# B

# Linear Algebra: Matrices

# TABLE OF CONTENTS

## §B.1. Matrices

This Appendix introduces the concept of a *matrix*. Consider a set of scalar quantities arranged in a rectangular array containing $m$ rows and $n$ columns:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}. \tag{B.1}$$

This will be called a *rectangular matrix* of order $m$ by $n$, or, briefly, an $m \times n$ matrix. Not every rectangular array is a matrix; to qualify as such it must obey the operational rules discussed below.

The quantities $a_{ij}$ are called the *entries* or *components* of the matrix.[1] As in the case of vectors, the term "matrix element" will be avoided to lessen the chance of confusion with finite elements. The two subscripts identify the row and column, respectively.

Matrices are conventionally identified by **bold uppercase** letters such as **A**, **B**, etc. The entries of matrix **A** may be denoted as $A_{ij}$ or $a_{ij}$, according to the intended use. Occassionally we shall use the short-hand component notation

$$\mathbf{A} = [a_{ij}]. \tag{B.2}$$

**Example B.1**. The following is a $2 \times 3$ numerical matrix:

$$\mathbf{B} = \begin{bmatrix} 2 & 6 & 3 \\ 4 & 9 & 1 \end{bmatrix} \tag{B.3}$$

This matrix has 2 rows and 3 columns. The first row is $(2, 6, 3)$, the second row is $(4, 9, 1)$, the first column is $(2, 4)$, and so on.

In some contexts it is convenient or useful to display the number of rows and columns. If this is so we will write them underneath the matrix symbol.[2] For the example matrix (B.3) we would show

$$\underset{2\times 3}{\mathbf{B}} \tag{B.4}$$

**Remark B.1**. Matrices should not be confused with determinants.[3] A determinant is a number associated with square matrices ($m = n$), defined according to the rules stated in §B.5

As in the case of vectors, the components of a matrix may be real or complex. If they are real numbers, the matrix is called *real*, and *complex* otherwise. For the present exposition all matrices will be assumed real unless otherwise stated.

---

[1] Of these two terms, "entry" is preferred when talking about the computer implementation, whereas "component" is preferred when connecting matrices to physical entities.

[2] A convention introduced in Berkeley courses by Ray Clough. It is particularly convenient in blackboard expositions.

[3] This confusion is apparent in the literature of the period 1860–1920.

## §B.1.1. Square Matrices

The case $m = n$ is important in practical applications. Such matrices are called *square matrices* of order $n$. Matrices for which $m \neq n$ are called non-square. The term "rectangular" is also used in this context, but this is slightly fuzzier because squares are special cases of rectangles.

Square matrices enjoy properties that are not shared by non-square matrices, such as the symmetry and antisymmetry conditions defined below. Furthermore many operations, such as taking determinants and computing eigenvalues, are only defined for square matrices.

**Example B.2.**

$$\mathbf{C} = \begin{bmatrix} 12 & 6 & 3 \\ 8 & 24 & 7 \\ 2 & 5 & 11 \end{bmatrix} \tag{B.5}$$

is a square matrix of order 3.

Consider a square matrix $\mathbf{A} = [a_{ij}]$ of order $n \times n$. Its $n$ components $a_{ii}$ form the *main diagonal*, which runs from top left to bottom right. The *cross diagonal* or *antidiagonal* runs from bottom left to upper right. The main diagonal of the example matrix (B.5) is {12, 24, 11} whereas its cross diagonal is {2, 24, 3}.

Entries that run parallel to and above (below) the main diagonal form superdiagonals (subdiagonals). For example, {6, 7} is the first superdiagonal of the matrix (B.5).

The *trace* of a $n \times n$ square matrix $\mathbf{A}$, denoted trace($\mathbf{A}$), is the sum of its diagonal coefficients:

$$\operatorname{trace}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{B.6}$$

## §B.1.2. Symmetry and Antisymmetry

Square matrices for which $a_{ij} = a_{ji}$ are called *symmetric about the main diagonal* or simply *symmetric*. Square matrices for which $a_{ij} = -a_{ji}$ are called *antisymmetric*[4]. The diagonal entries of an antisymmetric matrix must be zero.

**Example B.3.** The following is a symmetric matrix of order 3:

$$\mathbf{S} = \begin{bmatrix} 11 & 6 & 1 \\ 6 & 3 & -1 \\ 1 & -1 & -6 \end{bmatrix}. \tag{B.7}$$

The following is an antisymmetric matrix of order 4:

$$\mathbf{W} = \begin{bmatrix} 0 & 3 & -1 & -5 \\ -3 & 0 & 7 & -2 \\ 1 & -7 & 0 & 0 \\ 5 & 2 & 0 & 0 \end{bmatrix}. \tag{B.8}$$

[4] The alternative names *skew-symmetric* and *antimetric* are also used in the literature.

### §B.1.3. Special Matrices

The *null* matrix, written $\mathbf{0}$, is the matrix all of whose components are zero.

**Example B.4.** The null matrix of order $2 \times 3$ is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{B.9}$$

The *identity matrix*, written $\mathbf{I}$, is a square matrix all of which entries are zero except those on the main diagonal, which are ones.

**Example B.5.** The identity matrix of order 4 is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{B.10}$$

A *diagonal matrix* is a square matrix all of which entries are zero except for those on the main diagonal, which may be arbitrary.

**Example B.6.** The following matrix of order 4 is diagonal:

$$\mathbf{D} = \begin{bmatrix} 14 & 0 & 0 & 0 \\ 0 & -6 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \tag{B.11}$$

A short hand notation which lists only the diagonal entries is sometimes used for diagonal matrices to save writing space. This notation is illustrated for the above matrix:

$$\mathbf{D} = \mathbf{diag}\,[\,14 \quad -6 \quad 0 \quad 3\,]. \tag{B.12}$$

An *upper triangular matrix* is a square matrix in which all entries underneath the main diagonal vanish. A *lower triangular matrix* is a square matrix in which all entries above the main diagonal vanish. These forms are important in the solution of linear systems of equations.

**Example B.7.** Here are examples of each kind:

$$\mathbf{U} = \begin{bmatrix} 6 & 4 & 2 & 1 \\ 0 & 6 & 4 & 2 \\ 0 & 0 & 6 & 4 \\ 0 & 0 & 0 & 6 \end{bmatrix}, \qquad \mathbf{L} = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 10 & 4 & 0 & 0 \\ -3 & 21 & 6 & 0 \\ -15 & -2 & 18 & 7 \end{bmatrix}. \tag{B.13}$$

An upper triangular matrix with all diagonal entries equal to 1 is called a *unit upper triangular matrix*. Similarly, a lower triangular matrix with all diagonal entries equal to 1 is called a *unit upper triangular matrix*.

### §B.1.4. *Are Vectors a Special Case of Matrices?

Consider the 3-vector **x** and a $3 \times 1$ matrix **X** with the same components:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}. \tag{B.14}$$

in which $x_1 = x_{11}$, $x_2 = x_{22}$ and $x_3 = x_{33}$. Are **x** and **X** the same thing? If so we could treat column vectors as one-column matrices and dispense with the distinction. Indeed in many contexts a column vector of order $n$ may be treated as a matrix with a single column, i.e., as a matrix of order $n \times 1$. Similarly, a row vector of order $m$ may be sometimes treated as a matrix with a single row, i.e., as a matrix of order $1 \times m$.

There are operations, however, for which the analogy does *not* carry over, and one must consider vectors as different from matrices. The dichotomy is reflected in our notational conventions of lower versus upper case.

Another practical distinction between matrices and vectors can be cited: although we speak of "matrix algebra" as embodying vectors as special cases of matrices, in practice the quantities of primary interest to the engineer are vectors rather than matrices. For example, a structural engineer may be interested in displacement vectors, force vectors, vibration eigenvectors, buckling eigenvectors, etc. In finite element analysis even stresses and strains are often arranged as vectors[5] although they are actually tensors.

On the other hand, matrices are rarely the quantities of primary interest: they work silently in the background where they are normally engaged in operating on vectors.

## §B.2. Elementary Matrix Operations

### §B.2.1. Equality

Two matrices **A** and **B** of same order $m \times n$ are said to be *equal* if and only if all of their components are equal: $a_{ij} = b_{ij}$, for all $i = 1, \ldots m,\ j = 1, \ldots n$. We then write $\mathbf{A} = \mathbf{B}$. If the inequality test fails for at least one entry, the matrices are said to be *unequal* and we write $\mathbf{A} \neq \mathbf{B}$.

Two matrices of different order cannot be compared for equality or inequality.

There are specialized tests for greater-than or less-than, but these are not considered here.

### §B.2.2. Transposition

The *transpose* of a $m \times n$ real matrix $\mathbf{A} = [a_{ij}]$ is a $n \times m$ matrix denoted by $\mathbf{A}^T$, which satifies

$$\mathbf{A}^T = [a_{ji}]. \tag{B.15}$$

The rows of $\mathbf{A}^T$ are the columns of **A**, and the rows of **A** are the columns of $\mathbf{A}^T$. Obviously the transpose of $\mathbf{A}^T$ is again **A**, that is, $(\mathbf{A}^T)^T = \mathbf{A}$.

**Example B.8.**

$$\mathbf{A} = \begin{bmatrix} 5 & 7 & 0 \\ 1 & 0 & 4 \end{bmatrix}, \qquad \mathbf{A}^T = \begin{bmatrix} 5 & 1 \\ 7 & 0 \\ 0 & 4 \end{bmatrix}. \tag{B.16}$$

The transpose of a square matrix is also a square matrix. The transpose of a symmetric matrix **A** is equal to the original matrix, that is $\mathbf{A} = \mathbf{A}^T$. The negated transpose of an antisymmetric matrix matrix **A** is equal to the original matrix, that is $\mathbf{A} = -\mathbf{A}^T$.

The equivalent operation for a complex matrix involves conjugation and is not considered here.

---

[5] Through the "casting" operation described in Chapter 11.

**Example B.9**.

$$\mathbf{A} = \begin{bmatrix} 4 & 7 & 0 \\ 7 & 1 & 2 \\ 0 & 2 & 3 \end{bmatrix} = \mathbf{A}^T, \qquad \mathbf{W} = \begin{bmatrix} 0 & 7 & 0 \\ -7 & 0 & -2 \\ 0 & 2 & 0 \end{bmatrix} = -\mathbf{W}^T \tag{B.17}$$

### §B.2.3. Addition, Subtraction and Scaling

The simplest operation acting on two matrices is *addition*. The sum of two matrices of the same order, $\mathbf{A}$ and $\mathbf{B}$, is written $\mathbf{A} + \mathbf{B}$ and defined to be the matrix

$$\mathbf{A} + \mathbf{B} \stackrel{\text{def}}{=} [a_{ij} + b_{ij}]. \tag{B.18}$$

Like vector addition, matrix addition is commutative: $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$, and associative: $\mathbf{A}+(\mathbf{B}+\mathbf{C}) = (\mathbf{A}+\mathbf{B})+\mathbf{C}$. For $n = 1$ or $m = 1$ the operation reduces to the addition of two column or row vectors, respectively.

For matrix subtraction, replace $+$ by $-$ in the definition (B.18).

**Example B.10**. The sum of

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 0 \\ 4 & 2 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 6 & 3 & -3 \\ 7 & -2 & 5 \end{bmatrix} \quad \text{is} \quad \mathbf{A}+\mathbf{B} = \begin{bmatrix} 7 & 0 & -3 \\ 11 & 0 & 4 \end{bmatrix}. \tag{B.19}$$

Multiplication of a matrix $\mathbf{A}$ by a scalar $c$ is defined by means of the relation

$$c\,\mathbf{A} \stackrel{\text{def}}{=} [c a_{ij}] \tag{B.20}$$

That is, each entry of the matrix is multiplied by $c$. This operation is called *scaling* of a matrix. If $c = 0$, the result is the null matrix. Division of a matrix by a nonzero scalar $c$ is equivalent to multiplication by $(1/c)$.

**Example B.11**.

$$\text{If} \quad \mathbf{A} = \begin{bmatrix} 1 & -3 & 0 \\ 4 & 2 & -1 \end{bmatrix}, \qquad 3\,\mathbf{A} = \begin{bmatrix} 3 & -9 & 0 \\ 12 & 6 & -3 \end{bmatrix}. \tag{B.21}$$

## §B.3. Matrix Products

### §B.3.1. Matrix by Vector Product

Before describing the general matrix product of two matrices, let us consider the particular case in which the second matrix is a column vector. This so-called *matrix-vector product* merits special attention because it occurs frequently in the applications. Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix, $\mathbf{x} = \{x_j\}$ a column vector of order $n$, and $\mathbf{y} = \{y_i\}$ a column vector of order $m$. The matrix-vector product is symbolically written

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{B.22}$$

to mean the linear transformation

$$y_i \stackrel{\text{def}}{=} \sum_{j=1}^{n} a_{ij} x_j \stackrel{\text{sc}}{=} a_{ij} x_j, \qquad i = 1, \ldots, m. \tag{B.23}$$

The symbol $\Sigma$ may be omitted if using the summation convention, as ilustrated above after $\stackrel{\text{sc}}{=}$.

**Example B.12**. The product of a $2 \times 3$ matrix and a 3-vector is a 2-vector:

$$\begin{bmatrix} 1 & -3 & 0 \\ 4 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + (-3) \times 2 + 0 \times 3 \\ 4 \times 1 + 2 \times 2 + (-1) \times 3 \end{bmatrix} = \begin{bmatrix} -5 \\ 5 \end{bmatrix} \tag{B.24}$$

This product definition is not arbitrary but emanates from the analytical and geometric properties of entities represented by matrices and vectors.

For the product definition to make sense, the column dimension of the matrix $\mathbf{A}$ (called the pre-multiplicand) must equal the dimension of the vector $\mathbf{x}$ (called the post-multiplicand). For example, the reverse product $\mathbf{x}\,\mathbf{A}$ would not make sense unless $m = n = 1$.

If the row dimension $m$ of $\mathbf{A}$ is one, the matrix formally reduces to a row vector, and the matrix-vector product reduces to the inner product defined by equation (A.15) of Appendix A. The result of this operation is a one-dimensional vector or scalar. Thus we see that the present definition properly embodies previous cases.

The associative and commutative properties of the matrix-vector product fall under the rules of the more general matrix-matrix product discussed next.

### §B.3.2. Matrix by Matrix Product and Matrix Powers

We now pass to the most general matrix-by-matrix product, and consider the operations involved in computing the product $\mathbf{C}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{C} = \mathbf{A}\,\mathbf{B}. \tag{B.25}$$

Here $\mathbf{A} = [a_{ij}]$ is a matrix of order $m \times n$, $\mathbf{B} = [b_{jk}]$ is a matrix of order $n \times p$, and $\mathbf{C} = [c_{ik}]$ is a matrix of order $m \times p$. The entries of the result matrix $\mathbf{C}$ are defined by the formula

$$c_{ik} \stackrel{\text{def}}{=} \sum_{j=1}^{n} a_{ij} b_{jk} \stackrel{\text{sc}}{=} a_{ij} b_{jk}, \quad i = 1, \ldots, m, \quad k = 1, \ldots, p. \tag{B.26}$$

We see that the $(i, k)^{th}$ entry of $\mathbf{C}$ is computed by taking the *inner product* of the $i^{th}$ row of $\mathbf{A}$ with the $k^{th}$ column of $\mathbf{B}$. For this definition to work and the product be possible, *the column dimension of* $\mathbf{A}$ *must be the same as the row dimension of* $\mathbf{B}$. Matrices that satisfy this rule are said to be *product-conforming*, or *conforming* for short. If the two matrices do not conform, their product is undefined. The following mnemonic notation often helps in remembering this rule:

$$\underset{m \times p}{\mathbf{C}} = \underset{m \times n}{\mathbf{A}} \; \underset{n \times p}{\mathbf{B}} \tag{B.27}$$

For the matrix-by-vector case treated in the preceding subsection, $p = 1$.

Matrix $\mathbf{A}$ is called the pre-multiplicand and is said to *premultiply* $\mathbf{B}$. Matrix $\mathbf{B}$ is called the post-multiplicand and is said to *postmultiply* $\mathbf{A}$. This careful distinction on which matrix comes first is a consequence of the absence of commutativity: even if $\mathbf{B}\,\mathbf{A}$ exists (it only does if $m = n$), it is not generally the same as $\mathbf{AB}$.

If $\mathbf{A} = \mathbf{B}$, the product $\mathbf{A}\,\mathbf{A}$ is called the *square* of $\mathbf{A}$ and is denoted by $\mathbf{A}^2$. Note that for this definition to make sense, $\mathbf{A}$ must be a square matrix — else the factors would not be conforming. Similarly, $\mathbf{A}^3 = \mathbf{AAA} = \mathbf{A}^2\mathbf{A} = \mathbf{AA}^2$. Other positive-integer powers can be analogously defined.

The above definition *does not* encompass negative powers. For inverse, $\mathbf{A}^{-1}$ denotes the *inverse* of $\mathbf{A}$, which is studied later. The general power $\mathbf{A}^m$, where $m$ can be a real or complex scalar, can be defined with the help of the matrix spectral form — an advanced topic covered in Appendix E.

A square matrix $\mathbf{A}$ that satisfies $\mathbf{A} = \mathbf{A}^2$ is called *idempotent*. We shall see later that this condition characterizes the so-called projector matrices. A square matrix $\mathbf{A}$ whose $p^{th}$ power is the null matrix is called *p-nilpotent*.

**Remark B.2**. For *hand* computations, the matrix product is most conveniently organized by the so-called Falk's scheme:

$$\begin{array}{cc} & \begin{bmatrix} b_{11} & \cdots & b_{ik} & \cdots & b_{1p} \\ \vdots & \ddots & \downarrow & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} & \cdots & b_{np} \end{bmatrix} \\ \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \to & a_{in} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} & \begin{bmatrix} & & & & \\ & & \vdots & & \\ & \cdots & c_{ik} & & \\ & & & & \\ & & & & \end{bmatrix} \end{array}. \tag{B.28}$$

Each entry in row $i$ of $\mathbf{A}$ is multiplied by the corresponding entry in column $k$ of $\mathbf{B}$ (note the arrows), and the products are summed and stored in the $(i,k)^{th}$ entry of $\mathbf{C}$.

**Example B.13**. To illustrate Falk's scheme, let us form the product $\mathbf{C} = \mathbf{AB}$ of the following matrices

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 \\ 4 & -1 & 5 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 2 & 1 & 0 & -5 \\ 4 & 3 & -1 & 0 \\ 0 & 1 & -7 & 4 \end{bmatrix} \tag{B.29}$$

The matrices are conforming because the column dimension of $\mathbf{A}$ and the row dimension of $\mathbf{B}$ are the same (3). We arrange the computations as shown below:

$$\begin{array}{cccc} & & \begin{bmatrix} 2 & 1 & 0 & -5 \\ 4 & 3 & -1 & 0 \\ 0 & 1 & -7 & 4 \end{bmatrix} & = \mathbf{B} \\ \mathbf{A} = & \begin{bmatrix} 3 & 0 & 2 \\ 4 & -1 & 5 \end{bmatrix} & \begin{bmatrix} 6 & 5 & -14 & -7 \\ 4 & 6 & -34 & 0 \end{bmatrix} & = \mathbf{C} = \mathbf{AB} \end{array}. \tag{B.30}$$

Here $3 \times 2 + 0 \times 4 + 2 \times 0 = 6$ and so on.

### §B.3.3. Matrix Product Properties

*Associativity*. The associative law is verified:

$$\mathbf{A}\,(\mathbf{B}\,\mathbf{C}) = (\mathbf{A}\,\mathbf{B})\,\mathbf{C}. \tag{B.31}$$

Hence we may delete the parentheses and simply write $\mathbf{A}\,\mathbf{B}\,\mathbf{C}$.

*Distributivity*. The distributive law also holds: If $\mathbf{B}$ and $\mathbf{C}$ are matrices of the same order, then

$$\mathbf{A}\,(\mathbf{B}+\mathbf{C}) = \mathbf{A}\,\mathbf{B} + \mathbf{A}\,\mathbf{C}, \qquad \text{and} \qquad (\mathbf{B}+\mathbf{C})\,\mathbf{A} = \mathbf{B}\,\mathbf{A} + \mathbf{C}\,\mathbf{A}. \tag{B.32}$$

*Commutativity*. The commutativity law of scalar multiplication does not generally hold. If $\mathbf{A}$ and $\mathbf{B}$ are square matrices of the same order, then the products $\mathbf{A}\,\mathbf{B}$ and $\mathbf{B}\,\mathbf{A}$ are both possible but in general $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$. If $\mathbf{A}\,\mathbf{B} = \mathbf{B}\,\mathbf{A}$, the matrices $\mathbf{A}$ and $\mathbf{B}$ are said to *commute*. One important case is when $\mathbf{A}$ and $\mathbf{B}$ are diagonal. In general $\mathbf{A}$ and $\mathbf{B}$ commute if they share the same eigensystem.

**Example B.14.** Matrices

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} a-\beta & b \\ b & c-\beta \end{bmatrix}, \tag{B.33}$$

commute for any $a, b, c, \beta$. More generally, $\mathbf{A}$ and $\mathbf{B} = \mathbf{A} - \beta\,\mathbf{I}$ commute for any square matrix $\mathbf{A}$.

*Transpose of a Product*. The transpose of a matrix product is equal to the product of the transposes of the operands taken in reverse order:

$$(\mathbf{A}\,\mathbf{B})^T = \mathbf{B}^T\,\mathbf{A}^T. \tag{B.34}$$

The general transposition formula for an arbitrary product sequence is

$$(\mathbf{A}\,\mathbf{B}\,\mathbf{C}\ldots\mathbf{M}\,\mathbf{N})^T = \mathbf{N}^T\,\mathbf{M}^T\ldots\mathbf{C}^T\,\mathbf{B}^T\,\mathbf{A}^T. \tag{B.35}$$

*Congruent Transformation*. If $\mathbf{B}$ is a *symmetric* matrix of order $m$ and $\mathbf{A}$ is an arbitrary $m \times n$ matrix, then

$$\mathbf{S} = \mathbf{A}^T\,\mathbf{B}\,\mathbf{A}. \tag{B.36}$$

is a symmetric matrix of order $n$. Such an operation is called a *congruent* or *congruential* transformation. It is often found in finite element analysis when changing coordinate bases because such a transformation preserves certain key properties (such as symmetry).

*Loss of Symmetry*. The product of two symmetric matrices is not generally symmetric.

*Null Matrices may have Non-null Divisors*. The matrix product $\mathbf{AB}$ can be zero although $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$. Likewise it is possible that $\mathbf{A} \neq \mathbf{0}$, $\mathbf{A}^2 \neq \mathbf{0}$, $\ldots$, but $\mathbf{A}^p = \mathbf{0}$.

## §B.4. Bilinear and Quadratic Forms

Let $\mathbf{x}$ and $\mathbf{y}$ be two column vectors of order $n$, and $\mathbf{A}$ a real square $n \times n$ matrix. Then the following triple product produces a *scalar* result:

$$s = \underset{1\times n}{\mathbf{y}^T}\ \underset{n\times n}{\mathbf{A}}\ \underset{n\times 1}{\mathbf{x}} \tag{B.37}$$

This is called a *bilinear form*. Matrix $\mathbf{A}$ is called the *kernel* of the form.

Transposing both sides of (B.37) and noting that the transpose of a scalar does not change, we obtain the result

$$s = \mathbf{x}^T\,\mathbf{A}^T\,\mathbf{y} = \mathbf{y}^T\,\mathbf{A}\,\mathbf{x}. \tag{B.38}$$

If $\mathbf{A}$ is symmetric and vectors $\mathbf{x}$ and $\mathbf{y}$ coalesce, *i.e.* $\mathbf{A}^T = \mathbf{A}$ and $\mathbf{x} = \mathbf{y}$, the bilinear form becomes a *quadratic form*

$$s = \mathbf{x}^T\,\mathbf{A}\,\mathbf{x}. \tag{B.39}$$

Transposing both sides of a quadratic form reproduces the same equation.

**Example B.15**. The kinetic energy of a dynamic system consisting of three point masses $m_1$, $m_2$, $m_3$ moving in one dimension with velocities $v_1$, $v_2$ and $v_3$, respectively, is

$$T = \tfrac{1}{2}(m_1 v_1^2 + m_2 v_2^2 + m_3 v_3^2). \tag{B.40}$$

This can be expressed as the quadratic form

$$T = \tfrac{1}{2}\mathbf{v}^T\,\mathbf{M}\,\mathbf{v}, \tag{B.41}$$

in which

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}. \tag{B.42}$$

Here $\mathbf{M}$ denotes the system mass matrix whereas $\mathbf{v}$ is the system velocity vector.

The following material discusses more specialized properties of matrices, such as determinants, inverses, rank and orthogonality. These apply only to *square* matrices unless othewise stated.

## §B.5. Determinants

The *determinant* of a *square* matrix $\mathbf{A} = [a_{ij}]$ is a number denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$, through which important properties such as matrix singularity or its eigenvalue spectrum can be compactly characterized. This number is defined as the following function of the matrix elements:

$$|\mathbf{A}| = \det(\mathbf{A}) = \pm \prod a_{1j_1}\, a_{2j_2}\, \ldots\, a_{n\,j_n}, \tag{B.43}$$

where the column indices $j_1,\ j_2,\ \ldots\ j_n$ are taken from the set $\{1, 2, \ldots n\}$, with no repetitions allowed. The plus (minus) sign is taken if the permutation $(j_1\ j_2\ \ldots\ j_n)$ is even (odd).

**Example B.16**. If the order $n$ does not exceed 3, closed forms of the determinant are practical. For a $2 \times 2$ matrix,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{B.44}$$

For a $3 \times 3$ matrix,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}. \tag{B.45}$$

**Remark B.3**. The concept of determinant is not applicable to non-square matrices or to vectors. Thus the notation $|\mathbf{x}|$ for a vector $\mathbf{x}$ can be reserved for its magnitude (as in Appendix A) without risk of confusion.

**Remark B.4**. Inasmuch as the product (B.43) contains $n!$ terms, the calculation of $|\mathbf{A}|$ from the definition is impractical for general matrices whose order exceeds 3 or 4. For example, if $n = 10$, the product (B.43) contains $10! = 3,628,800$ terms, each involving 9 multiplications, so over 30 million floating-point operations would be required to evaluate $|\mathbf{A}|$ according to that definition. A more practical method based on matrix decomposition is described in Remark B.3.

### §B.5.1. Determinant Properties

Some useful rules associated with the calculus of determinants are listed next.

I. Rows and columns can be interchanged without affecting the value of a determinant. Consequently

$$|\mathbf{A}| = |\mathbf{A}^T|. \tag{B.46}$$

II. If two rows, or two columns, are interchanged the sign of the determinant is reversed. For example:

$$\begin{vmatrix} 3 & 4 \\ 1 & -2 \end{vmatrix} = - \begin{vmatrix} 1 & -2 \\ 3 & 4 \end{vmatrix}. \tag{B.47}$$

III. If a row (or column) is changed by adding to or subtracting from its elements the corresponding elements of any other row (or column) the determinant remains unaltered. For example:

$$\begin{vmatrix} 3 & 4 \\ 1 & -2 \end{vmatrix} = \begin{vmatrix} 3+1 & 4-2 \\ 1 & -2 \end{vmatrix} = \begin{vmatrix} 4 & 2 \\ 1 & -2 \end{vmatrix} = -10. \tag{B.48}$$

IV. If the elements in any row (or column) have a common factor $\alpha$ then the determinant equals the determinant of the corresponding matrix in which $\alpha = 1$, multiplied by $\alpha$. For example:

$$\begin{vmatrix} 6 & 8 \\ 1 & -2 \end{vmatrix} = 2 \begin{vmatrix} 3 & 4 \\ 1 & -2 \end{vmatrix} = 2 \times (-10) = -20. \tag{B.49}$$

V. When at least one row (or column) of a matrix is a linear combination of the other rows (or columns) the determinant is zero. Conversely, if the determinant is zero, then at least one row and one column are linearly dependent on the other rows and columns, respectively. For example, consider

$$\begin{vmatrix} 3 & 2 & 1 \\ 1 & 2 & -1 \\ 2 & -1 & 3 \end{vmatrix}. \tag{B.50}$$

This determinant is zero because the first column is a linear combination of the second and third columns:

$$\text{column 1} = \text{column 2} + \text{column 3}. \tag{B.51}$$

Similarly, there is a linear dependence between the rows, which is given by the relation

$$\text{row 1} = \tfrac{7}{8}\,\text{row 2} + \tfrac{4}{5}\,\text{row 3}. \tag{B.52}$$

VI. The determinant of an upper triangular or lower triangular matrix is the product of the main diagonal entries. For example,

$$\begin{vmatrix} 3 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 4 \end{vmatrix} = 3 \times 2 \times 4 = 24. \tag{B.53}$$

This rule is easily verified from the definition (B.43) because all terms vanish except $j_1 = 1$, $j_2 = 2, \ldots\ j_n = n$, which is the product of the main diagonal entries. Diagonal matrices are a particular case of this rule.

VII. The determinant of the product of two square matrices is the product of the individual determinants:

$$|\mathbf{A}\,\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|. \tag{B.54}$$

The proof requires the concept of triangular decomposition, which is covered in the Remark below. This rule can be generalized to any number of factors. One immediate application is to matrix powers: $|\mathbf{A}^2| = |\mathbf{A}||\mathbf{A}| = |\mathbf{A}|^2$, and more generally $|\mathbf{A}^n| = |\mathbf{A}|^n$ for integer $n$.

VIII. The determinant of the transpose of a matrix is the same as that of the original matrix:

$$|\mathbf{A}^T| = |\mathbf{A}|. \tag{B.55}$$

This rule can be directly verified from the definition of determinant, and also as direct consequence of Rule I.

**Remark B.5**. Rules VI and VII are the key to the practical evaluation of determinants. Any square nonsingular matrix $\mathbf{A}$ (where the qualifier "nonsingular" is explained in §B.3) can be decomposed as the product of two triangular factors

$$\mathbf{A} = \mathbf{L}\,\mathbf{U}, \tag{B.56}$$

in which $\mathbf{L}$ is unit lower triangular and $\mathbf{U}$ is upper triangular. This is called a LU triangularization, LU factorization or LU decomposition. It can be carried out in $O(n^3)$ floating point operations. According to rule VII:

$$|\mathbf{A}| = |\mathbf{L}|\,|\mathbf{U}|. \tag{B.57}$$

According to rule VI, $|\mathbf{L}| = 1$ and $|\mathbf{U}| = u_{11}u_{22}\ \ldots\ u_{nn}$. The last operation requires only $O(n)$ operations. Thus the evaluation of $|\mathbf{A}|$ is dominated by the effort involved in computing the factorization (B.56). For $n = 10$, that effort is approximately $10^3 = 1000$ floating-point operations, compared to approximately $3 \times 10^7$ from the naive application of the definition (B.43), as noted in Remark B.4. Thus the LU-based method is roughly 30, 000 times faster for that modest matrix order, and the ratio increases exponentially for large $n$.

### §B.5.2. Cramer's Rule and Homogeneous Systems

Cramer's rule provides a recipe for solving linear algebraic equations directly in terms of determinants. Let the system of equations be — as usual — denoted by

$$\mathbf{A}\,\mathbf{x} = \mathbf{y}, \tag{B.58}$$

in which $\mathbf{A}$ is a given $n \times n$ matrix, $\mathbf{y}$ is a given $n \times 1$ vector, and $\mathbf{x}$ is the $n \times 1$ vector of unknowns. The explicit form of (B.58) is Equation (A.1) of Appendix A, with $n = m$.

A closed form solution for the components $x_1, x_2 \ldots, x_n$ of $\mathbf{x}$ in terms of determinants is

$$x_1 = \frac{\begin{vmatrix} y_1 & a_{12} & a_{13} & \ldots & a_{1n} \\ y_2 & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \\ y_n & a_{n2} & a_{n3} & \ldots & a_{nn} \end{vmatrix}}{|\mathbf{A}|}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & y_1 & a_{13} & \ldots & a_{1n} \\ a_{21} & y_2 & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \\ a_{n1} & y_n & a_{n3} & \ldots & a_{nn} \end{vmatrix}}{|\mathbf{A}|}, \ \ldots \tag{B.59}$$

The rule can be remembered as follows: in the numerator of the quotient for $x_j$, replace the $j^{th}$ column of $\mathbf{A}$ by the right-hand side $\mathbf{y}$.

This method of solving simultaneous equations is known as *Cramer's rule*. Because the explicit computation of determinants is impractical for $n > 3$ as explained in Remark C.3, direct use of the rule has practical value only for $n = 2$ and $n = 3$ (it is marginal for $n = 4$). But such small-order systems arise often in finite element calculations at the *Gauss point level*; consequently implementors should be aware of this rule for such applications. In addition the rule may be advantageous for *symbolic* systems; see Example below.

**Example B.17.** Solve the $3 \times 3$ linear system

$$\begin{bmatrix} 5 & 2 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 3 \end{bmatrix}, \tag{B.60}$$

by Cramer's rule:

$$x_1 = \frac{\begin{vmatrix} 8 & 2 & 1 \\ 5 & 2 & 0 \\ 3 & 0 & 2 \end{vmatrix}}{\begin{vmatrix} 5 & 2 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{vmatrix}} = \frac{6}{6} = 1, \quad x_2 = \frac{\begin{vmatrix} 5 & 8 & 1 \\ 3 & 5 & 0 \\ 1 & 3 & 2 \end{vmatrix}}{\begin{vmatrix} 5 & 2 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{vmatrix}} = \frac{6}{6} = 1, \quad x_3 = \frac{\begin{vmatrix} 5 & 2 & 8 \\ 3 & 2 & 5 \\ 1 & 0 & 3 \end{vmatrix}}{\begin{vmatrix} 5 & 2 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{vmatrix}} = \frac{6}{6} = 1. \tag{B.61}$$

**Example B.18.** Solve the $2 \times 2$ linear algebraic system

$$\begin{bmatrix} 2+\beta & -\beta \\ -\beta & 1+\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \tag{B.62}$$

by Cramer's rule:

$$x_1 = \frac{\begin{vmatrix} 5 & -\beta \\ 0 & 1+\beta \end{vmatrix}}{\begin{vmatrix} 2+\beta & -\beta \\ -\beta & 1+\beta \end{vmatrix}} = \frac{5+5\beta}{2+3\beta}, \quad x_2 = \frac{\begin{vmatrix} 2+\beta & 5 \\ -\beta & 0 \end{vmatrix}}{\begin{vmatrix} 2+\beta & -\beta \\ -\beta & 1+\beta \end{vmatrix}} = \frac{5\beta}{2+3\beta}. \tag{B.63}$$

**Remark B.6.** Creamer's rule importance has grown in *symbolic computations* carried out by computer algebra systems. This happens when the entries of $\mathbf{A}$ and $\mathbf{y}$ are algebraic expressions. For example the example system (B.62). In such cases Cramer's rule may be competitive with factorization methods for up to moderate matrix orders, for example $n \leq 20$. The reason is that determinantal products may be simplified on the fly.

One immediate consequence of Cramer's rule is what happens if

$$y_1 = y_2 = \ldots = y_n = 0. \tag{B.64}$$

A linear equation system with a null right hand side

$$\mathbf{A}\,\mathbf{x} = \mathbf{0}, \tag{B.65}$$

is called a *homogeneous system*. From the rule (B.59) we see that if $|\mathbf{A}|$ is nonzero, all solution components are zero, and consequently the only possible solution is the trivial one $\mathbf{x} = \mathbf{0}$. The case in which $|\mathbf{A}|$ vanishes is discussed in the next section.

## §B.6. Singular Matrices, Rank

If the determinant $|\mathbf{A}|$ of a $n \times n$ square matrix $\mathbf{A} \equiv \mathbf{A}_n$ is zero, then the matrix is said to be *singular*. This means that at least one row and one column are linearly dependent on the others. If this row and column are removed, we are left with another matrix, say $\mathbf{A}_{n-1}$, to which we can apply the same criterion. If the determinant $|\mathbf{A}_{n-1}|$ is zero, we can remove another row and column from it to get $\mathbf{A}_{n-2}$, and so on. Suppose that we eventually arrive at an $r \times r$ matrix $\mathbf{A}_r$ whose determinant is nonzero. Then matrix $\mathbf{A}$ is said to have *rank* $r$, and we write $\text{rank}(\mathbf{A}) = r$.

If the determinant of $\mathbf{A}$ is nonzero, then $\mathbf{A}$ is said to be *nonsingular*. The rank of a nonsingular $n \times n$ matrix is equal to $n$.

Obviously the rank of $\mathbf{A}^T$ is the same as that of $\mathbf{A}$ since it is only necessary to transpose "row" and "column" in the definition.

The notion of rank can be extended to rectangular matrices; as covered in the textbooks cited in **Notes and Bibliography**. That extension, however, is not important for the material covered here.

**Example B.19**. The $3 \times 3$ matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}, \tag{B.66}$$

has rank $r = 3$ because $|\mathbf{A}| = -3 \neq 0$.

**Example B.20**. The matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}, \tag{B.67}$$

already used as example in §B.5.1, is singular because its first row and column may be expressed as linear combinations of the others as shown in (B.51) and (B.52). Removing the first row and column we are left with a $2 \times 2$ matrix whose determinant is $2 \times 3 - (-1) \times (-1) = 5 \neq 0$. Thus (B.67) has rank $r = 2$.

### §B.6.1. Rank Deficiency

If the square matrix $\mathbf{A}$ is supposed to be of rank $r$ but in fact has a smaller rank $\bar{r} < r$, the matrix is said to be *rank deficient*. The number $r - \bar{r} > 0$ is called the *rank deficiency*.

**Example B.21**. Suppose that the *unconstrained* master stiffness matrix $\mathbf{K}$ of a finite element has order $n$, and that the element possesses $b$ independent rigid body modes. Then the expected rank of $\mathbf{K}$ is $r = n - b$. If the actual rank is less than $r$, the finite element model is said to be rank-deficient. This is usually undesirable.

**Example B.22**. An an illustration of the foregoing rule, consider the two-node, 4-DOF, Bernoulli-Euler plane beam element stiffness derived in Chapter 12.

$$\mathbf{K} = \frac{EI}{L^3} \begin{bmatrix} 12 & 6L & -12 & 6L \\ & 4L^2 & -6L & 2L^2 \\ & & 12 & -6L \\ \textit{symm} & & & 4L^2 \end{bmatrix}, \tag{B.68}$$

in which $EI$ and $L$ are positive scalars. It may be verified (for example, through an eigenvalue analysis) that this $4 \times 4$ matrix has rank 2. The number of rigid body modes is 2, and the expected rank is $r = 4 - 2 = 2$. Consequently this FEM model is rank sufficient.

### §B.6.2. Rank of Matrix Sums and Products

In finite element analysis matrices are often built through sum and product combinations of simpler matrices. Two important rules apply to "rank propagation" through those combinations.

The rank of the product of two square matrices $\mathbf{A}$ and $\mathbf{B}$ cannot exceed the smallest rank of the multiplicand matrices. That is, if the rank of $\mathbf{A}$ is $r_a$ and the rank of $\mathbf{B}$ is $r_b$,

$$\text{rank}(\mathbf{A}\,\mathbf{B}) \leq \min(r_a, r_b). \tag{B.69}$$

In particular, if either matrix is singular, so is the product. Regarding sums: the rank of a matrix sum cannot exceed the sum of ranks of the summand matrices. That is, if the rank of $\mathbf{A}$ is $r_a$ and the rank of $\mathbf{B}$ is $r_b$,

$$\text{rank}(\mathbf{A} + \mathbf{B}) \leq r_a + r_b. \tag{B.70}$$

Rule (B.70) is the basis for studying the effect of the number of Gauss points on the rank of an element stiffness matrix being built through numerical integration, as discussed in Chapter 19.

### §B.6.3. Singular Systems: Particular and Homogeneous Solutions

Having introduced the notion of rank we can now discuss what happens to the linear system (B.58) when the determinant of $\mathbf{A}$ vanishes, meaning that its rank is less than $n$. If so, (B.58) has either no solution or an infinite number of solutions. Cramer's rule is of limited or no help in this situation.

To discuss this case further we note that if $|\mathbf{A}| = 0$ and the rank of $\mathbf{A}$ is $r = n - d$, where $d \geq 1$ is the *rank deficiency*, then there exist $d$ nonzero independent vectors $\mathbf{z}_i$, $i = 1, \ldots d$ such that

$$\mathbf{A}\,\mathbf{z}_i = \mathbf{0}. \tag{B.71}$$

These $d$ vectors, suitably orthonormalized, are called *null eigenvectors* of $\mathbf{A}$, and form a basis for its *null space*.

Let $\mathbf{Z}$ denote the $n \times d$ matrix obtained by collecting the $\mathbf{z}_i$ as columns. If $\mathbf{y}$ in (B.58) is in the *range* of $\mathbf{A}$, that is, there exists an nonzero $\mathbf{x}_p$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}_p$, its general solution is

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h = \mathbf{x}_p + \mathbf{Z}\,\mathbf{w}, \tag{B.72}$$

where $\mathbf{w}$ is an arbitrary $d \times 1$ weighting vector. This statement can be easily verified by substituting this solution into $\mathbf{A}\mathbf{x} = \mathbf{y}$ and noting that $\mathbf{A}\mathbf{Z}$ vanishes.

The components $\mathbf{x}_p$ and $\mathbf{x}_h$ are called the *particular* and *homogeneous* portions respectively, of the total solution $\mathbf{x}$. (The terminology: *particular solution* and *particular solution*, are often used.) If $\mathbf{y} = \mathbf{0}$ only the homogeneous portion remains.

If $\mathbf{y}$ is not in the range of $\mathbf{A}$, system (B.58) does not generally have a solution in the conventional sense, although least-square solutions can usually be constructed. The reader is referred to the many textbooks in linear algebra for further details.

### §B.6.4. *Rank of Rectangular Matrices

The notion of rank can be extended to rectangular matrices, real or complex, as follows. Let $\mathbf{A}$ be $m \times n$. Its *column range space* $\mathcal{R}(\mathbf{A})$ is the subspace spanned by $\mathbf{Ax}$ where $\mathbf{x}$ is the set of all complex $n$-vectors. Mathematically: $\mathcal{R}(\mathbf{A}) = \{\mathbf{Ax} : \mathbf{x} \in C^n\}$. The rank $r$ of $\mathbf{A}$ is the dimension of $\mathcal{R}(\mathbf{A})$.

The *null space* $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$ is the set of $n$-vectors $\mathbf{z}$ such that $\mathbf{A}\,\mathbf{z} = \mathbf{0}$. The dimension of $\mathcal{N}(\mathbf{A})$ is $n - r$.

Using these definitions, the product and sum rules (B.69) and (B.70) generalize to the case of rectangular (but conforming) $\mathbf{A}$ and $\mathbf{B}$. So does the treatment of linear equation systems $\mathbf{A}\,\mathbf{x} = \mathbf{y}$ in which $\mathbf{A}$ is rectangular. Such systems often arise in the fitting of observation and measurement data.

In finite element methods, rectangular matrices appear in change of basis through congruential transformations, and in the treatment of multifreedom constraints.

## §B.7. Matrix Inversion

The *inverse* of a square nonsingular matrix $\mathbf{A}$ is represented by the symbol $\mathbf{A}^{-1}$ and is defined by the relation

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{I}. \tag{B.73}$$

The most important application of the concept of inverse is the solution of linear systems. Suppose that, in the usual notation, we have

$$\mathbf{A}\,\mathbf{x} = \mathbf{y}. \tag{B.74}$$

Premultiplying both sides by $\mathbf{A}^{-1}$ we get the inverse relationship

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{y}. \tag{B.75}$$

More generally, consider the matrix equation for multiple ($m$) right-hand sides:

$$\underset{n\times n}{\mathbf{A}}\;\underset{n\times m}{\mathbf{X}} = \underset{n\times m}{\mathbf{Y}}, \tag{B.76}$$

which reduces to (B.74) for $m = 1$. The inverse relation that gives $\mathbf{X}$ as function of $\mathbf{Y}$ is

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}. \tag{B.77}$$

In particular, the solution of

$$\mathbf{A}\,\mathbf{X} = \mathbf{I}, \tag{B.78}$$

is $\mathbf{X} = \mathbf{A}^{-1}$. Practical methods for computing inverses are based on directly solving this equation; see Remark below.

### §B.7.1. Explicit Computation of Inverses

The explicit calculation of matrix inverses is seldom needed in large matrix computations. But occasionally the need arises for the explicit inverse of small matrices that appear in element level computations. For example, the inversion of Jacobian matrices at Gauss points, or of constitutive matrices.

A general formula for elements of the inverse can be obtained by specializing Cramer's rule to (B.78). Let $\mathbf{B} = [b_{ij}] = \mathbf{A}^{-1}$. Then

$$b_{ij} = \frac{A_{ji}}{|\mathbf{A}|}, \tag{B.79}$$

in which $A_{ji}$ denotes the so-called *adjoint* of entry $a_{ij}$ of $\mathbf{A}$. The adjoint $A_{ji}$ is defined as the determinant of the submatrix of order $(n-1) \times (n-1)$ obtained by deleting the $j^{th}$ row and $i^{th}$ column of $\mathbf{A}$, multiplied by $(-1)^{i+j}$.

This direct inversion procedure is useful only for small matrix orders, say 2 or 3. In the examples below the explicit inversion formulas for second and third order matrices are listed.

**Example B.23**. For order $n = 2$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{22} \end{bmatrix}, \tag{B.80}$$

in which $|\mathbf{A}|$ is given by (B.44).

**Example B.24**. For order $n = 3$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}, \tag{B.81}$$

where

$$\begin{aligned} b_{11} &= \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, & b_{21} &= -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}, & b_{31} &= \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \\ b_{12} &= -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, & b_{22} &= \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, & b_{32} &= -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \\ b_{13} &= \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, & b_{23} &= -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}, & b_{33} &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \end{aligned} \tag{B.82}$$

in which $|\mathbf{A}|$ is given by (B.45).

**Example B.25**.

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 3 & 0 \\ -1 & 2 & 5 \end{bmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{36} \begin{bmatrix} 15 & -3 & 3 \\ -10 & 14 & -2 \\ -1 & 5 & 7 \end{bmatrix}. \tag{B.83}$$

If the order exceeds 3, the general inversion formula based on Cramer's rule becomes rapidly useless because it displays combinatorial complexity as noted in a previous Remark. For numerical work it is preferable to solve (B.78) after $\mathbf{A}$ is factored. Those techniques are described in detail in linear algebra books; see also Remark below.

### §B.7.2. Some Properties of the Inverse

I. If $\mathbf{A}$ is nonsingular, the determinant of $\mathbf{A}^{-1}$ is the inverse of the determinant of $\mathbf{A}$

II. If $\mathbf{A}^{-1}$ exists, the inverse of its transpose is equal to the transpose of the inverse:

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T, \tag{B.84}$$

because $(\mathbf{A}\mathbf{A}^{-1}) = (\mathbf{A}\mathbf{A}^{-1})^T = (\mathbf{A}^{-1})^T\,\mathbf{A}^T = \mathbf{I}$.

III. The inverse of a symmetric matrix is also symmetric. Because of the previous rule, $(\mathbf{A}^T)^{-1} = \mathbf{A}^{-1} = (\mathbf{A}^{-1})^T$, hence $\mathbf{A}^{-1}$ is also symmetric.

IV. The inverse of a matrix product is the reverse product of the inverses of the factors:

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{B.85}$$

which is easily proven by substitution. This rule generalizes to an arbitrary number of factors.

V. For a diagonal matrix $\mathbf{D}$ in which all diagonal entries are nonzero, $\mathbf{D}^{-1}$ is again a diagonal matrix with entries $1/d_{ii}$. The verification is straightforward.

VI. If $\mathbf{S}$ is a block diagonal matrix:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{33} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{S}_{nn} \end{bmatrix} = \mathrm{diag}\,[\,\mathbf{S}_{ii}\,], \tag{B.86}$$

in which all $\mathbf{S}_{jj}$ are nonsingular, the inverse matrix is also block diagonal and is given by

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{S}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{33}^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{S}_{nn}^{-1} \end{bmatrix} = \mathrm{diag}\,[\,\mathbf{S}_{ii}^{-1}\,]. \tag{B.87}$$

VII. The inverse of an upper triangular matrix is also an upper triangular matrix. The inverse of a lower triangular matrix is also a lower triangular matrix. Both inverses can be computed in $O(n^2)$ floating-point operations.

VIII. An antisymmetric matrix is always singular.

**Remark B.7**. The practical numerical calculation of inverses is based on triangular factorization.[6] Given a nonsingular $n \times n$ matrix $\mathbf{A}$, calculate its LU factorization $\mathbf{A} = \mathbf{L}\,\mathbf{U}$, which can be obtained in $O(n^3)$ operations. Then solve the linear triangular systems:

$$\mathbf{U}\mathbf{Y} = \mathbf{I}, \qquad \mathbf{L}\mathbf{X} = \mathbf{Y}, \tag{B.88}$$

and the computed inverse $\mathbf{A}^{-1}$ appears in $\mathbf{X}$. One can overwrite $\mathbf{I}$ with $\mathbf{Y}$ and $\mathbf{Y}$ with $\mathbf{X}$. The whole process can be completed in $O(n^3)$ floating-point operations. For symmetric matrices the alternative decomposition $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$, where $\mathbf{L}$ is unit lower triangular and $\mathbf{D}$ is diagonal, is generally preferred to save computing time and storage.

[6] See References at the end of the Appendix.

## §B.8. Characteristic Polynomial and Eigenvalues

Let **A** be a square $n \times n$ matrix, and $\lambda$ a (real or complex) scalar variable. The determinant of $\mathbf{A} - \lambda\mathbf{I}$ is a polynomial in $\lambda$ of order $n$:

$$P(\lambda) = \det(\mathbf{A} - \lambda\,\mathbf{I}) = C_0 + C_1\lambda + \ldots + C_n\,\lambda^n, \tag{B.89}$$

which is called the *charactristic polynomial* of **A**. The $C_i$ are the *characteristic coefficients*. It is easy shown that $C_0 = \det(\mathbf{A})$ and $C_{n-1} = (-1)^{n-1}\,\mathrm{trace}(\mathbf{A})$. The $n$ roots of $P(\lambda) = 0$ are the *eigenvalues* of **A**. These may be complex even if **A** is real. This topic is developed in Appendix E.

**Example B.26**. The characteristic polynomial of the matrix **A** in (B.83) is

$$P(\lambda) = \det(\mathbf{A} - \lambda\,\mathbf{I}_3) = \begin{bmatrix} 3-\lambda & 1 & -1 \\ 2 & 3-\lambda & 0 \\ -1 & 2 & 5-\lambda \end{bmatrix} = 36 - 36\,\lambda + 11\,\lambda^2 - \lambda^3. \tag{B.90}$$

The roots of $P(\lambda) = 0$, in ascending order, are $\lambda_1 = 2$, $\lambda_2 = 3$ and $\lambda_3 = 6$. These are the eigenvalues of **A**. It may be verified that the eigenvalues of $\mathbf{A}^{-1}$ are $1/2$, $1/3$ and $1/6$. This illustrates a general rule: the eigenvalues of a nonsingular **A** and its inverse $\mathbf{A}^{-1}$ are reciprocal, as further discussed in Appendix E.

### §B.8.1. *Similarity Transformations and Invariants

Let again **A** be a square $n \times n$ matrix, whereas **T** denotes a *nonsingular* $n \times n$ matrix. The transformation

$$\mathbf{B} = \mathbf{T}^{-1}\,\mathbf{A}\,\mathbf{T}, \tag{B.91}$$

is called a *similarity transformation*. **A** and **B** are said to be *similar matrices*. The determinants of **A** and **B** are identical since $|\mathbf{B}| = |\mathbf{T}|\,|\mathbf{A}|\,(1/|\mathbf{T}|)$ — see Rule VII in §B.5.1. Likewise the determinants of $\mathbf{A} - \lambda\,\mathbf{I}$ and $\mathbf{B} - \lambda\,\mathbf{I}$ stay the same. This implies that the characteristic polynomials of **A** and **B** are identical, whence their eigenvalues coincide.

A *matrix invariant* or simply *invariant* is any expression that does not change under a similarity transformation. The characteristic coefficients furnish an obvious example since $P(\lambda)$ is unchanged. In particular $C_0 = \mathrm{trace}(\mathbf{A})$ and $C_n = \det(\mathbf{A})$ are invariants. The $n$ coefficients $C_n$ of $P(\lambda)$ are said to form an *invariant basis*. Any expression that is only a function of those coefficients, such as the eigenvalues, is also an invariant.

### §B.8.2. *Congruential Transformations

A matrix transformation that often appears in FEM developments is the *congruential transformation* (also called *congruent*):

$$\mathbf{B} = \mathbf{T}^T\,\mathbf{A}\,\mathbf{T}. \tag{B.92}$$

Matrices **B** and **A** related through (B.92) are said to be *congruent*. This operation becomes a similarity transformation if $\mathbf{T}^T = \mathbf{T}^{-1}$, or $\mathbf{T}^T\mathbf{T} = \mathbf{I}$, which for real **A** characterizes an *orthogonal* matrix.

Aside from the special case of **T** being orthogonal, a congruential transformation does not preserve eigenvalues (or generally, any invariants). But (B.92) can be generalized to product-conforming rectangular **T** matrices, whereas (B.91) cannot. Another key difference: if **A** is real symmetric, a congruential transformation preserves symmetry, whereas a similarity transformation with a non-orthogonal matrix destroys it.

### §B.8.3. *The Cayley-Hamilton Theorem

The Cayley-Hamilton theorem is of great importance in advanced matrix theory. It states that *a square matrix satisfies its own characteristic equation*. Symbolically: $P(\mathbf{A}) = \mathbf{0}$.

**Example B.27.** To see the theorem in action, consider the matrix $\mathbf{A}$ given in (B.83). Its characteristic polynomial was computed in (B.90) to be $P(\lambda)36 - 36\,\lambda + 11\,\lambda^2 - \lambda^3$. Replacing $\lambda^m$ by $\mathbf{A}^m$ (if $m = 0$, by $\mathbf{I}_3$) we can check that

$$P(\mathbf{A}) = 36\mathbf{I}_3 - 36\,\mathbf{A} + 11\,\mathbf{A}^2 - \mathbf{A}^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{0}. \tag{B.93}$$

## §B.9. *The Inverse of a Sum of Matrices

The formula for the inverse of a matrix product: $(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$ is not too different from its scalar counterpart: $(ab)^{-1} = (1/a)(1/b) = (1/b)(1/a)$, except that factor order matters. On the other hand, formulas for matrix sum inverses in terms of the summands are considerably more involved, and there are many variants. We consider here the expression of $(\mathbf{A} + \mathbf{B})^{-1}$ where both $\mathbf{A}$ and $\mathbf{B}$ are square and $\mathbf{A}$ is nonsingular. We begin from the identity introduced by Henderson and Searle in their review article [385]:

$$(\mathbf{I} + \mathbf{P})^{-1} = (\mathbf{I} + \mathbf{P})^{-1}\,(\mathbf{I} + \mathbf{P} - \mathbf{P}) = \mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1}\,\mathbf{P}, \tag{B.94}$$

in which $\mathbf{P}$ is a square matrix. Using (B.94) we may develop $(\mathbf{A} + \mathbf{B})^{-1}$ as follows:

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^{-1} &= \left(\mathbf{A}(\mathbf{I} + \mathbf{A}^{-1}\,\mathbf{B})^{-1}\right)^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\,\mathbf{B})^{-1}\,\mathbf{A}^{-1}\mathbf{B} \\ &= \left(\mathbf{I} - (\mathbf{I} + \mathbf{A}^{-1}\,\mathbf{B})^{-1}\mathbf{A}^{-1}\right)\,\mathbf{A}^{-1} = \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\,\mathbf{B})^{-1}\,\mathbf{A}^{-1}\,\mathbf{B}\,\mathbf{A}^{-1}. \end{aligned} \tag{B.95}$$

Here $\mathbf{B}$ may be singular. If $\mathbf{B} = \mathbf{0}$, it reduces to $\mathbf{A}^{-1} = \mathbf{A}^{-1}$. The check $\mathbf{B} = \beta\,\mathbf{A}$ also works. The last expression in (B.95) may be further transformed by matrix manipulations as

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\,\mathbf{B})^{-1}\,\mathbf{A}^{-1}\,\mathbf{B}\,\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\,(\mathbf{I} + \mathbf{B}\,\mathbf{A}^{-1})^{-1}\,\mathbf{B}\,\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\,(\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\,\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\,\mathbf{A}^{-1}(\mathbf{I} + \mathbf{B}\,\mathbf{A}^{-1})^{-1}. \end{aligned} \tag{B.96}$$

In all of these forms $\mathbf{B}$ may be singular (or null). If $\mathbf{B}$ is also invertible, the third expresion in (B.96) may be transformed to the a commonly used variant

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\,\mathbf{A}^{-1}. \tag{B.97}$$

The case of singular $\mathbf{A}$ may be handled using the notion of generalized inverses. This is a topic beyond the scope of this course, which may be studied, e.g., in the textbooks [83,129,686]. The special case of $\mathbf{B}$ being of low rank merges with the Sherman-Morrison and Woodbury formulas, covered below.

The Sherman-Morrison formula discussed below gives the inverse of a matrix modified by a rank-one matrix. The Woodbury formula extends it to a modification of arbitrary rank. In structural analysis these formulas are of interest for problems of *structural modifications*, in which a finite-element (or, in general, a discrete model) is changed by an amount expressable as a low-rank correction to the original model.

### §B.9.1. *The Sherman-Morrison and Woodbury Formulas

Let $\mathbf{A}$ be a square $n \times n$ invertible matrix, whereas $\mathbf{u}$ and $\mathbf{v}$ are two $n$-vectors and $\beta$ an arbitrary scalar. Assume that $\sigma = 1 + \beta\, \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u} \neq 0$. Then

$$\left(\mathbf{A} + \beta\, \mathbf{u}\, \mathbf{v}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{\beta}{\sigma}\, \mathbf{A}^{-1}\, \mathbf{u}\, \mathbf{v}^T\, \mathbf{A}^{-1}. \tag{B.98}$$

When $\beta = 1$ this is called the Sherman-Morrison formula after [740]. (For a history of this remarkable expression and its extensions, which are important in many applications such as statistics and probability, see the review paper by Henderson and Searle cited previously.) Since any rank-one correction to $\mathbf{A}$ can be written as $\beta \mathbf{u}\mathbf{v}^T$, (B.98) gives the rank-one change to its inverse. The proof can be done by direct multiplication; see Exercise B.15.

For practical computation of the change one solves the linear systems $\mathbf{A}\mathbf{a} = \mathbf{u}$ and $\mathbf{A}\mathbf{b} = \mathbf{v}$ for $\mathbf{a}$ and $\mathbf{b}$, using the known $\mathbf{A}^{-1}$. Compute $\sigma = 1 + \beta \mathbf{v}^T \mathbf{a}$. If $\sigma \neq 0$, the change to $\mathbf{A}^{-1}$ is the dyadic $-(\beta/\sigma)\mathbf{a}\mathbf{b}^T$.

Consider next the case where $\mathbf{U}$ and $\mathbf{V}$ are two $n \times k$ matrices with $k \leq n$ and $\beta$ an arbitrary scalar. Assume that the $k \times k$ matrix $\boldsymbol{\Sigma} = \mathbf{I}_k + \beta\, \mathbf{V}^T \mathbf{A}^{-1} \mathbf{U}$, in which $\mathbf{I}_k$ denotes the $k \times k$ identity matrix, is invertible. Then

$$\left(\mathbf{A} + \beta \mathbf{U}\mathbf{V}^T\right)^{-1} = \mathbf{A}^{-1} - \beta\, \mathbf{A}^{-1} \mathbf{U} \boldsymbol{\Sigma}^{-1} \mathbf{V}^T \mathbf{A}^{-1}. \tag{B.99}$$

This is called the Woodbury formula, after [890]. It reduces to the Sherman-Morrison formula (B.98) if $k = 1$, in which case $\boldsymbol{\Sigma} \equiv \sigma$ is a scalar.

### §B.9.2. *Formulas for Modified Determinants

Let $\widetilde{\mathbf{A}}$ denote the adjoint of $\mathbf{A}$. Taking the determinants from both sides of $\mathbf{A} + \beta \mathbf{u}\mathbf{v}^T$ one obtains

$$|\, \mathbf{A} + \beta \mathbf{u}\mathbf{v}^T\, | = |\, \mathbf{A}| + \beta\, \mathbf{v}^T\, \widetilde{\mathbf{A}}\, \mathbf{u}. \tag{B.100}$$

If $\mathbf{A}$ is invertible, replacing $\widetilde{\mathbf{A}} = |\mathbf{A}|\, \mathbf{A}^{-1}$ this becomes

$$|\mathbf{A} + \beta\, \mathbf{u}\mathbf{v}^T\, | = |\mathbf{A}|\, (1 + \beta\, \mathbf{v}^T \mathbf{A}^{-1}\, \mathbf{u}). \tag{B.101}$$

Similarly, one can show that if $\mathbf{A}$ is invertible, and $\mathbf{U}$ and $\mathbf{V}$ are $n \times k$ matrices,

$$|\, \mathbf{A} + \beta\, \mathbf{U}\mathbf{V}^T\, | = |\mathbf{A}|\ \, |\, \mathbf{I}_k + \beta\, \mathbf{V}^T \mathbf{A}^{-1} \mathbf{U}|. \tag{B.102}$$

### §B.9.3. *Inversion of Low Order Symmetric Matrix Sums

The inversion formulas given here are important in recent developments in continuum mechanics. Let $\mathbf{A}$ be a $n \times n$ positive definite symmetric matrix with $n = 2, 3$, $\mathbf{I}$ the identity matrix of order $n$ and $c$ a scalar. We want to express $\mathbf{A} + c\, \mathbf{I})^{-1}$ directly in terms of $\mathbf{A}$, $\mathbf{I}$, $c$ and the invariants of $\mathbf{A}$ provided by its characteristic coefficients. The derivations are carried out in [402], and make essential use of the Cayley-Hamilton theorem.

*Two Dimensional Case*. Here $n = 2$. The invariants are $I_A = \text{trace}(\mathbf{A})$ and $II_A = \det(\mathbf{A})$. Then

$$(\mathbf{A} + c\, \mathbf{I})^{-1} = -\frac{1}{c\,(c + I_A) + II_A}\, (\mathbf{A} - (c + I_A)\, \mathbf{I})\,. \tag{B.103}$$

*Three Dimensional Case*. Here $n = 3$. The invariants are $I_A = \text{trace}(\mathbf{A})$, $II_A = (\text{trace}(\mathbf{A}^2) - I_A^2)/2$, and $III_A = \det(\mathbf{A})$. Then

$$(\mathbf{A} + c\, \mathbf{I})^{-1} = \frac{1}{c[c(c + I_A) + II_A] + III_A} \left(\mathbf{A}^2 - (c + I_A)\, \mathbf{A} + [c(c + I_A) + II_A]\, \mathbf{I}\right). \tag{B.104}$$

## §B.10. *Matrix Orthogonality

Let $\mathbf{A}$ and $\mathbf{B}$ be two product-conforming real matrices. For example, $\mathbf{A}$ is $k \times m$ whereas $\mathbf{B}$ is $m \times n$. If their product is the null matrix

$$\mathbf{C} = \mathbf{A}\,\mathbf{B} = \mathbf{0}, \tag{B.105}$$

the matrices are said to be *orthogonal*. This is the generalization of the notions of vector orthogonality discussed in the previous Appendix.

### §B.10.1. *Matrix Orthogonalization Via Projectors

The *matrix orthogonalization* problem can be stated as follows. Product conforming matrices $\mathbf{A}$ and $\mathbf{B}$ are given but their product is not zero. How can $\mathbf{A}$ be orthogonalized with respect to $\mathbf{B}$ so that (B.105) is verified? Suppose that $\mathbf{B}$ is $m \times n$ with $m \geq n$ and that $\mathbf{B}^T\mathbf{B}$ is nonsingular (equivalently, $\mathbf{B}$ has full rank).[7] Then form the $m \times m$ *orthogonal projector matrix*, or simply *projector*

$$\boxed{\mathbf{P}_B = \mathbf{I} - \mathbf{B}\,(\mathbf{B}^T\,\mathbf{B})^{-1}\,\mathbf{B}^T.} \tag{B.106}$$

in which $\mathbf{I}$ is the $m \times m$ identity matrix. Since $\mathbf{P}_B = \mathbf{P}_B^T$, the projector is square symmetric.

Note that

$$\mathbf{P}_B\,\mathbf{B} = \mathbf{B} - \mathbf{B}\,(\mathbf{B}^T\,\mathbf{B})^{-1}\,(\mathbf{B}^T\mathbf{B}) = \mathbf{B} - \mathbf{B} = \mathbf{0}. \tag{B.107}$$

It follows that $\mathbf{P}_B$ projects $\mathbf{B}$ onto its null space. Likewise $\mathbf{B}^T\,\mathbf{P}_B = \mathbf{0}$. Postmultiplying $\mathbf{A}$ by $\mathbf{P}_B$ yields

$$\boxed{\tilde{\mathbf{A}} = \mathbf{A}\,\mathbf{P}_B = \mathbf{A} - \mathbf{A}\,\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\,\mathbf{B}^T.} \tag{B.108}$$

Matrix $\tilde{\mathbf{A}}$ is called the *projection* of $\mathbf{A}$ onto the null space of $\mathbf{B}$.[8] It is easily verified that $\tilde{\mathbf{A}}$ and $\mathbf{B}$ are orthogonal:

$$\tilde{\mathbf{A}}\,\mathbf{B} = \mathbf{A}\,\mathbf{B} - \mathbf{A}\,\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\,(\mathbf{B}^T\mathbf{B}) = \mathbf{A}\,\mathbf{B} - \mathbf{A}\,\mathbf{B} = \mathbf{0}. \tag{B.109}$$

Consequently, forming $\tilde{\mathbf{A}}$ via (B.106) and (B.108) solves the orthogonalization problem.

If $\mathbf{B}$ is square and nonsingular, $\tilde{\mathbf{A}} = \mathbf{0}$, as may be expected. If $\mathbf{B}$ has more columns than rows, that is $m < n$, the projector (B.106) cannot be constructed since $\mathbf{B}\,\mathbf{B}^T$ is necessarily singular. A similar difficulty arises if $m \geq n$ but $\mathbf{B}^T\mathbf{B}$ is singular. Such cases require treatment using generalized inverses, which is a topic beyond the scope of this Appendix.[9]

In some applications, notably FEM, matrix $\mathbf{A}$ is square symmetric and it is desirable to preserve symmetry in $\tilde{\mathbf{A}}$. That can be done by pre-and postmultiplying by the projector:

$$\boxed{\tilde{\mathbf{A}} = \mathbf{P}_B\,\mathbf{A}\,\mathbf{P}_B.} \tag{B.110}$$

Since $\mathbf{P}_B^T = \mathbf{P}_B$, the operation (B.110) is a congruential transformation, which preserves symmetry.

---

7 If you are not sure what “singular”, “nonsingular” and “rank” mean or what $(.)^{-1}$ stands for, please read §B.7.

8 In contexts such as control and signal processing, $\mathbf{P}_B$ is called a *filter* and the operation (B.108) is called *filtering*.

9 See e.g., the textbooks [83,686].

### §B.10.2. *Orthogonal Projector Properties

The following properties of the projector (B.106) are useful when checking out computations. Forming its square as

$$\begin{aligned}\mathbf{P}_B^2 &= \mathbf{P}_B\,\mathbf{P}_B = \mathbf{I} - 2\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\,\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\\ &= \mathbf{I} - 2\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T + \mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \mathbf{I} - \mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \mathbf{P}_B,\end{aligned} \tag{B.111}$$

shows that the projector matrix is idempotent. Repeating the process one sees that $\mathbf{P}_B^n = \mathbf{P}_B$, in which $n$ is an arbitrary nonnegative integer.

If $\mathbf{B}$ is $m \times n$ with $m \geq n$ and full rank $n$, $\mathbf{P}_B$ has $m - n$ unit eigenvalues and $n$ zero eigenvalues. This is shown in the paper [258], in which various applications of orthogonal projectors and orthogonalization to multilevel FEM computations are covered in detail.

**Notes and Bibliography**

Much of the material summarized here is available in expanded form in linear algebra textbooks. Highly reputable ones include Bellman [73], Golub and Van Loan [335], Householder [410], and Strang [773]. As books of historical importance we may cite Aitken [9], Muir [555], Turnbull [833], and Wilkinson and Reinsch [876]. For references focused on the eigenvalue problem, see Appendix E.

A comprehensive online catalog of matrix formulas is available in [917].

For inverses of matrix sums, there are two SIAM Review articles: [359,385]. For an historical account of the topic and its close relation to the Schur complement, see the bibliography in Appendix P.

## Homework Exercises for Appendix B: Matrices

In all ensuing exercises matrices and vectors are assumed *real* unless otherwise stated.

**EXERCISE B.1** [A:10] Given the three matrices

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 1 & 0 \\ -1 & 2 & 3 & 1 \\ 2 & 5 & -1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & -2 \\ 1 & 0 \\ 4 & 1 \\ -3 & 2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & -3 & 2 \\ 2 & 0 & 2 \end{bmatrix} \tag{EB.1}$$

compute the product $\mathbf{D} = \mathbf{ABC}$ by hand using Falk's scheme. *Hint:* do $\mathbf{BC}$ first, then premultiply that by $\mathbf{A}$.

**EXERCISE B.2** [A:10] Given the square matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ -4 & 2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 3 & 0 \\ 1 & -2 \end{bmatrix} \tag{EB.2}$$

verify by direct computation that $\mathbf{A\,B} \neq \mathbf{B\,A}$.

**EXERCISE B.3** [A:10] Given the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ -1 & 2 \\ 2 & 0 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 3 & -1 & 4 \\ -1 & 2 & 0 \\ 4 & 0 & 0 \end{bmatrix} \tag{EB.3}$$

(note that $\mathbf{B}$ is symmetric) compute $\mathbf{S} = \mathbf{A}^T\,\mathbf{B}\,\mathbf{A}$, and verify that $\mathbf{S}$ is symmetric.

**EXERCISE B.4** [A:10] Given the square matrices

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 0 & 3 \\ 3 & -2 & -5 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 3 & -6 & -3 \\ 7 & -14 & -7 \\ -1 & 2 & 1 \end{bmatrix} \tag{EB.4}$$

verify that $\mathbf{A\,B} = \mathbf{0}$ although $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$. Is $\mathbf{B\,A}$ also null?

**EXERCISE B.5** [A:10] Given the square matrix

$$\mathbf{A} = \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix} \tag{EB.5}$$

in which $a, b, c$ are real or complex scalares, show by direct computation that $\mathbf{A}^2 \neq \mathbf{0}$ but $\mathbf{A}^3 = \mathbf{0}$.

**EXERCISE B.6** [A:10] Can a diagonal matrix be antisymmetric?

**EXERCISE B.7** [A:20] Prove the matrix product transposition rule (B.34). *Hint:* call $\mathbf{C} = (\mathbf{AB})^T$, $\mathbf{D} = \mathbf{B}^T\mathbf{A}^T$, and use the matrix product definition (B.26) to show that the generic entries of $\mathbf{C}$ and $\mathbf{D}$ agree.

**EXERCISE B.8** [A:20] If $\mathbf{A}$ is an arbitrary $m \times n$ real matrix, show: (a) both products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ exist, and (b) both are square and symmetric. *Hint*: for (b) use the symmetry condition $\mathbf{S} = \mathbf{S}^T$ and (**?**).

**EXERCISE B.9** [A:15] Show that $\mathbf{A}^2$ only exists if and only if $\mathbf{A}$ is square.

**EXERCISE B.10** [A:20] If $\mathbf{A}$ is square and antisymmetric, show that $\mathbf{A}^2$ is symmetric. *Hint*: start from $\mathbf{A} = -\mathbf{A}^T$ and apply the results of Exercise B.8.

**EXERCISE B.11** [A:15] If $\mathbf{A}$ is a square matrix of order $n$ and $c$ a scalar, show that $\det(c\mathbf{A}) = c^n \det \mathbf{A}$.

**EXERCISE B.12** [A:20] Let $\mathbf{u}$ and $\mathbf{v}$ denote real $n$-vectors normalized to unit length, so that $\mathbf{u}^T\,\mathbf{u} = 1$ and $\mathbf{v}^T\mathbf{v} = 1$, and let $\mathbf{I}$ denote the $n \times n$ identity matrix. Show that

$$\det(\mathbf{I} - \mathbf{u}\mathbf{v}^T) = 1 - \mathbf{v}^T\mathbf{u} \tag{EB.6}$$

**EXERCISE B.13** [A:25] Let $\mathbf{u}$ denote a real $n$-vector normalized to unit length so that $\mathbf{u}^T\,\mathbf{u} = 1$, while $\mathbf{I}$ denotes the $n \times n$ identity matrix. Show that

$$\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T \tag{EB.7}$$

is orthogonal: $\mathbf{H}^T\mathbf{H} = \mathbf{I}$, and idempotent: $\mathbf{H}^2 = \mathbf{H}$. This matrix is called a *elementary Hermitian*, a *Householder matrix*, or a *reflector*. It is a fundamental ingredient of many linear algebra algorithms; for example the QR algorithm for finding eigenvalues.

**EXERCISE B.14** [A:20] The *trace* of a $n \times n$ square matrix $\mathbf{A}$, denoted trace($\mathbf{A}$) is defined by (B.6). Show that if the entries of $\mathbf{A}$ are real,

$$\text{trace}(\mathbf{A}^T\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 \tag{EB.8}$$

This is the same as the Frobenius norm $||\mathbf{A}||_F$ (incorrectly called Euclidean norm by some authors).

**EXERCISE B.15** [A:25] Prove the Sherman-Morrison formula (B.98) by direct matrix multiplication.

**EXERCISE B.16** [A:30] Prove the Woodbury formula (B.99) by considering the following block bordered system

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V}^T & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix} \tag{EB.9}$$

in which $\mathbf{I}_k$ and $\mathbf{I}_n$ denote the identity matrices of orders $k$ and $n$, respectively. Solve (EB.9) two ways: eliminating first $\mathbf{B}$ and then $\mathbf{C}$, and eliminating first $\mathbf{C}$ and then $\mathbf{B}$. Equate the results for $\mathbf{B}$.

**EXERCISE B.17** [A:20] Show that the eigenvalues of a real symmetric square matrix are real, and that the eigenvectors are real vectors.

**EXERCISE B.18** [A:25] Let the $n$ real eigenvalues $\lambda_i$ of a real $n \times n$ symmetric matrix $\mathbf{A}$ be classified into two subsets: $r$ eigenvalues are nonzero whereas $n - r$ are zero. Show that $\mathbf{A}$ has rank $r$.

**EXERCISE B.19** [A:15] Show that if $\mathbf{A}$ is positive definite, $\mathbf{A}\mathbf{x} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$.

**EXERCISE B.20** [A:20] Show that for any real $m \times n$ matrix $\mathbf{A}$, $\mathbf{A}^T\mathbf{A}$ is nonnegative. (Existence is proven in a previous Exercise).

**EXERCISE B.21** [A:15] Show that a triangular matrix is normal if and only if it is diagonal.

**EXERCISE B.22** [A:20] Let $\mathbf{A}$ be a real orthogonal matrix. Show that all of its eigenvalues $\lambda_i$, which are generally complex, have unit modulus.

**EXERCISE B.23** [A:15] Let $\mathbf{A}$ and $\mathbf{B}$ be two real symmetric matrices. If the product $\mathbf{A}\,\mathbf{B}$ is symmetric, show that $\mathbf{A}$ and $\mathbf{B}$ commute.

**EXERCISE B.24** [A:30] Let $\mathbf{A}$ and $\mathbf{T}$ be real $n \times n$ matrices, with $\mathbf{T}$ nonsingular. Show that $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ and $\mathbf{A}$ have the same eigenvalues. (This is called a similarity transformation in linear algebra).

**EXERCISE B.25** [A:40] (Research level) Let $\mathbf{A}$ be $m \times n$ and $\mathbf{B}$ be $n \times m$. Show that the nonzero eigenvalues of $\mathbf{AB}$ are the same as those of $\mathbf{BA}$ (Kahan).

**EXERCISE B.26** [A:25] Let $\mathbf{A}$ be real skew-symmetric, that is, $\mathbf{A} = -\mathbf{A}^T$. Show that all eigenvalues of $\mathbf{A}$ are either purely imaginary or zero.

**EXERCISE B.27** [A:30] Let $\mathbf{A}$ be real skew-symmetric, that is, $\mathbf{A} = -\mathbf{A}^T$. Show that $\mathbf{U} = (\mathbf{I}+\mathbf{A})^{-1}(\mathbf{I}-\mathbf{A})$, called a Cayley transformation, is orthogonal.

**EXERCISE B.28** [A:30] Let $\mathbf{P}$ be a real square matrix that satisfies

$$\mathbf{P}^2 = \mathbf{P}. \tag{EB.10}$$

Such matrices are called *idempotent*, and also *orthogonal projectors*. Show that (a) all eigenvalues of $\mathbf{P}$ are either zero or one; (b) $\mathbf{I} - \mathbf{P}$ is also a projector; (c) $\mathbf{P}^T$ is also a projector.

**EXERCISE B.29** [A:35] The necessary and sufficient condition for two square matrices to commute is that they have the same eigenvectors.

**EXERCISE B.30** [A:30] A matrix whose elements are equal on any line parallel to the main diagonal is called a Toeplitz matrix. (They arise in finite difference or finite element discretizations of regular one-dimensional grids.) Show that if $\mathbf{T}_1$ and $\mathbf{T}_2$ are any two Toeplitz matrices, they commute: $\mathbf{T}_1\mathbf{T}_2 = \mathbf{T}_2\mathbf{T}_1$. Hint: do a Fourier transform to show that the eigenvectors of any Toeplitz matrix are of the form $\{e^{i\omega nh}\}$; then apply the previous Exercise.