of vectors $v_1, \ldots, v_j$. We notice that S_0 is a sublist of S_1, which is a sublist of S_2, and so on. In this increasing list of lists $S_0, S_1, S_2, \ldots, S_n$, the first list S_0 is linearly independent and the last list S_n is linearly dependent. Therefore, somewhere the switching from linearly independent to linearly dependent happens. That is, there exists a $k \in \{1, \ldots, n\}$ such that all of $S_0, S_1, \ldots, S_{k-1}$ are linearly independent and S_k is linearly dependent. Then the list $v_1, \ldots, v_{k-1}$ is linearly independent and $v_k \in \operatorname{span}\{v_1, \ldots, v_{k-1}\}$. ∎

In the proof of Theorem 1.14, the number k is the least number i such that $\{v_1, \ldots, v_i\}$ is linearly dependent. In fact, the proof is only a detailed explanation of this idea.

Theorem 1.14 says that if a list of vectors is linearly dependent, then either the first vector is the zero vector, or there exists a vector in the list which is a linear combination of the previous ones. Notice that considering a linearly dependent set as an ordered set (a list) has this bonus. In a linearly dependent list, a vector which depends linearly on the previous ones need not be unique. Further, such a vector can change if we choose a different ordering on the given set.

Using this simple result, we discover a little secret that is shared by spanning sets and linearly independent sets.

***Theorem* 1.15**
Let V be a vector space. Let $A = \{u_1, \ldots, u_m\}$ be a linearly independent set and let $B = \{v_1, \ldots, v_n\}$ be a spanning set of V. Then $m \le n$.

Proof Since A is linearly independent, each u_i is a nonzero vector. Assume that $m > n$. Then we have vectors $u_{n+1}, \ldots, u_m$ in A. Since B is a spanning set, $u_1 \in \operatorname{span}(B)$. Thus, the list B_1 of vectors

$$u_1,\ v_1,\ v_2,\ \ldots,\ v_n$$

is linearly dependent. By Theorem 1.14, we have a vector in this list which is in the span of the previous ones. Notice that u_1 is not such a vector since $u_1 \ne 0$. So, let v_k be such a vector. Now, remove this v_k from B_1 to obtain the list C_1 of vectors

$$u_1,\ v_1,\ \ldots,\ v_{k-1},\ v_{k+1},\ \ldots,\ v_n.$$

Notice that $\operatorname{span}(C_1) = \operatorname{span}(B_1) = \operatorname{span}(B) = V$. By including u_2, enlarge the list C_1 to B_2 of vectors

$$u_2,\ u_1,\ v_1,\ \ldots,\ v_{k-1},\ v_{k+1},\ \ldots,\ v_n.$$

Again, B_2 is linearly dependent. Then one of the vectors in B_2 is a linear combination of the previous ones. Such a vector is neither u_1 nor u_2, since $\{u_1, \ldots, u_n\}$ is linearly independent. Thus, let v_j be such a vector. Remove this v_j from B_2 to obtain a list C_2. Now, $\operatorname{span}(C_2) = \operatorname{span}(B_2) = \operatorname{span}(B) = V$.

Continue this process of introducing a u and removing a v for n steps. Finally, v_n is removed and we end up with the list C_n of vectors

$$u_n,\ u_{n-1},\ \ldots,\ u_2,\ u_1,$$

which spans V. Then $u_{n+1} \in \operatorname{span}(C_n)$. This is a contradiction since A is linearly independent. Therefore, our assumption that $m > n$ is wrong. ∎

Exercises for § 1.5

1. In each of the following, a vector space V and a subset S of V are given. Determine whether S is linearly dependent; and if it is, express one of the vectors in S as a linear combination of the remaining vectors.
   (a) $V = \mathbb{R}^3,\ S = \{(1,0,-1),(2,5,1),(0,-4,3)\}$.
   (b) $V = \mathbb{R}^3,\ S = \{(1,2,3),(4,5,6),(7,8,9)\}$.
   (c) $V = \mathbb{C}^3,\ S = \{(1,-3,-2),(-3,1,3),(2,5,7)\}$.
   (d) $V = \mathbb{C}^3,\ S = \{(1,3,2),(3,1,3),(1,2,3),(4,7,5)\}$.
   (e) $V = \mathbb{F}_3[t],\ S = \{t^2 - 3t + 5,\ t^3 + 2t^2 - t + 1,\ t^3 + 3t^2 - 1\}$.
   (f) $V = \mathbb{F}_3[t], S = \{-2t^3 - t^2 + 3t + 2, t^3 - 2t^2 + 3t + 1, t^3 + t^2 + 3t\}$.
   (g) $V = \mathbb{F}_3[t], S = \{6t^3 - 3t^2 + t + 2, t^3 - t^2 + 2t + 3, 2t^3 + t^2 - 3t + 1\}$.
   (h) $V =$ the set of all functions from $\mathbb{R}$ to $\mathbb{R}$, $S = \{2,\ \sin^2 t,\ \cos^2 t\}$.
   (i) $V =$ the set of all functions from $\mathbb{R}$ to $\mathbb{R}$, $S = \{1,\ \sin t,\ \sin 2t\}$.
   (j) $V =$ the set of all functions from $\mathbb{R}$ to $\mathbb{R}$, $S = \{\sin t,\ \cos t,\ \sin 2t\}$.
2. Is $\{(1,0),(1,1),(2,2)\}$ linearly dependent? Is $(1,0) \notin \operatorname{span}\{(1,1),(2,2)\}$?
3. Let S be a subset of a vector space V. Suppose some $v \in S$ is not a linear combination of other vectors in S. Is S linearly independent?
4. Give three linearly dependent vectors in $\mathbb{R}^2$ such that none is a scalar multiple of another.
5. Show that in $\mathbb{R}^2$, $\{(a,b),(c,d)\}$ is linearly independent iff $ad - bc \neq 0$.
6. Prove the following:
   (a) Each subset of a linearly independent set is linearly independent.
   (b) Each superset of a linearly dependent set is linearly dependent.
   (c) The union of two linearly dependent sets is linearly dependent.
   (d) The intersection of two linearly independent sets is linearly independent.
7. Construct examples to show that the following statements are false:
   (a) Each subset of a linearly dependent set is linearly dependent.

(b) Each superset of a linearly independent set is linearly independent.
(c) The union of any two linearly independent sets is linearly independent.
(d) The intersection of any two linearly dependent sets is linearly dependent.

8. Let A and B be subsets of a vector space. Prove or disprove:
   (a) If $\operatorname{span}(A) \cap \operatorname{span}(B) = \{0\}$, then $A \cup B$ is linearly independent.
   (b) If $A \cup B$ is linearly independent, then $\operatorname{span}(A) \cap \operatorname{span}(B) = \{0\}$.
9. Show that in $\mathbb{R}^2$, any three vectors are linearly dependent.
10. Show that any four polynomials in $\mathbb{F}_2[t]$ are linearly dependent.
11. Show that in the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$, the set of functions $\{e^t,\ te^t,\ t^3e^t\}$ is linearly independent.
12. Is $\{\sin t,\ \sin 2t,\ \sin 3t,\ \ldots,\ \sin nt\}$ linearly independent in $C[-\pi,\ \pi]$?
13. Let $p_1(t),\ldots,p_r(t)$ be polynomials with coefficients from $\mathbb{F}$ such that $\deg p_1 < \deg p_2 < \cdots < \deg p_r$. Show that $\{p_1(t),\ldots,p_r(t)\}$ is linearly independent in $\mathbb{F}[t]$.

1.6 Basis

A spanning set of a vector space may contain a vector which is in the span of the other vectors in the set. Throwing away such a vector leaves a spanning set, again. That is, in a spanning set, there may be redundancy. On the other hand, a linearly independent set may fail to span the vector space. That is, in a linearly independent set there may be deficiency. We would like to have a set of vectors which is neither redundant nor deficient.

Let V be a vector space over $\mathbb{F}$. A linearly independent subset that spans V is called a **basis** of V. A basis depends on the underlying field since linear combinations depend on the field.

A vector space may have many bases. For example, both $\{1\}$ and $\{2\}$ are bases of $\mathbb{R}$. In fact $\{x\}$, for any nonzero $x \in \mathbb{R}$, is a basis of $\mathbb{R}$. However, $\{0\}$ has the unique basis $\varnothing$.

Example 1.16

1. Recall that in $\mathbb{R}^2$, the vectors $e_1 = (1,0)$ and $e_2 = (0,1)$ are linearly independent. They also span $\mathbb{R}^2$ since $(a,b) = ae_1 + be_2$ for any $a,b \in \mathbb{R}$. Therefore, $\{e_1,e_2\}$ is a basis of $\mathbb{R}^2$.

2. The set $\{(1,1),(1,2)\}$ is a basis of $\mathbb{R}^2$. Reason? Since neither of them is a scalar multiple of the other, the set is linearly independent.

   To see that the given set of vectors span $\mathbb{R}^2$, we must show that each vector $(a,b) \in \mathbb{R}^2$ can be expressed as a linear combination of these vectors. So, we ask whether the equation $(a,b) = \alpha(1,1) + \beta(1,2)$ has a solution for α and β. The requirement amounts to $\alpha + \beta = a$ and $\alpha + 2\beta = b$. We see that $\beta = b - a$ and $\alpha = 2a - b$ do the job. Hence $\{(1,1),(1,2)\}$ is a basis of $\mathbb{R}^2$.
3. Let $V = \{(a,b) \in \mathbb{R}^2 : 2a - b = 0\}$. Clearly, $(1,2) \in V$. If $(a,b) \in V$, then $b = 2a$. That is, $(a,b) = (a,2a) = a(1,2)$. So, $V = \operatorname{span}\{(1,2)\}$. Also, $\{(1,2)\}$ is linearly independent. Therefore, it is a basis of V.
4. The set $\{(1,0,0),(0,1,0),(0,0,1)\}$ as a subset of $\mathbb{R}^3$ is a basis of $\mathbb{R}^3$. Also, the same set as a subset of $\mathbb{C}^3$ is a basis of $\mathbb{C}^3$.
5. Let $V = \{(a,b,c) \in \mathbb{R}^3 : a - 2b + c = 0\}$. Let $(a,b,c) \in V$. Then $a = 2b - c$; that is, $(a,b,c) = (2b - c, b, c) = b(2,1,0) + c(-1,0,1)$. So, V is spanned by $\{(2,1,0),(-1,0,1)\}$. Further, it is a linearly independent subset of V. Therefore, it is a basis of V. □

Example 1.17
Recall that e_j in $\mathbb{F}^n$ has the jth component as 1 and all other components as 0. The set $\{e_1,\ldots,e_n\}$ is a basis of $\mathbb{F}^n$. As an ordered set, this basis is called the **Standard Basis** of $\mathbb{F}^n$. □

We also write the ordered set $\{e_1,\ldots,e_n\}$ for the standard basis of $\mathbb{F}^{n\times 1}$, where each e_j is taken as a column vector. Similarly, we write the standard basis of $\mathbb{F}^{1\times n}$ as $\{e_1,\ldots,e_n\}$, where each e_i is taken as a row vector. The context will clarify whether it is a column vector or a row vector, and the particular n.

We now show formally that a basis neither has redundancy nor has deficiency in spanning the vector space. We use the following terminology.

A subset B of a vector space V is a *maximal* linearly independent set means that B is linearly independent in V, and each proper superset of B is linearly dependent.

Similarly, a subset B of V is a *minimal* spanning set of V means that B spans V, and each proper subset of B fails to span V.

***Theorem* 1.18**
Any subset of a vector space is a basis iff it is a minimal spanning set iff it is a maximal linearly independent set.

Proof Let B be a subset of a vector space V.

Suppose that B is a basis of V. Then B is a spanning set of V. If B is not a minimal spanning set of V, then there exists a vector $v \in B$ such that $B \setminus \{v\}$ is also a spanning set. In particular, $v \in \operatorname{span}(B \setminus \{v\})$. This contradicts the assumption that B is linearly independent. Therefore, each basis of V is a minimal spanning set of V.

Let B be a minimal spanning set. If B is linearly dependent, then we have a vector $v \in B$ such that $v \in \operatorname{span}\{B \setminus \{v\})$. Then $\operatorname{span}(B \setminus \{v\}) = \operatorname{span}(B) = V$. This contradicts minimality of B. Hence, B is linearly independent. If B is not maximal linearly independent, then there exists a vector w such that $B \cup \{w\}$ is linearly independent. In that case, $w \notin \operatorname{span}(B)$. This contradicts the assumption that B is a spanning set. Therefore, each minimal spanning set of V is a maximal linearly independent set of V.

Assume that B is a maximal linearly independent set of V. If B does not span V, then there exists $x \in V$ such that $x \notin \operatorname{span}(B)$. In that case, $B \cup \{x\}$ is linearly independent. This contradicts the maximality of B. That is, B spans V; and hence B is a basis of V. Therefore, each maximal linearly independent set of V is a basis of V. ∎

Any vector $(a, b) \in \mathbb{R}^2$ can be written uniquely as $a(1,0) + b(0,1)$. Also, (a, b) can be written as a linear combination of the vectors $(1,1)$, $(-1,2)$, and $(1,0)$; but not uniquely. For example,

$$(2,3) = 3(1,1) + 0(-1,2) + -1(1,0) = 1(1,1) + 1(-1,2) + 2(1,0).$$

Here, both the sets $\{(1,0),(0,1)\}$ and $\{(1,1),(-1,2),(1,0)\}$ are spanning sets of $\mathbb{R}^2$. We see that neither $(1,0)$ is in the span of $(0,1)$, nor $(0,1)$ is in the span of $(1,0)$. Considering the second set, we find that $(1,0)$ is in the span of $\{(1,1),(-1,2)\}$. Reason?

$$(1,0) = 2(1,1) + 1(-1,2).$$

In this case, uniqueness of writing a vector as a linear combination breaks down because one of the vectors in the set is a linear combination of others. This is another reason why we are interested in a basis.

To express the intended idea formally, we require an ordering of the vectors in a basis. Recall that when we consider the set $\{v_1, \ldots, v_n\}$ as an ordered set, we mean that v_1 is the first vector, v_2 is the second vector, etc, and finally, v_n is the nth vector in the ordered set.

***Theorem* 1.19**
Let $v_1, \ldots, v_n$ be vectors in a vector space V. The ordered set $B = \{v_1, \ldots, v_n\}$ is a basis of V iff for each $v \in V$ there exists a unique n-tuple of scalars $(\alpha_1, \ldots, \alpha_n)$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$.

Proof Suppose the ordered set B is a basis of V. Let $v \in V$. As $\operatorname{span}(B) = V$, there exist scalars $\alpha_1, \ldots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. For uniqueness, suppose

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n.$$

Then $(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n = 0$. Due to linear independence of B,

$$\alpha_1 = \beta_1, \ \ldots, \ \alpha_n = \beta_n.$$

This proves uniqueness of the n-tuple $(\alpha_1, \ldots, \alpha_n)$.

Conversely, suppose for each $v \in V$, there exists a unique n-tuple of scalars $(\alpha_1, \ldots, \alpha_n)$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then $V \subseteq \operatorname{span}(B) \subseteq V$. That is, B spans V. Further, let $\beta_1 v_1 + \cdots + \beta_n v_n = 0$. Now, $0\,v_1 + \cdots + 0\,v_n = 0$. By uniqueness of the n-tuple of scalars in the linear combination of the zero vector, we have $\beta_1 = \cdots = \beta_n = 0$. That is, B is linearly independent. Hence, B is a basis of V. ∎

A basis can be extracted from a finite spanning set.

***Theorem* 1.20**
If a vector space has a finite spanning set, then every such finite spanning set contains a basis.

Proof Let $S = \{v_1, \ldots, v_n\}$ be a spanning set of a vector space V. Consider S as an ordered set. If S is linearly independent, then S is a basis. Otherwise, due to Theorem 1.14, there exists a vector v_k which is in the span of $v_1, \ldots, v_{k-1}$. Delete v_k from S and apply the same check again. Repeatedly throwing away those vectors from S which are linearly depending on the previous ones we end up at a basis. ∎

Example 1.21
Let $B = \{(1,0,1), (1,2,1), (2,2,2), (0,2,0)\}$ and let $V = \operatorname{span}(B)$. We see that $(1,2,1)$ is not a scalar multiple of $(1.0,1)$. Next, $(2,2,2) = (1,0,1) + (1,2,1)$. Removing the vector $(2,2,2)$ from B, we obtain

$$B_1 = \{(1,0,1), (1,2,1), (0,2,0)\}.$$

Here, $V = \operatorname{span}(B) = \operatorname{span}(B_1)$. Next, $(0,2,0) = -(1,0,1) + (1,2,1)$. Removing $(0,2,0)$ from B_1, we end up with

$$B_2 = \{(1,0,1), (1,2,1)\}.$$

Notice that $V = \operatorname{span}(B) = \operatorname{span}(B_1) = \operatorname{span}(B_2)$, and B_2 is linearly independent. Thus B_2 is a basis of V. □

Similarly, by enlarging a linearly independent set we may also end up with a spanning set, keeping linear independence preserved.

***Theorem* 1.22**
If a vector space has a finite spanning set, then every linearly independent set can be extended to a basis.

Proof Let $S = \{u_1, \ldots, u_n\}$ be a spanning set of a vector space V. Let $B = \{v_1, \ldots, v_m\}$ be a linearly independent subset of V. If B also spans V, then it is a basis of V. Otherwise, construct the ordered set $C = \{v_1, \ldots, v_m, u_1, \ldots, u_n\}$. Now, C is a linearly dependent spanning set. Since B is linearly independent, by Theorem 1.14, some u_i is a linear combination of the earlier vectors in C. Throw away all such vectors one-by-one. The remaining set so constructed from C is a basis of V; and it is an extension of B. ∎

Example 1.23
Let $B = \{(1,0,1), (1,2,1), (2,2,2), (0,2,0)\}$ and let $V = \operatorname{span}(B)$, as in the last example. The vector $(2,-2,2) \in V$, since $(2,-2,2) = 3(1,0,1) - (1,2,1)$.

For extending the set $\{(2,-2,2)\}$ to a basis, we construct the spanning set

$$C = \{(2,-2,2), (1,0,1), (1,2,1), (2,2,2), (0,2,0)\}.$$

Notice that C is a spanning set of V since its subset B is a spanning set. Further, we see that $\{(2,-2,2), (1,0,1)\}$ is linearly independent. And,

$$\begin{aligned}
(1,2,1) &= (-1)(2,-2,2) + 3(1,0,1), \\
(2,2,2) &= (-1)(2,-2,2) + 4(1,0,1), \\
(0,2,0) &= (-1)(2,-2,2) + 2(1,0,1).
\end{aligned}$$

Deleting these vectors from C, we end up with the basis $\{(2,-2,2), (1,0,1)\}$ of V, which is an extension of $\{(2,-2,2)\}$. □

Exercises for § 1.6

1. Determine which of the following sets form bases for $\mathbb{F}_2[t]$?
   (a) $\{1,\ 1+t,\ 1+t+t^2\}$.
   (b) $\{1+2t+t^2,\ 3+t^2,\ t+t^2\}$.
   (c) $\{1+2t+3t^2,\ 4-5t+6t^2,\ 3t+t^2\}$.
   (d) $\{-1-t-2t^2,\ 2+t-2t^2,\ 1-2t+4t^2\}$.
2. Is $\{1+t^n,\ t+t^n,\ \ldots,\ t^{n-1}+t^n,\ t^n\}$ a basis of $\mathbb{F}_n[t]$?
3. Let $\{x,\ y,\ z\}$ be a basis of a vector space V. Are the following sets also bases of V?
   (a) $\{x+y,\ y+z,\ z+x\}$ (b) $\{x-y,\ y-z,\ z-x\}$
4. Find a basis for the subspace $\{(a,b,c) \in \mathbb{R}^3 : a+b+c = 0\}$ of $\mathbb{R}^3$.
5. Find a basis of $V = \{(a_1, \ldots, a_5) \in \mathbb{C}^5 : a_1 + a_3 - a_5 = 0 = a_2 - a_4 = 0\}$.
6. Let $V = \{p(t) \in \mathbb{R}_2[t] : p(0) + 2p'(0) + 3p''(0) = 0\}$. Show that V is a vector space and find a basis for it.
7. Extend the set $\{1+t^2,\ 1-t^2\}$ to a basis of $\mathbb{F}_3[t]$.
8. Let $u_1 = 1$ and let $u_j = 1 + t + t^2 + \cdots + t^{j-1}$ for $j = 2,3,4,\ldots$ Is $\{u_1, \ldots, u_n\}$ a basis of $\mathbb{F}_n[t]$? Is $\{u_1, u_2, \ldots\}$ a basis of $\mathbb{F}[t]$?
9. Construct three bases for $\mathbb{R}^3$ so that no two of them have a common vector.

1.7 Dimension

Theorem 1.20 implies that if a vector space has a finite spanning set, then all its bases are finite. We show something more.

***Theorem* 1.24**
If a vector space has a finite spanning set, then each basis has the same number of elements.

Proof Let V be a vector space having a finite spanning set. Let B and E be two bases of V having m and n number of vectors. Consider B as a linearly independent set and E as a spanning set. Then $m \le n$, due to Theorem 1.15. Now, consider E as a linearly independent set and B as a spanning set. Then $n \le m$. Therefore, $m = n$. ∎

We give a name to this number, the number of vectors in a basis.

Let V be a vector space having a finite spanning set. Then the number of elements in a basis is called the **dimension** of V; and it is denoted by $\dim(V)$.

Example 1.25

1. $\dim(\mathbb{R}) = 1$; $\dim(\mathbb{R}^n) = n$; $\dim(\mathbb{C}) = 1$; $\dim(\mathbb{C}^n) = n$; $\dim(\mathbb{F}_n[t]) = n+1$.
2. The dimension of the zero space $\{0\}$ is 0. Reason? $\varnothing$ is a basis of $\{0\}$.
3. If $\mathbb{C}$ is considered as a vector space over $\mathbb{R}$, then $\dim(\mathbb{C}) = 2$. For instance, $\{1, i\}$ is a basis of the real vector space $\mathbb{C}$.
4. The real vector space $\mathbb{C}^n$ has dimension $2n$. Can you construct a basis of $\mathbb{C}^n$ considered as a vector space over $\mathbb{R}$?
5. The real vector space $\mathbb{C}_n[t]$ has dimension $2(n+1)$.

□

A vector space which has a finite basis is called a **finite dimensional** vector space. We also write $\dim(V) < \infty$ to express the fact that "V is finite dimensional". Due to Theorem 1.20, each vector space having a finite spanning set is finite dimensional. The dimension of a finite dimensional vector space is a non-negative integer.

A vector space which does not have a finite basis, is called an **infinite dimensional** vector space; and we sometimes express this fact by writing $\dim(V) = \infty$. Thus, a vector space is infinite dimensional iff none of its finite subsets is its basis iff no finite subset of it is a spanning set.

For instance, $\mathbb{F}^n$ and $\mathbb{F}_n[t]$ are finite dimensional vector spaces over $\mathbb{F}$, whereas $\mathbb{F}^\infty$ is infinite dimensional.

If a vector space contains an infinite linearly independent subset, then due to Theorem 1.15, no finite subset of it can be a spanning set. Thus, such a vector space is necessarily infinite dimensional.

Example 1.26

1. The set of all polynomials, $\mathbb{F}[t]$, is an infinite dimensional vector space. Reason? Suppose the dimension is finite, say $\dim(\mathbb{F}[t]) = n$. Then any set of $n+1$ vectors is linearly dependent. But $\{1, t, \ldots, t^n\}$ is linearly independent! Notice that $\{1, t, t^2, \ldots\}$ is a basis of $\mathbb{F}_n[t]$.
2. $C[a,\ b]$ is an infinite dimensional vector space. Reason? Take the collection of functions $\{f_n : f_n(t) = t^n \text{ for all } t \in [a,\ b];\ n = 0, 1, 2, \ldots\}$. Then $\{f_0, f_1, \ldots, f_n\}$ is linearly independent for every n. So, $C[a,\ b]$ cannot have a finite basis.

$\square$

We will not study infinite dimensional vector spaces, though occasionally, we will give an example to illustrate a point.

Inter-dependence of the notions of spanning set, linear independence, and basis can be seen using the notion of dimension. The following theorems (Theorems 1.27-1.28) state some relevant facts; their proofs are easy. For a finite set S, we write $|S|$ for the number of elements in S.

***Theorem* 1.27**
Let S be a finite subset of a finite dimensional vector space V.

(1) *S is a basis of V iff S is a spanning set and* $|S| = \dim(V)$.

(2) *S is a basis of V iff S is linearly independent and* $|S| = \dim(V)$.

(3) *If* $|S| < \dim(V)$, *then S does not span V.*

(4) *If* $|S| > \dim(V)$, *then S is linearly dependent.*

***Theorem* 1.28**
Let U be a subspace of a finite dimensional vector space V.

(1) *U is a proper subspace of V iff* $\dim(U) < \dim(V)$.

(2) **(Basis Extension)** *Each basis of U can be extended to a basis of V.*

Given two subspaces U and W of a finite dimensional vector space, we have two other subspaces, $\operatorname{span}(U \cap W)$ and $U + W$. What about their dimensions?

Example 1.29

1. Let $U = \{(a,b) \in \mathbb{R}^2 : 2a - b = 0\}$ and $W = \{(a,b) \in \mathbb{R}^2 : a + b = 0\}$. For these subspaces of $\mathbb{R}^2$, we see that

$$U \cap W = \{(a,b) : 2a - b = 0 = a + b\} = \{0\}.$$

$$U + W = \{(a, 2a) + (c, -c) : a, c \in \mathbb{R}\} = \{(a + c, 2a - c) : a, c \in \mathbb{R}\}.$$

Further, if $\alpha, \beta \in \mathbb{R}$, then

$$(\alpha,\, \beta) = \Big(\frac{\alpha+\beta}{3} + \frac{2\alpha-\beta}{3},\ 2\frac{\alpha+\beta}{3} - \frac{2\alpha-\beta}{3}\Big).$$

That is, each vector in $\mathbb{R}^2$ can be expressed in the form $(a+c, 2a-c)$. Thus $U+W=\mathbb{R}^2$.

Here, $\dim(U\cap W)+\dim(U+W) = 2 = \dim(U)+\dim(W)$.

2. Consider the following subspaces of $\mathbb{R}^3$:

$$U = \{(a,b,c) \in \mathbb{R}^3 : a+b+c=0\};\ W = \{(a,b,c) \in \mathbb{R}^3 : a-b-c=0\}.$$

We determine (some) bases for U, W, $U\cap W$ and $U+W$.

It is easy to see that a basis for U is $\{(1,0,-1),(1,-1,0)\}$; and a basis for W is $\{(1,1,0),(1,0,1)\}$.

Next, solving $a+b+c=0$ and $a-b-c=0$, we see that $a=0$ and $c=-b$. Thus $U\cap W = \{(0,b,-b) : b\in\mathbb{R}\}$ has a basis $\{(0,1,-1).\}$

Next, $U+W = \{(a,b,-a-b)+(\beta+\gamma,\beta,\gamma) : a,b,\beta,\gamma\in\mathbb{R}\}$. Any vector $(c_1,c_2,c_3)\in\mathbb{R}^3$ can be expressed as $(a,b,-a-b)+(\beta+\gamma,\beta,\gamma)$ with

$$a=0,\ b=\frac{1}{2}(-c_1+c_2-c_3),\ \beta=\frac{1}{2}(c_1-c_2+c_3),\ \gamma=\frac{1}{2}(c_1-c_2-c_3).$$

That is, $U+W=\mathbb{R}^3$.

Thus $\dim(U\cap W)+\dim(U+W) = 1+3 = \dim(U)+\dim(W)$. ∎

Since $U\cap W$ is a subspace of both U and W, and each of U and W is a subspace of $U+W$, we obtain

$$\dim(U\cap W) \ \le\ \dim(U), \dim(W) \ \le\ \dim(U+W).$$

However, Example 1.29 hints at something more.

***Theorem* 1.30**
Let U and W be finite dimensional subspaces of a vector space V. Then

$$\dim(U\cap W)+\dim(U+W) = \dim(U)+\dim(W).$$

Proof Since $U\cap W$ is a subspace of U, $\dim(U\cap W)<\infty$. Let $B=\{u_1,\ldots,u_n\}$ be a basis of $U\cap W$. (Here, if $U\cap W=\{0\}$, we take $B=\varnothing$; and thus $n=0$.) By the basis extension theorem, there are bases $B\cup C$ and $B\cup D$ for U and W, respectively, where $C=\{v_1,\ldots,v_k\}$ and $D=\{w_1,\ldots,w_m\}$, for some $k,m\ge 0$. Notice that no v is in $U\cap W$, and no w is in $U\cap W$. Also, no v is in W.

Reason? If $v_j \in W$, then $v_j \in U \cap W$, which is wrong. Similarly, no w is in U. Therefore, $B \cup C$ has $n+k$ vectors, and $B \cup D$ has $n+m$ vectors. Also, the set

$$E = B \cup C \cup D = \{u_1, \ldots, u_n, v_1, \ldots, v_k, w_1, \ldots, w_m\}$$

has exactly $n+k+m$ vectors. We show that E is a basis of $U+W$.

Let $z \in U+W$. Then there exist $x \in U$ and $y \in W$ such that $x+y \in U+W$. Further, there exist scalars $\alpha_1, \ldots, \alpha_{n+k},\ \beta_1, \ldots, \beta_{n+m}$ such that

$$\begin{aligned} x &= \alpha_1 u_1 + \cdots + \alpha_n u_n + \alpha_{n+1} v_1 + \cdots + \alpha_{n+k} v_k, \\ y &= \beta_1 u_1 + \cdots + \beta_n u_n + \beta_{n+1} w_1 + \cdots + \beta_{n+m} w_m. \end{aligned}$$

Then

$$z = x + y = \sum_{i=1}^{n} (\alpha_i + \beta_i) u_i + \sum_{j=1}^{k} \alpha_{n+j} v_j + \sum_{\ell=1}^{m} \beta_{n+l} w_\ell \in \operatorname{span}(E).$$

So, $U+W \subseteq \operatorname{span}(E)$. As $E \subseteq U+W$, we have $\operatorname{span}(E) \subseteq U+W$. Consequently, $\operatorname{span}(E) = U+W$.

To prove the linear independence of E, let

$$\alpha_1 u_1 + \cdots + \alpha_n u_n + \beta_1 v_1 + \cdots + \beta_k v_k + \gamma_1 w_1 + \cdots + \gamma_m w_m = 0.$$

Then

$$\alpha_1 u_1 + \cdots + \alpha_n u_n + \beta_1 v_1 + \cdots + \beta_k v_k = -\gamma_1 w_1 - \cdots - \gamma_m w_m.$$

The left hand side is a vector in U, and the right hand side is a vector in W. Therefore, both are in $U \cap W$. Since B is a basis for $U \cap W$, we have

$$-\gamma_1 w_1 - \cdots - \gamma_m w_m = a_1 u_1 + \cdots + a_n u_n$$

for some scalars $a_1, \ldots, a_n$. Thus

$$a_1 u_1 + \cdots + a_n u_n + \gamma_1 w_1 + \cdots + \gamma_m w_m = 0.$$

Since $B \cup D$ is linearly independent, $a_1 = \cdots = a_n = \gamma_1 = \cdots = \gamma_k = 0$. Substituting the values of γ_i's, we get

$$\alpha_1 u_1 + \cdots + \alpha_n u_n + \beta_1 v_1 + \cdots + \beta_k v_k = 0.$$

Since $B \cup C$ is linearly independent, $\alpha_1 = \cdots = \alpha_n = \beta_1 = \cdots = \beta_k = 0$. That is,

$$\alpha_1 = \cdots = \alpha_n = \beta_1 = \cdots = \beta_k = \gamma_1 = \cdots \gamma_k = 0.$$

Hence E is linearly independent.

Now that E is a basis of $U+W$, we have

$$\dim(U \cap W) + \dim(U+W) = n + n + m + k = \dim(U) + \dim(W). \qquad \blacksquare$$

It thus follows that two distinct planes through the origin in $\mathbb{R}^3$ intersect on a straight line.

Exercises for § 1.7

1. Describe all subspaces of $\mathbb{R}^3$.
2. Find bases and dimensions of the following subspaces of $\mathbb{R}^5$:
   (a) $\{(a_1, a_2, a_3, a_4, a_5) \in \mathbb{R}^5 : a_1 - a_3 - a_4 = 0\}$.
   (b) $\{(a_1, a_2, a_3, a_4, a_5) \in \mathbb{R}^5 : a_2 = a_3 = a_4,\ a_1 + a_5 = 0\}$.
   (c) Span of the set of vectors $(1, -1, 0, 2, 1), (2, 1, -2, 0, 0), (0, -3, 2, 4, 2), (2, 4, 1, 0, 1), (3, 3, -4, -2, -1)$ and $(5, 7, -3, -2, 0)$.
3. Find $\dim(\text{span}\{1 + t^2, -1 + t + t^2, -6 + 3t, 1 + t^2 + t^3, t^3\})$ in $\mathbb{F}_3[t]$.
4. Let U and W be subspaces of a finite dimensional vector space. Prove that if $\dim(U \cap W) = \dim(U)$, then $U \subseteq W$.
5. Show that if U and W are subspace of $\mathbb{R}^9$ with $\dim(U) = 5 = \dim(W)$, then $U \cap W \neq \{0\}$.
6. Let $U = \text{span}\{(1, 2, 3), (2, 1, 1)\}$ and $W = \text{span}\{(1, 0, 1), (3, 0, -1)\}$. Find a basis for $U \cap W$. Also, find $\dim(U + W)$.
7. Compute $\dim(U)$, $\dim(V)$, $\dim(U + V)$ and $\dim(U \cap V)$, where

$$\begin{aligned} U &= \{(a_1, \ldots, a_{50}) \in \mathbb{R}^{50} : a_i = 0 \text{ when 3 divides } i\} \\ V &= \{(a_1, \ldots, a_{50}) \in \mathbb{R}^{50} : a_i = 0 \text{ when 4 divides } i\}. \end{aligned}$$

8. Let V be the vector space of all functions from $\{1, 2, 3\}$ to $\mathbb{R}$. Consider each polynomial in $\mathbb{R}[t]$ as a function from $\{1, 2, 3\}$ to $\mathbb{R}$. Is the set of vectors $\{t, t^2, t^3, t^4, t^5\}$ linearly independent in V?
9. Let S be a set consisting of n elements. Let V be the set of all functions from S to $\mathbb{R}$. Show that V is a vector space of dimension n.
10. Prove Theorems 1.27-1.28.
11. Let V be the set of all functions $f(t)$ having a power series expansion for $|t| < 1$. Show that V is an infinite dimensional vector space.

1.8 Extracting a basis

Have you solved Exercise 2(c) of the last section? If not, try it now. A question of computational importance is that given a finite list of vectors, how do we construct a basis for the subspace spanned by these vectors?

Given a list of vectors $v_1, \ldots, v_n$ in a vector space V, we need to systematically eliminate those vectors which are linear combinations of the previous ones in the list, using Theorem 1.14. So that the span of the shortened list will be the same as that of the given list. For example, if $v_i = 0$, then v_i is a linear combination of others; also, $v_i \in \operatorname{span}(\varnothing)$. That is, a zero vector can safely be dropped from the list.

The notion of elementary operations help in achieving this, at least in $\mathbb{F}^n$. For this purpose we define an elementary operation. Let $v_1, \ldots, v_m$ be vectors in a vector space V. Let $1 \le i \ne j \le m$. The **elementary operation** $E_\alpha[i, j]$ applied on the list of vectors $v_1, \ldots, v_m$ produces a list of vectors $w_1, \ldots, w_m$ where $w_i = v_i + \alpha v_j$ and $w_k = v_k$ for $k \ne i$. It adds to the ith vector in the list α times the jth vector; keeping other vectors as they are.

Example 1.31
Consider the list of vectors v_1, v_2, v_3, v_4 in $\mathbb{R}^4$, where

$$v_1 = (1, 1, 0, -1),\ v_2 = (2, -1, 1, 0),\ v_3 = (1, 2, -2, 1),\ v_4 = (2, -2, 3, -2).$$

An elementary operation $E_2[1, 3]$ applied on v_1, v_2, v_3, v_4 produces the list $v_1 + 2v_3, v_2, v_3, v_4$. Here, $w_1 = v_1 + 2v_3$, $w_2 = v_2$, $w_3 = v_3$, and $w_4 = v_4$. That is, $w_1 = (5, -1, 2, -1),\ w_2 = (2, -1, 1, 0),\ w_3 = (1, 2, -2, 1),\ w_4 = (2, -2, 3, -2)$. $\square$

The elementary operation $E_\alpha[i, j]$ applied on the list $v_1, \ldots, v_m$ changes the ith vector v_i only. The new vector is $w_i = v_i + \alpha v_j$. Then $v_i = w_i - \alpha v_j$. It shows that

$$w_k \in \operatorname{span}\{v_1, \ldots, v_m\} \text{ and } v_k \in \operatorname{span}\{w_1, \ldots, w_m\} \text{ for each } k \in \{1, \ldots, m\}.$$

Thus, an elementary operation applied on a list of vectors does not change the span. Using this, we can have a strategy of extracting a basis for the span of given vectors $v_1, \ldots, v_m$. Our aim is to apply elementary operations repeatedly, and bring in as many zero entries as possible in the vectors so that it will be easier to see linear independence.

We write all vectors one below the other, as a rectangular array; and then look for a nonzero entry in the first column. If found, we fix that row and then use elementary operations to zero-out entries in that column and other rows. Then proceed with the next column, and so on. The following example illustrates this strategy.

Example 1.32
Using elementary operations, extract a basis for

$$U = \operatorname{span}\{(0, 0, 0, 0), (0, 1, 2, 3), (1, 0, 2, 4), (-1, 1, 0, -1), (3, 2, 1, -1)\}.$$

We write the vectors one below another, thus forming a rectangular array of 5 rows and 4 columns; and then apply suitable elementary operations.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 4 \\ -1 & 1 & 0 & -1 \\ 3 & 2 & 1 & -1 \end{bmatrix} \xrightarrow{R1} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & -5 & -13 \end{bmatrix} \xrightarrow{R2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -9 & -19 \end{bmatrix}$$

In the first step, we look at the first column for a nonzero entry. The first nonzero entry occurs in the third row. Then we use appropriate elementary operations so that the first entries in all other rows except the third row become 0. That is, we need only to apply elementary operations on fourth row onwards. Here, $R1 = E_1[4,3]$, $E_{-3}[5,3]$.

Next, we look at the entries in the second column. The first nonzero entry occurs at the second row. This row has not been chosen earlier. Using elementary operations we zero out the entries in the second column and third row onwards. Here, $R_2 = E_{-1}[4,2]$, $E_{-2}[5,2]$.

Next, we look at the third column. The first nonzero entry except the third and the second rows, occurs in the fifth row. The first and fourth row entries in the same third column are already zero. So, the process stops here.

The nonzero rows in the resulting matrix are linearly independent, and they give the basis. Here, such a basis for U is given by

$$\{(0,1,2,3),\ (1,0,2,3),\ (0,0,-9,-19)\}.$$

Notice that there is scope for zeroing out the entries in the third and fourth rows of our last matrix by using elementary operations. However, we do not require it.

Our final basis also says that the corresponding vectors in the original list, i.e., v_2, v_3, v_5 form a basis for U. You can then unfold the elementary operations to determine how the vectors in the original list depend linearly on the vectors that correspond to the final basis vectors. For example, $row(4)$ in the final array is a zero vector and it has been obtained by using $E_{-1}[4,2]$. Before it, $row(4)$ has been obtained by using $E_1[4,3]$. Thus,

$$0 = v_4 - v_2 + v_3.$$

Had v_2 been changed, the final equation would have been complicated; however, from the operations, it can always be determined. Notice that we do nothing about v_1 since it is the zero vector. $\square$

Exercises for § 1.8

1. In $\mathbb{R}^3$, what is $\dim(\operatorname{span}\{e_1+e_2, e_2+e_3, e_3+e_1\})$?

2. Find a basis and dimension of the subspace of $\mathbb{R}^5$ that is spanned by the vectors $(1,-1,0,2,1)$, $(2,1,-2,0,0)$, $(0,-3,2,4,2)$, $(3,3,-4,-2,-1)$, $(2,4,1,0,1)$ and $(5,7,-3,-2,0)\}$.
3. In each of the following subspaces U and W of a vector space V, determine the bases and dimensions of U, W, $U+W$ and of $U \cap W$.
   (a) $V = \mathbb{R}^3$, $U = \text{span}\{(1,2,3),\ (2,1,1)\}$, $W = \text{span}\{(1,0,1),\ (3,0,-1)\}$.
   (b) $V = \mathbb{R}^4$, $U = \text{span}\{(1,0,2,0),(1,0,3,0),\}$,
   $W = \text{span}\{(1,0,0,0),(0,1,0,0),(0,0,1,1)\}$.
   (c) $V = \mathbb{C}^4$, $U = \text{span}\{(1,0,0,2),(3,1,0,2),(7,0,5,2)\}$,
   $W = \text{span}\{(1,0,3,2),(10,4,14,8),(1,1,-1,-1)\}$.
4. Let $U = \{(a,b,c,d) \in \mathbb{R}^4 : b = -a\}$ and $W = \{(a,b,c,d) \in \mathbb{R}^4 : c = -a\}$. Find the dimensions of the subspaces U, W, $U+W$ and $U \cap W$ of $\mathbb{R}^4$.
5. Determine $\dim\left(\text{span}\{1+t^2, -1+t+t^2, -6+3t, 1+t^2+t^3, t^3\}\right)$.
6. Let V be a vector space. Let $v_1, \ldots, v_n \in V$. Show the following:
   (a) If $v_1, \ldots, v_n$ span V, then $v_1, v_2 - v_1, \ldots, v_n - v_1$ span V.
   (b) If $v_1, \ldots, v_n$ are linearly independent, then $v_1, v_2 - v_1, \ldots, v_n - v_1$ are linearly independent.

2

Inner Product Spaces

2.1 Inner Products

Vectors in the plane or in the usual three dimensional space are defined as entities having certain lengths and directions. In the last chapter we have only abstracted the notion of a vector via the operations of addition and scalar multiplication. We wish to give directions and lengths to our abstract vectors in a vector space. Notice that directions are relatively fixed by defining angle between vectors. In the plane, both angle and length are defined by the dot product. For instance, writing $\|x\|$ for the length of a vector x, we have

$$\|x\|^2 = x \cdot x, \quad \cos\big(\angle(x, y)\big) = \frac{x \cdot y}{\|x\|\,\|y\|}.$$

We will refer to the dot product as an *inner product*. As usual, we will define this inner product abstractly, as a certain map satisfying some properties. In fact, we take up those fundamental properties of the dot product in connection with the already available operations of addition and scalar multiplication.

An **inner product** on a vector space V over $\mathbb{F}$ is a map from $V \times V$ to $\mathbb{F}$, which associates a pair of vectors $x, y \in V$ to a scalar $\langle x, y\rangle$ in $\mathbb{F}$ satisfying the following properties:

(1) For each $x \in V$, $\langle x, x\rangle \geq 0$.

(2) For each $x \in V$, $\langle x, x\rangle = 0$ iff $x = 0$.

(3) For all $x, y, z \in V$, $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

(4) For each $\alpha \in \mathbb{F}$ and for all $x, y \in V$, $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

(5) For all $x, y \in V$, $\langle y, x\rangle = \overline{\langle x, y\rangle}$.

A vector space with an inner product on it is called an **inner product space**.

Read the abbreviation *ips* as ‘inner product space’. An ips with the underlying field as $\mathbb{R}$ is called a *real ips* or an *Euclidean space*. An ips with $\mathbb{F} = \mathbb{C}$ is called a *complex ips* or a *unitary space*. Notice that in a real ips, the fifth property reads as $\langle y, x\rangle = \langle x, y\rangle$ for all vectors x and y.

A ready made example of an ips is $\mathbb{R}^2$, where the dot product is the inner product. Similarly, $\mathbb{R}^3$ is a Euclidean space with the usual dot product as the inner product.

Example 2.1

1. For $x = (a_1, \ldots, a_n), y = (b_1, \ldots, b_n) \in \mathbb{R}^n$, $\langle x, y\rangle = \sum_{j=1}^n a_j b_j$ defines an inner product. It is called the *standard inner product* on $\mathbb{R}^n$.
2. For $x = (a_1, \ldots, a_n), y = (b_1, \ldots, b_n) \in \mathbb{C}^n$, $\langle x, y\rangle = \sum_{j=1}^n a_j \overline{b_j}$ defines an inner product. It is called the *standard inner product* on $\mathbb{C}^n$. Notice that $\langle x, y\rangle = \sum_{j=1}^n a_j b_j$ is not an inner product on $\mathbb{C}^n$.
3. In $\mathbb{R}_n[t]$, for $p(t) = a_0 + a_1 t + \cdots + a_n t^n,\ q(t) = b_0 + b_1 t + \cdots + b_n t^n$, take $\langle p, q\rangle = \sum_{i=0}^n a_i b_i$. This defines an inner product on $\mathbb{R}_n[t]$.
4. In $\mathbb{C}_n[t]$, for $p(t) = a_0 + a_1 t + \cdots + a_n t^n,\ q(t) = b_0 + b_1 t + \cdots + b_n t^n$, take $\langle p, q\rangle = \sum_{i=0}^n a_i \overline{b}_i$. This defines an inner product on $\mathbb{C}_n[t]$.
5. Let $t_1, t_2, \ldots, t_{n+1}$ be distinct real numbers. For any $p, q \in \mathbb{R}_n[t]$, define $\langle p, q\rangle = \sum_{i=1}^{n+1} p(t_i) q(t_i)$. This is an inner product on $\mathbb{R}_n[t]$.
6. Consider each polynomial $p \in \mathbb{R}_n[t]$ as a function from $[0,1]$ to $\mathbb{R}$. For $p, q \in \mathbb{R}_n[t]$, define $\langle p, q\rangle = \int_0^1 p(t) q(t)\, dt$. This is an inner product on $\mathbb{R}_n[t]$.
7. Let V be a finite dimensional real vector space. Let $B = \{u_1, u_2, \ldots, u_n\}$ be an ordered basis for V. For $x = \sum_{i=1}^n \alpha_i u_i,\ y = \sum_{i=1}^n \beta_i u_i$, define $\langle x, y\rangle_B = \sum_{i=1}^n \alpha_i \beta_i$. This is an inner product on V.
8. Let V be a finite dimensional complex vector space with an ordered basis $B = \{u_1, u_2, \ldots, u_n\}$. For $x = \sum_{i=1}^n \alpha_i u_i$, $y = \sum_{i=1}^n \beta_i u_i$, define $\langle x, y\rangle_B = \sum_{i=1}^n \alpha_i \overline{\beta}_i$. This is an inner product on V.
9. For $f, g \in C[a,b]$, define $\langle f, g\rangle = \int_a^b f(t) g(t)\, dt$. This gives an inner product on $C[a,b]$. $\square$

***Theorem* 2.2**
Let V be an ips. For all $x, y, z \in V$ and for each $\alpha \in \mathbb{F}$,

$$\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle \text{ and } \langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle.$$

Proof $\langle x, y+z\rangle = \overline{\langle y+z, x\rangle} = \overline{\langle y, x\rangle + \langle z, x\rangle} = \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} = \langle x, y\rangle + \langle x, z\rangle$.

$$\langle x, \alpha y\rangle = \overline{\langle \alpha y, x\rangle} = \overline{\alpha \langle y, x\rangle} = \overline{\alpha}\, \overline{\langle y, x\rangle} = \overline{\alpha}\, \langle x, y\rangle.$$ ∎

Let V be an ips. For any $x \in V$, the length of x, also called the **norm** of x is defined as

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Notice that $\langle x, x\rangle \geq 0$ implies that $\|x\|$ is a real number.

In any ips V over $\mathbb{F}$, the norm satisfies the following properties:

1. For each $x \in V$, $\|x\| \geq 0$.
2. For each $x \in V$, $x = 0$ iff $\|x\| = 0$.
3. For each $x \in V$ and for each $\alpha \in \mathbb{F}$, $\|\alpha x\| = |\alpha| \|x\|$.

Other most used properties of the norm, in an ips, are proved in the following theorem.

***Theorem* 2.3**
For all vectors x, y in an ips, the following are true:

(1) **(Parallelogram Law)** $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

(2) **(Cauchy-Schwartz Inequality)** $|\langle x, y\rangle| \leq \|x\| \, \|y\|$.
Further, equality holds iff one of x, y is a scalar multiple of the other.

(3) **(Triangle Inequality)** $\|x+y\| \leq \|x\| + \|y\|$.

(4) **(Reverse Triangle Inequality)** $\big| \|x\| - \|y\| \big| \leq \|x-y\|$.

Proof (1) $\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle$. Similarly, expand $\|x-y\|^2$ and complete the proof.

(2) If $y = 0$, then $\langle x, y\rangle = 0 = \|x\| \, \|y\|$. Assume that $y \neq 0$. Set $\alpha = \dfrac{\langle x, y\rangle}{\langle y, y\rangle}$. Then $\overline{\alpha} = \dfrac{\langle y, x\rangle}{\langle y, y\rangle}$; consequently, $\overline{\alpha}\langle x, y\rangle = \alpha\langle y, x\rangle = \alpha\overline{\alpha}\langle y, y\rangle = \dfrac{|\langle x, y\rangle|^2}{\|y\|^2}$. Now,

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

Therefore, $|\langle x, y\rangle| \leq \|x\| \, \|y\|$. Next, equality holds iff $y = 0$ or $x = \alpha y$ iff one is a scalar multiple of the other.

(3) $$\begin{aligned} \|x+y\|^2 = \langle x+y, x+y\rangle &= \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 \\ &= \|x\|^2 + 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \end{aligned}$$

(4) Using (3), we have $\|x\| \leq \|x-y\| + \|y\|$. Thus $\|x\| - \|y\| \leq \|x-y\|$. Similarly, $\|y\| - \|x\| \leq \|x-y\|$. Then the required inequality follows. ∎

Exercises for § 2.1

1. Check that the maps given in Example 2.1 are inner products on the respective vector spaces.
2. Check in each case below, whether the given map $\langle \cdot \rangle$ is an inner product on the given vector space:
   (a) $\langle x, y\rangle = ac$ for $x = (a, b)$, $y = (c, d)$ in $\mathbb{R}^2$.

(b) $\langle x, y\rangle = ac$ for $x = (a, b)$, $y = (c, d)$ in $\mathbb{C}^2$.
(c) $\langle x, y\rangle = ac - bd$ for $x = (a, b)$, $y = (c, d)$ in $\mathbb{R}^2$.
(d) $\langle p, q\rangle = \int_0^1 p'(t)q(t)\, dt$ for $p, q \in \mathbb{R}[t]$.
(e) $\langle x, y\rangle = \int_0^1 x'(t)y'(t)\, dt$ for $x, y \in C^1[0,1]$.
(f) $\langle x, y\rangle = x(0)y(0) + \int_0^1 x'(t)y'(t)\, dt$ for $x, y \in C^1[0,1]$.
(g) $\langle f, g\rangle = \int_0^{1/2} f(t)g(t)$ for $f, g \in C[0,1]$.

3. Let B be a basis for a finite dimensional ips V. Let $y \in V$ be such that $\langle x, y\rangle = 0$ for all $x \in B$. Show that $y = 0$.
4. Let V be a complex ips. Show that $\text{Re}\langle ix, y\rangle = -\text{Im}\langle x, y\rangle$ for all $x, y \in V$.
5. Let V be an inner product space, and let $x, y \in V$. Show the following:
   (a) $\|x\| \geq 0$.
   (b) $x = 0$ iff $\|x\| = 0$.
   (c) $\|\alpha x\| = |\alpha|\|x\|$, for all $\alpha \in \mathbb{F}$.
   (d) $\|x + \alpha y\| = \|x - \alpha y\|$ for all $\alpha \in \mathbb{F}$ iff $\langle x, y\rangle = 0$.
   (e) If $\|x + y\| = \|x\| + \|y\|$, then at least one of x, y is a scalar multiple of the other.

2.2 Orthonormal basis

In the presence of an inner product, we can define the non-obtuse angle between two vectors. Let x and y be nonzero vectors in an ips. The **angle between** x and y, denoted by $\theta(x, y)$, is defined by

$$\cos\theta(x, y) = \frac{|\langle x, y\rangle|}{\|x\|\,\|y\|}.$$

Notice that the angle is well defined due to Cauchy-Schwartz inequality. We will mainly work with a particular case of the angle, that is, when it is $\pi/2$.

Let x, y be vectors in an ips. We say that x is **orthogonal** to y iff $\langle x, y\rangle = 0$. When x is orthogonal to y, we write $x \perp y$.

Thus the zero vector is orthogonal to every vector, and $x \perp y$ implies $y \perp x$.

Example 2.4

1. Let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. If $i \neq j$, then $e_i \perp e_j$.
2. In $\mathbb{R}_2[t]$, with the inner product as $\langle p(t), q(t)\rangle = \int_{-1}^1 p(t)q(t)\, dt$, the vectors 1 and t are orthogonal to each other.

3. In $\mathbb{F}_n[t]$ with the inner product as
$$\langle a_0 + a_1 t + \cdots + a_n t^n, b_0 + b_1 t + \cdots + b_n t^n \rangle = a_0 \overline{b}_0 + a_1 \overline{b}_1 + \cdots + a_n \overline{b}_n,$$
the polynomials t^i and t^j are orthogonal to each other, provided $i \neq j$.
4. In $C[0, 2\pi]$, define $\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt$. Now, $\int_0^{2\pi} \cos mt \sin nt \, dt = 0$ for $m \neq n$. Hence for $m \neq n$, $\cos mt \perp \sin nt$. □

Theorem 2.5* (*Pythagoras)
Let x and y be vectors in an ips V.
(1) *If $x \perp y$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.*
(2) *If V is a real ips, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ implies $x \perp y$.*

Proof (1) $\langle x + y, x + y \rangle = \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2$. Since $x \perp y$, we have $\langle x, y \rangle = 0 = \langle y, x \rangle$. Then the equality follows.

(2) Suppose that V is a real ips. Then $\langle x, y \rangle = \langle y, x \rangle$. Now, $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ implies $\langle x, y \rangle = 0$. ∎

For a complex ips, the conclusion of the second statement in Pythagoras theorem need not be true. For example, consider $V = \mathbb{C}$, a complex ips with $\langle x, y \rangle = x\overline{y}$, as usual. Then $\|1 + i\|^2 = (1 + i)\overline{(1 + i)} = 1 + 1 = \|1\|^2 + \|i\|^2$. But $\langle 1, i \rangle = 1 \cdot (-i) = -i \neq 0$.

We extend the notion of orthogonality to a set of vectors.

Let S be a nonempty subset of nonzero vectors in an ips V. S is called an **orthogonal set** in V iff for all $x, y \in S$ with $x \neq y$, we have $x \perp y$.

S is called an **orthonormal** set in V iff S is an orthogonal set in V, and $\|x\| = 1$ for each $x \in S$.

Example 2.6
1. The standard basis of $\mathbb{R}^n$ is an orthonormal set.
2. Consider $C[0, 2\pi]$ as a real ips with $\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)\, dt$ as the inner product. The set of functions $\{\cos mt : m \in \mathbb{N}\}$ is an orthogonal set in $C[0, 2\pi]$. But $\int_0^{2\pi} \cos^2 t \, dt \neq 1$. Hence, the set is not orthonormal.
3. However, $\{(\cos mt)/\sqrt{\pi} : m \in \mathbb{N}\}$ is an orthonormal set in $C[0, 2\pi]$.
4. Any singleton set $\{v\}$ for $v \neq 0$, is an orthogonal set in any ips V. □

Notice that if $\{v_1, \ldots, v_n\}$ is an orthonormal set, then for $1 \leq i, j \leq n$,
$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$
Further, if $\{v_1, \ldots, v_n\}$ is an orthogonal set, then $\{v_1/\|v_1\|, \ldots, v_n/\|v_n\|\}$ is an orthonormal set.

Each orthogonal or orthonormal set in Example 2.6 is linearly independent; this is not a mere coincidence.

***Theorem* 2.7**
Every orthogonal (orthonormal) set is linearly independent.

Proof Let S be an orthogonal set in an ips V. Let $n \in \mathbb{N}$, $v_1, \ldots, v_n \in S$, and let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Suppose $\sum_{i=1}^n \alpha_i v_i = 0$. Let $1 \le j \le n$. Taking inner product with v_j, and using the fact that $\langle v_i, v_j \rangle = 0$ for $i \neq j$, we have

$$0 = \Big\langle \sum_{i=1}^n \alpha_i v_i, v_j \Big\rangle = \sum_{i=1}^n \alpha_i \langle v_i, v_j \rangle = \alpha_j \langle v_j, v_j \rangle.$$

Since $v_j \neq 0$, $\alpha_j = 0$. Therefore, S is linearly independent. ∎

Like maximal linearly independent sets, a *maximal orthogonal set* is an orthogonal set which when enlarged by including a vector not in the set, the enlarged set is no more orthogonal. In general, a maximal orthogonal set is called an *orthogonal basis* for the ips. Similarly, an *orthonormal basis* is a maximal orthonormal set. In infinite dimensional ips, an orthonormal basis may not span the space. However, in a finite dimensional ips, an orthogonal or an orthonormal basis spans the space; and hence, it is a basis in the conventional sense. Since we are concerned with finite dimensional ips only, we define orthogonal and orthonormal bases as in the following.

Let V be a finite dimensional ips. An orthogonal set which is also a basis for V is called an **orthogonal basis** of V. Similarly, an orthonormal set which is also a basis for V is called an **orthonormal basis** of V.

In case, we have an orthonormal basis for an ips, each vector can be expressed as a linear combination of the basis vectors, where the coefficients are just the inner products of the vector with the basis vectors. Moreover, the length of any vector can be expressed in a nice way.

***Theorem* 2.8**
Let $B = \{v_1, \ldots, v_n\}$ be an orthonormal basis of an ips V. Let $x \in V$. Then

(1) **(Fourier Expansion)** $x = \sum_{j=1}^n \langle x, v_j \rangle v_j$;

(2) **(Parseval's Identity)** $\|x\|^2 = \sum_{j=1}^n |\langle x, v_j \rangle|^2$.

Proof (1) Since B is a basis of V, $x = \sum_{i=1}^n \alpha_i v_i$ for some scalars α_i. Now,

$$\langle x, v_j \rangle = \Big\langle \sum_{i=1}^n \alpha_i v_i, v_j \Big\rangle = \sum_{i=1}^n \alpha_i \delta_{ij} = \alpha_j \quad \text{for } 1 \le j \le n.$$

Therefore, $x = \sum_{i=1}^n \alpha_i v_i = \sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \langle x, v_j \rangle v_j$.

(2) Using (1), we have

$$\begin{aligned}\|x\|^2 &= \Big\langle \sum_{j=1}^{n} \alpha_j v_j, \sum_{i=1}^{n} \alpha_i v_i \Big\rangle = \sum_{j=1}^{n} \alpha_j \Big(\sum_{i=1}^{n} \overline{\alpha}_i \langle v_j, v_i \rangle \Big) \\ &= \sum_{j=1}^{n} \alpha_j \Big(\sum_{i=1}^{n} \overline{\alpha}_i \delta_{ji} \Big) = \sum_{j=1}^{n} \alpha_j \overline{\alpha}_j = \sum_{j=1}^{n} |\alpha_j|^2 = \sum_{j=1}^{n} |\langle x, v_j \rangle|^2.\end{aligned}$$

∎

If the finite set B is not a basis for the ips V, then instead of an equality in Parseval's identity, we have an inequality.

***Theorem* 2.9 (Bessel's Inequality)**
Let $E = \{u_1, \ldots, u_m\}$ be an orthonormal set in an ips V. Let $x \in V$. Then $\sum_{j=1}^{m} |\langle x, u_j \rangle|^2 \le \|x\|^2$.

Proof Consider the ips $U = \operatorname{span}\{u_1, \ldots, u_m\}$, which is a subspace of V. Now, E is an orthonormal basis of U. Let $y = \sum_{j=1}^{m} \langle x, u_j \rangle u_j$. Now, $y \in U$. By Fourier expansion, $y = \sum_{j=1}^{m} \langle y, u_j \rangle u_j$. Since $\{u_1, \ldots, u_m\}$ is a basis of U, by Theorem 1.19, we have $\langle x, u_j \rangle = \langle y, u_j \rangle$ for $1 \le j \le m$. That is, $x - y \perp u_j$ for $1 \le j \le m$.

Then $x - y \perp y$. By Pythagoras' theorem and Parseval's identity, we obtain

$$\|x\|^2 = \|x-y\|^2 + \|y\|^2 \ge \|y\|^2 = \sum_{j=1}^{m} |\langle y, u_j \rangle|^2 = \sum_{j=1}^{n} |\langle x, u_j \rangle|^2.$$

∎

The vector y in the proof of Bessel's inequality has geometric significance. For illustartion, take U as the xy-plane, V as $\mathbb{R}^3$, and $x = (1, 2, 3)$. Choose the standard basis $\{e_1, e_2\}$ as the orthonormnal basis for U. Then

$$\langle x, e_1 \rangle = 1,\ \langle x, e_2 \rangle = 2, \quad y = 1\, e_1 + 2\, e_2 = (1, 2, 0).$$

The vector y is the orthogonal projection of x on U.

We may view Bessel's inequality in a different way. Imagine extending the orthonormal basis E of the subspace U to a basis B of the ips V. If B is also orthonormal, then Bessel's inequality would follow from Parseval's identity in a trivial way. But is this extension of an orthonormal basis of a subspace to the parent space always possible? We address this question in the next section.

Exercises for § 2.2

1. Consider $\mathbb{R}^4$ as a real ips with the standard inner product. Let $W = \{x \in \mathbb{R}^4 : x \perp (1, 0, -1, 1), x \perp (2, 3, -1, 2)\}$. Show that W is a subspace of $\mathbb{R}^4$. Further, find a basis for W.

2. For vectors x, y in an ips V, prove that $x + y \perp x - y$ iff $\|x\| = \|y\|$.
3. Consider the standard basis for $\mathbb{R}^n$. Take $x = (a_1, \ldots, a_n)$. Verify Fourier expansion and Prseval's identity.
4. Take $x = (n, n-1, \ldots, 1) \in \mathbb{R}^n$. Consider the orthonormal set $\{e_2, \ldots, e_{n-1}\}$. Verify Bessel's inequality.
5. Let $\{u_1, u_2, \ldots, u_k\}$ be an orthogonal set in a real ips V. Let $a_1, \ldots, a_k \in \mathbb{R}$. Is it rue that $\|\sum_{i=1}^k a_i u_i\|^2 = \sum_{i=1}^k |a_i|^2 \|u_i\|^2$?
6. Formulate a converse of Parseval's identity and prove it.
7. In the proof of Bessel's inequality, by plugging in the expression for y, show directly that $\langle x - y, y \rangle = 0$.
8. For a subset S of an ips V, define $S^\perp = \{x \in V : \langle x, u \rangle = 0, \text{ for all } u \in S\}$. Show the following:
   (a) $V^\perp = \{0\}$, and $\{0\}^\perp = V$.
   (b) $S^\perp$ is a subspace of V, and $S \subseteq S^{\perp\perp}$.
9. Let W be a subspace of a finite dimensional ips V. Show that $W^{\perp\perp} = W$.
10. Show that $\{\sin t,\ \sin(2t),\ \ldots, \sin(mt)\}$ is linearly independent in $C[0, 2\pi]$.

2.3 Gram-Schmidt orthogonalization

Given two linearly independent vectors u_1, u_2 on the plane how do we construct two orthogonal vectors v_1, v_2 such that $\operatorname{span}\{u_1, u_2\} = \operatorname{span}\{v_1, v_2\}$?

Keep $v_1 = u_1$. Project u_2 on u_1, and subtract the result from u_2 to get v_2. Now, $v_2 \perp v_1$. It is easy to see that the span condition is also satisfied.

We may continue this process of taking projections in n dimensions. Its general version is called the *Gram-Schmidt orthogonalization* process. We formulate and prove this process in the following theorem.

***Theorem* 2.10**
Let $u_1, \ldots, u_n$ be linearly independent vectors in an ips V. Construct the vectors $v_1, \ldots, v_n$ as follows:

$$\begin{aligned} v_1 &= u_1 \\ v_k &= u_k - \frac{\langle u_k, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_k, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 - \cdots - \frac{\langle u_k, v_{k-1} \rangle}{\langle v_{k-1}, v_{k-1} \rangle} v_{k-1} \quad \textit{for } k > 1. \end{aligned}$$

Then $\operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\{u_1, \ldots, u_n\}$, *and* $\{v_1, \ldots, v_n\}$ *is orthogonal.*

Proof We use induction on n. For $n = 1$, $v_1 = u_1$. Thus span$\{v_1\}$ = span$\{u_1\}$. Moreover, $\{u_1\}$ is linearly independent implies that $u_1 \neq 0$. So, $v_1 = u_1 \neq 0$. Then $\{v_1\}$ is an orthogonal set.

Assume that the conclusion of the theorem is true for $n = m$. We show that it is true for $n = m + 1$.

By assumption, span$\{v_1, \ldots, v_m\}$ = span$\{u_1, \ldots, u_m\}$. Let v be a linear combination of $v_1, \ldots, v_m, v_{m+1}$. Then v is a linear combination of $u_1, \ldots, u_m$, v_{m+1}. As v_{m+1} is a linear combination of $v_1, \ldots, v_m, u_{m+1}$, it is a linear combination of $u_1, \ldots, u_m, u_{m+1}$. Thus v is a linear combination of $u_1, \ldots, u_m, u_{m+1}$. Similarly, if u is a linear combination of $u_1, \ldots, u_m, u_{m+1}$, then it is a linear combination of $v_1, \ldots, v_m, u_{m+1}$. Again, u_{m+1} is a linear combination of $v_1, \ldots, v_{m+1}$ implies that u is a linear combination of $v_1, \ldots, v_{m+1}$. This proves that span$\{v_1, \ldots, v_{m+1}\}$ = span$\{u_1, \ldots, u_{m+1}\}$.

By assumption, $\{v_1, \ldots, v_m\}$ is orthogonal. Let $1 \leq j \leq m$. Then for any $i \neq j$, $1 \leq i \leq m$, $\langle v_i, v_j \rangle = 0$. We obtain

$$\langle v_{m+1}, v_j \rangle = \Big\langle u_{m+1} - \sum_{i=1}^{m} \frac{\langle u_{m+1}, v_i \rangle}{\langle v_i, v_i \rangle} v_i, v_j \Big\rangle = \langle u_{m+1}, v_j \rangle - \frac{\langle u_{m+1}, v_j \rangle}{\langle v_j, v_j \rangle} \langle v_j, v_j \rangle = 0.$$

Therefore, $\{v_1, \ldots, v_m, v_{m+1}\}$ is an orthogonal set. ∎

Starting from a basis for any ips V, Gram-Schmidt process yields an orthogonal basis for V.

Example 2.11
The vectors $u_1 = (1, 1, 0)$, $u_2 = (0, 1, 1)$, $u_3 = (1, 0, 1)$ form a basis for $\mathbb{F}^3$. Apply Gram-Schmidt Orthogonalization to obtain an orthogonal basis of $\mathbb{F}^3$.

$$\begin{aligned}
v_1 &= (1, 1, 0). \\
v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = (0, 1, 1) - \frac{(0, 1, 1) \cdot (1, 1, 0)}{(1, 1, 0) \cdot (1, 1, 0)} (1, 1, 0) \\
&= (0, 1, 1) - \tfrac{1}{2}(1, 1, 0) = (-\tfrac{1}{2}, \tfrac{1}{2}, 1). \\
v_3 &= u_3 - \frac{\langle u_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle u_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 \\
&= (1, 0, 1) - (1, 0, 1) \cdot (1, 1, 0)(1, 1, 0) - (1, 0, 1) \cdot (-\tfrac{1}{2}, \tfrac{1}{2}, 1)\,(-\tfrac{1}{2}, \tfrac{1}{2}, 1) \\
&= (1, 0, 1) - \tfrac{1}{2}(1, 1, 0) - \tfrac{1}{3}(-\tfrac{1}{2}, \tfrac{1}{2}, 1) = (-\tfrac{2}{3}, \tfrac{2}{3}, -\tfrac{2}{3}).
\end{aligned}$$

The required orthogonal basis of $\mathbb{F}^3$ is $\{(1, 1, 0), (-\frac{1}{2}, \frac{1}{2}, 1), (-\frac{2}{3}, \frac{2}{3}, -\frac{2}{3})\}$. □

Example 2.12
The vectors $u_1 = 1$, $u_2 = t$, $u_3 = t^2$ form a linearly independent set in the ips of all polynomials considered as functions from $[-1, 1]$ to $\mathbb{R}$; with the inner product as $\langle p(t), q(t) \rangle = \int_{-1}^{1} p(t)q(t)\, dt$. Gram-Schmidt Process yields:

$$\begin{aligned}
v_1 &= u_1 = 1. \\
v_2 &= u_2 - \frac{\langle u_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 = t - \frac{\int_{-1}^{1} t\,dt}{\int_{-1}^{1} dt} 1 = t. \\
v_3 &= u_3 - \frac{\langle u_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle u_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2 = t^2 - \frac{\int_{-1}^{1} t^2\,dt}{\int_{-1}^{1} dt} 1 - \frac{\int_{-1}^{1} t^3\,dt}{\int_{-1}^{1} t^2\,dt} t = t^2 - \tfrac{1}{3}.
\end{aligned}$$

The set $\{1,\, t,\, t^2 - \frac{1}{3}\}$ is an orthogonal set in this ips. Moreover, span$\{1,t,t^2\}$ = span$\{1,t,t^2 - \frac{1}{3}\}$. ∎

In fact, orthogonalization of $\{1,t,t^2,t^3,t^4,\ldots\}$ with the above inner product gives the Legendre Polynomials.

Observe that an orthogonal set can be made orthonormal by dividing each vector by its norm. We thus obtain the following result.

***Theorem* 2.13**
An orthonormal set in a finite dimensional ips can be extended to an orthonormal basis. Every finite dimensional ips has an orthonormal basis.

Notice that if you apply Gram-Schmidt orthogonalization on a linearly dependent set, then it will generate the zero vector. That is, if $\{u_1,\ldots,u_k\}$ is linearly independent and u_{k+1} is a linear combination of $u_1,\ldots,u_k$, then v_{k+1} will turn out to be the zero vector. Ignoring such zero vectors in the Gram-Schmidt process leads to an orthogonal basis for the span of the given vectors. This is how Gram-Schmidt orthogonalization process is used to extract a basis from a finite spanning set.

Exercises for § 2.3

1. Consider $\mathbb{R}^3$ with the standard inner product. Apply Gram-Schmidt process on the given set of vectors.
   (a) $\{(1,2,0),\ (2,1,0),\ (1,1,1)\}$.
   (b) $\{(1,1,1),\ (1,-1,1),\ (1,1,-1)\}$.
   (c) $\{(0,1,1),\ (0,1,-1),\ (-1,1,-1)\}$.
2. Consider $\mathbb{R}^3$ with the standard inner product. In each of the following, find a vector of norm 1 which is orthogonal to the given two vectors:
   (a) $(2,1,0),\ (1,2,1)$.
   (b) $(1,2,3),\ (2,1,-2)$.
   (c) $(0,2,-1),\ (-1,2,-1)$.

3. Consider $\mathbb{R}^4$ with the standard inner product. In each of the following, find the set of all vectors orthogonal to both u and v.
   (a) $u = (1,2,0,1)$, $v = (2,1,0,-1)$.
   (b) $u = (1,1,1,0)$, $v = (1,-1,1,1)$.
   (c) $u = (0,1,1,-1)$, $v = (0,1,-1,1)$.
4. Consider $\mathbb{C}^3$ with the standard inner product. Find an orthonormal basis for the subspace spanned by the vectors $(1,0,i)$ and $(2,1,1+i)$.
5. Consider the polynomials $u_0(t) = 1$, $u_1(t) = t$, $u_2(t) = t^2$ in $\mathbb{R}_2[t]$. Using Gram-Schmidt orthogonalization, find orthogonal polynomials obtained from u_1, u_2, u_3 with respect to the following inner products:
   (a) $\langle p,q\rangle = \int_0^1 p(t)q(t)\,dt$.
   (b) $\langle p,q\rangle = \int_{-1}^1 p(t)q(t)\,dt$.
   (c) $\langle p,q\rangle = \int_{-1}^0 p(t)q(t)\,dt$.
6. Equip $\mathbb{R}_3[t]$ with the inner product $\langle f,g\rangle = \int_0^1 f(t)g(t)dt$.
   (a) Find the set of all vectors orthogonal to the constant polynomials.
   (b) Apply Gram-Schmidt process to the ordered basis $\{1, t, t^2, t^3\}$.

2.4 Best approximation

In $\mathbb{R}^3$, given a plane and a point not on the plane, we are interested in finding a point on the plane that is closest to the given point. Such a point on the plane may be told to approximate the given point from the plane.

Of course, the phrase 'closest' is meaningful when we have the notion of distance. If u and v are two vectors in an ips, the *distance* between them may be defined as $\|u-v\|$. Given a subspace U of an ips V, and a vector $v \in V$, we look for a vector in u that minimizes the distance $\|v-x\|$ while x varies over the subspace U.

Let U be a subspace of an ips V. Let $v \in V$. A vector $u \in U$ is called a **best approximation of v from U** iff $\|v-u\| \le \|v-x\|$ for each $x \in U$.

Example 2.14
Let U be the plane $\{(a,b,c) \in \mathbb{R}^3 : a+b+c = 0\}$. Find a best approximation of the point $(1,1,1)$ from U.

Suppose (α,β,γ) is a best approximation of $(1,1,1)$ from U. Such a point satisfies $\alpha+\beta+\gamma = 0$ and minimizes the distance $\|(1,1,1)-(\alpha,\beta,\gamma)\|$. Substituting $\gamma = -\alpha-\beta$, we look for $\alpha,\beta \in \mathbb{R}$ so that

$$f(\alpha,\beta) = \left((1-\alpha)^2 + (1-\beta)^2 + (1+\alpha+\beta)^2\right)^{1/2}$$

is minimum. Simplifying the expression for $f(\alpha,\beta)$, we see that it is equivalent to minimizing

$$g(\alpha,\beta)=\alpha^2+\beta^2+\alpha\beta.$$

Then by using the methods of functions of two variables calculus, we may determine the required best approximation as $(0,0,0)$. $\square$

As it happens a best approximation in an ips can always be obtained in an easier way than applying calculus of several variables.

***Theorem* 2.15**
Let U be a subspace of an ips V. Let $v\in V$. A vector $u\in U$ is a best approximation of v from U iff $v-u\perp x$ for each $x\in U$. Moreover, a best approximation is unique.

Proof Let $u\in U$ satisfy $v-u\perp x$ for each $x\in U$. If $x\in U$, then $x-u\in U$. Thus for each $x\in U$, By Pythagoras' Theorem,

$$\|v-x\|^2=\|(v-u)+(u-x)\|^2=\|v-u\|^2+\|u-x\|^2\ge\|v-u\|^2.$$

Therefore, u is a best approximation of v from U.

Conversely, suppose that u is a best approximation of v. Then

$$\|v-u\|\le\|v-x\|\text{ for each }x\in U. \tag{2.1}$$

Let $y\in U$. We want to show that $\langle v-u,y\rangle=0$. For $y=0$, clearly $\langle v-u,y\rangle=0$. For $y\ne 0$, let $\alpha=\langle v-u,y\rangle/\|y\|^2$. Then

$$\langle v-u,\alpha y\rangle=\overline{\alpha}\langle v-u,y\rangle=|\alpha|^2\|y\|^2,\quad \langle\alpha y,v-u\rangle=\overline{\langle v-u,\alpha y\rangle}=|\alpha|^2\|y\|^2.$$

Notice that $u+\alpha y\in U$. From (2.1), we have

$$\begin{aligned}\|v-u\|^2 &\le \|v-u-\alpha y\|^2=\langle v-u-\alpha y,v-u-\alpha y\rangle\\ &= \|v-u\|^2-\langle v-u,\alpha y\rangle-\langle\alpha y,v-u\rangle+\alpha\overline{\alpha}\langle y,y\rangle\\ &= \|v-u\|^2-|\alpha|^2\|y\|^2.\end{aligned}$$

Hence, $|\alpha|^2\|y\|^2=0$. As $y\ne 0$, $|\alpha|^2=0$. It follows that $\langle v-u,y\rangle=0$.

To see the uniqueness of a best approximation, suppose that $u,w\in U$ are best approximations to v. Then $\|v-u\|\le\|v-w\|$ and $\|v-w\|\le\|v-u\|$. So, $\|v-u\|=\|v-w\|$.

Moreover, $w-u\in U$. Therefore, by what we have just proved, $v-w\perp w-u$. By Pythagoras' theorem,

$$\|v-u\|^2=\|(v-w)+(w-u)\|^2=\|v-w\|^2+\|w-u\|^2=\|v-u\|^2+\|w-u\|^2.$$

Thus, $\|w-u\|^2=0$. That is, $w=u$. ∎

Observe that Theorem 2.15 does not guarantee the existence of a best approximation. We show that if U is finite dimensional subspace of an ips V, then corresponding to each vector $v \in V$, there exists a unique best approximation u to v from U; and such a vector u can be given in a closed form using an orthonormal basis for U.

***Theorem* 2.16**
Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis for a subspace U of an ips V. Let $v \in V$. Then $u = \sum_{i=1}^n \langle v, u_i \rangle u_i$ is the best approximation of v from U.

Proof Write $u := \sum_{i=1}^n \langle v, u_i \rangle u_i$. Since $u \in U$, by Fourier expansion, we have $u = \sum_{i=1}^n \langle u, u_i \rangle u_i$. Due to Theorem 1.19, $\langle v, u_i \rangle = \langle u, u_i \rangle$ for $1 \le i \le n$. That is, $\langle v - u, u_i \rangle = 0$ for each $i \in \{1, \ldots, n\}$.

Now, suppose $x \in U$. There exist scalars $a_1, \ldots, a_n$ such that $x = \sum_{i=1}^n a_i u_i$. Then $\langle v - u, x \rangle = \sum_{i=1}^n \overline{a}_i \langle v - u, u_i \rangle = 0$. That is, $v - u \perp x$ for each $x \in U$. By Theorem 2.15, u is the best approximation of v from U. ∎

The orthogonality condition $v - u \perp x$ for each $x \in U$ in Theorem 2.15 is equivalent to $v - u \perp u_j$ for each j whenever $\{u_1, \ldots, u_n\}$ is a spanning set for U. This is helpful in computing the best approximation, without using an orthonormal basis.

Suppose $\{u_1, \ldots, u_n\}$ is any basis of U. Write the best approximation of v from U as $u = \sum_{j=1}^n \beta_j u_j$ with unknown scalars β_j. Then using the orthogonality condition, we have $\langle v - \sum_{j=1}^n \beta_j u_j, u_i \rangle = 0$. This way, the scalars β_j are determined from the linear system

$$\sum_{j=1}^{n} \langle u_j, u_i \rangle \beta_j = \langle v, u_i \rangle \quad \text{for } i = 1, \ldots, n.$$

Example 2.17

1. For the best approximation of $v = (1, 0) \in \mathbb{R}^2$ from $U = \{(a, a) : a \in \mathbb{R}\}$, we look for a point (α, α) so that $(1, 0) - (\alpha, \alpha) \perp (\beta, \beta)$ for all β. That is, we look for an α so that $(1 - \alpha, -\alpha) \cdot (1, 1) = 0$. Or, $1 - \alpha - \alpha = 0$. It leads to $\alpha = 1/2$. The best approximation here is $(1/2, 1/2)$.
2. Reconsider Example 2.14. We require a vector $(\alpha, \beta, \gamma) \in U = \{(a, b, c) \in \mathbb{R}^3 : a + b + c = 0\}$ which is the best approximation to $(1, 1, 1)$. A basis for U is given by $\{(1, -1, 0),\ (0, 1, -1)\}$. The orthogonality condition in Theorem 2.15 implies that

$$(1, 1, 1) - (\alpha, \beta, \gamma) \perp (1, -1, 0), \quad (1, 1, 1) - (\alpha, \beta, \gamma) \perp (0, 1, -1).$$

These equations along with the fact $(\alpha, \beta, \gamma) \in U$ give

$$1 - \alpha - 1 + \beta = 0, \quad 1 - \beta - 1 + \gamma = 0, \quad \alpha + \beta + \gamma = 0.$$

That is, $\alpha = \beta = \gamma = 0$. Therefore, $(0,0,0)$ is the required best approximation.

Alternatively, an orthonormal basis for U is given by $\{u_1, u_2\}$, where $u_1 = (1/\sqrt{2}, -1/\sqrt{2}, 0)$ and $u_2 = (1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6})$. With $v = (1,1,1)$, we obtain $\langle v, u_1\rangle = 0 = \langle v, u_2\rangle$. The best approximation of v from U is given by $u = \langle v, u_1\rangle u_1 + \langle v, u_2\rangle u_2 = (0,0,0)$.

3. In $\mathbb{R}[t]$ with $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, what is the best approximation of t^2 from $\mathbb{R}_1[t]$?

   Write the best approximation of t^2 from $\mathbb{R}_1[t]$ as $\alpha + \beta t$. The orthogonality condition asks us to determine $\alpha, \beta \in \mathbb{R}$ so that $t^2 - (\alpha + \beta t) \perp 1$ and $t^2 - (\alpha + \beta t) \perp t$. That is,

$$\int_0^1 (t^2 - \alpha - \beta t)\,dt = 0 = \int_0^1 (t^3 - \alpha t - \beta t^2)\,dt.$$

   This gives $\frac{1}{3} - \alpha - \frac{\beta}{2} = 0 = \frac{1}{4} - \frac{\alpha}{2} - \frac{\beta}{3}$ leading to $\alpha = -\frac{1}{6}$ and $\beta = 1$. Therefore, the best approximation is $-\frac{1}{6} + t$. $\square$

Observe that the best approximation u of a vector $v \in V$ from U is the orthogonal projection of v on U. It is the same vector u that we have used in the proof of Bessel's inequality.

Exercises for § 2.4

Find the best approximation of $v \in V$ from U in the following:

1. $V = \mathbb{R}^3$, $v = (1,2,1)$, $U = \operatorname{span}\{(3,1,2), (1,0,1)\}$.
2. $V = \mathbb{R}^3$, $v = (1,2,1)$, $U = \{(\alpha, \beta, \gamma) \in \mathbb{R}^3 : \alpha + \beta + \gamma = 0\}$.
3. $V = \mathbb{R}^4$, $v = (1,0,-1,1)$, $U = \operatorname{span}\{(1,0,-1,1), (0,0,1,1)\}$.
4. $V = \mathbb{R}_3[t]$, $v = t^3$, $U = \operatorname{span}\{1, 1+t, 1+t^2\}$, $\langle p, q\rangle = \int_0^1 p(t)q(t)\,dt$.
5. $V = C[-1,1]$, $v(t) = e^t$, $U = \mathbb{R}_4[t]$, $\langle f, g\rangle = \int_{-1}^1 f(t)g(t)\,dt$.

3

Linear Transformations

3.1 Linear maps

The interesting maps in a domain of mathematical discourse are those maps which preserve structures. Since a vector space gets its structure from the two operations of addition and scalar multiplication, we are interested in maps that preserve these operations.

Let V and W be vector spaces over the same field $\mathbb{F}$. A function $T : V \to W$ is said to be a **linear transformation** (or a linear map, or a linear operator) iff for all $x, y \in V$ and for each $\alpha \in \mathbb{F}$,

$$T(x + y) = T(x) + T(y) \text{ and } T(\alpha x) = \alpha T(x).$$

A linear transformation from V to V is called a **linear operator on** V.

A linear transformation from V to $\mathbb{F}$ is called a **linear functional**.

Observe that the two conditions in the definition of a linear transformation amount to the single condition

$$T(x + \alpha y) = T(x) + \alpha T(y) \quad \text{for all } x, y \in V, \text{ and for each } \alpha \in \mathbb{F}.$$

Example 3.1

1. Let V be a vector space. The map $T : V \to V$ defined by $T(v) = 0$ for each $v \in V$ is a linear operator on V; it is called the *zero operator*.
2. Let V be a vector space. The map $T : V \to V$ defined by $T(v) = v$ is a linear operator on V; it is called the *identity operator*.
3. Let V be a vector space. Let α be any scalar. Then the map $T : V \to V$ defined by $T(v) = \alpha v$ is a linear operator on V.
4. Define the map $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(a, b, c) = (2a + b, b - c)$. Then T is a linear transformation.
5. The map $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(a, b) = (a + b, 2a - b, a + 6b)$ is a linear transformation.

6. For $1 \le i \le m$ and $1 \le j \le n$, let $a_{ij} \in \mathbb{F}$. Define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(\beta_1, \ldots, \beta_n) = \Big(\sum_{j=1}^{n} a_{1j}\beta_j, \ldots, \sum_{j=1}^{n} a_{mj}\beta_j\Big).$$

   Then, T is a linear transformation.
7. Fix $\phi \in [0, 2\pi]$. For any $x = (r\cos\theta, r\sin\theta) \in \mathbb{R}^2$, the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(x) = \big(r\cos(\theta+\phi), r\sin(\theta+\phi)\big)$$

   is the rotation by the angle ϕ. It is a linear operator on $\mathbb{R}^2$.
8. Fix $\phi \in [0, 2\pi]$. For any $x = (r\cos\theta, r\sin\theta) \in \mathbb{R}^2$, the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(x) = \big(r\cos(\phi-\theta), r\sin(\phi-\theta)\big)$$

   is the reflection on the line making an angle $\phi/2$ with $\mathbb{R} \times \{0\}$, the x-axis. T is a linear operator on $\mathbb{R}^2$.
9. For $1 \le j \le n$, define $T_j : \mathbb{R}^n \to \mathbb{R}$ by $T_j(a_1, \ldots, a_n) = a_j$. Then T_j is a linear functional. In general, let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Define the map $T : \mathbb{R}^n \to \mathbb{R}$ by $T(a_1, \ldots, a_n) = \sum_{i=1}^{n} \alpha_i a_i$. Then T is a linear functional.
10. Let V be a vector space with basis $\{v_1, \ldots, v_n\}$. Given any vector v, there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Fix any $i \in \{1, \ldots, n\}$. Define the map $T_i : V \to \mathbb{F}$ by $T_i(v) = \alpha_i$. Then T_i is a linear functional. Reason?

    Any $v \in V$ can be written as $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, where α_i is a unique scalar for each $i \in \{1, \ldots, n\}$. Thus $T_i(v) = \alpha_i$ defines a function T_i from V to $\mathbb{F}$. To see that T_i is a linear functional, let $v, w \in V$. There exist unique scalars $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = \beta_1 v_1 + \cdots + \beta_n v_n$. Then

$$T_i(v+w) = T_i\Big(\sum_{j=1}^{n} (\alpha_j + \beta_j) v_j\Big) = \alpha_i + \beta_i = T_i(v) + T_i(w).$$

    Similarly the other condition can be verified.
11. Let $n \in \mathbb{N}$. Define the map $T : \mathbb{F}_n[t] \to \mathbb{F}_{n-1}[t]$ by $T(p(t)) = p'(t)$, the derivative of $p(t)$ with respect to t. Then T is a linear transformation.
12. Let $\alpha \in [a,\ b]$. Define the function $T_\alpha : C[a,\ b] \to \mathbb{F}$ by $T_\alpha(f) = f(\alpha)$. Verify that T_α is a linear functional.
13. Let the function $T : C^1[a,\ b] \to C[a,\ b]$ be defined by $T(f) = f'$. Then T is a linear transformation. (Here $C^k[a,b]$ is the vector space of all k-times continuously differentiable functions from $[a,b]$ to $\mathbb{R}$.)
14. Let the map $T : C^1[a,\ b] \to C[a,\ b]$ be defined by $T(f) = \alpha f + \beta f'$, where α, β are fixed scalars. Verify that T is a linear transformation.

15. Let S and T be linear transformations from V to W. Let $\alpha, \beta \in \mathbb{F}$. Then the map $A : V \to W$ defined by $A(v) = \alpha S(v) + \beta T(v)$ is a linear transformation.

$\square$

Remember that we talk of a linear transformation from V to W only when both V and W are vector spaces over the same field $\mathbb{F}$, be it $\mathbb{R}$ or $\mathbb{C}$.

Convention 3.1 When we say that $T : V \to W$ is a linear transformation, we assume that V and W are vector spaces over the same field $\mathbb{F}$.

***Theorem* 3.2**
Let $T : V \to W$ be a linear transformation. Then the following are true:

(1) $T(0) = 0$.

(2) *For all $u, v \in V$, $T(u - v) = T(u) - T(v)$.*

(3) *For any $n \in \mathbb{N}$, for all $v_1, \ldots, v_n \in V$ and for all scalars $\alpha_1, \ldots, \alpha_n$, $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$.*

Proof (1) $T(0) + T(0) = T(0 + 0) = T(0) = T(0) + 0$. Hence $T(0) = 0$.
(2) $T(u) = T(u - v + v) = T(u - v) + T(v)$. Hence $T(u - v) = T(u) - T(v)$.
(3) It follows by induction. ∎

Every map that 'looks linear' need not be a linear transformation. For instance, the map $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = 2x + 3$ is not a linear transformation since $T(0) \neq 0$.

We show that composition of two (hence any finite number of) linear transformations is a linear transformation.

***Theorem* 3.3**
Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then $S \circ T : U \to W$ is a linear transformation.

Proof Recall that the map $S \circ T$ is defined by $(S \circ T)(u) = S(T(u))$ for $u \in U$. Let $x, y \in U$ and let $\alpha \in \mathbb{F}$. Now,

$$\begin{aligned} (S \circ T)(x + \alpha y) &= S(T(x + \alpha y)) = S(T(x) + \alpha T(y)) \\ &= S(T(x)) + \alpha S(T(y)) = (S \circ T)(x) + \alpha (S \circ T)(y). \end{aligned}$$

Therefore, $(S \circ T)$ is a linear transformation. ∎

In what follows we will abbreviate $T(x)$ to Tx and $S \circ T$ to ST, whenever it does not harm our understanding.

Exercises for § 3.1

1. In each of the following determine whether $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation:
   (a) $T(a, b) = (1, b)$.
   (b) $T(a, b) = (a, a^2)$.
   (c) $T(a, b) = (\sin a, 0)$.
   (d) $T(a, b) = (|a|, b)$.
   (e) $T(a, b) = (a + 1, b)$.
   (f) $T(a, b) = (2a + b, a + b^2)$.
2. What are the linear operators on $\mathbb{R}$?
3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with $T(1, 0) = (1, 4)$ and $T(1, 1) = (2, 5)$. What is $T(2, 3)$? Is T one-one?
4. Let $S : C^1[0, 1] \to C[0, 1]$ and $T : C[0, 1] \to \mathbb{R}$ be defined by $S(u) = u'$ and $T(v) = \int_0^1 v(t)\, dt$. Find, if possible, ST and TS. Are they linear transformations?
5. Does there exist a linear operator on $\mathbb{R}^2$ which maps the square with corners at $(-1, -1)$, $(1, -1)$, $(1, 1)$, $(-1, 1)$ onto the square with corners at $(-1, 0)$, $(1, 0)$, $(1, 2)$, $(-1, 2)$?
6. Give a linear transformation T from V onto $\mathbb{F}^2$, where $\dim(V) = 2$.

3.2 Action on a basis

Consider the linear transformation $D : \mathbb{R}_3[t] \to \mathbb{R}_2[t]$ defined by $D(p(t)) = p'(t)$. We know that $D(t^3) = 3t^2$, $D(t^2) = 2t$, $D(t) = 1$ and $D(1) = 0$. We may then use its linearity to obtain $D(p(t))$ for any polynomial $p(t) \in \mathbb{R}_3[t]$. For instance,

$$D(a + bt + ct^2 + dt^3) = aD(1) + bD(t) + cD(t^2) + dD(t^3) = b + 2ct + 3dt^2.$$

Similarly, let $T : \mathbb{R}^3 \to \mathbb{R}$ be a linear transformation with $T(1, 0, 0) = 2$ and $T(0, 1, 0) = -1$. Since $(2, 3, 0) = 2(1, 0, 0) + 3(0, 1, 0)$,

$$T(2, 3, 0) = 2T(1, 0, 0) + 3T(0, 1, 0) = 2 \times 2 + 3 \times (-1) = 1.$$

Can we construct a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(1, 0, 0) = (1, 0)$? In fact, we can give infinitely many such linear transformations. For instance, if $\alpha \in \mathbb{R}$, then the map T given by $T(a, b, c) = (a, \alpha b)$ is a linear transformation.

We may also give infinitely many linear transformations $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(1,0,0) = (1,0)$ and $T(0,1,0) = (0,1)$. For instance, corresponding to each $\alpha \in \mathbb{R}$ define T by $T(a,b,c) = (a+\alpha c, b+\alpha c)$.

However, there exists only one linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(1,0,0) = (1,0)$, $T(0,1,0) = (0,1)$, and $T(0,0,1) = (1,1)$. Reason?

$$\begin{aligned} T(a,b,c) &= a\,T(1,0,0) + b\,T(0,1,0) + c\,T(0,0,1) \\ &= a\,(1,0) + b\,(0,1) + c\,(1,1) = (a+c, b+c). \end{aligned}$$

What are the information required to describe a linear transformation? The following theorem provides an answer, which we state informally and explain its formal meaning in the proof.

***Theorem* 3.4**
A linear transformation is uniquely determined from its action on a basis.

Proof Let V and W be vector spaces. Let $B = \{v_1, \ldots, v_n\}$ be a basis of V. Let $w_1, \ldots, w_n \in W$. We want to show that there exists a unique linear transformation $T : V \to W$ with $T(v_1) = w_1,\ T(v_2) = w_2,\ \ldots,\ T(v_n) = w_n$. Notice that the vectors $w_1, \ldots, w_n$, which are the images of the basis vectors, need neither be linearly independent, nor be distinct.

For the existence of such a map, we construct one from V to W, and then prove that this map is a linear transformation.

Let $x \in V$. Then $x = a_1 v_1 + \cdots + a_n v_n$ for unique scalars $a_1, \ldots, a_n$. Define

$$T(x) = a_1 w_1 + \cdots + a_n w_n.$$

Due to uniqueness of the scalars, this map is well-defined. We must verify the two defining conditions of a linear transformation.

Let $u, v \in V$. Then $u = b_1 v_1 + \cdots + b_n v_n$ and $v = c_1 v_1 + \cdots + c_n v_n$ for some scalars b_i, c_j. Now, $u + v = (b_1 + c_1) v_1 + \cdots + (b_n + c_n) v_n$. Thus

$$\begin{aligned} T(u+v) &= (b_1+c_1) w_1 + \cdots + (b_n + c_n) w_n \\ &= (b_1 w_1 + \cdots + b_n w_n) + (c_1 w_1 + \cdots + c_n w_n) = T(u) + T(v). \end{aligned}$$

Similarly, for any scalar α,

$$\begin{aligned} T(\alpha u) &= T(\alpha b_1 v_1 + \cdots + \alpha b_n v_n) = \alpha b_1 w_1 + \cdots + \alpha b_n w_n \\ &= \alpha (b_1 w_1 + \cdots + b_n w_n) = \alpha T(u). \end{aligned}$$

Therefore T is a linear transformation.

For uniqueness of T, suppose $S : V \to W$ is a linear transformation with $S(v_1) = w_1,\ S(v_2) = w_2,\ \ldots,\ S(v_n) = w_n$. That is,

$$S(v_i) = T(v_i) \quad \text{for } 1 \le i \le n.$$

We show that $S = T$. For this, let $y \in V$. We have scalars $\alpha_i \in \mathbb{F}$ such that $y = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then

$$\begin{aligned} S(y) &= S(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 S(v_1) + \cdots + \alpha_n S(v_n) \\ &= \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = T(\alpha_1 v_2 + \cdots + \alpha_n v_n) = T(y). \end{aligned}$$

Hence $S = T$ as maps from V to W. ∎

Example 3.5

1. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation such that $T(1,0,0) = (2,3)$, $T(0,1,0) = (-1,4)$ and $T(0,0,1) = (5,-3)$. Then

$$\begin{aligned} T(3,-4,5) &= 3T(1,0,0) + (-4)T(0,1,0) + 5T(0,0,1) \\ &= 3(2,3) + (-4)(-1,4) + 5(5,-3) = (35,-22). \end{aligned}$$

2. Construct a linear transformation $T : \mathbb{R}^2 \to \{(a,b,c) : a - b - c = 0\}$.

   We start with a basis, say, $\{v_1 = (1,0),\ v_2 = (0,1)\}$ of $\mathbb{R}^2$. Let $W := \{(a,b,c) : a - b - c = 0\}$. Choose any two vectors in W, for instance, $w_1 = (1,1,0)$ and $w_2 = (1,0,1)$. We want $T(1,0) = w_1$ and $T(0,1) = w_2$. Thus define

$$\begin{aligned} T(a,b) &= T(a(1,0) + b(0,1)) = aT(1,0) + bT(0,1) \\ &= aw_1 + bw_2 = a(1,1,0) + b(1,0,1) = (a+b,a,b). \end{aligned}$$

   This is one of the required linear transformations from $\mathbb{R}^2$ to W. Find another linear transformation.

3. Construct a linear transformation $T : \mathbb{R}_3[t] \to \mathbb{R}^2$ such that

$$T(1+t) = (1,2),\ T(1-t) = (1,1),\ T(1-t^2) = (2,1).$$

   For such a linear transformation T, we have

$$\begin{aligned} T(1) &= \tfrac{1}{2}(T(1+t) + T(1-t)) = \tfrac{1}{2}((1,2) + (1,1)) = (1, \tfrac{3}{2}). \\ T(t) &= T(1+t) - T(1) = (1,2) - (1, \tfrac{3}{2}) = (0, \tfrac{1}{2}). \\ T(t^2) &= T(1) - T(1-t^2) = (1, \tfrac{3}{2}) - (2,1) = (-1, \tfrac{1}{2}). \end{aligned}$$

   We are free to choose $T(t^3)$. For convenience, we choose $T(t^3) = (0,0)$. Then a required linear transformation is given by

$$\begin{aligned} T(a + bt + ct^2 + dt^3) &= aT(1) + bT(t) + cT(t^2) + dT(t^3) \\ &= (a - c, \tfrac{1}{2}(3a + b + c)). \end{aligned}$$

   Find another linear transformation satisfying the given conditions. □

Exercises for § 3.2

1. In each of the following, determine whether a linear transformation T exists. If such a linear transformation exists, then construct one.
   (a) $T:\mathbb{R}^2\to\mathbb{R}^2$; $T(1,1)=(1,-1)$, $T(0,1)=(-1,1)$, $T(2,-1)=(1,0)$.
   (b) $T:\mathbb{R}^2\to\mathbb{R}^3$; $T(1,1)=(1,0,2)$, $T(2,3)=(1,-1,4)$.
   (c) $T:\mathbb{R}^3\to\mathbb{R}^2$; $T(1,0,3)=(1,1)$, $T(-2,0,-6)=(2,1)$.
   (d) $T:\mathbb{R}^3\to\mathbb{R}^2$; $T(1,1,0)=(0,0)$, $T(0,1,1)=(1,1)$, $T(1,0,1)=(1,0)$.
   (e) $T:\mathbb{R}_3[t]\to\mathbb{R}$; $T(a+bt^2)=0$ for any $a,b\in\mathbb{R}$.
   (f) $T:\mathbb{R}_n[t]\to\mathbb{R}$; $T(p(x))=p(\alpha)$ for a fixed $\alpha\in\mathbb{R}$, and T is onto.
   (g) $T:C^1[0,1]\to\mathbb{R}$; $T(u)=\int_0^1(u(t))^2\,dt$.
   (h) $T:C^1[0,1]\to\mathbb{R}^2$; $T(u)=\left(\int_0^1 u(t)\,dt,\ u'(0)\right)$.
2. Can you construct a linear operator on $\mathbb{R}^2$ that maps the square with corners at $(-1,-1)$, $(1,-1)$, $(1,1)$ and $(-1,1)$ onto the square with corners at $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$?
3. Let V and W be real ips. Let $T:V\to W$ be a linear transformation. Prove that for all $x,y\in V$, $\langle Tx,Ty\rangle=\langle x,y\rangle$ iff $\|Tx\|=\|x\|$.

3.3 Range space and null space

Recall that a function $f:X\to Y$ is called a one-one (injective) function iff for all $w,x\in X$, $f(w)=f(x)$ implies that $w=x$. The function f is called an onto (surjective) function iff its range $\{f(x):x\in X\}$ is equal to its co-domain Y.

When a linear transformation is one-one, the zero vector is mapped to the zero vector only. Similarly, the range provides an answer to the question whether the linear transformation is an onto map.

Let $T:V\to W$ be a linear transformation. Then

$N(T)=\{v\in V:T(v)=0\}$ is called the **null space** of T;

$R(T)=\{T(v):v\in V\}$ is called the **range space** of T.

The null space of T is also called the *kernel* of T. We should justify as to why these two sets of vectors are called *spaces*.

***Theorem* 3.6**
Let $T:V\to W$ be a linear transformation. Then $N(T)$ is a subspace of V and $R(T)$ is a subspace of W.

Proof Since $T(0)=0$, $0\in N(T)$; so $N(T)$ is a nonempty subset of V. Let $u,v\in N(T)$ and let α be a scalar. Then $T(u)=T(v)=0$. Consequently,

$T(u+\alpha v) = T(u) + \alpha T(v) = 0$. That is, $u + \alpha v \in N(T)$. Therefore, $N(T)$ is a subspace of V.

Again, since $T(0) = 0$, $R(T)$ is a nonempty subset of W. Let $x, y \in R(T)$ and let β be a scalar. Then there exist $u, v \in V$ such that $x = T(u)$ and $y = T(v)$. Now, $x + \beta y = T(u) + \beta T(v) = T(u + \beta v)$. That is, $x + \beta y \in R(T)$. Therefore, $R(T)$ is a subspace of W. ∎

Since the null space and the range space of a linear transformation are vector spaces, they have some dimensions. We thus give names to these numbers.

Let $T : V \to W$ be a linear transformation. Then $\text{null}(T) := \dim(N(T))$ is called the **nullity** of T; and $\text{rank}(T) := \dim(R(T))$ is called the **rank** of T.

Example 3.7

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(a, b, c) = (a + b, a - c)$. To determine null($T$), suppose $T(a, b, c) = (0, 0)$, then $a = -b$ and $a = c$. Therefore

$$N(T) = \{(a, b, c) \in \mathbb{R}^3 : a = -b = c\} = \{(a, -a, a) : a \in \mathbb{R}\}.$$

A basis for $N(T)$ is $\{(1, -1, 1)\}$. Therefore, $\text{null}(T) = 1$.

Any vector in $R(T)$ is of the form $(a + b,\ a - c)$ for $a, b, c \in \mathbb{R}$. Since

$$(a + b, a - c) = a(1, 1) + b(1, 0) + c(0, -1),$$

$R(T) = \text{span}\{(1, 1), (1, 0), (0, -1)\}$. Now, $(1, 1) \in \text{span}\{(1, 0), (0, -1)\}$; and the set $\{(1, 0), (0, -1)\}$ is linearly independent. Therefore, a basis for $R(T)$ is $\{(1, 0), (0, -1)\}$. Consequently, $\text{rank}(T) = 2$. □

Since a linear transformation is completely determined by its action on a basis, the images of the basis vectors should span its range space. We show this formally in the following theorem.

***Theorem* 3.8**

Let $T : V \to W$ be a linear transformation. Let B be a basis of V. Then $R(T) = \text{span}\{T(v) : v \in B\}$.

Proof Let $w \in R(T)$. There exists $u \in V$ such that $T(u) = w$. Since B is a basis of V, there exist scalars $\alpha_1, \ldots, \alpha_n$ and vectors $v_1, \ldots, v_n \in B$ such that $u = \alpha_1 v + \cdots + \alpha_n v_n$. Then

$$\begin{aligned} w &= T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) \\ &\in \text{span}\{T(v_1), \ldots, T(v_n)\}. \end{aligned}$$

That is, $R(T) \subseteq \text{span}\{T(v) : v \in B\}$.

Conversely, for each $v \in B$, $T(v) \in R(T)$. As $R(T)$ is a vector space, we have $\text{span}\{T(v) : v \in B\} \subseteq R(T)$. ∎

As we see a linear transformation is one-one or onto can be characterized in terms of its nullity and rank.

***Theorem* 3.9**
Let $T : V \to W$ be a linear transformation.

(1) *T is one-one iff* $N(T) \subseteq \{0\}$ *iff* $N(T) = \{0\}$ *iff* $\mathrm{null}(T) = 0$.

(2) *T is an onto map iff* $W \subseteq \mathrm{span}\{Tv : v \in B\}$ *for any basis B of V iff* $\mathrm{rank}(T) = \dim(W)$.

Proof (1) Assume that T is one-one. If $x \in N(T)$, then $T(x) = 0$, which is equal to $T(0)$. Now, $T(x) = T(0)$ implies $x = 0$. That is, $N(T) \subseteq \{0\}$.

Conversely, suppose $N(T) \subseteq \{0\}$. Let $T(x) = T(y)$. Then $T(x-y) = 0$. So, $x - y \in N(T)$. That is, $x - y = 0$. Therefore, T is one-one.

Other equivalences are trivial.

(2) As $R(T) \subseteq W$, the linear transformation T is onto iff $W \subseteq R(T)$. Let B be any basis of V. From Theorem 3.8, it follows that $R(T) = \mathrm{span}\{Tv : v \in B\}$. Therefore, T is an onto map iff $W \subseteq \mathrm{span}\{Tv : v \in B\}$. The other equivalence is trivial. ∎

Look at Example 3.7. Does it suggest any connection between the rank and the nullity of a linear transformation?

***Theorem* 3.10 (Rank-nullity)**
Let V be a finite dimensional vector space. Let $T : V \to W$ be a linear transformation. Then $\mathrm{rank}(T) + \mathrm{null}(T) = \dim(V)$.

Proof If $T = 0$, the zero map, then $R(T) = \{0\}$ and $N(T) = V$. Clearly, the dimension formula holds. So, assume that T is a nonzero linear transformation. The null space $N(T)$ is a subspace of the finite dimensional vector space V. So, let $B = \{v_1, \ldots, v_k\}$ be a basis of $N(T)$. [It includes the case of $N(T) = \{0\}$. In this case, $B = \varnothing$; and we take $k = 0$, which is the number of vectors in B.] Extend B to a basis $B \cup \{w_1, \ldots, w_n\}$ for V. Let $E = \{T(w_1), \ldots, T(w_n)\}$. We show that E is a basis of $R(T)$.

By Theorem 3.8, $R(T) = \mathrm{span}\{T(v_1), \ldots, T(v_k), T(w_1), \ldots, T(w_n)\}$. Since $T(v_i) = 0$ for each $i \in \{1, \ldots, k\}$, $R(T) = \mathrm{span}\{T(w_1), \ldots, T(w_n)\} = \mathrm{span}(E)$.

For linear independence of E, let $b_1, \ldots, b_n$ be scalars such that

$$b_1 T(w_1) + \cdots + b_n T(w_n) = 0.$$

Then $T(b_1 w_1 + \cdots + b_n w_n) = 0$. So, $b_1 w_1 + \cdots + b_n w_n \in N(T)$. Since B is a basis of $N(T)$, there exist scalars $a_1, \ldots, a_k$ such that

$$b_1 w_1 + \cdots + b_n w_n = a_1 v_1 + \cdots + a_k v_k.$$

That is,

$$a_1v_1 + \cdots + a_kv_k - b_1w_1 - \cdots - b_nw_n = 0.$$

Since $B \cup \{w_1, \ldots, w_n\}$ is a basis of V, we have $a_1 = \cdots = a_k = b_1 = \cdots = b_n = 0$. Hence E is linearly independent.

Now that B is a basis for $N(T)$, E is a basis for $R(T)$, and $B \cup \{w_1, \ldots, w_n\}$ is a basis for V, we have $\mathrm{rank}(T) + \mathrm{null}(T) = n + k = \dim(V)$. ∎

Example 3.11

Define $T : \mathbb{R}_2[t] \to \mathbb{R}^4$ by $T(a + bt + ct^2) = (a - b,\ b - c,\ c + a,\ -2a)$. What are $\mathrm{rank}(T)$ and $\mathrm{null}(T)$?

We find that $R(T) = \{(a - b,\ b - c,\ c + a,\ -2a) : a, b, c \in \mathbb{R}\}$. Since

$$(a - b,\ b - c,\ c + a,\ -2a) = a(1, 0, 1, -2) + b(-1, 1, 0, 0) + c(0, -1, 1, 0),$$

the vectors $(1, 0, 1, -2)$, $(-1, 1, 0, 0)$ and $(0, -1, 1, 0)$ span $R(T)$.

Alternatively, we start with a basis of $\mathbb{R}_2[t]$, say $\{1, t, t^2\}$. Now,

$$T(1) = (1, 0, 1, -2),\ T(t) = (-1, 1, 0, 0),\ T(t^2) = (0, -1, 1, 0).$$

By Theorem 3.8, $R(T) = \mathrm{span}\{(1, 0, 1, -2),\ (-1, 1, 0, 0),\ (0, -1, 1, 0)\}$.

To check linear independence, suppose

$$a(1, 0, 1, -2) + b(-1, 1, 0, 0) + c(0, -1, 1, 0) = 0.$$

Then $(a - b,\ b - c,\ c + a,\ -2a) = (0, 0, 0, 0)$. Solving the ensuing linear equations $a - b = 0,\ b - c = 0,\ c + a = 0,\ -2a = 0$, we find that $a = b = c = 0$. So, $B = \{(1, 0, 1, -2),\ (-1, 1, 0, 0),\ (0, -1, 1, 0)\}$ is linearly independent. That is, B is a basis of $R(T)$. Hence, $\mathrm{rank}(T) = 3$.

To compute $\mathrm{null}(T)$, let $v = (a + bt + ct^2) \in N(T)$. Then

$$T(v) = (a - b,\ b - c,\ c + a,\ -2a) = (0, 0, 0, 0).$$

As earlier, it implies that $a = b = c = 0$. That is, $v = 0$. So, $N(T) = \{0\}$; and then $\mathrm{null}(T) = 0$.

Notice that using Rank-nullity theorem, we can conclude that $\mathrm{rank}(T) = \dim(\mathbb{R}_2[t]) - \mathrm{null}(T) = 3$. This is a shorter way to compute $\mathrm{rank}(T)$. □

Further Rank-nullity theorem implies the following:

> If $m < n$, then no linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ is onto.
> If $m > n$, then no linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ is one-one.

Exercises for § 3.3

1. Determine rank(T) and null(T) in each of the following cases:
   (a) $T : \mathbb{R}^2 \to \mathbb{R}^2$; $T(a,b) = (a-b, 2b)$.
   (b) $T : \mathbb{R}^3 \to \mathbb{R}^2$; $T(a,b,c) = (a-b, 2c)$.
   (c) $T : \mathbb{R}^2 \to \mathbb{R}^3$; $T(a,b) = (a+b, a-b, 0)$.
   (d) $T : \mathbb{R}^2 \to \mathbb{R}^3$; $T(a,b) = (a+b, 0, 2b-a)$.
   (e) $T : \mathbb{R}^2 \to \mathbb{R}^3$; $T(a,b) = (a+b, a+2b, 2a-2b)$.
   (f) $T : \mathbb{R}_2[t] \to \mathbb{R}_3[t]$; $T(p(t)) = t\,p(t) + p'(t)$.
2. Let V be the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$, having derivatives of all orders. Let $T : V \to V$ be the differential operator: $Tx = x'$. What is $N(T)$?
3. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(a,b,c) = (b+c,\ a+b+2c,\ a+c)$.
   (a) Determine the null space $N(T)$ and the range space $R(T)$.
   (b) Let $S = \{v \in \mathbb{R}^3 : T(v) = (1,2,3)\}$. Express the subset S of $\mathbb{R}^3$ in the form $S = \{u + x : x \in N(T)\}$ for a particular vector $u \in S$.
4. Let V and W be finite dimensional vector spaces. Let $T : V \to W$ be a linear transformation. Give reasons for the following:
   (a) rank$(T) \le \min\{\dim(V), \dim(W)\}$.
   (b) T is onto implies $\dim(W) \le \dim(V)$.
   (c) T is one-one implies $\dim(V) \le \dim(W)$.
   (d) $\dim(V) > \dim(W)$ implies T is not one-one.
   (e) $\dim(V) < \dim(W)$ implies T is not onto.
5. Give an example for each of the following:
   (a) A linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $N(T) = R(T)$.
   (b) Linear transformations $S \ne T$ with $N(S) = N(T)$ and $R(S) = R(T)$.
6. Let U be a subspace of a finite dimensional vector space V.
   (a) Does there exist a liner operator T on V such that $R(T) = U$?
   (b) Does there exist a liner operator T on V such that $N(T) = U$?
7. Let $T : V \to W$ be a linear transformation and let $\{v_1, \ldots, v_n\}$ be a basis of V. Show the following:
   (a) T is one-one iff $T(v_1), \ldots, T(v_n)$ are linearly independent.
   (b) T is onto iff span$\{T(v_1), \ldots, T(v_n)\} = W$.
   (c) T is one-one and onto iff the vectors $T(v_1), \ldots, T(v_n)$ form a basis for W.
8. Show that a subset $\{v_1, \ldots, v_n\}$ of a vector space V is linearly independent iff the map $f : \mathbb{F}^n \to V$ defined by $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = \alpha_1 v_1 + \cdots + \alpha_n v_n$ is one-one.
9. Let $T : V \to V$ be a linear operator such that $T^2 = T$. Let I denote the identity operator. Prove that $R(T) = N(I - T)$ and $N(T) = R(I - T)$.
10. Let $S : U \to V$ and $T : V \to W$ be linear transformations where U, V and W are of finite dimensions. Show that rank$(TS) \le \min\{\text{rank}(T),\ \text{rank}(S)\}$.

3.4 Isomorphisms

Recall that a function $f : X \to Y$ is one-one and onto iff there exists a unique function $g : Y \to X$ such that $g \circ f = I_X$ and $f \circ g = I_Y$. Here, the map I_X is the identity map on X and similarly, I_Y is the identity map on Y. In such a case, the function f is said to be invertible, and its inverse, which is the function g, is denoted by f^{-1}.

We give a name to a one-one onto linear transformation.

A one-one and onto linear transformation is called an **isomorphism**.

Two vector spaces are called **isomorphic** to each other iff there exists an isomorphism from one to the other.

Example 3.12

1. Define the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(a,b) = (a-b, 2a+b) \quad \text{for } (a,b) \in \mathbb{R}^2.$$

   If $(a,b) \in N(T)$, then $T(a,b) = (0,0)$. It leads to $a-b = 0 = 2a+b$. Thus, $a = b = 0$. That is, $N(T) \subseteq \{0\}$. Therefore T is one-one.

   By Rank-nullity theorem $\operatorname{rank}(T) = 2 - \operatorname{null}(T) = 2 - 0 = 2$. Thus $R(T) = \mathbb{R}^2$; consequently, T is an onto map.

   Therefore, T is an isomorphism.

2. $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(a_1, \ldots, a_n) = (a_1, a_1 + a_2, \ldots, a_1 + \cdots + a_n)$ is an isomorphism.
3. $T : \mathbb{R}_n[t] \to \mathbb{R}^{n+1}$ defined by $T(a_0 + a_1 t + \cdots + a_n t^n) = (a_0, \ldots, a_n)$ is an isomorphism.

□

***Theorem* 3.13**
The inverse of an isomorphism is also an isomorphism.

Proof Let $T : V \to W$ be an isomorphism. Since T is one-one and onto, its inverse T^{-1} exists, and it is also one-one and onto. We show that T^{-1} is a linear transformation. So, let $x, y \in W$ and let α be a scalar. Since T is onto, there exist $u, v \in V$ such that $T(u) = x$ and $T(v) = y$. Then $T(u + \alpha v) = T(u) + \alpha T(v) = x + \alpha y$. That is, $T^{-1}(x + \alpha y) = u + \alpha v = T^{-1}(x) + \alpha T^{-1}(y)$. Therefore, T^{-1} is a linear transformation. ∎

Each vector space is isomorphic to itself. Reason? The identity map is an isomorphism. If V is isomorphic to W, then the inverse of an isomorphism from V to W is an isomorphism from W to V. That is, W is also isomorphic to V. Further, Theorem 3.3 implies that a composition of isomorphisms is an isomorphism. Thus, if U is isomorphic to V and V is isomorphic to W, then U is also isomorphic to W. Therefore, 'is isomorphic to' is an equivalence relation on the set of all vector spaces. This is the reason we have been talking of "an isomorphism between spaces" and "two spaces being isomorphic".

The following result is a corollary to Rank-nullity theorem.

***Theorem* 3.14**
Let V and W be finite dimensional vector spaces with $\dim(V) = \dim(W)$. *Let $T : V \to W$ be a linear transformation. Then T is an isomorphism iff T is one-one iff T it is onto.*

Proof T is one-one iff $\text{null}(T) = 0$ iff $\text{rank}(T) = \dim(V)$ iff $\text{rank}(T) = \dim(W)$ iff T is onto. ∎

Example 3.15
Let $W = \{(a, b, c, d) \in \mathbb{R}^4 : a + b + c + d = 0\}$. Define $T : \mathbb{R}_2[t] \to W$ by

$$T(a + bt + ct^2) = (a - b,\ b - c,\ c + a,\ -2a).$$

The map T is well defined since $(a-b)+(b-c)+(c+a)+(-2a) = 0$ shows that $(a - b, b - c, c + a, -2a) \in W$. It is also easy to verify that T is a linear transformation. We look for $N(T)$ and $R(T)$.

As in Example 3.11, $N(T) \subseteq \{0\}$. So, $\text{null}(T) = 0$. Thus $\text{rank}(T) = 3$. Also, $\dim(W) = 3$, as $\{(1, -1, 0, 0),\ (1, 0, -1, 0),\ (1, 0, 0, -1)\}$ is a basis of W. Hence T is an isomorphism.

To see that $W = R(T)$ directly we may proceed as follows. We start with a basis of W, say, $\{(1, 0, 0, -1),\ (0, 1, 0, -1),\ (0, 0, 1, -1)\}$. Are these basis vectors in $R(T)$? For the first basis vector, we look for a polynomial $a + bt + ct^2$ such that $T(a + bt + ct^2) = (a - b,\ b - c,\ c + a,\ -2a) = (1, 0, 0, -1)$.

It leads to $a = \frac{1}{2},\ b = -\frac{1}{2},\ c = -\frac{1}{2}$. We verify.

$$T(\tfrac{1}{2} - \tfrac{1}{2}t - \tfrac{1}{2}t^2) = (\tfrac{1}{2} + \tfrac{1}{2}, -\tfrac{1}{2} + \tfrac{1}{2}, -\tfrac{1}{2} + \tfrac{1}{2}, -2 \times \tfrac{1}{2}) = (1, 0, 0, -1).$$

Proceeding similarly for the other two basis vectors, we see that

$$T(\tfrac{1}{2} + \tfrac{1}{2}t - \tfrac{1}{2}t^2) = (\tfrac{1}{2} - \tfrac{1}{2}, \tfrac{1}{2} + \tfrac{1}{2}, \tfrac{1}{2} - \tfrac{1}{2}, -2 \times \tfrac{1}{2}) = (0, 1, 0, -1).$$

$$T(\tfrac{1}{2} + \tfrac{1}{2}t + \tfrac{1}{2}t^2) = (\tfrac{1}{2} - \tfrac{1}{2}, \tfrac{1}{2} - \tfrac{1}{2}, \tfrac{1}{2} + \tfrac{1}{2}, -2 \times \tfrac{1}{2}) = (0, 0, 1, -1).$$

Therefore, $W = \text{span}\{(1, 0, 0, -1), (0, 1, 0, -1), (0, 0, 1, -1)\} \subseteq R(T) \subseteq W$. □

Clearly, a one-one linear transformation is an isomorphism from its domain space to its range space.

The dimensions of two vector spaces provide complete information as to when an isomorphism may exist from one to the other.

***Theorem* 3.16**

A vector space is isomorphic to a finite dimensional vector space iff both the spaces have equal dimensions.

Proof Let V and W be vector spaces, where V is of finite dimension. Let $T : V \to W$ be an isomorphism. Then $N(T) = \{0\}$ and $R(T) = W$. By Rank-nullity theorem, $\dim(V) = \text{null}(T) + \text{rank}(T) = 0 + \text{rank}(T) = \dim(W)$.

Conversely, suppose $\dim(V) = \dim(W) = n$. Let $B = \{v_1, \ldots, v_n\}$ be a basis of V and let $E = \{w_1, \ldots, w_n\}$ be a basis of W. Using Theorem 3.4, define the linear transformation $T : V \to W$ by $T(v_i) = w_i$ for $1 \le i \le n$.

Since B and E span V and W respectively, T is onto. By Theorem 3.14, T is an isomorphism. ∎

It thus follows that a finite dimensional vector space cannot be isomorphic to an infinite dimensional vector space.

Theorem 3.16 says that isomorphisms preserve dimension. The following result adds a little flavour to this; and it is quite useful. We state it informally; and convey its precise meaning in the proof.

***Theorem* 3.17**

Isomorphisms preserve rank and nullity of a linear transformation.

Proof Let $P : U \to V$, $T : V \to W$ and $Q : W \to X$ be linear transformations. Suppose P and Q are isomorphisms. Our goal is to show that

$$\text{rank}(QTP) = \text{rank}(T), \quad \text{null}(QTP) = \text{null}(T). \tag{3.1}$$

Notice that $TP : U \to W$ and $QT : V \to X$ are linear transformations. We first prove the following:

1. $R(TP) = R(T)$;
2. $N(QT) = N(T)$;
3. $R(QT)$ is isomorphic to $R(T)$; and
4. $N(TP)$ is isomorphic to $N(T)$.

(1) Let $x \in R(TP)$. There exists $u \in U$ such that $x = T(P(u))$. Clearly, $x \in R(T)$. Conversely, let $y \in R(T)$. There exists $v \in V$ such that $y = T(v)$. Then $y = T(v) = (TP)(P^{-1}(v)) \in R(TP)$. Hence $R(TP) = R(T)$.

(2) Let $u \in N(QT)$. Then $QT(u) = 0$ implies $T(u) = Q^{-1}(QT(u)) = 0$. That is, $u \in N(T)$. Conversely, let $z \in N(T)$. Then $T(z) = 0$. Clearly, $Q(T(z)) = 0$. So, $z \in N(QT)$. Therefore, $N(QT) = N(T)$.

(3) If $w \in R(T)$, then $w = Tv$ for some $v \in V$. Thus $Q(w) = (QT)(v)$ says that $Q(w) \in R(QT)$. Using this we define the map $A : R(T) \to R(QT)$ by $A(w) = Q(w)$ for $w \in R(T)$. Notice that A is the same map as Q with its domain restricted to the subspace $R(T)$ of W. Thus A is a one-one linear

transformation. Each vector in $R(QT)$ is in the form $Q(w)$ for some $w \in R(T)$. Hence, $R(A) = R(QT)$. Thus, A is an onto map. Therefore, $R(T)$ and $R(QT)$ are isomorphic.

(4) If $u \in N(TP)$, then $T(P(u)) = 0$. Thus, $P(u) \in N(T)$. Using this we define the map $B : N(TP) \to N(T)$ by $B(u) = P(u)$ for $u \in N(TP)$. Notice that B is the restriction of the isomorphism P with its domain as the subspace $N(TP)$ of U. Thus B is a one-one linear transformation. Now, if $v \in N(T)$, then $T(v) = 0$ implies that $(TP)(P^{-1}(v)) = 0$. That is, $P^{-1}(v) \in N(TP)$. Then, $v = P(P^{-1}(v)) = B(P^{-1}(v))$ shows that each vector in $N(T)$ is in the form $B(u)$ for some $u \in N(TP)$. That is, B is an onto map. We conclude that B is an isomorphism from $N(TP)$ to $N(T)$. Consequently, $N(TP)$ and $N(T)$ are isomorphic.

From (1) and (3) it follows that $\text{rank}(TP) = \text{rank}(T) = \text{rank}(QT)$. Similarly, from (2) and (4) it follows that $\text{null}(QT) = \text{null}(T) = \text{null}(TP)$. Now, (3.1) follows. ∎

Observe that in Theorem 3.17, if all the vector spaces are finite dimensional, then Rank-nullity theorem and (1)-(2) would complete its proof. Also, in such a case, rank is preserved iff nullity is preserved.

For finite dimensional spaces, the following result provides a converse to Theorem 3.17.

***Theorem* 3.18**
Let U, V, W and X be finite dimensional vector spaces with $\dim(U) = \dim(V)$ and $\dim(W) = \dim(X)$. Let $T : V \to W$ and $S : U \to X$ be linear transformations. If $\text{rank}(S) = \text{rank}(T)$, then there exist isomorphisms $P : U \to V$ and $Q : W \to X$ such that $S = QTP$.

The conditions $\dim(U) = \dim(V)$ and $\dim(W) = \dim(X)$ imply that there exist isomorphisms $P : U \to V$ and $Q : W \to X$. However, the composition formula $S = QTP$ may not hold, in general. The theorem says that such a composition formula holds for some isomorphisms P and Q when $\text{rank}(S) = \text{rank}(T)$.

Proof Suppose that $\dim(U) = \dim(V) = n$, $\dim(W) = \dim(X) = m$, and $\text{rank}(T) = \text{rank}(S) = r$. Then $\text{null}(T) = \text{null}(S) = n - r$.

Choose a basis $\{v_1, \ldots, v_{n-r}\}$ for $N(T)$, which is a subspace of V. Extend this basis to $\{v_1, \ldots, v_{n-r}, \ldots, v_n\}$ for V. Similarly, choose a basis $\{u_1, \ldots, u_{n-r}\}$ for $N(S)$, which is a subspace of U. Extend this to a basis $\{u_1, \ldots, u_{n-r}, \ldots, u_n\}$ for U. Then define the vectors

$$w_1 = Tv_{n-r+1}, \ldots, w_r = Tv_n; \quad x_1 = Su_{n-r+1}, \ldots, x_r = Su_n.$$

As in the proof of Rank-nullity theorem, $\{w_1, \ldots, w_r\}$ is a basis for $R(T)$, and

$\{x_1, \ldots, x_r\}$ is a basis for $R(S)$. Then extend these to bases $\{w_1, \ldots, w_r, \ldots, w_m\}$ for W, and $\{x_1, \ldots, x_r, \ldots, x_m\}$ for X. Define the linear transformations $P : U \to V$ and $Q : W \to X$ by specifying their actions on the bases of U and W as in the following:

$$P(u_1) = v_1, \ldots, P(u_n) = v_n; \quad Q(w_1) = x_1, \ldots, Q(w_n) = x_n.$$

Since each basis vector v_i of V is in $R(P)$, $V \subseteq R(P)$; that is, P is onto. By Theorem 3.14, P is an isomorphism. Similarly, Q is also an isomorphism.

For $1 \leq i \leq n-r$, $Q(T(P(u_i))) = Q(T(v_i)) = Q(0) = 0 = S(u_i)$.

For $n-r < i \leq n$, $Q(T(P(u_i))) = Q(T(v_i)) = Q(w_i) = x_i = S(u_i)$.

Since the linear transformations QTP and S act the same way on each of the basis vectors $u_1, \ldots, u_n$ of U, we conclude that $QTP = S$. ∎

Exercises for § 3.4

1. Let $\{u_1, \ldots, u_n\}$ be an ordered basis of a vector space V. Prove that the linear transformation $T : V \to \mathbb{F}^n$ given by $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = (\alpha_1, \ldots, \alpha_n)$ is an isomorphism.
2. Find an isomorphism between $\mathbb{F}^{n+1}$ and $\mathbb{F}_n[t]$ other than that in Example 3.12.
3. Show that a linear transformation $T : V \to W$ is an isomorphism iff for every basis $\{v_1, \ldots, v_n\}$ of V, the vectors $T(v_1), \ldots, T(v_n)$ form a basis of W.
4. Let $T : V \to \mathbb{R}$ be a nonzero linear transformation, where $V \neq \{0\}$. Prove or disprove: T is onto iff $\mathrm{null}(T) = \dim(V) - 1$.
5. Let V and W be finite dimensional vector spaces, and let $S, T : V \to W$ be linear transformations. Show that $\mathrm{rank}(S) = \mathrm{rank}(T)$ iff there exist isomorphisms $P : V \to V$ and $Q : W \to W$ such that $S = QTP$.
6. Let $T : V \to \mathbb{F}^n$ be an isomorphism. Let $\langle\, , \rangle$ be the standard inner product on $\mathbb{F}^n$. Show that $\langle x, y\rangle_T := \langle T(x), T(y)\rangle$ for all $x, y \in V$, defines an inner product on V.
7. Let $T : V \to W$ and $S : W \to V$ be linear transformations, where V and W are finite dimensional vector spaces. Show the following:
   (a) If $\dim(V) = \dim(W)$, then $ST = I_V$ implies that $TS = I_W$.
   (b) If $\dim(V) \neq \dim(W)$, then $ST = I_V$ does not necessarily imply that $TS = I_W$.
8. Give an example of an infinite dimensional vector space V and linear operators S, T on V such that $ST = I$ but $TS \neq I$.

3.5 Adjoint of a Linear Transformation

Recall that for ease in reading we write $T(x)$ as Tx and $S \circ T$ as ST. We also write inner products on different spaces using the same notation $\langle\ ,\ \rangle$.

In inner product spaces, equality of two linear transformations can be shown by using the inner products.

***Theorem* 3.19**
Let V be a vector space, W be an inner product space, and let $S, T : V \to W$ be linear transformations. Then

$$S = T \ \textit{ iff } \ \langle Sv, w\rangle = \langle Tv, w\rangle \textit{ for all } v \in V,\ w \in W.$$

Proof If $S = T$, then $\langle Sv, w\rangle = \langle Tv, w\rangle$ for all $v \in V,\ w \in W$. Conversely, suppose $\langle Sv, w\rangle = \langle Tv, w\rangle$ for all $v \in V,\ w \in W$. In particular, with $w = (S-T)v$, we obtain $\langle (S-T)v, (S-T)v\rangle = 0$. It implies that for each $v \in V$, $(S-T)v = 0$. That is, $S = T$. ∎

Linear transformations on finite dimensional inner product spaces give rise to new linear transformations that work backward. This can be explicitly exhibited in the presence of an orthonormal basis for the domain space.

***Theorem* 3.20**
Let V and W be inner product spaces, where V is of finite dimension. Then, corresponding to each linear transformation $T : V \to W$, there exists a unique linear transformation $S : W \to V$ such that

$$\langle Tv, w\rangle = \langle v, Sw\rangle \quad \textit{for all } \ v \in V,\ w \in W.$$

Proof Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of V. Let $v \in V$. By Fourier expansion, we have $v = \sum_{i=1}^{n} \langle v, v_i\rangle v_i$. Then $Tv = \sum_{i=1}^{n} \langle v, v_i\rangle Tv_i$.

Let $w \in W$. We compute $\langle Tv, w\rangle$.

$$\begin{aligned} \langle Tv,\ w\rangle &= \sum_{i=1}^{n} \langle v, v_i\rangle\langle Tv_i, w\rangle = \sum_{i=1}^{n} \langle Tv_i, w\rangle\langle v, v_i\rangle \\ &= \sum_{i=1}^{n} \langle v,\ \overline{\langle Tv_i, w\rangle}\, v_i\rangle = \Big\langle v,\ \sum_{i=1}^{n} \langle w, Tv_i\rangle\, v_i \Big\rangle. \end{aligned} \tag{3.2}$$

This suggests that we define the map $S : W \to V$ by

$$S(w) = \sum_{i=1}^{n} \langle w, Tv_i\rangle\, v_i \quad \text{for each } w \in W.$$

Is S a linear transformation? Let $x, y \in W$ and let α be a scalar. Now,

$$\begin{aligned} S(x+\alpha y) &= \sum_{i=1}^{n} \langle x+\alpha y, Tv_i\rangle v_i = \sum_{i=1}^{n} (\langle x, Tv_i\rangle + \alpha\langle y, Tv_i\rangle) v_i \\ &= \sum_{i=1}^{n} \langle x, Tv_i\rangle v_i + \alpha \sum_{i=1}^{n} \langle y, Tv_i\rangle v_i = S(x) + \alpha S(y). \end{aligned}$$

Therefore, S is a linear transformation. Moreover, (3.2) says that

$$\langle Tv, w\rangle = \langle v, Sw\rangle \quad \text{for all } v \in V,\ w \in W.$$

For uniqueness of S, let $Q : W \to V$ be a linear transformation such that

$$\langle Tv, w\rangle = \langle v, Qw\rangle \quad \text{for all } v \in V,\ w \in W.$$

Then for all $v \in V$ and $w \in W$, we obtain

$$\langle Qw, v\rangle = \overline{\langle Tv, w\rangle} = \overline{\langle v, Sw\rangle} = \langle Sw, v\rangle.$$

By Theorem 3.19, we conclude that $Q = S$. ∎

In view of Theorem 3.20, we give a name to the linear transformation S that corresponds to T.

Let V and W be inner product spaces, where V is of finite dimension. Let $T : V \to W$ be a linear transformation. The linear transformation $T^* : W \to V$ defined by

$$\langle Tx, y\rangle = \langle x, T^* y\rangle \quad \text{for all } x \in V,\ y \in W$$

is called the **adjoint of** T.

The proof of Theorem 3.20 supplies an explicit representation of the adjoint in the presence of an orthonormal basis. If $\{v_1, \ldots, v_n\}$ is an orthonormal basis of V, and $T : V \to W$ is a linear transformation, then $T^* : W \to V$ is given by

$$T^*(w) = \sum_{i=1}^{n} \langle w, T(v_i)\rangle v_i \quad \text{for each } w \in W. \tag{3.3}$$

Notice that in (3.3), the inner product is the inner product of W.

Example 3.21

Consider the spaces $\mathbb{R}^3$ and $\mathbb{R}^4$ with their standard bases and the standard inner products. Define the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ by

$$T(a, b, c, d) = (a + c,\ b - 2c + d,\ a - b + c - d).$$

For obtaining $T^* : \mathbb{R}^3 \to \mathbb{R}^4$, we proceed as follows:

$$\begin{aligned}\big\langle (a,b,c,d), T^*(\alpha,\beta,\gamma)\big\rangle &= \big\langle T(a,b,c,d),(\alpha,\beta,\gamma)\big\rangle \\ &= \langle (a+c, b-2c+d, a-b+c-d), (\alpha,\beta,\gamma)\rangle \\ &= (a+c)\alpha + (b-2c+d)\beta + (a-b+c-d)\gamma \\ &= a(\alpha+\gamma) + b(\beta-\gamma) + c(\alpha-2\beta+\gamma) + d(\beta-\gamma) \\ &= \big\langle (a,b,c,d), (\alpha+\gamma, \beta-\gamma, \alpha-2\beta+\gamma, \beta-\gamma)\big\rangle.\end{aligned}$$

Therefore, $T^*(\alpha,\beta,\gamma) = (\alpha+\gamma,\ \beta-\gamma,\ \alpha-2\beta+\gamma,\ \beta-\gamma)$.

Alternatively, using the standard basis as the orthonormal basis for $\mathbb{R}^4$, we obtain

$$T(e_1) = (1,0,1),\ T(e_2) = (0,1,-1),\ T(e_3) = (1,-2,1),\ T(e_4) = (0,1,-1).$$

Then, for each $w = (\alpha,\beta,\gamma) \in \mathbb{R}^3$, we have

$$\begin{aligned}T^*(w) &= \sum_{i=1}^{4} \langle w, T(e_i)\rangle e_i \\ &= (\alpha+\gamma)e_1 + (\beta-\gamma)e_2 + (\alpha-2\beta+\gamma)e_3 + (\beta-\gamma)e_4 \\ &= (\alpha+\gamma,\ \beta-\gamma,\ \alpha-2\beta+\gamma,\ \beta-\gamma).\end{aligned}$$

□

In the following theorems, we prove some useful facts about the adjoint.

***Theorem* 3.22**
Let U, V and W be finite dimensional inner product spaces. Let $S : U \to V$ and $T, T_1, T_2 : V \to W$ be linear transformations. Let $I : V \to V$ be the identity operator and let $\alpha \in \mathbb{F}$. Then

$$(T_1+T_2)^* = T_1^* + T_2^*,\ (\alpha T)^* = \overline{\alpha} T^*,\ (T^*)^* = T,\ I^* = I,\ (TS)^* = S^*T^*.$$

Proof $\langle x, (T_1+T_2)^* y\rangle = \langle (T_1+T_2)x, y\rangle = \langle T_1 x, y\rangle + \langle T_2 x, y\rangle = \langle x, T_1^* y\rangle + \langle x, T_2^* y\rangle = \langle x, T_1^* y + T_2^* y\rangle$. Therefore, $(T_1+T_2)^* = T_1^* + T_2^*$. Other equalities are proved similarly. ∎

***Theorem* 3.23**
Let V and W be finite dimensional inner product spaces, and let $T : V \to W$ be a linear transformation. Then the following are true:

(1) $N(T^*T) = N(T),\ \ N(TT^*) = N(T^*)$.

(2) $\operatorname{rank}(T^*) = \operatorname{rank}(T^*T) = \operatorname{rank}(TT^*) = \operatorname{rank}(T)$.

(3) $R(T^*T) = R(T^*),\ \ R(TT^*) = R(T)$.

(4) *If* $\dim(V) = \dim(W)$, *then* $\operatorname{null}(T^*) = \operatorname{null}(T^*T) = \operatorname{null}(TT^*) = \operatorname{null}(T)$.

Proof (1) Let $x \in N(T)$. Then $Tx = 0$; so, $T^*Tx = 0$. That is, $x \in N(T^*T)$. Conversely, if $y \in N(T^*T)$, then $T^*Ty = 0$. Taking inner product with y, we have $\langle y, T^*Ty\rangle = 0$. It follows that $\langle Ty, Ty\rangle = 0$. That is, $Ty = 0$. So, $y \in N(T)$. Therefore, $N(T^*T) = N(T)$. Replacing T with T^*, and using the fact that $(T^*)^* = T$, we have $N(TT^*) = N(T^*)$.

(2) From (1) it follows that $\text{null}(T^*T) = \text{null}(T)$. Then Rank-nullity theorem implies that $\text{rank}(T^*T) = \text{rank}(T)$. Similarly, $\text{rank}(TT^*) = \text{rank}(T^*)$.

Now, if $y \in R(T^*T)$, then $y = T^*(Tx)$ for some $x \in V$. That is, $y \in R(T^*)$. So, $R(T^*T)$ is a subspace of $R(T^*)$.

Therefore, $\text{rank}(T) = \text{rank}(T^*T) \le \text{rank}(T^*)$. Similarly, with T^* instead of T, we obtain $\text{rank}(T^*) = \text{rank}(TT^*) \le \text{rank}(T)$. Combining these, we get the required equalities.

(3) We have already seen that $R(T^*T)$ is a subspace of $R(T^*)$; and (2) says that their dimensions are equal. Therefore, $R(T^*T) = R(T^*)$.

With T^* instead of T, we obtain $R(TT^*) = R(T)$.

(4) We have $\text{null}(T^*) = \dim(W) - \text{rank}(T^*)$, $\text{null}(T) = \dim(V) - \text{rank}(T)$, and $\dim(V) = \dim(W)$. Then (2) implies $\text{null}(T^*) = \text{null}(T)$. Now, other equalities follow from (1). ∎

Basing on the adjoint, special types of linear operators on a finite dimensional ips can be defined.

A linear operator T on a finite dimensional ips V is called

1. **self-adjoint** iff $T^* = T$;
2. **normal** iff $T^*T = TT^*$;
3. **unitary** iff $T^*T = TT^* = I$;
4. **orthogonal** iff $T^*T = TT^* = I$ and V is a real ips;
5. **isometric** iff $\|Tx\| = \|x\|$ for each $x \in V$.

We will come across these types of linear operators at various places. For now, we observe that each self-adjoint linear operator is normal, and each unitary linear operator is normal and invertible. Similarly, it can be shown that a linear operator on a finite dimensional ips is unitary iff it is isometric.

Exercises for § 3.5

1. Fix a vector u in an ips V. Define a linear functional T on V by $Tv = \langle v, u\rangle$ for $v \in V$. What is $T^*(\alpha)$ for a scalar α?
2. Define an operator T on $\mathbb{F}^n$ by $T(a_1, a_2, \ldots, a_n) = (0, a_1, \ldots, a_{n-1})$. What is $T^*(a_1, \ldots, a_n)$?
3. Prove that a linear operator T on a finite dimensional ips is invertible iff T^* is invertible. In that case, show that $(T^*)^{-1} = (T^{-1})^*$.

4. For any subset S of an ips V, write $S^\perp = \{x \in V : x \perp u \text{ for each } u \in S\}$. Prove that if U is a subspace of a finite dimensional ips V, then

   (a) $V = U \oplus U^\perp$. (b) $(U^\perp)^\perp = U$.

5. *Fundamental subspaces*: Let $T : V \to W$ be a linear operator, where V and W are finite dimensional ips. Prove the following:

   (a) $N(T) = R(T^*)^\perp$. (b) $N(T^*) = R(T)^\perp$.
   (c) $R(T) = N(T^*)^\perp$. (d) $R(T^*) = N(T)^\perp$.
   (e) $V = N(T) \oplus N(T)^\perp = R(T) \oplus R(T)^\perp = N(T^*) \oplus N(T^*)^\perp = R(T^*) \oplus R(T^*)^\perp$.

6. Give an example of a linear operator which is

   (a) normal but not self-adjoint. (b) normal but not unitary.
   (c) unitary but not self-adjoint. (d) self-adjoint but not unitary.

7. Show that on a finite dimensional ips, a linear operator is unitary iff it is isometric.

8. Define $f : \mathbb{C}^3 \to \mathbb{C}$ by $f(a,b,c) = (a+b+c)/3$. Find a vector $y \in \mathbb{C}^3$ such that $f(x) = \langle x, y\rangle$ for each $x \in \mathbb{C}^3$.

9. (*Riesz Representation*) Let $f : V \to \mathbb{F}$ be a linear functional on a finite dimensional ips V. Prove that there exists a unique $y \in V$ such that $f(x) = \langle x, y\rangle$ for all $x \in V$. [Hint: If $\{v_1, \ldots, v_n\}$ is an orthonormal basis of V, then $y = \sum_{i=1}^n \overline{f(v_i)}\, v_i$.]

10. Let T be a linear operator on a finite dimensional ips. Using Theorem 3.23 show that if $T^*T = I$, then $TT^* = I$.

4

Matrices

4.1 Matrix of a linear transformation

Let V be a vector space of dimension n over $\mathbb{F}$. As dimension of $\mathbb{F}^{n\times 1}$ is also n, V and $\mathbb{F}^{n\times 1}$ are isomorphic. In fact, if $B = \{v_1, \ldots, v_n\}$ is an ordered basis of V, and $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{F}^{n\times 1}$, then the map that takes v_i to e_i provides the required natural isomorphism. If $v \in V$ is any vector, then there exist unique scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that $v = a_1 v_1 + \cdots + a_n v_n$. Then this natural isomorphism from V to $\mathbb{F}^{n\times 1}$ is given by

$$v \mapsto [v]_B := (a_1, \ldots, a_n)^{\mathrm{t}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

The vector $[v]_B$ is called the **coordinate vector of** v **with respect to** B. The coordinate vector is well-defined since B is an ordered basis for V ensures that the scalars $a_1, \ldots, a_n$ are uniquely determined from v.

We write the map $v \mapsto [v]_B$ as $[\ ]_B : V \to \mathbb{F}^{n\times 1}$, and $[\ ]_B(v)$ as $[v]_B$. Also, we call the map $[\ ]_B$ as the **coordinate vector map** with respect to the ordered basis B of V.

Example 4.1

1. Consider the ordered basis $B = \{(1,-1), (1,0)\}$ for $\mathbb{F}^2$. Then
   $(0,1) = -1(1,-1) + 1(1,0)$. Thus $[(0,1)]_B = (-1,1)^{\mathrm{t}}$.
2. Consider the ordered basis $B = \{(1,0), (1,-1)\}$ for $\mathbb{F}^2$. Then
   $(0,1) = 1(1,0) + -1(1,-1)$. Thus $[(0,1)]_B = (1,-1)^{\mathrm{t}}$.
3. Consider the ordered basis $B = \{1, 1+t, 1+t^2\}$ for $\mathbb{F}_2[t]$. Then
   $1+t+t^2 = -1(1) + 1(1+t) + 1(1+t^2)$. Thus $[1+t+t^2]_B = (-1,1,1)^{\mathrm{t}}$. □

The coordinate vectors would be possibly different if we alter the positions of the basis vectors in the ordered basis. Further, we speak of a coordinate vector map with respect to a given ordered basis of a finite dimensional vector space.

***Theorem* 4.2**
Each coordinate vector map is an isomorphism.

Proof Let V be a vector space with an ordered basis $\{v_1,\ldots,v_n\}$. Let $x,y\in V$. There exist unique scalars $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ such that

$$x=\alpha_1v_1+\cdots+\alpha_nv_n,\quad y=\beta_1v_1+\cdots\beta_nv_n.$$

Then $[x]_B=(\alpha_1,\ldots,\alpha_n)^t$ and $[y]_B=(\beta_1,\ldots,\beta_n)^t$. Thus, for any $c\in\mathbb{F}$,

$$\begin{aligned}[x+cy]_B &= [(\alpha_1+c\beta_1)v_1+\cdots+(\alpha_n+c\beta_n)v_n]_B\\ &= (\alpha_1+c\beta_1,\ldots,\alpha_n+c\beta_n)^t=[x]_B+c[y]_B.\end{aligned}$$

Thus the map $[\ ]_B$ is a linear transformation from V to $\mathbb{F}^{n\times1}$.

Let $v\in N([\ ]_B)$. Then $[v]_B=0\in\mathbb{F}^{n\times1}$. Thus $v=0v_1+\cdots+0v_n=0$. That is, $N([\ ]_B)\subseteq\{0\}$. Therefore, $[\ ]_B$ is one-one. By Theorem 3.14, $[\ ]_B$ is an isomorphism. ∎

Thus we say that an ordered basis of an n-dimensional vector space V brings in a coordinate system by inducing an isomorphism between V and $\mathbb{F}^{n\times1}$.

Let V and W be finite dimensional vector spaces over $\mathbb{F}$ with respective ordered bases as $B=\{v_1,\ldots v_n\}$ and $E=\{w_1,\ldots,w_m\}$. Let $T:V\to W$ be a linear transformation. Let $v\in V$. We have unique scalars $\alpha_1,\ldots,\alpha_n$ such that $v=\alpha_1v_1+\cdots+\alpha_nv_n$. Then $T(v)=\alpha_1T(v_1)+\cdots+\alpha_nT(v_n)$. It follows that

$$[T(v)]_E=\alpha_1[T(v_1)]_E+\cdots+\alpha_n[T(v_n)]_E.$$

That is, the coordinate vector of $T(v)$ can be obtained once we know the coordinate vectors of $T(v_1),\ldots,T(v_n)$ with respect to E. Suppose the coordinate vectors of $T(v_j)$ are given as follows:

$$[T(v_1)]_E=\begin{bmatrix}a_{11}\\ \vdots\\ a_{m1}\end{bmatrix},\ \cdots,\ [T(v_j)]_E=\begin{bmatrix}a_{1j}\\ \vdots\\ a_{mj}\end{bmatrix},\ \cdots,\ [T(v_n)]_E=\begin{bmatrix}a_{1n}\\ \vdots\\ a_{mn}\end{bmatrix}$$

for scalars a_{ij}. Notice that this is equivalent to expressing $T(v_j)$ in terms of the basis vectors $w_1,\ldots,w_m$ as in the following:

$$\begin{aligned}T(v_1) &= a_{11}w_1+\cdots+a_{m1}w_m\\ &\ \ \vdots\\ T(v_n) &= a_{1n}w_1+\cdots+a_{mn}w_m.\end{aligned}$$

We put together the coordinate vectors $[T(v_j)]_E$ in that order to obtain the following array of scalars a_{ij} :

$$\Big[\, [T(v_1)]_E \cdots [T(v_j)]_E \cdots [T(v_n)]_E \,\Big] = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ & & \vdots & & \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ & & \vdots & & \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}.$$

Such an array of mn scalars a_{ij} is called an $m \times n$ **matrix** with entries from $\mathbb{F}$. The set of all $m \times n$ matrices with entries from $\mathbb{F}$ is denoted by $\mathbb{F}^{m\times n}$. The scalar at the ith row and the jth column of a matrix is called its (i, j)th **entry**. By writing

$$A = [a_{ij}] \in \mathbb{F}^{m\times n}$$

we mean that A is an $m \times n$ matrix with its (i, j)th entry as a_{ij}.

Two matrices in $\mathbb{F}^{m\times n}$ are equal iff their corresponding entries are equal.

We summarize and give a notation. Let V and W be vector spaces over $\mathbb{F}$ with dimensions n and m, and ordered bases $B = \{v_1, \ldots v_n\}$ and $E = \{w_1, \ldots, w_m\}$, respectively. Let $T : V \to W$ be a linear transformation. Let $[T(v_j)]_E$ denote the coordinate vector of $T(v_j)$ with respect to E. The **matrix** $[T]_{E,B}$ **of** T **with respect to the ordered bases** B **and** E is given by

$$[T]_{E,B} := \Big[[T(v_1)]_E \ \ldots \ [T(v_n)]_E \Big].$$

Now we know how to construct the matrix of a linear transformation. We first express the images of the basis vectors of the domain space in terms of the basis vectors of the co-domain space as in the following:

$$\begin{aligned} T(v_1) &= a_{11}w_1 + \cdots + a_{i1}w_i + \cdots + a_{m1}w_m \\ &\ \ \vdots \\ T(v_j) &= a_{1j}w_1 + \cdots + a_{ij}w_i + \cdots a_{mj}w_m \\ &\ \ \vdots \\ T(v_n) &= a_{1n}w_1 + \cdots + a_{in}w_i + \cdots a_{mn}w_m. \end{aligned}$$

Then

$$[T]_{E,B} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ & & \vdots & & \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ & & \vdots & & \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}.$$

Caution: Mark which a_{ij} goes where.

Example 4.3

In the following, we consider all bases as ordered bases.

1. Let $B = \{e_1, e_2\}$ and $E = \{f_1, f_2, f_3\}$ be the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(a,b) = (2a-b, a+b, b-a)$. Then

$$\begin{aligned} T(e_1) &= (2,1,-1) &= 2f_1 + 1f_2 + (-1)f_3 \\ T(e_2) &= (-1,1,1) &= (-1)f_1 + 1f_2 + 1f_3. \end{aligned}$$

   Therefore $[T]_{E,B} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$.

   Further, $[(a,b)]_B = (a,b)^{\mathrm{t}}$ and $[T(a,b)]_E = (2a-b, a+b, -a+b)^{\mathrm{t}}$.

2. Consider the linear transformation $D : \mathbb{R}_3[t] \to \mathbb{R}_2[t]$; $D(p(t)) = p'(t)$. Choose bases $B = \{1, t, t^2, t^3\}$ and $E = \{1, t, t^2\}$ for $\mathbb{R}_3[t]$ and $\mathbb{R}_2[t]$, respectively. Then

$$\begin{aligned} D(1) &= 0\times 1 + 0\times t + 0\times t^2 \\ D(t) &= 1\times 1 + 0\times t + 0\times t^2 \\ D(t^2) &= 0\times 1 + 2\times t + 0\times t^2 \\ D(t^3) &= 0\times 1 + 0\times t + 3\times t^2. \end{aligned}$$

   Therefore, $[D]_{E,B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.

   Here, $[a + bt + ct^2 + dt^3]_B = (a,b,c,d)^{\mathrm{t}}$ and $D(a + bt + ct^2 + dt^3)]_E = (b, 2c, 3d)^{\mathrm{t}}$.

3. With the same linear transformation D and the basis B for $\mathbb{R}_3[t]$ as in (2), let $E = \{1, 1+t, 1+t^2\}$. Then

$$\begin{aligned} D(1) &= 0\times 1 + 0\times (1+t) + 0\times (1+t^2) \\ D(t) &= 1\times 1 + 0\times (1+t) + 0\times (1+t^2) \\ D(t^2) &= -2\times 1 + 2\times (1+t) + 0\times (1+t^2) \\ D(t^3) &= -3\times 1 + 0\times (1+t) + 3\times (1+t^2). \end{aligned}$$

   Therefore, $[D]_{E,B} = \begin{bmatrix} 0 & 1 & -2 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.

   Observe that

$$\begin{aligned} D(a + bt + ct^2 + dt^3) &= b + 2ct + 3dt^2 \\ &= (b - 2c - 3d)\times 1 + 2c(1+t^2) + 3d(1+t^2). \end{aligned}$$

So, $[a+bt+ct^2+dt^3)]_B = (a,b,c,d)^{\rm t}$, and $[D(a+bt+ct^2+dt^3)]_E = (b-2c-3d, 2c, 3d)^{\rm t}$.

4. Let $B = \{1, 1+t, t+t^2\}$ and $E = \{1, t, t+t^2, t^2+t^3\}$ be bases for $\mathbb{R}_2[t]$ and $\mathbb{R}_3[t]$, respectively. Let $T : \mathbb{R}_2[t] \to \mathbb{R}_3[t]$ be the linear transformation given by $T(p(t)) = \int_0^t p(s)ds$. Then

$$\begin{aligned}
T(1) = \quad & \int_0^t ds & = 0\times 1 + 1\times t + 0(t+t^2) + 0(t^2+t^3)\\
T(1+t) = \quad & \int_0^t (1+s)\,ds & = 0\times 1 + \tfrac{1}{2}\times t + \tfrac{1}{2}(t+t^2) + 0(t^2+t^3)\\
T(t+t^2) = \quad & \int_0^t (s+s^2)\,ds & = 0\times 1 - \tfrac{1}{6}\times t + \tfrac{1}{6}(t+t^2) + \tfrac{1}{3}(t^2+t^3).
\end{aligned}$$

Therefore, $[T]_{E,B} = \begin{bmatrix} 0 & 0 & 0\\ 1 & 1/2 & -1/6\\ 0 & 1/2 & 1/6\\ 0 & 0 & 1/3 \end{bmatrix}$.

The coordinate vectors of a typical vector in $\mathbb{R}_2[t]$ and its image are:

$$\begin{aligned}
v &= a+bt+ct^2 = (a-b+c)\times 1 + (b-c)(1+t) + c(t+t^2)\\
T(v) &= T(a+bt+ct^2) = \int_0^t (a+bs+cs^2)\,ds = at + \tfrac{b}{2}t^2 + \tfrac{c}{3}t^3\\
&= 0\times 1 + (a - \tfrac{b}{2} + \tfrac{c}{3})t + (\tfrac{b}{2} - \tfrac{c}{3})(t+t^2) + \tfrac{c}{3}(t^2+t^3).
\end{aligned}$$

Therefore, $[v]_B = \begin{bmatrix} a-b+c\\ b-c\\ c \end{bmatrix}$ and $[T(v)]_E = \begin{bmatrix} 0\\ a-b/2+c/3\\ b/2-c/3\\ c/3 \end{bmatrix}$. $\square$

As we see, a linear transformation $T : V \to W$ with fixed ordered bases B for V, and E for W, gives rise to a matrix. We can also construct back the linear transformation from such a given matrix. For, suppose $B = \{v_1, \ldots, v_n\}$ and $E = \{w_1, \ldots, w_m\}$. Let $v \in V$. There exist unique scalars $\beta_1, \ldots, \beta_n$ such that $v = \sum_{j=1}^n \beta_j v_j$. Then

$$T(v) = \sum_{j=1}^n \beta_j T(v_j) = \sum_{j=1}^n \beta_j (a_{1j}w_1 + \cdots + a_{mj}w_m).$$

We thus say that the matrix $[T]_{E,B}$ *represents* the linear transformation T.

Exercises for § 4.1

1. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(a,b) = (a-b, a, 2b+b)$. Let B be the standard basis of $\mathbb{R}^2$. Take the ordered bases $C = \{(1,2),(2,3)\}$ and $D = \{(1,1,0),(0,1,1),(2,2,3)\}$ for $\mathbb{R}^3$. Compute $[T]_{D,B}$ and $[T]_{D,C}$.

2. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(a,b,c) = (b+c, c+a, a+b)$. Determine $[T]_{E,B}$ where
   (a) $B = \{(1,0,0),(0,0,1),(0,1,0)\}$, $E = \{(0,0,1),(1,0,0),(0,1,0)\}$.
   (b) $B = \{(0,0,1),(1,0,0),(0,1,0)\}$, $E = \{(1,0,0),(0,0,1),(0,1,0)\}$.
   (c) $B = \{(1,1,-1),(-1,1,1),(1,-1,1)\}$,
   $E = \{(-1,1,1),(1,-1,1),(1,1,-1)\}$.
3. Define $T : \mathbb{R}_2[t] \to \mathbb{R}$ by $T(f) = f(2)$. Compute the matrix of T with respect to the standard bases of the spaces.
4. Let $T : \mathbb{F}_2[t] \to \mathbb{F}_3[t]$ be defined by $T(a+bt+ct^2) = at+bt^2+ct^3$. Take $B = \{1+t, 1-t, t^2\}$ and $E = \{1, 1+t, 1+t+t^2, t^3\}$ as ordered bases for $\mathbb{F}_2[t]$ and $\mathbb{F}_3[t]$, respectively. Then what is $[T]_{E,B}$?
5. Let $T : \mathbb{R}_2[t] \to \mathbb{R}_3[t]$ be given by $T(p(t)) = t\,p(t)$. Consider the ordered bases $B = \{1+t,\ 1-t,\ t^2\}$ and $E = \{1,\ 1+t,\ 1+t+t^2,\ t^3\}$ for $\mathbb{R}_2[t]$ and $\mathbb{R}_3[t]$, respectively. Find the matrix $[T]_{E,B}$.
6. Let $B = \{u_1, \ldots, u_n\}$ be an orthonormal ordered basis of an ips V. Let T be a linear operator on V. Show that the (i,j)th entry of $[T]_{B,B}$ is equal to $\langle Tu_j, u_i\rangle$.

4.2 Matrix operations

How are the coordinate vectors of v, $T(v)$ and the matrix of T related? Look back at Example 4.3, where we computed the coordinate vector of a typical vector and also that of its image. In the first problem there, with $v = (a,b)$, we have seen that

$$[T]_{E,B} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad [v]_B = \begin{bmatrix} a \\ b \end{bmatrix}, \quad [T(v)]_E = \begin{bmatrix} 2a-b \\ a+b \\ -a+b \end{bmatrix}.$$

In the second problem with $v = a+bt+ct^2+dt^3$, and $T = D$, we had

$$[T]_{E,B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad [v]_B = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad [T(v)]_E = \begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}.$$

In third problem, with $v = a+bt+ct^2+dt^3$ and $T = D$, we had

$$[T]_{E,B} = \begin{bmatrix} 0 & 1 & -2 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad [v]_B = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad [T(v)]_E = \begin{bmatrix} b-2c-3d \\ 2c \\ 3d \end{bmatrix}.$$

In the fourth problem, we had obtained

$$[T]_{E,B} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1/2 & -1/6 \\ 0 & 1/2 & 1/6 \\ 0 & 0 & 1/3 \end{bmatrix}, \quad [v]_B = \begin{bmatrix} a-b+c \\ b-c \\ c \end{bmatrix}, \quad [T(v)]_E = \begin{bmatrix} 0 \\ a - b/2 + c/3 \\ b/2 - c/3 \\ c/3 \end{bmatrix}.$$

Can you see how do we obtain the column vector $[T(v)]_E$ from the matrix $[T]_{E,B}$ and the column vector $[v]_B$? The rule is simple. The first component of $[T(v)]_E$ is the dot product of the first row of $[T]_{E,B}$ with the column vector $[v]_B$. The second component of $[T(v)]_E$ is the dot product of the second row of $[T]_{E,B}$ with the column vector $[v]_B$; and so on.

This suggests a rule of multiplying an $m \times n$ matrix with a column vector having n components. We can generalize this rule to multiplying an $m \times n$ matrix with another $n \times k$ matrix. For instance, consider multiplying

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{with} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}.$$

Multiplying A with the first column of B gives the column vector

$$\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \end{bmatrix}.$$

And multiplying A with the second column of B gives the column vector

$$\begin{bmatrix} a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \end{bmatrix}.$$

Putting them together in the same order, we obtain the matrix

$$\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \end{bmatrix}.$$

The **product** AB of two matrices $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ and $B = [b_{jk}] \in \mathbb{F}^{n\times r}$ is given by $AB = [c_{ik}] \in \mathbb{F}^{m\times r}$, where $c_{ik} = a_{i1}b_{1k} + \cdots + a_{in}b_{nk}$.

It says that the (i,k)th entry of AB is the dot product of the ith row of A and the jth column of B. It follows that the jth column of a matrix $A \in \mathbb{F}^{m\times n}$ is equal to Ae_j, where e_j is the standard basis vector of $\mathbb{F}^{n\times 1}$.

We may verify the following properties of matrix multiplication:

If $A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times r}$ and $C \in \mathbb{F}^{r\times p}$, then $(AB)C = A(BC)$.

If $A, B \in \mathbb{F}^{m\times n}$ and $C \in \mathbb{F}^{n\times r}$, then $(A+B)C = AB + AC$.

If $A \in \mathbb{F}^{m\times n}$ and $B, C \in \mathbb{F}^{n\times r}$, then $A(B+C) = AB + AC$.

Thus, matrix multiplication is associative, and it distributes over addition. However, it is not commutative: When AB is defined, BA may not be defined; even when both AB and BA are defined, they may not be of the same size; and even if they are of the same size, they need not be equal. For example,

$$\begin{bmatrix}1 & 2\\ 2 & 3\end{bmatrix}\begin{bmatrix}0 & 1\\ 2 & 3\end{bmatrix} = \begin{bmatrix}4 & 7\\ 6 & 11\end{bmatrix} \quad \text{but} \quad \begin{bmatrix}0 & 1\\ 2 & 3\end{bmatrix}\begin{bmatrix}1 & 2\\ 2 & 3\end{bmatrix} = \begin{bmatrix}2 & 3\\ 8 & 13\end{bmatrix}.$$

It does not mean that AB is never equal to BA.

Unlike numbers, product of two nonzero matrices can be a zero matrix. For example,

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

If $A \in \mathbb{F}^{m\times n}$ and $x \in \mathbb{F}^{n\times 1}$, then $Ax \in \mathbb{F}^{m\times 1}$. Thus, a system of m linear equations with n unknowns can be written as

$$Ax = b \quad \text{where} \quad b \in \mathbb{F}^{m\times 1}.$$

A matrix in $\mathbb{F}^{n\times n}$ is called a **square matrix** of order n.

Powers of square matrices can be defined inductively by taking

$$A^0 = I \quad \text{and} \quad A^n = AA^{n-1} \text{ for } n \in \mathbb{N}.$$

Using the multiplication of a matrix with a column vector, we may view a matrix as a linear transformation. For, suppose $A \in \mathbb{F}^{m\times n}$. Define the function $T_A : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ by

$$T_A(x) = Ax \quad \text{for each } x \in \mathbb{F}^{n\times 1}.$$

This is a linear transformation since

$$T_A(x + \alpha y) = A(x + \alpha y) = Ax + \alpha Ay = T_A(x) + \alpha T_A(y).$$

We use the symbol A itself for the linear transformation T_A and remember that $A(x)$ is Ax.

Convention 4.1 A matrix $A \in \mathbb{F}^{m\times n}$ is viewed as the linear transformation $A : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ defined by $A(x) = Ax$ for $x \in \mathbb{F}^{n\times 1}$.

Conversely, any linear transformation from $\mathbb{F}^{n\times 1}$ to $\mathbb{F}^{m\times 1}$ is a matrix multiplication. To see this, suppose $T : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ is a linear transformation. Let $e_1, \ldots, e_n$ be the standard basis vectors of $\mathbb{F}^{n\times 1}$. Construct the matrix $A \in \mathbb{F}^{m\times n}$ by taking the vector Te_i as its ith column; that is,

$$A = \begin{bmatrix} T(e_1) & \cdots & T(e_n) \end{bmatrix}.$$

If $x \in \mathbb{F}^{n\times 1}$, then we have scalars $a_1, \ldots, a_n$ such that $x = a_1e_1 + \cdots + a_ne_n$. Then

$$T(x) = a_1T(e_1) + \cdots + a_nT(e_n) = A \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = Ax.$$

The following theorem generalizes this observation to the situation of abstract finite dimensional vector spaces by using the coordinate vectors.

***Theorem* 4.4**
Let $B = \{v_1, \ldots, v_n\}$ and $E = \{w_1, \ldots, w_m\}$ be ordered bases for the vector spaces V and W, respectively. Let $T : V \to W$ be a linear transformation. Then $[T]_{E,B}$ is the unique matrix in $\mathbb{F}^{m\times n}$ such that for each $x \in V$,

$$[Tx]_E = [T]_{E,B}[x]_B.$$

Proof Let $x = a_1v_1 + \cdots + a_nv_n$. Then $Tx = a_1Tv_1 + \cdots + a_nTv_n$. Let $[T]_{E,B} = [b_{ij}] \in \mathbb{F}^{m\times n}$. Then we have

$$[x]_B = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad [Tv_j]_E = \begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix} \quad \text{for } j = 1, \ldots, n.$$

$$[Tx]_E = a_1 \begin{bmatrix} b_{11} \\ \vdots \\ b_{m1} \end{bmatrix} + \cdots + a_n \begin{bmatrix} b_{1n} \\ \vdots \\ b_{mn} \end{bmatrix} = \begin{bmatrix} b_{11}a_1 + \cdots + b_{1n}a_n \\ \vdots \\ b_{m1}a_1 + \cdots + b_{mn}a_n \end{bmatrix} = [T]_{E,B}[x]_B.$$

For the uniqueness of $[T]_{E,B}$, let $A \in \mathbb{F}^{m\times n}$ satisfy $[Tx]_E = A[x]_B$ for each $x \in V$. Let $j \in \{1, \ldots, n\}$. Take $x = v_j$, the jth basis vector in B. Then, $[v_j]_B = e_j$, the jth standard basis vector of $\mathbb{F}^{n\times 1}$. Consequently,

the jth column of $[T]_{E,B} = [T(v_j)]_E = A[v_j]_B = Ae_j =$ the jth column of A.

Therefore, $[T]_{E,B} = A$. ∎

Let $S, T : V \to W$ be linear transformations and let α be a scalar. Then the maps $S+T : V \to W$ and $\alpha S : V \to W$ defined by

$$(S+T)(x) = S(x)+T(x), \quad (\alpha S)(x) = \alpha S(x) \quad \text{for all } x \in V$$

are linear transformations. (Prove!) Analogously, addition of two matrices and multiplication of a matrix with a scalar are defined entry-wise.

Let $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{F}^{m\times n}$ be two matrices. Let $\alpha \in \mathbb{F}$. Then we define $A+B = [c_{ij}] \in \mathbb{F}^{m\times n}$ and $\alpha A = [d_{ij}] \in \mathbb{F}^{m\times n}$, where

$$c_{ij} = a_{ij} + b_{ij}, \quad \text{and} \quad d_{ij} = \alpha a_{ij}.$$

The **zero matrix** in $\mathbb{F}^{m\times n}$ has all entries as 0; it is written as 0. We see that for all matrices $A \in \mathbb{F}^{m\times n}$,

$$A+0 = 0+A = A.$$

For $A = [a_{ij}]$, the matrix $-A \in \mathbb{F}^{m\times n}$ is taken as one whose (i, j)th entry is $-a_{ij}$. Thus

$$-A = (-1)A \quad \text{and} \quad A+(-A) = -A+A = 0.$$

We also abbreviate $A+(-B)$ to $A-B$, as usual.

***Theorem* 4.5**
Let B and E be ordered bases for the finite dimensional vector spaces V and W, respectively. Let S, $T : V \to W$ be linear transformations. Let $\alpha \in \mathbb{F}$. Then

$$[S+T]_{E,B} = [S]_{E,B} + [T]_{E,B}, \quad \textit{and} \quad [\alpha T]_{E,B} = \alpha[T]_{E,B}.$$

Proof Let $x \in V$. Since the coordinate vector of a sum is the sum of the coordinate vectors, we have

$$\begin{aligned} [(S+T)(x)]_E &= [S(x)+T(x)]_E = [S(x)]_E + [T(x)]_E \\ &= [S]_{E,B}\,[x]_B + [T]_{E,B}\,[x]_B = \big([S]_{E,B} + [T]_{E,B}\big)\,[x]_B. \end{aligned}$$

By Theorem 4.4, $[S+T]_E = [S]_{E,B} + [T]_{E,B}$.

Similarly, it follows that $[\alpha T]_{E,B} = \alpha[T]_{E,B}$. ∎

Convention 4.1 suggests that $AB(x) = A(B(x))$; that is, the matrix product AB should correspond to the composition map AB. This is made precise in the following theorem.

***Theorem* 4.6**
Let B, C and D be ordered bases for the finite dimensional vector spaces U, V and W, respectively. Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then $[TS]_{D,B} = [T]_{D,C}\,[S]_{C,B}$.

Proof Suppose $B = \{u_1, \ldots, u_n\}$. Let e_j be the jth standard basis vector of $\mathbb{F}^{n\times 1}$. Then, for each $j \in \{1, \ldots, n\}$, $[u_j]_B = e_j$. Now, for each such j,

$$\begin{aligned}[TS]_{D,B}\, e_j &= [TS]_{D,B}[u_j]_B = [(TS)(u_j)]_D = [T(S(u_j))]_D \\ &= [T]_{D,C}[S(u_j)]_C = [T]_{D,C}[S]_{C,B}[u_j]_B = [T]_{D,C}[S]_{C,B}\, e_j.\end{aligned}$$

That is, the jth column of $[TS]_{D,B}$ is same as the jth column of $[T]_{D,C}[S]_{C,B}$ for each such j. Therefore, $[TS]_{D,B} = [T]_{D,C}[S]_{C,B}$. ∎

As you see, the addition of matrices, a scalar multiple of a matrix, and product of two matrices are defined in such a way that as linear transformations, they correspond to the sum of linear transformations, a scalar multiple of a linear transformation, and the composition of two linear transformations, respectively.

We consider a particular case. Suppose T is an isomorphism from V to W, where $\dim(V) = n = \dim(W)$. Then its inverse, written as T^{-1} is an isomorphism from W to V. Now, fix ordered bases B for V, and E for W. Then $T^{-1}T = I$, the identity map on V. If $B = \{v_1, \ldots, v_n\}$, then

$$(T^{-1}T)(v_j) = v_j = 0\,v_1 + \cdots + 0\,v_{j-1} + 1\,v_j + 0\,v_{j+1} + \cdots + 0\,v_n.$$

Then the $n\times n$ matrix $[T^{-1}T]_{B,B}$ has the jth column as $e_j \in \mathbb{F}^{n\times 1}$. Similarly, the ith column of $[TT^{-1}]_E$ is also e_i, where E is an ordered basis of W.

The matrix $I_n = [\delta_{ij}] \in \mathbb{F}^{n\times n}$ is called the **identity matrix** of order n.

A matrix $A \in \mathbb{F}^{n\times n}$ is called **invertible** iff there exists a matrix $B \in \mathbb{F}^{n\times n}$ such that $AB = BA = I_n$. Such a matrix B is denoted by A^{-1}, and is called an (the) **inverse** of A.

Thus the identity matrix of order n is the $n\times n$ matrix whose jth column is e_j. If $A \in \mathbb{F}^{m\times n}$, then $AI_n = A$ and $I_m A = A$. We will write an identity matrix as I, without subscript, unless a context demands otherwise.

It is easy to see that when a matrix A is invertible its inverse A^{-1} is unique. For, if B and C are two inverses of A, then

$$C = CI = C(AB) = (CA)B = IB = B.$$

An isomorphism maps a basis onto a basis. Looking at an $n\times n$ matrix as a linear transformation, we see that the images of the standard basis vectors are the columns of the matrix. It then follows that

a square matrix is invertible iff its columns are linearly independent.

Further, we observe that the matrix representation of the identity map with respect to the same basis in both copies of the vector space is the identity matrix.

We talk of invertibility of square matrices only; and all square matrices are not invertible. For example, I is invertible but 0 is not. If $AB = 0$ for nonzero square matrices A and B, then neither A nor B is invertible. Further, if both A and B are invertible, then AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Invertible matrices play a crucial role in solving linear systems uniquely. If A is invertible, then the solution of $Ax = b$ can be written as $x = A^{-1}b$. If A is not invertible, then $Ax = b$ may have infinitely many solutions or no solutions at all. We will prove these facts later in a relatively easy and conclusive manner, including the case of A being a non-square matrix.

Now, we look for the connection between the matrix representation of the inverse of an isomorphism and that of the isomorphism itself.

***Theorem* 4.7**

Let B and E be ordered bases for finite dimensional vector spaces V and W, respectively. Let $T : V \to W$ be an isomorphism. Then

$$[T^{-1}]_{B,E} = \left([T]_{E,B}\right)^{-1}.$$

Proof Due to Theorem 4.6, $[T^{-1}]_{B,E}[T]_{E,B} = [T^{-1}T]_{B,B} = [I]_{B,B} = I$. Similarly, $[T]_{E,B}[T^{-1}]_{B,E} = [TT^{-1}]_{E,E} = [I]_{E,E} = I$. ∎

Denote the set of all linear transformations from V to W by $\mathcal{L}(V,W)$. It is easy to verify that $\mathcal{L}(V,W)$ is a vector space over the same underlying field with the addition and scalar multiplication of linear transformations as mentioned earlier. Analogously, $\mathbb{F}^{m\times n}$ is a vector space with addition and scalar multiplication of matrices.

For $i = 1,\ldots,m$; $j = 1,\ldots,n$, let E_{ij} be the $m\times n$ matrix with its (i,j)th entry as 1 and all other entries 0. Then $\{E_{ij} : i = 1\ldots,m;\ j = 1,\ldots,n\}$ is a basis of $\mathbb{F}^{m\times n}$; prove it! Therefore, $\dim(\mathbb{F}^{m\times n}) = mn$. We regard E_{11} as the first, E_{12} as the second, proceeding further, E_{1n} as the nth, then E_{21} as the $(n+1)$th, and so on; so that this basis can be made into an ordered basis. The ordered basis

$$\{E_{11},\ldots,E_{1n},\ \ldots,\ E_{i1},\ldots,E_{ij},\ldots,E_{in},\ \ldots,\ E_{m1},\ldots,E_{mn}\}$$

is called the **standard basis of** $\mathbb{F}^{m\times n}$.

Once ordered bases for V and W are fixed, the matrix representation provides an isomorphism from $\mathcal{L}(V,W)$ to $\mathbb{F}^{m\times n}$, where $n = \dim(V)$ and $m = \dim(W)$. What is the linear transformation $T_{ij} \in \mathcal{L}(V,W)$ that corresponds to the matrix E_{ij}?

In addition to the structure of a vector space, $\mathcal{L}(V,V)$ is equipped with the composition of linear transformations. Correspondingly, in the vector space $\mathbb{F}^{n\times n}$, the operation of taking product of two matrices is available. The composition of linear transformations and the matrix product are associative, and satisfy the property of *bilinearity*. This means

$$A(\alpha B+\beta C)=\alpha AB+\beta AC \quad \text{and} \quad (\alpha B+\beta C)A=\alpha BA+\beta CA$$

for scalars α,β and linear transformations or matrices A,B,C. We thus say that both $\mathcal{L}(V,V)$ and $\mathbb{F}^{n\times n}$ are *linear algebras*, from which the subject derives its name.

Exercises for § 4.2

1. Prove that matrix multiplication is associative and distributive over addition. That is, if the products make sense, then $(AB)C=A(BC)$, $A(B+C)=AB+AC$ and $(A+B)C=AC+BC$.
2. Let $A\in\mathbb{C}^{m\times k}$ and $B\in\mathbb{C}^{k\times n}$ be matrices. Let $1\le i\le m$ and $1\le j\le n$. Show the following:
   (a) The ith row of AB is equal to (the ith row of A) times B.
   (b) The jth column of AB is equal to A times (the jth column of B).
3. Let $A\in\mathbb{F}^{m\times k}$ and let $B\in\mathbb{F}^{k\times n}$. Show the following:
   (a) Each column of AB is a linear combination of columns of A.
   (b) Each row of AB is a linear combination of rows of B.
4. Let $A=\begin{bmatrix}1&1&0\\0&1&2\\0&0&1\end{bmatrix}$. Show that $A^n=\begin{bmatrix}1&n&n(n-1)\\0&1&2n\\0&0&1\end{bmatrix}$ for $n\ge 0$.
5. Show that $\mathbb{F}^{m\times n}$ is a vector space with the operations of addition of two matrices and multiplying a scalar to a matrix.
6. For $i=1,2,\ldots,m$, and $j=1,2,\ldots,n$, let $E_{ij}\in\mathbb{F}^{m\times n}$ be the matrix whose (i,j)th entry is 1 and all other entries are 0. Show that the set of all such E_{ij} is a basis for $\mathbb{F}^{m\times n}$.
7. Is $A=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix},\begin{bmatrix}0&1\\1&0\end{bmatrix}\right\}$ linearly independent in $\mathbb{F}^{2\times 2}$?
8. Let $U=\left\{\begin{bmatrix}a&-a\\b&c\end{bmatrix}: a,b,c\in\mathbb{R}\right\}$; $V=\left\{\begin{bmatrix}a&b\\-a&c\end{bmatrix}: a,b,c\in\mathbb{R}\right\}$.
   (a) Prove that U and V are subspaces of $\mathbb{R}^{2\times 2}$.
   (b) Find bases, and hence dimensions, of $U\cap V$, U, V, and $U+V$.
9. Let $A\in\mathbb{F}^{m\times n}$. Let $0\in\mathbb{F}^{m\times 1}$ be the zero vector. Is the set of all $x\in\mathbb{F}^{1\times n}$ with $Ax^{t}=0$ a subspace of $\mathbb{F}^{1\times n}$?

10. Find, if possible, a basis $\{A_1, A_2, A_3, A_4\}$ for the vector space $\mathbb{R}^{2\times 2}$, where $A_i^2 = A_i$ for $i \in \{1,2,3,4\}$.
11. Let $A \in \mathbb{F}^{m\times n}$. Show the following:
    (a) $U = \{x \in \mathbb{F}^{n\times 1} : Ax = 0\}$ is a subspace of $\mathbb{F}^{n\times 1}$.
    (b) $W = \{Ax \in \mathbb{F}^{m\times 1} : x \in \mathbb{F}^{n\times 1}\}$ is a subspace of $\mathbb{F}^{m\times 1}$.
12. View a matrix $A \in \mathbb{F}^{m\times n}$ as a linear transformation from $\mathbb{F}^{n\times 1}$ to $\mathbb{F}^{m\times 1}$. Show that $R(T)$ is the span of the columns of A. Conclude that $\operatorname{rank}(A)$ is the maximum number of linearly independent columns of A.
13. Let $u_1, \ldots, u_n$ be the columns of $A \in \mathbb{F}^{m\times n}$. Let $b \in \mathbb{F}^{m\times 1}$. Is it true that the linear system of equations $Ax = b$ has a solution vector $x \in \mathbb{F}^{n\times 1}$ iff $b \in \operatorname{span}\{u_1, \ldots, u_n\}$?
14. Let $A = [a_{1j}] \in \mathbb{R}^{n\times n}$ and let $w_1, \ldots, w_n$ be the n columns of A. Let $\{u_1, \ldots, u_n\}$ be linearly independent in $\mathbb{R}^{n\times 1}$. Define vectors $v_1, \ldots, v_n$ by $v_j = a_{1j}u_1 + \ldots + a_{nj}u_n$, for $j = 1, 2, \ldots, n$. Show that $\{v_1, \ldots, v_n\}$ is linearly independent iff $\{w_1, \ldots, w_n\}$ is linearly independent.
15. Let $A = [a_{ij}] \in \mathbb{F}^{m\times n}$, where $m < n$. Show that there exists $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ such that $a_{i1}\alpha_1 + a_{i2}\alpha_2 + \cdots + a_{in}\alpha_n = 0$, for all $i = 1, \ldots, m$.
16. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \end{bmatrix}$. Show that A is invertible iff $a \neq 0$. In that case, A^{-1} is equal to $(1/a)(bI + cA - A^2)$.
17. In each case below, determine the values of α so that the matrix is invertible. Find the inverse of the matrix if possible.
$$\begin{bmatrix} \alpha & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & \alpha \\ 1 & \alpha \end{bmatrix}, \quad \begin{bmatrix} 1 & \alpha \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & \alpha \end{bmatrix}.$$
18. Is $T : \mathbb{R}^{3\times 3} \to \mathbb{R}^{2\times 2}$ defined below a linear transformation?
$$T\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 2a_{11} - a_{12} & a_{13} + 2a_{12} \\ 0 & 0 \end{bmatrix}$$
19. Let T be a linear operator on $\mathbb{C}^{2\times 2}$ defined by $T(X) = \begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix} X$. What is $\operatorname{rank}(T)$?
20. Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ be ordered bases for V and W, respectively. Using these bases, construct a basis with mn linear transformations from V to W for $\mathcal{L}(V, W)$.
21. Let $A \in \mathbb{F}^{m\times n}$. Define the linear transformation $T : \mathbb{F}^{n\times p} \to \mathbb{F}^{m\times p}$ by $T(X) = AX$ for $X \in \mathbb{F}^{n\times p}$. Express $\operatorname{rank}(T)$ and $\operatorname{null}(T)$ in terms of m, n, p, $\operatorname{rank}(A)$ and $\operatorname{null}(A)$.

4.3 Special types of matrices

Some types of matrices occur very frequently in applications. We wish to review those here; and also consider some more matrix operations.

If $A = [a_{ij}]$ is an $m \times n$ matrix, the entries a_{ii} for $i \in \{1, \ldots, \min\{m, n\}\}$ are called the **diagonal entries** of A. The **diagonal** of the matrix consists of all diagonal entries.

A matrix $A \in \mathbb{F}^{m \times n}$ is said to be **upper triangular** iff all entries below its diagonal are zero. That is, $A = [a_{ij}]$ is upper triangular when $a_{ij} = 0$ for $i > j$. In writing such a matrix, we simply do not show the zero entries below the diagonal.

Similarly, a matrix is called **lower triangular** iff all entries above its diagonal are zero. Both upper triangular and lower triangular matrices are referred to as **triangular** matrices.

A square matrix $A \in \mathbb{F}^{n \times n}$ is called a **diagonal matrix** iff all off-diagonal entries of A are 0.

A diagonal matrix is both upper triangular and lower triangular. A diagonal matrix $A \in \mathbb{F}^{n \times n}$ is called a **scalar matrix** if its diagonal entries are all same. That is, a scalar matrix is of the form αI for some $\alpha \in \mathbb{F}$. Scalar matrices commute with all square matrices of the same order. That is, if $A, B \in \mathbb{F}^{n \times n}$ and A is a scalar matrix, then $AB = BA$. Conversely, if $A \in \mathbb{F}^{n \times n}$ is such that $AB = BA$ for all $B \in \mathbb{F}^{n \times n}$, then A must be a scalar matrix. This fact is not at all obvious, and you should try proving it!

If $A = [a_{ij}] \in \mathbb{F}^{m \times n}$, then its **transpose** is a matrix in $\mathbb{F}^{n \times m}$, defined by

$$A^{\mathrm{t}} = [b_{ij}], \quad \text{where} \quad b_{ij} = a_{ji}.$$

That is, the jth column of A^{t} is the column vector $(a_{j1}, \; \cdots, \; a_{jn})^{\mathrm{t}}$ formed from the jth row of A. The rows of A are the columns of A^{t} and the columns of A are the rows of A^{t}, seen entry-wise. Notice that the transpose notation goes well with our style of writing a column vector as the transpose of a row vector. Transpose of a lower triangular matrix is an upper triangular matrix and vice versa.

If $A = [a_{ij}] \in \mathbb{F}^{m \times n}$, then its **conjugate transpose** is a matrix in $\mathbb{F}^{n \times m}$, defined by

$$A^* = [b_{ij}], \quad \text{where} \quad b_{ij} = \overline{a_{ji}}.$$

That is, the jth column of A^{t} is the column vector $(\overline{a_{j1}}, \; \cdots, \; \overline{a_{jn}})^{\mathrm{t}}$. Here, the bar above a_{ij} means the conjugate of the complex number a_{ij}. (Recall $\overline{\alpha + i\beta} = \alpha - i\beta$.) If $a_{ij} \in \mathbb{R}$, then $\overline{a_{ij}} = a_{ij}$. The conjugate transpose of A is also called the **adjoint** of A. Obviously, $I^{\mathrm{t}} = I^* = I$. In addition, the following are true:

If $A \in \mathbb{F}^{m\times n}$, then $(A^{\mathrm{t}})^{\mathrm{t}} = A$ and $(A^*)^* = A$.

If $A, B \in \mathbb{F}^{m\times n}$, then $(A+B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$ and $(A+B)^* = A^* + B^*$.

If $A \in \mathbb{F}^{m\times n}$ and $\alpha \in \mathbb{F}$, then $(\alpha A)^{\mathrm{t}} = \alpha A^{\mathrm{t}}$ and $(\alpha A)^* = \overline{\alpha} A^*$.

If $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times r}$, then $(AB)^{\mathrm{t}} = B^{\mathrm{t}} A^{\mathrm{t}}$ and $(AB)^* = B^* A^*$.

If $A \in \mathbb{F}^{n\times n}$ is invertible, then $(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}}$ and $(A^*)^{-1} = (A^{-1})^*$.

The standard inner product on $\mathbb{F}^{n\times 1}$ can now be written as $\langle x, y\rangle = y^* x$. Similarly, in $\mathbb{F}^n$, $\langle x, y\rangle = x y^*$.

Adjoints of matrices behave in a very predictable way with respect to the standard inner product in $\mathbb{F}^{n\times 1}$.

***Theorem* 4.8**
Let $A \in \mathbb{F}^{m\times n}$, $x \in \mathbb{F}^{n\times 1}$, and let $y \in \mathbb{F}^{m\times 1}$. Then $\langle Ax, y\rangle = \langle x, A^* y\rangle$ and $\langle A^* y, x\rangle = \langle y, Ax\rangle$.

Proof In $\mathbb{F}^{r\times 1}$, $\langle u, v\rangle = v^* u$. Thus $\langle Ax, y\rangle = y^* A x = (A^* y)^* x = \langle x, A^* y\rangle$. The second equality follows from the first. ∎

We can see the above theorem in a slightly generalized manner. It has to do something with orthonormal bases. Let $B = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Let $u, v \in V$. Then $u = \sum_{i=1}^n \langle u, v_i\rangle v_i$, $v = \sum_{j=1}^n \langle v, v_j\rangle v_j$, and

$$\langle u, v\rangle = \sum_{i=1}^n \sum_{j=1}^n \langle u, v_i\rangle \langle v, v_j\rangle \langle v_i, v_j\rangle = \sum_{i=1}^n \langle u, v_i\rangle \langle v, v_i\rangle = [u]_B \cdot [v]_B.$$

An orthonormal basis converts the inner product to the dot product. Using an orthonormal basis in a finite dimensional inner product space amounts to working in $\mathbb{F}^{n\times 1}$ with the dot product. Moreover, orthonormal bases allow writing the entries of the matrix representation of a linear transformation by using the inner products; see the following theorem.

***Theorem* 4.9**
Let $B = \{v_1, \ldots, v_n\}$ and $E = \{w_1, \ldots, w_m\}$ be ordered bases of the inner product spaces V and W, respectively. Let $T : V \to W$ be a linear transformation. If E is an orthonormal basis of W, then the (ij)th entry of $[T]_{E,B}$ is equal to $\langle T v_j, w_i\rangle$.

Proof For $1 \le i \le m$ and $1 \le j \le n$, let a_{ij} denote the (ij)th entry of the matrix $[T]_{E,B}$. Then $T v_j = a_{1j} w_1 + \cdots + a_{mj} w_m = \sum_{k=1}^m a_{kj} w_k$. Since E is orthonormal, $\langle T v_j, w_i\rangle = \big\langle \sum_{k=1}^m a_{kj} w_k,\ w_i\big\rangle = a_{ij} \langle w_i, w_i\rangle = a_{ij}$. ∎

Notice that In Theorem 4.9 the basis for V need not be orthonormal.

We find some connection between the matrix representation of the adjoint, and the adjoint of the matrix representation, of a linear transformation.

***Theorem* 4.10**
Let $B = \{v_1, \ldots, v_n\}$ and $E = \{w_1, \ldots, w_m\}$ be orthonormal ordered bases of the inner product spaces V and W, respectively. Let $T : V \to W$ be a linear transformation. Then $[T^*]_{B,E} = ([T]_{E,B})^*$.

Proof Suppose $1 \le i \le m$ and $1 \le j \le n$. Denote the (ij)th entry of $[T]_{E,B}$ by a_{ij} and that of $[T^*]_{B,E}$ by b_{ij}. By Theorem 4.9,

$$b_{ij} = \langle T^* w_j, v_i \rangle = \overline{\langle v_i, T^* w_j \rangle} = \overline{\langle T v_i, w_j \rangle} = \overline{a_{ji}}.$$

Therefore, $[T^*]_{B,E} = ([T]_{E,B})^*$. ∎

If the bases are not orthonormal, then the adjoint of the matrix representation may not represent the adjoint of the linear transformation.

Example 4.11
Let $u_1 = (1,1,0)$, $u_2 = (1,0,1)$ and $u_3 = (0,1,1)$. Consider $E = \{u_1, u_2, u_3\}$ as a basis of $\mathbb{R}^3$, and the standard basis $B = \{e_1, e_2, e_3, e_4\}$ of $\mathbb{R}^4$. Use the standard inner products (the dot products) on these spaces. Consider the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$T(a,b,c,d) = (a+c,\ b-2c+d,\ a-b+c-d).$$

The computation in Example 3.21 shows that $T^* : \mathbb{R}^3 \to \mathbb{R}^4$ is given by

$$T^*(\alpha,\beta,\gamma) = (\alpha+\gamma,\ \beta-\gamma,\ \alpha-2\beta+\gamma,\ \beta-\gamma).$$

For the matrix representations of T and of T^*, we proceed as follows:

$$\begin{array}{lclclcl}
Te_1 & = & T(1,0,0,0) & = & (1,0,1) & = & 0u_1 + 1u_2 + 0u_3 \\
Te_2 & = & T(0,1,0,0) & = & (0,1,-1) & = & 1u_1 - 1u_2 + 0u_3 \\
Te_3 & = & T(0,0,1,0) & = & (1,-2,1) & = & -1u_1 + 2u_2 - 1u_3 \\
Te_4 & = & T(0,0,0,1) & = & (0,1,-1) & = & 1u_1 - 1u_2 + 0u_3
\end{array}$$

$$\begin{array}{lclclcl}
T^*u_1 & = & T^*(1,1,0) & = & (1,1,-1,1) & = & 1e_1 + 1e_2 - 1e_3 + 1e_4 \\
T^*u_2 & = & T^*(1,0,1) & = & (2,-1,2,-1) & = & 2e_1 - 1e_2 + 2e_3 - 1e_4 \\
T^*u_3 & = & T^*(0,1,1) & = & (1,0,-1,0) & = & 1e_1 + 0e_2 - 1e_3 + 0e_4
\end{array}$$

Therefore, the matrices are

$$[T]_{E,B} = \begin{bmatrix} 0 & 1 & -1 & 1 \\ 1 & -1 & 2 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad [T^*]_{B,E} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 0 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Notice that E is not an orthonormal basis, and $[T^*]_{B,E} \neq ([T]_{E,B})^*$. □

A square matrix A is called **hermitian** iff $A^* = A$. And A is called **skew hermitian** iff $A^* = -A$.

A hermitian matrix with real entries satisfies $A^t = A$; and accordingly, such a matrix is called a **real symmetric** matrix. In general, if $A^t = A$, then A is called **symmetric**, and if $A^t = -A$, then A is called **skew symmetric**.

Let A be a square matrix. Since $A + A^t$ is symmetric and $A - A^t$ is skew symmetric, every square matrix can be written as a sum of a symmetric matrix and a skew symmetric matrix:

$$A = \tfrac{1}{2}(A + A^t) + \tfrac{1}{2}(A - A^t).$$

Similar rewriting is possible with hermitian and skew hermitian matrices:

$$A = \tfrac{1}{2}(A + A^*) + \tfrac{1}{2}(A - A^*).$$

A square matrix A is called **normal** iff $A^*A = AA^*$. A square matrix A is called **unitary** iff $A^*A = I = AA^*$. In addition, if all entries of A are real, then A is called an orthogonal matrix. That is, an **orthogonal matrix** is a matrix with real entries satisfying $A^tA = I = AA^t$.

A matrix which is hermitian, or real-symmetric, or unitary, or orthogonal is a normal matrix. But a normal matrix need not be any of these types.

The 2×2 orthogonal matrices have some geometrical significance. Let $A = [a_{ij}]$ be an orthogonal matrix of order 2. Then $A^tA = I$ implies

$$a_{11}^2 + a_{21}^2 = 1 = a_{12}^2 + a_{22}^2, \quad a_{11}a_{12} + a_{21}a_{22} = 0.$$

Thus, there exist $\alpha, \beta \in \mathbb{R}$ such that $a_{11} = \cos\alpha$, $a_{21} = \sin\alpha$, $a_{12} = \cos\beta$, $a_{22} = \sin\beta$, and $\cos(\alpha - \beta) = 0$. Hence, A is in one of the following forms:

$$O_1 := \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \quad O_2 := \begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{bmatrix}.$$

Let $\vec{u}$ be the vector in the plane that starts at the origin and ends at the point (a, b). Writing the point (a, b) as a column vector $[a,\ b]^t$, we see that the matrix product $O_1[a,\ b]^t$ is the endpoint of the vector obtained by rotating the vector $\vec{u}$ by an angle ϕ. Thus, O_1 is said to be a *rotation by an angle* ϕ.

Similarly, $O_2[a\ b]^t$ gives a point obtained by reflecting (a, b) along a straight line that makes an angle $\phi/2$ with the x-axis. Accordingly, O_2 is called a *reflection by an angle* $\phi/2$.

We show that unitary and orthogonal matrices preserve the inner product and the norm.

***Theorem* 4.12**
Let $A \in \mathbb{F}^{n\times n}$ be a unitary or an orthogonal matrix.

(1) *For each pair of vectors x, y, $\langle Ax, Ay\rangle = \langle x, y\rangle$.*
(2) *For each vector x, $\|Ax\| = \|x\|$.*
(3) *The columns of A form an orthonormal basis for $\mathbb{F}^{n\times 1}$.*
(4) *The rows of A form an orthonormal basis for $\mathbb{F}^{n}$.*

Proof Assume that A is unitary; that is, $A^*A = AA^* = I$.

(1) Now, $\langle Ax, Ay\rangle = \langle x, A^*Ay\rangle = \langle x, y\rangle$.

(2) Taking $x = y$ in (1) we obtain $\|Ax\| = \|x\|$.

(3) Since $A^*A = I$, the ith row of A^* multiplied with the jth column of A gives δ_{ij}. However, this product is simply the inner product of the jth column of A with the ith column of A.

(4) It follows from (3). Also, considering $AA^* = I$, we get this result.

When A is orthogonal, we take transpose instead of the adjoint. ∎

The connection between special types of linear operators and their matrix representations can be stated in the presence of an orthonormal basis.

***Theorem* 4.13**
Let T be a linear operator on a finite dimensional inner product space V. Let B be an orthonormal ordered basis of V.

(1) *T is self-adjoint iff $[T]_{B,B}$ is hermitian.*
(2) *T is normal iff $[T]_{B,B}$ is normal.*
(3) *T is unitary iff $[T]_{B,B}$ is unitary.*

Proof (1) Due to Theorem 4.10, $[T^*]_{B,B} = [T]^*_{B,B}$. If T is self-adjoint, then $T^* = T$. It follows that $[T]^*_{B,B} = [T^*]_{B,B} = [T]_{B,B}$. That is, $[T]_{B,B}$ is hermitian. Conversely, if $[T]^*_{B,B} = [T]_{B,B}$, then for each $v \in V$,

$$[T^*v]_B = [T^*]_{B,B}[v]_B = [T]^*_{B,B}[v]_B = [T]_{B,B}[v]_B = [Tv]_B.$$

It then follows that $T^*v = Tv$ for each $v \in V$. That is, $T^* = T$.
Proofs of (2)-(4) are similar to that of (1). ∎

The proof of Theorem 4.13(1) reveals that T is self-adjoint iff for each orthonormal basis B of V, $[T^*]_{B,B} = [T]_{B,B}$ iff for some orthonormal basis B of V, $[T^*]_{B,B} = [T]_{B,B}$. Similar comments hold for the statements in (2)-(4). Further, if V is a real inner product space, then 'hermitian' may be replaced by 'real symmetric' and 'unitary' by 'orthogonal'.

Exercises for § 4.3

1. Let $B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$.
   (a) Define $T : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ by $T(A) = A^{\mathrm{t}}$. Compute $[T]_{B,B}$.
   (b) Define $T : \mathbb{R}_2[t] \to \mathbb{R}^{2\times 2}$ by $T(f) = \begin{bmatrix} f'(0) & 2f(1) \\ 0 & f'(3) \end{bmatrix}$. Compute $[T]_{B,D}$, where $D = \{1,\ t,\ t^2\}$.
   (c) Define $T : \mathbb{R}^{2\times 2} \to \mathbb{R}$ by $T(A) =$ the sum of diagonal entries of A. Compute $[T]_{\{1\},B}$.
2. Is the set of all $n \times n$ hermitian matrices a subspace of $\mathbb{C}^{n\times n}$?
3. Show that the vector space of all 2×2 hermitian matrices with complex entries is a real vector space isomorphic to $\mathbb{R}^4$.
4. Determine the dimension of the vector space of matrices (over $\mathbb{R}$ or $\mathbb{C}$ or both, as appropriate) with complex entries of all
   (a) $n \times n$ matrices.
   (b) symmetric $n \times n$ matrices.
   (c) skew-symmetric $n \times n$ matrices.
   (d) hermitian $n \times n$ matrices.
   (e) upper triangular $n \times n$ matrices.
   (f) diagonal $n \times n$ matrices.
   (g) scalar $n \times n$ matrices.
5. Let A be an invertible matrix. Show the following:
   (a) If A is symmetric, then A^{-1} is symmetric.
   (b) If A is hermitian, then A^{-1} is hermitian.
   (c) If A is lower triangular, then A^{-1} is lower triangular.
   (d) If A is upper triangular, then A^{-1} is upper triangular.
6. Show that the product of two orthogonal matrices of the same order is an orthogonal matrix.
7. Let $A, B \in \mathbb{F}^{n\times n}$ Prove the following:
   (a) Let A and B be hermitian matrices. Show that AB is hermitian iff $AB = BA$.
   (b) If A is hermitian, then B^*AB is hermitian.
   (c) If B is invertible and B^*AB is hermitian, then A is hermitian.
8. Show that $W := \{A \in \mathbb{C}^{2\times 2} : A^* + A = 0\}$ is a subspace of $\mathbb{C}^{2\times 2}$. Find a basis for W.
9. Construct a normal matrix which is neither hermitian nor unitary.
10. Find a real normal matrix which is neither symmetric nor orthogonal.

4.4 Trace and determinant

The sum of all diagonal entries of a square matrix is called the **trace** of the matrix. That is, if $A = [a_{ij}] \in \mathbb{F}^{n\times n}$, then

$$\operatorname{tr}(A) = a_{11} + \cdots + a_{nn} = \sum_{k=1}^{n} a_{kk}.$$

The trace satisfies the following properties:

If $A \in \mathbb{F}^{n\times n}$ and $\alpha \in \mathbb{F}$, then $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr}(A)$.

If $A \in \mathbb{F}^{n\times n}$, then $\operatorname{tr}(A^{\mathrm{t}}) = \operatorname{tr}(A)$ and $\operatorname{tr}(A^*) = \overline{\operatorname{tr}(A)}$.

If $A, B \in \mathbb{F}^{n\times n}$, then $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ and $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Let $A = [a_{ij}] \in \mathbb{F}^{n\times n}$. Then $\operatorname{tr}(A^*A) = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2$. Therefore, $\operatorname{tr}(A) = 0$ iff $A = 0$. Obviously, $\operatorname{tr}(I_n) = n$.

Another helpful parameter of square matrices is called the determinant. The **determinant** of a square matrix A, denoted by $\det(A)$, is defined by induction on the order of the matrix, as in the following:

If $A = [a_{11}] \in \mathbb{F}^{1\times 1}$, then $\det(A) = a_{11}$.

If $A = [a_{ij}] \in \mathbb{F}^{n\times n}$ for $n > 1$, then $\det(A) = \sum_{j=1}^{n}(-1)^{j+1} a_{1j} \det(A_{1j})$, where A_{1j} is the matrix in $\mathbb{F}^{(n-1)\times(n-1)}$ obtained from A by deleting its first row and jth column.

When we show the entries of A explicitly, we write $\det(A)$ by enclosing the square array with two big vertical bars. For example, if $A = [a_{ij}] \in \mathbb{F}^{3\times 3}$, then

$$\begin{aligned}\det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).\end{aligned}$$

Using the determinant, a new square matrix, called the **adjugate** of A, denoted by $\operatorname{adj}(A)$, is formed in the following manner.

For $A \in \mathbb{F}^{1\times 1}$, $\operatorname{adj}(A) = I_1$, the 1×1 matrix whose sole entry is 1.

For $A \in \mathbb{F}^{n\times n}$, $n > 1$, let $A_{ij} \in \mathbb{F}^{(n-1)\times(n-1)}$ be the matrix obtained from A by removing the ith row and the jth column of A. Then

$$\operatorname{adj}(A) = [b_{ij}] \in \mathbb{F}^{n\times n}, \quad \text{where } \; b_{ij} = (-1)^{i+j}\det(A_{ji}).$$

That is, adjugate of A is the transpose of the matrix whose (i, j)th entry is $(-1)^{i+j}\det(A_{ij})$.

For instance, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

We list some properties of the determinant. Let $A \in \mathbb{F}^{n\times n}$.

1. If A is a triangular matrix, then $\det(A) = a_{11}a_{22}\cdots a_{nn}$.
2. If $n > 1$, then for any row index i, $\det(A) = \sum_{j=1}^{n}(-1)^{j+1}a_{ij}\det(A_{ij})$.
3. If all entries in a row of A are 0, then $\det(A) = 0$.
4. If two rows of A are identical, then $\det(A) = 0$.
5. If two rows of A are exchanged to get a matrix B, then $\det(B) = -\det(A)$.
6. If a scalar α is multiplied to a row of A to obtain the matrix B, then $\det(B) = \alpha\det(A)$.
7. If a row of A is replaced by that row plus a nonzero scalar α times another row to obtain the matrix B, then $\det(B) = \det(A)$.
8. Let $B, C \in \mathbb{F}^{n\times n}$. Let k be a row index. Suppose that
   (a) the kth row of A = the kth row of B + the kth row of C; and
   (b) for $i \neq k$, the ith row of A = the ith row of B = the ith row of C.

   Then $\det(A) = \det(B) + \det(C)$.
9. If $n > 1$, then for any column index j, $\det(A) = \sum_{i=1}^{n}(-1)^{i+1}a_{ij}\det(A_{ij})$.
10. Properties (3)-(8) are true if the word 'row' is replaced with 'column'.
11. $\det(A^{\mathrm{t}}) = \det(A)$ and $\det(A^*) = \overline{\det(A)}$.
12. If $B \in \mathbb{F}^{n\times n}$, then $\det(AB) = \det(A)\det(B) = \det(BA)$.
13. $A\,\mathrm{adj}(A) = \mathrm{adj}(A)\,A = \det(A)\,I$.
14. If $\det(A) \neq 0$, then A is invertible, and $A^{-1} = (\det(A))^{-1}\,\mathrm{adj}(A)$.

Proofs of these properties are easy but long; and we leave them as exercises.

Example 4.14

Show that the matrix $A = \begin{bmatrix} 1 & 4 & -6 \\ -1 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}$ is invertible; and find its inverse.

We find the determinants of A_{ij} as follows:

$$\det(A_{11}) = \begin{vmatrix} -1 & 3 \\ -2 & 3 \end{vmatrix} = 3,\ \det(A_{12}) = \begin{vmatrix} -1 & 3 \\ 1 & 3 \end{vmatrix} = -6,\ \det(A_{13}) = \begin{vmatrix} -1 & -1 \\ 1 & -2 \end{vmatrix} = 3,$$

$$\det(A_{21}) = \begin{vmatrix} 4 & -6 \\ -2 & 3 \end{vmatrix} = 0,\ \det(A_{22}) = \begin{vmatrix} 1 & -6 \\ 1 & 3 \end{vmatrix} = 9,\ \det(A_{23}) = \begin{vmatrix} 1 & 4 \\ 1 & -2 \end{vmatrix} = -6,$$

$$\det(A_{31}) = \begin{vmatrix} 4 & -6 \\ -1 & 3 \end{vmatrix} = 6,\ \det(A_{32}) = \begin{vmatrix} 1 & -6 \\ -1 & 3 \end{vmatrix} = -3,\ \det(A_{33}) = \begin{vmatrix} 1 & 4 \\ -1 & -1 \end{vmatrix} = 3.$$

Then $\det(A) = +(1)(3) - 4(-6) + (-6)(3) = 9$. Thus A is invertible. And

$$\mathrm{adj}(A) = \begin{bmatrix} 3 & 6 & 3 \\ 0 & 9 & 6 \\ 6 & 3 & 3 \end{bmatrix}^{\mathrm{t}} = \begin{bmatrix} 3 & 0 & 6 \\ 6 & 9 & 3 \\ 3 & 6 & 3 \end{bmatrix}.$$

$$A^{-1} = (\det(A))^{-1}\mathrm{adj}(A) = \frac{1}{9}\begin{bmatrix} 3 & 0 & 6 \\ 6 & 9 & 3 \\ 3 & 6 & 3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 2 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}.$$

You should verify that $A^{-1}A = AA^{-1} = I$. □

Of course, without using the adjugate, we can show that a matrix is invertible when its determinant is nonzero.

***Theorem* 4.15**
Let $A \in \mathbb{F}^{n\times n}$. Then, A is invertible iff $\mathrm{rank}(A) = n$ *iff* $\det(A) \neq 0$.

Proof The matrix A is invertible iff it is an isomorphism iff A maps $\mathbb{F}^{n\times 1}$ onto itself iff $\mathrm{rank}(A) = \dim(R(T)) = \dim(\mathbb{F}^{n\times 1}) = n$.

If $\mathrm{rank}(A) < n$, then, the columns of A are linearly dependent. That is, some column of A is a linear combination of other columns. Then $\det(A) = 0$.

Conversely, if $\mathrm{rank}(A) = n$, then A is invertible. Now, $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1$ shows that $\det(A) \neq 0$. ∎

Exercises for § 4.3

1. Let $V = \{A \in \mathbb{C}^{2\times 2} : \mathrm{tr}(A) = 0\}$. Show that V is a real vector space. What is $\dim(V)$?
2. Let $V = \{A \in \mathbb{C}^{n\times n} : \mathrm{tr}(A) = 0\}$ for $n > 2$. Is V a vector space?
3. Show that the map $T : \mathbb{F}^{n\times n} \to \mathbb{F}$ defined by $T(A) = \mathrm{tr}(A)$ is a linear functional. Show that T is an onto map. Also, show that the subspace of all matrices in $\mathbb{F}^{n\times n}$ with trace 0 has dimension $n^2 - 1$.
4. Construct a matrix $A \in \mathbb{R}^{2\times 2}$ with $\mathrm{tr}(A^2) < 0$.
5. In the following is $\langle\, , \,\rangle$ an inner product on the vector space V?
   (a) $\langle A, B\rangle = \mathrm{tr}(A + B)$ for A, B in $V = \mathbb{R}^{2\times 2}$.
   (b) $\langle A, B\rangle = \mathrm{tr}(A^{\mathrm{t}}B)$ for A, B in $V = \mathbb{R}^{3\times 3}$.
6. Construct matrices $A, B \in \mathbb{F}^{n\times n}$ so that $\mathrm{tr}(AB) \neq \mathrm{tr}(A)\mathrm{tr}(B)$.
7. Let $A \in \mathbb{F}^{m\times n}$ and let $B \in \mathbb{F}^{n\times m}$. Show that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.
8. Let $A, B \in \mathbb{C}^{n\times n}$. Show that $AB - BA \neq I$.
9. Let $C \in \mathbb{F}^{2\times 2}$. Show that $\mathrm{tr}(C) = 0$ iff there exist $A, B \in \mathbb{F}^{2\times 2}$ such that $C = AB - BA$.

10. Let $A = [a_{ij}] \in \mathbb{R}^{2\times 2}$. For $x, y \in \mathbb{R}^2$, let $f_A(x, y) = y^t Ax$. Show that f_A is an inner product on $\mathbb{R}^2$ iff $a_{12} = a_{21}$, $a_{11} > 0$, $a_{22} > 0$, and $a_{11}a_{22} - a_{12}a_{21} > 0$. [Hint: $(y^t Ax)^t = y^t Ax$ implies $a_{12} = a_{21}$. Next, $x^t Ax \ge 0$ gives a quadratic. Complete the square and argue.]
11. Let $A \in \mathbb{C}^{m\times m}$. Show that if $\operatorname{tr}(A^* A) = 0$, then $A = 0$. What would happen for matrices in $\mathbb{C}^{m\times n}$?
12. Let A be a square matrix such that $A^* A = A^2$. Prove that A is hermitian.
13. Prove all the properties of the determinant as mentioned in the text.
14. Let $A \in \mathbb{R}^{4\times 4}$ have all entries as integers. It is found that for some matrix $B \in \mathbb{R}^{4\times 4}$, $AB = 2C$, where $C \in \mathbb{R}^{4\times 4}$ has determinant 1. Then What is the maximum value of $\det(A) + \det(B)$?
15. Let $A \in \mathbb{F}^{n\times n}$. Let $E_{ij} \in \mathbb{F}^{n\times n}$ have the (i, j)th entry as 1 and all other entries 0. Show that if $AE_{ij} = E_{ij}A$ for all i, j with $1 \le i, j \le n$, then $A = \alpha I$ for some $\alpha \in \mathbb{F}$.
16. Let $A = [a_{ij}] \in \mathbb{F}^{n\times n}$, $B = \operatorname{diag}(1, 2, 3, \ldots, n)$, and let $C = [c_{ij}] \in \mathbb{F}^{n\times n}$, where $c_{i1} = c_{ii} = 1$ for $1 \le i \le n$, and all other c_{ij} are 0. Show the following:
    (a) If $AB = BA$, then $A = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.
    (b) If $\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})\, C = C \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, then $a_{ii} = a_{11}$ for each $i = 1, 2, \ldots, n$.
    (c) If $AM = MA$ for all invertible matrices $M \in \mathbb{F}^{n\times n}$, then $A = \alpha I$ for some $\alpha \in \mathbb{F}$.
17. Suppose $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times m}$. Show the following:
    (a) If $m = n$, then $AB = I$ implies that $BA = I$.
    (b) If $m \ne n$, then $AB = I_m$ does not necessarily imply that $BA = I_n$.

4.5 Change of basis

If we fix an ordered basis for an n-dimensional vector space V, then the coordinate vector map with respect to this basis provides an isomorphism from V to $\mathbb{F}^{n\times 1}$. Using such a coordinate vector map, we had seen how to represent a linear transformation by a matrix.

To recall, let $B = \{v_1, \ldots, v_n\}$ be an ordered basis of V. Let $C = \{w_1, \ldots, w_m\}$ be an ordered basis of W. Have standard bases for $\mathbb{F}^{n\times 1}$ and $\mathbb{F}^{m\times 1}$. The coordinate vector maps $[\ ]_B : V \to \mathbb{F}^{n\times 1}$ and $[\ ]_C : W \to \mathbb{F}^{m\times 1}$ are isomorphisms. Let $T : V \to W$ be a linear transformation and $[T]_{C,B}$ be its matrix representation. Then we have the following commutative diagram:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "{[\ ]_B}"', "\simeq"] & W \arrow[d, "\simeq"', "{[\ ]_C}"] \\ \mathbb{F}^{n\times 1} \arrow[r, "{[T]_{C,B}}"'] & \mathbb{F}^{m\times 1} \end{tikzcd}$$

It means

$$T = [\ ]_C^{-1} \circ [T]_{C,B} \circ [\ ]_B, \quad [T]_{C,B} = [\ ]_C \circ T \circ [\ ]_B^{-1}. \tag{4.1}$$

Also, $[\ ]_C \circ T = [T]_{C,B} \circ [\ ]_B$, which amounts to

$$[T(x)]_C = [T]_{C,B}\,[x]_B \quad \text{for each } x \in V. \tag{4.2}$$

We know that isomorphisms preserve rank and nullity. Since the coordinate vector maps are isomorphisms, we obtain the following theorem.

***Theorem* 4.16**
Let V and W be finite dimensional vector spaces with ordered bases B and C, respectively. Let $T : V \to W$ be a linear transformation. Then

$$\text{rank}(T) = \text{rank}([T]_{C,B}) \quad \textit{and} \quad \text{null}(T) = \text{null}([T]_{C,B}).$$

Since we are able to go from T to its matrix representation $[T]_{C,B}$ and back in a unique manner, it suggests that the *matrix representation* itself is some sort of isomorphism. It is easy to verify that the map $T \mapsto [T]_{C,B}$ is an isomorphism from $\mathcal{L}(V, W)$ to $\mathbb{F}^{m\times n}$.

It thus follows that $\dim(\mathcal{L}(V, W)) = mn$. Alternatively, a basis for $\mathcal{L}(V, W)$ can be constructed explicitly. Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ be ordered bases for the vector spaces V and W, respectively. Suppose $1 \le i \le n$ and $1 \le j \le m$. Define $T_{ij} : V \to W$ by $T_{ij}(v_i) = w_j$, $T_{ij}(v_k) = 0$ for $k \ne i$. Then show that the set $\{T_{ij} : i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is a basis for $\mathcal{L}(V, W)$.

We look at a particular case of the composition formulas in (4.1). Consider a vector space V of dimension n, with ordered bases $O = \{v_1, \ldots, v_n\}$ and $N = \{w_1, \ldots, w_n\}$. Consider the identity map I on V. Let us write V_O for the vector space V where we take the ordered basis as O. Similarly, write V_N for the same space but with the ordered basis N. Fix the standard basis $E = \{e_1, \ldots, e_n\}$ for $\mathbb{F}^{n\times 1}$. The matrix representation diagram now looks like

$$\begin{tikzcd} V_O \arrow[r, "I"] \arrow[d, "{[\ ]_O}"', "\simeq"] & V_N \arrow[d, "\simeq"', "{[\ ]_N}"] \\ \mathbb{F}^{n\times 1} \arrow[r, "{[I]_{N,O}}"'] & \mathbb{F}^{n\times 1} \end{tikzcd}$$

Here, $[\ ]_O$ maps each v_i to the corresponding e_i and $[\ ]_N$ maps each w_j to the corresponding e_j. Then (4.2) says that

$$[v]_N = [I\,v]_N = [I]_{N,O}\,[v]_O \quad \text{for each } v \in V.$$

Thus the $n \times n$ matrix $[I]_{N,O}$ is called the **change of basis** matrix. This matrix is obtained by collecting the scalars while expressing the basis vectors $v_i \in O$ as linear combinations of basis vectors $w_j \in N$.

Notice that $I_{N,O}$ is invertible, and $I_{N,O}^{-1} = I_{O,N}$.

Example 4.17

Consider two ordered bases for $\mathbb{R}^3$ such as $O = \{(1,0,1),(1,1,0),(0,1,1)\}$ and $N = \{(1,-1,1),(1,1,-1),(-1,1,1)\}$. Find the change of basis matrix $[I]_{N,O}$ and verify that $[v]_N = [I]_{N,O}[v]_O$ for the vector $v = (1,2,3)$.

We need to express each vector in O as a linear combination of vectors in N. Towards this, suppose

$$(1,0,1) = a(1,-1,1) + b(1,1,-1) + c(-1,1,1).$$

Then $a+b-c = 1,\ -a+b+c = 0,\ a-b+c = 1$. Solving these equations, we obtain $a = 1$, $b = \frac{1}{2}$, and $c = \frac{1}{2}$. So,

$$(1,0,1) = 1(1,-1,1) + \tfrac{1}{2}(1,1,-1) + \tfrac{1}{2}(-1,1,1).$$

Continuing with the second and third vectors in O, we obtain

$$(1,1,0) = \tfrac{1}{2}(1,-1,1) + 1(1,1,-1) + \tfrac{1}{2}(-1,1,1).$$

$$(0,1,1) = \tfrac{1}{2}(1,-1,1) + \tfrac{1}{2}(1,1,-1) + 1(-1,1,1).$$

Therefore the change of basis matrix is

$$[I]_{N,O} = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}.$$

For $v = (1,2,3)$, we obtain

$$v = (1,2,3) = 2(1,-1,1) + \tfrac{3}{2}(1,1,-1) + \tfrac{5}{2}(-1,1,1) \Rightarrow [v]_N = [2, 3/2, 5/2]^t.$$
$$v = (1,2,3) = 1(1,0,1) + 0(1,1,0) + 2(0,1,1) \Rightarrow [v]_O = [1,0,2]^t.$$

$$[I]_{N,O}[v]_O = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3/2 \\ 5/2 \end{bmatrix} = [v]_N. \qquad \square$$

In Example 4.17, construct the matrix B by taking its columns as the transposes of vectors in O, keeping the same order of the vectors as given in

O. Similarly, construct the matrix C by taking its columns as the transposes of vectors from N, again keeping the same order. We claim that $[I]_{N,O} = C^{-1}B$. Indeed, the following may be easily verified:

$$C\,[I]_{N,O} = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = B.$$

In general, if $V = \mathbb{F}^{n\times 1}$, then the change of basis matrix can be given in a closed from using the given basis vectors.

***Theorem* 4.18**
Let $O = \{v_1, \ldots, v_n\}$ and $N = \{w_1, \ldots, w_n\}$ be ordered bases for $\mathbb{F}^{n\times 1}$. Let $E = \{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{F}^{n\times 1}$. Then the change of basis matrices $[I]_{E,O}$ and $[I]_{N,O}$ are given by

$$[I]_{E,O} = [v_1 \ \cdots \ v_n] \quad \textit{and} \quad [I]_{N,O} = [w_1 \ \cdots w_n]^{-1}[v_1 \ \cdots \ v_n].$$

Proof Since $[v_j]_O = e_j$, we have

$$[I]_{E,O}\, e_j = [I]_{E,O}[v_j]_O = [I(v_j)]_E = [v_j]_E = v_j.$$

That is, the jth column of the matrix $[I]_{E,O}$ is simply v_j. So,

$$[I]_{E,O} = [v_1 \ \cdots \ v_n].$$

It is the matrix formed by putting the vectors $v_1, \ldots, v_n$ as columns in that order. Similarly,

$$[I]_{E,N} = [w_1 \ \cdots \ w_n].$$

Using these equalities, we obtain

$$[I]_{N,O} = [I]_{N,E}\,[I]_{E,O} = [I]_{E,N}^{-1}\,[I]_{E,O} = [w_1 \ \cdots w_n]^{-1}[v_1 \ \cdots \ v_n]. \quad \blacksquare$$

In fact, columns of any invertible $n \times n$ matrix form a basis for $\mathbb{F}^{n\times 1}$. Therefore, any invertible matrix is a change of basis matrix in this sense. It gives rise to the following generalization.

***Theorem* 4.19**
Let $A \in \mathbb{F}^{m\times n}$. Let $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_m\}$ be ordered bases for $\mathbb{F}^{n\times 1}$ and $\mathbb{F}^{m\times 1}$, respectively. Then $[A]_{C,B} = Q^{-1}AP$, where

$$Q = [w_1 \ \cdots \ w_m] \quad \textit{and} \quad P = [v_1 \ \cdots \ v_n].$$

Proof Take the standard bases for $\mathbb{F}^{n\times 1}$ and $\mathbb{F}^{m\times 1}$ as D and E, respectively. Now, $[I]_{D,B} = [v_1 \ \cdots \ v_n]$, $[I]_{E,C} = [w_1 \ \cdots \ w_m]$, and $[A]_{E,D} = A$. Then $[A]_{C,B} = [I]_{C,E}\,[A]_{E,D}\,[I]_{D,B} = [w_1 \ \cdots \ w_m]^{-1}\,A\,[v_1 \ \cdots \ v_n]$. $\blacksquare$

Theorem 4.19 says that a matrix $A \in \mathbb{F}^{m\times n}$ and a matrix $Q^{-1}AP$, where both $Q \in \mathbb{F}^{m\times m}$ and $P \in \mathbb{F}^{n\times n}$ are invertible, represent the same linear transformation with respect to ordered bases chosen in both the domain and the co-domain spaces. Note that a matrix $A \in \mathbb{F}^{m\times n}$ is equal to $[A]_{E',E}$, where E is the standard basis for $\mathbb{F}^{n\times 1}$ and E' is the standard basis for $\mathbb{F}^{m\times 1}$.

Example 4.20

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$ as a linear transformation from $\mathbb{R}^{3\times 1}$ to $\mathbb{R}^{2\times 1}$. It maps

$$v = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \text{to} \quad Av = \begin{bmatrix} a+2b+3c \\ b+c \end{bmatrix} \quad \text{for } a,b,c \in \mathbb{R}.$$

Choose ordered bases

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\} \text{ for } \mathbb{R}^{3\times 1}, \quad C = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\} \text{ for } \mathbb{R}^{2\times 1}.$$

Then

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad Q^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

We write $v = (a,b,c)^t$ as a linear combination of basis vectors from B, and $Av = (a+2b+3c,\ b+c)^t$ as a linear combination of basis vectors from C.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \tfrac{1}{3}(a+b+c)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \tfrac{1}{3}(-2a+b+c)\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + \tfrac{1}{3}(-a+2b-c)\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

$$\begin{bmatrix} a+2b+3c \\ b+c \end{bmatrix} = \tfrac{1}{2}(a+3b+4c)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \tfrac{1}{2}(a+b+2c)\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus their coordinate vectors with respect to bases B and C are as follows:

$$[v]_B = \frac{1}{3}\begin{bmatrix} a+b+c \\ -2a+b+c \\ -a+2b-c \end{bmatrix}, \quad [Av]_C = \frac{1}{2}\begin{bmatrix} a+3b+4c \\ a+2b+2c \end{bmatrix}.$$

Then the matrix representation of A with respect to the new bases is given by

$$[A]_{C,B} = Q^{-1}AP = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 8 & 3 & -1 \\ 4 & 1 & -1 \end{bmatrix}.$$

As it should happen, we see that

$$[A]_{C,B}\,[v]_B = \tfrac{1}{2}\begin{bmatrix} 8 & 3 & -1 \\ 4 & 1 & -1 \end{bmatrix} \tfrac{1}{3}\begin{bmatrix} a+b+c \\ -2a+b+c \\ -a+2b-c \end{bmatrix} = \tfrac{1}{2}\begin{bmatrix} a+3b+4c \\ a+2b+2c \end{bmatrix} = [Av]_C. \qquad \square$$

Exercises for § 4.5

1. Consider the ordered bases $O = \{[1,0,1]^t,\ [1,1,0]^t,\ [0,1,1]^t\}$ and $N = \{[1,-1,1]^t,\ [1,1,-1]^t,\ [-1,1,1]^t\}$ for $\mathbb{R}^{3\times 1}$. Let T be the linear operator on $\mathbb{R}^{3\times 1}$ whose matrix representation with respect to the standard basis of $\mathbb{R}^{3\times 1}$ is given by the matrix $\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.
   (a) Find the change of basis matrix $I_{N,O}$.
   (b) Find the matrix $[T]_{N,O}$.
   (c) Verify that $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_N = I_{N,O} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_O$ and $\left[T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right]_N = [T]_{N,O} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_O$.

2. Consider the ordered bases $O = \{u_1, u_2\}$ and $N = \{v_1, v_2\}$ for $\mathbb{F}^{2\times 1}$, where $u_1 = (1,1)^t$, $u_2 = (-1,1)^t$, $v_1 = (2,1)^t$, and $v_2 = (1,0)^t$. Let $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.
   (a) Compute $Q = [A]_{O,O}$ and $R = [A]_{N,N}$.
   (b) Find the change of basis matrix $P = I_{N,O}$.
   (c) Compute $S = PQP^{-1}$.
   (d) Is it true that $R = S$? Why?
   (e) If $S = [s_{ij}]$, verify that $Av_1 = s_{11}v_1 + s_{21}v_2,\ Av_2 = s_{12}v_1 + s_{22}v_2$.

3. Construct a matrix $A \in \mathbb{R}^{2\times 2}$, a vector $v \in \mathbb{R}^{2\times 1}$, and a basis $B = \{u_1, u_2\}$ for $\mathbb{R}^{2\times 1}$ satisfying $[Av]_B \neq A[v]_B$.

4. Let V and W be vector spaces with bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$, respectively. For $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, define $T_{ij} : V \to W$ by $T_{ij}(v_i) = w_j$, and $T_{ij}(v_k) = 0$ for $k \neq i$. Then show that the set of all these T_{ij} is a basis for $\mathcal{L}(V, W)$.

5. Let $B = \{u_1, \ldots, u_n\}$ and $E = \{v_1, \ldots, v_m\}$ be ordered bases of V and W, respectively. Let $T \in \mathcal{L}(V, W)$. Show the following:
   (a) T is one-one iff columns of $[T]_{E,B}$ are linearly independent.
   (b) T is onto iff columns of $[T]_{E,B}$ span $\mathbb{F}^{m\times 1}$.

6. Let $B = \{u_1, \ldots, u_n\}$ and $E = \{v_1, \ldots, v_m\}$ be ordered bases of V and W, respectively. Let $\{M_{ij} : i = 1\ldots, m;\ j = 1, \ldots, n\}$ be an ordered basis of $\mathbb{F}^{m\times n}$. Let $T_{ij} \in \mathcal{L}(V, W)$ be such that $[T_{ij}]_{E,B} = M_{ij}$. Then a basis for $\mathcal{L}(V, W)$ is given by $\{T_{ij} : i = 1\ldots, m;\ j = 1, \ldots, n\}$.

7. Let $\{u_1, \ldots, u_n\}$ be an ordered basis of an ips V. Show the following:
   (a) The matrix $[a_{ij}]$, where $a_{ij} = \langle u_i, u_j \rangle$, is invertible.
   (b) If $\alpha_1, \alpha_2, \ldots \alpha_n \in \mathbb{F}$, then there is exactly one vector $x \in V$ such that $\langle x, u_j \rangle = \alpha_j$, for $j = 1, 2, \ldots, n$.

8. Let T be a linear operator on a finite dimensional vector space V. Let B and C be ordered bases for V. Show that $\operatorname{tr}([T]_{C,C}) = \operatorname{tr}([T]_{B,B})$ and $\det([T]_{C,C}) = \det([T]_{B,B})$. Thus, define $\operatorname{tr}(T)$ and $\det(T)$.
9. Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis of an ips V. Show that for any $x, y \in V$, $\langle x, y\rangle = \sum_{k=1}^{n} \langle x, u_k\rangle\langle u_k, y\rangle$. Then deduce that there exists an isometric linear transformation T from V to $\mathbb{F}^n$.

4.6 Equivalence

Let $A, B \in \mathbb{F}^{m\times n}$. We say that B is **equivalent to** A iff there exist invertible matrices $P \in \mathbb{F}^{n\times n}$ and $Q \in \mathbb{F}^{m\times m}$ such that $B = Q^{-1}AP$.

As it happens, a change of bases in both the domain and co-domain spaces bring up an equivalent matrix that represents the original matrix viewed as a linear transformation.

It is easy to see that on $\mathbb{F}^{m\times n}$, the relation 'is equivalent to' is an equivalence relation. Observe that A and B are equivalent iff there exist invertible matrices P and Q of appropriate order such that $B = QAP$. There is an easy characterization of equivalence of two matrices.

***Theorem* 4.21 (Rank Theorem)**
Two matrices of the same size are equivalent iff they have the same rank.

Proof Let A and B be $m \times n$ matrices. We view them as linear transformations from $\mathbb{F}^{n\times 1}$ to $\mathbb{F}^{m\times 1}$. Observe that isomorphisms on $\mathbb{F}^{k\times 1}$ are simply invertible matrices. Now, using Theorems 3.17-3.18 we obtain the following:

A and B are equivalent
iff there exist invertible matrices $P \in \mathbb{F}^{n\times n}$, $Q \in \mathbb{F}^{m\times m}$ such that $B = QAP$
iff there exist isomorphisms P on $\mathbb{F}^{n\times n}$, and Q on $\mathbb{F}^{m\times 1}$ such that $B = QAP$
iff $\operatorname{rank}(B) = \operatorname{rank}(A)$. ∎

It is easy to construct an $m \times n$ matrix of rank r. For instance

$$E_r = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{F}^{m\times n}$$

is such a matrix, where I_r is the identity matrix of order r, and the other zero matrices are of appropriate size. We thus obtain the following result as a corollary to the Rank theorem.

***Theorem* 4.22 (Rank factorization)**
Let $A \in \mathbb{F}^{m\times n}$. *Then* $\operatorname{rank}(A) = r$ *iff* A *is equivalent to* $E_r \in \mathbb{F}^{m\times n}$.

Later, we will give a computational method to construct the matrices P and Q from a given matrix A so that $Q^{-1}AP = E_r$.

Now, taking transpose, we have $A^{\mathrm{t}} = P^{\mathrm{t}} E_r^{\mathrm{t}} (Q^{-1})^{\mathrm{t}}$. That is, A^{t} is equivalent to E_r^{t}. However, E_r^{t} has rank r. It thus follows that $\operatorname{rank}(A^{\mathrm{t}}) = \operatorname{rank}(A) = r$.

We know that the ith column of A is equal to $A(e_i)$. Thus, $R(A)$ is the subspace of $\mathbb{F}^{m\times 1}$ spanned by the columns of A. Then $\operatorname{rank}(A)$ is same as the maximum number of linearly independent columns of A. Then $\operatorname{rank}(A^{\mathrm{t}})$ is same as the maximum number of linearly independent rows of A. We have thus shown that these two numbers are equal. This fact is expressed by asserting that the *column rank* of a matrix and the *row rank* of a matrix are equal. Of course, this also follows from Theorem 3.23.

Further, if A is a square matrix, then $\operatorname{rank}(A^{\mathrm{t}}) = \operatorname{rank}(A)$ implies that A^{t} and A are equivalent matrices.

The rank factorization yields another related factorization. For a matrix $A \in \mathbb{F}^{m\times n}$ of rank r, we have invertible matrices $P \in \mathbb{F}^{n\times n}$ and $Q \in \mathbb{F}^{m\times m}$ such that $A = QE_rP^{-1}$. However, $E_r \in \mathbb{F}^{m\times n}$ can be written as

$$E_r = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r \\ 0 \end{bmatrix} \begin{bmatrix} I_r & 0 \end{bmatrix} \quad \text{with } \begin{bmatrix} I_r \\ 0 \end{bmatrix} \in \mathbb{F}^{m\times r},\ \begin{bmatrix} I_r & 0 \end{bmatrix} \in \mathbb{F}^{r\times n}.$$

Taking

$$B = Q\begin{bmatrix} I_r \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} I_r & 0 \end{bmatrix} P^{-1},$$

we see that $\operatorname{rank}(B) = r = \operatorname{rank}(C)$. We have thus have proved the following theorem.

***Theorem* 4.23 (Full rank factorization)**
Let $A \in \mathbb{F}^{m\times n}$ *be of rank* r. *Then there exist rank* r *matrices* $B \in \mathbb{F}^{m\times r}$ *and* $C \in \mathbb{F}^{r\times n}$ *such that* $A = BC$.

Exercises for § 4.6

1. Show that for any matrix $A \in \mathbb{F}^{m\times n}$, $\operatorname{rank}(A^*) = \operatorname{rank}(A)$.
2. Derive the Rank theorem for matrices from the Rank factorization of a matrix.
3. Show that $\operatorname{rank}(AB) \le \operatorname{rank}(A)$ for all matrices $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$.
4. Is it true that $\operatorname{rank}(AB) \le \operatorname{rank}(B)$ for all matrices $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$?
5. Show that if $A = BC$ is a full rank factorization of A, and D is an invertible matrix of suitable order, then $A = (BD)(D^{-1}C)$ is also a full rank factorization.

6. Let $A = B_1C_1$ and $A = B_2C_2$ be two full rank factorizations of A. Show that there exists an invertible matrix D of appropriate order such that $B_2 = B_1D$ and $C_2 = D^{-1}C_1$.

5

Systems of Linear Equations

5.1 Solvability

We discuss application of linear transformations to the solvability of systems of linear equations.

Let $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ and let $b \in \mathbb{F}^{m\times 1}$. Then $Ax = b$, written explicitly as

$$\begin{array}{ccccc} a_{11}x_1 & + \cdots & +a_{1n}x_n & = b_1 \\ \vdots & & \vdots & \\ a_{m1}x_1 & + \cdots & +a_{mn}x_n & = b_n \end{array}$$

is called a **system of linear equations** with m equations and n unknowns $x_1, \ldots, x_n$, where the coefficients a_{ij} are from $\mathbb{F}$. We abbreviate the phrase 'a system of linear equations' to ' a linear system'. When $b = 0$, the linear system is said to be **homogeneous.**

The **solution set** of the linear system $Ax = b$ is defined as

$$\mathrm{Sol}\,(A, b) = \{x \in \mathbb{F}^{n\times 1} : Ax = b\}.$$

Thus the system $Ax = b$ has a solution iff $\mathrm{Sol}\,(A, b) \neq \emptyset$.

We use the notation $[A|b]$ for the new matrix in $\mathbb{F}^{m\times(n+1)}$ obtained by taking the columns of A in the same order along with b as the $(n+1)$th column. We call $[A|b]$ an **augmented matrix**. The system $Ax = b$ is called **consistent** iff $\mathrm{rank}([A|b]) = \mathrm{rank}(A)$.

Recall that $\mathrm{rank}(A)$ is the dimension of $R(A)$, the reange space of A, which is the subspace of $\mathbb{F}^{m\times 1}$ spanned by the columns of A.

***Theorem* 5.1**
A linear system has a solution iff it is consistent.

Proof Let U be the subspace of $\mathbb{F}^{m\times 1}$ spanned by the columns of $[A|b]$. Notice that $R(A)$ is a subspace of U. Then,

$Ax = b$ has a solution

iff $b = Ax$ for some $x \in \mathbb{F}^{n\times 1}$
iff $b \in R(A)$
iff $R(A) = U$
iff $\text{rank}(A) = \dim(R(A)) = \dim(U) = \text{rank}([A|b])$. ∎

Recall that $N(A)$ is the null space of A, which is equal to the solution set of the homogeneous sytem. That is, $N(A) = \text{Sol}(A, 0)$. We connect $\text{Sol}(A, b)$ and $\text{Sol}(A, 0)$.

***Theorem* 5.2**
If u is a solution of $Ax = b$, then $\text{Sol}(A, b) = u + N(A) = \{u + x : x \in N(A)\}$.

Proof Let $Au = b$. If $v \in \text{Sol}(A, b)$, then $A(v-u) = Av - Au = b - b = 0$. That is, $v - u \in N(A)$. Then, $v \in u + N(A)$.

Conversely, let $x \in u + N(A)$. Then $x = u + y$ for some $y \in N(A)$. That is, $x = u + y$ for some y with $Ay = 0$. Now, $A(x) = A(u + y) = Au + Ay = Au = b$. Therefore, $u + x \in \text{Sol}(A, b)$. ∎

It means that any solution of $Ax = b$ can be obtained by taking a particular solution u of $Ax = b$ and then adding it to any solution of the homogeneous system $Ax = 0$.

As a corollary, we obtain the following result. It shows why a linear system has either no solutions, or a unique solution, or infinitely many solutions.

***Theorem* 5.3**
Let $A \in \mathbb{F}^{m\times n}$, and let $b \in \mathbb{F}^{m\times 1}$. Write $k = \text{null}(A) = n - \text{rank}(A)$.

(1) *The linear system $Ax = b$ has a unique solution iff $k = 0$ and $b \in R(A)$ iff* $\text{rank}([A|b]) = \text{rank}(A) = n$.

(2) *If u is a solution of $Ax = b$ and $\{v_1, \ldots, v_k\}$ is a basis for $N(A)$, then*

$$\text{Sol}(A, b) = \{u + \alpha_1 v_1 + \cdots + \alpha_k v_k : \alpha_1, \ldots, \alpha_k \in \mathbb{F}\}.$$

(3) *For $m = n$, $Ax = b$ has a unique solution iff $det(A) \neq 0$.*

We will discuss later how to compute the solution set of a given linear system.

When a system has a unique solution, a determinant formula can be given for obtaining the solution.

***Theorem* 5.4 (*Cramer's Rule*)**
Let $A \in \mathbb{F}^{n\times n}$ with $\det(A) \neq 0$, *and let $b \in \mathbb{F}^{n\times 1}$. Let $A_i[b]$ denote the matrix obtained from A by replacing its ith column with the vector b. Then the solution of $Ax = b$ is given by* $x_i = \dfrac{\det(A_i[b])}{\det(A)}$ *for* $1 \le i \le n$.

Proof Since $\det(A) \neq 0$, there exists a unique $x \in \mathbb{F}^{n\times 1}$ such that $Ax = b$. Let $x = (x_1, \ldots, x_n)^t$. Write $Ax = b$ as $x_1C_1 + \cdots + x_nC_n = b$, where C_j is the jth column of A. Next, move b to the left side to obtain

$$x_1C_1 + \cdots + (x_iC_i - b) + \cdots + x_nC_n = 0.$$

So, the column vectors $C_1, \ldots, x_iC_i - b_i, \ldots, C_n$ are linearly dependent. Thus, $\det(A_i[x_iC_i - b]) = 0$. Using Properties (9)-(10) of the the determinant, we get

$$\det[C_1, \ldots, x_iC_i, \ldots, C_n] - \det[C_1, \ldots, b_i, \ldots, C_n] = 0.$$

This is same as $x_i\det(A) - \det(A_i[b]) = 0$. ∎

Cramer's rule helps in studying the map $(A, b) \mapsto x$, when $\det(A) \neq 0$.

Exercises for § 5.1

1. Let $A = [a_{ij}]$ be an $m \times n$ matrix with $a_{ij} \in \mathbb{F}$ and $n > m$. Show that there exists $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n$ such that $\alpha_1 a_{i1} + \cdots + \alpha_n a_{in} = 0$ for all $i = 1, \ldots, m$. Interpret the result for linear systems.
2. Prove Theorem 5.3.
3. Suppose a linear homogeneous system $Ax = 0$ has m equations and n unknowns. Show the following:
   (a) If $\operatorname{rank}(A) = n$, then $Ax = 0$ has a unique solution, the trivial solution.
   (b) If $\operatorname{rank}(A) < n$, then $Ax = 0$ has infinitely many solutions.
   (c) If $m \geq n$, then both (a)-(b) are possible.
   (d) If $m < n$ then $Ax = 0$ has infinitely many solutions.
   (e) $Ax = 0$ has infinitely many solutions iff the columns of A are linearly dependent.
4. Let the linear system $Ax = b$ have m equations and n unknowns. Show the following:
   (a) If $\operatorname{rank}([A\,|\,b]) > \operatorname{rank}(A)$, then $Ax = b$ has no solutions.
   (b) If $\operatorname{rank}([A\,|\,b]) = \operatorname{rank}(A) = n$, then $Ax = b$ has a unique solution.
   (c) If $\operatorname{rank}([A\,|\,b]) = \operatorname{rank}(A) < n$, then $Ax = b$ has infinitely many solutions.
   (d) $Ax = b$ has at least one solution iff $b \in \operatorname{span}\{A_1, \ldots, A_n\}$.
   (e) $Ax = b$ has at most one solution iff $A_1, \ldots, A_n$ are linearly independent.
5. Let $A \in \mathbb{F}^{m\times n}$. Let b be a nonzero vector in $\mathbb{F}^{m\times 1}$ orthogonal to each column of A. Show that the linear system $Ax = b$ has no solutions.
6. Prove: If U is a subspace of $\mathbb{F}^n$ and $x \in \mathbb{F}^n$, then there exists a system of linear equations having n equations and n unknowns, with coefficients in $\mathbb{F}$, such that its solution set equals $x + U$.
7. Let $A \in \mathbb{F}^{n\times n}$. Let $U = \{X \in \mathbb{F}^{n\times n} : AX = 0\}$. What is $\dim(U)$ as a subspace of $\mathbb{F}^{n\times n}$?

5.2 Elementary row operations

In general, evaluation of the determinant of a matrix of order more than five is unmanageable. In this case, Cramer's rule cannot be used for computation. We discuss a more efficient method, called Gaussian elimination that uses elementary row operations of matrices.

The elementary row operations are a formalization of the usual way of solving linear systems, albeit, in a systematic way. While solving a system of linear equations, you go on replacing an equation with the sum of that and a multiple of another equation so that some of the unknowns are eliminated. The elementary operations do the same thing operating with the coefficients only so that writing would be easier.

In fact elementary operations can be seen as matrix products with special types of matrices, called elementary matrices. Thus we define elementary matrices as in the following:

An **elementary matrix** of order m is one of the following three types:

1. Type 1: $E[i,j]$ is the matrix obtained from I_m by exchanging the ith and the jth rows.
2. Type 2: $E_\alpha[i]$ is the matrix obtained from I_m by replacing the ith row with α times the ith row.
3. Type 3: $E_\alpha[i,j]$ is the matrix obtained from I_m by replacing the ith row with the ith row plus α times the jth row.

Notice that our symbolism does not reflect the order m of the elementary matrices; it should be understood from the context. An alternative way of presenting elementary matrices in closed form is given in Exercise 2.

Example 5.5
The following are instances of elementary matrices of order 3.

$$E[1,2] = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_{-1}[2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_2[3,1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}. \qquad \square$$

An elementary matrix of the type $E[i,j]$ is its own inverse. The matrix $E_\alpha[i]$ has the inverse $E_{1/\alpha}[i]$. And the inverse of $E_\alpha[i,j]$ is $E_{-\alpha}[i,j]$. Thus each elementary matrix is invertible; it is an isomorphism.

We observe that for a matrix $A \in \mathbb{F}^{m\times n}$, the following are true:

1. $E[i,j]\,A$ is the matrix obtained from A by exchanging its ith and jth rows.

2. $E_\alpha[i]\,A$ is the matrix obtained from A by replacing its ith row with α times the ith row.
3. $E_\alpha[i,j]\,A$ is the matrix obtained from A by replacing its ith row with the ith row plus α times the jth row.

We call each of these operations of pre-multiplying a matrix with an elementary matrix as an **elementary row operation**. In computations, we will write

$$A \xrightarrow{E} B$$

to mean that the matrix B has been obtained by an elementary row operation E; that is, $B = EA$. For instance,

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \xrightarrow{E_{-3}[3,1]} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{E_{-2}[2,1]} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = B.$$

Here, $B = E_{-2}[2,1]\,E_{-3}[3,1]\,A$; the products are in reverse order.

Often we will apply elementary operations in a sequence. In this way, the operations in the above computation could be shown as $E_{-3}[3,1]$, $E_{-2}[2,1]$.

Elementary row operations can be used to convert a matrix to nice forms. Towards this, we introduce some terminology.

The first, from left, nonzero entry in a nonzero row of a matrix is called a **pivot**. We denote a pivot in a row by putting a box around it.

A matrix $B = [b_{ij}] \in \mathbb{F}^{m\times n}$ is said to be in **row echelon form** iff the following conditions are satisfied:

1. For any two pivots b_{ij} and $b_{k\ell}$, if $i < k$ then $j < \ell$.
2. Each zero row has larger row index than that of any nonzero row.

The second condition says that all zero rows are at the bottom. A column of a matrix where a pivot occurs is called a **pivotal column**. Both the conditions imply that in any pivotal column, each entry below a pivot is equal to zero.

Example 5.6

The matrix $\begin{bmatrix} \boxed{1} & 2 & 3 & 0 \\ 0 & \boxed{2} & 0 & 1 \\ 0 & 0 & 0 & \boxed{5} \end{bmatrix}$ is in row echelon form but $\begin{bmatrix} 0 & \boxed{2} & 3 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{5} & 1 \end{bmatrix}$ is not in row echelon form. $\square$

Any matrix can be brought to row echelon form by using elementary row operations. In fact, we give an algorithm to do it. In the algorithm, we mark a specific sub-matrix of A as our search region R. This search region varies as the algorithm proceeds. In the algorithm the pivots are selected in stages; accordingly, a row where a pivot occurs is called a *pivotal row*.

Reduction to Row Echelon Form

1. Set the search region R to the whole matrix A.
2. If all entries in R are 0 or R is empty, then stop.
3. Else, let a_{ij} be the left most and top most entry of R and let $a_{k\ell}$ be the left most and top most nonzero entry of R.
4. If $i \neq k$, then update A by interchanging the ith and the kth rows of A.
5. In the updated matrix A, mark the (i, ℓ)th entry as the pivot; call the ith row as a pivotal row, and the ℓth column as a pivotal column.
6. Zero-out all entries in the pivotal column below the pivot by replacing each row below the pivotal row with one obtained by a suitable Type 3 elementary row operation using that row and the pivotal row.
7. Reset the search region R to the sub-matrix of R to the right and below the current pivot. Go to Step 2.

We call the matrix obtained from a given matrix A using the above algorithm as *its row echelon form*.

In the following, we show how A is reduced to its row echelon form B:

$$A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 3 & 5 & 5 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 3 & 5 & 5 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \xrightarrow{R1} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix}$$

$$\xrightarrow{R2} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & \boxed{2} & 1 & 1 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \end{bmatrix} \xrightarrow{R3} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & \boxed{2} & 1 & 1 \\ 0 & 0 & 0 & \boxed{3} \\ 0 & 0 & 0 & 0 \end{bmatrix} = B.$$

Here, $R1 = E_3[2,1], E_1[3,1], E_2[4,1]$; $R2 = E_2[3,2], E_3[4,2]$; $R3 = E_2[4,3]$. And, $B = E_2[4,3]\,E_3[4,2]\,E_2[3,2]\,E_2[4,1]\,E_1[3,1]\,E_3[2,1]\,A$.

Recall that the row rank of a matrix A is the maximum number of linearly independent rows of A, which is equal to the rank of A.

***Theorem* 5.7**

The rank of a matrix is the number of pivots in its row echelon form.

Proof Let A be a given matrix and let B be its row echelon form. Each elementary matrix is an isomorphism; and isomorphisms preserve rank. Thus $\text{rank}(A) = \text{rank}(B) =$ the row rank of $B =$ the number of pivots in B. ∎

Thus, a row echelon form of a matrix shows its rank.

Example 5.8

Look at the following reduction to row echelon form:

$$\begin{bmatrix} \boxed{1} & 1 & 0 & 1 \\ 0 & 1 & 1 & -1 \\ 2 & -1 & -3 & 5 \end{bmatrix} \xrightarrow{E_{-2}[3,1]} \begin{bmatrix} \boxed{1} & 1 & 0 & 1 \\ 0 & \boxed{1} & 1 & -1 \\ 0 & -3 & -3 & 3 \end{bmatrix} \xrightarrow{E_3[3,2]} \begin{bmatrix} \boxed{1} & 1 & 0 & 1 \\ 0 & \boxed{1} & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence, the rank of the original matrix is 2. ∎

Similar to elementary row operations, **Elementary Column Operations** are of the following three types:

1. Type 1: Interchange of two columns;
2. Type 2: Multiplication of a column by a nonzero scalar;
3. Type 3: Adding to a column a nonzero scalar multiple of another column.

Elementary column operations are achieved by post-multiplying a matrix with the transpose of the corresponding elementary matrix. Alternatively, we take the transpose of a matrix, apply elementary row operations, and then take transpose of the result. Thus we discuss elementary row operations in detail, and revert to elementary column operations, when necessary.

Exercises for § 5.2

1. Convert the following matrices into their row echelon forms, and then determine their ranks:

$$\begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{bmatrix}, \begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 & 0 & 1 & 1 \\ 3 & 5 & 5 & 1 & 1 & 2 \\ 4 & 6 & 7 & 1 & 2 & 3 \\ 1 & 5 & 0 & 5 & 1 & 2 \end{bmatrix}.$$

2. Recall that the matrix E_{ij} of order m is the $m \times m$ matrix whose (i, j)th entry is 1 and all other entries are 0. Show that the elementary matrices of order m can be given by the following:
   (a) $E[i, j] = I_m - E_{ii} - E_{jj} + E_{ij} + E_{ji}$ with $i \neq j$.
   (b) $E_\alpha[i] = I_m - E_{ii} + \alpha E_{ii}$, where α is a nonzero scalar.
   (c) $E_\alpha[i, j] = I_m + \alpha E_{ij}$, where α is a nonzero scalar and $i \neq j$.
3. Show that the matrices $E[i, j]A$, $E_\alpha[i]A$ and $E_\alpha[i, j]A$ match the corresponding matrices obtained from A by performing the intended row operations.
4. Show that $A(E[i, j])^t$, $A(E_\alpha[i])^t$ and $A(E_\alpha[i, j])^t$ are the matrices obtained from A by using elementary column operations of the corresponding type.
5. Show that $E[i, j]\, E_\alpha[i] = E_\alpha[j]\, E[i, j]$. What about other such products?

5.3 Row reduced echelon form

There is another simplified form of a matrix, though it involves a bit more work. A matrix $A \in \mathbb{F}^{m\times n}$ is said to be in **row reduced echelon form** (RREF) iff the following conditions are satisfied:

1. Each pivot is equal to 1.
2. For any two pivots a_{ij} and $a_{k\ell}$, if $i < k$ then $j < \ell$.
3. In a pivotal column, all entries other than the pivot are zero.
4. Each zero row has larger row index than that of any nonzero row.

Example 5.9

The matrix $\begin{bmatrix} \boxed{1} & 2 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$ is in row reduced echelon form whereas

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

are not in row reduced echelon form. $\square$

Observe that a matrix in RREF is also in row echelon form satisfying the added constraints that each pivotal column is equal to e_j for some j.

Any matrix can be brought to a matrix in RREF by using elementary row operations. We modify the algorithm for reduction to row echelon form by replacing its Stpe 6 with Steps 6A-6B as in the following:

Reduction to RREF

1-5. As in the Algorithm: Reduction to Row Echelon Form.

6A. Multiply the current pivotal row by the reciprocal of the pivot.

6B. Zero-out all entries except the pivot in the pivotal column by replacing each non-pivotal row with one obtained by a suitable Type 3 elementary row operation using that row and the pivotal row.

7. As in the Algorithm: Reduction to Row Echelon Form.

We will refer to the output of the above reduction algorithm as *the row reduced echelon form* (the RREF) of a given matrix.

Example 5.10
In the following the matrix A is reduced to its RREF:

$$A = \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 3 & 5 & 7 & 1 \\ 1 & 5 & 4 & 5 \\ 2 & 8 & 7 & 9 \end{bmatrix} \xrightarrow{R1} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix} \xrightarrow{E_{1/2}[2]} \begin{bmatrix} \boxed{1} & 1 & 2 & 0 \\ 0 & \boxed{1} & 1/2 & 1/2 \\ 0 & 4 & 2 & 5 \\ 0 & 6 & 3 & 9 \end{bmatrix}$$

$$\xrightarrow{R2} \begin{bmatrix} \boxed{1} & 0 & 3/2 & -1/2 \\ 0 & \boxed{1} & 1/2 & 1/2 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \end{bmatrix} \xrightarrow{E_{1/3}[3]} \begin{bmatrix} \boxed{1} & 0 & 3/2 & -1/2 \\ 0 & \boxed{1} & 1/2 & 1/2 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

$$\xrightarrow{R3} \begin{bmatrix} \boxed{1} & 0 & 3/2 & 0 \\ 0 & \boxed{1} & 1/2 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} = B.$$

Here, $R1 = E_{-3}[2,1], E_{-1}[3,1], E_{-2}[4,1]$; $R2 = E_{-1}[2,1], E_{-4}[3,2], E_{-6}[4,2]$; and $R3 = E_{1/2}[1,3], E_{-1/2}[2,3], E_{-6}[4,3]$. The matrix B is the RREF of A. Notice that

$$\begin{aligned} B &= E_{-6}[4,3]\,E_{-1/2}[2,3]\,E_{1/2}[1,3]\,E_{-1/3}[3]\,E_{-6}[4,2]\,E_{-4}[3,2]\,E_{-1}[2,1] \\ &\quad E_{-1/2}[2]\,E_{-2}[4,1]\,E_{-1}[3,1]\,E_{-3}[2,1]\,A. \end{aligned}$$

The products are in reverse order. ∎

We observe the following connections between linear combinations, linear dependence and linear independence of rows and columns of a matrix, and its RREF.

Observation 5.1 In the RREF of A suppose $R_{i1}, \ldots, R_{ir}$ are the rows of A which have become the nonzero rows in the RREF, and other rows have become the zero rows. Also, suppose $C_{j1}, \ldots, C_{jr}$ for $j_1 < \cdots < j_r$, are the columns of A which have become the pivotal columns in the RREF, other columns being non-pivotal. Then the following are true:

1. The rows $R_{i1}, \ldots, R_{ir}$ are linearly independent; and the other rows of A are linear combinations of $R_{i1}, \ldots, R_{ir}$.
2. The columns $C_{j1}, \ldots, C_{jr}$ have respectively become $e_1, \ldots, e_r$ in the RREF.
3. The columns $C_{j1}, \ldots, C_{jr}$ are linearly independent; and other columns of A are linear combinations of $C_{j1}, \ldots, C_{jr}$.
4. If $e_1, \ldots, e_k$ are all the pivotal columns in the RREF that occur to the left of a non-pivotal column, then the non-pivotal column is in the form $(a_1, \ldots, a_k, 0, \ldots, 0)^T$. Further, if a column C in A has become this non-pivotal column in the RREF, then $C = a_1 C_{j1} + \cdots + a_k C_{jk}$.

For instance, in Example 5.10, the third column in the RREF is given by

$$\tfrac{3}{2}(\text{first pivotal column}) + \tfrac{1}{2}(\text{the second pivotal column}).$$

Also,

> the third column of A is $\frac{3}{2}$ times the first column plus $\frac{1}{2}$ times the second column.

Recall the rank factorization, which states that a matrix A of rank r is equivalent to the matrix E_r of the same size, where

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

With the help of elementary row operations, we can actually construct the matrices P and Q so that $A = QE_rP^{-1}$.

Let $A \in \mathbb{F}^{m\times n}$ have rank r. Convert A to its RREF $C = EA$, where E is the suitable product of elementary matrices. Now, C has r number of pivots. The pivotal columns are $e_1, \ldots, e_r$; and the rows below the rth row in C are all zero rows. Therefore, each non-pivotal column is a linear combination of the r pivotal columns. Then, we use column operations to reduce the matrix EA to $E_r = EA\hat{E}^{\mathrm{t}}$.

Instead of the column operations, we may use the row operations on the transpose of the matrix $C = EA$. That is, we consider the matrix C^{t}. The pivotal columns of C are now the pivotal rows e_i^{t} in C^{t}. Each other row in C^{t} is a liner combination of these pivotal rows. Exchange the rows of C^{t} so that the first r rows of the new C^{t} are $e_1^{\mathrm{t}}, \ldots, e_r^{\mathrm{t}}$ in that order. Use suitable elementary row operations to zero-out all non-pivotal rows. This is possible since each non-pivotal row is a linear combination of $e_1^{\mathrm{t}}, \ldots, e_r^{\mathrm{t}}$. We obtain the matrix $\hat{E}C^{\mathrm{t}}$, where $\hat{E}$ is the suitable product of elementary matrices such that $\hat{E}C^{\mathrm{t}}$ has first r rows as $e_1^{\mathrm{t}}, \ldots, e_r^{\mathrm{t}}$ and all other rows are zero rows. Then taking transpose, we see that $C\hat{E}^{\mathrm{t}}$ is a matrix whose first r columns are $e_1, \ldots, e_r$ and all other columns are zero columns.

To summarize, we have obtained two matrices E and $\hat{E}$, which are products of elementary matrices such that

$$C\hat{E}^{\mathrm{t}} = EA\hat{E}^{\mathrm{t}} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = E_r.$$

Or that, $A = E^{-1}E_r(\hat{E}^{\mathrm{t}})^{-1}$.

Example 5.11

In Example 5.10, we had obtained B as the RREF of A. We illustrate how to convert B to E_r using column operations. For column operations, we use the transpose notation of the elementary matrices, which are intended to be multiplied on the right.

$$B \quad = \quad \begin{bmatrix} \boxed{1} & 0 & 3/2 & 0 \\ 0 & \boxed{1} & 1/2 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{E^{\mathrm{t}}_{-3/2}[3,1]} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 1/2 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{E^{\mathrm{t}}_{-1/2}[3,2]} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{E^{\mathrm{t}}[3,4]} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = E_3.$$

This computation says that $B\, E^{\mathrm{t}}_{-3/2}[3,1]\, E^{\mathrm{t}}_{-1/2}[3,2]\, E^{\mathrm{t}}[3,4] = E_3$. ∎

Go back to Section 1.8. With the help of reduction to RREF, you can extract bases for the span of given vectors in $\mathbb{F}^n$. For this purpose, you may use the row vectors as they are given, or take the transposes of those and work with the column vectors. Working with columns vectors has an advantage; the RREF shows how to express the linearly dependent columns as linear combinations of the linearly independent ones.

Exercises for § 5.3

1. Convert the following matrices into their row reduced echelon form:

$$\begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{bmatrix}, \begin{bmatrix} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 & 0 & 1 & 1 \\ 3 & 5 & 5 & 1 & 1 & 2 \\ 4 & 6 & 7 & 1 & 2 & 3 \\ 1 & 5 & 0 & 5 & 1 & 2 \\ 2 & 8 & 1 & 6 & 0 & 2 \end{bmatrix}.$$

2. Prove that a square matrix is invertible iff it is a product of elementary matrices.
3. Let $A \in \mathbb{F}^{n\times n}$ and let $E \in \mathbb{C}^{n\times n}$ be an elementary matrix. Show that $\det(EA) = \det(E)\det(A)$. Use this to prove that $\det(AB) = \det(A)\det(B)$ for matrices $A, B \in \mathbb{F}^{n\times n}$.
4. Call two matrices A and B of the same size to be *row equivalent* iff one can be brought to the other by elementary row operations. Show that A and B are row equivalent iff there exists an invertible matrix P such that $B = PA$.
5. Define *column equivalence* of two matrices analogous to row equivalence. Formulate and answer a similar question as in the previous exercise.
6. Is it true that two matrices are row equivalent iff they are column equivalent?

7. Let B and C be two $m \times n$ matrices in RREF. Show that if C can be obtained from B by using elementary row operations, then $C = B$.
8. Use the previous exercise to show that if B and C are obtained from A by elementary row operations and if B and C are in RREF, then $B = C$.
9. Let $A \in \mathbb{F}^{m \times n}$. Prove the following:
   (a) The rows of A are linearly independent iff A is an onto map iff for each $b \in \mathbb{F}^{m \times 1}$, the linear system $Ax = b$ has at least one solution.
   (b) The columns of A are linearly independent iff A is a one-one map iff for each $b \in \mathbb{F}^{m \times 1}$, the linear system $Ax = b$ has at most one solution.
10. Let $A \in \mathbb{F}^{m \times n}$ be of rank r, and let R be the RREF of A. Construct $B \in \mathbb{F}^{m \times r}$ by deleting from A the non-pivotal columns; and construct $C \in \mathbb{F}^{r \times n}$ from R by deleting the $n - r$ zero rows at the bottom of R. Show that $A = BC$ is a full rank factorization of A.

5.4 Gaussian and Gauss-Jordan elimination

Gaussian elimination is an application of converting the augmented matrix to its row echelon form for solving linear systems. Similarly, Gauss-Jordan elimination applies conversion of the augmented matrix to its RREF. We now discuss these systematic approaches to solving systems of linear equations.

***Theorem* 5.12**
Let $[A'|b']$ *be an augmented matrix obtained from the augmented matrix* $[A|b]$ *by elementary row operations. Then* $\mathrm{Sol}\,(A, b) = \mathrm{Sol}\,(A', b')$.

Proof Let B be an invertible matrix. If x satisfies $Ax = b$, then x also satisfies $BAx = Bb$, and vice-versa. That is, $\mathrm{Sol}\,(A, b) = \mathrm{Sol}\,(BA, Bb)$. Since each elementary matrix is invertible, and elementary operations amount to pre-multiplying both A and b with a product of elementary matrices, we see that $\mathrm{Sol}\,(A, b) = \mathrm{Sol}\,(A', b')$. ∎

In **Gaussian elimination** method for solving a linear system $Ax = b$, we start with the augmented matrix $[A|b]$. Using elementary row operations, we bring $[A|b]$ to its row echelon form $[A'|b']$. If a pivot occurs in b', then $\mathrm{rank}([A|b]) > \mathrm{rank}(A)$. Hence the system is inconsistent.

Otherwise, omitting the zero rows in $[A'|b']$, the system is written with the unknowns. The pivoted entries correspond to the **basic variables** and the other unknowns are taken as **free variables**. Using back-substitution, the basic variables are expressed in terms of the free variables. Then the free variables are given arbitrary values, in symbols, α_i leading to the solution set $\mathrm{Sol}\,(A,b)$.

Example 5.13
Consider the following linear system:

$$
\begin{aligned}
5x_1 + 2x_2 - 3x_3 + x_4 &= 7\\
x_1 - 3x_2 + 2x_3 - 2x_4 &= 11\\
3x_1 + 8x_2 - 7x_3 + 5x_4 &= 8
\end{aligned}
$$

We take the augmented matrix and reduce it to its row echelon form.

$$
\left[\begin{array}{rrrr|r} 5 & 2 & -3 & 1 & 7\\ 1 & -3 & 2 & -2 & 11\\ 3 & 8 & -7 & 5 & 8 \end{array}\right]
\xrightarrow{R1}
\left[\begin{array}{rrrr|r} \boxed{5} & 2 & -3 & 1 & 7\\ 0 & -17/5 & 13/5 & -11/5 & 48/5\\ 0 & 34/5 & -26/5 & 22/5 & 19/5 \end{array}\right]
$$

$$
\xrightarrow{R2}
\left[\begin{array}{rrrr|r} \boxed{5} & 2 & -3 & 1 & 7\\ 0 & \boxed{-17/5} & 13/5 & -11/15 & -48/5\\ 0 & 0 & 0 & 0 & \boxed{23} \end{array}\right].
$$

Here, $R_1 = E_{-1/5}[2,1]$, $E_{-3/5}[3,1]$ and $R2 = E_2[3,2]$. Since an entry in the b portion has become a pivot, the system is inconsistent.

Indeed, subtracting twice the second equation from the first, we obtain

$$3x_1 + 8x_2 - 7x_3 + 5x_4 = -15.$$

The third equation says that the same quantity should have been 8. Therefore, the system has no solutions. $\square$

Example 5.14
We change the last equation in the previous example to make it consistent. The only change is in the third component of b. The system now looks like:

$$
\begin{aligned}
5x_1 + 2x_2 - 3x_3 + x_4 &= 7\\
x_1 - 3x_2 + 2x_3 - 2x_4 &= 11\\
3x_1 + 8x_2 - 7x_3 + 5x_4 &= -15.
\end{aligned}
$$

Our goal is to solve the current linear system by using Gaussian elimination. The reduction to row echelon form of the augmented matrix uses the same row operations $R_1 = E_{-1/5}[2,1]$, $E_{-3/5}[3,1]$ and $R2 = E_2[3,2]$. The b-entry in the last row changes as follows:

$$\left[\begin{array}{cccc|c} 5 & 2 & -3 & 1 & 7 \\ 1 & -3 & 2 & -2 & 11 \\ 3 & 8 & -7 & 5 & 8 \end{array}\right] \xrightarrow{R1} \left[\begin{array}{cccc|c} \boxed{5} & 2 & -3 & 1 & 7 \\ 0 & -17/5 & 13/5 & -11/5 & 48/5 \\ 0 & 34/5 & -26/5 & 22/5 & -96/5 \end{array}\right]$$

$$\xrightarrow{R2} \left[\begin{array}{cccc|c} \boxed{5} & 2 & -3 & 1 & 7 \\ 0 & \boxed{-17/5} & 13/5 & -11/15 & 48/5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This expresses the fact that the third equation is redundant. Now, solving the new system in row echelon form is easier. We can back-substitute. Writing as linear equations, we have

$$\begin{aligned} \boxed{5}\,x_1 + 2x_2 - 3x_3 + x_4 &= 7 \\ \boxed{-\tfrac{17}{5}}\,x_2 + \tfrac{13}{5}x_3 - \tfrac{11}{5}x_4 &= \tfrac{48}{5}. \end{aligned}$$

In this reduced system, the basic variables are x_1 and x_2; and the unknowns x_3, x_4 are free variables. We assign x_3 to α_3 and x_4 to α_4. The last equation gives

$$\boxed{-\tfrac{17}{5}}\,x_2 = \tfrac{48}{5} - \tfrac{13}{5}\alpha_3 + \tfrac{11}{5}\alpha_4.$$

That is,

$$x_2 = -\tfrac{48}{17} + \tfrac{13}{17}\alpha_3 - \tfrac{11}{17}\alpha_4.$$

Substituting this in the first equation, we obtain

$$\boxed{5}\,x_1 = 7 - 2x_2 + 3x_3 - x_4 = \tfrac{225}{17} + \tfrac{25}{17}\alpha_3 + \tfrac{5}{17}\alpha_4.$$

That is,

$$x_1 = \tfrac{45}{17} + \tfrac{5}{17}\alpha_3 + \tfrac{1}{17}\alpha_4.$$

Writing in vector form, the solution set looks like

$$\left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 45/17 \\ -48/17 \\ 0 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 13/17 \\ 5/17 \\ 1 \\ 0 \end{bmatrix} + \alpha_4 \begin{bmatrix} 1/17 \\ -11/17 \\ 0 \\ 1 \end{bmatrix} \;:\; \alpha_3, \alpha_4 \text{ arbitrary} \right\}. \qquad \square$$

Example 5.15

Consider solving the following system of linear equations:

$$\begin{array}{rcrcrcrcrl} x_1 & + & x_2 & + & 2x_3 & + & & & x_5 & = 1 \\ 3x_1 & + & 5x_2 & + & 5x_3 & + & x_4 & + & x_5 & = 2 \\ 4x_1 & + & 6x_2 & + & 7x_3 & + & x_4 & + & 2x_5 & = 3 \\ x_1 & + & 5x_2 & + & & & 5x_4 & + & x_5 & = 2 \\ 2x_1 & + & 8x_2 & + & x_3 & + & 6x_4 & + & 0x_5 & = 2. \end{array}$$

We reduce the augmented matrix to its row echelon form as follows.

$$\left[\begin{array}{ccccc|c} 1 & 1 & 2 & 0 & 1 & 1 \\ 3 & 5 & 5 & 1 & 1 & 2 \\ 4 & 6 & 7 & 1 & 2 & 3 \\ 1 & 5 & 0 & 5 & 1 & 2 \\ 2 & 8 & 1 & 6 & 0 & 2 \end{array}\right] \xrightarrow{R1} \left[\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 0 & 1 & 1 \\ 0 & 2 & -1 & 1 & -2 & -1 \\ 0 & 2 & -1 & 1 & -2 & -1 \\ 0 & 4 & -2 & 5 & 0 & 1 \\ 0 & 6 & -3 & 6 & -2 & 0 \end{array}\right]$$

$$\xrightarrow{R2} \left[\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 0 & 1 & 1 \\ 0 & \boxed{2} & -1 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 4 & 3 \\ 0 & 0 & 0 & 3 & 4 & 3 \end{array}\right] \xrightarrow{E[3,4]} \left[\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 0 & 1 & 1 \\ 0 & \boxed{2} & -1 & 1 & -2 & -1 \\ 0 & 0 & 0 & \boxed{3} & 4 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 4 & 3 \end{array}\right]$$

$$\xrightarrow{R3} \left[\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 0 & 1 & 1 \\ 0 & \boxed{2} & -1 & 1 & -2 & -1 \\ 0 & 0 & 0 & \boxed{3} & 4 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Here, $R1 = E_{-3}[2,1], E_{-4}[3,1], E_{-1}[4,1], E_{-2}[5,1]$; $R2 = E_{-1}[3,2], E_{-2}[4,2], E_{-3}[5,2]$; and $R3 = E_{-1}[5,3]$.

The equations now look like

$$\begin{array}{rrrrrcr} x_1 & +x_2 & +2x_3 & & +x_5 & = & 1 \\ & 2x_2 & -x_3 & +x_4 & -2x_5 & = & -1 \\ & & & 3x_4 & +4x_5 & = & 3. \end{array}$$

The basic variables are x_1, x_2, x_4 and the free variables are x_3 and x_5. Using back substitution, we express basic variables in terms of the free variables:

$$\begin{aligned} x_4 &= 1 - \tfrac{4}{3}x_5 \\ x_2 &= \tfrac{1}{2}\big(-1 + x_3 - x_4 + 2x_5\big) = -1 + \tfrac{1}{2}x_3 + \tfrac{5}{3}x_5 \\ x_1 &= 1 - x_2 - 2x_3 + x_5 = 2 - \tfrac{5}{2}x_3 - \tfrac{8}{3}x_5. \end{aligned}$$

The free variables x_3 and x_5 can take any arbitrary values. So, the solution set is given by

$$\mathrm{Sol}\,(A,b) = \left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} -5/2 \\ 1/2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_5 \begin{bmatrix} -8/3 \\ 5/3 \\ 0 \\ -4/3 \\ 1 \end{bmatrix} : \alpha_3, \alpha_5 \in \mathbb{F} \right\}.$$

Here, we have taken $x_3 = \alpha_3$ and $x_5 = \alpha_5$, which can be any scalar in $\mathbb{F}$. ∎

In **Gauss-Jordan elimination**, we reduce a matrix to its RREF instead of row echelon form. Then, we do not require the back substitution phase. Directly the solution can be written down from the RREF of the augmented matrix. We illustrate this method for the last example.

Example 5.16
Consider solving the linear system in Example 5.15. In Gauss-Jordan elimination, the reduction of the augmented matrix to RREF goes as follows:

$$\left[\begin{array}{ccccc|c} 1 & 1 & 2 & 0 & 1 & 1 \\ 3 & 5 & 5 & 1 & 1 & 2 \\ 4 & 6 & 7 & 1 & 2 & 3 \\ 1 & 5 & 0 & 5 & 1 & 2 \\ 2 & 8 & 1 & 6 & 0 & 2 \end{array}\right] \xrightarrow{R1} \left[\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 0 & 1 & 1 \\ 0 & 2 & -1 & 1 & -2 & -1 \\ 0 & 2 & -1 & 1 & -2 & -1 \\ 0 & 4 & -2 & 5 & 0 & 1 \\ 0 & 6 & -3 & 6 & -2 & 0 \end{array}\right]$$

$$\xrightarrow{R2} \left[\begin{array}{ccccc|c} \boxed{1} & 0 & 5/2 & -1/2 & 2 & 3/2 \\ 0 & \boxed{1} & -1/2 & 1/2 & -1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 4 & 3 \\ 0 & 0 & 0 & 3 & 4 & 3 \end{array}\right]$$

$$\xrightarrow{R3} \left[\begin{array}{ccccc|c} \boxed{1} & 0 & 5/2 & 0 & 8/3 & 2 \\ 0 & \boxed{1} & -1/2 & 0 & -1/3 & -1 \\ 0 & 0 & 0 & \boxed{1} & 4/3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Here, $R1 = E_{-3}[2,1],\ E_{-4}[3,1],\ E_{-1}[4,1],\ E_{-2}[5,1]$;
$R2 = E_{1/2}[2],\ E_{-1}[1,2],\ E_{-2}[3,2],\ E_{-4}[4,2],\ E_{-6}[5,2]$; and
$R3 = E[3,4,\ ]; E_{1/2}[1,3],\ E_{-1/2}[2,3],\ E_{-3}[5,3]$.

The equations now look like

$$\begin{array}{rrrrrl} x_1 & & +\frac{5}{2}x_3 & & +\frac{8}{3}x_5 & = \ \ 2 \\ & x_2 & -\frac{1}{2}x_3 & & -\frac{1}{3}x_5 & = -1 \\ & & & x_4 & +\frac{4}{3}x_5 & = \ \ 1. \end{array}$$

The basic variables are x_1, x_2, x_4 and the free variables are x_3 and x_5. Assigning the free avraibales arbitrary values, say, $x_3 = \alpha$ and $x_5 = \beta$, we have

$$\begin{aligned} x_1 &= 2 - \tfrac{5}{2}\alpha - \tfrac{8}{3}\beta \\ x_2 &= -1 + \tfrac{1}{2}\alpha + \tfrac{1}{3}\beta \\ x_3 &= \alpha \\ x_4 &= 1 - \tfrac{4}{3}\beta \\ x_5 &= \beta. \end{aligned}$$

Hence the solution set is

$$\mathrm{Sol}\,(A,b) = \left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -5/2 \\ 1/2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -8/3 \\ 5/3 \\ 0 \\ -4/3 \\ 1 \end{bmatrix} : \alpha, \beta \in \mathbb{F} \right\}.$$

In fact, you can write the solution set from the RREF of the augmented matrix quite mechanically instead of rewriting as a set of equations. $\square$

Exercises for § 5.4

1. Using Gaussian elimination, determine whether the following systems of linear equations are consistent. If consistent, then find the solution set. Next, use Gauss-Jordan elimination to do the same.

   (a) $x_1 - x_2 + 2x_3 - 3x_4 = 7$, $4x_1 + 3x_3 + x_4 = 9$, $2x_1 - 5x_2 + x_3 = -2$, $3x_1 - x_2 - x_3 + 2x_4 = -2$.

   (b) $x_1 - x_2 + 2x_3 - 3x_4 = 7$, $4x_1 + 3x_3 + x_4 = 9$, $2x_1 - 5x_2 + x_3 = -2$, $3x_1 - x_2 - x_3 + 2x_4 = -2$.

2. Using Gaussian Elimination, and also Gauss-Jordan elimination, find all possible values of $k \in \mathbb{R}$ such that the following system of linear equations has more than one solution:

$$x + y + 2z - 5w = 3,\ 2x + 5y - z - 9w = -3,$$
$$x - 2y + 6z - 7w = 7,\ 2x + 2y + 2z + kw = -4.$$

3. Determine the values of $k \in \mathbb{R}$ so that the system of linear equations

$$x + y - z = 1,\ 2x + 3y + kz = 3,\ x + ky + 3z = 2$$

   has (a) no solution, (b) infinitely many solutions, (c) exactly one solution.

4. Let $A \in \mathbb{F}^{m\times n}$ and let $b \in \mathbb{F}^{m\times 1}$. Write an algorithm to obtain $\mathrm{Sol}\,(A,b)$ from the RREF of $[A|b]$. [Hint: Look at Example 5.16 and then think what to do when $m \neq n$.]

5. Discuss the number of solutions of the linear system

$$x + 2y + 3z = 1,\ x - ay + 21z = -2,\ 3x + 7y + az = b$$

   for all possible values of the scalars a and b.

6. By using Gaussian elimination, determine conditions on the scalars a and b so that the vectors $(1,a,4)$, $(1,3,1)$ and $(0,2,b)$ are linearly dependent. [Hint: Write the vectors as columns of a matrix A and use Gaussian elimination on $[A|0]$.]

5.5 Least squares solution

In applications, we often neglect small parameters to arrive at a model. Also, reading of instruments is never exact. These considerations lead to inexact data. Suppose that after admitting such inaccuracies in our model we reached at a linear system of equations, which is inconsistent. How do we solve such a linear system for a required solution?

If x is any suggested solution to the equation $Ax = b$, then we look at the *residual* $\|Ax - b\|$. If x is indeed a solution of $Ax = b$, then the residual is equal to 0. Thus we take up the heuristic of minimizing the residual for arriving at a vector which may be best among the suggested solutions.

Let $T : U \to V$ be a linear transformation, where U is a vector space and V is an ips. A vector $u \in U$ is called a **least squares solution** of the equation $Tx = y$ iff $\|Tu - y\| \le \|Tw - y\|$ for all $w \in U$.

Notice that if $u \in U$ is a least squares solution of $Tx = y$, then Tu is the best approximation of y from $R(T)$. Thus a least squares solution of $Tx = y$ is also called a *best approximate solution*.

To determine a least squares solution of $Tx = y$, we need to determine the best approximation v of y from $R(T)$. And then we would find an appropriate $u \in U$ which satisfies $Tu = v$. Of course, $v \in R(T)$ guarantees that there exists such a $u \in U$. This strategy along with Theorems 2.15-2.16 yield the following result.

***Theorem* 5.17**
Let $T : U \to V$ be a linear transformation, where U is a subspace of an ips V. Let $y \in V$. Then the following are true:

(1) *If $R(T)$ is finite dimensional, then $Tx = y$ has a least squares solution.*

(2) *A vector $u \in U$ is a least squares solution of $Tx = y$ iff $Tu - y \perp z$ for each $z \in R(T)$.*

(3) *A least squares solution is unique iff T is one-one.*

In case of matrices, we have the following simplification.

***Theorem* 5.18**
Let $A \in \mathbb{F}^{m\times n}$, and let $b \in \mathbb{F}^{m\times 1}$. A vector $u \in \mathbb{F}^{n\times 1}$ is a least squares solution of the system of linear equations $Ax = b$ iff $A^*Au = A^*b$.

Proof Let $u_1, \ldots, u_n$ be the columns of A. These vectors span $R(A)$. Using Theorem 5.17, we see that

u is a least squares solution of $Ax = b$

iff $\langle Au - b, u_i \rangle = 0$ for $i = 1, \ldots, n$
iff $u_i^*(Au - b) = 0$ for $i = 1, \ldots, n$
iff $A^*(Au - b) = 0$
iff $A^*Au = A^*b$. ∎

Example 5.19

Suppose $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $u = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

We see that $A \in \mathbb{R}^{2\times 2}$ and $A^{\mathrm{t}}Au = A^{\mathrm{t}}b$. Therefore, u is a least squares solution of $Ax = b$. Notice that $Ax = b$ has no solution. □

A least squares solution can be written in a simplified form by using the QR-factorization, which stems from Gram-Schmidt orthogonalization. To see this, we first present this matrix factorization.

The **QR-factorization** of a matrix $A \in \mathbb{F}^{m\times n}$ is the determination of a matrix $Q \in \mathbb{F}^{m\times n}$ with orthonormal columns, and an upper triangular matrix $R \in \mathbb{F}^{n\times n}$ such that $A = QR$.

***Theorem* 5.20**

Each matrix with linearly independent columns has a QR-factorization, where R is invertible. Consequently, $R = Q^*A$.

Proof Let $u_1, \ldots, u_n$ be the columns of $A \in \mathbb{F}^{m\times n}$. Suppose the columns are linearly independent. It ensures that $m \geq n$. Use Gram-Schmidt process and orthonormalize to obtain the orthonormal vectors $v_1, \ldots, v_n$. We know that for each $k \in \{1, \ldots, n\}$, $\operatorname{span}\{u_1, \ldots, u_k\} = \operatorname{span}\{v_1, \ldots, v_k\}$.

In particular, $u_k \in \operatorname{span}\{v_1, \ldots, v_k\}$. Hence there exist scalars a_{ij} such that the following equalities hold:

$$\begin{aligned} u_1 &= a_{11}v_1 \\ u_2 &= a_{12}v_1 + a_{22}v_2 \\ &\vdots \\ u_n &= a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{nn}v_n. \end{aligned}$$

Since the vectors $u_1, \ldots, u_n$ are linearly independent, the scalars $a_{11}, \ldots, a_{nn}$ are nonzero. Put $a_{ij} = 0$ for $i > j$. Write $R = [a_{ij}]$ and $Q = [v_1, \cdots, v_n]$. Then the above equalities give

$$A = [u_1, \cdots, u_n] = QR.$$

Here, $Q \in \mathbb{F}^{m\times n}$ has orthonormal columns; and $R \in \mathbb{F}^{n\times n}$ is upper triangular. Moreover, R is invertible since the diagonal entries a_{ii} in R are nonzero.

Moreover, the inner product in $\mathbb{F}^{n\times 1}$ is given by $\langle u, v\rangle = v^* u$. Therefore, Q has orthonormal columns means that $Q^*Q = I$. Then $QR = A$ implies that $R = Q^*A$. ∎

Notice that $Q^*Q = I$ does not imply that QQ^* is I, in general. In case A is a square matrix, $QQ^* = I$ and thus Q is unitary.

Example 5.21

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$. Gram-Schmidt process on the columns of A followed by orthonormalization yields the following:

$$w_1 = u_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \|w_1\|^2 = w_1^{\rm t} w_1 = 2, \quad v_1 = \frac{1}{\|w_1\|} w_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}.$$

$$w_2 = u_2 - (v_1^{\rm t} u_2) v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad v_2 = \frac{1}{\|w_2\|} w_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Therefore,

$$Q = [v_1 \;\; v_2] = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}.$$

Then $R = Q^*A = Q^{\rm t}A = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 1 \end{bmatrix}$. Verify that $A = QR$. □

The QR-factorization can be used to express a least squares solution in closed form.

***Theorem* 5.22**

Let $A \in \mathbb{F}^{m\times n}$ have linearly independent columns, and let $b \in \mathbb{F}^{m\times 1}$. Then the least squares solution of $Ax = b$ is unique; and it is given by $u = R^{-1}Q^*b$, where $A = QR$ is a QR-factorization of A.

Proof Since A has linearly independent columns, $\mathrm{rank}(A) = n$. It implies that $\mathrm{null}(A) = n - \mathrm{rank}(A) = 0$. Thus, as a linear transformation, A is one-one. By Theorem 5.17(3), we have a unique least squares solution of $Ax = b$.

Using Theorem 5.20, take $u = R^{-1}Q^*b$. Now,

$$A^*Au = R^*Q^*QRR^{-1}Q^*b = R^*Q^*y = A^*b.$$

That is, u satisfies the equation $A^*Ax = A^*b$. Therefore, u is the least squares solution. ∎

Assume that A has linearly independent columns. Is the least squares solution u a solution of $Ax = b$? We have $Au = QRR^{-1}Q^*b = QQ^*b$. As seen earlier, this is not necessarily equal to b; and then u need not be a solution of $Ax = b$. However, u is also a solution of $Ax = b$ iff Q has orthonormal rows. In that case, A must be a square matrix.

But, if a solution v exists for $Ax = b$, then $v = u$. Reason? If $Av = b$, then $QRv = b$ implies that $Rv = Q^*b$. Hence, $v = u$.

Notice that the linear system $Ru = Q^*b$ is easy to solve since R is upper triangular.

Example 5.23

Consider computing the least squares solution of the system $Ax = b$, where A is the matrix in Example 5.21, and $b = (1,2,3)^t$. We have seen that

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = QR, \quad Q = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix}, \quad R = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 1 \end{bmatrix}.$$

If u is the least squares solution of $Ax = b$, then u satisfies $Ru = Q^*b$, or

$$\sqrt{2}u_1 + \sqrt{2}u_2 = 2\sqrt{2}, \quad u_2 = 2.$$

It is a triangular system, which is solved by back-substitution to obtain $u_1 = 0$ and $u_2 = 2$.

Alternatively, the least squares solution $u = (u_1, u_2)^t$ satsifies $A^*A = A^*b$, that is,

$$2u_1 + 2u_2 = 4, \quad 2u_1 + 3u_2 = 6.$$

Solving, we obtain the same solution $u_1 = 0$ and $u_2 = 2$. $\square$

Exercises for § 5.5

1. Find least squares solutions for the following systems:
   (a) $3x + y = 1,\ x + 2y = 0,\ 2x - y = -2.$
   (b) $x + y + z = 0,\ -x + z = 1,\ x - y = -1,\ y - z = -2.$
2. Find a QR-factorization of each of the following matrices:

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

3. Let $A \in \mathbb{F}^{m \times n}$ have linearly independent columns. Show that A has a unique QR-factorization where the upper triangular matrix R has positive diagonal entries.

4. Let $A \in \mathbb{R}^{m\times n}$ have linearly independent columns. Let $b \in \mathbb{R}^{m\times 1}$. Show that there exists a unique $x \in \mathbb{R}^{n\times 1}$ such that $A^*Ax = A^*b$.
5. Let $A \in \mathbb{F}^{m\times n}$ be a matrix of rank m. Show the following:
   (a) There exist an invertible matrix $B \in \mathbb{F}^{m\times m}$ and a matrix $C \in \mathbb{F}^{m\times n}$ such that $A = BC$ and C^*C is a diagonal matrix.
   (b) There exist an invertible matrix $B \in \mathbb{F}^{m\times m}$ and a matrix $C \in \mathbb{F}^{m\times n}$ such that $A = BC$ and $C^*C = I$.

6

Spectral Representation

6.1 Eigenvalues and eigenvectors

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. As a linear operator on $\mathbb{R}^2$, it transforms straight lines to straight lines. Find a straight line that is mapped to itself by A. We see that

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

Thus, the line $\{(a, a) : a \in \mathbb{R}\}$ never moves. So also the line $\{(a, -a) : a \in \mathbb{R}\}$.

Observe that

$$A \begin{bmatrix} a \\ a \end{bmatrix} = 1 \begin{bmatrix} a \\ a \end{bmatrix}, \quad A \begin{bmatrix} a \\ -a \end{bmatrix} = (-1) \begin{bmatrix} a \\ -a \end{bmatrix}.$$

In general, if a straight line remains invariant under a linear operator T, then the image of any point on the straight line must be a point *on the same* straight line. That is, $T(x)$ must be a scalar multiple of x. Notice that $T(0) = 0$, which is a scalar multiple of 0, anyway.

Let T be a linear operator on a vector space V over $\mathbb{F}$. A scalar $\lambda \in \mathbb{F}$ is called an **eigenvalue** of T iff there exists a nonzero vector $v \in V$ such that $Tv = \lambda v$. Such a vector v is called an **eigenvector of** T **for** (or associated with, or corresponding to) the eigenvalue λ.

Example 6.1

1. Let T be the linear operator on $\mathbb{R}^2$ given by $T(a, b) = (a, a + b)$. We have $T(0, 1) = (0, 0 + 1) = 1\,(0, 1)$. Thus the vector $(0, 1)$ is an eigenvector associated with the eigenvalue 1 of T. Is $(0, b)$ also an eigenvector associated with the same eigenvalue 1?
2. Let the linear operator T on $\mathbb{R}^2$ be given by $T(a, b) = (-b, a)$. If λ is an eigenvalue of T with an eigenvector (a, b), then $(-b, a) = (\lambda a, \lambda b)$ implies that $b = -\lambda a$ and $a = \lambda b$. It gives $b = -\lambda^2 b$, or $b(1 + \lambda^2) = 0$. Since $\mathbb{R}^2$ is a real vector space, $\lambda \in \mathbb{R}$. Then $1 + \lambda^2 \neq 0$. Hence $b = 0$. This leads to $a = 0$. That is, $(a, b) = (0, 0)$. But this impossible as an eigenvector is nonzero. Therefore, we conclude that T does not have an eigenvalue.

3. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be given by $T(a,b) = (-b,a)$. As in (2), if λ is an eigenvalue of T with an eigenvector (a,b), then $b(1+\lambda^2) = 0$ and $b = -\lambda a$. If $b = 0$, then $a = 0$, which is not possible as $(a,b) \neq (0,0)$. Thus $1+\lambda^2 = 0$. Hence $\lambda = \pm i$. It is easy to verify that the eigenvalue $\lambda = i$ is associated with an eigenvector $(1,-i)$ and the eigenvalue $\lambda = -i$ is associated with an eigenvector $(1,i)$.
4. The linear operator $T : \mathbb{F}[t] \to \mathbb{F}[t]$ defined by $T(p(t)) = tp(t)$ has no eigenvector and no eigenvalue, since for a polynomial $p(t)$, $tp(t) \neq \alpha p(t)$ for any $\alpha \in \mathbb{F}$.
5. Let $T : \mathbb{R}[t] \to \mathbb{R}[t]$ be defined by $T(p(t)) = p'(t)$, where we interpret each $p \in \mathbb{R}[t]$ as a function from the open interval $(0,1)$ to $\mathbb{R}$. Since derivative of a constant polynomial is 0, which equals 0 times the constant polynomial, all nonzero constant polynomials are eigenvectors of T associated with the eigenvalue 0. □

***Theorem* 6.2**
Let $T : V \to V$ be a linear operator, and let λ be a scalar. Then

(1) *A nonzero vector $v \in V$ is an eigenvector of T for the eigenvalue λ iff $v \in N(T-\lambda I)$;*

(2) *λ is an eigenvalue of T iff $T-\lambda I$ is not one-one.*

Proof (1) A nonzero vector v is an eigenvector of T for the eigenvalue λ iff $Tv = \lambda v$ iff $(T-\lambda I)v = 0$ iff $v \in N(T-\lambda I)$.

(2) If λ is an eigenvalue of T then there exists a nonzero vector $v \in V$ such that $v \in N(T-\lambda I)$. Thus $T-\lambda I$ is not one-one. Conversely, if $T-\lambda I$ is not one-one, then there exist distinct vectors $u, w \in V$ such that $(T-\lambda I)(u) = (T-\lambda I)w$. Then $T(u-w) = \lambda(u-w)$, where $u-w \neq 0$. It follows that λ is an eigenvalue of T with an eigenvector as $u-w$. ∎

The results on eigenvalues of operators are applicable to square matrices, in particular. Further, eigenvalues of an operator and that of its matrix representation coincide.

***Theorem* 6.3**
Let T be a liner operator on a finite dimensional vector space V over $\mathbb{F}$. Let B be an ordered basis of V. Then, $\lambda \in \mathbb{F}$ is an eigenvalue of T with an associated eigenvector v iff λ is an eigenvalue of the matrix $[T]_{B,B} \in \mathbb{F}^{n\times n}$ with an associated eigenvector $[v]_B \in \mathbb{F}^{n\times 1}$.

Proof Let $\lambda \in \mathbb{F}$ be an eigenvalue of T. There exists a nonzero vector $v \in V$ such that $Tv = \lambda v$. Then $[v]_B \neq 0$ and $[T]_{B,B}[v]_B = [Tv]_B = [\lambda v]_B = \lambda[v]_B$.

Conversely, let $\lambda \in \mathbb{F}$ be an eigenvalue of $[T]_{B,B}$ with an associated eigenvector u. Suppose $B = \{v_1, \ldots, v_n\}$. Then $u \in \mathbb{F}^{n\times 1}$. Let $u = (a_1, \ldots, a_n)^{\mathrm{t}}$. Write

$v = a_1v_1 + \cdots + a_nv_n$. Then $u = [v]_B$. Now, $u \neq 0$ implies $v \neq 0$. Further, $[\lambda v]_B = \lambda u = [T]_{B,B}u = [T]_{B,B}[v]_B = [Tv]_B$ implies that $Tv = \lambda v$. ∎

Theorem 6.3 allows us to go back and forth from a linear operator on a finite dimensional vector space to its matrix representation with respect to any basis while considering problems about its eigenvalues and eigenvectors.

Exercises for § 6.1

1. Show that the eigenvalues of a triangular matrix (upper or lower) are the entries on the diagonal.

2. Let A be an $n \times n$ matrix and let α be a scalar such that each row (or each column) sums to α. Show that α is an eigenvalue of A.

3. Let $T : V \to V$ be a linear operator . Prove the following:

   (a) If λ is an eigenvalue of T then λ^k is an eigenvalue of T^k.
   (b) If λ is an eigenvalue of T and $\alpha \in \mathbb{F}$, then $\lambda + \alpha$ is an eigenvalue of $T + \alpha I$.
   (c) Let $p(t) = a_0 + a_1t + \ldots + a_kt^k \in \mathbb{F}[t]$. If λ is an eigenvalue of T then $p(\lambda)$ is an eigenvalue of $p(T) := a_0I + a_1T + \ldots + a_kT^k$.

   Are all the eigenvalues of $p(T)$ of the form $p(\lambda)$, where λ is an eigenvalue of T?

4. Let $T : V \to V$ be an isomorphism, where V is a finite dimensional vector space. Let λ be a nonzero scalar. Show that λ is an eigenvalue of T iff $1/\lambda$ is an eigenvalue of T^{-1}.

5. Can any nonzero vector in any non-trivial vector space be an eigenvector of some linear operator?

6. Given a scalar λ, can any nonzero vector in any non-trivial vector space be an eigenvector associated with the eigenvalue λ of some linear operator?

7. Construct $A, B \in \mathbb{R}^{2\times 2}$ such that λ is an eigenvalue of A, μ is an eigenvalue of B but $\lambda\mu$ is not an eigenvalue of AB.

8. Let S and T be linear operators on V. Let λ and μ be eigenvalues of S and T, respectively. What is wrong with the following argument?

   > $\lambda\mu$ an eigenvalue of $S \circ T$ because, if $S(x) = \lambda x$ and $T(x) = \mu x$, then $(S \circ T)x = S(\mu x) = \mu S(x) = \mu\lambda x = \lambda\mu x$.

6.2 Characteristic polynomial

Eigenvalues of a matrix can be seen as zeros of a certain polynomial. Due to Theorem 6.3, the same should be true for a linear operator.

***Theorem* 6.4**
Let T be a linear operator on a finite dimensional vector space V. Let B be an ordered basis of V, and let $A = [T]_{B,B}$ be the matrix representation of T with respect to B. Then a scalar λ is an eigenvalue of T iff $\det(A - \lambda I) = 0$.

Proof Let $\dim V = n$. With respect to the basis B, the matrix of $T - \lambda I$ is $A - \lambda I$. Now, $T - \lambda I$ is not one-one iff $(A - \lambda I)x = 0$ has a nonzero solution iff $\operatorname{rank}(A - \lambda I) < n$ iff $\det(A - \lambda I) = 0$, due to Theorem 4.15. ∎

In Theorem 6.4, it looks as though the equation $\det([T]_{B,B} - \lambda I) = 0$ is not affected by the choice of a basis B for the vector space. In fact, if we choose another ordered basis, then the matrix of T with respect to the new basis can be written as $P^{-1}[T]_{B,B}P$ for some invertible matrix P; see Theorem 4.19. In that case,

$$\det(P^{-1}[T]_{B,B}P - \lambda I) = \det(P^{-1}([T]_{B,B} - \lambda I)P) = \det([T]_{B,B} - \lambda I).$$

Therefore, $\det([T]_{B,B} - \lambda I)$ as a polynomial in λ, is independent of the particular matrix representation of T.

Further, if A is an $n \times n$ matrix, then $\det(A - tI)$ is a polynomial in t with its leading term as $(-1)^n t^n$. To make it a monic polynomial, that is, when the leading term has the coefficient as 1, we multiply it with $(-1)^n$. We know that such a monic polynomial is not affected by a change of basis. Thus we give a name to it.

Let T be a linear operator on a vector space V of dimension n. Let A be a matrix representation of T with respect to some ordered basis of V. The polynomial $(-1)^n \det(A - tI)$ in the variable t is called the **characteristic polynomial** of T; and it is denoted by $\chi_T(t)$.

Due to Theorem 6.4, the eigenvalues of T are precisely the zeros of its characteristic polynomial $\chi_T(t)$, which are in the underlying field.

Example 6.5

1. In Example 6.1(1), we had $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(a,b) = (a, a+b)$. Let us take the standard basis $E = \{e_1, e_2\}$ for $\mathbb{R}^2$. Then

$$[T]_{E,E} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \chi_T(t) = (-1)^2 \begin{vmatrix} 1-t & 0 \\ 1 & 1-t \end{vmatrix} = (t-1)^2.$$

The eigenvalues of T are the zeros of $\chi_T(t)$ which are scalars in the underlying field. That is, 1 is the only eigenvalue of T. Solving $T(a,b) = 1\,(a,b)$, we get eigenvectors $(0,b)$ for $b \neq 0$.

2. For the linear operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(a,b) = (-b,a)$, in Example 6.1(2), fix the standard basis E of $\mathbb{R}^2$. Then

$$[T]_{E,E} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \chi_T(t) = (-1)^2 \begin{vmatrix} -t & -1 \\ 1 & -t \end{vmatrix} = t^2 + 1.$$

If λ is an eigenvalue of T, then $\lambda \in \mathbb{R}$ and $\lambda^2 + 1 = 0$. But $\lambda^2 + 1 \neq 0$ for any $\lambda \in \mathbb{R}$. Therefore, T does not have an eigenvalue. $\square$

Let $T : V \to V$ be a linear operator, where V is a finite dimensional vector space over $\mathbb{F}$. The field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. Thus the coefficients of powers of t in $\chi_T(t)$ are complex numbers, in general. Then all zeros of $\chi_T(t)$ are in $\mathbb{C}$. This fact is a consequence of the *fundamental theorem of algebra*, which states the following:

> Each polynomial of degree n with complex coefficients has n number of complex zeros, counting multiplicities.

An eigenvalue of T is a zero of the characteristic polynomial $\chi_T(t)$ which lies in the underlying field.

Example 6.6

The rotation by an angle θ on the plane $\mathbb{R}^2$ is given by the linear operator

$$T_\theta : \mathbb{R}^2 \to \mathbb{R}^2, \quad T_\theta(a,b) = (a\cos\theta - b\sin\theta,\ a\sin\theta + b\cos\theta).$$

Taking the standard basis B for $\mathbb{R}^2$, we have

$$[T_\theta]_{B,B} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Then $\chi_T(t) = (t - \cos\theta)^2 + \sin^2\theta$. Its zeros are $\cos\theta \pm i\sin\theta$. We find that if θ is not a multiple of π, then all these zeros are non-real. In this case, T_θ does not have an eigenvalue.

Indeed, if θ is not a multiple of π, the rotation does not fix any straight line through the origin. $\square$

It follows from the fundamental theorem of algebra that if $p(t)$ is a polynomial with real coefficients, then its complex zeros come in conjugate pairs. That is, if $\lambda = \alpha + i\beta,\ \beta \neq 0$ is a zero of such a polynomial $p(t)$, then so is $\overline{\lambda} = \alpha - i\beta$. Further, it implies that a polynomial of odd degree has a real zero.

Therefore, if $\mathbb{F} = \mathbb{C}$, then each zero of the characteristic polynomial is an eigenvalue of T. When $\mathbb{F} = \mathbb{R}$, the non-real zeros of the characteristic polynomial are not eigenvalues of T. It thus follows that each linear operator on a finite dimensional complex vector space has an eigenvalue; and each linear operator on an odd dimensional real vector space has an eigenvalue.

For matrices, we need to be a bit more careful. Let A be an $n \times n$ matrix. If at least one entry of A is a complex number with nonzero imaginary part, then A is viewed as a linear operator on $\mathbb{C}^{n\times 1}$. In this case, the eigenvalues of A are precisely the zeros of the characteristic polynomial. On the other hand, if A has only real entries, we may view it as a linear operator on $\mathbb{C}^{n\times 1}$ or on $\mathbb{R}^{n\times 1}$. In the former case, all zeros of the characteristic polynomial are its eigenvalues. In the latter case, only the real zeros are the eigenvalues of A.

Thus, for convenience, we view a matrix $A \in \mathbb{F}^{n\times n}$ as a matrix in $\mathbb{C}^{n\times n}$. Accordingly, each zero of the characteristic polynomial of A is called a **complex eigenvalue** of A. Notice that in general, an eigenvector corresponding to a complex eigenvalue is a vector in $\mathbb{C}^{n\times 1}$.

For instance, in Example 6.1(3), the matrix $[T]_{E,E}$ has complex eigenvalues i and $-i$. The corresponding eigenvectors are $(1, -i)^{\mathrm{t}}$ and $(1, i)^{\mathrm{t}}$, which are in $\mathbb{C}^{2\times 1}$. Similarly, the rotation matrix $[T_\theta]_{B,B}$ in Example 6.6 has complex eigenvalues $\cos\theta \pm i\sin\theta$ with eigenvectors $[1,\ \mp i]^{\mathrm{t}}$.

Recall that a polynomial $p(t)$ has λ as a *zero of multiplicity* m means that $(t-\lambda)^m$ divides the polynomial $p(t)$ but $(t-\lambda)^{m+1}$ does not divide $p(t)$. Accordingly, if $A \in \mathbb{F}^{n\times n}$ and λ is a complex eigenvalue of A, where λ has multiplicity m as a zero of $\chi_A(t)$, then we say that the **algebraic multiplicity** of the eigenvalue λ of A is m.

When we speak of all complex eigenvalues of A *counting multiplicities,* we are concerned with the list of all complex eigenvalues, where each one is repeated as many times as its algebraic multiplicity. For instance, an $n \times n$ matrix A has n number of complex eigenvalues, counting multiplicities. You should interpret the following results in this sense.

***Theorem* 6.7**

(1) *A square matrix and its transpose have the same complex eigenvalues, counting multiplicities.*

(2) *Eigenvalues of a triangular (upper or lower) and of a diagonal matrix are precisely its diagonal entries.*

(3) *The determinant of a square matrix is the product of all its complex eigenvalues, counting multiplicities.*

(4) *The trace of a square matrix is the sum of all its complex eigenvalues, counting multiplicities.*

Proof Let A be a square matrix of order n.

(1) Now, $\chi_{A^{\mathrm{t}}}(t) = \det(A^{\mathrm{t}} - tI) = \det((A-tI)^{\mathrm{t}}) = \det(A-tI) = \chi_A(t)$.

(2) In all these cases, $\chi_A(t) = (-1)^n \det(A-tI) = (t-a_{11})\cdots(t-a_{nn})$.

(3) Let $\lambda_1, \ldots, \lambda_n$ be the n complex eigenvalues of A, counting multiplicities. Then

$$\chi_A(t) = (-1)^n \det(A - tI) = (t - \lambda_1) \cdots (t - \lambda_n).$$

Put $t = 0$. It gives $\det(A) = \lambda_1 \cdots \lambda_n$.

(4) Let $A = [a_{ij}]$. Expand $\det(A - tI)$; and look at the coefficient of t^{n-1} in this expansion. We obtain

$$\text{Coeff. of } t^{n-1} \text{ in } \det(A - tI) = (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn}) = (-1)^{n-1}\mathrm{tr}(A).$$

But $\det(A - tI) = (-1)^n \chi_A(t) = (\lambda_1 - t) \cdots (\lambda_n - t)$. So,

$$\text{Coeff. of } t^{n-1} \text{ in } \det(A - tI) = (-1)^{n-1}(\lambda_1 + \cdots + \lambda_n).$$

Therefore, $\lambda_1 + \cdots + \lambda_n = tr(A)$. ∎

***Theorem* 6.8 (Cayley-Hamilton)**

Any linear operator on a finite dimensional vector space satisfies its characteristic polynomial.

Proof Let $T : V \to V$ be a linear operator, where B is an ordered basis of the finite dimensional vector space V. As we know $\chi_T(t) = \chi_A(t)$, where $A = [T]_{B,B}$. Also, $T = [\ ]_B^{-1} \circ A \circ [\ ]_B$; see Section 4.5.

Recall that for any square matrix C, $\det(C)I = C\,\mathrm{adj}(C)$, where $\mathrm{adj}(C)$ is the adjugate of C. Using this result on the matrix $A - tI$, we have

$$\chi_A(t)\, I = (-1)^n \det(A - tI)\, I = (-1)^n (A - tI)\,\mathrm{adj}\,(A - tI).$$

The entries in $\mathrm{adj}\,(A - tI)$ are polynomials in t of degree at most $n-1$. Thus there exist $n \times n$ matrices $A_0, A_1, , \ldots, A_{n-1}$ such that

$$\mathrm{adj}\,(A - tI) = A_0 + tA_1 + \cdots + t^{n-1}A_{n-1}.$$

Then $\chi_A(t)I = (-1)^n(A - tI)(A_0 + tA_1 + \cdots t^{n-1}A_{n-1})$.

Notice that this is an identity in polynomials, where the coefficients of t^j are matrices. Substituting t by any matrix of the same order will satisfy the equation. In particular, substitute A for t to obtain $\chi_A(A) = 0$.

Since $[\ ]_B : V \to \mathbb{F}^{n\times 1}$ is an isomorphism, it is easy to see that for any polynomial $p(t)$, $p([\ ]_B^{-1} \circ A \circ [\ ]_B) = [\ ]_B^{-1} \circ p(A) \circ [\ ]_B$. In particular, taking $p(t) = \chi_T(t)$ and using the fact that $\chi_T(t) = \chi_A(t)$, we obtain

$$\chi_T(T) = \chi_A([\ ]_B^{-1} \circ A \circ [\ ]_B) = [\ ]_B^{-1} \circ \chi_A(A) \circ [\ ]_B = 0.$$ ∎

However, the characteristic polynomial is not the only polynomial that is satisfied by the operator. For example, the identity operator $I : V \to V$, defined by $I(v) = v$, on an n-dimensional vector space V has the characteristic polynomial $p(t) = (t-1)^n$. But I satisfies the polynomial $q(t) = t - 1$, which is not the characteristic polynomial in case $n > 1$. In this regard, the following result is often helpful.

***Theorem* 6.9**
If a linear operator on a finite dimensional vector space satisfies a polynomial, then its eigenvalues are from among the zeros of the polynomial.

Proof Let T be a linear operator on an n-dimensional vector space V over $\mathbb{F}$. Let $p(t)$ be a polynomial with coefficients from $\mathbb{F}$ such that $p(T) = 0$, the zero operator. Let λ be an eigenvalue of T. We show that $p(\lambda) = 0$.

Suppose $v \in V$ is an eigenvector corresponding to the eigenvalue λ of T. Let $\alpha, \beta \in \mathbb{F}$. Write the identity operator on V as I. We see that

$$(\alpha T + \beta I)v = \alpha T v + \beta v = (\alpha\lambda + \beta)v, \quad T^2 v = T(\lambda v) = \lambda T v = \lambda^2 v.$$

It then follows by induction that $T^k v = \lambda^k v$ for any $k \in \mathbb{N}$; consequently, $p(A)v = p(\lambda)v$. Since $p(T) = 0$. we have $p(\lambda)v = 0$. As $v \neq 0$, we conclude that $p(\lambda) = 0$. ∎

Example 6.10
Let $A \in \mathbb{R}^{3\times 3}$ be such that $A^2 = 3A - 2I$ and $\det(A) = 4$. Then what is $\mathrm{tr}(A)$?

A satisfies the polynomial $p(t) = t^2 - 3t + 2$. Since $p(t) = (t-2)(t-1)$, which has zeros as 2 and 1, the eigenvalues of A can be 2 or 1. Now that A is a matrix of order 3, taking into account the repetition of eigenvalues, the possibilities are 2, 2, 2 or 2, 2, 1 or 2, 1, 1 or 1, 1, 1. As $\det(A)$ is equal to the product of the eigenvalues of A; and it is 4, the eigenvalues are 2, 2 and 1. So $\mathrm{tr}(A)$, which is equal to the sum of the eigenvalues, is 5. □

Recall that the adjoint of a linear operator gives rise to special types of operators. We wish to find out the nature of eigenvalues and eigenvectors of these special types of operators. It will help to consider matrices first.

***Theorem* 6.11**
Let A be a square matrix. Let λ be any complex eigenvalue of A.

(1) *If A is hermitian or real symmetric, then $\lambda \in \mathbb{R}$.*
(2) *If A is hermitian, then eigenvectors corresponding to distinct eigenvalues are orthogonal.*
(3) *If A is real symmetric, then a real eigenvector corresponding to the eigenvalue λ exists.*
(4) *If A is skew-hermitian or real skew-symmetric, then λ is purely imaginary or zero.*
(5) *If A is a unitary or an orthogonal matrix, then $|\lambda| = 1$ and $|\det(A)| = 1$.*

Proof (1) Suppose $A^* = A$. Let $v \in \mathbb{C}^{n\times 1}$ be an eigenvector corresponding to the complex eigenvalue λ of A. Now, $Av = \lambda v$ implies

$$\begin{aligned}\langle Av, v\rangle &= \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.\\ \langle Av, v\rangle &= \langle A^* v, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle.\end{aligned}$$

Since $\langle v, v\rangle \neq 0$, it follows that $\lambda = \overline{\lambda}$. That is, λ is real.

(2) Let u, v be eigenvectors corresponding to distinct eigenvalues λ, μ of A. The equalities $A^* = A,\ Au = \lambda u,\ Av = \mu v$ and λ, μ are real imply that

$$\begin{aligned}\langle Au, v\rangle &= \langle \lambda u, v\rangle = \lambda\langle u, v\rangle.\\ \langle Au, v\rangle &= \langle A^* u, v\rangle = \langle u, Av\rangle = \langle u, \mu v\rangle = \overline{\mu}\langle u, v\rangle = \mu\langle u, v\rangle.\end{aligned}$$

Then $(\lambda - \mu)\langle u, v\rangle = 0$. As $\lambda \neq \mu$, we conclude that u and v are orthogonal.

(3) Let A be real-symmetric. By (1), $\lambda \in \mathbb{R}$. Let $v \in \mathbb{C}^{n\times 1}$ be an eiegnevector of A corresponding to the eigenvalue λ. Write $v = x + iy$, where $x, y \in \mathbb{R}^{n\times 1}$. Comparing the real and imaginary parts in $A(x + iy) = \lambda(x + iy)$, we have

$$Ax = \lambda x, \quad Ay = \lambda y.$$

Since $x + iy \neq 0$, at least one of x or y is nonzero. Such a nonzero vector is a real eigenvector of A associated with the eigenvalue λ.

(4) Suppose $A^* = -A$. Let v be an eigenvector associated with the complex eigenvalue λ of A. As in the proof of (1), it follows that

$$\langle Av, v\rangle = \lambda\langle v, v\rangle = -\overline{\lambda}\langle v, v\rangle.$$

Since $v \neq 0$, we have $\lambda = -\overline{\lambda}$. That is, λ is purely imaginary or zero.

(5) Suppose $A^*A = AA^* = I$. Let v be an eigenvector associated with the complex eigenvalue λ of A. Now, $Av = \lambda v$ implies that

$$\begin{aligned}\langle Av, Av\rangle &= \langle v, A^*Av\rangle = \langle v, v\rangle.\\ \langle Av, Av\rangle &= \langle \lambda v, \lambda v\rangle = \lambda\overline{\lambda}\langle v, v\rangle = |\lambda|^2\langle v, v\rangle.\end{aligned}$$

That is, $(1 - |\lambda|^2)\langle v, v\rangle = 0$. As $v^*v \neq 0$, we have $|\lambda| = 1$.

Further, since $\det(A)$ is the product of all complex eigenvalues of A counting multiplicities, $|\det(A)| = 1$. ∎

We remark that a hermitian matrix need not have a real eigenvector though all eigenvalues are real. For example, the matrix $\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$ has eigenvalues 1 and -1 with corresponding linearly independent eigenvectors $[i\ 1]^t$ and $[-i\ 1]^t$. It does not have any real eigenvector.

We may use the results in Theorem 6.11 to prove similar facts about linear operators on finite dimensional inner product spaces.

***Theorem* 6.12**

Let T be a linear operator on a finite dimensional inner product space.

(1) *If T is self-adjoint, then it has an eigenvalue; all eigenvalues of T are real; and eigenvectors corresponding to distinct eigenvalues are orthogonal.*

(2) *If T is unitary, then each of its eigenvalues has absolute value* 1.

Proof Let $T : V \to V$ be a linear operator, where V is a finite dimensional ips. Let B be an orthonormal ordered basis of V. Assume that T is self-adjoint. Then the matrix $[T]_{B,B}$ is hermitian. By Theorems 6.3 and 6.11(1) it follows that all eigenvalues of T are real. Similarly, it follows that eigenvectors corresponding to distinct eigenvalues are orthogonal.

For the existence of an eigenvalue, notice that if V is a complex ips, then T has an eigenvalue. So, let V be a real ips. Then $[T]_{B,B}$ is real symmetric. This matrix has eigenvalues, and they are real. By Theorem 6.11(2), each such eigenvalue λ is associated with a real eigenvector u. By Theorem 6.3, this λ is an eigenvalue of T with an associated eigenvector $v \in V$, where $[v]_B = u$. This proves (1). Similarly, (2) is proved. ∎

As we know, the determinant of an orthogonal operator (matrix) is either 1 or -1. However, every orthogonal operator need not have an eigenvalue.

Example 6.13

Consider the linear operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(a, b) = \frac{1}{\sqrt{2}}(a - b, a + b)$. Its matrix representation with respect to the standard basis E of $\mathbb{R}^2$ is

$$[T]_{E,E} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

It is easy to verify that $[T]_{E,E}$ is an orthogonal matrix. Its characteristic polynomial is

$$\chi_T(t) = t^2 - \sqrt{2}\,t + 1.$$

Since $\chi_T(t)$ has no real zeros, T has no eigenvalues. You can verify that $[T]_{E,E}$, and hence T, has eigenvalues as $(1 \pm i)/\sqrt{2}$. □

In Example 6.13, T is the rotation in the plane by an angle of $\pi/4$. In fact, any rotation T_θ of Example 6.6, where θ is not a multiple of π, provides such an example.

Exercises for § 6.2

1. Find the characteristic polynomial, the eigenvalues and the associated eigenvectors for the matrices given below.

$$\begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix}, \begin{bmatrix} -2 & 0 & 3 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & 17 & 8 \end{bmatrix}.$$

2. Find the eigenvalues and associated eigenvectors of the differentiation operator $d/dt : \mathbb{R}_3[t] \to \mathbb{R}_3[t]$.
3. If you know the characteristic polynomial of $A \in \mathbb{F}^{n\times n}$, then how do you determine whether A is invertible or not? If A is invertible, then how do you compute A^{-1} using the characteristic polynomial of A?
4. Is it true that each real skew symmetric matrix has an eigenvalue?
5. If x and y are eigenvectors corresponding to distinct eigenvalues of a real symmetric matrix of order 3, then show that the cross product of x and y is a third eigenvector linearly independent with x and y.
6. Show that the eigenvectors corresponding to distinct eigenvalues of a real symmetric matrix are orthogonal.
7. Let V and W be vector spaces, each of dimension n. Let $\phi : V \to W$ be an isomorphism and let $T : V \to V$ be a linear operator. Let $p(t)$ be any polynomial. Prove that $p(\phi^{-1} \circ T \circ \phi) = \phi^{-1} \circ p(T) \circ \phi$.

6.3 Schur triangularization

Eigenvalues and eigenvectors can be used to represent a linear operator on an inner product space in a nice form. One of them is the unitary triangularization of Schur.

***Theorem* 6.14 (Schur Triangularization)**
For each linear operator on a finite dimensional complex ips, there exists an orthonormal ordered basis for the ips with respect to which the matrix of the linear operator is upper triangular.

Notice that in such a case, the eigenvalues of $T : V \to V$ are precisely the diagonal elements in the upper triangular matrix. Moreover, $B = \{v_1, \ldots, v_n\}$ is an ordered basis of V such that $[T]_{B,B}$ is upper triangular is equivalent to the following conditions:

$$Tv_k \in \operatorname{span}\{v_1, \ldots, v_k\} \quad \text{for } 1 \le k \le n.$$

Proof Let $\dim(V) = n$. For $n = 1$, let $\{u\}$ be a basis of V. Then $u \neq 0$. Take $v = u/\|u\|$. Then $B = \{v\}$ is an orthonormal basis of V. It follows that $Tv \in V$, and hence $Tv = \alpha v$ for some $\alpha \in \mathbb{C}$. So, $Tv \in \operatorname{span}\{v\}$.

To apply induction, assume that the statement holds for all linear operators on each complex ips of dimension less than n. Let T be a linear operator on a complex ips V of dimension n. Let $\lambda \in \mathbb{C}$ be an eigenvalue of T with an associated eigenvector u. Take $v_1 = u/\|u\|$; so that v_1 is an eigenvector of T of norm 1 associated with the eigenvalue λ. Extend the set $\{v_1\}$ to obtain a basis for V. Then use Gram-Schmidt process to obtain an orthogonal (or orthonormal) basis $\{v_1, u_2, \ldots, u_n\}$ for V.

Let $U = \operatorname{span}\{u_2, \ldots, u_n\}$. Then v_1 is orthogonal to each vector in U. Let $x \in U$. As $Tx \in V$, there exist unique scalars $a_1, a_2, \ldots, a_n \in \mathbb{C}$ such that

$$Tx = a_1 v_1 + a_2 u_2 + \cdots + a_n u_n.$$

Define $S : U \to U$ by

$$S(x) = a_2 u_2 + \cdots + a_n u_n.$$

Clearly, for $u \in U$, $S(u) \in U$. We show that S is a linear operator. For this, let $y, z \in U$, and let $\alpha \in \mathbb{C}$. There exist unique scalar b_i, c_i such that

$$\begin{aligned} Ty &= b_1 v_1 + b_2 u_2 + \cdots + b_n u_n, \quad Tz = c_1 v_1 + c_2 v_2 + \cdots + c_n u_n; \\ T(y + \alpha z) &= (b_1 + \alpha c_1) v_1 + (b_2 u_2 + \cdots + b_n u_n) + \alpha (c_2 u_2 + \cdots + c_n u_n). \end{aligned}$$

We see that

$$S(y + \alpha z) = (b_2 u_2 + \cdots + b_n u_n) + \alpha (c_2 u_2 + \cdots + c_n u_n) = S(y) + \alpha S(z).$$

Therefore, $S : U \to U$ is a linear operator. It satisfies

$$Tx = a_1 v_1 + S(x),$$

where the scalar $a_1 \in \mathbb{C}$ and the vector $S(x) \in U$ are uniquely determined from T and the vector $x \in U$.

By the induction hypothesis, there exists an orthonormal ordered basis $\{v_2, \ldots, v_n\}$ for U such that

$$S(v_j) \in \operatorname{span}\{v_2, \ldots, v_j\} \quad \text{for } 2 \le j \le n.$$

Let $B = \{v_1, v_2, \ldots, v_n\}$. Since $\|v_1\| = 1$, v_1 is orthogonal to each vector in U, and $\{v_1, \ldots, v_n\}$ is an orthonormal set. We see that B is an orthonormal ordered basis for V.

To see that this is the required basis, we compute the T-images of the basis vectors as follows: (Here, α_i are some suitable scalars.)

$$\begin{aligned} T(v_1) &= \lambda v_1 \in \operatorname{span}\{v_1\}, \\ T(v_2) &= \alpha_1 v_1 + S(v_2) \in \operatorname{span}\{v_1, v_2\}, \\ &\vdots \\ T(v_j) &= \alpha_j v_1 + S(v_j) \in \operatorname{span}\{v_1, v_2, \ldots, v_j\}, \quad \text{for } 2 \le j \le n. \end{aligned}$$

This completes the proof. ∎

Schur triangularization involves the choice of a basis for V so that the linear operator T on V is represented by an upper triangular matrix. When $V = \mathbb{C}^{n\times 1}$, the linear operator T is a matrix in $\mathbb{C}^{n\times n}$. We discuss this particular case.

Let $A \in \mathbb{C}^{n\times n}$. Let $E = \{v_1, \ldots, v_n\}$ be a basis for $\mathbb{C}^{n\times 1}$ such that $[A]_{E,E}$ is upper triangular. From Theorem 4.19 it follows that $P^{-1}AP$ is upper triangular, where $P = [v_1 \ \cdots \ v_n]$. Notice that this is a stronger type of equivalence.

Let $A, B \in \mathbb{F}^{n\times n}$. B is called **similar to** A iff there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$.

B is called **unitarily similar** to A iff there exists a unitary matrix $P \in \mathbb{F}^{n\times n}$ such that $B = P^{-1}AP$.

When $\mathbb{F} = \mathbb{R}$, B is called **orthogonally similar** to A iff there exists an orthogonal matrix $P \in \mathbb{R}^{n\times n}$ such that $B = P^{-1}AP$.

Clearly, similarity is an equivalence relation on the set of square matrices of the same order. So, we speak of matrices similar to each other, etc. If two matrices are similar, then they are also equivalent. However, converse does not hold. For instance, the identity matrix of order n is similar to only itself. But the identity matrix is equivalent to any invertible matrix. Thus, equivalence may not preserve eigenvalues. In contrast, the following theorem shows that similarity preserves eigenvalues.

***Theorem* 6.15**
Similar square matrices have the same eigenvalues, counting multiplicities.

Proof Let $A, P, B \in \mathbb{F}^{n\times n}$ such that P is invertible and $B = P^{-1}AP$. Then

$$\begin{aligned}\chi_B(t) &= (-1)^n \det(P^{-1}AP - tI) = (-1)^n \det(P^{-1}(A - tI)P) \\ &= (-1)^n \det(P^{-1})\det(A - tI)\det(P) = (-1)^n \det(A - tI) = \chi_A(t).\end{aligned}$$

From this, the result follows. ∎

Though equivalence of matrices is easily characterized by the rank theorem, similarity involves much more. We will proceed towards that goal, but slowly.

For matrices, Theorem 6.14 may now be stated as follows.

***Theorem* 6.16 (*Schur Triangularization*)**
Each complex square matrix is unitarily similar to an upper triangular matrix; and each real square matrix having no non-real eigenvalues is orthogonally similar to an upper triangular matrix.

The inductive construction of a unitary matrix described in the proof of Schur triangularization amounts to the following.

Assume that for all $B \in \mathbb{C}^{m\times m}$, $m \geq 1$, we have a unitary matrix $Q \in \mathbb{C}^{m\times m}$ such that Q^*BQ is upper triangular. Let $A \in \mathbb{C}^{(m+1)\times(m+1)}$ and let $\lambda \in \mathbb{C}$ be an eigenvalue of A with an associated eigenvector u. Consider $\mathbb{C}^{(m+1)\times 1}$ as an ips with the usual inner product $\langle w, z\rangle = z^*w$. Let $v = u/\|u\|$, so that v is an eigenvector of A of norm 1 associated with the eigenvalue λ. Extend the set $\{v\}$ to obtain an orthonormal ordered basis $E = \{v, u_2, \ldots, u_{m+1}\}$ for $\mathbb{C}^{(m+1)\times 1}$. Here, you may have to use an extension of a basis, and then Gram-Schmidt orthonormalization process. Now, construct the matrix $R \in \mathbb{C}^{(m+1)\times(m+1)}$ by taking these basis vectors as its columns, in that order; that is, let

$$R = \begin{bmatrix} v & u_2 & \cdots & u_{m+1} \end{bmatrix}.$$

Since E is an orthonormal set, R is unitary. With respect to the basis E, the matrix representation of A is $R^{-1}AR = R^*AR$. Using the standard basis for $\mathbb{C}^{(m+1)\times 1}$, we see that the first column of R^*AR is

$$R^*ARe_1 = R^*Av = R^{-1}\lambda v = \lambda R^{-1}v = \lambda R^{-1}Re_1 = \lambda e_1.$$

Then R^*AR can be written in the following block form:

$$R^*AR = \begin{bmatrix} \lambda & x \\ 0 & C \end{bmatrix},$$

where $0 \in \mathbb{C}^{m\times 1}$, $C \in \mathbb{C}^{m\times m}$ and $x = [v^*Av_1 \ \cdots \ v^*Av_m] \in \mathbb{C}^{1\times m}$.

Notice that if $m = 1$, the construction is complete. For $m > 1$, by induction hypothesis, we have a matrix $S \in \mathbb{C}^{m\times m}$ such that S^*CS is upper triangular. Then take

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

Since S is unitary, direct computation shows that P is unitary. Moreover,

$$P^*AP = \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix}^* R^*AR \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & S^* \end{bmatrix}\begin{bmatrix} \lambda & x \\ 0 & C \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} \lambda & y \\ 0 & S^*CS \end{bmatrix}$$

for some $y \in \mathbb{C}^{1\times m}$. Since S^*CS is upper triangular, so is P^*AP. The construction is complete.

If A is a real matrix, and all its complex eigenvalues turn out to be real, then we use the transpose instead of the adjoint every where in the above construction. Thus, P is an orthogonal matrix.

Example 6.17

Consider the matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 2 & 3 & 0 \\ -1 & -1 & 1 \end{bmatrix}$ for Schur triangularization.

We find that $\chi_A(t) = (t-1)^2(t-4)$. All complex eigenvalues of A are real. Thus there exists an orthogonal matrix P such that P^tAP is upper triangular. To determine such a matrix P, we take one of the eigenvalues, say 1. An associated eigenvector of norm 1 is $v = (0,0,1)^t$. We extend $\{v\}$ to an orthonormal basis for $\mathbb{R}^{3\times 1}$. For convenience, we take the (ordered) orthonormal basis as

$$\{(0,0,1)^t,\ (1,0,0)^t,\ (0,1,0)^t\}.$$

Taking the basis vectors as columns, we form the matrix R as follows:

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

We then find that

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

Now, we try to triangularize the matrix $C = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$. It has eigenvalues 1 and 4. The eigenvector of unit norm associated with the eigenvalue 1 is $(1/\sqrt{2},\ -1/\sqrt{2})^t$. We extend it to an orthonormal basis

$$\{(1/\sqrt{2},\ -1/\sqrt{2})^t,\ (1/\sqrt{2},\ 1/\sqrt{2})^t\}$$

for $\mathbb{R}^{2\times 1}$. Then we construct the matrix S by taking these basis vectors as its columns, that is,

$$S = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

We find that $S^tCS = \begin{bmatrix} 1 & -1 \\ 0 & 4 \end{bmatrix}$, which is an upper triangular matrix. Then

$$P = R\begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}.$$

Now, $P^tAP = \begin{bmatrix} 1 & 0 & -\sqrt{2} \\ 0 & 1 & -1 \\ 0 & 0 & 4 \end{bmatrix}$, which is upper triangular. ∎

Notice that there is nothing sacred about being *upper* triangular. For, given a matrix $A \in \mathbb{C}^{n\times n}$, consider using Schur triangularization of A^*. There exists a unitary matrix P such that P^*A^*P is upper triangular. Then taking adjoint, we have P^*AP is lower triangular. That is, each square matrix is unitarily similar to a lower triangular matrix.

Analogously, a real square matrix having no non-real eigenvalues is also orthogonally similar to a lower triangular matrix. We remark that the lower triangular form of a matrix need not be the transpose or the adjoint of its upper triangular form.

Neither the unitary matrix P nor the upper triangular matrix P^*AP in Schur triangualrization is unique. That is, there can be unitary matrices P and Q such that $P \neq Q$, $P^*AP \neq Q^*AQ$, and both P^*AP and Q^*AQ are upper triangular. The non-uniqueness stems from the choices involved in the associated eigenvectors and in extending this to an orthonormal basis. For instance, in Example 6.17, if you extend $\{(0,0,1)^{\mathrm{t}}\}$ to the ordered orthonormal basis

$$\{(0,0,1)^{\mathrm{t}},\ (0,1,0)^{\mathrm{t}},\ (1,0,0)^{\mathrm{t}}\},$$

then you end up with (verify)

$$P = \begin{bmatrix} 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}, \quad P^{\mathrm{t}}AP = \begin{bmatrix} 1 & 0 & -\sqrt{2} \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}.$$

Exercises for § 6.3

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(a,b) = (5a + 7b, -2a - 4b)$. Determine an ordered basis of $\mathbb{R}^2$ with respect to which, the matrix of T is upper triangular.
2. Find Schur triangularization of the following matrices:
$$\begin{bmatrix} 7 & -2 \\ 12 & -3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 13 & 8 & 8 \\ -1 & 7 & -2 \\ -1 & -2 & 7 \end{bmatrix}.$$
3. Compute A^{50} for the 3×3 matrix in Exercise 2.
[Hint: Write the upper triangular matrix as a diagonal matrix plus another matrix; and then use the binomial theorem.]
4. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of an $n \times n$ matrix A. Prove that
$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \leq \mathrm{tr}(A^*A).$$
5. For a matrix $A = [a_{ij}] \in \mathbb{C}^{n\times n}$, let $c = \max\{|a_{ij}| : 1 \leq i, j \leq n\}$. Show that $|\det(A)| \leq c^n n^{n/2}$.
[Hint: Use the previous exercise and AM-GM inequality.]
6. Let the scalar λ appear exactly m times in the diagonal of a square upper triangular matrix A. Prove that $\mathrm{null}((A - \lambda I)^m) = m$.
7. Let λ be an eigenvalue of a matrix $A \in \mathbb{C}^{n\times n}$ having multiplicity m. Then prove that $\mathrm{null}((A - \lambda I)^m) = m$.

6.4 Diagonalizability

An upper triangular matrix similar to a given hermitian matrix takes a better form. For, suppose B is an upper triangular matrix similar to a hermitian matrix A. Then B is hermitian and upper triangular. Now, B^* is lower triangular; and then $B^* = B$ forces B to be a diagonal matrix. A similar statement holds for normal matrices.

Suppose B is a normal upper triangular matrix. If B is a 1×1 matrix, then it is diagonal. Otherwise, assume that (induction hypothesis) all $n \times n$ normal upper triangular matrices are diagonal. Let B be an $(n+1) \times (n+1)$ normal upper triangular matrix. Then we may write B in block form as

$$B = \begin{bmatrix} a & x \\ 0 & C \end{bmatrix},$$

where $x \in \mathbb{F}^{1\times n}$, $0 \in \mathbb{F}^{n\times 1}$, and $C \in \mathbb{F}^{n\times n}$ is an upper triangular matrix. Then

$$\begin{aligned} B^*B &= \begin{bmatrix} \overline{a} & 0^* \\ x^* & C^* \end{bmatrix} \begin{bmatrix} a & x \\ 0 & C \end{bmatrix} = \begin{bmatrix} |a|^2 & \overline{a}\,x \\ ax^* & x^*x + C^*C \end{bmatrix}, \\ BB^* &= \begin{bmatrix} a & x \\ 0 & C \end{bmatrix} \begin{bmatrix} \overline{a} & 0^* \\ x^* & C^* \end{bmatrix} = \begin{bmatrix} |a|^2 + xx^* & xC^* \\ Cx^* & CC^* \end{bmatrix}. \end{aligned}$$

Now, $B^*B = BB^*$ implies that $xx^* = 0$ and $x^*x + C^*C = CC^*$. If $x = [b_1 \ \cdots \ b_n]$, then $xx^* = |b_1|^2 + \cdots + |b_n|^2$. So, $xx^* = 0$ implies that $x = 0$. Thus $C^*C = CC^*$. That is,

$$B = \begin{bmatrix} a & 0 \\ 0 & C \end{bmatrix},$$

where C is an $n \times n$ normal upper triangular matrix. By our assumption, C is a diagonal matrix. Therefore, B is a diagonal matrix.

Observation 6.1 A normal upper triangular matrix is diagonal. In particular, a hermitian upper triangular matrix is diagonal.

Using Observation 6.1 and Schur triangularization, we see that each normal matrix is similar to a diagonal matrix.

A linear operator T on a finite dimensional inner product space V is called **diagonalizable** iff there exists an ordered basis B for V such that $[T]_{B,B}$ is a diagonal matrix.

The linear operator T is called **unitarily diagonalizable** iff there exists an orthonormal basis B for V such that $[T]_{B,B}$ is a diagonal matrix.

***Theorem* 6.18 (*Spectral theorem*)**
A linear operator on a finite dimensional ips is unitarily diagonalizable iff it is a normal operator. In particular, each self-adjoint linear operator on a finite dimensional ips is unitarily diagonalizable.

Proof Let $T : V \to V$ be a linear operator, where V is an ips of dimension n. Assume that T is normal. Schur triangularization implies that there exists an orthonormal ordered basis B for V such that $[T]_{B,B}$ is upper triangular. By Theorem 4.13, $[T]_{B,B}$ is a normal matrix. By Observation 6.1, $[T]_{B,B}$ is a diagonal matrix.

Conversely, suppose $[T]_{B,B}$ is a diagonal matrix, where B is an orthonormal ordered basis for V. Since a diagonal matrix is necessarily normal, $[T]_{B,B}$ is normal. Again, Theorem 4.13 implies that T is normal. ∎

Theorem 6.19 is called the *spectral theorem* since it uses the eigenvalues of the linear operator. Notice that the diagonal matrix that represents a normal linear operator T has its diagonal entries as the eigenvalues of T. In case, T is self-adjoint, its eigenvalues are real.

The spectral theorem for matrices may be stated as follows.

***Theorem* 6.19 (*Spectral theorem*)**
A matrix is unitarily diagonalizable iff it is normal. In particular, each hermitian matrix is unitarily diagonalizable, and each real-symmetric matrix is orthogonally diagonalizable.

Since a change of basis leads to similarity of matrices, it follows that a square matrix A is unitarily diagonalizable iff there exists a unitary matrix U such that

$$U^{-1}AU = U^*AU = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

where λ_is are the eigenvalues of A. Similarly, **orthogonally diagonalizable** means that such a matrix U has also real entries.

Example 6.20

The matrix $A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$ is real symmetric. It has eigenvalues $-1,\ 2,\ 2$, with respective eigenvectors

$$\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}.$$

They form an orthonormal basis for $\mathbb{R}^3$. Taking

$$P = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix},$$

we see that $P^{-1} = P^{\mathrm{t}}$ and $P^{-1}AP = P^{\mathrm{t}}AP = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. □

There can be non-normal but diagonalizable matrices. For such a matrix A, there cannot exist a unitary matrix U such that $U^{-1}AU$ is a diagonal matrix.

When $P^{-1}AP$ is a diagonal matrix, we say that A is **diagonalized by** P. In this case, we have scalars $\lambda_1, \ldots, \lambda_n$ such that

$$P^{-1}AP = D = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n).$$

Then $AP = PD$ implies that the ith column of P is an eigenvector corresponding to the eigenvalue λ_i of A. Since P is invertible, its columns form a basis for $\mathbb{F}^{n\times 1}$.

Conversely, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of a matrix $A \in \mathbb{F}^{n\times n}$ with corresponding eigenvectors $v_1, \ldots, v_n$ which form a basis for $\mathbb{F}^{n\times 1}$, then construct the matrices

$$P = [v_1 \ \cdots \ v_n] \in \mathbb{F}^{n\times n}, \quad D = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n).$$

The matrix P is invertible; and

$$APe_i = Av_i = \lambda_i v_i = \lambda_i Pe_i = PDe_i.$$

That is, $AP = PD$. Thus $P^{-1}AP = D$. That is, P is diagonalized by P.

Similarly, when a linear operator $T : V \to V$ is diagonalizable, we have a basis $B = \{v_1, \ldots, v_n\}$ for V such that $[T]_{B,B} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$ for scalars λ_i. In this case, $Tv_i = \lambda_i v_i$ for $1 \le i \le n$. Therefore, B consists of eigenvectors of T. Conversely, if $B = \{v_1, \ldots, v_n\}$ is a basis for V, where $Tv_i = \lambda_i$, then $[T]_{B,B} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$. This leads to the following result.

***Theorem* 6.21**
A linear operator T on a finite dimensional vector space is diagonalizable iff there exists a basis for V consisting of eigenvectors of T. In particular, a matrix $A \in \mathbb{F}^{n\times n}$ is diagonalizable iff there exists a basis of $\mathbb{F}^{n\times 1}$ consisting of eigenvectors of A.

So, we ask: when are there n linearly independent eigenvectors of a linear operator on a vector space of dimension n? The spectral theorem provides a partial answer, covering most cases that come up in applications. Another useful and easy partial answer is provided by the following theorem.

***Theorem* 6.22**

Eigenvectors associated with distinct eigenvalues of a linear operator on a finite dimensional vector space are linearly independent. In particular, each linear operator on a vector space of dimension n having n distinct eigenvalues is diagonalizable.

Proof Let $T : V \to V$ be a linear operator, where V is a vector space of dimension n. Let $\lambda_1, \ldots, \lambda_n$ be the distinct eigenvalues of T with corresponding eigenvectors as $v_1, \ldots, v_n$. We use induction on $k \in \{1, \ldots, n\}$.

For $k = 1$, since $v_1 \neq 0$, $\{v_1\}$ is linearly independent. Lay out the induction hypothesis: for $k = m$ suppose $\{v_1, \ldots, v_m\}$ is linearly independent. Now, for $k = m + 1$, suppose

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m + \alpha_{m+1} v_{m+1} = 0. \tag{6.1}$$

Then, $T(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m + \alpha_{m+1} v_{m+1}) = 0$ gives (since $T v_i = \lambda_i v_i$)

$$\alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \cdots + \alpha_m \lambda_m v_m + \alpha_{m+1} \lambda_{m+1} v_{m+1} = 0.$$

Multiply (6.1) with λ_{m+1} and subtract from the last equation to get

$$\alpha_1(\lambda_1 - \lambda_{m+1}) v_1 + \cdots + \alpha_m(\lambda_m - \lambda_{m+1}) v_m = 0.$$

By the Induction Hypothesis, $\alpha_i(\lambda_i - \lambda_{m+1}) = 0$. This implies each $\alpha_i = 0$, for $1 \leq i \leq m$. Then, (6.1) yields $\alpha_{m+1} v_{m+1} = 0$. Since $v_{m+1} \neq 0$, $\alpha_{m+1} = 0$. This completes the proof of linear independence of eigenvectors associated with distinct eigenvalues.

For the particular case, suppose T has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Then the associated eigenvectors $v_1, \ldots, v_n$ are linearly independent. Since $\dim(V) = n$, these vectors form a basis for V. By Theorem 6.21, the linear operator T is diagonalizable. ∎

For matrices, we thus conclude the following: Eigenvectors associated with distinct complex eigenvalues of a matrix are linearly independent. In particular, if a matrix of order n has n distinct complex eigenvalues, then it is similar to a diagonal matrix. The diagonalizing matrix is, in general, a matrix with complex entries.

If λ is an eigenvalue of a linear operator T, then its associated eigenvector u is a solution of $Tu = \lambda u$. Thus the maximum number of linearly independent eigenvectors associated with the eigenvalue λ is $\dim(N(T - \lambda I))$. This number and the algebraic multiplicity of λ have certain relations with the diagonalizability of T.

Let λ be an eigenvalue of a linear operator T on a finite dimensional vector space V. The number $\dim(N(T - \lambda I))$ is called the **geometric multiplicity** of λ.

Recall that the number m such that $(t-\lambda)^m$ divides $\chi_T(t)$ and $(t-\lambda)^{m+1}$ does not divide $\chi_T(t)$ is the algebraic multiplicity of λ.

***Theorem* 6.23**
The geometric multiplicity of any eigenvalue of a linear operator on a finite dimensional vector space is less than or equal to its algebraic multiplicity.

Proof Let V be a vector space of dimension n. Let T be a linear operator on V, and let λ be an eigenvalue of T. Suppose the geometric multiplicity of λ is k. Then we have k number of linearly independent eigenvectors of T associated with this eigenvalue λ, and no more. Let these eigenvectors be $v_1, \ldots, v_k$. Extend these to an ordered basis $B = \{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V. Then

$$Tv_1 = \lambda v_1, \ \ldots, \ Tv_k = \lambda v_k.$$

And for $j > k$, Tv_j can be any linear combination of $v_1, \ldots, v_n$. That is, the matrix of T with respect to B is given by

$$A := [T]_{B,B} = \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix},$$

where I_k is the identity matrix of order k and $C \in \mathbb{C}^{k\times(n-k)}$, $D \in \mathbb{C}^{(n-k)\times(n-k)}$ are some matrices. Now,

$$\chi_T(t) = \chi_A(t) = (t-\lambda)^k p(t)$$

for some polynomial $p(t)$ of degree $n-k$. Clearly, the algebraic multiplicity of λ is at least k. ∎

***Theorem* 6.24**
A linear operator T on a vector space of dimension n is diagonalizable iff the geometric multiplicity of each eigenvalue of T is equal to its algebraic multiplicity iff the sum of the geometric multiplicities of all eigenvalues of T is n.

Proof Suppose $T : V \to V$ is diagonalizable, where $\dim(V) = n$. Then there exists an ordered basis B of V consisting of eigenvectors of T such that $[T]_{B,B}$ is a diagonal matrix. If λ is an eigenvalue of T of algebraic multiplicity m, then in this diagonal matrix there are exactly m number of entries equal to λ. In the basis B there are exactly m number of eigenvectors associated with λ. Therefore, the geometric multiplicity of λ is m.

Conversely, suppose that the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity. Then corresponding to each eigenvalue λ, we have exactly that many linearly independent eigenvectors as its algebraic multiplicity. Collecting together these eigenvalues, we get n linearly independent eigenvectors; which form a basis for $\mathbb{F}^{n\times 1}$. Therefore, A is diagonalizable.

The second 'iff' statement follows since geometric multiplicity of each eigenvalue is at most its algebraic multiplicity. ∎

Example 6.25
Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. We see that $\chi_A(t) = \chi_B(t) = (t-1)^2$. The eigenvalue $\lambda = 1$ has algebraic multiplicity 2 for both A and B.

For geometric multiplicities, we solve $Ax = x$ and $By = y$.

Now, $Ax = x$ gives $x = x$, which is satisfied by the linearly independent vectors $(1,0)^{\mathrm{t}}$ and $(0,1)^{\mathrm{t}}$. Thus, $N(A - \lambda I)$ has dimension 2, which is equal to the geometric multiplicity of the eigenvalue 1 for A. Also, we see that A is diagonalizable; in fact, it is already a diagonal matrix.

$Bx = x$ with $x = (a,b)^{\mathrm{t}}$ gives $a + b = a$ and $b = b$. That is, $b = 0$ and a can be any complex number. For instance, $x = (1,0)^{\mathrm{t}}$. Now, $\dim(N(B - \lambda I)) = 1$. The geometric multiplicity of the eigenvalue 1 of B is 1, whereas its algebraic multiplicity is 2. Therefore, B is not diagonalizable. □

Exercises for § 6.4

1. Diagonalize the matrices: $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{bmatrix}$.
2. Which of the following linear operators are diagonalizable? If T is diagonalizable, find the ordered basis B and the diagonal matrix $[T]_{B,B}$.
   (a) $T : \mathbb{R}^3 \to \mathbb{R}^3;\ T(a,b,c) = (a+b+c,\ a+b-c,\ a-b+c)$.
   (b) $T : \mathbb{R}^3 \to \mathbb{R}^3;\ Te_1 = 0,\ Te_2 = e_1,\ Te_3 = e_2$.
   (c) $T : \mathbb{R}^3 \to \mathbb{R}^3;\ Te_1 = e_2,\ Te_2 = e_3,\ Te_3 = 0$.
   (d) $T : \mathbb{R}^3 \to \mathbb{R}^3;\ Te_1 = e_3,\ Te_2 = e_2,\ Te_3 = e_1$.
   (e) $T : \mathbb{F}_3[t] \to \mathbb{F}_3[t];\ T(a + bt + ct^2 + dt^3) = b + 2ct + 3dt^2$.
   (f) $T : \mathbb{F}_3[t] \to \mathbb{F}_3[t];\ T(p(t)) = dp/dt$.
   (g) $T : \mathbb{F}_3[t] \to \mathbb{F}_3[t];\ T(p(t)) = \int_0^{t} p(s)\,ds$.
3. Give three 3×3 matrices which cannot be diagonalized.
4. Which of the following matrices are diagonalizable?
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix},\quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},\quad \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},\quad \begin{bmatrix} 3/2 & -1/2 & 0 \\ -1/2 & 3/2 & 0 \\ 1/2 & -1/2 & 1 \end{bmatrix}.$$
5. For the diagonalizable matrices in Exercise 4, can you find eigenvectors of the matrix that form a basis for (a) $\mathbb{C}^{3\times 1}$ and (b) $\mathbb{R}^{3\times 1}$?

6.5 Jordan form

We now know that all linear operators (matrices) cannot be diagonalized. Non-diagonalizability means that we cannot have a basis $\{v_1, \ldots, v_n\}$ for the underlying space so that $T(v_j) = \lambda_j v_j$. In that case, we would like to have a basis which would bring the matrix of the linear operator to a nearly diagonal form. Specifically, if possible, we would try to construct an ordered basis $\{v_1, \ldots, v_n\}$ such that

$$T(v_j) = \lambda_j v_j \quad \text{or} \quad T(v_j) = v_{j-1} + \lambda_j v_j \quad \text{for each } j \in \{1, \ldots, n\}.$$

Notice that the matrix representation of T with respect to such a basis would possibly have nonzero entries on the diagonal and on the super diagonal (entries above the diagonal); all other entries being 0.

Let λ be an eigenvalue of a linear operator T on a finite dimensional complex vector space V with an associated eigenvector v_1. A **Jordan string** for λ is a list of nonzero vectors $v_1, \ldots, v_k$ such that

$$T(v_1) = \lambda_1 v_1,\ T(v_2) = v_1 + \lambda v_2,\ \ldots,\ T(v_k) = v_{k-1} + \lambda v_k;$$
$$T(v) \neq v_k + \lambda v \text{ for any } v \in V.$$

Equivalently,

$$(T - \lambda I)v_1 = 0,\ (T - \lambda I)v_2 = v_1,\ \ldots,\ (T - \lambda I)v_k = v_{k-1};$$
$$(T - \lambda I)v \neq v_k \text{ for any } v \in V.$$

Such a Jordan string $v_1, \ldots, v_k$ is said to start with v_1 and end with v_k. The number k is called the **length** of the Jordan string.

Example 6.26
Consider the linear operator $T : \mathbb{C}^5 \to \mathbb{C}^5$ given by

$$T(a, b, c, d, e) = (a, a + b, b + c, d, d + e).$$

With respect to the standard basis of $\mathbb{C}^5$, T is represented by the matrix

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

Since this matrix is lower triangular with all diagonal entries as 1, the only eigenvalue of T is 1 with algebraic multiplicity 5. Notice that

$$(T - I)(a, b, c, d, e) = (0, a, b, 0, d).$$

To find the associated eigenvectors, we solve $(T - I)(a, b, c, d, e) = 0$. We obtain $a = b = d = 0$ and c, e are arbitrary. The geometric multiplicity of the eigenvalue 1 is 2. The two linearly independent eigenvectors are

$$v_1 = (0, 0, 1, 0, 0), \quad w_1 = (0, 0, 0, 0, 1).$$

There are two types of Jordan strings for the eigenvalue 1, one starting with v_1 and the other starting with w_1.

For a Jordan string starting with v_1, we need to solve $(T - I)v_2 = v_1$. If $v_2 = (a, b, c, d, e)$, then the equation is

$$(0, a, b, 0, d) = (0, 0, 1, 0, 0).$$

This implies $a = 0$, $b = 1$, $d = 0$; leaving c, e arbitrary. This has again two linearly independent solutions obtained by taking $c = 1, e = 0$ or, $c = 0, e = 1$. We take the first one. That is,

$$v_2 = (0, 1, 1, 0, 0).$$

Next, we solve $(T - I)v_3 = v_2$. If $v_3 = (a, b, c, d, e)$, then the equation is

$$(0, a, b, 0, d) = (0, 1, 1, 0, 0).$$

It gives $a = 1$, $b = 1$, $d = 0$ leaving c, e arbitrary. Once more we have two linearly independent solutions; we take up one of them, say,

$$v_3 = (1, 1, 1, 0, 0).$$

Further, we set up $(T - I)v_4 = v_3$. If $v_4 = (a, b, c, d, e)$, then

$$(0, a, b, 0, d) = (1, 1, 1, 0, 0).$$

It does not have a solution; and we stop here with the Jordan string v_1, v_2, v_3.

Notice that we could have chosen v_2 differently, and once again, v_3 could have been chosen in a different way.

Now, with the eigenvector $w_1 = (0, 0, 0, 0, 1)$, we proceed similarly. Suppose $w_2 = (a, b, c, d, e)$ satisfies $(T - I)w_2 = w_1$. Then

$$(0, a, b, 0, d) = (0, 0, 0, 0, 1).$$

That is, $a = 0$, $b = 0$, $d = 1$ and c, e arbitrary. We take one of the two possibilities to obtain

$$w_2 = (0, 0, 1, 1, 0).$$

Next, we set up $(A-I)w_3 = w_2,\ w_3 = (a,b,c,d,e)$; to obtain

$$(0,a,b,0,d) = (0,0,1,1,0).$$

This equation has no solutions; we stop with the Jordan string w_1, w_2.

What happens if we apply $(A-I)$ successively to the vectors v_3 and w_2? Notice that $(T-I)(a,b,c,d,e) = (0,a,b,0,d)$. Therefore, we see that

$$\begin{aligned}
(A-I)v_3 &= (A-I)(1,1,1,0,0) = (0,1,1,0,0) = v_2,\\
(A-I)^2 v_3 &= (A-I)v_2 = (A-I)(0,1,1,0,0) = (0,0,1,0,0) = v_1,\\
(A-I)^3 v_3 &= (A-I)v_1 = (A-I)(0,0,1,0,0) = (0,0,0,0,0).
\end{aligned}$$

$$\begin{aligned}
(A-I)w_2 &= (A-I)(0,0,1,1,0) = (0,0,0,0,1) = w_1,\\
(A-I)^2 w_2 &= (A-I)w_1 = (A-I)(0,0,0,1,0) = (0,0,0,0,0).
\end{aligned}$$

That is, $(A-I)^3 v_3 = 0 = (A-I)^2 w_2$. $\square$

Let us look at constructing v_2 from v_1 in a Jordan string for an eigenvalue λ. Suppose that $N(T-\lambda I)$ is a proper subspace of $N(T-\lambda I)^2$, and we have already chosen a nonzero vector v_1 from $N(T-\lambda I)$. We know that there exist nonzero vectors in $N(T-\lambda I)^2 \setminus N(T-\lambda I)$. But there is no guarantee that such a vector v_2 exists that may satisfy the equation $(T-\lambda I)v_2 = v_1$. If this happens, then $v_1 \in R(T-\lambda I)$. Conversely, if $v_1 \in R(T-\lambda I)$, then such a vector v_2 exists. This suggests that

> for the construction of a Jordan string, we must choose the starting vector v_1 from $N(T-\lambda I) \cap R(T-\lambda I)$.

Notice that if $N(T-\lambda I) \cap R(T-\lambda I) = \{0\}$, then such a choice is not possible; and then the Jordan string with starting vector $v_1 \in N(T-\lambda I)$ will have only v_1 in it.

***Theorem* 6.27 (Jordan Strings)**
Let T be a linear operator on a finite dimensional complex vector space V. Then there exists a basis of V, which is a disjoint union of Jordan strings for eigenvalues of T. Further, if λ is an eigenvalue of T, then in this basis, the number of Jordan strings for λ is equal to the geometric multiplicity of λ.

Proof We use induction on n. For $n = 1$, let v be an eigenvector associated with the eigenvalue λ of T. Then $Tv = \lambda v$. As $\dim(V) = 1$, $\{v\}$ is a basis of V. If $x \in V$, then $x = \alpha v$ for some scalar α. Now,

$$(T-\lambda I)x = Tx - \lambda x = \alpha T v - \lambda \alpha v = \alpha\lambda v - \alpha\lambda v = 0 \neq v.$$

Therefore, the Jordan string with v consists of this single vector v. Since $\{v\}$ is a basis of V, the statements of the theorem hold true.

Lay out the induction hypothesis that for all complex vector spaces of dimension less than n, the statements are true. Let $T : V \to V$ be a linear operator, where V is a complex vector space of dimension n. Let λ be an eigenvalue of T. Then $\mathrm{null}(T - \lambda I) \geq 1$. Write $U = R(T - \lambda I)$. By the rank nullity theorem, $r = \dim(U) = \mathrm{rank}(T - \lambda I) < n$. If $x \in U$, then $(T - \lambda I)x \in R(T - \lambda I) = U$. Thus, the restriction of $T - \lambda I$ to U is a linear operator. Call this restriction linear operator as S. That is, S is the linear operator given by

$$S : U \to U, \quad Sx = (T - \lambda I)(x), \quad \text{for each } x \in U = R(T - \lambda I).$$

We now break the proof into two cases.

Case 1: Assume that $N(T - \lambda I) \cap U \neq \{0\}$. Then there exists a nonzero vector $x \in U$ such that $(T - \lambda I)x = 0$. So, $Sx = 0 = 0\,x$. Thus 0 is an eigenvalue of S, and $s = \mathrm{null}(S) \geq 1$. Notice that s is the geometric multiplicity of the eigenvalue 0 of S. By the induction hypothesis, there exists an ordered basis $E = \{v_1, \ldots, v_r\}$ of U, which is a disjoint union of Jordan strings. Moreover, corresponding to the eigenvalue 0 of S, there are s number of Jordan strings listed in the basis E starting with a vector from $N(S)$.

We first look at any possible nonzero eigenvalue of S. Let $x_1, \ldots, x_j$ be a Jordan string for a nonzero eigenvalue μ of S listed among $v_1, \ldots, v_r$. Then

$$\begin{gathered} Sx_1 = \mu x_1,\ Sx_2 = x_1 + \mu x_2,\ \ldots, Sx_j = x_{j-1} + \mu x_j; \\ Sx \neq x_j + \mu x \text{ for any } x \in U. \end{gathered} \tag{6.2}$$

As $S = T - \lambda I$ on U, it follows that

$$\begin{gathered} Tx_1 = (\lambda + \mu)x_1,\ Tx_2 = x_1 + (\lambda + \mu)x_2,\ \ldots, Tx_j = x_{j-1} + (\lambda + \mu)x_j; \\ Tx \neq x_j + (\lambda + \mu)x \text{ for any } x \in U. \end{gathered}$$

Look at the last inequality. Can it happen that there exists a vector $y \in V$ such that $Ty = x_j + (\lambda + \mu)y$? If it so happens, then $(T - \lambda I)y = x_j + \mu y$. Then $x_j + \mu y \in R(T - \lambda I) = U$. But $x_j \in U$. Therefore, $y \in U$. This will contradict (6.2). Hence the last inequality is replaced by

$$Tx \neq x_j + (\lambda + \mu)x \quad \text{for any } x \in V.$$

And, we conclude that any Jordan string for an eigenvalue $\mu \neq 0$ of S listed among $v_1, \ldots, v_r$ is a Jordan string for an eigenvalue $\lambda + \mu$ of T.

Next, we look at any Jordan string for the eigenvalue 0. Any such Jordan string looks like a list of vectors $u_1, \ldots, u_k$ in U with

$$Su_1 = 0,\ Su_2 = u_1,\ \ldots, Su_k = u_{k-1};\ Sx \neq u_k \text{ for any } x \in U. \tag{6.3}$$

The vectors u_i are from the set $v_1, \ldots, v_r$. Then (6.3) implies that

$$Tu_1 = \lambda u_1,\ Tu_2 = u_1 + \lambda u_2,\ \ldots, Tu_k = u_{k-1} + \lambda u_k.$$

Since $u_k \in U = R(T - \lambda I)$, there exists a vector $u_{k+1} \in V$ such that

$$(T - \lambda I)u_{k+1} = u_k.$$

That is, $Tu_{k+1} = u_k + \lambda u_{k+1}$. Further, if for some $x \in V$, $(T - \lambda I)x = u_{k+1}$, then $u_{k+1} \in R(T - \lambda I) = U$. Then $Su_{k+1} = (T - \lambda I)(u_{k+1}) = u_k$ contradicts the inequality in (6.3). Hence

$$Tx \neq u_{k+1} + \lambda x \quad \text{for any } x \in V.$$

Therefore, we have an enlarged Jordan string $u_1, \ldots, u_k, u_{k+1}$ for the eigenvalue λ of T. Observe that such a vector u_{k+1} need not be in $N(T - \lambda I)$. This way, starting with each Jordan string of the eigenvalue 0 of S, we end up with an enlarged Jordan string for the eigenvalue λ of T; and they are s in number.

Each of these enlarged Jordan strings has length one more than the corresponding one listed in $v_1, \ldots, v_r$, where one vector has been added to each Jordan string, at the end. Let $w_1, \ldots, w_s$ be the added vectors. The set $\{v_1, \ldots, v_r,\ w_1, \ldots, w_s\}$ is now a disjoint union of Jordan strings with s number of Jordan strings for the eigenvalue λ of T. These s number of Jordan strings start with a vector from $N(T - \lambda I) \cap U$ and end with the vectors $w_1, \ldots, w_s$. And the other Jordan strings of nonzero eigenvalues of S are kept as they are, accounting for Jordan strings of eigenvalues of T other than λ.

The starting vectors of these s number of enlarged Jordan strings form a basis for $N(T - \lambda I) \cap U$. Extend the set of these starting vectors to a basis of $N(T - \lambda I)$. Since $\operatorname{null}(T - \lambda I) = n - r$, we obtain $n - r - s$ number of linearly independent vectors $z_1, \ldots, z_{n-r-s} \in N(T - \lambda I) \setminus U$ so that these vectors and the starting vectors of the enlarged Jordan strings form a basis for $N(T - \lambda I)$. Notice that if $n = r + s$, then we do not need any such z_i. Further, $(T - \lambda I)x \neq z_i$ for any $x \in V$, since otherwise, z_i would be a vector in U. Therefore, each of these vectors z_i is a Jordan string of length 1 for the eigenvalue λ of T.

Now, the set $C = \{v_1, \ldots, v_r,\ w_1, \ldots, w_s,\ z_1, \ldots, z_{n-r-s}\}$ is a disjoint union of Jordan strings for the eigenvalues of T. We claim that C is linearly independent. To prove this, suppose that for scalars α_i, β_j and γ_ℓ,

$$\alpha_1 v_1 + \cdots + \alpha_r v_r + \beta_1 w_1 + \cdots + \beta_s w_s + \gamma_1 z_1 + \cdots + \gamma_{n-r-s} z_{n-r-s} = 0. \quad (6.4)$$

Apply $T - \lambda I$ to both the sides. Since $(T - \lambda I)z_\ell = 0$ for each ℓ, we have

$$\alpha_1 (T - \lambda I)v_1 + \cdots + \alpha_r (T - \lambda I)v_r + \beta_1 (T - \lambda I)w_1 + \cdots + \beta_s (T - \lambda I)w_s = 0.$$

We look at the effect of applying $(T - \lambda I)$ in three stages. First, look at the starting vectors of the Jordan strings. They are from $N(T - \lambda I)$. If v is a such a vector, then $(T - \lambda I)v = 0$. Thus, all starting vectors of the Jordan strings vanish from the sum.

Second, look at all other v_i in the Jordan strings. Since $(T - \lambda I)v_i = v_{i-1}$, each such v_i in the Jordan string will have coefficient as α_{i+1} in the sum instead of the previous α_i. This will reintroduce the starting vectors of the Jordan strings with updated coefficients. Further, the vectors with which Jordan strings end are absent in the sum.

Third, look at $(T - \lambda I)w_j$. Each w_j is the last vector in an enlarged Jordan string. Thus $w_j = (T - \lambda I)v_p$ for some p; this v_p is the vector with which a Jordan string ends. Thus these vectors v_p are reintroduced in the sum with coefficients as β_j. Further, the vectors w_j are absent in the sum.

Therefore, the simplified sum is a liner combination of all v_i with updated coefficients where vectors v_p with which a Jordan string ends has the coefficient β_j of the corresponding next vector w_j. In this sum all αs do not occur, but all βs occur as coefficients of these vectors v_p.

For instance, if the list v_1, v_2, v_3, w_1 is an enlarged Jordan string, and another Jordan string starts from v_4, then after applying $(T - \lambda I)$ the part for v_1, v_2, v_3, w_1 in the sum is

$$\alpha_1(T - \lambda I)v_1 + \alpha_2(T - \lambda I)v_2 + \alpha_3(T - \lambda I)v_3 + \beta_1(T - \lambda I)w_1.$$

Since $(T - \lambda I)v_1 = 0$, $(T - \lambda I)v_2 = v_1$, $(T - \lambda I)v_3 = v_2$ and $(T - \lambda I)w_1 = v_3$, this portion in the simplified sum would look like

$$\alpha_2 v_1 + \alpha_3 v_2 + \beta_1 v_3.$$

Coming back to the proof, we have shown that $\{v_1, \ldots, v_r\}$ is linearly independent. Thus, all scalars in the simplified sum are 0. In particular, each β_j is 0. Then (6.4) simplifies to

$$\alpha_1 v_1 + \cdots + \alpha_r v_r = -\gamma_1 z_1 - \cdots - \gamma_{n-r-s} z_{n-r-s}.$$

Here, the left hand side is a vector in U and the right hand side is a vector in $\operatorname{span}(N(T - \lambda I) \setminus U)$. Since $U \cap \operatorname{span}(N(T - \lambda I) \setminus U) = \{0\}$, the vector on each side of the above equation is 0. However, each of the sets $\{v_1, \ldots, v_r\}$ and $\{z_1, \ldots, z_{n-r-s}\}$ is linearly independent. We conclude that each α_i is equal to zero, and each γ_ℓ is equal to 0. This proves our claim.

Now the set $B := \{v_1, \ldots, v_r,\ w_1, \ldots, w_s,\ z_1, \ldots, z_{n-r-s}\}$ is a linearly independent set having exactly n vectors. So, it is a basis for V, which is a disjoint union of Jordan strings of eigenvalues of T.

Case 2: Suppose that $N(T - \lambda I) \cap U = \{0\}$. Then there exists no nonzero vector in U such that $(T - \lambda I)x = 0$; consequently, 0 is not an eigenvalue of S. But λ is an eigenvalue of T anyway. Going through the proof in the first case, we find that each Jordan string among the vectors $v_1, \ldots, v_r$ is a Jordan string for a nonzero eigenvalue of S. The starting vectors of these Jordan strings are not from $N(T - \lambda I)$. Thus, we do not need to enlarge the set of vectors $v_1, \ldots, v_r$ by adjoining any w_j. we rather take $\{z_1, \ldots, z_{n-r}\}$ as a basis for $N(T - \lambda I)$. Each vector z_j is a Jordan string on its own. Essentially, in the previous construction, we take $s = 0$. The proof of the first case holds for this case also.

This completes the proof of the first statement. For the second statement about the number of Jordan strings, notice that each $z_\ell \in N(T - \lambda I)$ is itself a Jordan string of length 1 for the eigenvalue λ of T. (Such a situation arises when $n > r + s$.) Thus for the eigenvalue λ, we have in total $n - r - s + s = n - r$ number of Jordan strings in B. Since $n - r = \text{null}(T - \lambda I)$, the number of Jordan strings for the eigenvalue λ of T is the geometric multiplicity of λ. Jordan strings for other eigenvalues of T remain unchanged. ∎

The inductive construction in the proof of Theorem 6.27 goes as follows. Let λ be an eigenvalue of $T : V \to V$. Write $U := R(T - \lambda I)$; and define the linear operator $S : U \to U$ as the restriction of $T - \lambda I$ to U. Without loss of generality, suppose the first s Jordan strings below (each Jordan string in each line) correspond to the eigenvalue 0 of S; and other Jordan strings may correpond to other eigenvalues of S.

$$
\begin{array}{cccc}
v_{11}, & v_{12}, & \cdots & v_{1n_1} \\
 & & \vdots & \\
v_{s1}, & v_{s2}, & \cdots & v_{sn_s} \\
 & & \vdots & \\
v_{k1}, & v_{k2}, & \cdots & v_{kn_k}
\end{array}
$$

When 0 is not an eigenvalue of S, the number s is equal to 0. If such a Jordan string corresponds to an eigenvalue $\mu - \lambda$ of S, then it also corresponds to the eigenvalue μ of T.

Here, $n_1 + \cdots + n_k = r = \text{rank}(T - \lambda I) = \dim(U)$. The first vectors of the first s Jordan strings are in $N(S)$. That is,

$$v_{11}, \ldots, v_{s1}, \ldots, v_{k1} \in N(T - \lambda I) \cap R(T - \lambda I).$$

Next, we construct w_is such that

$$(T - \lambda I) w_i = v_{in_i} \quad \text{for } 1 \le i \le s.$$

Also we construct a basis $\{z_1, \ldots, z_{n-r-s}\}$ for $\text{span}\,(N(T - \lambda I) \setminus R(T - \lambda I))$. The enlarged basis for V consisting of Jordan strings of eigenvalues of T may be arranged as follows:

$$\begin{array}{ccccc} v_{11}, & v_{12}, & \cdots & v_{1n_1}, & w_1 \\ & & \vdots & & \\ v_{s1}, & v_{s2}, & \cdots & v_{sn_s}, & w_s \\ & & \vdots & & \\ v_{k1}, & v_{k2}, & \cdots & v_{kn_k} & \\ z_1 & & & & \\ \vdots & & & & \\ z_{n-r-s} & & & & \end{array}$$

With respect to a basis that consists of disjoint union of Jordan strings, how does a matrix representation of a linear operator look like?

Let T be a linear operator on a complex vector space V of dimension n having distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, whose geometric multiplicities are $s_1, \ldots, s_k$, and algebraic multiplicities $m_1, \ldots, m_k$, respectively. Then we have a basis in which there are s_i number of Jordan strings for λ_i. We choose these s_i Jordan strings in some order, so that we may talk of first Jordan string for λ_i, second Jordan string for λ_i, and so on. We also take the eigenvalues in some order, say, λ_i is the ith eigenvalue. The vectors in any Jordan string are already ordered. Then we construct an ordered basis from these Jordan strings by listing all vectors (in order) in the first Jordan string for λ_1; next, the vectors from the second Jordan string for λ_1, and so on. This completes the list of m_1 vectors for λ_1. Next, we list all vectors in order from the first Jordan string for λ_2, and so on. After the list is complete we obtain an ordered basis $B = \{v_1, \ldots, v_n\}$ of V. Notice that for any v_j in this basis, we have

$$T(v_j) = \lambda v_j \quad \text{or} \quad T(v_j) = v_{j-1} + \lambda v_j$$

for an eigenvalue λ of T. For instance, if the first Jordan string for λ_1 is of length ℓ, then

$$T(v_1) = \lambda_1 v_1,\ T(v_2) = v_1 + \lambda_1 v_2,\ \ldots, T(v_\ell) = v_{\ell-1} + \lambda_1 v_\ell,$$

and after this the next Jordan string starts. In the matrix representation of T with respect to B, this will correspond to a block of order ℓ having diagonal entries as λ_1 and super diagonal entries (entries above the diagonal) as 1.

Then the next Jordan string will give rise to another block of diagonal entries λ_1 and super diagonal entries as 1. This block will have the order as the length of the Jordan string. This way, when all the Jordan strings for λ_1 are complete, another similar block will start with a Jordan string for λ_2, and so on.

With respect to this basis B, the linear operator T will have the matrix representation in the form

$$[T]_{B,B} = \operatorname{diag}(J_1, J_2, \ldots, J_k), \tag{6.5}$$

where each J_i is again a block diagonal matrix looking like

$$J_i = \operatorname{diag}(\tilde{J}_1(\lambda_i), \tilde{J}_2(\lambda_i), \ldots, \tilde{J}_{s_i}(\lambda_i)),$$

wit s_i as the geometric multiplicity of the eigenvalue λ_i. Each matrix $\tilde{J}_j(\lambda_i)$ here has the form

$$\tilde{J}_j(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

The missing entries are all 0. Such a matrix $\tilde{J}_j(\lambda_i)$ is called a **Jordan block** with diagonal entries λ_i. The order of this Jordan block is the length of the corresponding Jordan string for the eigenvalue λ_i. In the matrix $[T]_{B,B}$, the number of Jordan blocks with diagonal entries λ_i is the number of Jordan strings for the eigenvalue λ_i, which is equal to the geometric multiplicity of the eigenvalue λ_i. Any matrix which is in the block diagonal form (6.5) is said to be in **Jordan form**. Using Theorem 6.27 we obtain the following result.

***Theorem* 6.28 (Jordan form)**
Let T be a linear operator on a finite dimensional complex vector space V. Then there exists a basis B for V such that $[T]_{B,B}$ is in Jordan form, whose diagonal entries are the eigenvalues of T. The number of Jordan blocks in $[T]_{B,B}$ with diagonal entry λ is the geometric multiplicity of λ. Moreover, the number $m_k(\lambda)$ of Jordan blocks of order k with diagonal entry λ, is given by

$$m_k(\lambda) = \operatorname{rank}((T - \lambda I)^{k-1}) - 2\operatorname{rank}((T - \lambda I)^k) + \operatorname{rank}((T - \lambda I))^{k+1}$$

for $1 \le k \le n$. Consequently, the Jordan form of T is unique up to a permutation of the blocks.

In the formula for $m_k(\lambda)$, we use the convention that for any matrix B of order n, B^0 is the identity matrix of order n.

Proof Existence of Jordan form and the statement about the number of Jordan blocks with diagonal entry as λ follow from Theorem 6.27.

For the formula for $m_k(\lambda)$, let λ be an eigenvalue of T. Write $J := [T]_{B,B}$. Suppose $1 \le k \le n$. Observe that $[T - \lambda I]_{B,B} = J - \lambda I$. From Theorem 4.16

it follows that $\text{rank}((T-\lambda I)^i) = \text{rank}((J-\lambda I)^i)$ for each i. Therefore, it is enough to prove the formula for J instead of T.

We use induction on n. In the basis case, $J = [\lambda]$. Here, $k = 1$; $m_k(\lambda) = m_1(\lambda) = 1$. On the right hand side, due to the convention,

$$(J-\lambda I)^{k-1} = I = [1],\ (J-\lambda I)^k = [0]^1 = [0],\ (J-\lambda I)^{k+1} = [0]^2 = [0].$$

So, the formula holds for $n = 1$.

Lay out the induction hypothesis that for all matrices J in Jordan form of order less than n, the formula holds. Let J be a matrix of order n, which is in Jordan form. We consider two cases.

Case 1: Let J have a single Jordan block corresponding to λ. That is,

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & \lambda \end{bmatrix}, \quad J-\lambda I = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & 0 \end{bmatrix}.$$

Here $m_1(\lambda) = 0,\ m_2(\lambda) = 0, \ldots, m_{n-1}(\lambda) = 0$ and $m_n(\lambda) = 1$. By direct computation, we see that $(J-\lambda I)^2$ has 1 on the super-super-diagonal, and 0 elsewhere. Proceeding similarly for higher powers of $J-\lambda I$, we obtain

$$\text{rank}(J-\lambda I) = n-1,\ \text{rank}((J-\lambda I)^2) = n-2,\ \ldots, \text{rank}((J-\lambda I)^i = n-i,$$
$$\text{rank}((J-\lambda I)^n) = 0,\ \text{rank}((J-\lambda I)^{n+1} = 0,\ \ldots.$$

$$\begin{aligned} \text{Then for } k < n,\quad & \text{rank}((J-\lambda I)^{k-1}) - 2\,\text{rank}((J-\lambda I)^k) + \text{rank}((J-\lambda I)^{k+1}) \\ & = (n-(k-1)) - 2(n-k) + (n-k-1) = 0. \end{aligned}$$

$$\begin{aligned} \text{And for } k = n,\quad & \text{rank}((J-\lambda I)^{k-1}) - 2\,\text{rank}((J-\lambda I)^k) + \text{rank}((J-\lambda I)^{k+1}) \\ & = (n-(n-1)) - 2\times 0 + 0 = 1 = m_n(\lambda). \end{aligned}$$

Case 2: Suppose J has more than one Jordan block corresponding to λ. By reordering of blocks, we assume that the first Jordan block in J corresponds to λ and has order $r < n$. Then $J-\lambda I$ can be written in block form as

$$J-\lambda I = \begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix},$$

where C is the Jordan block of order r with diagonal entries as 0, and D is the matrix of order $n-r$ consisting of other blocks of $J-\lambda I$. Then, for any j,

$$(J-\lambda I)^j = \begin{bmatrix} C^j & 0 \\ 0 & D^j \end{bmatrix}.$$

Therefore,

$$\operatorname{rank}(J-\lambda I)^j = \operatorname{rank}(C^j) + \operatorname{rank}(D^j).$$

Write $m_k(\lambda, C)$ and $m_k(\lambda, D)$ for the number of Jordan blocks of order k for the eigenvalue λ that appear in C and in D, respectively. Then

$$m_k(\lambda) = m_k(\lambda, C) + m_k(\lambda, D).$$

By the induction hypothesis,

$$\begin{aligned} m_k(\lambda, C) &= \operatorname{rank}(C^{k-1}) - 2\operatorname{rank}(C^k) + \operatorname{rank}(C)^{k+1}, \\ m_k(\lambda, D) &= \operatorname{rank}(D^{k-1}) - 2\operatorname{rank}(D^k) + \operatorname{rank}(D)^{k+1}. \end{aligned}$$

It then follows that

$$m_k(\lambda) = \operatorname{rank}((J-\lambda I)^{k-1}) - 2\operatorname{rank}((J-\lambda I)^k) + \operatorname{rank}((J-\lambda I))^{k+1}.$$

Since the number of Jordan blocks of order k corresponding to each eigenvalue of T is uniquely determined, the Jordan form of T is also uniquely determined up to a permutation of blocks. ∎

Taking V as $\mathbb{C}^{n\times 1}$, we obtain the matrix version of Theorem 6.28.

***Theorem* 6.29 (*Jordan form*)**
Each matrix $A \in \mathbb{C}^{n\times n}$ is similar to a matrix in Jordan form whose diagonal entries are the eigenvalues of A. In a Jordan form, the number of Jordan blocks with diagonal entry λ is the geometric multiplicity of λ. The Jordan form of A is unique up to a permutation of Jordan blocks.

To obtain a Jordan form of a given matrix, we may construct a basis consisting of Jordan strings. Alternatively, we may use the formula for m_k in Theorem 6.28.

Example 6.30

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$ of Example 6.26.

There, we had constructed the Jordan strings

$$v_1 = (0,0,1,0,0)^{\mathrm t},\ v_2 = (0,1,1,0,0)^{\mathrm t},\ v_3 = (1,1,1,0,0)^{\mathrm t};$$
$$w_1 = (0,0,0,0,1)^{\mathrm t},\ w_2 = (0,0,1,1,0)^{\mathrm t}.$$

In the ordered basis $\{v_1, v_2, v_3, w_1, w_2\}$, the matrix of A is given by $J = P^{-1}AP$ with P as the matrix whose columns are these basis vectors. That is, the

Jordan form of A is

$$J = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Alternatively, we compute the ranks of successive powers of $A - \lambda I$. Notice that the only eigenvalue of A is 1. We find that

$$\operatorname{rank}(A - I) = 3,\ \operatorname{rank}((A - I)^2) = 1,\ \operatorname{rank}((A - I)^{3+i}) = 0 \text{ for } i \geq 0.$$

Then

$$m_1(1) = 5 - 2 \times 3 + 1 = 0, \quad m_2(1) = 3 - 2 \times 1 + 0 = 1$$
$$m_3(1) = 1 - 2 \times 0 + 0 = 1, \quad m_4(1) = 0 = m_5(1).$$

From this information, J can be constructed. Verify that it is the same as computed above, up to a permutation of the blocks. ∎

In the Jordan strings, observe that the vectors used are from $N\big((T - \lambda I)^j\big)$. Such a vector is called a **generalized eigenvector** corresponding to the eigenvalue λ of T. For a matrix A, the similarity matrix P in $J = P^{-1}AP$ has the columns as the vectors from the basis in which J represents the matrix A. These vectors are specific generalized eigenvectors of A.

The uniqueness of a Jordan form can be made exact by first ordering the eigenvalues of A and then arranging the blocks corresponding to each eigenvalue, which are now together, in some order, say in ascending order of their size. In doing so, the Jordan form of any matrix becomes unique. Such a Jordan form is called the **Jordan canonical form** of a matrix. It then follows that if two matrices are similar, then they have the same Jordan canonical form. Moreover, uniqueness also implies that two dissimilar matrices will have different Jordan canonical forms. Therefore, Jordan form characterizes similarity of matrices. It implies the following:

> Two square matrices A and B of the same order are similar iff they have the same eigenvalues, and for each eigenvalue λ, for each $j \in \mathbb{N}$, $\operatorname{rank}(A - \lambda I)^j = \operatorname{rank}(B - \lambda I)^j$.

As an application of Jordan form, we will show that each matrix is similar to its transpose. Let $A \in \mathbb{C}^{n \times n}$. We know that a scalar λ is an eigenvalue of A iff it is an eigenvalue of A^{t}. Further, $\operatorname{rank}(A^{t}) = \operatorname{rank}(A)$. Thus, for any eigenvalue λ of A and for any j, we have $\operatorname{rank}\big((A - \lambda I)^j\big) = \operatorname{rank}\big((A^{t} - \lambda I)^j\big)$. Consequently, A and A^{t} are similar.

It also follows from the Jordan form that if m is the algebraic multiplicity of an eigenvalue λ, then one can always choose m linearly independent generalized eigenvectors; see Exercise 7 of Section 6.3. Further, the following is guaranteed (See Exercise 8.):

> If the linear system $(A-\lambda I)^k x = 0$ has $r < m$ number of linearly independent solutions, then $(A-\lambda I)^{k+1}$ has at least $r+1$ number of linearly independent solutions.

This result is often more useful in computing the exponential of a matrix rather than using explicitly the Jordan form, which is comparatively difficult to compute.

Exercises for § 6.5

1. In Example 6.26, explore the other choices for v_2, v_3, and w_2. Construct the corresponding Jordan strings.
2. Let $A \in \mathbb{F}^{n\times n}$. Let $B = P^{-1}AP$ for an invertible matrix $P \in \mathbb{F}^{n\times n}$. Show that if $\{v_1, \ldots, v_m\}$ is a basis of $N(A)$, then $\{Pv_1, \ldots, Pv_m\}$ is a basis of $N(B)$. This would prove directly that null(B) = null(A); consequently, rank(B) = rank(A).
3. Determine the Jordan forms of the following matrices:
   (a) $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} -2 & -1 & -3 \\ 4 & 3 & 3 \\ -2 & 1 & -1 \end{bmatrix}$.
4. Determine the matrix $P \in \mathbb{C}^{3\times 3}$ such that $P^{-1}AP$ is in Jordan form, where A is the matrix in Exercise 3(b).
5. Let A be the 5×5 matrix whose first row is $(0,1,1,0,1)$, the second row is $(0,0,1,1,1)$, and all other rows are zero rows. Find the Jordan form of A.
6. Let A be a 7×7 matrix with $\chi_A(t) = (t-2)^4(3-t)^3$. Suppose that in the Jordan form of A, the largest block for each of the eigenvalues is 2. Show that there are only two possible Jordan forms of A; and determine those Jordan forms.
7. Let λ be an eigenvalue of a matrix $A \in \mathbb{C}^{n\times n}$. Let m be the algebraic multiplicity of λ. Prove that null$((A-\lambda I)^m) = m$.
   [Hint: This is Exercise 7 of § 6.3. But use Jordan form now.]
8. Let λ be an eigenvalue of a matrix $A \in \mathbb{C}^{n\times n}$. Let m be the algebraic multiplicity of λ. For each $k \in \mathbb{N}$, if null$((A-\lambda I)^k) < m$, then show that null$((A-\lambda I))^k <$ null$((A-\lambda I)^{k+1}$.
   [Hint: Show that $N(A-\lambda I)^i \subseteq N(A-\lambda I)^{i+1}$. Then use Exercise 7.]
9. Using the Jordan form of a matrix show that a matrix A is diagonalizable iff for each eigenvalue of A, its geometric multiplicity is equal to its algebraic multiplicity.

10. Let $A \in \mathbb{C}^{n\times n}$. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of A having algebraic multiplicities $m_1, \ldots, m_k$, respectively. Prove that A is similar to a matrix of the form $\operatorname{diag}(B_1, \ldots, B_k)$, where B_i is of order m_i. This is called the block diagonal form of A.
[Hint: Try proving it without using Jordan form.]

11. For a matrix $A \in \mathbb{C}^{n\times n}$, suppose B is an ordered basis of $\mathbb{C}^{n\times 1}$ such that $[A]_{B,B}$ is in Jordan form. Reorder the basis B to obtain a basis C so that $[A]_{C,C} = [A]^{\mathrm{t}}_{B,B}$. Then conclude that A is similar to A^{t}.

12. The matrix $A = \begin{bmatrix} 2 & 2 & 3 \\ 1 & 3 & 3 \\ -1 & -2 & -3 \end{bmatrix}$ has the Jordan form $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Thus there exist vectors $v_1, v_2, v_3 \in \mathbb{F}^{3\times 1}$ such that $Av_1 = v_1$, $Av_2 = v_1 + v_2$ and $Av_3 = v_3$. Now, $N(A - I)$ has a basis $\{x_1, x_2\}$, where $x_1 = (1, -2, 0)^{\mathrm{t}}$ and $x_2 = (-3, 0, 1)^{\mathrm{t}}$. Taking $v_1 = x_1$, we see that $Av_2 = x_1 + v_2$ has no solutions. Also, by taking $v_1 = x_2$, we find that $Av_2 = x_2 + v_2$ has no solutions. Why does it happen?
[Hint: With the basis $\{(1, 1, -1)^{\mathrm{t}}, (-3, 0, 1)^{\mathrm{t}}\}$ we get the Jordan form.]

6.6 Singular value decomposition

Let $T : V \to W$ be a linear transformation, where V and W are inner product spaces with $\dim(V) = n$ and $\dim(W) = m$. There are two natural self-adjoint operators associated with T, namely, $T^*T : V \to V$ and $TT^* : W \to W$. Suppose $\lambda \in \mathbb{R}$ is an eigenvalue of T^*T with an associated eigenvector $v \in V$. Then $T^*Tv = \lambda v$ implies that

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\|v\|^2.$$

Since $\|v\| > 0$, we see that $\lambda \geq 0$. If there are r number of positive eigenvalues of T^*T, for $0 \leq r \leq n$, then all eigenvalues can be arranged in a decreasing list such as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n.$$

The non-negative square roots of eigenvalues of the self-adjoint linear operator $T^*T : V \to V$ are called the **singular values** of T. If there are r number of positive eigenvalues of T^*T, then the singular values of T are written as

$$s_1 \geq s_2 \geq \cdots \geq s_r > 0 = s_{r+1} = \cdots = s_n.$$

If $\lambda > 0$ is an eigenvalue of T^*T with an associated eigenvector v, then $T^*Tv = \lambda v$. It yields $(TT^*)(Tv) = \lambda(Tv)$. Now, $\lambda v \neq 0$ implies $Tv \neq 0$. Hence $\lambda > 0$ is also an eigenvalue of TT^* with an associated eigenvector Tv. Similarly, it follows that each positive eigenvalue of TT^* is also an eigenvalue of T^*T.

Further, the spectral theorem implies that the self-adjoint linear operator T^*T is represented by a diagonal matrix. In such a diagonal matrix, the only nonzero entries are $s_1^2, \ldots, s_r^2$. Therefore, T^*T has rank r. It then follows that $s_1^2, \ldots, s_r^2$ are the only positive eigenvalues of the self-adjoint linear operator TT^*. There are $n - r$ number of zero eigenvalues of T^*T, where as TT^* has $m - r$ number of zero eigenvalues.

The following theorem gives much more information than this by representing T in terms of its singular values.

***Theorem* 6.31 (Singular value decomposition, SVD)**
Let V and W be inner product spaces of dimensions n and m, respectively. Let $T : V \to W$ be a linear transformation with r positive singular values $s_1 \geq \ldots \geq s_r$. Then there exist orthonormal ordered bases $B = \{v_1, \ldots, v_n\}$ for V and $E = \{w_1, \ldots, w_m\}$ for W such that

$$[T]_{E,B} = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{m \times n} \quad \textit{with } S := \operatorname{diag}(s_1, \ \ldots, \ s_r) \in \mathbb{C}^{r \times r}.$$

Further, the vectors v_i and w_j satisfy the following:

(1) *Each v_i is an eigenvector of T^*T.*

(2) *$\{v_1, \ldots, v_r\}$ is an orthonormal basis of $R(T^*)$.*

(3) *$\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis of $N(T)$.*

(4) *$\{w_1, \ldots, w_r\}$ is an orthonormal basis of $R(T)$.*

(5) *$\{w_{r+1}, \ldots, w_m\}$ is an orthonormal basis of $N(T^*)$.*

(6) *Each w_j is an eigenvector of TT^*.*

Proof We construct the bases B and E meeting the requirements in (1)-(6). Finally, we show the matrix representation of T.

(1) The eigenvalues of T^*T are $s_1^2, \ldots, s_r^2$, and $n - r$ number of 0s. By the spectral theorem, the self-adjoint linear operator T^*T is diagonalizable. So, there exists an orthonormal ordered basis $B := \{v_1, \ldots, v_n\}$ for V such that

$$T^*T\, v_i = s_i^2 v_i \ \text{ for } i = 1, \ldots, r; \quad T^*T\, v_i = 0\, v_i = 0 \ \text{ for } i = r + 1, \ldots, n.$$

Thus each v_i is an eigenvector of T^*T.

(2) For $1 \leq i \leq r$, $v_i = (s_i^2)^{-1} T^*T v_i$. It shows that $v_i \in R(T^*T)$. Further, $\operatorname{rank}(T^*T) = r$. So, $\{v_1, \ldots, v_r\}$ is an orthonormal basis of $R(T^*T)$. Due to

Theorem 3.23, $R(T^*T) = R(T^*)$. Therefore, $\{v_1, \ldots, v_r\}$ is an orthonormal basis of $R(T^*)$.

(3) For $r + 1 \le i \le n$, $T^*Tv_i = 0$. It follows that $v_i \in N(T^*T)$. Further, $\text{null}(T^*T) = n - r$. Thus $\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis for $N(T^*T)$. By Theorem 3.23, $N(T) = N(T^*T)$. Hence $\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis of $N(T)$.

(4) For $i = 1, \ldots, r$, define $w_i = (s_i)^{-1}Tv_i$. Then

$$\begin{aligned}\langle w_i, w_j\rangle &= (s_i s_j)^{-1}\langle Tv_i, Tv_j\rangle = (s_i s_j)^{-1}\langle v_i, T^*Tv_j\rangle \\ &= (s_i s_j)^{-1}\langle v_i, s_j^2 v_j\rangle = (s_i)^{-1}s_j\langle v_i, v_j\rangle.\end{aligned}$$

Since $\{v_1, \ldots, v_r\}$ is an orthonormal set, $\{w_1, \ldots, w_r\}$ is also an orthonormal set. For $1 \le j \le r$, clearly, $w_j \in R(T)$. Further, $\text{rank}(T) = \text{rank}(T^*T) = r$. Therefore, $\{w_1, \ldots, w_r\}$ is an orthonormal basis of $R(T)$.

(5) Extend the orthonormal set $\{w_1, \ldots, w_r\}$ to an orthonormal basis

$$E = \{w_1, \ldots, w_r, w_{r+1}, \ldots, w_n\}$$

for W. Let $r + 1 \le j \le m$. Then $T^*w_j \in R(T^*)$. Since $\{v_1, \ldots, v_r\}$ is a basis for $R(T^*)$, there exist scalars $\alpha_1, \ldots, \alpha_r$ such that $T^*w_j = \alpha_1 v_1 + \cdots + \alpha_r v_r$. So,

$$TT^*w_j = \alpha_1 Tv_1 + \cdots + \alpha_r Tv_r = \alpha_1 s_1 w_1 + \cdots + a_r s_r w_r.$$

As $\{w_1, \ldots, w_r\}$ is linearly independent, and s_i are positive, it follows that $\alpha_1 = \cdots = \alpha_r = 0$. Then $T^*w_j = 0$. That is, $w_j \in N(T^*)$ for $r + 1 \le j \le m$. But $\text{null}(T^*) = m - \text{rank}(T^*) = m - \text{rank}(T^*T) = m - r$. Hence, $\{w_{r+1}, \ldots, w_m\}$ is an orthonormal basis of $N(T^*)$.

(6) For $1 \le j \le r$, $TT^*w_j = (s_j)^{-1}TT^*Tv_j = (s_j)^{-1}Ts_j^2 v_j = s_j^2 w_j$.
For $r + 1 \le j \le m$, $T^*w_j = 0$ implies that $TT^*w_j = 0\,w_j$.
Therefore, each w_j is an eigenvector of TT^*.

Towards the matrix representation of T, notice that

$$Tv_j = s_j w_j \text{ for } j = 1, \ldots, r; \quad Tv_j = 0 \text{ for } j = r + 1, \ldots, n.$$

Therefore, $[T]_{E,B} = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$ with $S = \text{diag}\,(s_1, \ldots, s_r)$. ∎

The matrix interpretation of SVD may be formulated as in the following.

***Theorem* 6.32 (SVD)**
Let $A \in \mathbb{C}^{m\times n}$ be of rank r. Then there exist unitary matrices $P \in \mathbb{C}^{m\times m}$ and $Q \in \mathbb{C}^{n\times n}$ such that

$$A = P\Sigma Q^*, \quad \Sigma := \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{m\times n}, \quad S := \text{diag}\,(s_1, \ldots, s_r) \in \mathbb{C}^{r\times r},$$

where $s_1 \ge \ldots \ge s_r$ are the positive singular values of A.

In Theorem 6.32, the columns of P are eigenvectors of AA^*, that form an orthonormal basis for $\mathbb{C}^{m\times 1}$; and these are called the *left singular vectors* of A. Further, the columns of Q are eigenvectors of A^*A that form an orthonormal basis for $\mathbb{C}^{n\times 1}$; and these are called the *right singular vectors* of A.

A singular value decomposition depends on the choice of orthonormal bases. For example, just by multiplying ± 1 to an already constructed one we obtain a different orthonormal basis. Thus, SVD is not unique.

Also, it can be shown that when $A \in \mathbb{R}^{m\times n}$, the matrices P and Q can be chosen to have real entries.

In the product $P\Sigma Q^*$, there are 0 blocks in case $m \neq n$. We may thin out certain 0 blocks and obtain the same product. Let $A \in \mathbb{C}^{m\times n}$ have rank r.

Case 1: Suppose $m < n$. We delete the last $n-m$ columns from Σ to obtain Σ_1; and delete the last $n-m$ rows from Q^* to obtain Q_1^*. The deleted columns and rows do not contribute anything to the product. Therefore, we have

$$A = P\Sigma_1 Q_1^*,\ \Sigma_1 = \operatorname{diag}(s_1,\ldots,s_r,0,\ldots,0) \in \mathbb{C}^{m\times m},\ Q_1 = [v_1\ \cdots\ v_m] \in \mathbb{C}^{n\times m}.$$

Here, P is unitary, and Q_1 has orthonormal columns.

Case 2: Suppose $m > n$. We delete the last $m-n$ columns of P to obtain P_2, and delete the last $m-n$ rows of Σ to obtain Σ_2. We see that

$$A = P_2\Sigma_2 Q^*,\ P_2 = [w_1\ \cdots w_n] \in \mathbb{C}^{m\times n},\ \Sigma_2 = \operatorname{diag}(s_1,\ldots,s_r,0,\ldots,0) \in \mathbb{C}^{n\times n}.$$

Here, the columns of P_2 are orthonormal, while Q is unitary.

Case 3: When $m = n$, we keep all of P, Σ, Q as they are, in $A = P\Sigma Q$.

These three forms of SVD are called the **thin SVD** of A.

A further tightening in deleting the 0 blocks in the product $P\Sigma Q^*$ is possible. Suppose $A \in \mathbb{C}^{m\times n}$ has rank r. Observe that the columns $r+1$ onwards in the matrices P and Q in the product PAQ^* produce the 0 blocks. We may then delete the last $m-r$ columns of P, the last $n-r$ columns of Q, and curtail Σ to its first $r\times r$ block. Thus we obtain

$$A = \tilde{P}S\tilde{Q}^*, \quad \text{with} \quad \tilde{P} = [w_1\ \cdots\ w_r], \quad \tilde{Q} = [v_1\ \cdots\ v_r].$$

Here, the matrices $\tilde{P} \in \mathbb{C}^{m\times r}$ and $\tilde{Q} \in \mathbb{C}^{n\times r}$ have orthonormal columns, and $S = \operatorname{diag}(s_1,\ldots,s_r) \in \mathbb{C}^{r\times r}$. This simplified decomposition of the matrix A is called the **tight SVD** of A.

In the tight SVD, the matrices $A \in \mathbb{C}^{m\times n}$, $\tilde{P} \in \mathbb{C}^{m\times r}$, $S \in \mathbb{C}^{r\times r}$ and $\tilde{Q}^* \in \mathbb{C}^{r\times n}$ are all of rank r. Write $B = \tilde{P}S$ and $C = S\tilde{Q}^*$ to obtain

$$A = B\tilde{Q}^* = \tilde{P}C,$$

where $B \in \mathbb{C}^{m\times r}$ and $C \in \mathbb{C}^{r\times n}$ are also of rank r. It shows that each $m\times n$ matrix of rank r can be written as a product of an $m\times r$ matrix with an $r\times n$ matrix each of rank r. This way, the *full rank factorization* of a matrix follows from the tight SVD.

Example 6.33

Obtain SVD, tight SVD, and full rank factorizations of $A = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix}$.

The matrix $A^*A = \begin{bmatrix} 24 & -12 \\ -12 & 6 \end{bmatrix}$ has eigenvalues $\lambda_1 = 30$ and $\lambda_2 = 0$. Thus $s_1 = \sqrt{30}$. It is easy to check that $\mathrm{rank}(A) = 1$ as the first column of A is -2 times the second column. Solving the equations $A^*A(a,b)^{\mathrm{t}} = 30(a,b)^{\mathrm{t}}$, that is,

$$24a - 12b = 30a, \ -12a + 6b = 30b,$$

we obtain a solution as $a = -2, \ b = 1$. So, a unit eigenvector corresponding to the eigenvalue 30 is

$$v_1 = \tfrac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

For the eigenvalue $\lambda_2 = 0$, the equations are

$$24a - 12b = 0, \ -12a + 6b = 0.$$

Thus a unit eigenvector orthogonal to v_1 is

$$v_2 = \tfrac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Then,

$$w_1 = \tfrac{1}{\sqrt{30}} A v_1 = \tfrac{1}{\sqrt{30}} \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} = \tfrac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}.$$

Notice that $\|w_1\| = 1$. We extend $\{w_1\}$ to an orthonormal basis of $\mathbb{C}^{3\times 1}$. It is

$$\left\{ w_1 = \tfrac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}, \quad w_2 := \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad w_3 ;= \tfrac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

Next, we take w_1, w_2, w_3 as the columns of P and v_1, v_2 as the columns of Q to obtain an SVD of A as

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ -2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{30} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}^* .$$

For the tight SVD, $\tilde{P}$ has the r columns as the the first r columns of P, $\tilde{Q}$ has the the r columns as the first r columns of Q, and S is the usual $r \times r$ block consisting of positive singular values of A as the diagonal entries. With $r = \text{rank}(A) = 1$, we thus have the tight SVD as

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix} \begin{bmatrix} \sqrt{30} \end{bmatrix} \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}^* .$$

In the tight SVD, using associativity of matrix product, we get the full rank factorizations as

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} -\sqrt{5} \\ \sqrt{5} \\ -2\sqrt{5} \end{bmatrix} \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}^* = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix} \begin{bmatrix} -2/\sqrt{6} \\ \sqrt{6} \end{bmatrix}^* .$$

You should check that the columns of Q are eigenvectors of AA^*. □

Observe that when we write an $m \times n$ matrix A of rank r in its SVD form $A = P\Sigma Q^*$, the columns of P are the eigenvectors of the matrix AA^* associated with the eigenvalues $s_1^2, \ldots, s_r^2, 0, \ldots, 0$. Similarly, the columns of Q are the eigenvectors of the matrix A^*A associated with the same eigenvalues. In the former case, there are $m - r$ zero eigenvalues and in the latter case, they are $n - r$ in number. Writing the ith column of P as w_i and the jth column of Q as v_j, SVD amounts to writing A as

$$A = P\Sigma Q^* = s_1 w_1 v_1^* + \cdots + s_r w_r v_r^* .$$

Each matrix $w_k v_k^*$ here is of rank 1. This means that if we know the r positive singular values of A and we know their corresponding left and right singular vectors, we know A completely. This is particularly useful when A is a very large matrix of low rank. No wonder, SVD is used in image processing, various compression algorithms, and in principal components analysis. Next to the theory of linear equations, SVD is the most important tool for applications. We will see another application of SVD in representing a matrix in a very useful and elegant manner.

Exercises for § 6.6

1. Let $A \in \mathbb{C}^{m \times n}$. Show that the positive singular values of A^* are precisely the positive singular values of A.
2. Prove that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of an $n \times n$ hermitian matrix, then its singular values are $|\lambda_1|, \ldots, |\lambda_n|$.

3. Compute the singular value decomposition of the following matrices:
$$\begin{bmatrix} 2 & -2 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 2 \\ -2 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 2 \\ 2 & 0 & -5 \\ 3 & 0 & 0 \end{bmatrix}.$$
4. Show that the matrices $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$ are similar but they have different singular values.
5. Show that if s is a singular values of a matrix A, then there exists a nonzero vector x such that $\|Ax\| = s\|x\|$.
6. Show that a matrix $A \in \mathbb{C}^{m\times n}$ is of rank 1 iff $A = uv$ for some vectors $u \in \mathbb{C}^{m\times 1}$ and $v \in \mathbb{C}^{1\times n}$.
7. Show that a scalar $\lambda > 0$ is an eigenvalue of T^*T iff it is an eigenvalue of TT^*, without using SVD.
8. In an SVD of a linear operator $T : V \to W$, show that the orthonormal bases $\{v_1, \ldots, v_n\}$ for V and $\{w_1, \ldots, w_m\}$ for W can be chosen in such a way that for $1 \le i \le r$, $w_i = (s_i)^{-1}Av_i$ and $v_i = (s_i)^{-1}A^*w_i$.
9. Let $A \in \mathbb{C}^{m\times n}$ have positive singular values $s_1 \ge \cdots \ge s_r$. Let S be the unit circle in $\mathbb{C}^{n\times 1}$. That is, $S = \{x \in \mathbb{C}^{n\times 1} : \ \|x\| = 1\}$. Show that $s_1 = \max\{\|Ax\| : x \in S\}$ and $s_r = \min\{\|Ax\| : x \in S\}$.
10. Let $A \in \mathbb{C}^{m\times n}$ have the positive singular values $s_1, \ldots, s_r$.
   Let $A = P \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} Q^*$ be an SVD of A, where $S = \operatorname{diag}(s_1, \ldots, s_r)$.
   Define $A^\dagger = P \begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^*$. Prove the following:
   (a) For any $b \in \mathbb{C}^{m\times 1}$, $A^\dagger b$ is a least squares solution of $Ax = b$.
   (b) $A^\dagger$ satisfies the following:
   $$(AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A, \quad AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger.$$
   (c) If $A^\dagger$ is any matrix in $\mathbb{C}^{n\times m}$ that satisfies the equalities in (b), then it is unique.

   The matrix $A^\dagger$ is called the *generalized inverse* of A.

6.7 Polar decomposition

Linear operators on an inner product space behave like complex numbers, in many respects. Recall that a complex number is written as $z = re^{i\theta}$, where r

is a non-negative real number which may be thought of as a stretching factor, and $e^{i\theta}$ is a rotation. We aim towards representing a linear operator as a product of a *positive semidefinite* matrix and a unitary matrix.

Let T be a linear operator on a finite dimensional inner product space V. We say that T is **positive semi-definite** iff T is self-adjoint, and $\langle Tx, x\rangle \geq 0$ for each $x \in V$. T is called **positive definite** iff T is self adjoint, and $\langle Tx, x\rangle > 0$ for each nonzero $x \in V$.

Similarly, a matrix $A \in \mathbb{C}^{m\times}$ is called *positive semi-definite* iff A is hermitian and $x^*Ax \geq 0$ for each $x \in \mathbb{C}^{n\times 1}$; and A is called *positive definite* iff A is hermitian and $x^*Ax > 0$ for each nonzero $x \in \mathbb{C}^{n\times 1}$.

Examples of positive semi-definite linear operators are abundant. For, if T is any linear operator on a finite dimensional ips, then both T^*T and TT^* are positive semi-definite.

***Theorem* 6.34 (Polar Decomposition)**
Let $T : V \to W$ be a linear operator, where V and W are inner product spaces of dimensions n and m, respectively. Then there exist positive semi-definite linear operators P on W, Q on V, and a linear operator $U : V \to W$ such that

$$T = PU = UQ,$$

where $P^2 = TT^*$, $Q^2 = T^*T$, $UU^* = I$ for $m \leq n$, $U^*U = I$ for $m \geq n$, and U is unitary for $m = n$.

Proof Let $\operatorname{rank}(T) = r$. Suppose $s_1 \geq \cdots \geq s_r$ are the positive singular values of T. For uniformity in notation, let $s_i = 0$ for $i > r$. Due to the SVD of T, there exist orthonormal bases $\{v_1, \ldots, v_n\}$ for V and $\{w_1, \ldots, w_m\}$ for W such that

$$TT^*w_i = s_i^2 w_i \text{ for } 1 \leq i \leq m, \quad T^*Tv_i = s_i^2 v_i \text{ for } 1 \leq i \leq n,$$
$$Tv_i = s_i w_i \text{ for } 1 \leq i \leq r, \quad Tv_i = 0 \text{ for } r+1 \leq i \leq n.$$

Let $v \in V$ and $w \in W$. By Fourier expansion,

$$v = \sum_{i=1}^{n} \langle v, v_i\rangle v_i, \quad w = \sum_{j=1}^{m} \langle w, w_j\rangle w_j.$$

Then the above equalities imply that

$$TT^*w = \sum_{i=1}^{r} s_i^2 \langle w, w_i\rangle w_i, \quad T^*T = \sum_{i=1}^{r} s_i^2 \langle v, v_i\rangle v_i, \quad Tv = \sum_{i=1}^{r} s_i \langle v, v_i\rangle w_i.$$

Let $\ell = \min\{m,n\} \geq r$. Define linear operators $P : W \to W$, $Q : V \to V$, and $U : V \to W$ by

$$Pw = \sum_{i=1}^{r} s_i \langle w, w_i\rangle w_i,\ Qv = \sum_{i=1}^{r} s_i \langle v, v_i\rangle v_i,\ Uv = \sum_{i=1}^{\ell} \langle v, v_i\rangle w_i \ \text{ for } v \in V,\ w \in W.$$

Notice that for $1 \leq i \leq r$, $P(w_i) = s_i w_i$; for $i > r$, $P(w_i) = 0$, $Q(v_i) = 0$; and for $i > \ell$, $U(v_i) = 0$. Due to the formula for the adjoint in (3.3), we have

$$\begin{aligned}
P^* w &= \sum_{i=1}^{m} \langle w, Pw_i\rangle w_i = \sum_{i=1}^{r} \langle w, s_i w_i\rangle w_i = \sum_{i=1}^{r} s_i \langle w, w_i\rangle w_i = Pw.\\
\langle Pw, w\rangle &= \sum_{i=1}^{r} s_i \langle w, w_i\rangle\langle w_i, w\rangle = \sum_{i=1}^{r} s_i |\langle w, w_i\rangle|^2 = \sum_{i=1}^{m} s_i |\langle w, w_i\rangle|^2 \geq 0.\\
P^2 w &= P\Big(\sum_{i=1}^{r} s_i \langle w, w_i\rangle w_i\Big) = \sum_{i=1}^{r} s_i \langle w, w_i\rangle P w_i = \sum_{i=1}^{r} s_i^2 \langle w, w_i\rangle w_i = TT^* w.\\
PUv &= P\Big(\sum_{i=1}^{\ell} \langle v, v_i\rangle w_i\Big) = \sum_{i=1}^{\ell} \langle v, v_i\rangle P w_i = \sum_{i=1}^{r} s_i \langle v, v_i\rangle w_i = Tv.
\end{aligned}$$

Hence P is positive semi-definite, $P^2 = TT^*$, and $PU = T$. Similarly, it follows that Q is positive semi-definite, $Q^2 = T^*T$, and $UQ = T$. It remains to verify the conditions on U in different cases.

Case 1: Let $m \leq n$. Then $\ell = m$. Let $w \in W$. For $1 \leq i \leq m$, using (3.3), we have

$$U^* w = \sum_{j=1}^{n} \langle w, Uv_j\rangle v_j = \sum_{j=1}^{m} \langle w, w_j\rangle v_j, \quad Uv_i = \sum_{j=1}^{m} \langle v_i, v_j\rangle w_j = w_i.$$

$$UU^* w = U\Big(\sum_{j=1}^{m} \langle w, w_j\rangle v_j\Big) = \sum_{j=1}^{m} \langle w, w_j\rangle U v_j = \sum_{j=1}^{m} \langle w, w_j\rangle w_j = w.$$

That is, $UU^* = I$.

Case 2: Let $m \geq n$. Then $\ell = n$. Let $v \in V$. For $1 \leq i \leq n$, using (3.3) again, we obtain

$$U^* w_i = \sum_{j=1}^{n} \langle w_i, Uv_j\rangle v_j = \sum_{j=1}^{n} \langle w_i, w_j\rangle v_j = v_i.$$

$$U^* U v = U^*\Big(\sum_{i=1}^{n} \langle v, v_i\rangle w_i\Big) = \sum_{i=1}^{n} \langle v, v_i\rangle U^* w_i = \sum_{i=1}^{n} \langle v, v_i\rangle v_i = v.$$

That is, $U^*U = I$.

When $m = n$, both (1) and (2) are true. Therefore, U is unitary. ∎

The matrix interpretation of Theorem 6.34 is straight forward.

***Theorem* 6.35 (Polar Decomposition)**
Each matrix $A \in \mathbb{C}^{m\times n}$ can be written as $A = PU = UQ$, where $P \in \mathbb{C}^{m\times m}$ and $Q \in \mathbb{C}^{n\times n}$ are positive semi-definite, $P^2 = AA^*$, $Q^2 = A^*A$, $UU^* = I$ for $m \le n$, $U^*U = I$ for $m \ge n$, and U is unitary for $m = n$.

The construction of polar decomposition of matrices from SVD may be summarized as follows:

If $A \in \mathbb{C}^{m\times n}$ has SVD as $A = BDE^*$, then $A = PU = UQ$, where

$$\begin{array}{llll}
m = n: & U = BE^*, & P = BDB^*, & Q = EDE^*. \\
m < n: & U = BE_1^*, & P = BD_1B^*, & Q = E_1D_1E_1^*. \\
m > n: & U = B_1E^*, & P = B_1D_2B_1^*, & Q = ED_2E^*.
\end{array}$$

Here, E_1 is constructed from E by taking its first m columns; D_1 is constructed from D by taking its first m columns; B_1 is constructed from B by taking its first n columns; and D_2 is constructed from D by taking its first n rows.

Example 6.36

Consider the matrix $A = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix}$ of Example 6.33. We had obtained its SVD as $A = BDE^*$, where

$$B = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ -2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}, \quad D = \begin{bmatrix} \sqrt{30} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}.$$

We construct the matrices B_1 by taking first two columns of B, and D_2 by taking first two rows of D, as in the following:

$$B_1 = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{6} & 1/\sqrt{2} \\ -2/\sqrt{6} & 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} \sqrt{30} & 0 \\ 0 & 0 \end{bmatrix}.$$

Then

$$U = B_1E^* = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 & 1 \\ 1 & \sqrt{3} \\ -2 & 0 \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \frac{1}{\sqrt{30}}\begin{bmatrix} 2+\sqrt{3} & -1+2\sqrt{3} \\ -2+\sqrt{3} & 1+2\sqrt{3} \\ 4 & -2 \end{bmatrix},$$

$$P = BDB_1^* = \sqrt{5}\begin{bmatrix} -1 & 0 \\ 1 & 0 \\ -2 & 0 \end{bmatrix}\frac{1}{\sqrt{6}}\begin{bmatrix} -1 & 1 & -2 \\ \sqrt{3} & \sqrt{3} & 0 \end{bmatrix} = \frac{\sqrt{5}}{\sqrt{6}}\begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{bmatrix},$$

$$Q = ED_2E^* = \sqrt{6}\begin{bmatrix} -2 & 0 \\ 1 & 0 \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix} = \frac{\sqrt{6}}{\sqrt{5}}\begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}.$$

As expected we find that

$$PU = \frac{\sqrt{5}}{\sqrt{6}}\begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{bmatrix}\frac{1}{\sqrt{30}}\begin{bmatrix} 2+\sqrt{3} & -1+2\sqrt{3} \\ -2+\sqrt{3} & 1+2\sqrt{3} \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} = A.$$

$$UQ = \frac{1}{\sqrt{30}}\begin{bmatrix} 2+\sqrt{3} & -1+2\sqrt{3} \\ -2+\sqrt{3} & 1+2\sqrt{3} \\ 4 & -2 \end{bmatrix}\frac{\sqrt{6}}{\sqrt{5}}\begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} = A. \qquad \square$$

The polar decomposition can be computed without using SVD. Notice that in $A = PU = UQ$, the matrices P and Q satisfy $P^2 = AA^*$ and $Q^2 = A^*A$. If $A \in \mathbb{C}^{m\times n}$, then $AA^* \in \mathbb{C}^{m\times m}$ and $A^*A \in \mathbb{C}^{n\times n}$ are hermitian matrices with eigenvalues as $s_1^2, \ldots, s_r^2, 0, \ldots, 0$. Due to the Spectral theorem, there exists a matrix C such that $AA^* = C^* \operatorname{diag}(s_1^2, \ldots, s_r^2, 0, \ldots, 0)\, C$. Then, we construct

$$P = C \operatorname{diag}(s_1, \ldots, s_r, 0, \ldots, 0)\, C^*.$$

Here, the matrix C consists of orthonormal eigenvectors of AA^* corresponding to the eigenvalues $s_1^2, \ldots, s_r^2, 0, \ldots, 0$. Similarly, we construct

$$Q = F^* \operatorname{diag}(s_1, \ldots, s_r, 0, \ldots, 0)\, F,$$

where F consists of orthonormal eigenvectors of A^*A corresponding to the eigenvalues $s_1^2, \ldots, s_r^2, 0, \ldots, 0$. Finally, the Us can be computed by solving the linear systems $A = PU$ and $A = UQ$.

In this method, the two instances of U in the equalities $A = PU = UQ$ may differ since they depend on the choices of orthonormal eigenvectors of AA^* and A^*A. In case, A is invertible, you would end up at the same U.

Exercises for § 6.7

1. Determine the polar decompositions of the matrix A of Example 6.36 by diagonalizing AA^* and A^*A as mentioned in the text.

2. Let $A \in \mathbb{C}^{m\times n}$ with $m < n$. Prove that there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and a matrix $P \in \mathbb{C}^{m\times n}$ such that $A = PU$.
3. Give a direct proof of Theorem 6.34(3) analogous to that of (2). You may have to partition B instead of E in the SVD: $A = BDE$.
4. Prove Theorem 6.34(2-3) by using thin SVD.
5. Derive singular value decomposition from the polar decomposition.
6. Let T be a self-adjoint liner operator on a finite dimensional ips V. Let $\alpha, \beta \in \mathbb{R}$ be such that $\alpha^2 < 4\beta$. Show that $T^2 + \alpha T + \beta I$ is positive semi-definite.
7. Show that if T is a linear operator on a finite dimensional complex ips V, then the condition $\langle Tx, x\rangle \geq 0$ for each $x \in V$ implies that T is self-adjoint. [Hint: First show that $\langle Tx, x\rangle = 0$ for each $x \in V$ implies that $T = 0$.]
8. Let V be an ips. Show that if for all $x, y, z \in V$, $\langle x, y\rangle = \langle z, y\rangle$, then $x = z$.
9. *Polarization identity*: Let V be a real ips. Show that for all $x, y \in V$, $\langle x, y\rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right)$.
10. *Polarization identity*: Let V be a complex ips. Show that for all $x, y \in V$, $\langle x, y\rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2 + i\|u + iv\|^2 - i\|u - iv\|^2\right)$.
11. Let S and T be linear operators on an ips V. Show that if for all $x \in V$, $\langle Sx, x\rangle = \langle Tx, x\rangle$, then $S = T$.
12. Let U be a linear operator on a finite dimensional ips V. Use the previous exercise to show that if $\|Ux\| = \|x\|$ for all $x \in V$, then $U^*U = I$.
13. Let S and T be linear operators on a finite dimensional ips V. We say that S is a *square root* of T iff $S^2 = T$. Prove that each positive semi-definite linear operator on a finite dimensional ips has a unique positive semi-definite square root.
14. Let T be a linear operator on a finite dimensional ips V. Show that the following are equivalent:
    (a) T is positive semi-definite.
    (b) T is self-adjoint and each eigenvalue of T is non-negative.
    (c) T has a positive semi-definite square root.
    (d) T has a self-adjoint square root.
    (e) There exists a linear operator S on V such that $T = S^*S$.

Bibliography

[1] S. Axler, *Linear Algebra Done Right*, Springer Int. ed., Indian Reprint, 2013.

[2] A. F. Filippov, A short proof of the theorem on reduction of a matrix to Jordan form, *Vestnik Moskov Univ. Ser. I Mat. Meh.* 26 (1971), no. 2, 18-19, MR 43, No. 4839.

[3] P. R. Halmos, *Finite Dimensional Vector Spaces*, Springer Int. Ed., Indian Reprint, 2013.

[4] K. Hoffman, R. Kunze, *Linear Algebra, 2nd Ed.*, Prentice Hall Inc., 1971.

[5] R. Horn, C. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[6] K. Jänich, *Linear Algebra*, Undergraduate Texts in Math., Springer, 1994.

[7] S. J. Leon, *Linear Algebra with Applications, 9th Ed.*, Pearson, 2014.

[8] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM Pub., 2000.

[9] M. T. Nair, A. Singh, *Linear Algebra,* Springer India, 2018.

[10] A. Singh, *Introduction to Matrix Theory,* Ane Books, 2017.

Index

$N(T)$, 56
$R(T)$, 56
$\mathbb{F}^{m\times n}$, 73, 82
$\mathcal{L}(V,W)$, 82
Sol (A,b), 103

adjoint, 67, 86
adjugate, 91
Algebraic multiplicity, 130
angle between vectors, 39
augmented matrix, 103

basic variables, 114
basis, 23
Bessel's inequality, 42
best approximation, 46

Cauchy Schwartz Inequality, 38
Cayley-Hamilton theorem, 131
change of basis, 96
characteristic polynomial, 128
column vector, 6
complex eigenvalue, 130
complex vector space, 7
conjugate transpose, 85
consistent system, 103
coordinate vector, 71
coordinate vector map, 71
Cramer's rule, 104

determinant, 91
diagonal, 85
diagonalizable, 141
diagonalized by, 143
diagonal entries, 85
dimension, 28

eigenvalue, 125
eigenvector, 125
elementary column operation, 109
elementary matrix, 106
elementary operation, 33
elementary row operation, 107
equivalent matrices, 100
Euclidean space, 36

finite dimensional, 28
Fourier expansion, 41
free variables, 114
Full rank factorization, 101
full rank factorization, 164
fundamental subspaces, 70

Gauss-Jordan elimination, 118
Gaussian elimination, 114
generalized eigenvector, 158
geometric multiplicity, 144

homogeneous system, 103

identity operator, 50
infinite dimensional, 28
inner product, 36
inner product space, 36
inverse of, 81
invertible, 81
ips, 36
isometric, 69
isomorphic, 61
isomorphism, 61

Jordan block, 155
Jordan form, 155
Jordan string, 147

least squares, 120
length of a Jordan string, 147
linearly dependent, 18
linearly independent, 18
linear combination, 14
linear equations, 103
linear functional, 50
linear map, 50
linear operator, 50
linear transformation, 50

matrix, 73
  diagonal, 85
  hermitian, 88
  identity, 81
  lower triangular, 85
  normal, 88
  of linear map, 73
  orthogonal, 88
  real symmetric, 88
  scalar, 85
  skew hermitian, 88
  skew symmetric, 88
  standard basis, 82
  symmetric, 88
  triangular, 85
  unitary, 88
  upper triangular, 85
  zero, 80
maximal, 24
minimal, 24

norm, 37
normal, 69
nullity, 57
null space, 56

orthogonal, 69
orthogonally diagonalizable, 142
orthogonal basis, 41
orthogonal set, 40
orthogonal vectors, 39
orthonormal basis, 41
orthonormal set, 40

parallelogram law, 38
Parseval's identity, 41
pivot, 107
pivotal column, 107
pivotal row, 108
polarization identity, 171
polar decomposition, 167, 169
positive definite, 167
positive semi-definite, 167
powers of matrices, 78
Product of matrices, 77
Pythagoras' theorem, 40

QR-factorization, 121

range space, 56
rank, 57
Rank-nullity theorem, 58
Rank factorization, 101
Rank theorem, 100
real vector space, 7
Reverse triangle inequality, 38
Riesz representation, 70
row echelon form, 107
row reduced echelon form, 110
row vector, 6
RREF, 110
RREF reduction, 110

scalar, 7
Schur triangularization, 137
self-adjoint, 69
singular values, 160
solution set, 103
span, 14
spanned by, 16
spanning set, 16
Square matrix, 78
square root, 171

standard basis, 24
standard inner product, 37
subspace, 11
sum of subsets, 16
SVD, 161

Theorem
    Basis extension, 29
    Bessel's inequality, 42
    Cauchy Schwartz, 38
    Cayley-Hamilton, 131
    Fourier expansion, 41
    Full rank factorization, 101
    Jordan form, 155
    Jordan strings, 149
    Parallelogram law, 38
    Parseval's identity, 41
    Polar decomposition, 167, 169
    Rank-nullity, 58
    Rank factorization, 101
    Rank theorem, 100
    Reverse triangle ineq., 38
    Schur triangularization, 135
    SVD, 161, 162
    Triangle inequality, 38
thin SVD, 163
tight SVD, 163
trace, 91
transpose, 85
triangle inequality, 38

unitarily diagonalizable, 141
unitary, 69
unitary space, 36

vector, 7
vector space, 6

zero operator, 50