A CONCISE TREATMENT of

LINEAR ALGEBRA

with DIFFERENTIAL EQUATIONS

M. Erol Sezer
Bilkent University, Ankara

Preface

Linear algebra is not only a powerful mathematical theory; it is also a useful computational tool supported by abundant computer software. It provides the theoretical foundations of the concept of linearity as well as efficient methods for solving problems formulated within this framework. For these reasons it has found applications in many diverse fields outside mathematics, ranging over various engineering disciplines, computer science, economics, statistics, and more.

This book is intended to introduce the theory and applications of linear algebra to engineering students, especially those in electrical engineering. It is motivated by the observation that, although the concept of linearity is introduced and widely used in many basic and core courses of a typical engineering curriculum, only a few graduates gain a full understanding of the fundamental role it plays in formulation and solution of many engineering problems. Only in a high-level graduate course does a student learn that matrices, linear differential operators and transfer functions, all being linear transformations between suitably constructed linear spaces, are essentially the same. Only then can he/she get a full grasp of the meaning of the impulse response of a dynamical system and the Fourier transform of a signal, and relate the harmonic content of a periodic signal and the modes of an electromagnetic field or a vibrating structure to the coordinates of a vector in space.

The main objective of this book is to provide students of electrical engineering a firm understanding of the concept of linearity at an early stage of their program. It is built upon a rigorous treatment of vector spaces and linear transformations, which are motivated by linear systems of algebraic equations and first and second order linear differential equations. A second objective is to provide the students with a knowledge of theoretical and operational aspects of matrix algebra that will be sufficient for their undergraduate and early graduate curricula. Finally, the third objective is to introduce linear differential equations as a useful application of linear algebra while providing the students with elementary material that they will need in a concurrent study of dynamical circuits and systems.

This book is primarily a text on linear algebra supplemented with linear differential equations. Although merging linear differential equations in a text on linear algebra is observed to be pedagogically useful in connecting different concepts, the book is not intended to serve as a text on differential equations. This is evident from its contents, as many important topics covered in an introductory course on differential equations, such as series solutions and introductory partial differential equations are left out; other topics such as first order nonlinear differential equations, numerical solution techniques, and Laplace transforms are mentioned only briefly.

The main difficulty I faced in preparing the manuscript was to make a choice for the content and presentation between conflicting alternatives: Amount of material to be included versus suitability for a one-semester or two-quarter course, emphasis on theory versus operational aspects, mathematical formalism versus explanation of implications of the results, and finally examples versus exercises.

- I preferred a self-contained text to one tailored for a specific course. I included most of the material that engineering students might need throughout their undergraduate curriculum, so that they can use it as a reference book even if not everything in the text is taught in a specific course. Another reason for this choice was to make the text suitable for self-study. (However, I offer some suggestions about what might be included in a one- or two-quarter/semester course.) To help the instructor/reader in selecting the material to be covered in a quick treatment, I indicated more advanced and/or abstract topics (including examples) with an asterisk (*). However, this does not mean that all the marked sections or examples can be omitted without destroying the completeness of the book.

- Although I gave priority to a logical development of the theory, I explained, whenever possible, what the theory is good for. For example, the basic definitions and properties of matrices are followed by systems of linear equations to help the reader appreciate the convenience and power inherent in the matrix formulation of linear equations. I encouraged the student to use MATLAB to solve matrix-related problems starting immediately in the first chapter without substituting it for the underlying theory. I avoided integrating MATLAB with the text, leaving that to the instructor. However, I provided sufficient exercises to demonstrate MATLAB's power and, to a lesser degree, its limitations.

- Although I avoided a Definition-Theorem-Proof structure to make the reading easier and less dry, I did not state any result without a proof except when:

  - I thought the reader could provide a fairly straightforward proof independently by simply imitating the steps of similar results for which a formal proof is given, or

  - In few instances, the proof of the stated result is beyond the scope of the book, but its implications are too significant to omit. In such cases I tried to provide insight into the meaning and the implications of the stated result.

  I endeavored to develop the material from the concrete to the abstract, in most instances explaining the need to introduce a new definition or result. Rather than leaving the often difficult task of establishing a connection between a new result and a previous one to the reader, I attempted to explain the logic behind the development, often by referring to previous examples.

- I provided sufficient examples to explain the theoretical development. However, I avoided multiple similar examples except when they emphasize different aspects of the same concept. Instead of filling the pages with redundant examples, I preferred to include many exercises with hints to formulation and/or solution. Again, I avoided similar exercises that differ only in the numerical values involved.

The first three chapters of the book contain the most essential material of a first course on linear algebra, where the concepts of vector spaces and linear transformations are built upon linear algebraic and differential equations.

Chapter 1 is a self-contained treatment of simple matrix algebra, where basic definitions and operations on matrices are introduced and systems of linear algebraic equations are studied. Properties of matrix addition and scalar multiplication are stated in a manner consistent with the corresponding properties of vector addition and scalar multiplication to prepare the student for the more abstract concepts to follow. The Gaussian elimination is introduced not only as a systematic approach to solving linear equations, but also as a theoretical tool leading to the concepts of rank and particular and complementary solutions, which in turn pave the road to a more abstract treatment of linear equations in terms of the kernel and image of the associated linear transformation. Through simple examples and exercise problems the student is urged to use MATLAB to check the results of their hand calculations and to digest the idea that a matrix is a data unit (like a number) on which they can perform some operations.

In Chapter 2 first and second order linear differential equations are studied with emphasis on the constant coefficient case. The three objectives of the chapter are (i) to provide the students with solution techniques for simple differential equations that they can immediately start utilizing in concurrent courses on circuits or dynamical systems, (ii) to further prepare the student for linear transformations by repeating the concepts of particular and complementary solutions in a different context and by introducing linear differential operators, and (iii) to introduce the basics of numerical solution techniques so that the student can begin to use MATLAB or other software, and at the same time, to give an idea of linear difference equations, which involve yet another type of linear transformation.

Chapter 3 contains an abstract treatment of vector spaces and linear transformations based on the ideas introduced in the preceding two chapters. The concept of a vector space is extended beyond the familiar n-spaces with the aim of unifying linear algebraic and differential equations under a common framework. By interpreting an n-vector as a function defined over a finite domain, the student is prepared for function spaces. The concept of basis is given special emphasis to establish the link between abstract vectors and the more familiar n-vectors, as well as between abstract linear transformations and matrices. Discrete Fourier series are introduced as an example of representation of vectors of a finite-dimensional vector space with respect to a fixed basis. This chapter also contains some more advanced topics such as inverse transformations, direct sum decompositions, and projections.

Chapter 4 introduces rank and inverse of matrices. Rank is defined in terms of the row and column spaces. Left, right, two-sided and generalized inverses are based on elementary matrices without reference to determinants. Concepts of equivalence and similarity are related to change of basis. The LU decomposition is studied as a natural and useful application of elementary operations. Determinants are considered mainly for traditional reasons to mention the role they play in the solution of linear equations with square coefficient matrices.

Chapter 5 deals with the eigenvalues, eigenvectors and diagonalization of square matrices from a geometric perspective. The diagonalization problem is related to the decomposition of the underlying vector space into invariant subspaces to motivate the much more advanced Jordan form. The chapter concludes with a treatment of functions of a matrix, the main objective being to define the exponential matrix function that will be needed in the study of systems of linear differential equations.

In Chapter 6 we return to linear differential equations. As opposed to the traditional approach of treating nth order linear differential equations and systems of first order linear differential equations separately, and in that order, systems of differential equations are studied first and the results developed in that context are used to derive the corresponding results for nth order differential equations. This is consistent with the matrix theoretic approach of the text to the treatment of linear problems, which relates the abstract concepts of bases, direct sum decomposition of vector spaces, and the Jordan form to solutions of a homogeneous system of linear differential equations and their modal decomposition. The method of undetermined coefficients is included as a practical way of solving linear differential equations with special forcing functions that are common in engineering applications.

Chapter 7 treats normed and inner product spaces with emphasis on the concepts of orthogonality and orthogonal projections, where the Gram-Schmidt orthogonalization process, the least-squares problem and the Fourier series are formulated as applications of the projection theorem.

Chapter 8 deals with unitary and Hermitian matrices for the purpose of presenting such useful applications as quadratic forms and the singular-value decomposition, which is related to the least-squares problem and matrix norms.

I was able to cover most of the material in a 56-class-hour one-semester course I taught to a class of select students at the Electrical Engineering Department of Bilkent University. However, an average class of second or third year students would need two quarters (about 60 class hours) to cover the essential material. For such a course, Sections 2.6, 3.4.3, 3.6 and 4.5 may be omitted, and the material in Sections 1.5, 3.4.2, 5.4, 6.2 and Chapter 8 may be discussed briefly. For a one-semester course I suggest omitting this material completely.

The text was developed over some years of my experience with teaching linear algebra at the Middle East Technical University and Bilkent University. My long time friend and colleague Özay Oral and I prepared some lecture notes to meet the demand for a text for a combined course on Linear Algebra and Differential Equations. Although the present text is completely different from those lecture notes, both in its approach and in contents, it would not have come to fruition without those initial efforts. I am indebted to Özay for his motivation and encouragement that led first to the lecture notes and eventually to the present version of the text. Thanks are also due to my colleagues at the Middle East Technical and Bilkent Universities for their suggestions and constructive criticisms.

M. Erol Sezer
sezer@ee.bilkent.edu.tr

Contents

Chapter 1
Matrices and Systems of Linear Equations

1.1 Basic Matrix Definitions

An $m \times n$ (read "*m-by-n*") **matrix** is an array of mn elements of a field F arranged in m rows and n columns as

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

A matrix with m rows and n columns is said to be of **order** (size, dimension) $m \times n$. We denote matrices with uppercase letters and their elements with corresponding lowercase letters, and use the notation $A = [\,a_{ij}\,]_{m\times n}$ to describe an $m \times n$ matrix where a_{ij} typifies the element in the ith row and the jth column. When the order of A need not be specified we simply write $A = [\,a_{ij}\,]$. The set of all $m \times n$ matrices with elements from F is denoted by $\mathsf{F}^{m\times n}$. Throughout the book we will assume that the underlying field F is either R (in which case A is a real matrix) or C (in which case A is a complex matrix).[1]

A $1 \times n$ matrix is called a **row matrix** or a **row vector**, and an $m \times 1$ matrix is called a **column matrix** or a **column vector**.[2] We denote row and column vectors with boldface lowercase letters. Thus

$$\mathbf{x} = [\,x_1 \;\; x_2 \;\; \cdots \;\; x_n\,]$$

is a $1 \times n$ row vector, and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

[1] Definition of a field and a brief review of complex numbers are given in Appendix A. Since the field of real numbers is a subfield of complex numbers, every real matrix can also be viewed as a complex matrix.

[2] The use of the term "vector" for a column or a row matrix is justified in Chapter 3.

is an $m \times 1$ column vector, which we also denote as

$$\mathbf{y} = \mathrm{col}\,[\,y_1, y_2, \ldots, y_m\,]$$

to save space. Note that a single column or row index suffices to denote elements of a row or a column vector.

An $n \times n$ matrix is called a **square matrix** of order n. The sum of the diagonal elements $a_{11}, \ldots, a_{nn}$ of a square matrix $A = [\,a_{ij}\,]_{n\times n}$ is called the **trace** of A, denoted tr (A):

$$\mathrm{tr}\,(A) = \sum_{i=1}^{n} a_{ii}$$

A square matrix $D = [\,d_{ij}\,]_{n\times n}$ in which $d_{ij} = 0$ for all $i \neq j$ is called a **diagonal** matrix, denoted

$$D = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix} = \mathrm{diag}\,[\,d_{11}, d_{22}, \ldots, d_{nn}\,]$$

A square matrix $L = [\,l_{ij}\,]_{n\times n}$ in which $l_{ij} = 0$ whenever $i < j$ is called a **lower triangular** matrix for the obvious reason that all the elements located above its diagonal are zero. Similarly, a square matrix $U = [\,u_{ij}\,]_{n\times n}$ with $u_{ij} = 0$ whenever $i > j$ is called an **upper triangular** matrix.

Example 1.1

The array

$$A = \begin{bmatrix} 0 & \sqrt{2} & -1 \\ 3 & e & \ln 5 \end{bmatrix}$$

is a 2×3 real matrix with elements $a_{11} = 0, a_{12} = \sqrt{2}, \ldots, a_{23} = \ln 5$.

The array

$$B = \begin{bmatrix} 1+2i & -3 & -1+i \\ 0 & -3i & 5 \\ -1 & -2 & 3+2i \end{bmatrix}$$

is a complex square matrix of order 3 with

$$\mathrm{tr}\,(B) = (1+2i) + (-3i) + (3+2i) = 4+i$$

The matrices

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 0 & 4 & 5 \end{bmatrix}, \quad D = \begin{bmatrix} 3+i & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are lower triangular and diagonal, respectively.

1.2 Basic Matrix Operations

1.2.1 Matrix Addition and Scalar Multiplication

Let $A = [\, a_{ij} \,]$ and $B = [\, b_{ij} \,]$ be two matrices in $\mathsf{F}^{m \times n}$.

A and B are said to be **equal**, denoted $A = B$, if $a_{ij} = b_{ij}$ for all (i, j).[3] Note that it would be meaningless to talk about equality of two matrices unless they are of the same size and their elements are comparable, that is, they are also of the same type.

If $a_{ij} = 0$ for all (i, j), then A is called an $m \times n$ **zero matrix** (null matrix), denoted $O_{m \times n}$, or simply O if the order is known. That is, $O_{m \times n} = [\, 0 \,]_{m \times n}$.

The **sum** of A and B, denoted $A + B$, is defined in terms of their elements as

$$A + B = [\, a_{ij} + b_{ij} \,]_{m \times n}$$

That is, the (i, j)th element of $A + B$ is the sum of the corresponding elements of A and B. The subtraction operation is defined in terms of addition as

$$A - B = A + (-B)$$

where

$$-B = [\, -b_{ij} \,]$$

Note that, like equality, addition and subtraction operations are defined only for matrices belonging to the same class, and that if $A \in \mathsf{F}^{m \times n}$ and $B \in \mathsf{F}^{m \times n}$, then $A + B \in \mathsf{F}^{m \times n}$, $-B \in \mathsf{F}^{m \times n}$, and therefore, $A - B \in \mathsf{F}^{m \times n}$.

Any element of the field F over which the matrices of concern are defined is called a **scalar**. The **scalar product** of a matrix A with a scalar c, denoted cA, is also defined element-by-element as

$$cA = [\, ca_{ij} \,]_{m \times n}$$

Thus the (i, j)th element of cA is c times the corresponding element of A. It follows from the definition that $cA \in \mathsf{F}^{m \times n}$. Clearly,

$$(-1)A = -A$$

Example 1.2

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

and

$$(1+i) \begin{bmatrix} 1-i & 0 \\ 1 & -1+2i \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1+i & -3+i \end{bmatrix}$$

The first example above shows that we can add a real matrix to a complex matrix by treating it as a complex matrix. Similarly, a real matrix can be multiplied with a complex scalar, and a complex matrix with a real scalar.

[3] We will use the phrase "for all (i, j)" to mean "for all $1 \le i \le m$ and $1 \le j \le n$" if the ranges of the indices i (1 to m) and j (1 to n) are known.

We list below some properties of matrix addition and scalar multiplication.

A1. $A + B = B + A$

A2. $(A + B) + C = A + (B + C)$

A3. $A + O = A$

A4. $A + (-A) = O$

S1. $c(dA) = (cd)A$

S2. $(c + d)A = cA + dA$

S3. $c(A + B) = cA + cB$

S4. $1A = A$, where 1 is the multiplicative identity of F.

Note that in property S2 the same symbol "+" is used to denote both the addition of the scalars c and d, and also the addition of the matrices cA and dA. This should cause no confusion as which operation is meant is clear from the operands. Similarly, in property S1, cd represents the product of the scalars c and d, whereas dA represents the scalar multiplication of the matrix A with the scalar d, although no symbol is used to denote either of these two different types of multiplications.[4]

The properties above follow from the properties of addition and multiplication of the scalars involved. For example, property S2 can be proved as

$$\begin{array}{rcccccl} (c+d)A & = & [\,(c+d)a_{ij}\,] & = & [\,ca_{ij} + da_{ij}\,] & & \\ & = & [\,ca_{ij}\,] + [\,da_{ij}\,] & = & c\,[\,a_{ij}\,] + d\,[\,a_{ij}\,] & = & cA + dA \end{array}$$

Proofs of the other properties are left to the reader.

From the basic properties above we can derive further useful properties of matrix addition and scalar multiplication. For example,

$$\begin{array}{rcl} A + B = O & \Longrightarrow & B = -A \\ A + B = A + C & \Longrightarrow & B = C \\ cA = O & \Longrightarrow & c = 0 \quad \text{or} \quad A = O \end{array}$$

We finally note that if $A_1, A_2, \ldots, A_k \in \mathsf{F}^{m\times n}$ then because of property A2, an expression of the form $A_1 + A_2 + \cdots + A_k$ unambiguously defines a matrix in $\mathsf{F}^{m\times n}$.

1.2.2 Transpose of a Matrix

Let A be an $m \times n$ matrix. The $n \times m$ matrix obtained by interchanging the rows and columns of A is called the **transpose** of A, denoted A^t. Thus if $A = [\,a_{ij}\,]_{m\times n}$ then $A^t = B = [\,b_{ij}\,]_{n\times m}$ where $b_{ij} = a_{ji}$ for all (i, j). From the definition it follows that

$$(A^t)^t = A$$

[4]If we used a symbol for scalar multiplication, say "·", then property S1 would be stated as

$$c \cdot (d \cdot A) = (cd) \cdot A$$

If A is a complex matrix then its conjugate transpose (obtained by transposing A and replacing every element with its complex conjugate, or vice versa) is called the **Hermitian adjoint** of A, denoted A^h. Thus if $A = [\, a_{ij} \,]_{m\times n}$ then $A^h = C = [\, c_{ij} \,]_{n\times m}$ where $c_{ij} = a^*_{ji}$ for all (i, j). Again, the definition implies that

$$(A^h)^h = A$$

Note that if A is real then $A^h = A^t$. Hence all properties concerning the Hermitian adjoint of a complex matrix are valid for the transpose of a real matrix. For this reason, from now on we will state and prove such properties only for the Hermitian adjoint. For example, the properties

$$\begin{aligned} (A+B)^h &= A^h + B^h \\ (cA)^h &= c^* A^h \end{aligned}$$

involving complex matrices A and B and a complex scalar c can be shown in one or two steps. We can then safely state without proof that

$$\begin{aligned} (A+B)^t &= A^t + B^t \\ (cA)^t &= cA^t \end{aligned}$$

for real matrices A and B and a real scalar c.

Clearly, the transpose or Hermitian adjoint of a row vector is a column vector, and vice versa. Also, $D^t = D$ and $D^h = D^*$ for any diagonal matrix D. Finally,

$$O^h_{m\times n} \;=\; O^t_{m\times n} \;=\; O_{n\times m}$$

whether O is treated as a real or as a complex matrix.

A square matrix A is called **symmetric** if $A^t = A$, and **skew-symmetric** if $A^t = -A$. Thus $A = [\, a_{ij} \,]_{n\times n}$ is symmetric if and only if $a_{ij} = a_{ji}$ for all (i, j), and skew-symmetric if and only if $a_{ij} = -a_{ji}$ for all (i, j). Note that if A is skew-symmetric then $a_{ii} = -a_{ii}$, which requires that the diagonal elements should be zero.

A complex square matrix is called **Hermitian** if $A^h = A$, and **skew-Hermitian** if $A^h = -A$. Clearly, a real Hermitian matrix is symmetric, and a real skew-Hermitian matrix is skew-symmetric. Further properties of Hermitian matrices are dealt with in Exercises 1.14-1.16.

Example 1.3

The transpose and the Hermitian adjoint of the matrix B in Example 1.1 are

$$B^t = \begin{bmatrix} 1+2i & 0 & -1 \\ -3 & -3i & -2 \\ -1+i & 5 & 3+2i \end{bmatrix}, \quad B^h = \begin{bmatrix} 1-2i & 0 & -1 \\ -3 & 3i & -2 \\ -1-i & 5 & 3-2i \end{bmatrix}$$

The transpose of the lower triangular matrix C in Example 1.1 is the upper triangular matrix

$$C^t = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 5 \end{bmatrix}$$

The matrices

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 0 & 2 \\ -1 & 2 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$$

are symmetric, skew-symmetric, and Hermitian, respectively.

1.2.3 Matrix Multiplication

Let $A = [\,a_{ij}\,]_{m\times n}$ and $B = [\,b_{ij}\,]_{p\times q}$. If A has exactly as many columns as B has rows, that is, if $n = p$, then A and B are said to be **compatible** for the product AB. The product is then defined to be an $m \times q$ matrix $AB = C = [\,c_{ij}\,]_{m\times q}$ whose elements are

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

That is, the (i,j)th element of the product is the sum of the ordered products of the ith row elements of A with the jth column elements of B.[5] Thus it takes n multiplications to compute a single element of the product, and therefore, mnq multiplications to compute C.

Example 1.4

Let

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & -1 & 1 & 3 \\ 1 & 0 & -1 & 2 \end{bmatrix}$$

Since A is 2×3 and B is 3×4, the product $C = AB$ is defined, and is a 2×4 matrix. Some of the elements of C are found as

$$\begin{aligned} c_{11} &= 1\cdot 1 + (-1)\cdot 0 + 2\cdot 1 &&= 3 \\ c_{12} &= 1\cdot 2 + (-1)\cdot(-1) + 2\cdot 0 &&= 3 \\ c_{24} &= 3\cdot(-1) + 0\cdot 3 + 1\cdot 2 &&= -1 \end{aligned}$$

Computing other elements of C similarly, we obtain

$$C = \begin{bmatrix} 3 & 3 & -3 & 0 \\ 4 & 6 & -1 & -1 \end{bmatrix}$$

On the other hand, the product BA is not defined.

Example 1.5

Let

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

Then

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} -1 \\ 5 \end{bmatrix} \\ \mathbf{y}A &= \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \\ \mathbf{x}\mathbf{y} &= \begin{bmatrix} -2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} \end{aligned}$$

Other pairwise products are not defined.

[5] Implicit in the definition of matrix multiplication is the assumption that elements of A can be multiplied with those of B, which requires that they belong to the same field. However, we can multiply a complex matrix with a real one.

Examples 1.4 and 1.5 illustrate that matrix multiplication is not commutative. If A is $m \times n$ and B is $n \times q$, then AB is an $m \times q$ matrix, but BA is not defined unless $q = m$. If $q = m$, that is, when B is $n \times m$, then both AB and BA are defined, but AB is $m \times m$, while BA is $n \times n$. Even when A and B are both $n \times n$ square matrices, so that both AB and BA are defined and are $n \times n$ matrices, in general $AB \neq BA$. For example, if

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{1.1}$$

then $AB = O$, but $BA = A$. Therefore, the order in which two matrices are multiplied is important. In the product AB, B is said to be premultiplied (or multiplied from left) with A, and A is said to be postmultiplied (or multiplied from right) with B.

Two square matrices A and B of the same order are said to **commute** if $AB = BA$. Obviously, every square matrix commutes with itself. If A is a square matrix then we denote the product AA by A^2. Higher order powers of A are defined recursively as $A^{k+1} = AA^k\,, \quad k = 1, 2, \ldots$.

We now state several properties of matrix multiplication.

M1. $(AB)C = A(BC)$

M2. $A(B + C) = AB + AC$
$(A + B)C = AC + BC$

M3. $OA = O$
$AO = O$

M4. $(AB)^h = B^h A^h$

In stating these properties, we implicitly assume that the products involved are defined. Note that the two properties in item M2, as well as those in item M3, are different (that is, one does not follow from the other), as matrix multiplication is not commutative.

The properties above follow directly from definitions. For example, letting $A = [\,a_{ij}\,]_{m\times n}$, $B = [\,b_{ij}\,]_{n\times q}$ and $C = [\,c_{ij}\,]_{n\times q}$, we have

$$\begin{aligned} A(B + C) &= [\,a_{ij}\,][\,b_{ij} + c_{ij}\,] \\ &= [\sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})\,] \\ &= [\sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj}\,] \\ &= [\sum_{k=1}^{n} a_{ik}b_{kj}\,] + [\sum_{k=1}^{n} a_{ik}c_{kj}\,] \\ &= AB + AC \end{aligned}$$

Some usual properties of multiplication of scalars (like commutativity) do not hold for matrices. For example, $AB = O$ does not necessarily imply that $A = O$ or $B = O$ as for the matrices in (1.1). Similarly, $AB = AC$ does not necessarily imply that $B = C$.

Finally, as in the case of matrix addition, an expression of the form $A_1A_2\cdots A_k$ is unambiguous due to property M1, and can be evaluated by computing pairwise products of adjacent matrices in any sequence without changing the original order of the matrices. For example, the product $ABCD$ can be evaluated as

$$((AB)C)D) \quad \text{or} \quad (AB)(CD) \quad \text{or} \quad (A(BC))D \quad \text{or} \quad A((BC)D) \quad \text{or} \quad A(B(CD))$$

However, a careful reader might observe that one of these equivalent expressions might be easier to compute depending on the order of the matrices (see Exercise 1.9).

Multiplication with a diagonal matrix

Let

$$D = [\, d_{ij} \,] = \operatorname{diag}[\, d_1, d_2, \ldots, d_n \,]$$

Also, let $A = [\, a_{ij} \,]_{m\times n}$, and consider the product $C = AD = [\, c_{ij} \,]_{m\times n}$. Since $d_{kj} = 0$ for $k \neq j$ and $d_{jj} = d_j$,

$$c_{ij} = \sum_{k=1}^{n} a_{ik}d_{kj} = a_{ij}d_{jj} = a_{ij}d_j$$

Thus

$$AD = \begin{bmatrix} a_{11}d_1 & a_{12}d_2 & \cdots & a_{1n}d_n \\ a_{21}d_1 & a_{22}d_2 & \cdots & a_{2n}d_n \\ \vdots & \vdots & & \vdots \\ a_{m1}d_1 & a_{m2}d_2 & \cdots & a_{mn}d_n \end{bmatrix}$$

that is, the product AD is obtained simply by scaling the columns of A with the diagonal elements of D.

Similarly, if $B = [\, b_{ij} \,]_{n\times p}$ then

$$DB = \begin{bmatrix} d_1b_{11} & d_1b_{12} & \cdots & d_1b_{1p} \\ d_2b_{21} & d_2b_{22} & \cdots & d_2b_{2p} \\ \vdots & \vdots & & \vdots \\ d_nb_{n1} & d_nb_{n2} & \cdots & d_nb_{np} \end{bmatrix}$$

that is, the product DB is obtained by scaling the rows of B with the diagonal elements of D.

Now consider a diagonal matrix having all 1's on its diagonal. Such a matrix is called an **identity** matrix, denoted I or I_n if the order needs to be specified. That is,

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n} = \operatorname{diag}[\, 1, 1, \ldots, 1 \,]$$

Applying the above results about the product of a matrix with a diagonal matrix to $D = I$ we obtain

M5. $IA = A$
$AI = A$

as further properties of matrix multiplication. In fact, it is because of these properties that I is called an identity matrix: It acts like a multiplicative identity.

From the definition it follows that

$$I^h = I^t = I$$

Note that the jth column of an $n \times n$ identity matrix is an $n \times 1$ vector having all 0's except a 1 in the jth position. Because of their special structure columns of an identity matrix deserve a special notation: We denote the jth column by $\mathbf{e}_j$. Thus

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Then, obviously, the rows of I_n are $\mathbf{e}_1^t, \mathbf{e}_2^t, \ldots, \mathbf{e}_n^t$.

1.3 Partitioned Matrices

Let A be any matrix. By deleting some of the rows and some of the columns of A we obtain a smaller matrix called a **submatrix** of A.

Let us partition the rows of an $m \times n$ matrix A into p groups of size $m_1, \ldots, m_p$, and the columns into q groups of size $n_1, \ldots, n_q$, where

$$\sum_{i=1}^{p} m_i = m, \quad \sum_{j=1}^{q} n_j = n$$

Calling the submatrix of A consisting of the ith group of m_i rows and the jth group of n_j columns A_{ij}, we can represent A as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix} = [\, A_{ij} \,]$$

Each submatrix A_{ij} in the above representation is called a **block** of A.[6] In general, the blocks A_{ij} are of different order; however, all blocks in the same row block have the same number of rows, and all blocks in the same column block have the same number of columns. Note also that only the rows (but not the columns) or only the

[6]Representing a partitioned matrix A in terms of its blocks does not mean that A is a matrix with elements themselves being matrices. However, the blocks of a partitioned matrix may be treated just like its elements in some matrix operations as explained later in this section.

columns (but not the rows) of a matrix may be partitioned. Below are some examples of partitioned matrices.

$$A = \left[\begin{array}{cc|c} * & * & 0 \\ \hline 1 & 0 & * \\ 0 & 1 & * \\ \hline 0 & 1 & * \end{array}\right] = \begin{bmatrix} A_{11} & O \\ I & A_{22} \\ \mathbf{e}_2^t & A_{33} \end{bmatrix}$$

$$B = \left[\begin{array}{c|cc} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{array}\right] = [\, B_1 \;\; B_2 \,]$$

$$I_n = \left[\begin{array}{c|c|c|c} 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & 1 \end{array}\right] = [\, \mathbf{e}_1 \;\; \mathbf{e}_2 \;\; \cdots \;\; \mathbf{e}_n \,]$$

$$= \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 1 \end{array}\right] = \begin{bmatrix} \mathbf{e}_1^t \\ \mathbf{e}_2^t \\ \vdots \\ \mathbf{e}_n^t \end{bmatrix}$$

Note how partitioning of A is used to display its special structure and how the identity matrix can be expressed in terms of its columns or rows.

If A and B are matrices of the same order and are partitioned in exactly the same way so that their corresponding blocks are of the same order, then the sum $A + B$ can be obtained by adding the corresponding blocks. That is, if

$$A = [A_{ij}] \quad \text{and} \quad B = [B_{ij}]$$

where the blocks A_{ij} and B_{ij} are of the same order for all (i, j), then

$$A + B = [A_{ij} + B_{ij}]$$

If A and B are partitioned matrices compatible for the product AB, then the product can be obtained by treating the blocks of A and B as if they were their elements. That is, if $A = [\, A_{ij} \,]$ and $B = [\, B_{ij} \,]$ then $AB = C = [\, C_{ij} \,]$, where

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

Of course, this requires that the blocks A_{ik} and B_{kj} be compatible for the products $A_{ik}B_{kj}$ for all (i, k, j). In other words, columns of A must be partitioned in exactly the same way as the rows of B are partitioned.

Block multiplication is useful in expressing matrix products in a compact form. For example, if A is an $m \times n$ matrix and B is an $n \times q$ matrix partitioned into its columns as

$$B = [\, \mathbf{b}_1 \;\; \mathbf{b}_2 \;\; \cdots \;\; \mathbf{b}_q \,]$$

where each $\mathbf{b}_j$ is an $n \times 1$ column vector, then the product AB can be obtained by block multiplication as

$$AB = A\,[\,\mathbf{b}_1 \;\; \mathbf{b}_2 \;\; \cdots \;\; \mathbf{b}_q\,] = [\,A\mathbf{b}_1 \;\; A\mathbf{b}_2 \;\; \cdots \;\; A\mathbf{b}_q\,]$$

Observe that the product is also partitioned into its columns, the jth column being an $m \times 1$ vector obtained by premultiplying the jth column of B with A.

Similarly, if A is partitioned into its rows as

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{bmatrix}$$

where each $\boldsymbol{\alpha}_i$ is a $1 \times n$ row vector, then the product AB can be expressed as

$$AB = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{bmatrix} B = \begin{bmatrix} \boldsymbol{\alpha}_1 B \\ \boldsymbol{\alpha}_2 B \\ \vdots \\ \boldsymbol{\alpha}_m B \end{bmatrix}$$

Now the product is partitioned into its rows, the ith row being a $1 \times q$ vector obtained by postmultiplying the ith row of A with B.

If both A and B are partitioned as above (A into its rows and B into its columns), then $AB = C = [\,C_{ij}\,]_{m \times q}$, where each block $C_{ij} = \boldsymbol{\alpha}_i \mathbf{b}_j$ is a scalar. In fact, $C_{ij} = c_{ij}$, the (i,j)th element of C, as expected. This is actually how matrix multiplication is defined. The (i,j)th element of the product AB is the product of the ith row of A with the jth column of B.

Alternatively, we may choose to partition A into its columns and B into its rows as

$$A = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n\,]\,, \quad B = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix}$$

where now each $\mathbf{a}_i$ is an $m \times 1$ column vector and each $\boldsymbol{\beta}_j$ is a $1 \times q$ row vector. Then the product AB consists of only one block given as

$$AB = \mathbf{a}_1\boldsymbol{\beta}_1 + \mathbf{a}_2\boldsymbol{\beta}_2 + \ldots + \mathbf{a}_n\boldsymbol{\beta}_n$$

where each product term $\mathbf{a}_i\boldsymbol{\beta}_i$ is an $m \times q$ matrix.

As a final property of partitioned matrices we note that if $A = [\,A_{ij}\,]$ then $A^h = [\,A_{ji}^h\,]$ as the reader can verify by examples.

1.4 Systems of Linear Equations

An $m \times n$ system of linear equations, or an $m \times n$ **linear system**, is a set of m equations in n unknown variables $x_1, x_2, \ldots, x_n$, written as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{1.2}$$

where the coefficients a_{ij} and the constants b_i are fixed scalars. Letting

$$A = [\, a_{ij} \,], \quad \mathbf{x} = \mathrm{col}\,[\, x_1, x_2, \ldots, x_n \,], \quad \mathbf{b} = \mathrm{col}\,[\, b_1, b_2, \ldots, b_n \,]$$

the system in (1.2) can be written in matrix form as

$$A\mathbf{x} = \mathbf{b} \tag{1.3}$$

A is called the **coefficient matrix** of (1.3). If $\mathbf{b}=\mathbf{0}$, then the system (1.3) is said to be **homogeneous**.

An $n \times 1$ column vector $\mathbf{x} = \boldsymbol{\phi}$ is called a **solution** of (1.3) if $A\boldsymbol{\phi} = \mathbf{b}$. A system may have no solution, a unique solution, or more than one solution. If it has at least one solution, it is said to be **consistent**, otherwise, **inconsistent**. A homogeneous system is always consistent as it has at least the **trivial** (null) solution $\mathbf{x} = \mathbf{0}$.

We are interested in the following problems associated with a linear system:

a) Determine whether the system is consistent.

b) If it is consistent

   i. determine if it has a unique solution or many solutions,

   ii. if it has a unique solution find it,

   iii. if it has many solutions, find a solution or all solutions.

c) If it is inconsistent find $\mathbf{x} = \boldsymbol{\phi}$ that is closest to being a solution.

In this section we will deal with problems (a) and (b), leaving (c) to Chapter 7.

Example 1.6

The system

$$\begin{array}{ccccc} x_1 & - & x_2 & = & 1 \\ x_1 & + & x_2 & = & 5 \end{array} \tag{1.4}$$

has a unique solution

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \tag{1.5}$$

which can be obtained graphically as illustrated in Figure 1.1(a). Each of the equations describes a straight line in the x_1x_2 plane, and the solution is their intersection point.

The equations of the system

$$\begin{array}{ccccc} x_1 & - & x_2 & = & 1 \\ 2x_1 & - & 2x_2 & = & -6 \end{array}$$

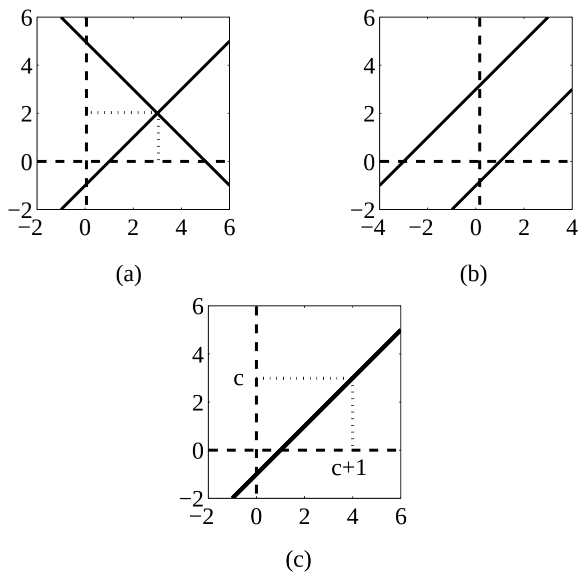

Figure 1.1: Geometric representation of systems in Example 1.6

describe parallel lines in the x_1x_2 plane as illustrated in Figure 1.1(b). Since there is no point common to both lines, the system has no solution.

On the other hand, the equations of the system

$$\begin{array}{rcrcl} x_1 & - & x_2 & = & 1 \\ 2x_1 & - & 2x_2 & = & 2 \end{array}$$

are proportional, and describe the same line shown in Figure 1.1(c). Since any point on this line satisfies both equations, the system has infinitely many solutions. To characterize these solutions we choose one of the variables, say x_2, arbitrarily as $x_2 = c$, and determine the other variable from either of the equations as $x_1 = 1 + c$. Thus we obtain a one-parameter family of solutions described as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1+c \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The geometric interpretation of equations and their solution(s) can easily be generalized to systems containing three variables x_1, x_2 and x_3, where each equation defines a plane in the $x_1x_2x_3$ space, and any point common to all planes, if it exists, defines a solution (see Exercise 1.28). However, when there are more than three variables, the geometric interpretation loses much of its appeal (for we can not visualize what an equation in four or more variables describes), and we resort to pure algebraic methods. One such method is the elimination method, which we recall by reconsidering the example above.

Example 1.7

Let us consider the system in (1.4). One way of finding the solution is to express one of the variables in terms of the other by using one of the equations, and then to substitute this expression into the other equation. For example, by expressing x_1 in terms of x_2 using the first equation we get

$$x_1 = 1 + x_2$$

Substituting this expression for x_1 into the second equation and rearranging the terms we obtain

$$2x_2 = 4 \tag{1.6}$$

The last equation gives $x_2 = 2$, and from the expression for x_1 we get

$$x_1 = 1 + x_2 = 1 + 2 = 3$$

thus obtaining the solution in (1.5).

What we did by expressing x_1 in terms of x_2 and substituting this expression into the second equation was to eliminate x_1 from the second equation. We could do this simply by subtracting the first equation from the second, which would give (1.6) directly.

Alternatively, we could eliminate x_2 from the second equation by adding the first equation to it. This would give

$$2x_1 = 6$$

from which we would obtain $x_1 = 3$. Then from the first equation we would get

$$x_2 = x_1 - 1 = 3 - 1 = 2$$

reaching the same solution.

For the simple example considered above, it makes no difference whether we eliminated x_1 or x_2. However, to solve larger equations we need to be more systematic in the elimination process, especially if we are using a computer program to do the job. A systematic procedure is based on transforming the given system into a simpler equivalent system in which variables can be solved one after the other by successive substitutions as we explain by the following example.

Example 1.8

Let us solve the system

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & - & x_3 & = & 1 \\ -2x_1 & + & 8x_2 & - & x_3 & = & 5 \\ 2x_1 & - & 6x_2 & + & 2x_3 & = & -4 \end{array} \tag{1.7}$$

by using the elimination method.

We first eliminate the variable x_1 from all equations except one. Since it appears in all three equations, we can associate it with any one of them and eliminate from the other two. Associating x_1 arbitrarily with the first equation, we eliminate it from the second and third equations by adding 2 times the first equation to the second and -2 times the first equation to the third. After these manipulations the equations become

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & - & x_3 & = & 1 \\ & & 4x_2 & - & 3x_3 & = & 7 \\ & - & 2x_2 & + & 4x_3 & = & -6 \end{array} \tag{1.8}$$

Next we eliminate x_2 from one of the last two equations. This can be done by associating x_2 with the second equation and eliminating it from the third (by adding

1/2 times the second equation to the third). Alternatively, we may associate x_2 with the third equation and eliminate it from the second (by adding 2 times the third equation to the second). To avoid dealing with fractions we choose the latter. However, before the elimination we first interchange the second and third equations so that the equation with which x_2 is associated comes before those from which it is to be eliminated. This gives

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & - & x_3 & = & 1 \\ & - & 2x_2 & + & 4x_3 & = & -6 \\ & & 4x_2 & - & 3x_3 & = & 7 \end{array} \tag{1.9}$$

Now we add 2 times the second equation to the third and obtain

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & - & x_3 & = & 1 \\ & - & 2x_2 & + & 4x_3 & = & -6 \\ & & & & 5x_3 & = & -5 \end{array} \tag{1.10}$$

The system in (1.10) has a triangular shape which allows us to solve the unknown variables starting from the last equation and working backwards. From the last equation we obtain

$$x_3 = -5/5 = -1$$

Substituting the value of x_3 into the second equation we find

$$x_2 = (-1/2)(-6 - 4x_3) = (-1/2)(-6 + 4) = 1$$

Finally, substituting the values of x_2 and x_3 into the first equation we get

$$x_1 = 1 + 2x_2 + x_3 = 1 + 2 - 1 = 2$$

Thus we obtain the solution of the system (1.7) as

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \tag{1.11}$$

Instead of finding x_2 and x_1 by successive backward substitutions, we can continue elimination of the variables in (1.10) in the reverse order. Starting with the system in (1.10), we first multiply the last equation with 1/5 to normalize the coefficient of x_3 to 1, and then add 1 and -4 times the resulting equation to the first and the second equations to eliminate x_3 from the first two equations. After these operations the system becomes

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

Now we scale the second equation with $-1/2$, add 2 times the resulting equation to the first to eliminate x_2 from the first equation, and thus obtain

$$\begin{array}{rcrcrcr} x_1 & & & & & = & 2 \\ & & x_2 & & & = & 1 \\ & & & & x_3 & = & -1 \end{array} \tag{1.12}$$

Note that in the last two steps we not only eliminated the variables x_2 and x_3 from the first two equations but also normalized their coefficients to 1. (Coefficient of x_1 in the first equation was already 1 at the start, so we need not do anything about it.) The final system in (1.12) is so simple that it displays the solution.

Backward elimination need not wait until forward elimination is completed; they can be performed simultaneously. Consider the system in (1.9) in which x_1 is already eliminated from the second and third equations. Scaling the second equation with $-1/2$, and adding 2 and -4 times the resulting equation to the first and the third equations, we eliminate x_2 not only from the third equation but also from the first equations, and get

$$\begin{array}{ccccccc} x_1 & & & - & 5x_3 & = & 7 \\ & & x_2 & - & 2x_3 & = & 3 \\ & & & & 5x_3 & = & -5 \end{array}$$

Now, multiplying the third equation with $1/5$ and adding 5 and 2 times the resulting equation to the first and second equations, we eliminate x_3 from these equations and end up with (1.12). However, a careful reader may observe that it is not smart to perform forward and backward eliminations simultaneously (see Exercise 1.37).

Example 1.8 illustrates the recursive nature of the elimination method. After the elimination of x_1, the last two equations in (1.8) form a 2×2 system

$$\begin{array}{rcrcr} 4x_2 & - & 3x_3 & = & 7 \\ -2x_2 & + & 4x_3 & = & -6 \end{array} \tag{1.13}$$

which is obviously easier to solve than the original 3×3 system. Once x_2 and x_3 are solved from (1.13), x_1 can easily be found from the first equation in (1.8) by substitution. Now the process can be repeated for (1.13) to further reduce it to a simpler system. This is exactly what we do when we eliminate x_2 from one of the equations in (1.13) to reach the triangular system in (1.10). Thus at every step of the elimination process, both the number of equations and the number of unknowns are reduced by at least one.[7]

Let us consider the system (1.7) in matrix form:

$$\begin{bmatrix} 1 & -2 & -1 \\ -2 & 8 & -1 \\ 2 & -6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -4 \end{bmatrix} \tag{1.14}$$

After the first two operations the equations take the form in (1.8), which has the matrix representation

$$\begin{bmatrix} 1 & -2 & -1 \\ 0 & 4 & -3 \\ 0 & -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -6 \end{bmatrix}$$

Apparently, the operation of adding 2 times the first equation to the second equation is reflected as adding 2 times the first row of the coefficient matrix in (1.14) to its second row, and at the same time, adding 2 times the first element of the column matrix on the right-hand side of (1.14) to its second element. Since the same operations are involved

[7] Although in Example 1.8 the number of equations and and the number of unknowns are reduced by exactly one, we will later see examples where either or both are reduced by more than one.

in the coefficient matrix and the column on the right-hand side, we conveniently form an **augmented matrix**

$$[\,A\ \ \mathbf{b}\,] = \left[\begin{array}{rrr|r} 1 & -2 & -1 & 1 \\ -2 & 8 & -1 & 5 \\ 2 & -6 & 2 & -4 \end{array}\right]$$

associated with the system in (1.7), and represent the operations leading to (1.12) as row operations on the augmented matrix as

$$\left[\begin{array}{rrr|r} 1 & -2 & -1 & 1 \\ -2 & 8 & -1 & 5 \\ 2 & -6 & 2 & -4 \end{array}\right] \begin{array}{c} 2\mathrm{R}_1 + \mathrm{R}_2 \to \mathrm{R}_2 \\ -2\mathrm{R}_1 + \mathrm{R}_3 \to \mathrm{R}_3 \\ \longrightarrow \end{array} \left[\begin{array}{rrr|r} 1 & -2 & -1 & 1 \\ 0 & 4 & -3 & 7 \\ 0 & -2 & 4 & -6 \end{array}\right]$$

$$\begin{array}{c} \mathrm{R}_2 \leftrightarrow \mathrm{R}_3 \\ 2\mathrm{R}_2 + \mathrm{R}_3 \to \mathrm{R}_3 \\ \longrightarrow \end{array} \left[\begin{array}{rrr|r} 1 & -2 & -1 & 1 \\ 0 & -2 & 4 & -6 \\ 0 & 0 & 5 & -5 \end{array}\right]$$

$$\begin{array}{c} (1/5)\mathrm{R}_3 \to \mathrm{R}_3 \\ \mathrm{R}_3 + \mathrm{R}_1 \to \mathrm{R}_1 \\ -4\mathrm{R}_3 + \mathrm{R}_2 \to \mathrm{R}_2 \\ \longrightarrow \end{array} \left[\begin{array}{rrr|r} 1 & -2 & 0 & 0 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

$$\begin{array}{c} -(1/2)\mathrm{R}_2 \to \mathrm{R}_2 \\ 2\mathrm{R}_2 + \mathrm{R}_1 \to \mathrm{R}_1 \\ \longrightarrow \end{array} \left[\begin{array}{rrr|r} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

The first two steps of the above operations correspond to forward elimination of the variables, and the last two steps correspond to backward elimination. The notation "$\mathrm{R}_i \leftrightarrow \mathrm{R}_j$" denotes interchange of the ith and the jth rows, "$c\mathrm{R}_i \to \mathrm{R}_i$" denotes multiplication of the ith row by a nonzero scalar c, and "$c\mathrm{R}_i + \mathrm{R}_j \to \mathrm{R}_j$" denotes addition of c times the ith row to the jth row. Note that after the operation $\mathrm{R}_2 \leftrightarrow \mathrm{R}_3$ at the second step above, R_2 and R_3 denote the current second and third rows.

The following three types of operations on the rows of the augmented matrix are involved in the elimination process.

I: Interchange any two rows

II: Multiply any row by a nonzero scalar

III: Add any scalar multiple of a row to another row

These operations performed on the rows of a matrix are called **elementary row operations**. An $m \times n$ matrix B is said to be **row equivalent** to an $m \times n$ matrix A if it can be obtained from A by a finite number of elementary row operations. Clearly, to every elementary row operation there corresponds an inverse elementary row operation of the same kind. For example, the inverse of adding c times the ith row to the jth row is to add $-c$ times the ith row to the jth row. If B is obtained from A by a single elementary row operation, then A can be restored from B by performing the inverse operation. Thus if B is row equivalent to A, then A is row equivalent to B, and we say that A and B are row equivalent.[8] Two $m \times n$ systems of linear equations are said to be **equivalent** if their augmented matrices are row equivalent. Two equivalent systems either have the same solution(s), or are both inconsistent. We have already used this fact in Example 1.8 to find the solution of a system. Below we consider another example.

Example 1.9

Find the value of the parameter q such that the system

$$\begin{bmatrix} 1 & 1 & -1 & 2 & 0 \\ 2 & 2 & -2 & 3 & 2 \\ -1 & -1 & 3 & -4 & 2 \\ 1 & 1 & 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \\ 3 \\ q \end{bmatrix} \tag{1.15}$$

is consistent, and then find all solutions.

We first associate x_1 with the first equation and eliminate it from the last three equations. The operations involved in the elimination of x_1 are represented by elementary row operations on the augmented matrix as

$$\left[\begin{array}{rrrrr|r} 1 & 1 & -1 & 2 & 0 & -1 \\ 2 & 2 & -2 & 3 & 2 & -3 \\ -1 & -1 & 3 & -4 & 2 & 3 \\ 1 & 1 & 2 & 1 & -1 & q \end{array}\right]$$

$$\begin{array}{r} -2\mathrm{R}_1 + \mathrm{R}_2 \rightarrow \mathrm{R}_2 \\ \mathrm{R}_1 + \mathrm{R}_3 \rightarrow \mathrm{R}_3 \\ -\mathrm{R}_1 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrrrr|r} 1 & 1 & -1 & 2 & 0 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 2 & -2 & 2 & 2 \\ 0 & 0 & 3 & -1 & -1 & q+1 \end{array}\right]$$

At this point we observe that incidentally x_2 is also eliminated from the last three equations. We then continue with the elimination of the next variable, x_3, which appears in the third and fourth equations, and must be associated with one of them. We associate x_3 with the third equation and interchange the second and third equations. The rest of the elimination process is straightforward, and is summarized below.

$$\begin{array}{r} \mathrm{R}_2 \leftrightarrow \mathrm{R}_3 \\ -(3/2)\mathrm{R}_2 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrrrr|r} 1 & 1 & -1 & 2 & 0 & -1 \\ 0 & 0 & 2 & -2 & 2 & 2 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 2 & -4 & q-2 \end{array}\right]$$

$$\begin{array}{ccccc} \downarrow & & \downarrow & \downarrow & \end{array}$$

$$\begin{array}{r} 2\mathrm{R}_3 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrrrr|r} 1 & 1 & -1 & 2 & 0 & -1 \\ 0 & 0 & 2 & -2 & 2 & 2 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & q-4 \end{array}\right] \tag{1.16}$$

[8] An equivalence relation defined on a set $\mathbf{S}$, denoted $\equiv$, is a reflexive, symmetric and transitive relation among the elements of $\mathbf{S}$. That is, for all $a, b, c \in \mathbf{S}$, $a \equiv a$, if $a \equiv b$ then $b \equiv a$, and if $a \equiv b$ and $b \equiv c$ then $a \equiv c$. In this sense, row equivalence is indeed an equivalence relation on $\mathsf{F}^{m \times n}$. An equivalence relation partitions $\mathbf{S}$ into disjoint subsets, called equivalence classes, such that every element of the set belongs to one and only one equivalence class and any two equivalent elements belong to the same equivalence class. Hence row equivalence partitions $\mathsf{F}^{m \times n}$ into equivalence classes such that any two matrix in the same equivalence class can be obtained from each other by a finite sequence of elementary row operations.

$$\begin{array}{c} -\mathrm{R}_3 \to \mathrm{R}_3 \\ -2\mathrm{R}_3 + \mathrm{R}_1 \to \mathrm{R}_1 \\ 2\mathrm{R}_3 + \mathrm{R}_2 \to \mathrm{R}_2 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrrrr|r} 1 & 1 & -1 & 0 & 4 & -3 \\ 0 & 0 & 2 & 0 & -2 & 4 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & q-4 \end{array}\right]$$

$$\begin{array}{ccccc} \downarrow & & \downarrow & \downarrow & \\ \end{array}$$

$$\begin{array}{c} (1/2)\mathrm{R}_2 \to \mathrm{R}_2 \\ \mathrm{R}_2 + \mathrm{R}_1 \to \mathrm{R}_1 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrrrr|r} 1 & 1 & 0 & 0 & 3 & -1 \\ 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & q-4 \end{array}\right] \tag{1.17}$$

In the above sequence of elementary row operations, steps leading to (1.16) correspond to forward elimination of the variables associated with the columns marked by arrows, and those leading to (1.17) correspond to scaling of the equations and backward elimination of the same variables.

The last equation associated with the augmented matrix in (1.17) is

$$0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4 + 0 \cdot x_5 = q - 4$$

from which we observe that if $q \neq 4$ then the system is inconsistent as this equation is never satisfied. On the other hand, if $q = 4$ then this equation is a trivial identity $0 = 0$ for any choice of the variables, and can be discarded. Then using the first three equations we express the variables associated with the marked columns of the augmented matrix in terms of the others as

$$\begin{aligned} x_1 &= -1 - x_2 - 3x_5 \\ x_3 &= 2 + x_5 \\ x_4 &= 1 + 2x_5 \end{aligned} \tag{1.18}$$

From (1.18) we see that we can choose the variables x_2 and x_5 arbitrarily, and calculate x_1, x_3 and x_4 from these relations to obtain a solution. Letting $x_2 = c_1$ and $x_5 = c_2$, where $c_1, c_2 \in \mathbf{R}$ are arbitrary, and calculating x_1, x_3 and x_4 from (1.18), we obtain the solution in parametric form as

$$\begin{aligned} x_1 &= -1 - c_1 - 3c_2 \\ x_2 &= c_1 \\ x_3 &= 2 + c_2 \\ x_4 &= 1 + 2c_2 \\ x_5 &= c_2 \end{aligned}$$

or equivalently, as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} \tag{1.19}$$

Note that we could obtain (1.18) and hence the solution in (1.19) from the augmented matrix in (1.16) by back substitution of the marked variables.

Unlike the system in (1.7), which has the unique solution given in (1.11), the system in (1.15) is either inconsistent (if $q \neq 4$) or has infinitely many solutions as given in (1.19). For example,

$$\mathbf{x} = \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

are two solutions corresponding to the choice of the arbitrary constants as $(c_1, c_2) = (0, 0)$ and $(c_1, c_2) = (1, -1)$, respectively.

Example 1.9 outlines a general procedure to determine whether a given system is consistent and to find solutions when it is consistent. It is based on transforming the coefficient matrix A into a simple form by performing elementary row operations on the augmented matrix. We now give a precise definition of what we mean by a simple form.

An $m \times n$ matrix R is said to be in **row echelon form** if it has the following characteristics.

i) First r rows of R are nonzero, and the remaining $m - r$ rows are zero for some $0 \leq r \leq m$.

ii) The first nonzero element in each of the first $r - 1$ rows lies to the left of the first nonzero element in the subsequent row. (If $r = 0$ or $r = 1$ this item does not apply.)

The number of nonzero rows, r, is called the **row rank** of R. The first nonzero element of each of the first r rows is called the **leading entry** of its row, and the column which contains the leading entry of the ith row is called the ith **basic column**. Thus, if $1 \leq i < p \leq r$, then the ith basic column lies to the left of the pth basic column. Note that this requirement is equivalent to R having all zero elements below a jagged diagonal defined by the leading entries.

A matrix R in row echelon form is said to be in **reduced row echelon form** if it satisfies the following additional conditions.

iii) Each leading entry is a 1.

iv) The ith leading entry is the only nonzero element in the ith basic column.

For example, the matrices

$$R_1 = \begin{bmatrix} 0 & 1 & -1 & 2 & -4 \\ 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad R_2 = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are in row echelon form with $r(R_1) = 3$ and $r(R_2) = 2$, and the matrix

$$R_3 = \begin{bmatrix} 1 & -3 & 0 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in reduced row echelon form. As further examples, the reader can check that the coefficient matrix associated with the augmented matrix in (1.16) is in row echelon form, and the coefficient matrix associated with the augmented matrix in (1.17) is in reduced row echelon form.

An algorithm to bring a given $m \times n$ matrix into reduced row echelon form by means of elementary row operations is given in Table 1.1. The algorithm, which is known as **Gaussian elimination**, simply imitates the steps involved in Example 1.9.

Table 1.1: Gaussian Elimination Algorithm

1.	Set $r = 0$
	[Forward Elimination]
2.	For $j = 1 : n$
3.	Find $r < p \leq m$ such that $a_{pj} \neq 0$. If none, increment j
4.	$r \leftarrow r + 1$
5.	$j_r \leftarrow j$
6.	If $p > r$, $\mathrm{R}_p \leftrightarrow \mathrm{R}_r$
7.	For $i = r + 1 : m$
8.	$\mu_{ij} \leftarrow a_{ij}/a_{rj}$
9.	$\mathrm{R}_i \leftarrow \mathrm{R}_i - \mu_{ij}\mathrm{R}_r$
10.	End
11.	End
	[Backward Elimination]
12.	For $p = r : 2$
13.	$\mathrm{R}_p \leftarrow (1/a_{pj_p})\mathrm{R}_p$
14.	For $i = 1 : p - 1$
15.	$\mathrm{R}_i \leftarrow \mathrm{R}_i - a_{ij_p}\mathrm{R}_p$
16.	End
17.	End
18.	$\mathrm{R}_1 \leftarrow (1/a_{1j_1})\mathrm{R}_1$

The algorithm returns the reduced row echelon form of A written over A, the rank of the reduced row echelon form (r), and the column indices of the basic columns ($j_1, \ldots, j_r$). In steps 1-11, forward elimination is performed and A is brought to a row echelon form. The nonzero element a_{pj} found in Step 3 of the algorithm is called a **pivot element** of the jth column. After incrementing r in Step 4, the pivot element is brought to the (r, j)th position by a row interchange in Step 6 (if it is not already there), and it becomes the leading entry of its row. In steps 7-10, the pivot element is used to nullify the elements below it. In steps 12-18, leading entries are normalized to unity and A is brought into reduced row echelon form by means of backward elimination.[9]

Whether a system of linear equations is consistent can be determined from the reduced row echelon form of the augmented matrix of the system. If it is consistent, all solutions can be obtained in parametric form by choosing the non-basic variables arbitrarily and expressing the basic variables in terms of the non-basic variables. (The basic and non-basic variables are the unknowns corresponding to the basic and non-basic columns of the coefficient matrix.)

As illustrated in Example 1.9, the general form of the solution is

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

where $\nu = n - r$ is the number of non-basic variables. $\boldsymbol{\phi}_p$ and $\boldsymbol{\phi}_c$ are called a **particular solution** and the **complementary solution**, respectively. Obviously, when $r = n$, i.e., when there are no non-basic variables that can be chosen arbitrarily, then the complementary solution does not exist. The significance of particular and complementary solutions is studied in the next section.

Depending on the choice of the pivot element, Gaussian Elimination may produce different matrices in row echelon form at the end of step 11. However, upon completion of the algorithm, we end up with a unique matrix in reduced row echelon form. In fact, uniqueness of the reduced row echelon form is not specific to Gaussian elimination, but is a result of the fact that any given matrix A is row equivalent to a unique reduced row echelon matrix, which we define as the reduced row echelon form of A. In other words, independent of the algorithm used, if a finite sequence of elementary row operations on A results in a reduced row echelon matrix R, then R is the unique reduced row echelon form of A.[10] Consequently, all matrices which are row equivalent to A have the same reduced row echelon form, which can be interpreted as a convenient representative of its equivalence class.

The **row rank** of a matrix, denoted $r(A)$, is defined to be the row rank of its unique reduced row echelon form. Thus all row equivalent matrices have the same row rank. From the reduced row echelon form we deduce that if A is $m \times n$, then not only $r(A) \leq m$ but also $r(A) \leq n$.

Example 1.10

In Example 1.9 we obtained a row echelon form of the coefficient matrix in (1.15) as in (1.16). A different row echelon form of the same matrix can be obtained by choosing different pivots as

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

[9] Some books use the term "Gaussian elimination" to refer to the forward elimination process only, and call the complete algorithm in Table 1.1 the **Gauss-Jordan algorithm**.

[10] We shall prove this fact in Chapter 4.

$$
\begin{array}{r}
\mathrm{R}_3 \leftrightarrow \mathrm{R}_1 \\
2\mathrm{R}_1 + \mathrm{R}_2 \rightarrow \mathrm{R}_2 \\
\mathrm{R}_1 + \mathrm{R}_3 \rightarrow \mathrm{R}_3 \\
\mathrm{R}_1 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
-1 & -1 & 3 & -4 & 2 \\
0 & 0 & 4 & -5 & 6 \\
0 & 0 & 2 & -2 & 2 \\
0 & 0 & 5 & -3 & 1
\end{bmatrix}
$$

$$
\begin{array}{r}
\mathrm{R}_3 \leftrightarrow \mathrm{R}_2 \\
-2\mathrm{R}_2 + \mathrm{R}_3 \rightarrow \mathrm{R}_3 \\
-(5/2)\mathrm{R}_2 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
-1 & -1 & 3 & -4 & 2 \\
0 & 0 & 2 & -2 & 2 \\
0 & 0 & 0 & -1 & 2 \\
0 & 0 & 0 & 2 & -4
\end{bmatrix}
$$

$$
\begin{array}{r}
2\mathrm{R}_3 + \mathrm{R}_4 \rightarrow \mathrm{R}_4 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
-1 & -1 & 3 & -4 & 2 \\
0 & 0 & 2 & -2 & 2 \\
0 & 0 & 0 & -1 & 2 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\tag{1.20}
$$

However, if we perform backward elimination on the matrix in (1.20) we end up with the same matrix in (1.17). This illustrates the uniqueness of the reduced row echelon form.

Observe that both row echelon forms in (1.16) and (1.20) as well as the reduced row echelon form in (1.17) have the same row rank, $r = 3$, which is the row rank of the coefficient matrix in (1.15).

Example 1.11

While computing the reduced row echelon form of a matrix by hand, simplifying the matrix as much as possible before selecting the pivot element may help avoid dealing with complex numbers or fractions. Consider the matrix

$$
\begin{bmatrix}
1+i & i & i \\
2+i & 1+i & 1+2i
\end{bmatrix}
$$

A straightforward application of the Gaussian elimination algorithm requires that either $(2+i)/(1+i)$ times the first row be subtracted from the second row or $(1+i)/(2+i)$ times the second row be subtracted from the first row at the first step. Subsequent steps require similar operations with complex numbers. However, much of the complex arithmetic can be avoided at the expense of more row operations as

$$
\begin{bmatrix}
1+i & i & i \\
2+i & 1+i & 1+2i
\end{bmatrix}
\quad
\begin{array}{r}
-\mathrm{R}_1 + \mathrm{R}_2 \rightarrow \mathrm{R}_2 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
1+i & i & i \\
1 & 1 & 1+i
\end{bmatrix}
$$

$$
\begin{array}{r}
\mathrm{R}_1 \leftrightarrow \mathrm{R}_2 \\
-\mathrm{R}_1 + \mathrm{R}_2 \rightarrow \mathrm{R}_2 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
1 & 1 & 1+i \\
i & -1+i & -1
\end{bmatrix}
$$

$$
\begin{array}{r}
-i\mathrm{R}_1 + \mathrm{R}_2 \rightarrow \mathrm{R}_2 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
1 & 1 & 1+i \\
0 & -1 & -i
\end{bmatrix}
$$

$$
\begin{array}{r}
-\mathrm{R}_2 \leftrightarrow \mathrm{R}_2 \\
\mathrm{R}_1 - \mathrm{R}_2 \rightarrow \mathrm{R}_1 \\
\longrightarrow
\end{array}
\quad
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & i
\end{bmatrix}
$$

* 1.5 Solution Properties of Linear Equations

Suppose that the $m \times n$ system in (1.3) is transformed into

$$R\mathbf{x} = \mathbf{d} \tag{1.21}$$

by elementary row operations, where R is the reduced row echelon form of A. Let $r(A) = r(R) = r$.

We first consider the general case where $r < \min\{m, n\}$.

Partitioning the rows of R into two groups consisting of the nonzero and zero rows, (1.21) can be written as

$$\begin{bmatrix} F \\ O \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{1.22}$$

where F is an $r \times n$ matrix consisting of the nonzero rows of R, O is the $(m-r) \times n$ zero matrix, and $\mathbf{p}$ and $\mathbf{q}$ are the corresponding blocks of $\mathbf{d}$.

Observe that if $\tilde{\mathbf{x}}$ is obtained from $\mathbf{x}$ by reordering its components and $\tilde{F}$ is obtained from F by the same reordering of its columns, then (1.22) is equivalent to

$$\begin{bmatrix} \tilde{F} \\ O \end{bmatrix} \tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{1.23}$$

In particular, let the reordering of the components of $\mathbf{x}$ be such that the basic variables (corresponding to the basic columns of F) occupy the first r positions, and the non-basic variables occupy the last $n - r$ positions. That is,

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$$

where $\mathbf{u}$ contains the basic variables, and $\mathbf{v}$ contains the non-basic variables. Since the jth basic column of F contains all 0's except a 1 in the jth position, the basic columns of F make up the matrix I_r. Hence, $\tilde{F}$ is of the form

$$\tilde{F} = [\, I_r \;\; H \,]$$

where H consists of the non-basic columns of F. With $\tilde{F}$ and $\tilde{\mathbf{x}}$ partitioned this way, (1.23) can be written as

$$\begin{bmatrix} I & H \\ O & O \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{1.24}$$

Referring to the system in Example 1.9, (1.22) corresponds to

$$\left[\begin{array}{ccccc} 1 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \left[\begin{array}{c} -1 \\ 2 \\ 1 \\ \hline q-4 \end{array}\right] \tag{1.25}$$

which can be written down from the augmented matrix in (1.17). The basic variables are x_1, x_3 and x_4, and the non-basic variables are x_2 and x_5. Reordering the variables so that the basic variables appear before the non-basic variables, and performing the same reordering on the columns of the coefficient matrix, (1.25) becomes

$$\left[\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_3 \\ x_4 \\ \hline x_2 \\ x_5 \end{array}\right] = \left[\begin{array}{c} -1 \\ 2 \\ 1 \\ \hline q-4 \end{array}\right] \tag{1.26}$$

which is in the form of (1.24).

From (1.24) we immediately observe that if $\mathbf{q} \neq \mathbf{0}$, then the system (1.21), and therefore, the system (1.3) are inconsistent. This happens if and only if the row rank of the augmented matrix $[\,A \;\; \mathbf{b}\,]$ is larger than the row rank of the coefficient matrix A (see Exercise 1.34). Thus we obtain our first result: If $r[\,A \;\; \mathbf{b}\,] > r$, then the system (1.3) is inconsistent.

On the other hand, if $\mathbf{q} = \mathbf{0}$, that is, if $r[\,A \;\; \mathbf{b}\,] = r$, then discarding the last $m-r$ trivial equations, (1.24) is reduced to

$$[\,I \;\; H\,]\left[\begin{array}{c} \mathbf{u} \\ \mathbf{v} \end{array}\right] = \mathbf{u} + H\mathbf{v} = \mathbf{p} \tag{1.27}$$

which can be rewritten as

$$\mathbf{u} = \mathbf{p} - H\mathbf{v} = \mathbf{p} - v_1\mathbf{h}_1 - \cdots - v_\nu\mathbf{h}_\nu \tag{1.28}$$

where $\nu = n - r$ and $\mathbf{h}_i$ are the columns of H. Equation (1.28) specifies the basic variables in terms of the non-basic variables, which can be chosen arbitrarily. Letting $v_1 = c_1, \ldots, v_\nu = c_\nu$, where $c_1, \ldots, c_\nu$ are arbitrary constants, we express $\mathbf{v}$ as

$$\mathbf{v} = \left[\begin{array}{c} c_1 \\ \vdots \\ c_\nu \end{array}\right] = c_1\mathbf{e}_1 + \cdots + c_\nu\mathbf{e}_\nu$$

where $\mathbf{e}_i$ is the ith column of I_ν. Then $\mathbf{u}$ is obtained from (1.28) as

$$\mathbf{u} = \mathbf{p} + c_1(-\mathbf{h}_1) + \cdots + c_\nu(-\mathbf{h}_\nu)$$

Hence we obtain the following general expression for the solution in terms of the renamed variables.

$$\left[\begin{array}{c} \mathbf{u} \\ \mathbf{v} \end{array}\right] = \left[\begin{array}{c} \mathbf{p} \\ \mathbf{0} \end{array}\right] + c_1\left[\begin{array}{r} -\mathbf{h}_1 \\ \mathbf{e}_1 \end{array}\right] + \cdots + c_\nu\left[\begin{array}{r} -\mathbf{h}_\nu \\ \mathbf{e}_\nu \end{array}\right] \tag{1.29}$$

or in more compact form as

$$\tilde{\mathbf{x}} = \tilde{\boldsymbol{\phi}}_p + c_1\tilde{\boldsymbol{\phi}}_1 + \cdots + c_\nu\tilde{\boldsymbol{\phi}}_\nu = \tilde{\boldsymbol{\phi}}_p + \tilde{\boldsymbol{\phi}}_c \tag{1.30}$$

where

$$\tilde{\boldsymbol{\phi}}_p = \left[\begin{array}{c} \mathbf{p} \\ \mathbf{0} \end{array}\right] \quad \text{and} \quad \tilde{\boldsymbol{\phi}}_i = \left[\begin{array}{r} -\mathbf{h}_i \\ \mathbf{e}_i \end{array}\right], \quad i = 1, \ldots, \nu$$

Reversing the reordering and using the original names for the variables, the solution above is expressed as

$$\mathbf{x} = \boldsymbol{\phi}_p + c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu = \boldsymbol{\phi}_p + \boldsymbol{\phi}_c \tag{1.31}$$

where $c_1, \ldots, c_\nu$ are arbitrary constants.

To illustrate expressions (1.30) and (1.31) we refer to the system in Example 1.9 again, which has been simplified to (1.26). We first observe that the system is consistent if and only if $\mathbf{q} = q - 4 = 0$, in which case $r[\,A \;\; \mathbf{b}\,] = r(A) = 3$. Assuming so and discarding the last equation, we rewrite (1.26) in the form of (1.28) as

$$\begin{bmatrix} x_1 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} - x_2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - x_5 \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix} \tag{1.32}$$

Now letting $x_2 = c_1, x_5 = c_2$, (1.32) gives the solution in the form of (1.30) as

$$\left[\begin{array}{c} x_1 \\ x_3 \\ x_4 \\ \hline x_2 \\ x_5 \end{array}\right] = \left[\begin{array}{c} -1 \\ 2 \\ 1 \\ \hline 0 \\ 0 \end{array}\right] + c_1 \left[\begin{array}{c} -1 \\ 0 \\ 0 \\ \hline 1 \\ 0 \end{array}\right] + c_1 \left[\begin{array}{c} -3 \\ 1 \\ 2 \\ \hline 0 \\ 1 \end{array}\right] \tag{1.33}$$

where

$$\tilde{\boldsymbol{\phi}}_p = \left[\begin{array}{c} -1 \\ 2 \\ 1 \\ \hline 0 \\ 0 \end{array}\right], \quad \tilde{\boldsymbol{\phi}}_1 = \left[\begin{array}{c} -1 \\ 0 \\ 0 \\ \hline 1 \\ 0 \end{array}\right], \quad \tilde{\boldsymbol{\phi}}_2 = \left[\begin{array}{c} -3 \\ 1 \\ 2 \\ \hline 0 \\ 1 \end{array}\right] \tag{1.34}$$

Finally restoring the original order of the variables we get the solution in (1.19), which has the form of (1.31).

Having studied the general case, we now consider other possible cases.

If $r = m < n$, then (1.24) is already in the form of (1.27) with $r[\,A \;\; \mathbf{b}\,] = r(A) = m$. Thus the system is consistent, and the solution is given by (1.31).

If $r = n < m$, then (1.24) is reduced to

$$\begin{bmatrix} I \\ O \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \tag{1.35}$$

Again, the system is consistent if and only if $\mathbf{q} = \mathbf{0}$, or equivalently, if and only if $r[\,A \;\; \mathbf{b}\,] = r(A) = n$. In this case, (1.35) is further reduced to

$$\mathbf{x} = \mathbf{p} \tag{1.36}$$

which displays a unique solution.

Finally, if $r = m = n$, then the system is consistent, and has a unique solution given in (1.36).

The solution expression in (1.31) consists of two parts: The first part, $\boldsymbol{\phi}_p$, is fixed and is due to the right-hand side of the system. If $\mathbf{b} = \mathbf{0}$, that is, if the system is homogeneous, then $\mathbf{p} = \mathbf{0}$, and therefore, $\boldsymbol{\phi}_p = \tilde{\boldsymbol{\phi}}_p = \mathbf{0}$. The second part, $\boldsymbol{\phi}_c$, exists if and only if $r < n$. If $r = n$, then $\boldsymbol{\phi}_c = \mathbf{0}$, and $\mathbf{x} = \boldsymbol{\phi}_p$ would be the unique solution. (As a consequence, if the system is homogeneous and if $r = n$, then the only solution is the null solution $\mathbf{x} = \mathbf{0}$.)

Since $\boldsymbol{\phi}_p = \mathbf{0}$ when $\mathbf{b} = \mathbf{0}$, we conclude that when $r < n$

$$\boldsymbol{\phi}_c = c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu$$

is a nontrivial solution of the associated homogeneous system

$$A\mathbf{x} = \mathbf{0} \tag{1.37}$$

Since it contains ν arbitrary constants, $\boldsymbol{\phi}_c$ defines a family of solutions of (1.37). For every fixed choice of the arbitrary constants we get a member of this family. In particular, each $\boldsymbol{\phi}_i$ is a member of this family, $\boldsymbol{\phi}_1$ corresponding to the choice $c_1 = 1, c_2 = \cdots = c_\nu = 0$, $\boldsymbol{\phi}_2$ corresponding to the choice $c_1 = 0, c_2 = 1, c_3 = \cdots = c_\nu = 0$, etc. These solutions have the property that none of them can be expressed in terms of the others, and are said to be **linearly independent**. This follows from the fact that each $\mathbf{e}_i$ contains a single 1 at a different position, so that no $\mathbf{e}_i$ can be expressed in terms of the others. Then the same must be true for $\tilde{\boldsymbol{\phi}}_i$, and therefore, for $\boldsymbol{\phi}_i$. The reader should examine the expressions for $\tilde{\boldsymbol{\phi}}_1$ and $\tilde{\boldsymbol{\phi}}_2$ in (1.34) to verify this fact. The significance of the solutions $\boldsymbol{\phi}_i$ being linearly independent is that they can not be combined to simplify $\boldsymbol{\phi}_c$, which implies that for every choice of the arbitrary constants $c_1, \ldots, c_\nu$ we get a different solution of the homogeneous equation (see Exercise 1.39).[11]

Like $\boldsymbol{\phi}_c$, the expression in (1.31) also defines a family of solutions of the non-homogeneous system in (1.3). Any member of this family, obtained by assigning fixed values to the arbitrary constants $c_1, \ldots, c_\nu$ is called a **particular solution**. A simple particular solution is obtained by choosing $c_1 = \cdots = c_\nu = 0$ to be

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

The second part of the expression in (1.31), $\boldsymbol{\phi}_c$, is called a **complementary solution** of (1.3), because by adding to $\boldsymbol{\phi}_p$ any member of the family defined by $\boldsymbol{\phi}_c$, we obtain another particular solution of (1.3).

Finally, we note that since any solution of (1.3) must satisfy (1.28), it must be of the form in (1.31). In other words, the family characterized by the expression in (1.31) contains all solutions of (1.3). For this reason, we call this expression a **general solution** of (1.3).

---

[11] The concept of linear independence will be discussed in Chapter 3. For the time being it suffices to know that if $\boldsymbol{\phi}_i$ were not linearly independent, then one of them, say the kth one, would be expressed in terms of the others as

$$\boldsymbol{\phi}_k = \sum_{i \neq k} \alpha_i \boldsymbol{\phi}_i$$

Then $\boldsymbol{\phi}_c$ would reduce to

$$\boldsymbol{\phi} = \sum_{i \neq k} (c_i + c_k\alpha_i)\boldsymbol{\phi}_i = \sum_{i \neq k} c_i' \boldsymbol{\phi}_i$$

containing only $\nu - 1$ arbitrary constants, and one degree of freedom would be lost.

Thus we not only showed that the system is consistent if $r[\,A \;\; \mathbf{b}\,] = r(A)$, but also gave a systematic procedure to find a general form of the solution. We summarize these results as a theorem.

Theorem 1.1 *Let A be an $m \times n$ matrix with $r(A) = r$.*

a) The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is consistent, and

i. if $r = n$, then the only solution is the trivial solution $\mathbf{x} = \mathbf{0}$,

ii. if $r < n$, then there exist $\nu = n - r$ linearly independent solutions $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu$, and $\mathbf{x} = c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu$ is a solution for every choice of the arbitrary constants $c_1, \ldots, c_\nu \in \mathbf{R}$.

b) The non-homogeneous system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $r[\,A \;\; \mathbf{b}\,] = r$, in which case

i. if $r = n$, then there exists a unique solution $\mathbf{x} = \boldsymbol{\phi}_p$,

ii. if $r < n$, then $\mathbf{x} = \boldsymbol{\phi}_p + c_1\boldsymbol{\phi}_1 + \ldots + c_\nu\boldsymbol{\phi}_\nu$ is a solution for every choice of the arbitrary constants $c_1, \ldots, c_\nu \in \mathbf{R}$, where $\nu = n - r$, $\boldsymbol{\phi}_p$ is any particular solution, and $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu$ are the linearly independent solutions of the associated homogeneous system.

The analysis above shows that existence and uniqueness of solution of a given system of linear equations cannot be deduced from the number of equations (m) and the number of unknowns (n) alone. Indeed, Example 1.9 illustrates that if a system contains more unknowns than equations that does not necessarily imply that the system has a solution. The converse is not true either: A system that contains more equations than unknowns may still have a solution as we illustrate by the following example.

Example 1.12

Check if the system

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & 3x_3 & = & 11 \\ -x_1 & + & 3x_2 & - & 2x_3 & = & -11 \\ 2x_1 & - & 3x_2 & + & 5x_3 & = & 18 \\ -x_1 & + & x_2 & - & 2x_3 & = & -7 \end{array}$$

is consistent, and if so, find the solution.

We form the augmented matrix and simplify it as described below.

$$\left[\begin{array}{rrr|r} 1 & -2 & 3 & 11 \\ -1 & 3 & -2 & -11 \\ 2 & -3 & 5 & 18 \\ -1 & 1 & -2 & -7 \end{array}\right] \quad \begin{array}{r} \mathrm{R}_1 + \mathrm{R}_2 \to \mathrm{R}_2 \\ -2\mathrm{R}_1 + \mathrm{R}_3 \to \mathrm{R}_3 \\ \mathrm{R}_1 + \mathrm{R}_4 \to \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrr|r} 1 & -2 & 3 & 11 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & -4 \\ 0 & -1 & 1 & 4 \end{array}\right]$$

$$\begin{array}{r} 2\mathrm{R}_2 + \mathrm{R}_1 \to \mathrm{R}_1 \\ -\mathrm{R}_2 + \mathrm{R}_3 \to \mathrm{R}_3 \\ \mathrm{R}_2 + \mathrm{R}_4 \to \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrr|r} 1 & 0 & 5 & 11 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -2 & -4 \\ 0 & 0 & 2 & 4 \end{array}\right]$$

$$\begin{array}{r} -(1/2)\mathrm{R}_3 \to \mathrm{R}_3 \\ -5\mathrm{R}_3 + \mathrm{R}_1 \to \mathrm{R}_1 \\ -\mathrm{R}_3 + \mathrm{R}_2 \to \mathrm{R}_2 \\ -\mathrm{R}_3 + \mathrm{R}_4 \to \mathrm{R}_4 \\ \longrightarrow \end{array} \quad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

From the last augmented matrix we find that the given system is consistent, and has a unique solution

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

Although a system that contains more equations than unknowns may be consistent, the reader should not expect to come across such systems often. Indeed, unless the elements of A and $\mathbf{b}$ are somehow related, such a system will almost always be inconsistent (see Exercise 1.38). However, certain problems lead to systems with more equations than unknowns, which, by the nature of the problem, are consistent (see Exercise 1.46).

1.6 Numerical Considerations

Computers use limited space to represent numbers, that is, they have finite precision. This may lead to numerical errors in evaluating expressions that involve operations with several numbers. Anyone who tries to evaluate the expression

$$(1/3) \cdot 3 - 1$$

with a hand calculator would probably get an answer like 1×10^{-6} or 1×10^{-12} depending on the precision of the calculator, instead of the exact answer 0. The reason is that the number $1/3$ cannot be represented exactly by the calculator (no matter how many digits are used), so that when it is multiplied with 3 the result will be slightly different from 1, hence the difference from 1 slightly different from zero. If the result of these operations is later used in other expressions, the errors accumulate, possibly reaching unacceptable levels.

Since operations with matrices involve many successive operations with scalars (as in the case of Gaussian elimination), one should be alert about numerical errors that might result from a sequence of such operations, and try to avoid them as much as possible.

Example 1.13

Consider the linear system

$$\begin{array}{rcrcl} -0.001\,x_1 & + & 1.000\,x_2 & = & 1.000 \\ 1.000\,x_1 & + & 1.000\,x_2 & = & 2.000 \end{array}$$

Eliminating x_1 from the second equation we obtain

$$\begin{array}{rcrcl} -0.001\,x_1 & + & 1.000\,x_2 & = & 1.000 \\ & & 1001.\,x_2 & = & 1002. \end{array}$$

from which we obtain the exact solution

$$x_2 = \frac{1002}{1001} = 1 + \frac{1}{1001}\,, \quad x_1 = \frac{1000}{1001} = 1 - \frac{1}{1001}$$

Now suppose we try to solve the same system using a calculator that can represent numbers with three significant digits only. If we proceed with eliminating x_1 as above, we obtain the system

$$\begin{array}{rcrcl} -0.001\,x_1 & + & 1.000\,x_2 & = & 1.000 \\ & & 1000.\,x_2 & = & 1000. \end{array}$$

because both 1001 and 1002 are represented as 1000. in our calculator. We then obtain

$$x_2 = 1.000\,, \quad x_1 = 0.000$$

which is far from being a solution.

What happened is that the information in the second equation was lost when the first equation is multiplied by 1000 and added to the second. That is why the "computed" solution above satisfies the first equation but not the second.

Fortunately, however, the problem can be overcome simply by interchanging the equations. Gaussian elimination applied to the reordered system

$$\begin{array}{rcrcl} 1.000\,x_1 & + & 1.000\,x_2 & = & 2.000 \\ -0.001\,x_1 & + & 1.000\,x_2 & = & 1.000 \end{array}$$

produces

$$\begin{array}{rcrcl} 1.000\,x_1 & + & 1.000\,x_2 & = & 2.000 \\ & & 1.000\,x_2 & = & 1.000 \end{array}$$

in the calculator, from which the solution is computed as

$$x_2 = 1.000\,, \quad x_1 = 1.000$$

The "computed" solution is acceptable now.

The situation in the above example is similar in nature to computation of

$$(1/3) \cdot 3 - 1$$

Just like rephrasing this expression as

$$(1 \cdot 3)/3 - 1$$

eliminates the error, reordering the equations before applying Gaussian elimination reduces error to an acceptable level. The error in the first attempt originates from choosing a very small pivot at the first step, which results in a very large multiplier that erases the information in the second equation. Reordering the equations leads to a choice of a larger pivot that does not cause a significant loss of information. This strategy (of picking as large a pivot as possible among the candidates) is called **partial pivoting**.

Although pivoting can handle many difficult situations, there are problems that are inherently "bad", and error is unavoidable.

Example 1.14

The system

$$\begin{bmatrix} 0.9900 & 0.9800 \\ 0.9800 & 0.9700 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.970 \\ 1.950 \end{bmatrix} \tag{1.38}$$

has the exact solution

$$\mathbf{x}_e = \begin{bmatrix} 1.000 \\ 1.000 \end{bmatrix}$$

Gaussian elimination with four-digit floating-point arithmetic reduces the augmented matrix to

$$\left[\begin{array}{rr|r} 0.9900 & 0.9800 & 1.970 \\ 0.0000 & -0.0001 & .0000 \end{array}\right]$$

The reduced system has the solution

$$\mathbf{x}_c = \begin{bmatrix} 1.990 \\ .0000 \end{bmatrix}$$

which is nowhere near the exact solution. Interchanging the equations as in the previous example is of no use; we end up with a similar erroneous result.

Suppose we did not know the exact solution, and wanted to check the "computed" solution $\mathbf{x}_c$ by substituting it into the original system. We would get

$$\begin{bmatrix} 0.9900 & 0.9800 \\ 0.9800 & 0.9700 \end{bmatrix} \begin{bmatrix} 1.990 \\ .0000 \end{bmatrix} = \begin{bmatrix} 1.970 \\ 1.950 \end{bmatrix}$$

which is the same as the right-hand side of (1.38) up to the fourth significant digits. What is more interesting is that when we evaluate $A\mathbf{x}$ for

$$\mathbf{x}_1 = \begin{bmatrix} 6.910 \\ -4.970 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -4.870 \\ 6.930 \end{bmatrix}$$

which are totally unrelated to each other and to the exact solution, we get

$$A\mathbf{x}_1 = \begin{bmatrix} 1.970 \\ 1.949 \end{bmatrix} \quad \text{and} \quad A\mathbf{x}_2 = \begin{bmatrix} 1.970 \\ 1.951 \end{bmatrix}$$

which differ from the right-hand side of (1.38) only in the fourth significant digit. Apparently, our check is not reliable.

The numerical difficulties encountered in the solution of the system in (1.38) stem from the fact that the lines described by the equations of the system are almost parallel. Although they intersect at a point whose coordinates are specified by the exact solution $\mathbf{x}_e$, the points with coordinates defined by $\mathbf{x}_c$, $\mathbf{x}_1$ and $\mathbf{x}_2$ are not far from these lines either. Unfortunately, the problem is inherent in the system, and no cure (other than increasing the precision) is available. Such systems are said to be **ill-conditioned**.

1.7 Exercises

1. Study the tutorial in Appendix D. Experiment with MATLAB, and learn
   (a) how to input a real and a complex matrix,
   (b) basic matrix operations (addition, multiplication, transposition),
   (c) how to create special matrices (identity matrix, zero matrix, diagonal matrix, etc.),
   (d) how to construct a partitioned matrix from given blocks and how to extract a submatrix from a given matrix,
   (e) how to create and execute an M-file.

2. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

   (a) Compute $\mathbf{x}^t\mathbf{x}$, $\mathbf{y}^t\mathbf{y}$, $\mathbf{x}\mathbf{x}^t$, $\mathbf{y}\mathbf{y}^t$, $\mathbf{x}^t\mathbf{y}$, $\mathbf{y}^t\mathbf{x}$, $\mathbf{x}\mathbf{y}^t$, $\mathbf{y}\mathbf{x}^t$, $A\mathbf{x}$, $\mathbf{x}^tA$, $A\mathbf{y}$, $\mathbf{y}^tA$, $\mathbf{x}^tA\mathbf{x}$, $\mathbf{x}^tA\mathbf{y}$, $\mathbf{y}^tA\mathbf{x}$ and $\mathbf{y}^tA\mathbf{y}$.

   (b) Use MATLAB to find the products in part (a).

   (c) Compute $\mathrm{tr}\,(A), \mathrm{tr}\,(\mathbf{x}\mathbf{x}^t), \mathrm{tr}\,(\mathbf{x}\mathbf{y}^t), \mathrm{tr}\,(\mathbf{y}\mathbf{x}^t), \mathrm{tr}\,(\mathbf{y}\mathbf{y}^t)$.

3. Show, by an example, that $AB = AC$ does not imply that $B = C$.

4. (a) Show that if $A \in \mathsf{F}^{m\times n}$ and $B \in \mathsf{F}^{n\times m}$ then $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$.

   (b) Show that if $\mathbf{x}, \mathbf{y} \in \mathsf{F}^{n\times 1}$ then

$$\mathrm{tr}\,(\mathbf{x}\mathbf{y}^t) = \sum_{i=1}^{n} x_i y_i$$

5. (a) Show that $A^pA^q = A^{p+q} = A^qA^p$.

   (b) Use MATLAB to verify the result in part (a) for

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

   and several p and q.

6. Let $A, B \in \mathsf{R}^{n\times n}$. Determine under what condition

$$(A+B)^2 = A^2 + 2AB + B^2$$

7. Show that if $A = \mathrm{diag}\,[\,d_1, \ldots, d_n\,]$, then $A^k = \mathrm{diag}\,[\,d_1^k, \ldots, d_n^k\,]$.

8. (a) Prove that the product of two lower (upper) triangular matrices is also lower (upper) triangular.

   (b) Verify the result in part (a) by computing the product of two arbitrarily chosen 3×3 upper triangular matrices using MATLAB .

9. If A, B and C are 100×2, 2×100 and 100×10 matrices, would you compute the product ABC as $(AB)C$ or as $A(BC)$? Why?

10. Show that the matrices

$$A = \begin{bmatrix} I_n & O \\ C & I_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} I_n & O \\ D & I_n \end{bmatrix}$$

    commute.

11. Find a general expression for A^n for the A matrix in Exercise 1.10.

12. Let

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

    (a) Obtain a general formula for A^n.

    (b) Verify your formula by calculating A^n for $n = 2, 3, 4, 5$ using MATLAB.

13. Let

$$A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

   (a) Use MATLAB to compute A^5 and A^{10}. Would you expect A^n to blow up or converge to a finite limit matrix as $n \to \infty$?

   (b) Use MATLAB to compute A^n for $n = 1, 2, \ldots$ until the maximum element in absolute value of $A^n - A^{n-1}$ is smaller than a sufficiently small number, say 10^{-6}.

14. Let $A = B + iC$, where $B, C \in \mathbf{R}^{n \times n}$. Show that A is Hermitian if and only if B is symmetric and C is skew-symmetric. State and prove a corresponding result for A to be skew-Hermitian.

15. What can you say about the diagonal elements of a Hermitian and a skew-Hermitian matrix?

16. Show that $A^h A$ is a Hermitian matrix for any $A \in \mathbf{C}^{m \times n}$. State and prove a corresponding result for $A \in \mathbf{R}^{m \times n}$.

17. Let $\mathbf{e}_i$ denote the ith column of I_n.

   (a) Interpret the products

$$\mathbf{e}_i^t \mathbf{e}_i\,, \quad \mathbf{e}_i^t \mathbf{e}_j\,, \quad \mathbf{e}_i \mathbf{e}_j^t\,, \quad \mathbf{e}_i^t A\,, \quad A\mathbf{e}_j\,, \quad \mathbf{e}_i^t A \mathbf{e}_j$$

   (b) Use MATLAB to verify your interpretation by calculating the above products for the A matrix in Exercise 1.2 and for $i, j = 1, 2, 3$.

18. Let

$$Q_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n \times n}$$

   (a) Obtain a general expression for $Q_n^k,\ k = 1, \ldots, m \geq n$.

   (b) Use MATLAB to verify your result for $n = 4$ and $k = 1, \ldots, 5$.

19. Let $A \in \mathbf{R}^{m \times n}$, $B \in \mathbf{R}^{n \times m}$ and Q_n be as in Exercise 1.14.

   (a) Express AQ_n^k in terms of the columns of A.

   (b) Express $Q_n^k B$ in terms of the rows of B.

   (c) Construct the matrices Q_5, $A = [\,10i + j\,]_{3 \times 5}$ and $B = [\,10i + j\,]_{5 \times 4}$ in MATLAB, and calculate AQ_5^k and $Q_5^k B$ for $k = 1, \ldots, 5$.

20. Let A be a 10×4 matrix partitioned into its columns as

$$A = [\,\mathbf{a}_1\ \ \mathbf{a}_2\ \ \mathbf{a}_3\ \ \mathbf{a}_4\,]$$

   and let Q be such that

$$AQ = [\,\mathbf{a}_4\ \ \mathbf{a}_3\ \ \mathbf{a}_1\ \ \mathbf{a}_2\,]$$

   (a) Find Q

   (b) Find AQ^{25} in terms of the columns of A

21. For each of the following $(A, \mathbf{b})$ pairs find the reduced row echelon form of the augmented matrix by hand, and check your result using the MATLAB command `rref`. Using the reduced row echelon form of the augmented matrix determine if the system $A\mathbf{x} = \mathbf{b}$ is consistent, and if so, find the general solution.

(a) $$A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & -1 & -6 \\ 5 & -3 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

(b) $$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -4 \\ 2 & 4 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

(c) $$A = \begin{bmatrix} 5 & -6 & 1 \\ 2 & -3 & 1 \\ 4 & -3 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}$$

(d) $$A = \begin{bmatrix} 1 & 0 & 2 & -1 \\ 1 & 4 & 2 & 7 \\ 2 & -2 & 4 & -6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$$

(e) $$A = \begin{bmatrix} 2 & -4 & -3 & -4 \\ -1 & 2 & 2 & 3 \\ 1 & -2 & -1 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

(f) $$A = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & 1 & -3 & -1 \\ 4 & -1 & 1 & 1 \\ 1 & 2 & -5 & -2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

(g) $$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

(h) $$A = \begin{bmatrix} 1+i & 0 & -i & 1 \\ 0 & 1 & 0 & -1 \\ -i & -1+i & 1 & 1 \\ 1 & i & 1-i & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ i \end{bmatrix}$$

22. Attempt to solve the linear systems in Exercise 1.21 by using the MATLAB command `x=A\b`, and interpret the results.

23. Find the value of the scalar p such that the system

$$\begin{bmatrix} 1 & -2 & 3 \\ 2 & -4 & 7 \\ 1 & -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ p \end{bmatrix}$$

is consistent, and then find the general solution.

24. Repeat Exercise 1.23 for the system

$$\begin{bmatrix} 2 & -1 & 1 \\ -1 & -1 & -6 \\ 5 & -3 & 1 \\ 1 & -2 & p \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

25. For the following pair, determine all values of the constants p and q such that the system $A\mathbf{x} = \mathbf{b}$ has (a) no solution, (b) infinitely many solutions, (c) a unique solution.

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 4 & -3 & 2 \\ -2 & -1 & p \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ q \end{bmatrix}$$

26. Determine geometrically if the following systems are consistent, and if so, find their solutions.

(a)
$$\begin{array}{rcrcr} x_1 & - & x_2 & = & 1 \\ x_1 & + & x_2 & = & 5 \\ x_1 & - & 3x_2 & = & 3 \end{array}$$

(b)
$$\begin{array}{rcrcr} x_1 & - & x_2 & = & 1 \\ x_1 & + & x_2 & = & 5 \\ x_1 & - & 3x_2 & = & -3 \end{array}$$

27. Write the equation of a straight line passing through the points $(x_1, y_1) \neq (x_2, y_2)$ in the xy plane. Hint: A straight line in the xy plane is described by an equation of the form $ax + by + c = 0$ in the most general case. Translate the problem into solving a system of two linear equations in the unknowns a, b and c.

28. Consider a square pyramid with base vertices at $\mathbf{v}_1 = (0,0,0)$, $\mathbf{v}_2 = (2,0,0)$, $\mathbf{v}_3 = (2,2,0)$ and $\mathbf{v}_4 = (0,2,0)$, and the tip at $\mathbf{v}_0 = (1,1,1)$ in the $x_1x_2x_3$ space. Write equations of the four planar faces of the pyramid and obtain the solution of the 4×3 system consisting of these equations. Is the answer what you expect?

29. Write a MATLAB program to implement the Gaussian Elimination Algorithm in Table 1.1, and save it for your future use. Use your program to obtain the row echelon and reduced row echelon forms of the matrices in Exercise 1.21.

30. Find the reduced row echelon form of the augmented matrix in (1.16) if $q \neq 4$.

31. Let the $m \times n$ matrix A have rank $r < m$ and the reduced row echelon form

$$\begin{bmatrix} R \\ O \end{bmatrix}$$

where R is $r \times n$ and O is $(m - r) \times n$. Find the rank and the reduced row echelon form of

$$B = \begin{bmatrix} A \\ A \end{bmatrix}$$

32. (a) Define **elementary column operations** on a matrix by imitating the definition of elementary row operations.

    (b) Give a precise definition of **column equivalence** of matrices.

    (c) Define **column echelon form** and **reduced column echelon form** of a matrix.

    (d) Define column rank of a matrix.

33. (a) Explain how the Gaussian Elimination algorithm of Section 1.4 can be modified to obtain the reduced column echelon form of a matrix.

    (b) Show that the reduced column echelon form of a matrix A is the transpose of the reduced row echelon form of A^t.

34. Let $r(A) = r$ and

$$[\,A\ \ \mathbf{b}\,] \longrightarrow [\,R\ \ \mathbf{d}\,] = \begin{bmatrix} F & \mathbf{p} \\ O & \mathbf{q} \end{bmatrix}$$

where R is the reduced row echelon form of A. Clearly, $r[\,A\ \ \mathbf{b}\,] = r$ if and only if $\mathbf{q} = \mathbf{0}$. Find the reduced row echelon form of the augmented matrix and its rank when $r[\,A\ \ \mathbf{b}\,] \neq r$.

35. (a) Construct the matrices

```
A=eye(3)+ones(3,3);b=[2;0;2]
```

in MATLAB, and solve the equation $A\mathbf{x} = \mathbf{b}$ by using your Gaussian elimination algorithm and also by the MATLAB command `x=A\b`.

(b) Repeat (a) for

```
A=eye(3)+i*ones(3,3);b=[0;-2;-1+i]
```

36. Calculate the total number of multiplication/division operations required to solve an $n \times n$ system by Gaussian elimination, assuming that a pivot can be chosen at every step. Include in your calculation divisions by 1, but exclude multiplications with 0. Hint: Consider the loop consisting of steps 7-10 of the Gaussian elimination algorithm applied to the augmented matrix. Elimination of x_r from each of the remaining $n - r$ equations requires 1 division to find the multiplier μ_{ir} at Step 8, and $n - r + 1$ multiplications to modify the ith row at Step 9. Thus the loop requires $(n-r)(n-r+2)$ multiplications/divisions, and the whole forward elimination process requires

$$\sum_{r=1}^{n-1} (n-r)(n-r+2)$$

such operations. Calculate the operations required by backward elimination similarly, and then find closed form expressions for the sums.

37. Repeat Exercise 1.36 if forward and backward substitutions are performed simultaneously, and explain why forward elimination followed by backward elimination is more efficient (in terms of the number of multiplication/division operations) than simultaneous elimination.

38. Use MATLAB command `M=rand(5,4)` to generate a 5×4 augmented matrix $M = [\,A\ \ \mathbf{b}\,]$ with random elements. Use either the MATLAB code written in Exercise 1.29 or MATLAB's build-in function `rref` to compute the reduced row echelon form of M, and determine if the associated system $A\mathbf{x} = \mathbf{b}$ is consistent. Repeat several times.

39. Let $\{\,\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu\,\}$ be given set of column vectors. Show that if

$$a_1\boldsymbol{\phi}_1 + \cdots + a_\nu\boldsymbol{\phi}_\nu = b_1\boldsymbol{\phi}_1 + \cdots + b_\nu\boldsymbol{\phi}_\nu$$

for two different ordered sets of scalars $(a_1, \ldots, a_\nu)$ and $(b_1, \ldots, b_\nu)$, then at least one of $\boldsymbol{\phi}_i$ can be expressed in terms of the others. This shows that if $\{\,\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\nu\,\}$ is linearly independent then different choices of the arbitrary constants $c_1, \ldots, c_\nu$ in the expression

$$\boldsymbol{\phi}_c = c_1\boldsymbol{\phi}_1 + \cdots + c_\nu\boldsymbol{\phi}_\nu$$

yield different vectors.

40. Show that if a linear system has two distinct solutions then it has infinitely many solutions. Hint: Let $\mathbf{x} = \boldsymbol{\phi}_1$ and $\mathbf{x} = \boldsymbol{\phi}_2$ be two distinct solutions of $A\mathbf{x} = \mathbf{b}$, and consider $A(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)$.

41. Suppose that

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

is the general solution of $A\mathbf{x} = \mathbf{b}$ and that $\mathbf{x} = \boldsymbol{\psi}$ is a particular solution of $A\mathbf{x} = \mathbf{c}$. Find the general solution of

$$A\mathbf{x} = 2\mathbf{b} - \mathbf{c}$$

42. Consider the system

$$\begin{bmatrix} 0.19 & 0.18 \\ 0.18 & 0.17 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.74 \\ 0.70 \end{bmatrix}$$

(a) Find the exact solution.

(b) Show that Gaussian Elimination with 3-digit floating point arithmetic results in an inconsistent system.

(c) Solve the system by using 4-digit floating point arithmetic.

(d) Solve the system by using 5-digit floating point arithmetic.

43. Repeat Exercise 1.42 for the system

$$\begin{bmatrix} 0.820 & 0.528 \\ 0.730 & 0.470 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.340 \\ 0.300 \end{bmatrix}$$

44. Consider the system

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 1 \\ x_1 & + & x_2 & + & \epsilon x_3 & = & -1 \\ x_1 & + & \epsilon x_2 & + & x_3 & = & -1 \end{array}$$

where ϵ is smaller than the precision of a calculator (that is, the calculator can represent ϵ alone, but rounds $1 + \epsilon$ to 1).

(a) Find the exact solution of the system.

(b) Show that, independent of the choice of the pivot elements, Gaussian elimination implemented on the calculator fails to produce a solution.

(c) Rewrite the equations in terms of new variables $z_1 = x_1$, $z_2 = x_1 + x_2$, $z_3 = x_1 + x_3$. Can you solve the resulting system with the same calculator?

45. Find all possible values of s such that the system

$$\begin{bmatrix} s & 1 & 0 \\ 0 & s & 1 \\ 0 & 1 & s \end{bmatrix} \mathbf{x} = \mathbf{0}$$

has a nontrival solution.

46. (Application) Consider the resistive electrical network shown in Figure 1.2(a), where v_1 and v_2 are the voltages supplied by external sources. The problem is to determine the voltages across and currents through all components of the network using Kirchoff's voltage and current laws and the voltage/current relations of the resistors. Kirchoff's voltage law states that the algebraic sum of the voltages across components that form a closed circuit is zero, and the current law states that the algebraic sum of currents through components that form a cut-set (a hypothetical line that separates the network into two disjoint parts) is zero. For convenience in identifying the circuits and cut-sets of the circuit, we associate with it a directed graph as shown in Figure 1.2(b), where

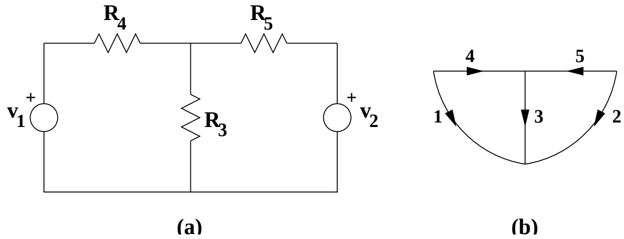

Figure 1.2: A resistive network

each edge corresponds to a component. Direction of the edges are assigned arbitrarily with the convention that if a current flows in the assigned direction then it has a positive value, and that if the voltage across the component drops in the assigned direction then it has a positive value. The voltage/current relation of a resistor is $v = Ri$.

(a) Identify all circuits in the network, and write the circuit equations (relating the voltages of the components in the circuit) using Kirchoff's voltage law. Hint: There are three circuits.

(b) Identify all cut-sets in the network, and write the cut-set equations (relating the currents of the components in the circuit) using Kirchoff's current law. Hint: There are six cut-sets.

(c) Use circuit and cut-set equations together with the voltage/current equations of the resistors to obtain a linear system in which all voltages and currents except v_1 and v_2 appear as unknowns to be solved in terms of v_1 and v_2.

(d) Solve the linear system you obtained above for the specific values $v_1 = 30V$, $v_2 = 60V$, $R_3 = 6K\Omega$, $R_4 = 2K\Omega$, $R_5 = 6K\Omega$. Show that although there are more equations than unknowns, the system is consistent and has a unique solution.

(e) Show that the solution above is independent of the value of R_4.

(f) Apparently not all the circuit and cut-set equations are independent (i.e., some of them can be obtained from the others, and are, therefore, redundant). Show that only two of the three circuit equations and only three of the six cut-set equations are independent. (Thus, together with the three voltage/current equations of the resistors, there are eight equations in eight unknowns.)

47. (Application) The diagram in Figure 1.3 shows the major pipelines of the water distribution network of a town, where q_1 and q_2 denote the supply flow rates (in thousand m^3/sec) into the network from two reservoirs, and q_3 and q_4 the outflow rates from the network to town. It is assumed that no water is stored anywhere in the network, so that $q_1 + q_2 = q_3 + q_4$. The variables f_i associated with each pipe denote the flow rate in the direction arbitrarily assigned to the pipe. (Thus a negative value indicates that the flow is in the reverse direction.)

    (a) Obtain a linear system in the variables f_i by equating the inflow rate at each node to the outflow rate.

    (b) Obtain a general solution of the system obtained in part (a). How many variables can be chosen arbitrarily?

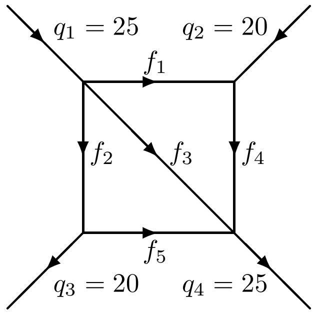

Figure 1.3: Water distribution network

(c) Suppose that the pipes have a limited capacity so that $-F \leq f_i \leq F$, where $F = 18$. Obtain a region in the parameter space of arbitrary parameters that appear in the general solution in which every combination of the parameters gives a solution that satisfy the capacity constraints.

(d) Find F such that the system has a unique solution. Find also the corresponding solution.

48. (Application) Figure 1.4 shows a planar structure consisting of two rigid pieces pinned to each other and to two fixed supports on the ground. Let α_1 and α_2 denote the interior angles of the pieces with the horizontal, and let the weights W_1 and W_2 of the pieces be represented as downward forces acting at their midpoints. The problem is to find the forces on the supports. Let F_{ix} and F_{iy}, $i = 1, 2, 3$, denote the horizontal and vertical components of the reaction forces on the pin joints. Since each rigid piece is in equilibrium, the net horizontal and vertical force as well as the net torque acting on each piece is zero.

    (a) Write three equations for each piece to describe the equilibrium conditions to obtain a linear system of a total of six equations in the six unknowns F_{ix} and F_{iy}, $i = 1, 2, 3$. Hint: Use the geometry of the structure.

    (b) Solve the system formulated in part (a) to find the forces on the supports in terms of W_1, W_2 and α_1, α_2.

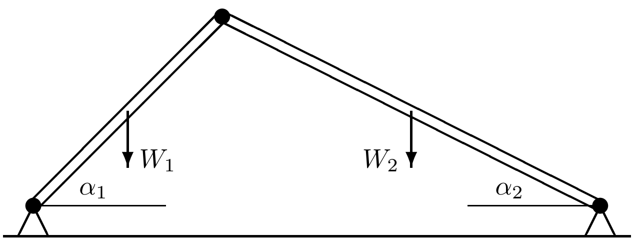

Figure 1.4: Rigid planar structure