2 Matrix Algebra

2.1 Noncommutative algebra of lists

We shall now formalize and consummate the notational initiatives taken in the previous chapter for the symbolic representation and manipulation of whole arrays, culminating, at least for square matrices, in a set of matrix relations and operations of sufficient mathematical structure to merit the distinction of being called *algebra*.

What is algebra? For real (or complex) numbers it is the embodiment of their formal properties under the twin operations of *addition* and *multiplication*. If a, b, c stand for three numbers, then it is true that:

1. Their addition is commutative, $a + b = b + a$.
2. Their addition is associative, $a + (b + c) = (a + b) + c$.
3. Their multiplication is commutative, $ab = ba$.
4. Their multiplication is associative, $(ab)c = a(bc)$.
5. Their multiplication is distributive with respect to addition, $a(b + c) = ab + ac$.
6. There exists a number 0, zero, such that $a + 0 = a$.
7. There exists a number $-a$, the additive inverse of a, such that $a + (-a) = 0$.
8. There exists a number 1, one, different from 0 such that $a1 = a$.
9. For $a \neq 0$ there exists a number a^{-1}, the multiplicative inverse of a, such that $aa^{-1} = 1$.

This is the algebra we know, and the one we hope and aspire to emulate for matrices.

Zero enters the above rules in addition only. Its universal multiplication property is deductive. Rule 6 says that $0 + 0 = 0$ and hence by rule 5 $a(0 + 0) = a0 + a0 = a0$ so that $a0 = 0$. From rule 8 we have that $1(-1) = -1$, and hence also $(-1)1 = -1$. Writing $-1(1 + (-1)) = 0$ we verify that $(-1)(-1) = 1$. From $a(1 + (-1)) = 0$ we obtain that $-a = -1a$. From rules 5 and 7 we have that $a(b + (-b)) = ab + a(-b) = 0$, and $a(-b) = -(ab)$. Similarly $-a(b + (-b)) = 0$ leads to $(-a)(-b) = ab$.

The rules of positive times negative is negative, and negative times negative is positive appear upon their introduction mysterious and arbitrary, which they are, but they save the unified algebra of real numbers.

The objects we deal with here are lists of numbers such as those for the unknowns or right-hand sides of a system of equations, that are *ordered sequentially* in long one-dimensional arrays; or lists of numbers such as the coefficients of a system of linear equations, that are ordered in two-dimensional tabular or *matrix* forms. We made a determined step in the direction of matrix algebra when we instituted the convention of designating an entire list by a single Roman letter – lower case for a one-index array, and upper case for a two-index array, or matrix.

From this unassuming though propitious beginning, matters evolve naturally and we symbolize relations between, and operations with, ordered arrays. Two matrices are *the same* or *equal*, we decide, if all their corresponding entries are equal. If A and B denote two matrices, then $A = B$ is written to express their sameness, and once the $=$ sign convention is enacted there is no ambiguity as to what $A = B$ means, provided we are informed that A and B are matrices.

Also, for matrices as for numbers, we shall define *two* operations, or rules of combination, that we name addition and multiplication, written in the style of numbers as $A+B$ and AB. In the abstract sense these definitions are arbitrary, but they are not careless, nor groundless, but arise naturally in the handling of systems of linear equations. In the context of these definitions we shall discover matrix analogies to 0 and 1, that we mark O and I, respectively.

Sum matrix C of the two matrices A and B, $C = A + B$, we define as holding the sums of the corresponding entries of A and B. Such matrix addition is simple enough to have properties in complete formal agreement with the algebra of numbers. For any matrix A there exists, we shall soon find, a unique additive inverse $-A$ so that $A + (-A) = \mathrm{O}$, where O is the empty or zero matrix, and we shall also confirm that matrix addition is commutative and associative. Matrix equation $X + A = B$ is formally solved by all this as $X + A + (-A) = B + (-A)$, $X + \mathrm{O} = B + (-A)$, $X = B + (-A)$, in the exact way of numbers, except that addition here is specifically that of tables or matrices, and so is equation.

Product matrix C of two matrices A and B, $C = AB$ is purposely defined as being combined matrix C in the system $Cx = f$, resulting from the *substitution* of system $y = Bx$ into system $f = Ay$ so that $A(Bx) = f$, or $A(Bx) = Cx$. This operation that we want to call matrix multiplication is too complex for a total analogy with number multiplication. There is a price to pay for the spare notation for such complex objects as matrices. Something has to be given up.

In many important respects matrix multiplication is still formally indistinguishable from that of numbers. It is associative, $A(BC) = (AB)C$, and distributive with respect to addition, $A(B + C) = AB + AC$. But it is, alas, *not commutative*. Generally $AB \neq BA$; substitution of $y = Bx$ into $f = Ay$ to have $A(Bx) = f$ is decidedly not the same as substitution of $y = Ax$ into $f = By$ to have $B(Ax) = f$.

Only in very special instances does $AB = BA$ hold for matrices, but to restrict the algebra to this narrow class of matrices would make it limited to the point of uselessness for our purpose. For a certain class of *square* matrices a multiplicative inverse A^{-1} is found such that $A^{-1}A = AA^{-1} = I$, but the situation is not as simple here as with numbers, where a^{-1} exists for all $a \neq 0$. A matrix multiplicative inverse, we shall see, does not exist for every square nonzero matrix.

To have a multiplicative inverse the matrix must possess intrinsic properties that require work and insight to discover. Moreover with matrices, $AB = \mathrm{O}$ can happen with both $A \neq \mathrm{O}$ and $B \neq \mathrm{O}$. It may even happen that $AB = \mathrm{O}$ while $BA \neq \mathrm{O}$. Having thoroughly familiarized oneself with the algebra of numbers, one must discipline oneself further to accept the exceptions to the rules of the *noncommutative algebra* of matrices.

But even in the absence of commutative multiplication, matrix algebra is still sufficiently versatile to do us immense service in rendering logical exactness to matrix statements by substituting symbols for words. The extension of algebra to objects more complex than numbers, and the realization that even under the compulsion to abandon commutativity, noncommutative algebra is still of great mathematical usefulness is one of the great intellectual accomplishments of modern times, in league with non-Euclidean geometry.

Linear algebra is at once practically and theoretically interesting. Linearity is omnipresent in every field of our endeavor. Every process and event that is not catastrophic is instantaneously linear in the parameters that describe it. Because linear algebra deals with objects of considerable inherent complexity, an unlimited mathematical richness unfolds as the subject is further and further explored. On account of both the practice and the theory, linear algebra is emerging as one of the most vigorous, vital, and illustrious of all mathematical disciplines.

2.2 Vector conventions

Dealing with systems of linear equations, we acquired the habit of writing lists of unknowns and right-hand sides in a *column*. It is standard practice to call the *columnwise ordered* list of n numbers

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{2.1}$$

a (*column*) *vector*. The reason for the name lies in the distant history of geometry and mechanics. Notational consistency peculiar to matrix algebra makes us differentiate between a *column* vector, that is written from top to bottom, and a *row* vector, that is listed from left to right. If a is a column vector then its *transpose*

$$a^T = [a_1 \ a_2 \ \dots \ a_n], \ a = [a_1 \ a_2 \ \dots \ a_n]^T \tag{2.2}$$

is a row vector.

Equality in eqs. (2.1) and (2.2) means *is*, and we commonly employ lower-case Roman letters to denote vectors. Numbers, or *scalars* are usually denoted by lower-case Greek letters.

Capital Romans are reserved for matrices. Numbers $a_1, a_2, \ldots, a_n$ are the *components* of vector a, and two vectors are equal if their corresponding components are equal, implying that order in the list is important. Vector equation $a = b$ is shorthand for n equations for the n components. Vector calculus begins with the following

Definitions.

1. *Vector equality:* $a = b$, $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ *is* $\begin{matrix} a_1 = b_1 \\ a_2 = b_2 \\ \vdots \\ a_n = b_n \end{matrix}$.

2. *Multiplication of a vector by a scalar:* $\alpha a = \alpha \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_n \end{bmatrix}$.

3. *Vector addition:* $a + b = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$.

4. *Zero vector:* $o = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

5. *Vector subtraction:* $a - b = a + (-1)b = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_n - b_n \end{bmatrix}$.

As we progress from vectors to matrices we shall find it important to distinguish between a row vector and a column vector. For the time being a vector is a mere one-dimensional string of components, and what we define for a is also true for a^T. But we shall never mix the two and we thus reject expressions such as $a + b^T$.

Theorem 2.1.

1. *Vector addition is commutative,* $a + b = b + a$.
2. *Vector addition is associative,* $(a + b) + c = a + (b + c)$.
3. *There is a unique vector* o*, the zero vector, such that* $a + o = a$.

4. To every vector a there corresponds a unique vector $-a$, the additive inverse, such that $a + (-a) = o$.

5. Multiplication of a vector by a scalar is associative, $\alpha(\beta a) = \beta(\alpha a) = \alpha\beta a$.

6. Multiplication of a vector by a scalar is distributive with respect to vector addition, $\alpha(a + b) = \alpha a + \alpha b$.

7. Multiplication of a vector by a scalar is distributive with respect to scalar addition, $(\alpha + \beta)a = \alpha a + \beta a$.

Proof. Proofs to the seven statements follow directly from the definitions, and the laws of the algebra of numbers. End of proof.

A rectangular matrix has m *rows* and n *columns*, and when this needs to be pointed out we shall write $A = A(m \times n)$. A matrix can be looked upon as an ordered set of n column vectors each with m components, or an ordered set of m row vectors each with n components. Consistently, a vector is a one-column matrix. For historical and mathematical reasons, including the recognition of the different character that vectors and matrices have in systems of linear equations, we retain the distinct terminology and notation for vectors and matrices.

As a symbol for the left-hand side of a system of linear equations we wrote Ax, and we want to name it now for what it appears to be – the *multiplication* of vector x by matrix A,

$$Ax = \begin{bmatrix} A_{11} & A_{12} & & A_{1n} \\ A_{21} & A_{22} & & A_{2n} \\ & & & \\ A_{m1} & A_{m2} & & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix} \tag{2.3}$$

or

$$(Ax)_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n, \; i = 1, 2, \ldots, m. \tag{2.4}$$

Matrix vector multiplication symbolizes the act of substituting x into Ax. The order in which the symbols are written is all-important, and xA is senseless.

Notationally, matters become somewhat unsettling with the introduction of *subscripted* vectors, but they are unavoidable. Rectangular matrix $A = A(m \times n)$ if considered as consisting of n column vectors a_j, or m row vectors ${a_i}^T$ is written as

$$A = [a_1 \ a_2 \ldots a_n] \quad , \quad A = \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ \vdots \\ {a_m}^T \end{bmatrix} \tag{2.5}$$

where row ${a_i}^T$ is not necessarily the transpose of column a_i.

In the extreme, matrix A may consist of only one column or one row. Still, we shall retain the distinct notation and terminology for the one-column and one-row matrices and keep writing a and a^T. The reasons for which we make this fine, and to a degree artificial, distinction between vectors and matrices and consider them as different mathematical objects will become clearer as we grow richer in algebraic understanding.

Recognizing a^Tx as the left-hand side of one equation in n unknowns, we write the product Ax in terms of the rows ${a_i}^T$ of A as

$$Ax = \begin{bmatrix} {a_1}^T x \\ {a_2}^T x \\ \vdots \\ {a_m}^T x \end{bmatrix} \quad . \tag{2.6}$$

Otherwise

$$Ax = x_1 \begin{bmatrix} A_{11} \\ A_{21} \\ \\ A_{m1} \end{bmatrix} + x_2 \begin{bmatrix} A_{12} \\ A_{22} \\ \\ A_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} A_{1n} \\ A_{2n} \\ \\ A_{mn} \end{bmatrix} = x_1a_1 + x_2a_2 + \ldots x_na_n \tag{2.7}$$

which is a *linear combination* of the n columns of A.

With subscripted vectors we must be on our guard and keep in mind that in eq. (2.7) $x_1, x_2, \ldots, x_n$ are *numbers*, the components of x, while $a_1, a_2, \ldots, a_n$ are *vectors*, the columns of A. Whenever we can help it we shall avoid the notational mixture of eq. (2.7) and use Greek letters for scalars.

The matrix vector multiplication of a one-row matrix by a column vector

$$a^Tb = [a_1 \ a_2 \qquad a_n] \begin{bmatrix} b_1 \\ b_2 \\ \\ b_n \end{bmatrix} = b^Ta = [b_1 \ b_2 \qquad b_n] \begin{bmatrix} a_1 \\ a_2 \\ \\ a_n \end{bmatrix} = a_1b_1 + a_2b_2 + \cdots + a_nb_n \tag{2.8}$$

is a very basic operation in matrix computations. It produces a scalar, and is called the *scalar, dot*, or *inner* product of vectors a and b.

Vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \tag{2.9}$$

or generally e_j for the vector with all components equal to zero, except for the jth which is 1, are useful for their property $x^T e_j = e_j^T x = x_j$. Also, if $A = [a_1\ a_2 \ldots a_n]$, then according to equation (2.7) $Ae_j = a_j$, and

$$e_i^T (Ae_j) = A_{ij}. \tag{2.10}$$

Theorem 2.2.

$$A(\alpha x + \alpha' x') = \alpha Ax + \alpha' Ax'. \tag{2.11}$$

Proof. This is an immediate result of the meaning of Ax, and the vector conventions. End of proof.

Theorem 2.2 expresses the inherent *linearity* of operation Ax.

We define:

1. The *zero* matrix O=O($m \times n$) as having entries $O_{ij} = 0$.
2. The *identity* matrix $I = I(n \times n)$ as having entries $I_{ii} = 1, I_{ij} = 0$.

For instance,

$$I = I(3 \times 3) = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = [e_1\ e_2\ e_3]. \tag{2.12}$$

We shall occasionally write I_n for $I = I(n \times n)$.

Theorem 2.3.

1. *There exists a unique matrix O, the zero matrix, so that $Ox = o$ for any vector x.*
2. *There exists a unique matrix I, the identity matrix, so that $Ix = x$ for any vector x.*

Proof.

1. If $\mathrm{O}_{ij} = 0$ then $\mathrm{O}x = o$ for any x. Conversely, according to Lemma 1.7, if $Ax = o$ for any x, then necessarily $A_{ij} = 0$.

2. If I is as in eq. (2.12), then $Ix = x$ for any x. Conversely if $Ax = Ix$ for any x, then according to Lemma 1.7 necessarily $A_{ij} - I_{ij} = 0$. End of proof.

Exercises

2.2.1. What is α if $\alpha x = o$. Discuss all possibilities.

2.2.2. Does A exist so that $Ax = x^T$?

2.2.3. If $x^Tx + y^Ty = 0$ what are x and y ?

2.2.4. For $a = [1\ \ -2\ 1]^T,\ b = [1\ 1\ 1]^T$ write $a^Tb,\ b^Ta,\ ab^T$, and ba^T. Are ab and a^Tb^T defined?

2.2.5. For $a = [1\ \ -3\ 2]^T,\ b = [-2\ \ -1\ 0]^T,\ c = [1\ 1\ \ -1]^T$ write $a^Tbc, ca^Tb,\ (b^Tc)(a^Tc)$, $b^T(ca^T)c$. Is abc^T defined?

2.2.6. Prove that $(ab^T)(pq^T) = (b^Tp)(aq^T)$, and that $a^T(bp^T)q = (a^Tb)(p^Tq)$.

2.2.7. Can $(b^Ta)(ab^T)$ be computed as $b^T(aa)b^T$? What about the associative law?

2.2.8. If a and b are two column vectors with n components, how many rows and columns do matrices $a^Tb + b^Ta$, $ab^T + ba^T,\ a^Tbb^Ta,\ ab^Tab^T,\ a^Tba$ and b^Taa^T have?

2.2.9. Is $a^Tb + ab^T$ defined?

2.2.10. For $a = [1\ \ -1\ 1]^T,\ b = [2\ 1\ \ -1]^T$, write ab^T and $(ab^T)^2$. Hint: do the algebra first.

2.2.11. For $= [1\ \ -1\ 1]^T,\ b = [1\ 1\ 1]^T$, write $(ab^T)^6$. Hint: do the algebra first.

2.2.12. For $a = [1\ 1\ 1]^T,\ b = [1\ \ -2\ 1]^T$, write $(aa^T + bb^T)^4$. Hint: do the algebra first and notice that $a^Tb = 0$.

2.2.13. Show that $(ab^T)^n = (b^Ta)^{n-1}ab^T$.

2.2.14. Show that if $A = uu^T + vv^T$, with $u^T u = v^T v = 1$, $u^T v = v^T u = \gamma$, then $A^3 = \alpha_1 uu^T + \alpha_2 uv^T + \alpha_3 vu^T + \alpha_4 vv^T$. Find $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in terms of γ.

2.2.15. Does there exist a vector v so that $vv^T = I$?

2.3 Matrix addition and multiplication

We find it convenient to define matrix relationships and operations in terms of their effect on an arbitrary vector.

Definitions.

1. *Matrices A and B are equal, $A = B$, if and only if $Ax = Bx$ for any vector x.*

2. *Matrix B is the scalar α times $A, B = \alpha A$, if and only if $Bx = \alpha Ax$, $(\alpha A)x = \alpha Ax$, for any vector x.*

3. *Matrix C is the sum of the two matrices A and B, $C = A + B$, if and only if $Cx = Ax + Bx$, $(A + B)x = Ax + Bx$, for any vector x.*

Theorem 2.4.

1. *Equality $A = B$ holds if and only if $A_{ij} = B_{ij}$.*

2. *Equality $B = \alpha A$ holds if and only if $B_{ij} = \alpha A_{ij}$.*

3. *Equality $C = A + B$ holds if and only if $C_{ij} = A_{ij} + B_{ij}$.*

Proof.

1. If $A_{ij} = B_{ij}$, then $Ax = Bx$ for all x. With $x = e_j$ we have that $Ax = Bx$ only if $a_j = b_j$, where a_j and b_j are the jth column of A and B, respectively. Hence from vector equation we have that $Ax = Bx$ for any x only if $A_{ij} = B_{ij}$.

2. If $B_{ij} = \alpha A_{ij}$, then $Bx = \alpha Ax$ for all x. From $Be_j = \alpha Ae_j$ we have that $b_j = \alpha a_j$, and hence equality holds only if $B_{ij} = \alpha A_{ij}$.

3. If $C_{ij} = A_{ij}+B_{ij}$, then $Cx = Ax+Bx$ for any x. With $x = e_j$ we have that $c_j = a_j+b_j$, and hence from vector addition it results that equality holds only if $C_{ij} = A_{ij} + B_{ij}$. End of proof.

In some instances as in geometry and mechanics, one prefers to give vectors an identity of their own separate from matrices. A vector is then essentially a one-dimensional list of components for which there is no point in distinguishing between a row and a column. In this case it matters little how the vector is written, and for typographical convenience it is commonly listed row-wise. But for matrices it matters. Matrices are inherently two-dimensional tables with rows and columns, and a matrix with n rows and one column does not equal a matrix with one row and n columns. We shall consistently look upon a vector as a one-column matrix.

Theorem 2.5.

1. *Matrix addition is commutative,* $A + B = B + A$.

2. *Matrix addition is associative,* $(A + B) + C = A + (B + C)$.

3. *The equality* $A+O= A$ *holds for any matrix* A *if and only if* O *is the zero matrix.*

4. *To every matrix* A *there corresponds a unique additive inverse* $-A$, *so that* $A + (-A) =$O.

5. $\alpha(\beta A) = (\alpha\beta)A$.

6. $(\alpha + \beta)A = \alpha A + \beta A$.

7. $\alpha(A + B) = \alpha A + \alpha B$.

Proof.

1. $(A + B)x = Ax + Bx = Bx + Ax = (B + A)x$.

2. $$\begin{aligned}((A + B) + C)x &= (A + B)x + Cx\\ &= Ax + Bx + Cx\\ &= Ax + (B + C)x\\ &= (A + (B + C))x.\end{aligned}$$

3. $(A + O)x = Ax + Ox = Ax$ implies that $Ox = o$, and O must be, by Theorem 2.3, the zero matrix.

4. $(A + B)x = Ox$ implies that $Ax = -Bx$. Choosing $x = e_j$ we have that $b_j = -a_j$ for the columns of B and A, and $B_{ij} = -A_{ij}$. Matrix B is unique, since if $A + B = O$ and $A + B' = O$, then $A + B' + B = B$ and $B' = B$.

5. $(\alpha(\beta A))x = \alpha((\beta A)x) = \alpha(\beta Ax) = \alpha\beta Ax$.

6. $((\alpha + \beta)A)x = (\alpha + \beta)Ax = \alpha Ax + \beta Ax = (\alpha A + \beta A)x$.

7. $(\alpha(A + B))x = \alpha(A + B)x = \alpha(Ax + Bx) = (\alpha A + \alpha B)x$.

End of proof.

Consider the consecutive linear systems $Ay = f$ and $Bx = y$. Symbolic *substitution* of one equation into the other, $A(Bx) = f$, results in the system $Cx = f$ with coefficient matrix C that we call the *product* of A and B. This is matrix multiplication and we put it formally in the

Definition. *Matrix C is the product of A and $B, C = AB$, if and only if $Cx = A(Bx)$, $(AB)x = A(Bx)$, for any vector x.*

To write the product in terms of entries A and B we have

Theorem 2.6. *The equality $C = AB,\ A = A(m \times k),\ B = B(k \times n),\ C(m \times n)$, holds if and only if*

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ik}B_{kj}. \tag{2.13}$$

Proof. For a detailed look at the product take the typical

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \end{bmatrix} \tag{2.14}$$

for which

$$Bx = \begin{bmatrix} B_{11}x_1 + B_{12}x_2 + B_{13}x_3 \\ B_{21}x_1 + B_{22}x_2 + B_{23}x_3 \end{bmatrix} \tag{2.15}$$

and

$A(Bx) =$

$$\begin{bmatrix} (A_{11}B_{11} + A_{12}B_{21})x_1 + (A_{11}B_{12} + A_{12}B_{22})x_2 + (A_{11}B_{13} + A_{12}B_{23})x_3 \\ (A_{21}B_{11} + A_{22}B_{21})x_1 + (A_{21}B_{12} + A_{22}B_{22})x_2 + (A_{21}B_{13} + A_{22}B_{22})x_3 \\ (A_{31}B_{11} + A_{32}B_{21})x_1 + (A_{31}B_{12} + A_{32}B_{22})x_2 + (A_{31}B_{13} + A_{32}B_{23})x_3 \end{bmatrix}. \tag{2.16}$$

Hence

$$C = AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} & A_{11}B_{13} + A_{12}B_{23} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} & A_{21}B_{13} + A_{22}B_{23} \\ A_{31}B_{11} + A_{32}B_{21} & A_{31}B_{12} + A_{32}B_{22} & A_{31}B_{13} + A_{32}B_{23} \end{bmatrix}. \tag{2.17}$$

End of proof.

Theorem 2.6 says that $C_{ij} = (AB)_{ij}$ is the product of the ith row of A by the jth column of B, $(AB)_{ij} = {a_i}^T b_j$, schematically

$$\begin{array}{cc} & \begin{array}{ccc} & j & \end{array} \\ \begin{array}{c} \\ \\ i \end{array} & \begin{bmatrix} & & \\ & & \\ & \times & \end{bmatrix} \end{array} = \begin{array}{cc} & \\ \begin{array}{c} \\ \\ i \end{array} & \begin{bmatrix} & & \\ & & \\ \times & \times & \times \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{ccc} & j & \end{array} \\ \begin{bmatrix} & \times & \\ & \times & \\ & \times & \end{bmatrix} \end{array}. \qquad (2.18)$$

That product AB exists does not mean that BA is possible, but even if both AB and BA exist, the product of the ith row of A and the jth column of B need not be the same as the product of the ith row of B and the jth column of A. The operation is too complicated for that, and *matrix multiplication is not commutative*. Generally $AB \neq BA$ and we need distinguish between B that *post-multiplies* A, and B that *pre-multiplies* A, or B that right-multiplies A and B that left-multiplies A. Substitution of a system with matrix B into a system with matrix A is not the same as substitution of a system with matrix A into a system with matrix B.

Matrices for which it does happen that $AB = BA$ are said to *commute*.

Setting $e_1, e_2, \ldots, e_n$ into $Cx = (AB)x = A(Bx)$ and recalling that $Ce_j = c_j$, the jth column of C, and $Be_j = b_j$, the jth column of B, we have that $c_j = Ab_j$. The jth column of $C = AB$ is the product of matrix A by the jth column of B, $C = [Ab_1\ Ab_2 \ldots Ab_n]$. Matrix multiplication becomes in this way an obvious generalization of matrix vector multiplication.

When writing AB we shall implicitly assume that A and B are *compatible*, that A has as many columns as B has rows, so that the product is possible.

For the zero matrix O and identity matrix I we verify that $AO = OA = O$, and $IA = AI = A$ for any A.

Apart from this we have

Theorem 2.7.

1. AO=O, OA =O for any matrix A if and only if O is the zero matrix.

2. $AI = A,\ IA = A$ *for any matrix A if and only if I is the identity matrix.*

3. $A(B + C) = AB + AC$.

4. $(A+B)C = AC + BC$.

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

6. $(\alpha A)(\beta B) = \alpha\beta AB$.

Proof.

1. We choose in $AX = X$ or $XA = X$ $A =$O and have that $X =$O.

2. The equations $AX = A$ or $XA = A$ can happen only if X is square. Equality is for any A, and if we choose $A = I$ we have $X = I$.

3. $$\begin{aligned}(A(B+C))x &= A((B+C)x) = A(Bx+Cx)\\ &= A(Bx) + A(Cx)\\ &= (AB)x + (AC)x\\ &= (AB+AC)x.\end{aligned}$$

4. The same as 3.

5. $A(BC)x = A((BC)x) = A(B(Cx)) = AB(Cx) = (AB)Cx$.

6. $(\alpha A)(\beta B)x = (\alpha A)(\beta(Bx)) = \alpha\beta(AB)x$.

End of proof.

For *specific* matrices it may well happen that $AX = O$ with $X \neq O$, and $AX = A$ with $X \neq I$. Consider for example

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ AX = O, \tag{2.19}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \ AX = A. \tag{2.20}$$

The product a^Tb of $a^T = a^T(1 \times n)$ and $b = b(n \times 1)$ is a (1×1) matrix, but the product $\alpha A, \ A = A(m \times n)$, is not the product of a (1×1) matrix α by an $(m \times n)$ matrix A. In this scheme of things we let numbers retain an existence separate from matrices, as vectors may do, and the operation of scalar times a matrix is not part of the matrix product operations. An A (1×1) matrix loses its status of matrix and becomes a number.

Multiplication of every entry of A by scalar α can be done in proper matrix form by pre- or post-multiplying A by the diagonal $D = \alpha I$

$$\begin{bmatrix} \alpha & & \\ & \alpha & \\ & & \alpha \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \begin{bmatrix} \alpha & \\ & \alpha \end{bmatrix} = \begin{bmatrix} \alpha A_{11} & \alpha A_{12} \\ \alpha A_{21} & \alpha A_{22} \\ \alpha A_{31} & \alpha A_{32} \end{bmatrix}. \tag{2.21}$$

Lack of commutativeness in matrix multiplication detracts from the algebra, but it is the associative law that saves it. One realizes how immensely important the law is in the manipulation of matrix products, and how poor is an algebra devoid of it.

The associative law allows us to write matrix products without parentheses, and we may use the notation $AA = A^2$, $A^2A^3 = A^5$. On the other hand, because matrix multiplication is not commutative $(AB)^2 \neq A^2B^2$, actually, $(AB)^2 = ABAB$, and $(A+B)^2 = A^2 + AB + BA + B^2$. We see in example 1 below that AB =O can happen with both $A \neq$ O and $B \neq$ O, in which case A and B are said to be *divisors of zero*. A homogeneous system of linear equations $Ax = o$, we know, can well have nonhomogeneous solutions even when A is square. To conclude that $x = o$ is the only solution requires that we have knowledge of a deeper property of A, namely that the system is equivalent to triangular form of type 1.

Examples.

1. $$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

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

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

4. $$AB = \begin{bmatrix} 1 & 3\alpha \\ 3\beta & -2 \end{bmatrix} \begin{bmatrix} 3 & \alpha \\ \beta & 2 \end{bmatrix} = \begin{bmatrix} 3 + 3\alpha\beta & 7\alpha \\ 7\beta & 3\alpha\beta - 4 \end{bmatrix}.$$

$$BA = \begin{bmatrix} 3 & \alpha \\ \beta & 2 \end{bmatrix} \begin{bmatrix} 1 & 3\alpha \\ 3\beta & -2 \end{bmatrix} = \begin{bmatrix} 3+3\alpha\beta & 7\alpha \\ 7\beta & 3\alpha\beta - 4 \end{bmatrix}$$

and $AB = BA$ for any α and β.

5. $$\begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} d_1A_{11} & d_1A_{12} & d_1A_{13} \\ d_2A_{21} & d_2A_{22} & d_2A_{23} \\ d_3A_{31} & d_3A_{32} & d_3A_{33} \end{bmatrix}.$$

6. $$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix} = \begin{bmatrix} d_1A_{11} & d_2A_{12} & d_3A_{13} \\ d_1A_{21} & d_2A_{22} & d_3A_{23} \\ d_1A_{31} & d_2A_{32} & d_3A_{33} \end{bmatrix}.$$

7. $$\begin{bmatrix} 1 & & \\ 3 & 2 & \\ -2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & & \\ -3 & 3 & \\ -2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & & \\ 0 & 6 & \\ -5 & 2 & 2 \end{bmatrix}.$$

8. $$\begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 2 \\ -1 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 5 & -5 \\ 4 & -8 & 9 \\ -2 & 5 & -6 \end{bmatrix}.$$

9. $$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ AA = A.$$

10. $$\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \begin{bmatrix} -1 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ -2 & 6 \\ 3 & -9 \end{bmatrix}, \ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ -2 & 1 & -3 \end{bmatrix}.$$

11. $$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \ AA = O.$$

12. $$\begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} -1 & \\ & -1 \end{bmatrix}, \ AA = -I.$$

In Chapter 1 we introduced the notions of a triangular system of equations, systems of type 1 and 0, systems of Hermite and echelon form, and the rank of a system. Obviously it is the particular form of the coefficient matrix only that is responsible for these distinctions. Now that the matrix has come into its own, it should be perfectly natural for us to speak about triangular matrices, matrices of type 0 and 1, Hermite and echelon matrices, and the rank of a matrix. One-dimensional arrays are too restricted to have such interesting forms.

The products $a^Tb = b^Ta$ of $a^T = a^T(1 \times n)$ and $b = b(n \times 1)$, or $b^T = b^T(1 \times n)$ and $a = a(n \times 1)$ is a (1×1) matrix, or scalar. But the product ab^T of $a = a(m \times 1)$ and $b^T = b^T(1 \times n)$ is an $(m \times n)$ matrix, as in previous example 10. Matrices $(ab^T)_{ij} = a_ib_j$, $(ba^T)_{ij} = b_ia_j$, with nonzero a and b are *rank-one* matrices. To verify this we write

$$ab^T = \begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ & & & \\ a_mb_1 & a_mb_2 & \cdots & a_mb_n \end{bmatrix} \tag{2.22}$$

and perform the elementary operations necessary to bring ab^T into reduced echelon form. Assume that $a_1 = 1$. If a_1 happens to be zero, then the rows of ab^T are interchanged to bring a nonzero entry of a to the top. Addition to each row of ab^T, $-a_i$ times the first row leaves us with

$$R = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \\ 0 & 0 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{2.23}$$

which is, since $b \neq o$, an echelon matrix of rank 1. In section 10 we shall show that the converse is also true, namely that every matrix A of rank one can be written as $A = ab^T$.

Matrix multiplication can be described in various ways in terms of the rows and columns of A and B. Let the columns of A, B, C be a_j, b_j, c_j, respectively, and the rows of A, B, C be ${a_i}^T, {b_i}^T, {c_i}^T$, respectively (notice that ${a_1}^T$ is *not* the transpose of a_1 here.)

We may write

$$A = [a_1 \ a_2 \ldots a_n] = [a_1 \ o \ o \ o] + [o \ a_2 \ o \ o] + \cdots + [o \ o \ o \ a_n] \tag{2.24}$$

or

$$A = a_1{e_1}^T + a_2{e_2}^T + \cdots + a_n{e_n}^T. \tag{2.25}$$

Similarly, in terms of the rows ${a_i}^T$

$$A = \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ \vdots \\ {a_m}^T \end{bmatrix} = \begin{bmatrix} {a_1}^T \\ o^T \\ o^T \\ o^T \end{bmatrix} + \begin{bmatrix} o^T \\ {a_2}^T \\ o^T \\ o^T \end{bmatrix} + \cdots + \begin{bmatrix} o^T \\ o^T \\ o^T \\ {a_m}^T \end{bmatrix} \tag{2.26}$$

or

$$A = e_1 {a_1}^T + e_2 {a_2}^T + \ldots + e_m {a_m}^T . \tag{2.27}$$

Let $A = A(m \times k)$ and $B = B(k \times n)$. Then in terms of eqs. (2.25) and (2.27) the product AB is written as

$$AB = (a_1 {e_1}^T + a_2 {e_2}^T + \cdots + a_k {e_k}^T)(e_1 {b_1}^T + e_2 {b_2}^T + \cdots + e_k {b_k}^T) \tag{2.28}$$

where $a_j = a_j(m \times 1)$, ${b_i}^T = {b_i}^T(1 \times n)$, $e_j = e_j(k \times 1)$ and ${e_i}^T = {e_i}^T(1 \times k)$. Since ${e_i}^T e_i = 1$ and ${e_i}^T e_j = 0$ if $i \neq j$, the product becomes

$$AB = AB(m \times n) = a_1 {b_1}^T + a_2 {b_2}^T + \cdots + a_k {b_k}^T \tag{2.29}$$

which is the sum of k rank-one matrices.

Otherwise we may write

$$AB = (e_1 {a_1}^T + e_2 {a_2}^T + \cdots e_m {a_m}^T)(b_1 {e_1}^T + b_2 {e_2}^T + \cdots b_n {e_n}^T) \tag{2.30}$$

where ${a_i}^T = {a_i}^T(1 \times k)$, $b_j = b_j(k \times 1)$, $e_j = e_j(m \times 1)$, and ${e_i}^T = {e_i}^T(1 \times n)$. The products ${a_i}^T b_j$ are scalars, while $e_i {e_j}^T$ are matrices with all entries equal to zero, except for one entry at the ith row and jth column that is 1,

$$e_2 {e_3}^T = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = [\, o \ o \ e_2 \,] . \tag{2.31}$$

Hence

$$AB = \begin{bmatrix} {a_1}^T b_1 & {a_1}^T b_2 & & {a_1}^T b_n \\ {a_2}^T b_1 & {a_2}^T b_2 & & {a_2}^T b_n \\ & & & \\ {a_m}^T b_1 & {a_m}^T b_2 & & {a_m}^T b_n \end{bmatrix}, \ (AB)_{ij} = {a_i}^T b_j , \tag{2.32}$$

and $(AB)_{ij}$ is the product of the ith *row* of A and the jth *column* of B.

Writing

$$AB = A(b_1 {e_1}^T + b_2 {e_2}^T + \cdots + b_n {e_n}^T) = (Ab_1){e_1}^T + (Ab_2){e_2}^T + \cdots + (Ab_n){e_n}^T \tag{2.33}$$

we have that

$$AB = [Ab_1 \ Ab_2 \ \cdots \ Ab_n] . \tag{2.34}$$

In terms of rows

$$AB = (e_1 {a_1}^T + e_2 {a_2}^T + \cdots + e_m a_m^T)B = e_1({a_1}^T B) + e_2({a_2}^T B) + \cdots + e_m(a_m^T B) \quad (2.35)$$

or

$$AB = \begin{bmatrix} {a_1}^T B \\ {a_2}^T B \\ \vdots \\ a_m^T B \end{bmatrix} \quad (2.36)$$

since ${e_i}^T A = {a_i}^T$ is the ith row of A.

Exercises

2.3.1. What is matrix A so that $Ax = -2x$ for any x?

2.3.2. Can matrix A be found so that $A(x + y) = x + 2y$ for any x and y?

2.3.3 For matrix

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

find α so that if $x_1^2 + x_2^2 = x_3^2$, then $y_1^2 + y_2^2 = y_3^2$, where $x = [x_1\ x_2\ x_3]^T$ and $y = Ax$.

2.3.4. Find scalars α, β, γ so that

$$\alpha \begin{bmatrix} 1 & -2 & 1 \\ -3 & 2 & 1 \\ -1 & -2 & 1 \end{bmatrix} + \beta \begin{bmatrix} -2 & 2 & 5 \\ 1 & 3 & -1 \\ -3 & 4 & -3 \end{bmatrix} + \gamma \begin{bmatrix} -1 & -2 & -3 \\ 4 & -2 & 1 \\ 2 & 2 & -3 \end{bmatrix}$$

is upper-triangular, but not zero.

2.3.5. For

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

write AB and BA.

2.3.6. Show that

$$\begin{bmatrix} 1 & 2 \\ -3 & -6 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 1 & -1 \end{bmatrix} = O.$$

2.3.7. Show that if $ABAB = O$, then $(BA)^3 = O$.

2.3.8. Show that if $(AB)^2 = \alpha(AB)$, and $BA = X$, then $X^2(X - \alpha I) = O$.

2.3.9. Show that if $AC = CB$, then A and B are square.

2.3.10. If $AB = BA$ and $CB = BC$, is it true that $CA = AC$?

2.3.11. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

be such that the row sum of A is α, $A_{11} + A_{12} = \alpha$ and $A_{21} + A_{22} = \alpha$, and the row sum of B is β, $B_{11} + B_{12} = \beta$ and $B_{21} + B_{22} = \beta$. Show that the row sum of AB is $\alpha\beta$.

2.3.12. Let A be such that $A_{ij} \geq 0$, and such that the row sums are all 1. A matrix having these properties is called *stochastic*. Show that the product of stochastic matrices is a stochastic matrix.

2.3.13. What does the interchange of rows i and j of A do to product AB? What does the interchange of columns i and j of B do to product AB?

2.3.14. Consider the matrix product $AB = C$. Add the elements in each *column* of A to have row vector a^T, and the elements in each *row* of B to similarly have column vector b. Show that $a^T b = \sum\limits_{i,j} C_{ij}$.

2.3.15. Find α and β so that

$$A = \begin{bmatrix} 1 & \alpha & \beta \\ & 1 & \alpha \\ & & 1 \end{bmatrix}, A^2 = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}.$$

2.3.16. Carry out the matrix multiplication

$$\begin{bmatrix} 1 & & \\ L_{21} & 1 & \\ L_{31} & L_{32} & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ L_{21} & 1 & \\ L_{31} & L_{32} & 1 \end{bmatrix}.$$

Is the product of lower-triangular matrices always a lower triangular matrix?

2.3.17. Show that for any lower triangular matrix $L = L(2 \times 2)$, $L_{21} \neq 0$, scalars α_0, α_1 exist so that $L^2 + \alpha_1 L + \alpha_0 I = O$. Write α_0 and α_1 in terms of L_{11}, L_{21}, L_{22}.

2.3.18. For matrix

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

find scalars $\alpha_0, \ \alpha_1$ so that $A^2 + \alpha_1 A + \alpha_0 I = O$.

2.3.19. Show that if

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

then $A^2 = 3A - I$. Use this to show further that $A^8 = \alpha A - \beta I$. Find α and β.

2.3.20. What are the conditions on α and β so that

$$X = \begin{bmatrix} -1 & \alpha \\ \beta & 0 \end{bmatrix}$$

solves the matrix polynomial equation $(X - 2I)(X + 3I) = X^2 + X - 6I = O$?

2.3.21. Show that for any given $A = A(2 \times 2)$, scalars α_0 and α_1 can be found so that

$$A^2 + \alpha_1 A + \alpha_0 I = O.$$

Write α_0 and α_1 in terms of $A_{11}, A_{12}, A_{21}, A_{22}$.

2.3.22. Let

$$L_1 = \begin{bmatrix} 0 & & \\ 1 & 3 & \\ -2 & -1 & 1 \end{bmatrix}, \ L_2 = \begin{bmatrix} 3 & & \\ -2 & 0 & \\ 1 & 3 & 2 \end{bmatrix}, \ L_3 = \begin{bmatrix} -3 & & \\ 1 & -2 & \\ 2 & 1 & 0 \end{bmatrix}.$$

Verify that

$$L_1 L_2 = \begin{bmatrix} 0 & & \\ -3 & 0 & \\ -3 & 3 & 2 \end{bmatrix}, \text{ and } \ L_1 L_2 L_3 = \begin{bmatrix} 0 & & \\ 9 & 0 & \\ 16 & -4 & 0 \end{bmatrix}.$$

2.3.23. Let

$$L = \begin{bmatrix} 0 & & \\ L_{21} & 0 & \\ L_{31} & L_{32} & 0 \end{bmatrix}.$$

Show that

$$L^2 = \begin{bmatrix} 0 & & \\ 0 & 0 & \\ L_{21}L_{32} & 0 & 0 \end{bmatrix} \text{ and } \ L^3 = O.$$

Matrix A for which $A^m = O$, $A^{m-1} \neq O$, m being a positive integer, is *nilpotent* of index m. When $m = 2$ the index is omitted.

2.3.24. For

$$N = \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ & & & & \end{bmatrix}$$

compute $N_2 = N^2,\ N_3 = N^3,\ N_4 = N^4,\ N_5 = N^5$.

2.3.25. Write all nonzero 2×2 nilpotent matrices.

2.3.26. Show that

$$N = \begin{bmatrix} & \alpha & \\ \beta & & -\alpha \\ & \beta & \end{bmatrix}$$

is nilpotent of index 3, $N^3 = O$.

2.3.27. Show that

$$\begin{bmatrix} & \alpha & \\ & & 1 \\ 1 & & \end{bmatrix}^3 = \begin{bmatrix} \alpha & & \\ & \alpha & \\ & & \alpha \end{bmatrix}.$$

2.3.28. For positive integer m compute $A^m,\ B^m,\ C^m$ and $(AB)^m$,

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

2.3.29. Compute

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^m, \begin{bmatrix} \alpha & & \\ & \beta & \\ & & \gamma \end{bmatrix}^m, \begin{bmatrix} & & \alpha \\ & \beta & \\ \gamma & & \end{bmatrix}^m, \begin{bmatrix} \alpha & 1 & \\ & \alpha & 1 \\ & & \alpha \end{bmatrix}^m$$

for positive integer m.

2.3.30. Find the lowest positive integer value of m so that

$$\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}^m + \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}^m = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^m$$

holds.

2.3.31. Let $\alpha = \alpha(\xi)$, $\beta = \beta(\xi)$ be functions of ξ, with derivatives α', α'' and β', β''. Show that if

$$A(\alpha) = \begin{bmatrix} \alpha & & \\ \alpha' & \alpha & \\ \frac{1}{2}\alpha'' & \alpha' & \alpha \end{bmatrix}$$

then $A(\alpha)\ A(\beta) = A(\alpha\beta)$.

2.3.32. If $A = ab^T$, what are the conditions on a and b so that $A^2 = \alpha A$?

2.3.33. Let $A = I - uv^T, u^T v = v^T u = \alpha$. What is α so that $A^2 = I$? $A^3 = I$? Show that if $A = I - uv^T, u^T v = 0$, then, $A^3 = I - 3uv^T$.

2.3.34. If $A = I - uu^T$ and $u^T u = 1$, show that $A^2 = A$. Such a matrix is called *idempotent*.

2.3.35. Show that if $A^2 = A$, then $(I - A)(I + A) = I - A$.

2.3.36. Show that if $AB = A$ and $BA = B$, then A and B are idempotent, that is, $A^2 = A, B^2 = B$.

2.3.37. Find the relationships between $\alpha, \beta, \gamma, \delta$ so that $A^2 = A$,

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$

2.3.38. Let

$$P = I - uu^T - vv^T$$

be with $u^T u = v^T v = 1$ and $u^T v = 0$. Show that $P^2 = P$.

2.3.39. Let

$$P = \frac{1}{1 - \alpha^2}(vv^T - \alpha vu^T),\ \alpha = u^T v$$

be with $u^T u = v^T v = 1$. Show that $P^2 = P$.

2.3.40. What are the conditions on u and v so that $(I + uv^T)^2 = I$.

2.3.41. If $A = I - \alpha uu^T, u^T u = 1$, for what α is $A^3 = A$.

2.3.42. Let u be a vector so that $u^T u = 1$. Find α and β so that $(\alpha I + \beta uu^T)^2 = 4I$.

2.3.43. Write all 2×2 matrices such that $A^2 = I$.

2.3.44. Write all matrices that commute with

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

namely all X so that $AX = XA$.

2.3.45. Show that all matrices X that commute with

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

that is, such that $AX = XA$, are of the form

$$X = \alpha \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} + \beta \begin{bmatrix} & 1 \\ -2 & -2 \end{bmatrix}$$

where α and β are arbitrary. Explain why specific α and β exist so that $A = X$. Show subsequently that $X = \alpha' I + \beta' A$. Explain in turn why specific α' and β' exist so that $A^2 = X = \alpha' I + \beta' A$.

2.3.46. Write all matrices X that commute with

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

2.3.47. For $A = I + 2uu^T$, $u^T u = 1$, find β in $B = I + \beta uu^T$ so that $AB = I$.

2.3.48. For $A = I + uu^T + vv^T$, $u^T u = v^T v = 1$, $u^T v = v^T u = 0$, find α and β in $B = I + \alpha uu^T + \beta vv^T$ so that $AB = I$.

2.3.49. What is α if $A = I + \alpha B$ and $B = I + \alpha A$?

2.3.50. Show that if A and B commute with C, then $I + \alpha A + \beta B$ commutes with C.

2.3.51. Prove that if $AB = BA$, then also $A^m B = BA^m$ for any positive integer m.

2.3.52. Solve the matrix equation

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

2.3.53. Show that if $A = I + S$ and $B = I - S$, then $AB = BA$

2.3.54. Show that if $AX = XA$ for any X, then $A = \alpha I$. Can it ever happen that matrix equation $AX = XA$ has a unique solution X?

2.3.55. Show that if $AD = DA$, D being a diagonal matrix and such that $D_{ii} \neq D_{jj}$ $i \neq j$, then A is diagonal.

2.3.56. Find matrix P so that

$$\begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{bmatrix} = \alpha I + \beta P + \gamma P^2.$$

What is P^3?

2.3.57. Write all matrices of the form

$$A = \begin{bmatrix} \alpha & \beta & \gamma \\ \beta & \alpha & \beta \\ \gamma & \beta & \alpha \end{bmatrix}, A \neq I$$

so that $A^2 = I$. Hint: write A as in the previous exercise.

2.3.58. Find matrix P so that

$$\begin{bmatrix} \alpha & \beta & \gamma \\ & \alpha & \beta \\ & & \alpha \end{bmatrix} = \alpha I + \beta P + \gamma P^2.$$

What is P^3?

2.3.59. Matrix

$$T = \begin{bmatrix} \alpha & \beta & \gamma & \delta \\ & \alpha & \beta & \gamma \\ & & \alpha & \beta \\ & & & \alpha \end{bmatrix}$$

is an upper-triangular *Töpliz* matrix. Is the product of two such Töpliz matrices a Töpliz matrix? Do Töpliz matrices commute? Use the result of the previous exercise.

2.3.60. Matrix

$$A = \begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{bmatrix}$$

is a 3×3 *circulant* matrix. Use the result of problem 2.3.56 to show that the product of two circulant matrices is a circulant matrix, and that they commute.

2.3.61. Show that, for circulant matrix

$$C = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}, \beta \neq 0$$

scalars c_1 and c_2 exist so that

$$C^2 + c_1 C + c_0 I = O.$$

Write c_1 and c_0 in terms of α and β. Extend the result to circulant $C(3 \times 3)$ and a matrix polynomial equation of degree 3.

2.3.62. Show that if $A = A(2 \times 2)$, and $A^3 = O$, then $A^2 = O$.

2.3.63. If $A^2 = O$ and trace $(A) = A_{11} + A_{22} + \cdots + A_{nn} = 0$, does that imply that $A = O$?

2.3.64. Show that no square A and B exist so that $AB - BA = I$.

2.3.65. Prove that if $A + B = O$ and $A^2 + B^2 = O$, then $A^2 = B^2 = O$.

2.3.66. Let A and B be idempotent. Prove that $AB + BA = O$ implies that $AB + (AB)^2 = O$, $BA + (BA)^2 = O$, and $(AB)^2 + (BA)^2 = O$.

2.3.67. Prove that if $A = A^2$ and $B = B^2$, then $A + B = (A + B)^2$ if and only if $AB = BA = O$.

2.4 Matrix transposition-symmetry

Matrices

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = [a_1 \ \ a_2] \quad \text{and} \quad A^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \end{bmatrix} = \begin{bmatrix} {a_1}^T \\ {a_2}^T \end{bmatrix} \tag{2.37}$$

where ${a_j}^T$ is truly the transpose of a_j, are the transpose of each other. Obviously if $A = A(m \times n)$, then $A^T = A^T(n \times m)$, and $(A^T)_{ij} = A_{ji}$.

Matrix A can equal its transpose only when square, and when this equality occurs the matrix is termed *symmetric*.

Definition. *If $A = A^T$, $A_{ij} = A_{ji}$, then matrix A is said to be symmetric.*

Symmetric matrices are common in the application of linear algebra to mathematical physics (nature is symmetrical), and they have outstanding properties that we shall discuss in the coming chapters.

Lemma 2.8.

1. $(x^T)^T = x$.
2. $(x + y)^T = x^T + y^T$.
3. $x^T y = y^T x$.
4. $(Ax)^T = x^T A^T$.
5. $(x^T A)^T = A^T x$.

Proof.

1. Obvious.
2. A direct consequence of the definitions of $x = y$ and $x^T + y^T$.
3. $x^T y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = y_1 x_1 + y_2 x_2 + \cdots + y_n x_n = y^T x$.

4. Let ${a_i}^T$ be the ith row of A. Then $(Ax)_i = {a_i}^T x$, $(Ax)_j^T = {a_j}^T x$. But $(x^T A^T)_j = x^T a_j$, where $a_j = ({a_j}^T)^T$, and hence since ${a_j}^T x = x^T a_j$, $(Ax)^T = x^T A^T$.

5. Same as 4.

End of proof.

Theorem 2.9.

1. $(A^T)^T = A$.
2. $(A + B)^T = A^T + B^T$.
3. $(\alpha A)^T = \alpha A^T$.
4. $(AB)^T = B^T A^T$.
5. $(A^m)^T = (A^T)^m \quad m$ integer > 0.

Proof. By Lemma 2.8,

1. $(A^T)^T x = (x^T A^T)^T = ((Ax)^T)^T = Ax$.

2. $(A+B)^T x = (x^T(A+B))^T = (x^T A + x^T B)^T$.

$$= (x^T A)^T + (x^T B)^T.$$

$$= A^T x + B^T x = (A^T + B^T)x.$$

3. Obvious.

4. $(AB)^T x = (x^T(AB))^T = ((x^T A)B)^T = B^T(x^T A)^T = B^T A^T x$.

5. A corollary of 4 when $A = B$. End of proof.

Theorem 2.9 implies that if A and B are symmetric then so is $A+B$, so is αA and so is A^m, where m is a positive integer. Statement 4, on the other hand, implies that *the product of two symmetric matrices is not symmetric* unless the matrices commute. Indeed if $A = A^T$ and $B = B^T$, then $(AB)^T = B^T A^T = BA \neq AB$. Notice that in previous example 8 both A and B are symmetric but AB is not.

We verify that

$$R = P^T AP \quad \text{and} \quad S = AB + BA \tag{2.38}$$

are symmetric if A and B are symmetric, and that

$$R = A^T A \quad \text{and} \quad S = A + A^T \tag{2.39}$$

are symmetric for any A.

The triple product $P^T AQ$, with $P = P(n \times m)$, $A = A(n \times n)$, and $Q = Q(n \times m)$

$$P^T AQ = \begin{bmatrix} {p_1}^T \\ {p_2}^T \\ \vdots \\ {p_m}^T \end{bmatrix} [\, Aq_1 \;\; Aq_2 \qquad Aq_m \,] = \begin{bmatrix} {p_1}^T Aq_1 & {p_1}^T Aq_2 & & {p_1}^T Aq_m \\ {p_2}^T Aq_1 & & & \\ & & & \\ {p_m}^T Aq_1 & & & {p_m}^T Aq_m \end{bmatrix} \tag{2.40}$$

or $(P^T AQ)_{ij} = {p_i}^T Aq_j$ is important. When $P = Q$, and $A = A^T$

$$(P^T AP)_{ij} = {p_i}^T Ap_j = {p_j}^T Ap_i. \tag{2.41}$$

Matrix A for which it is true that $A = -A^T$ is *skew symmetric*, and we readily verify that if A is skew symmetric, then $A_{ii} = 0$ for all i. Matrix $B = A - A^T$ is skew symmetric since $B^T = A^T - (A^T)^T = A^T - A = -B$.

Exercises

2.4.1. Show that if A and B are symmetric, then so is $C = ABA$.

2.4.2. Matrix A is symmetric if $A = A^T$, it is skew symmetric if $A = -A^T$. If matrix A is skew symmetric is matrix $B = A^2$ symmetric or skew symmetric? What about matrix $C = A^3$? Hint: consider B^T and C^T.

2.4.3. Let $A = -A^T, B = B^T$. Is matrix $C = ABA$ symmetric or skew symmetric? Let $A = A^T, B = -B^T$. Is matrix $C = ABA$ symmetric or skew symmetric?

2.4.4. Show that if $A^T A = AA^T = I$ and $B^T B = BB^T = I$, then $C = AB$ is such that $C^T C = CC^T = I$.

2.4.5. Find A if

$$A^T A = AA^T = \begin{bmatrix} 1 & \\ & 4 \end{bmatrix}.$$

2.4.6. For

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

find X so that $XA = A^T$.

2.4.7. Show that if A is nilpotent, $A^2 = O$, then so is A^T, and if A is idempotent,$A^2 = A$, then so is A^T.

2.4.8. Show that for

$$A = \begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{bmatrix},$$

$AA^T = I$ if $\alpha^2 + \beta^2 + \gamma^2 = 1$, and $\alpha\beta + \beta\gamma + \gamma\alpha = 0$.

2.4.9. Show that if $A = A^T$, then $A^2 = O$ implies that $A = O$.

2.4.10. Prove that if A and B are such that $A = A^T,\ B = B^T$ and $(A - B)^2 = O$, then $A = B$.

2.4.11. Show that if $A = A^T,\ B = B^T$ and $A^2 + B^2 = O$, then $A = B = O$.

2.4.12. Show that if X satisfies $AX = XA, A = A^T$,then so does X^T. Say that $A = A(2 \times 2) = A^T$. When does it happen that $X = X^T$?

2.4.13. Consider matrix equation $XD = DX$ where D is diagonal. Under what conditions on D is X diagonal?

2.4.14. Consider matrix equation $XD = D'X$ where D and D' are diagonal. Under what conditions on D and D' is $X = O$ the only solution?

2.4.15. Show that if tridiagonal matrix

$$T = \begin{bmatrix} a_1 & b_2 & & \\ c_2 & a_2 & b_3 & \\ & c_3 & a_3 & b_4 \\ & & c_4 & a_4 \end{bmatrix}$$

is irreducible, i,e. $b_i \neq 0$ and $c_i \neq 0$ for all i, then a diagonal matrix $D, D_{ii} = d_i \neq 0, D_{11} = 1$, exists so that DT is symmetric, $DT = (DT)^T$. Verify further that if symmetric tridiagonal T is irreducible, then a diagonal matrix $D, D_{ii} = d_i \neq 0, D_{11} = 1$, exists that accomplishes the reduction

$$\begin{bmatrix} 1 & & & \\ & d_2 & & \\ & & d_3 & \\ & & & d_4 \end{bmatrix} \begin{bmatrix} a_1 & b_2 & & \\ b_2 & a_2 & b_3 & \\ & b_3 & a_3 & b_4 \\ & & b_4 & a_4 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & d_2 & & \\ & & d_3 & \\ & & & d_4 \end{bmatrix} = \begin{bmatrix} a'_1 & 1 & & \\ 1 & a'_2 & 1 & \\ & 1 & a'_3 & 1 \\ & & 1 & a'_4 \end{bmatrix}.$$

2.4.16. If A is symmetric, $A = A^T$, and $B^2 = A$, does it imply that $B = B^T$? Write all $B = B(2 \times 2)$ so that $B^2 = I, B \neq \pm I$.

2.4.17. Let $A_i = A_i^T$, and so that $A_1 + A_2 + \cdots + A_m = I$. Show that if $A_i^2 = A_i$, then $A_i A_j = O \quad i \neq j$.

2.5 Partitioned matrices

It may be practically expedient or theoretically desirable to divide, or *partition*, a vector into *sub-vectors*, each with its own designating letter, say

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} f \\ - \\ g \end{bmatrix}. \tag{2.42}$$

Vector a in eq. (2.42) can be considered as having *vector* components.

We readily verify that if a and b are sub-vectors, then

$$\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} a + a' \\ b + b' \end{bmatrix}, \text{ and } \alpha \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \alpha a \\ \alpha b \end{bmatrix} \tag{2.43}$$

provided that the vectors are partioned *conformally*, so that the subvector additions $a + a'$ and $b + b'$ are meaningful.

Matrices may be similarly partitioned. For example

$$\left[\begin{array}{cc|ccc} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ \hline A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ A_{41} & A_{42} & A_{43} & A_{44} & A_{45} \end{array}\right] = \begin{bmatrix} E & F \\ & \\ G & H \end{bmatrix} \tag{2.44}$$

where $E = E(1 \times 2),\ F = F(1 \times 3),\ G = G(3 \times 2),\ H = H(3 \times 3)$ are four *submatrices* of matrix A.

Definition. *Square submatrix S is a diagonal (principal) submatrix of square matrix A if $S_{11} = A_{kk}$ for some k. Submatrix S is a leading diagonal submatrix of A if $S_{11} = A_{11}$.*

What makes partitioning so interesting is that the rules of matrix addition and multiplication remain valid when submatrices are instituted for scalar entries, provided addition and multiplication are matrix operations.

Theorem 2.10. *Let x, A and A' be conformally portioned. Then*

1. $$Ax = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} Ea + Fb \\ Ga + Hb \end{bmatrix}.$$

2. $$\alpha A = \begin{bmatrix} \alpha E & \alpha F \\ \alpha G & \alpha H \end{bmatrix}.$$

3. $$A^T = \begin{bmatrix} E^T & G^T \\ F^T & H^T \end{bmatrix}.$$

4. $$A + A' = \begin{bmatrix} E & F \\ G & H \end{bmatrix} + \begin{bmatrix} E' & F' \\ G' & H' \end{bmatrix} = \begin{bmatrix} E + E' & F + F' \\ G + G' & H + H' \end{bmatrix}.$$

5. $$AA' = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} E' & F' \\ G' & H' \end{bmatrix} = \begin{bmatrix} EE' + FG' & EF' + FH' \\ GE' + HG' & GF' + HH' \end{bmatrix}.$$

Proof. Statement 1 is a direct consequence of the definitions of matrix vector product and vector addition. To prove statement 5 we partition x as $x = [z^T o^T]^T$ and have that

$$AA'x = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} E'z \\ G'z \end{bmatrix} = \begin{bmatrix} EE'z + FG'z \\ GE'z + HG'z \end{bmatrix} = \begin{bmatrix} EE' + FG' \\ GE' + HG' \end{bmatrix} z \tag{2.45}$$

which establishes the first submatrix column of AA'. The second column is found with $x = [o^T z^T]^T$. The rest is left for an exercise. End of proof.

Exercises

2.5.1. If

$$A = \begin{bmatrix} I & O \\ B & -I \end{bmatrix}$$

what is A^2?

2.5.2. For $A = \begin{bmatrix} B & C \\ & I \end{bmatrix}$, compute $A^2, A^3, \ldots, A^n$.

2.5.3. Fix submatrices X and Y so that

$$\begin{bmatrix} I & X \\ & I \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} I & Y \\ & I \end{bmatrix} = \begin{bmatrix} A + B & \\ B & A - B \end{bmatrix}.$$

2.5.4. Fix submatrix X to render

$$\begin{bmatrix} I & \\ X & I \end{bmatrix} \begin{bmatrix} I & A \\ A^T & I \end{bmatrix} \begin{bmatrix} I & X^T \\ & I \end{bmatrix}$$

block diagonal.

2.6 Elementary matrices

Whenever possible, every matrix manipulation should be expressed in terms of the defined operations of addition and multiplication so that these manipulations may be included in the algebra. In this section we shall translate the elementary operations of the first chapter into matrix multiplications.

For vector x, an elementary operation consists of:

1. Multiplication of component x_i by α, $\alpha \neq 0$.

2. Addition to component x_i, α times component x_j.

3. Interchange of two components.

A square matrix that consistently performs one of these operations for *arbitrary* vector x, is an *elementary matrix*. Letter E is usually reserved to denote an elementary matrix. We shall consider at once the generalization of operations 1 and 2, whereby the ith component of x is replaced by a linear combination of all components,

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ & & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3' \\ x_4 \end{bmatrix}, \quad \begin{matrix} x_3' = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ \alpha_3 \neq 0 \end{matrix} \tag{2.46}$$

which we divide into the three particular cases

$$\begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ \alpha x_2 \\ x_3 \end{bmatrix} \quad \alpha \neq 0,$$

$$\begin{bmatrix} 1 & & \\ & 1 & \\ \alpha & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 + \alpha x_1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \alpha \\ & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + \alpha x_3 \\ x_2 \\ x_3 \end{bmatrix}, \tag{2.47}$$

done lastly with matrix E that can be written as

$$E = I + \alpha e_i {e_j}^T \tag{2.48}$$

and if $i > j$, E is *lower-triangular*, while if $i < j$, E is *upper-triangular*.

By the fact that ${e_i}^T x = x_i$, we write the *permutation* matrix that interchanges, say, components two and three of x as

$$P = e_1 {e_1}^T + e_2 {e_3}^T + e_3 {e_2}^T = [e_1 \, e_3 \, e_2] \tag{2.49}$$

so that, in this example

$$Px = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix}. \tag{2.50}$$

Permutation matrix P for the interchange of just two entries of x is symmetric, $P = P^T$, and we also have the interesting result that

$$PP = P^2 = I \tag{2.51}$$

which is the algebraic statement to the effect that repeated permutations return the entries of x to their natural order. Permutation matrix P that is the product of several successive permutations is not necessarily symmetric. For example, if matrix P_1 interchanges entries 2 and 3, and P_2 subsequently interchanges entries 3 and 4, then their successive application produces the compound permutation matrix

$$\begin{aligned} P = P_2 P_1 &= (e_1{e_1}^T + e_2{e_2}^T + e_3{e_4}^T + e_4{e_3}^T)(e_1{e_1}^T + e_2{e_3}^T + e_3{e_2}^T + e_4{e_4}^T) \\ &= e_1{e_1}^T + e_2{e_3}^T + e_3{e_4}^T + e_4{e_2}^T = [e_1 \ e_4 \ e_2 \ e_3] \end{aligned} \tag{2.52}$$

or

$$P = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & & 1 \\ & & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & & 1 & \\ & & & 1 \\ & 1 & & \end{bmatrix} \tag{2.53}$$

which is clearly nonsymmetric. However, if $P = P_k \cdots P_2 P_1$, then $P^T = P_1 P_2 \cdots P_k$, since ${P_j}^T = P_j$, and the compound permutation matrix P is such that

$$P^T P = PP^T = (P_k \cdots P_2 P_1)(P_1 P_2 \cdots P_k) = I. \tag{2.54}$$

Pre-multiplication of A by E, EA, causes every *column* of A to change according to eqs.(2.47) and (2.50), and we name this an elementary *row* operation on A,

$$\begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix} A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ \alpha A_{21} & \alpha A_{22} & \alpha A_{23} & \alpha A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{bmatrix} = \begin{bmatrix} {a_1}^T \\ \alpha {a_2}^T \\ {a_3}^T \end{bmatrix},$$

$$\begin{bmatrix} 1 & & \\ & 1 & \\ \alpha & & 1 \end{bmatrix} A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} + \alpha A_{11} & A_{32} + \alpha A_{12} & A_{33} + \alpha A_{13} & A_{34} + \alpha A_{14} \end{bmatrix} = \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ {a_3}^T + \alpha {a_1}^T \end{bmatrix}$$

$$\begin{bmatrix} 1 & & \alpha \\ & 1 & \\ & & 1 \end{bmatrix} A = \begin{bmatrix} A_{11} + \alpha A_{31} & A_{12} + \alpha A_{32} & A_{13} + \alpha A_{33} & A_{14} + \alpha A_{34} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{bmatrix} = \begin{bmatrix} {a_1}^T + \alpha {a_3}^T \\ {a_2}^T \\ {a_3}^T \end{bmatrix}$$

$$
\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix} A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{21} & A_{22} & A_{23} & A_{24} \end{bmatrix} = \begin{bmatrix} a_1{}^T \\ a_3{}^T \\ a_2{}^T \end{bmatrix}. \tag{2.55}
$$

Post-multiplication of A by an elementary matrix E, AE, is, in the same way, an elementary *column* operation on A,

$$
A \begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix} = \begin{bmatrix} A_{11} & \alpha A_{12} & A_{13} \\ A_{21} & \alpha A_{22} & A_{23} \\ A_{31} & \alpha A_{32} & A_{33} \\ A_{41} & \alpha A_{42} & A_{43} \end{bmatrix} = [\, a_1 \ \alpha a_2 \ a_3 \,],
$$

$$
A \begin{bmatrix} 1 & & \alpha \\ & 1 & \\ & & 1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} + \alpha A_{11} \\ A_{21} & A_{22} & A_{23} + \alpha A_{21} \\ A_{31} & A_{32} & A_{33} + \alpha A_{31} \\ A_{41} & A_{42} & A_{43} + \alpha A_{41} \end{bmatrix} = [\, a_1 \ a_2 \ a_3 + \alpha a_1 \,],
$$

$$
A \begin{bmatrix} 1 & & \\ & 1 & \\ \alpha & & 1 \end{bmatrix} = \begin{bmatrix} A_{11} + \alpha A_{13} & A_{12} & A_{13} \\ A_{21} + \alpha A_{23} & A_{22} & A_{23} \\ A_{31} + \alpha A_{33} & A_{32} & A_{33} \\ A_{41} + \alpha A_{43} & A_{42} & A_{43} \end{bmatrix} = [\, a_1 + \alpha a_3 \ a_2 \ a_3 \,],
$$

$$
A \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix} = \begin{bmatrix} A_{11} & A_{13} & A_{12} \\ A_{21} & A_{23} & A_{22} \\ A_{31} & A_{33} & A_{32} \\ A_{41} & A_{43} & A_{42} \end{bmatrix} = [\, a_1 \ a_3 \ a_2 \,]. \tag{2.56}
$$

Notice that if in the second row operation in eq. (2.55) is EA, then the corresponding column operation in eq. (2.56) is AE^T. Operation EAE^T on $A = A^T$ is a symmetric row and column operation. A formal proof that EAE^T is symmetric is given as

$$
(EAE^T)^T = (E^T)^T A^T E^T = EAE^T. \tag{2.57}
$$

Definition. *If E is an elementary matrix, then A and EA are row-equivalent matrices. Matrices A and AE are column-equivalent.*

Elementary matrices are a mere notational device. It is senseless to actually write down E and carry out the multiplication EA. What is theoretically important is to know that elementary row and column operations *can* be carried out by matrix multiplications. From the associative law of matrix multiplication we have the useful result that $E(AB) = (EA)B$, and $(AB)E = A(BE)$.

If no row interchanges are required, then forward elimination that turns $A = A(3 \times 3)$ into a row-equivalent upper-triangular matrix U is expressed as

$$E_3E_2E_1A = U \tag{2.58}$$

where

$$E_1 = \begin{bmatrix} 1 & & \\ \alpha_1 & 1 & \\ & & 1 \end{bmatrix}, \; E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ \alpha_2 & & 1 \end{bmatrix}, \; E_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & \alpha_3 & 1 \end{bmatrix} \tag{2.59}$$

and the compound total elementary operation is

$$E = E_3E_2E_1 = \begin{bmatrix} 1 & & \\ \alpha_1 & 1 & \\ \alpha_2 + \alpha_1\alpha_3 & \alpha_3 & 1 \end{bmatrix}. \tag{2.60}$$

More striking is the fact that

$$E_1E_2E_3 = \begin{bmatrix} 1 & & \\ \alpha_1 & 1 & \\ \alpha_2 & \alpha_3 & 1 \end{bmatrix}. \tag{2.61}$$

As a demonstration of the reasoning power that matrix algebra grants us, we algebraically demonstrate that row interchange is not an independent operation, but can be achieved through a succession of the first and second elementary operations of eq. (2.55). Indeed

$$\begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ & 1 & -1 \\ & & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}. \tag{2.62}$$

Compound elementary operation matrix E

$$\begin{bmatrix} 1 & & \\ \alpha_2 & 1 & \\ \alpha_3 & & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 + \alpha_2x_1 \\ x_3 + \alpha_3x_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \quad x_1 \neq 0 \tag{2.63}$$

clears all the entries in a column below the pivot at once. If P is a permutation matrix that swaps entries *below* the pivot, then there exists an E' such that

$$PE = E'P \tag{2.64}$$

where E' is an elementary matrix of the same form as E, $E' = PEP$. For instance

$$\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}\begin{bmatrix} 1 & & \\ \alpha_2 & 1 & \\ \alpha_3 & & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix} = \begin{bmatrix} 1 & & \\ \alpha_3 & 1 & \\ \alpha_2 & & 1 \end{bmatrix}. \tag{2.65}$$

We raise this issue here because it permits us to give a matrix proof to Theorem 1.22.

If the elementary operations to create an upper-triangular matrix are interspersed with permutations, then the permutation matrices can be migrated to the front of the sequence. Consider

$$\cdots E_3PE_2E_1A = U \tag{2.66}$$

where E_j is lower-triangular. The first permutation matrix interchanges equations below the second, and hence there are lower-triangular E_1' and E_2' such that

$$\cdots E_3PE_2E_1A = \cdots E_3{E_2}'PE_1A = \cdots E_3{E_2}'{E_1}'PA \tag{2.67}$$

and P is at the head of the sequence. Every other P can be lead this way to the front and

$$EPA = U \tag{2.68}$$

where E is lower-triangular and U upper-triangular.

When system $Ax = o$ is of rank r we shall say that matrix A is of rank r. We should, however, be more precise and say that the *row* rank of A is r, since matrix A in system $Ax = o$ is brought to Hermite form by means of row operations only. Now that matrices have assumed an independent existence, elementary operations are performed exclusively on them rather than on the system of equations; and for a matrix rows have no ascendency over columns. We must now give equal consideration to the *column* rank of A as to the *row* rank of A. Obviously the column rank of A equals the row rank of A^T. According to Theorem 1.16, row rank A^T equals row rank A, and hence *row rank A equals column rank A*, which is rank A.

General elementary row operation (2.46) applied to matrix A transforms it into

$$\begin{aligned} EA = \begin{bmatrix} {e_1}^T \\ {e_2}^T \\ y^T \\ {e_4}^T \end{bmatrix} A = \begin{bmatrix} 1 & & & \\ & 1 & & \\ y_1 & y_2 & y_3 & y_4 \\ & & & 1 \end{bmatrix} \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ {a_3}^T \\ {a_4}^T \end{bmatrix} &= \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ y_1{a_1}^T + y_2{a_2}^T + y_3{a_3}^T + y_4{a_4}^T \\ {a_4}^T \end{bmatrix} \\ &= \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ y^{\,T}A \\ {a_4}^T \end{bmatrix}, y_3 \neq 0 \end{aligned} \tag{2.69}$$

in which the third row of A is replaced by a linear combination of all rows of A. If $A^Ty = o$ and $y_3 \neq 0$, then since $y^TA = (A^Ty)^T$, the third row of EA becomes zero and is cleared. If the nontrivial solution to $A^Ty = o$ includes two independent unknowns, for instance if y_4 is also arbitrary, then we may set $y_4 \neq 0$ and create another elementary operation matrix E' through the replacement of the 4th row of I by y^T, with which we have

$$E'(EA) = \begin{bmatrix} {a_1}^T \\ {a_2}^T \\ o^T \\ o^T \end{bmatrix} . \tag{2.70}$$

No more rows of A can be cleared by elementary row operations, the row rank of A being two. If for any $y \neq o$, a linear combination of the rows of A does not vanish, $A^Ty \neq o$, then A is of full row rank. Matrix A is of full row rank if and only if $A^Ty = o$ has the sole solution $y = o$.

Clearly, the rows of A may be prearranged so that the rows wiped out by the elementary row operations are all at the bottom. In other words, there exists a permutation matrix P and a sequence of general elementary matrices as in eq. (2.69) so that

$$E_k \cdots E_2E_1PA = EPA = \begin{bmatrix} A' \\ O \end{bmatrix} \tag{2.71}$$

where A' is of full row rank. The number of (nonzero) rows in A' equals the (row) rank r of matrix A. In the same way

$$E'_m \cdots {E_2}'{E_1}'P'A^T = E'P'A^T = \begin{bmatrix} A'' \\ O \end{bmatrix} \tag{2.72}$$

which upon transposition becomes

$$AP'^TE'^T = [A''^T O] \tag{2.73}$$

and what we have for the rows of A we also have for the columns. Submatrix A'' is of full row rank equaling the row rank r of A^T. Submatrix A''^T is of full column rank equaling the column rank r of A. Combination of these row and column elementary operations leads to

Theorem 2.11. *For every matrix A of rank r there exist permutation matrices P and P', and compound row and column elementary operation matrices E and E' so that*

$$EPAP'E' = \begin{bmatrix} R & O \\ O & O \end{bmatrix} \tag{2.74}$$

where $R = R(r \times r)$ is of full rank r.

Only a square matrix can have full row rank and full column rank, and when this happens the matrix is of a distinguished nature as we shall see in the next two sections.

Exercises

2.6.1. What is the rank of

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

2.6.2. Matrix $A = A(2 \times n)$ has rows a_1^T and a_2^T such that $a_1^T a_1 = a_2^T a_2 = 1, a_1^T a_2 = a_2^T a_1 = 0$. Show that A is of full row rank. Hint: consider replacing a row by $\alpha_1 a_1^T + \alpha_2 a_2^T = o$.

2.6.3. Prove that if $N = N(n \times n)$ is a strictly upper-triangular nilpotent matrix,

$$N = \begin{bmatrix} 0 & N_{12} & N_{13} & N_{14} \\ & 0 & N_{23} & N_{24} \\ & & 0 & N_{34} \\ & & & 0 \end{bmatrix}$$

Then rank(N^2) <rank(N). Hint: Consider the eventualities of $N_{34} = 1$ and $N_{34} = 0$.

2.6.4. Prove that rank(A)= rank($A^T A$)= rank(AA^T).

2.6.5. Prove that the rank of the sum of m, rank one, matrices cannot exceed m.

2.6.6. Write all six (3×3) permutation matrices. Start with $P_1 = [e_1 \ e_2 \ e_3]$.

Matrix

$$A = \begin{bmatrix} 5 & 1 & 3 \\ 4 & 2 & 3 \\ 0 & 6 & 3 \end{bmatrix}$$

has this property that all its coefficients are non-negative, and all its row and column sums are equal, here to 9. Show that A may be written as the linear combination of the six permutation matrices,

$$A = \alpha_1 P_1 + \alpha_2 P_2 + \cdots + \alpha_6 P_6.$$

Find $\alpha_1, \alpha_2, \ldots, \alpha_6$.

2.6.7. Show that to every diagonal matrix D of type 0 there corresponds a permutation matrix P, $P^2 = P$, so that PDP may be written as the product of two nilpotent matrices,

$$P\begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix}P = \begin{bmatrix} 0 & & \\ 1 & 0 & \\ & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & \alpha & \\ & 0 & \beta \\ & & 0 \end{bmatrix}.$$

Write also the nilpotent factorizations

$$\begin{bmatrix} 2 & 1 & \\ -1 & 1 & \\ & & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & \gamma \\ & 0 & 1 \\ & & 0 \end{bmatrix}\begin{bmatrix} 0 & & \\ \alpha' & 0 & \\ \gamma' & \beta' & 0 \end{bmatrix},$$

and

$$\begin{bmatrix} L_{11} & L_{12} & \\ & L_{22} & \\ & & 0 \end{bmatrix} = \begin{bmatrix} 0 & \alpha & \gamma \\ & 0 & \beta \\ & & 0 \end{bmatrix}\begin{bmatrix} 0 & & \\ 1 & 0 & \\ & 1 & 0 \end{bmatrix}.$$

2.7 Right and left inverse matrices

Matrix algebra is taking shape and assuming a robust life of its own dissociated from systems of linear equations. Addition and multiplication of matrices are defined, the zero and identity matrices O and I are found, and the associative and distributive rules are confirmed. All that is left for us to do is explore and discover the multiplicative inverse. Because matrix multiplication is noncommutative, the twin questions of the existence and the uniqueness of an inverse requires careful examination.

A multiplicative inverse B to A is such that $BA = I$, or $AB = I$. The existence of B satisfying $BA = I$ would allow us to solve the linear system $Ax = f$ for any right-hand side f through the pre-multiplication $BAx = Bf$, $x = Bf$. In section 1.10 we trifled with the idea of an inverse, and even considered a way to compute it, but the subject calls for more serious consideration.

Because matrix multiplication is not commutative we are obligated to probe the actuality of a right inverse B to A such that $AB = I$, and a left inverse B' to A such that $B'A = I$. The principle of computing an inverse is rather simple. Take the right inverse. If $A = A(m \times n)$, then $B = B(n \times m)$, $I = I(m \times m)$,

$$AB = [Ab_1\ Ab_2 \ldots Ab_m] = I = [e_1\ e_2 \ldots e_m] \tag{2.75}$$

and the m columns of B are computed from the m systems

$$Ab_j = e_j \quad j = 1, 2, \ldots, m \tag{2.76}$$

that include a total of m^2 equations in the mn unknowns B_{ij}.

It is the existence and uniqueness of B that calls for theoretical attention.

Theorem 2.12. *A right inverse to A exists if and only if A is of full row rank. A left inverse to A exists if and only if A is of full column rank.*

Proof. A left inverse B to A is such that $BA = I$, becoming upon transposition $A^T B^T = I$. Hence, if B is a left inverse to A, then B^T is the right inverse to A^T and it is sufficient that we prove the theorem for the right inverse only.

The condition is sufficient. If A is of full row rank, then in the echelon form of $Ab_j = e_j$ every equation has at least one nonzero coefficient and solutions b_j exist for all j.

The condition is necessary. If A is not of full row rank, then a general elementary transformation matrix E as in eq. (2.69) exists so that EA has one zero row, say the kth. To compute the kth column of the right inverse B, b_k, we shall need to solve $Ab_k = e_k$ which is inconsistent. To uncover this, multiply both sides of the equation by E so as to have $EAb_k = Ee_k$. The left-hand side of the kth equation is zero but the right-hand side of the kth equation is $y_k \neq 0$ and the absurdity becomes evident. End of proof.

Theorem 2.12 states that rectangular matrix $A = A(m \times n),\ n > m$ has a right inverse $B = B(n \times m)$ if and only if A is of rank m. When it exists, the right inverse B to matrix A which has more columns than rows is not unique. A right inverse to $A = A(m \times n),\ m > n$ does not exist since the rank of a matrix with more rows than columns is at most n and it therefore cannot be of full row rank. If B is the right inverse of A, $AB = I$, then A is the left inverse of B. A rectangular matrix cannot have both left and right inverses. The distinction of having both left and right inverses belongs to square matrices only, which we deal with next.

Exercises

2.7.1. Write the right inverses of

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad A' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

2.7.2. Find all left inverses of

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

2.7.3. What is the condition on rectangular matrix A so that matrix equation $AX = A$ has the unique solution $X = I$?

2.7.4. Solve matrix equation $AXA = A$ for

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Argue that $AXA = A$ is soluble for any A. Hint: assume first that A has a right or a left inverse.

2.8 The inverse matrix

With square matrices we prefer the concept of triangular matrices of type 0 and 1 over the concept of rank.

Theorem 2.13. *A necessary and sufficient condition for a square matrix to have a right inverse and a left inverse is that it be equivalent to a triangular matrix of type 1.*

Proof. This is a restatement of Theorem 2.12. Square matrix A is of full row rank and full column rank if and only if it is row equivalent to a triangular matrix of type 1. End of proof.

Theorem 2.14. *The right and left inverses of a square matrix are equal and unique.*

Proof. According to Theorem 2.13 the existence of a right inverse B to A implies the existence of a left inverse B' to A so that $AB = B'A = I$. Hence $B'AB = B'$ and $B = B'$. If also $CA = I$, then $(CA)B = C(AB)$, and $C = B$. End of proof.

Definition. *A square matrix that possesses an inverse, that is invertible, is said to be nonsingular. If the matrix does not have an inverse, is not invertible, it is singular. The unique inverse of nonsingular* A *is denoted by* A^{-1}, $AA^{-1} = A^{-1}A = I$.

Now we know that if $AB =$O, and A is nonsingular, then B is necessarily zero, since $A^{-1}(AB) = A^{-1}$O implies $IB =$O and $B =$O. If B is nonsingular, then A is necessarily zero. In the case where both matrices are singular but not zero, they are divisors of zero. Notice that because matrix multiplication is noncommutative, A/B is meaningless as it fails to distinguish between AB^{-1} and $B^{-1}A$.

Theorem 2.15. *If* A *and* B *are nonsingular then:*

1. $(A^{-1})^{-1} = A$.
2. $(AB)^{-1} = B^{-1}A^{-1}$.
3. $(A^{-1})^T = (A^T)^{-1}$.
4. $A^{-1} = (A^{-1})^T = A^{-T}$, if $A = A^T$; *the inverse of a symmetric matrix is symmetric.*
5. $(\alpha A)^{-1} = 1/\alpha A^{-1}$, $\alpha \neq 0$.

Proof.

1. $$A^{-1}(A^{-1})^{-1} = I$$
$$AA^{-1}(A^{-1})^{-1} = AI$$
$$(A^{-1})^{-1} = A.$$

2. $$AB\ (AB)^{-1} = I$$
$$B^{-1}A^{-1}AB(AB)^{-1} = B^{-1}A^{-1}I$$
$$B^{-1}IB(AB)^{-1} = B^{-1}A^{-1}$$
$$(AB)^{-1} = B^{-1}A^{-1}.$$

3. $$(AA^{-1})^T = I = (A^{-1})^TA^T$$
$$I(A^T)^{-1} = (A^{-1})^T.$$

4. If $A = A^T$, then $A^{-1} = (A^T)^{-1} = (A^{-1})^T$.

5. From statement 6 of Theorem 2.7 it results that $(\alpha A)(1/\alpha A^{-1}) = I$. End of proof.

It may well happen in applied linear algebra that a matrix is known beforehand to be singular or nonsingular by the way it is set up, but if nothing is known previously about A, the matrix must be brought by elementary operations to triangular form before resolving the question of whether it is nonsingular or singular.

The statement that A^{-1} exists if and only if the determinant of A, $\det(A) \neq 0$ is occasionally useful, and so is

Theorem 2.16. *For square matrix A the following statements are equivalent:*

1. *Matrix A is nonsingular.*
2. *Homogeneous system $Ax = o$ has $x = o$ as the only solution.*
3. *Nonhomogeneous system $Ax = f$ has solution x for any f.*

Proof. Each statement holds if and only if matrix A is row-equivalent to a triangular matrix of type 1, and hence any single statement implies the other two statements. End of proof.

We may put the proof to Theorem 2.16 in a different perspective. According to eq.(2.7) the product of Ax creates a linear combination of the columns of A with factors $x_1, x_2, \ldots, x_n$. The existence of $x \neq o$ so that $Ax = o$ spells then a generalized elementary operation on the columns of A that annuls at least one of them, telling us in effect that A is not of full rank. On the other hand if $Ax \neq o$ for $x \neq o$, then A is of full rank.

With the aid of Theorem 2.16 we verify that

$$A = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 1 \end{bmatrix} \tag{2.77}$$

is singular since for $x = [1\ 1\ 1\ 1]^T$, $Ax = o$.

For rectangular $A = A(m \times n), m > n$, the fact that $Ax \neq o$ if $x \neq o$, means that A possesses a left inverse.

Statement 2 of Theorem 2.15 says that if AB is nonsingular, then so are A and B since $A^{-1} = B(AB)^{-1}, B^{-1} = (AB)^{-1}A$, and conversely, that the product of nonsingular matrices is a nonsingular matrix. No corresponding simple rank statement exists for the product of *rectangular* matrices. For instance

$$AB = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{2.78}$$

is singular even though B is of full column rank and A is of full row rank. But if in $C = AB$, A is not of full row rank, or B is not of full column rank, then C is singular, for otherwise $(C^{-1}A)B = A(BC^{-1}) = I$ contradicts Theorem 2.12.

Theorem 2.17

1. *For every square A there exists a scalar α so that $A + \alpha I$ is nonsingular.*
2. *Matrix $A^TA + I$ is nonsingular for any rectangular A.*
3. *Matrix A^TA is nonsingular if and only if A is of full column rank.*

Proof.

1. By choosing α large enough we may render every pivot of $A + \alpha I$ positive.

2. The proof is by contradiction. Suppose $x \neq o$ exists so that $A^TAx + x = o$. This implies that $x^TA^TAx + x^Tx = (Ax)^T(Ax) + x^Tx = 0$. Since $x^Tx > 0$ and $(Ax)^T(Ax) \geq 0$ their sum cannot vanish if $x \neq o$ and our assumption is wrong. By Theorem 2.16 the matrix is nonsingular.

3. Assume that $A^TAx = o$, $x \neq o$. This implies that $x^TA^TAx = (Ax)^T(Ax) = 0$, which happens if and only if $Ax = o$. If A is of full column rank $Ax = o$ implies $x = o$ in contradiction to our assumption and A^TA is nonsingular. If A is not of full rank, $Ax = o$ definitely has nontrivial solutions, thus our assumption that $A^TAx = o$ for some nonzero x is correct, and by Theorem 2.16 A^TA is singular. End of proof.

Theorem 2.18. *An elementary matrix is nonsingular.*

Proof. The first three elementary matrices in eq. (2.55) are triangular matrices of type 1. In fact

$$\begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & \alpha^{-1} & \\ & & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & & \\ & 1 & \\ \alpha & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ -\alpha & & 1 \end{bmatrix}, \begin{bmatrix} 1 & & \alpha \\ & 1 & \\ & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & -\alpha \\ & 1 & \\ & & 1 \end{bmatrix}. \tag{2.79}$$

Formally, if $E = I + \alpha e_i {e_j}^T$, then $E^{-1} = I - \alpha e_i {e_j}^T$, and

$$EE^{-1} = I - \alpha e_i {e_j}^T + \alpha e_i {e_j}^T - \alpha^2 e_i {e_j}^T e_i {e_j}^T = I \tag{2.80}$$

because $e_i^T e_j = 0$ if $i \neq j$. More generally

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ \alpha_1 & \alpha_2 & 1 & \alpha_4 \\ & & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ -\alpha_1 & -\alpha_2 & 1 & -\alpha_4 \\ & & & 1 \end{bmatrix}. \tag{2.81}$$

For compound permutation matrix $P = P_k \cdots P_2 P_1$ we have that $P^T = P_1 P_2 \cdots P_k$, and hence $P^T P = PP^T = I$, or $P^{-1} = P^T$. End of proof.

It is possible to invert a matrix piecemeal by partitioning. We write

$$\begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \begin{bmatrix} I & \\ & I \end{bmatrix} \tag{2.82}$$

then we separate the block matrix equation into its components

$$\begin{aligned} EP + FR = I \quad & EQ + FS = O \\ GP + HR = O \quad & GQ + HS = I \end{aligned} \tag{2.83}$$

from which we obtain

$$\begin{aligned} P = (E - FH^{-1}G)^{-1} \quad & R = -H^{-1}GP \\ S = (H - GE^{-1}F)^{-1} \quad & Q = -E^{-1}FS \end{aligned} \tag{2.84}$$

as the submatrices of A^{-1}.

In floating-point computations the distinction between singular and nonsingular matrices can be blurred by round-off errors. A matrix can be *nearly* singular in the sense that small changes in its entries can make it singular, and we imagine that in this case the variations in A^{-1} may be relatively large. A sensible discussion of this subject must wait for more groundwork on the theory of square matrices.

Clearly, the concept of the inverse is of central importance to the algebra of matrices, and we would think to the practice as well. Surprisingly, such is not the case. An inverse is rarely computed, and the linear system $Ax = f$ with arbitrary right-hand sides is not solved as $x = A^{-1}f$, as we shall see in the next three sections and further in Chapter 3.

Exercises

2.8.1. Solve the matrix equations $AX = B$, and $XA = B$ for unknown matrix X, assuming that matrix A is nonsingular and has an inverse. Beware of the fact that matrix multiplication in not commutative!

2.8.2. Solve the system of matrix equations

$$AX + BY = C$$

$$FX + Y = G.$$

for unknown matrices X and Y. You may assume that the matrices you need to invert are nonsingular.

2.8.3. Show that if $AB + A = I$, then A is nonsingular.

2.8.4. Let A, B be square matrices. Prove that if $AB = O$, then either A or B are singular.

2.8.5. If square matrices A and B are such that $AB = 2I$ what is BA? If $BA = 2I$, what is the inverse of $C = I + AB$? Hint: if $AB = I$, then also $BA = I$.

2.8.6. Show that the inverse of $A = I + \alpha uu^T, u^Tu = 1$, is of the form $A^{-1} = I + \beta uu^T$. Write β in terms of α. For what values of α is A singular?

2.8.7. Show that the inverse of $A = I + ab^T$ is of the form $A^{-1} = I + \alpha ab^T$. Write α in terms of vectors a and b.

2.8.8. Show that either $I + ab^T$ or $I - ab^T$ is nonsingular.

2.8.9. Prove that both $A = I + uv^T$, $B = I - uv^T$, $u^Tv = v^Tu = 0$, are nonsingular.

2.8.10. Show that if matrix N is nilpotent, $N^2 = O$, then $A = I + 2N$ has an inverse in the form $A^{-1} = I + \alpha N$. Find α. Also, that if A is idempotent, $A^2 = A$, then the inverse of $B = I + A$ is of the form $B^{-1} = I + \alpha A$.

2.8.11. Let $B = I + A$, $A^2 = \kappa I$. Show that $B^{-1} = \alpha I + \beta A$. Find α and β.

2.8.12. Prove that $C = I + AB, BA = O$ is nonsingular. Write the inverse of C.

2.8.13. Show that if $A^2 + 3A + I = O$, then $A^{-1} = -A - 3I$. Also that if $A^3 = I + A$, then A, $I + A$ and $A - I$ are nonsingular.

2.8.14. Prove that if $A^2 = A \neq I$, then A is singular.

2.8.15. Write all 6 nonsingular 2×2 matrices with entries 0 or 1.

2.8.16. Show that every square matrix can be written as the sum of two nonsingular matrices.

2.8.17. Show that matrices A, B, C

$$A = \begin{bmatrix} \times & \times & 0 & \times \\ \times & \times & 0 & \times \\ \times & \times & 0 & \times \\ \times & \times & 0 & \times \end{bmatrix}, \ B = \begin{bmatrix} \times & 1 & 1 & \times \\ \times & -2 & -2 & \times \\ \times & 3 & 3 & \times \\ \times & -1 & -1 & \times \end{bmatrix}, \ C = \begin{bmatrix} \times & 1 & \times & -2 \\ \times & -2 & \times & 4 \\ \times & 3 & \times & -6 \\ \times & -1 & \times & 2 \end{bmatrix}$$

are singular. Hint: find vector $x \neq o$ such that $Ax = o$.

2.8.18. Show that

$$A = \begin{bmatrix} & 1 & \times & \times & \times \\ & & & & 1 \\ & & & 1 & \times \\ 1 & \times & \times & \times & \times \\ & & 1 & \times & \times \end{bmatrix} \text{ and } B = \begin{bmatrix} \times & \times & \times & 1 & \\ \times & \times & 1 & & \\ \times & 1 & & & \\ 1 & & & & \\ & & & & 1 \end{bmatrix}$$

are nonsingular.

2.8.19. Under what conditions on α, β, γ is the *Vandermonde* matrix

$$V = \begin{bmatrix} 1 & 1 & 1 \\ \alpha & \beta & \gamma \\ \alpha^2 & \beta^2 & \gamma^2 \end{bmatrix}.$$

nonsingular?

2.8.20. Compute the inverse of the Pascal triangle matrix

$$P = \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}.$$

2.8.21. Compute the inverses of

$$A = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & 1 \\ 0 & 2 & 5 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 2 & 5 \\ 1 & 0 & 2 \\ -1 & 1 & 1 \end{bmatrix}, \; T = \begin{bmatrix} 2 & -1 & & \\ -1 & & -1 & \\ & -1 & & -1 \\ & & -1 & 1 \end{bmatrix}.$$

2.8.22. Compute the inverse of

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

2.8.23. Write the inverses of A, B, and C

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \alpha & & & & \\ & \alpha & & & \\ & & \alpha & & \\ & & & \alpha & \end{bmatrix}, \; B = \begin{bmatrix} & & & & \alpha \\ & & & \alpha & \\ & & \alpha & & \\ & \alpha & & & \\ \alpha & & & & \end{bmatrix}, \; C = \begin{bmatrix} 1 & & & & \\ \alpha & 1 & & & \\ & \alpha & 1 & & \\ & & \alpha & 1 & \\ & & & \alpha & 1 \end{bmatrix}.$$

2.8.24. Compute the inverse of matrix

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

for what values of α is A nonsingular?

2.8.25. For what value of α are matrices A, B and C

$$A = \begin{bmatrix} & 1 & 1 \\ 1 & & 1 \\ 1 & 1 & \alpha \end{bmatrix}, B = \begin{bmatrix} & 1 & \alpha \\ 1 & & 1 \\ \alpha & 1 & \end{bmatrix}, C = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix}$$

singular?

2.8.26. Show that

$$A = \begin{bmatrix} -1 & 1 & 3 \\ 5 & 7 & 9 \\ 11 & 13 & 15 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = B + \lambda C$$

is singular for any λ. Hint: find vector $x \neq o$ such that $Bx = Cx = o$.

2.8.27. Invert

$$A = \begin{bmatrix} \lambda & & & 1 \\ & \lambda & 1 & \\ & 1 & \lambda & \\ 1 & & & \lambda \end{bmatrix}, A = \begin{bmatrix} \lambda & & 1 \\ & 1+\lambda & \\ 1 & & \lambda \end{bmatrix}.$$

For what values of λ is A singular?

2.8.28. Invert

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \lambda \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

For what values of λ is A singular?

2.8.29. Show that

$$A = \begin{bmatrix} 0 & A_{12} & A_{13} \\ -A_{12} & 0 & A_{23} \\ -A_{13} & -A_{23} & 0 \end{bmatrix}$$

is singular. Hint: find vector $x \neq o$ such that $Ax = o$.

2.8.30. Consider matrix $A(n \times n)$ with entries all equal to 1, $A_{ij} = 1$ for all i and j. Show first that $A^2 = nA$, then that $(I - A)^{-1} = I - 1/(n-1)A$.

2.8.31. How does the interchange of rows i and j of A affect A^{-1}? How does the interchange of columns i and j of A affect A^{-1}? Hint: consider $P(AA^{-1} - I)P$, with permutation matrix P.

2.8.32. Show that if matrix H is in Hermite form (see sec. 1.7), then $H^2 = H$. Show further that if B is nonsingular and such that $BA = H$, then $ABA = A$.

2.8.33. Show that for

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

$H^2 = H$. Is the *Hermite* form of square matrix H characterized by $H^2 = H$ plus the condition that H is upper-triangular with 0's and 1's on the diagonal?

2.8.34. Show that

$$\begin{bmatrix} A & \\ & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & \\ & B^{-1} \end{bmatrix}.$$

2.8.35. Show that

$$\begin{bmatrix} I & A \\ & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -A \\ & I \end{bmatrix}.$$

2.8.36. Show that

$$\begin{bmatrix} A & B \\ & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ & C^{-1} \end{bmatrix}.$$

2.8.37. Show that

$$\begin{bmatrix} A & I \\ I & A \end{bmatrix}^{-1} = \begin{bmatrix} (I-A^2)^{-1} & \\ & (I-A^2)^{-1} \end{bmatrix} \begin{bmatrix} -A & I \\ I & -A \end{bmatrix}.$$

2.8.38. Show that

$$\begin{bmatrix} I & A \\ A & I \end{bmatrix}^{-1} = \begin{bmatrix} (I-A^2)^{-1} & \\ & (I-A^2)^{-1} \end{bmatrix} \begin{bmatrix} I & -A \\ -A & I \end{bmatrix}.$$

2.8.39. Write A^{-1} for

$$A = \begin{bmatrix} B & \alpha I \\ \alpha I & B^{-1} \end{bmatrix}.$$

For what values of α is A singular?

2.8.40. Show that

$$\begin{bmatrix} I & \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ O & D-CA^{-1}B \end{bmatrix}$$

and that

$$\begin{bmatrix} I & O \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & O \\ O & D-CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ O & I \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

2.8.41. Write the inverses of

$$A = \begin{bmatrix} I & o \\ a^T & 1 \end{bmatrix}, B = \begin{bmatrix} S & o \\ a^T & 1 \end{bmatrix},\ C = \begin{bmatrix} I & b \\ a^T & 1 \end{bmatrix},\ a^Tb = 0.$$

2.8.42. Let $A = A^T$ be such that all nontrivial solutions to $Bx = o$ are of the form $\alpha u,\ u^Tu = 1$. Show that under these conditions

$$A = \begin{bmatrix} B & u \\ u^T & 0 \end{bmatrix}$$

is nonsingular.

2.8.43. Show that

$$B = \begin{bmatrix} A & -Aq \\ -p^T A & p^T Aq \end{bmatrix}$$

is always singular. Hint: compute Bx for $x = [q^T \ 1]^T$.

2.8.44. The matrix in the partitioned system

$$\begin{bmatrix} A & a \\ & \\ a^T & \alpha \end{bmatrix} \begin{bmatrix} x \\ \\ x_n \end{bmatrix} = \begin{bmatrix} f \\ \\ f_n \end{bmatrix}$$

is called *bordered*. Assume that $A = A^T$ and show that

$$x_n = (f_n - f^T A^{-1} a)/(\alpha - a^T A^{-1} a), x = A^{-1} f - x_n A^{-1} a.$$

2.8.45. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad \begin{matrix} A_{11} + A_{12} = \alpha \\ A_{21} + A_{22} = \alpha. \end{matrix}$$

Show that if $B = A^{-1}$, then the row sums of B is $1/\alpha, \ \alpha \neq 0$.

2.8.46. Square matrix Q is *orthogonal* if $Q^T Q = I$, that is if $Q^{-1} = Q^T$. Show that $Q = (A^T A)^{1/2} A^{-1}$ is orthogonal. Matrix $B = (A^T A)^{1/2}$ is symmetric and such that $B^2 = A^T A$.

2.8.47. Show that if $A^T A = Q$ and $Q^T Q = I$, then A has a left inverse.

2.8.48. Show that if $Q = A^T A$ is orthogonal for square A, $Q^T Q = QQ^T = I$, then so is AA^T. Is it true for rectangular A?

2.8.49. Let $A = A(3 \times 3) = [a_1 \ a_2 \ a_3]$ be with nonzero columns and such that ${a_1}^T a_2 = {a_2}^T a_3 = {a_3}^T a_1 = 0$. Show that A is nonsingular. Hint: form $A^T A$.

2.8.50. Let $S = -S^T$. Show that if $Q = (I + S)^{-1}(I - S)$, then $QQ^T = I$.

2.8.51. Prove that if nonsingular A and B are such that $AA^T = BB^T$, then $B = AQ$, $QQ^T = I$.

2.8.52. Show that $B = I - A(A^T A)^{-1} A^T$ is idempotent, $B^2 = B$.

2.8.53. Let $C = A(B^T A)^{-1} B^T$. Show that $C^2 = C$.

2.8.54. Show that if A is nilpotent, then so is $B^{-1}AB$, and if A is idempotent, then so is $B^{-1}AB$.

2.8.55. Let square A be nilpotent. Show that $(A+A^T)^2 = AA^T + A^TA$. Also that if $A+A^T$ is nonsingular, then $PA + AQ = I$,

$$P = (A + A^T)^{-2}A^T, \; Q = A^T(A + A^T)^{-2}.$$

2.8.56. Show that for any positive integer m, $BA^mB^{-1} = (BAB^{-1})^m$.

2.8.57. Prove that if A is nonsingular, then $A+B$ and $I+A^{-1}B$ are either both nonsingular or both singular.

2.8.58. Show that if A, B and $A+B$ are nonsingular, then so is $A^{-1} + B^{-1}$ and

$$(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B = B(A + B)^{-1}A.$$

2.8.59. Show that $(I + BA)^{-1} = I - B(I + AB)^{-1}A$.

2.8.60. Show that if A, B, U and V are such that A, B and $V^TA^{-1}U - B^{-1}$ are nonsingular, then

$$(A - UBV^T)^{-1} = A^{-1} - A^{-1}UTV^TA^{-1}, \; T = (V^TA^{-1}U - B^{-1})^{-1}.$$

2.8.61. Show that for any square B,

$$(A + B)A^{-1}(A - B) = (A - B)A^{-1}(A + B) = A - BA^{-1}B.$$

2.8.62. Show that if $A = -A^T$, then $I + A$ is nonsingular. Hint: show that $(I + A)x \neq o$ if $x \neq o$.

2.8.63. Let $R = R(m \times m) = AB - \alpha I$, and $S = S(n \times n) = BA - \alpha I$ be nonsingular for $\alpha \neq 0$. Show that

$$S^{-1} = \frac{1}{\alpha}(BR^{-1}A - I) \; , \; R^{-1} = \frac{1}{\alpha}(AS^{-1}B - I).$$

2.8.64. Show that for rectangular A and B, $I + AB$ and $I + BA$ are either both singular or both nonsingular. Hint: If $I + AB$ is singular, then $x + ABx = o, x \neq o, Bx \neq 0$.

2.8.65. Let $(I - A)$ be nonsingular and show that

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^k + R_k\ ,\ R_k = (I - A)^{-1}\ A^{k+1}.$$

2.8.66. Let

$$A(\alpha) = \begin{bmatrix} 1 & & \\ \alpha & 1 & \\ \alpha^2 & 2\alpha & 1 \end{bmatrix}.$$

Show that

$$A(\alpha)A(\beta) = A(\beta)A(\alpha) = A(\alpha + \beta)$$

and that $A^{-1}(\alpha) = A(-\alpha)$.

2.8.67. Is the inverse of a circulant matrix circulant? Hint: write

$$\begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{bmatrix} = \alpha I + \beta P + \gamma P^2,\ P^3 = I.$$

2.8.68. Is the inverse of an upper-triangular Töpliz matrix an upper triangular Töpliz matrix? Hint: write

$$\begin{bmatrix} \alpha & \beta & \gamma \\ & \alpha & \beta \\ & & \alpha \end{bmatrix} = \alpha I + \beta T + \gamma T^2,\ T^3 = O.$$

2.8.69. Show that $B = (A^T A)^{-1} A^T$, and $C = A^T (AA^T)^{-1}$ are left and right inverses of A, $BA = I$ and $AC = I$, provided the needed inverses exist.

2.8.70. Let P be a column permutation matrix such that $AP = [A_1\ A_2]$ is with a nonsingular A_1. Show that every right inverse C of AP is of the block form

$$C = \begin{bmatrix} A_1^{-1}(I - A_2 Y) \\ Y \end{bmatrix}$$

where Y is arbitrary.

2.8.71. Show that if $E = uv^T$, then $EAE = \alpha E$ and $(EA)^2 = \alpha(EA)$. Find α in terms of u, v, A. Show that if A is nonsingular, then

$$(A+E)^{-1} = A^{-1} + \gamma A^{-1} E A^{-1}.$$

Find γ.

2.8.72. Let $A_1 = A + E_1$, $A_2 = A_1 + E_2$ with $E_1 = u_1 {v_1}^T$, $E_2 = u_2 {v_2}^T$. Establish the recursive formula

$$A_1^{-1} = A^{-1} + \gamma_1 A^{-1} E_1 A^{-1}, \; A_2^{-1} = A_1^{-1} + \gamma_2 A_1^{-1} E_2 A_1^{-1}$$

and write out γ_1 and γ_2.

2.8.73. Show that if matrices E_1 and E_2 are of rank 1, then inversion of $A_2 = A + E_1 + E_2$ can be done recursively by

$$A_1^{-1} = A_0^{-1} - \frac{1}{1+\tau_1} A_0^{-1} E_1 A_0^{-1}$$
$$A_2^{-1} = A_1^{-1} - \frac{1}{1+\tau_2} A_1^{-1} E_2 A_1^{-1}$$

where $A_0 = A, A_1 = A_0 + E_1, A_2 = A_1 + E_2, \tau_1 = \text{trace}(E_1), \tau_2 = \text{trace}(E_2)$, provided the needed inverses exist.

2.8.74. Let $u_1 = u_1(3 \times 1), u_2 = u_2(3 \times 1), u_3 = u_3(3 \times 1)$ be such that ${u_1}^T u_1 = {u_2}^T u_2 = {u_3}^T u_3 = 1$, and ${u_1}^T u_2 = {u_2}^T u_3 = {u_3}^T u_1 = 0$. Show that

$$A = A(3 \times 3) = u_1 {u_1}^T + u_2 {u_2}^T + u_3 {u_3}^T$$

is the identity matrix. Hint: Write $U = [u_1 \; u_2 \; u_3]$ and notice that $A = UU^T, U^T U = I$.

2.9 Positive definite matrices

We open our discussion by considering scalar functions of *variable* vectors.

Definition. *Let x and y be arbitrary vectors, a, a given vector, and A a given matrix. Scalar function $x^T a$ of variable vector x is a linear form of x. Scalar functions $x^T Ay$ and*

$y^T Ax$ of variable vectors x and y are bilinear forms of x and y. Scalar function $x^T Ax$ of variable vector x is a quadratic form of x.

Notice that in $x^T Ay$ and $y^T Ax$, A may be rectangular, but that it is square in $x^T Ax$. In case $x^T Ay = y^T Ax$ the bilinear form is said to be *symmetric.*

In expanded form

$$\begin{aligned} x^T Ay = A_{11}x_1y_1 + A_{12}x_1y_2 + \cdots + A_{1n}x_1y_n \\ + A_{21}x_2y_1 + A_{22}x_2y_2 + \cdots + A_{2n}x_2y_n \\ \vdots \\ + A_{m1}x_my_1 + A_{m2}x_my_2 + \cdots + A_{mn}x_my_n \end{aligned} \tag{2.85}$$

which includes the sum of all possible $A_{ij}x_iy_j$.

It is unorthodox though visually helpful to write the bilinear form $x^T Ay$ in border matrix fashion

$$\begin{array}{cc} & \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{bmatrix} \end{array} \tag{2.86}$$

to mark that x_i multiplies each entry in the ith row of A, and y_j each entry in the jth column.

Theorem 2.19. *A bilinear form is symmetric, $x^T Ay = y^T Ax$, for any x and y, if and only if $A = A^T$.*

Proof. Since the transpose of a scalar is the same scalar we have that $(x^T Ay)^T = y^T Ax$ if $A = A^T$. Choosing $x = e_i$ and $y = e_j$ we prove that $x^T Ay = y^T Ax$ only if $A_{ij} = A_{ji}$, or $A = A^T$. End of proof.

Theorem 2.20. *For any square matrix A*

$$x^T Ax = \frac{1}{2}x^T(A + A^T)x. \tag{2.87}$$

Proof.

$$\begin{aligned} x^T Ax &= \frac{1}{2}x^T(A + A^T)x + \frac{1}{2}x^T(A - A^T)x \\ &= \frac{1}{2}x^T(A + A^T)x \end{aligned} \tag{2.88}$$

since $x^T(A - A^T)x = 0$. End of proof.

Every matrix can be written as the sum of the symmetric $\frac{1}{2}(A + A^T)$ and the skew symmetric $\frac{1}{2}(A - A^T)$. In view of Theorem 2.20 we may consider quadratic form $x^T Ax$ as always with a symmetric A.

Definition. *Symmetric matrix $A = A^T$, for which the quadratic form is positive, $x^T Ax > 0$, for any $x \neq o$ is positive definite. If $x^T Ax \geq 0$ for all $x \neq o$, then A is positive semidefinite.*

In mathematical physics, positive definiteness is a concept related to energy, and positive definite matrices constitute an important class of matrices in applied computational linear algebra. We want to pay close attention to symmetric and positive definite matrices.

Examples.

1. Matrix

$$D = \begin{bmatrix} 2 & & \\ & 3 & \\ & & 1 \end{bmatrix} \tag{2.89}$$

is positive definite since

$$x^T Dx = 2{x_1}^2 + 3{x_2}^2 + {x_3}^2 > 0 \ \text{ if } x \neq o. \tag{2.90}$$

2. Matrix

$$A = \begin{bmatrix} 5 & -2 & 3 \\ -2 & 3 & 1 \\ 3 & 1 & 6 \end{bmatrix} \tag{2.91}$$

is positive definite since

$$x^T Ax = 2(x_1 - x_2)^2 + 3(x_1 + x_3)^2 + (x_2 + x_3)^2 + 2{x_3}^2 > 0 \ \text{ if } x \neq o. \tag{2.92}$$

3. Matrix

$$A = \begin{bmatrix} 2 & 2 & -1 \\ 2 & 1 & 2 \\ -1 & 2 & 4 \end{bmatrix} \tag{2.93}$$

is indefinite since for $x = [1\ \ -1\ 1]^T$, $x^T Ax = -2$, while for $x = [1\ 0\ 0]^T$, $x^T Ax = 2$.

Theorem 2.21. *Let A be symmetric and positive definite. Then matrix $B^T AB$ is symmetric and positive semidefinite if B does not have a left inverse, and is symmetric* *positive definite if B does have a left inverse. If A is symmetric and positive semidefinite, then B^TAB is symmetric and positive semidefinite.*

Proof. With $(B^TAB)^T = B^TA^T(B^T)^T = B^TAB$ we prove that the matrix is symmetric. Write $Bx = y$ so that $x^TB^TABx = y^TAy$. If B does not have a left inverse then there are vectors $x \neq o$ for which $Bx = y = o$, and $y^TAy \geq 0$ if $x \neq o$. Matrix B^TAB is then positive semidefinite. If B does have a left inverse, then for any $x \neq o$, also $y \neq o$, and $y^TAy > 0$ if $x \neq o$ since A is positive definite. Matrix B^TAB is then positive definite.

If A is positive definite then $x^TB^TABx = (Bx)^TA(Bx) \geq 0$ for any x and B^TAB is positive semidefinite. End of proof.

Notice that for a square B, *singular* and *nonsingular* can be instituted for not having a left inverse and having a left inverse, respectively.

Lemma 2.22. *The conditions necessary and sufficient for symmetric $A = A(2 \times 2)$ to be positive definite are that*

$$A_{11} > 0 \quad and \quad \det(A) = \begin{vmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{vmatrix} = A_{11}A_{22} - A_{12}^2 > 0. \tag{2.94}$$

Matrix A is positive semidefinite if and only if $A_{11} \geq 0,\ A_{22} \geq 0$ and $\det(A) \geq 0$.

Proof. Choosing $x = [1\ 0]^T$ we have that $x^TAx = A_{11}$, whereas with $x = [-A_{12}\ A_{11}]$ we have that $x^TAx = A_{11}\det(A)$, and hence the conditions are necessary. Then we write

$$\begin{aligned} x^TAx &= A_{11}^{-1}\left((A_{11}x_1 + A_{12}x_2)^2 + (A_{11}A_{22} - {A_{12}}^2){x_2}^2\right) \\ &= A_{22}^{-1}\left((A_{22}x_2 + A_{12}x_1)^2 + (A_{11}A_{22} - {A_{12}}^2){x_1}^2\right) \end{aligned} \tag{2.95}$$

and see that the conditions $A_{11} > 0, \det(A) > 0$ are sufficient for positive definiteness.

In case $A_{11} = 0,\ x^TAx = 2A_{12}x_1x_2 + A_{22}{x_2}^2$, and $x^TAx \geq 0$ if and only if $A_{12} = 0$ and $A_{22} \geq 0$.

If $A_{22} = 0$, then $x^TAx \geq 0$ if and only if $A_{12} = 0$ and $A_{11} \geq 0$. End of proof.

We use Lemma 2.22 to prove one of the most useful inequalities of all mathematics.

Theorem (Cauchy-Schwarz) 2.23. *Let $A = A^T$ be positive semidefinite, and B rectangular. Then for any two vectors x and y conforming in multiplication*

$$(x^TABy)^2 \leq (x^TAx)(y^TB^TABy) \tag{2.96}$$

or if $A = B = I$

$$(x^T y)^2 \leq (x^T x)(y^T y). \tag{2.97}$$

Proof. Let $X(n \times 2)$ have the two columns x_1 and x_2. According to Theorem 2.21

$$X^T A X = \begin{bmatrix} {x_1}^T A x_1 & {x_1}^T A x_2 \\ {x_2}^T A x_1 & {x_2}^T A x_2 \end{bmatrix} \tag{2.98}$$

is positive semidefinite. Hence, necessarily

$$({x_1}^T A x_1)({x_2}^T A x_2) - ({x_1}^T A x_2)({x_2}^T A x_1) \geq 0 \tag{2.99}$$

or since $A^T = A$

$$({x_1}^T A x_2)^2 \leq ({x_1}^T A x_1)({x_2}^T A x_2). \tag{2.100}$$

With $x_1 = x,\ x_2 = By$ we obtain the inequality. End of proof.

Corollary 2.24. *If* $A = A^T$ *is positive semidefinite and* $x^T A x = 0$*, then* $Ax = o$.

Proof. Matrix $X^T A X$, $X = [x_1\ x_2]$, is positive semidefinite and hence if $x_1^T A x_1 = 0$, then also $x_2^T (A x_1) = 0$ for any vector x_2. This can happen only if $A x_1 = o$. End of proof.

Theorem 2.25. *If* $x^T A x > 0$ *and* $x^T B x \geq 0$*, then* $x^T (A + B) x > 0$*. In particular, if* $A = A^T$ *is positive semidefinite, and* B *is rectangular, then* $I + B^T A B$ *is positive definite.*

Proof.

$$x^T (A + B) x = x^T A x + x^T B x > 0. \tag{2.101}$$

End of proof.

Positive definiteness is preserved under addition but not under multiplication. If A and B are positive definite then $AB + (AB)^T$ need not be positive definite. Indeed,

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 7 & -1 \\ -1 & 1 \end{bmatrix}, AB + (AB)^T = \begin{bmatrix} 30 & -11 \\ -11 & 4 \end{bmatrix} \tag{2.102}$$

are such that the first two matrices are positive definite, but not the third. However,

Theorem 2.26. *If* A *and* B *are symmetric positive semidefinite, then* $\epsilon I + AB$ *is nonsingular for any* $\epsilon > 0$.

Proof. Assume $\epsilon I + AB$ is singular so that $\epsilon x + ABx = o$ with $x \neq o$. Premultiplication by x^TB yields $\epsilon x^TBx + x^TBABx = 0$ and since both B and BAB are semidefinite each quadratic term must disappear and we have that $x^TBx = 0$ and, consequently $Bx = o$. But $\epsilon x + ABx = o$ implies then that $x = o$. Hence, if $x \neq o$, then $(\epsilon I + AB)x \neq o$ and the matrix is nonsingular. End of proof.

Theorem 2.27.

1. If A is symmetric and positive definite, then any diagonal submatrix of A is symmetric and positive definite.

2. If A is symmetric and positive definite, then $A_{ii} > 0$ for all i. If A is symmetric and positive semidefinite, then $A_{ii} \geq 0$ for all i, but if $A_{ii} = 0$ then the entire ith row and column are zero.

3. In a symmetric positive definite matrix the largest entry in magnitude is on the diagonal.

4. A symmetric positive definite matrix is nonsingular.

5. If A is symmetric positive definite, then so are A^2 and A^{-1}.

Proof.

1. Let A' be a principal submatrix of A and let x be partitioned conformally as $x = [o^T z^T o^T]^T$, $z \neq o$, so that $x^TAx = z^TA'z$. If A is symmetric positive definite, then $z^TA'z > 0$, for arbitrary z and A' is also symmetric positive definite.

2. Choosing $x = e_i$, we have that $x^TAx = {e_i}^TAe_i = A_{ii} \geq 0$ if A is positive semidefinite, and $A_{ii} > 0$ if A is positive definite. Continue with 3.

3. With $x = \alpha_1 e_i + \alpha_2 e_j$ we have that

$$\begin{aligned} x^TAx &= {\alpha_1}^2 {e_i}^TAe_i + 2\alpha_1\alpha_2 A{e_i}^TAe_j + {\alpha_2}^2{e_j}^TAe_j \\ &= {\alpha_1}^2 A_{ii} + 2\alpha_1\alpha_2 A_{ij} + {\alpha_2}^2 A_{jj} \geq 0 \end{aligned} \tag{2.103}$$

since ${e_i}^TAe_i = A_{ii}$, and ${e_i}^TAe_j = A_{ij}$. Assume first that A is symmetric positive definite. Then quadratic form (2.103) is positive and

$$A_{ii}A_{jj} - {A_{ij}}^2 > 0. \tag{2.104}$$

But the inequality is contradictory if $|A_{ij}|$ is the largest entry in magnitude. Hence the largest entry in magnitude is on the diagonal.

If A is symmetric and positive semidefinite, then $A_{ii}A_{jj} - {A_{ij}}^2 \geq 0$, and if $A_{ii} = 0,\ A_{ij} = A_{ji} = 0$ for all j.

4. The existence of $x \neq o$ such that $Ax = o$ would have meant that $x^T Ax = 0$, in contradiction with the assumption that A is symmetric positive definite. Hence $Ax = o$ only when $x = o$, and A is nonsingular.

5. Matrix A is positive definite and hence invertible, therefore $Ax \neq o$ if $x \neq o$, and consequently $x^T A^2 x = (Ax)^T(Ax) > 0$ if $x \neq o$. With $x = A^{-1}y,\ x^T Ax = y^T A^{-1} A A^{-1} y = y^T A^{-1} y$, which is positive for every nonzero y since $y \neq o$ implies $x \neq o$. End of Proof.

One of the most welcome and practically significant properties of symmetric positive semidefinite matrices is that Gauss forward elimination done on a system with such a matrix theoretically never requires row interchanges. If the matrix is positive definite, then all the pivots are positive, whereas if the matrix is positive semidefinite, then when a zero pivot is encountered the whole column (and row) is zero.

Theorem 2.28. *Gauss forward elimination done on $Ax = f$, with a symmetric positive semidefinite matrix A does not require row interchanges.*

Proof. We assume the first pivot to be nonzero, since otherwise the entire first row and entire first column of A would be zero. Elementary row operation

$$\begin{bmatrix} 1 & & & \\ \times & 1 & & \\ \times & & 1 & \\ \times & & & 1 \end{bmatrix} \begin{bmatrix} 1 & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & \times & \times & \times \\ \hline & & & \\ & & A' & \\ & & & \end{array}\right] \tag{2.105}$$

leaves submatrix A' symmetric and positive semidefinite. Indeed, submatrix A' is not affected by the symmetric *column* operation which also clears the first row,

$$EAE^T = \begin{bmatrix} 1 & o^T \\ & \\ o & A' \end{bmatrix}. \tag{2.106}$$

Since E is nonsingular, EAE^T is positive semidefinite and so is A'.

The second pivot (the first diagonal of A') is nonnegative, but if it is zero then the second row and second column of EAE^T are zero and we move to the third pivot. Symmetrical elimination produces exactly the same pivots that appear in the Gauss forward elimination. End of proof.

Theorem 2.28 has in addition important practical implications for the Gauss elimination of systems with a symmetric positive semidefinite matrix. Since A' remains symmetric, *elimination can be performed with only the lower half* of A. If A is symmetric but not positive semidefinite, row interchanges might be necessary during the forward elimination. To save the symmetry, a simultaneous symmetric column interchange must be carried out. This effectively interchanges one diagonal pivot with another diagonal pivot, but will not work if all diagonal entries of A are zero. A symmetric matrix with a zero diagonal need not necessarily be singular. For example

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}. \tag{2.107}$$

Fortunately, the monstrously large systems of linear equations that mathematical physics gives rise to are most often positive definite. Pivoting on such large systems would have been a programmer's nightmare. The solution algorithms are much simplified without the need, at least theoretically, to interchange equations.

Exercises

2.9.1. Given that

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

find α so that $x^TAx = 10$.

2.9.2. Is

$$A = \begin{bmatrix} 273 & 271 & 270 \\ 271 & 274 & 276 \\ 270 & 276 & 275 \end{bmatrix}$$

positive definite?

2.9.3. Let

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

Write

$$x^TAx = {x_1}^2 - 2x_1x_2 + \alpha {x_2}^2,$$
$$= (x_1 - x_2)^2 + (\alpha - 1){x_2}^2$$

and decide for what values of α is A positive definite. Do the same to

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

2.9.4. Given matrix A find B so that $x^TAx = x^TBx$ for all x,

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

2.9.5. Prove that if $x^TAx = x^Tx$ for any x, then $A = I + S, S = -S^T$.

2.9.6. Show that if $x^TAx = 0$ for any x, then A is skew symmetric, $A = -A^T$.

2.9.7. Prove that if $A = A(2 \times 2)$ and $B = B(2 \times 2)$ are symmetric and positive definite, then so is C, $C_{ij} = A_{ij}\ B_{ij}$.

2.9.8. Show that if $A = A(2\times 2) = A^T$ is positive definite, then $A_{11} - 2|A_{12}| + A_{22} > 0$. Also that if $A = A^T$ and $B = B^T$ are both positive definite, then $A_{11}B_{11} + 2A_{12}B_{12} + A_{12}B_{12} > 0$. Hint: use the fact that $x^TAx > 0$ with $x = [\sqrt{B_{11}} \quad \pm\sqrt{B_{22}}]^T$.

2.9.9. Show that

$$A = \begin{bmatrix} 1+\alpha & -1 \\ -1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1+\beta & -1 \\ -1 & 1 \end{bmatrix}$$

are positive definite if and only if $\alpha > 0$ and $\beta > 0$. Establish matrix

$$AB + BA = \begin{bmatrix} 2((1+\alpha)(1+\beta)+1) & -\alpha-\beta-4 \\ -\alpha-\beta-4 & 4 \end{bmatrix}$$

and show that $AB + BA$ is positive definite if and only if

$$6\alpha\beta > \alpha^2 + \beta^2 \qquad \text{or} \qquad 4\alpha\beta > (\alpha - \beta)^2.$$

Find α and β such that A and B are positive definite, but $AB + BA$ is not.

2.9.10. Write all square root matrices of

$$A = \begin{bmatrix} 4 & \\ & 9 \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 & 2 \\ & 1 \end{bmatrix}$$

that is, write B so that $B^2 = A$.

2.9.11. Consider matrix

$$B = A^2 = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

for some matrix A. Clearly $AB = BA = A^3$. Find all matrices X so that $A^2X = XA^2$, then among them all X so that $X^2 = A^2$. Are there other matrices Y so that $Y^2 = A^2$?

2.9.12. Prove that if A and B are positive definite and such that $A^2 = B^2$, and $AB = BA$, then $A = B$.

2.10 Triangular factorizations

It is more economical to solve a system of linear equations with a triangular matrix than a system with a full matrix, and the factorization of a square matrix into the product of triangular matrices is, for this reason, highly desirable. We have come close to it in eqs.(2.58) and (2.68).

Theorem 2.29. *For every square matrix A there exists a permutation matrix P so that*

$$PA = LU \tag{2.108}$$

where L is lower-triangular with unit diagonal elements, $L_{ii} = 1$, and where U is upper-triangular.

Proof. According to Theorem 1.22, row permutations exist for which forward Gauss elimination never encounters a zero pivot. Hence lower-triangular elementary matrices, as in eq.(2.59), exist with which

$$E_k \cdots E_2 E_1 (PA) = U \tag{2.109}$$

and

$$PA = E_1^{-1} E_2^{-1} \cdots E_k^{-1} U = LU \ , \ L_{ii} = 1 \tag{2.110}$$

by virtue of the basic properties of E_j. End of proof.

If A is singular then so is U, and it is of type 0. If A is nonsingular, then so is U, and it is of type 1. Moreover, if U is nonsingular and $U_{ii} \neq 0$, then $A = LU$ may be written as

$$A = LDU, \ L_{ii} = U_{ii} = 1 \tag{2.111}$$

where D is diagonal.

Theorem 2.30. *If A is nonsingular, then its LU, and hence also its LDU factorization is unique.*

Proof. Suppose it is not, and let $A = L_1U_1$ and $A = L_2U_2$ be two factorizations of the same nonsingular A. Then $L_1U_1 = L_2U_2$ and

$$L_2^{-1}L_1 = U_2U_1^{-1}. \tag{2.112}$$

The left-hand side of the above matrix equation is lower-triangular with unit diagonal entries while the right-hand side is upper-triangular. Both sides must then be diagonal and

$$L_2^{-1}L_1 = I\ ,\quad U_2U_1^{-1} = I \tag{2.113}$$

and it results that $L_2 = L_1, U_2 = U_1$. End of proof.

Corollary 2.31. *If symmetric nonsingular A admits the triangular factorization $A = LU, L_{ii} = 1$, then also $A = LDL^T$.*

Proof. Since A is nonsingular we may put the factorization in the form $A = LDU, L_{ii} = U_{ii} = 1$. The fact that $A = A^T$ implies that $LDU = U^TDL^T$, and by Theorem 2.30 $U = L^T$. End of proof.

For the actual computation of the lower-and upper-triangular matrices in $A = LU$, all we need is to assume this form and then compute the columns of L one by one taking the product of $L, L_{ii} = 1$, and U.

Examples.

1.
$$\begin{bmatrix} 2 & 1 \\ -2 & -4 \end{bmatrix} = \begin{bmatrix} 1 & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ & -3 \end{bmatrix}.$$
$$= \begin{bmatrix} 1 & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & \\ & -3 \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ & 1 \end{bmatrix}.$$

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

3.
$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ & 0 \end{bmatrix} = \begin{bmatrix} 1 & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & \alpha \end{bmatrix} \begin{bmatrix} 1 & 1 \\ & 0 \end{bmatrix}.$$

4.
$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \neq LU.$$

5.
$$\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & \alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \alpha \text{ arbitrary.}$$

Theorem 2.32. *Symmetric positive (semi)definite matrix A can be factored as*

$$A = LDL^T \tag{2.114}$$

where L is lower-triangular with $L_{ii} = 1$, and where D is diagonal and such that $D_{ii} > 0$ if A is positive definite, and $D_{ii} \geq 0$ if A is positive semidefinite.

Proof. No row interchanges are necessary, and the symmetric row and column elementary operations

$$E_k \cdots E_2 E_1 A {E_1}^T {E_2}^T \cdots {E_k}^T = D \tag{2.115}$$

produce diagonal D that holds the pivots. Hence

$$A = E_1^{-1} \cdots E_{k-1}^{-1} E_k^{-1} D E_k^{-T} E_{k-1}^{-T} \cdots E_1^{-T} \tag{2.116}$$

where E_j and E_j^{-1} are lower-triangular with unit diagonal entries. The product of elementary matrices E_j^{-1} is a lower-triangular matrix L with $L_{ii} = 1$. End of proof.

With

$$D^{\frac{1}{2}} = \begin{bmatrix} {D_{11}}^{\frac{1}{2}} & & & \\ & {D_{22}}^{\frac{1}{2}} & & \\ & & {D_{33}}^{\frac{1}{2}} & \\ & & & {D_{44}}^{\frac{1}{2}} \end{bmatrix} \tag{2.117}$$

so that $D = D^{1/2} D^{1/2}$, the factorization becomes

$$A = LD^{\frac{1}{2}} D^{\frac{1}{2}} L^T = (LD^{\frac{1}{2}})(LD^{\frac{1}{2}})^T \tag{2.118}$$

or, with the generic notation of L for a lower-triangular matrix,

$$A = LL^T. \tag{2.119}$$

Conversely, if $A = LDL^T,\ L_{ii} = 1$, then A is positive definite if $D_{ii} > 0$, and positive semi definite if $D_{ii} \geq 0$. Indeed

$$x^TAx = x^TLDL^Tx = y^TDy,\ y = L^Tx \tag{2.120}$$

and if $y \neq o$ then so is x, in which case $y^TDy > 0$, $y \neq o$, implies $x^TAx > 0,\ x \neq o$.

We may now characterize positive definite and symmetric matrices in a numerically convenient way.

Theorem 2.33. *Symmetric matrix A is positive definite if and only if it can be factored as $A = LDL^T$, where L is lower-triangular with $L_{ii} = 1$, and D is diagonal with $D_{ii} > 0$.*

Examples.

1. $$\begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ L_{21} & 1 & \\ L_{31} & L_{32} & 1 \end{bmatrix} \begin{bmatrix} D_{11} & & \\ & D_{22} & \\ & & D_{33} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} \\ & 1 & L_{32} \\ & & 1 \end{bmatrix}.$$

Moving columnwise we encounter one equation in only one unknown at a time, and we find that

$$\begin{gathered} D_{11} = 2,\ L_{21} = -\frac{1}{2},\ L_{31} = 0 \\ D_{22} = \frac{3}{2},\ L_{32} = -\frac{2}{3} \\ D_{33} = \frac{4}{3} \end{gathered} \tag{2.121}$$

and the matrix is positive definite.

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

and the matrix is positive semidefinite.

Theorem 2.34. *The LDL^T factorization of a symmetric positive definite matrix is unique.*

Proof. Positive definite matrix A is nonsingular and hence by Theorem 2.30 its LU, and consequently also its LDL^T, factorization is unique. End of proof.

When A is symmetric positive semidefinite the LDL^T factorization has a unique D, but no unique L.

Example.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1 & \times & \\ 1 & \times & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 0 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & \times & \times \\ & & 1 \end{bmatrix}. \tag{2.122}$$

For any matrix B for which $B^TB = I,\ B^{-1} = B^T$, holds, the factorization $A = LL^T$ of positive definite and symmetric A may be written as $A = LB^TBL^T = (BL^T)^T(BL^T) = C^TC$, and hence there are many ways of factoring A as $A = C^TC$ with a nonsingular C.

Recall that if P is a single permutation matrix, $P = P^T,\ PP = I$, as in eq.(2.50), then PAP that first interchanges rows i and j of A, then columns i and j of PA, causes diagonal entries A_{ii} and A_{jj} to be interchanged. For generalized elementary matrix E in eq.(2.46) this means that $PEP = E'$, E' being also such a matrix. For instance, interchange of rows 1 and 3 followed by an interchange of columns 1 and 3 results in

$$\begin{bmatrix} 1 & & \\ \alpha_1 & \alpha_2 & \alpha_3 \\ & & 1 \end{bmatrix} \rightarrow \begin{bmatrix} & & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ 1 & & \end{bmatrix} \rightarrow \begin{bmatrix} 1 & & \\ \alpha_3 & \alpha_2 & \alpha_1 \\ & & 1 \end{bmatrix}, \alpha_2 \neq 0, \tag{2.123}$$

while interchange of rows 1 and 2 followed by an interchange of columns 1 and 2 results in

$$\begin{bmatrix} 1 & & \\ \alpha_1 & \alpha_2 & \alpha_3 \\ & & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ 1 & & \\ & & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \alpha_2 & \alpha_1 & \alpha_3 \\ & 1 & \\ & & 1 \end{bmatrix}. \tag{2.124}$$

Hence, if P is a single permutation matrix and E a generalized elementary matrix, then for any matrix A, $P(EA) = (PEP)PA = E'PA$, meaning that a row of A replaced by a general elementary operation can be rearranged to appear wherever desired by a proper prearrangement of the rows of A.

Theorem 2.35.

1. For every matrix A there exists a permutation matrix P so that

$$PA = LR \tag{2.125}$$

where L is a nonsingular, $L_{ii} \neq 0$, lower triangular matrix and where R is in row echelon form.

2. A nonsingular matrix T exists for matrix A so that

$$A = TR \tag{2.126}$$

where R is in reduced echelon form. In case matrix A is square R becomes the Hermite form of A.

Proof. Consider matrix A and echelon form matrix R obtained from it by elementary row operations including row interchanges:

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix} \rightarrow \begin{array}{c} 4 \\ 2 \\ 1 \\ 3 \end{array} \begin{bmatrix} 1 & \times & \times & \times & \times \\ & & 1 & \times & \times \\ & & & & 1 \\ & & & & \end{bmatrix}, \; A \rightarrow R. \tag{2.127}$$

The *reduced* echelon form is obtained from R by back substitutions.

We show that if the rows of A are ordered in advance in their final order of appearance in R, then lower-triangular elementary row operation matrices exist that recreate R. In forming R every row of matrix A is sequentially replaced, starting from the bottom row, by an appropriate linear combination of all rows of A with the intent of having in place all the possible zeroes before the leading 1. We know from eq.(2.127) that a generalized elementary operation exists that voids original row 3 so that

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix} PA = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & & & & \end{bmatrix}, \alpha_4 \neq 0. \tag{2.128}$$

In the same manner the third row of R is reconstructed with

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ & & & 1 \end{bmatrix} \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & & & & \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & & & & 1 \\ & & & & \end{bmatrix}, \alpha_3 \neq 0, \tag{2.129}$$

but since the last row of R is zero we may arbitrarily set $\alpha_4 = 0$ in order to have a lower-triangular elementary matrix. Next we carry out

$$\begin{bmatrix} 1 & & & \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & & & & 1 \\ & & & & \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times & \times \\ & & 1 & \times & \times \\ & & & & 1 \\ & & & & \end{bmatrix}, \alpha_2 \neq 0, \tag{1.130}$$

but since we do not care for the entries to the right of the leading 1 in each row of R, we may set $\alpha_3 = \alpha_4 = 0$ to have a third lower-triangular elementary matrix. Continuing in this manner we end up with

$$E_1 E_2 E_3 E_4 (PA) = R \tag{2.131}$$

where each E_i is a nonsingular lower-triangular matrix. Hence $PA = E_4^{-1} E_3^{-1} E_2^{-1} E_1^{-1} R = LR$, considering that the inverse of a lower-triangular matrix is a lower-triangular matrix, and that the product of lower-triangular matrices is still a lower-triangular matrix. To obtain the next factorization we proceed in eq.(2.131) with more elementary back substitutions until $EPA = R$, with total elementary matrix E, turns R into the reduced form. Setting $EP = T^{-1}$ we obtain

$$A = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \begin{bmatrix} 1 & \times & & \times & \\ & & 1 & \times & \\ & & & & 1 \\ & & & & \end{bmatrix} = TR. \tag{2.132}$$

End of proof.

Corollary 2.36. *Matrix A of rank r can be written as the sum of at least r, rank-one, matrices.*

Proof. The rank of matrix R in the TR factorization of matrix A in eq.(2.132) is r. Let $t_1, t_2, \ldots, t_m$ be the columns of T, and ${r_1}^T, {r_2}^T, \ldots, {r_m}^T$ the rows of R. Then according to eq.(2.29)

$$A = t_1 {r_1}^T + t_2 {r_2}^T + \cdots + t_r {r_r}^T \tag{2.133}$$

since ${r_k}^T = o^T$ if $k > r$.

Elementary row operations reduce $t_1 {r_1}^T + t_2 {r_2}^T + \cdots t_r {r_r}^T$ to a matrix with at most r nonzero rows,

$$\begin{bmatrix} \times \\ \times \\ \times \\ \times \end{bmatrix} [\times\ \times] + \begin{bmatrix} \times \\ \times \\ \times \\ \times \end{bmatrix} [\times\ \times] \rightarrow \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \end{bmatrix} [\times\ \times] + \begin{bmatrix} \times \\ \times \\ 0 \\ 0 \end{bmatrix} [\times\ \times], \tag{2.134}$$

and hence the rank of the sum of r rank-one matrices cannot exceed r. Matrix A of rank r cannot be written therefore as the sum of less than r rank-one matrices. End of proof.

Corollary 2.37. *Matrix $A = A(m \times n)$ of rank r can be factored as $A = BC$, where $B = B(m \times r)$ and $C = C(r \times n)$ are both of rank r. This is a full rank factorization of A.*

Proof. According to eq.(2.133) matrix A may be written as

$$A = \begin{matrix} [\, t_1 \; t_2 \ldots t_r \,] \\ \\ \\ \\ \end{matrix} \begin{bmatrix} {r_1}^T \\ r_2^T \\ \vdots \\ {r_r}^T \end{bmatrix} = BC. \tag{2.135}$$

Matrix C is obviously of full row rank r, and so is matrix B that holds r columns of the full rank matrix T. End of proof.

Example.

$$\begin{bmatrix} 1 & 4 & 5 \\ -1 & -4 & -5 \\ 2 & 7 & 7 \\ -2 & -5 & -1 \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \\ \times & \times \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 2 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ -2 \end{bmatrix} [\, 1 \quad 4 \quad 5 \,] + \begin{bmatrix} 0 \\ 0 \\ -1 \\ 3 \end{bmatrix} [\, 0 \quad 1 \quad 3 \,] . \tag{2.136}$$

Exercises

2.10.1. Perform the LU factorization of $A,\ B,\ C$

$$A = \begin{bmatrix} 2 & -4 & 2 \\ -4 & 7 & -7 \\ 2 & -7 & 1 \end{bmatrix} ,\ B = \begin{bmatrix} 2 & -1 & \\ -1 & 4 & -1 \\ & -1 & 2 \end{bmatrix} ,\ C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

to determine which one of them is positive definite.

2.10.2. Write the LU factorization of

$$A = \begin{bmatrix} -2 & 1 & -2 \\ 4 & 1 & 5 \\ -6 & 6 & -5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & -1 \\ -4 & -2 & 3 \\ 2 & 1 & 1 \end{bmatrix} .$$

2.10.3. Write the factorization

$$\begin{bmatrix} 1 & & \\ \times & 1 & \\ \times & \times & 1 \end{bmatrix} \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 & -3 \\ -6 & 6 & -2 & 10 \\ 2 & -4 & 4 & -2 \end{bmatrix} .$$

2.10.4. Show that

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ & -1 & -2 \end{bmatrix}.$$

2.10.5. Write the full rank factorization of

$$A = \begin{bmatrix} 0 & 2 & -1 \\ 3 & 1 & 4 \\ 1 & 5 & -1 \end{bmatrix}.$$

2.10.6. Write the factorization

$$\begin{bmatrix} 1 & & \\ a & 1 & \\ b & a & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \alpha & 1 & \\ & \alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ \beta & 1 & \\ & \beta & 1 \end{bmatrix}.$$

Express α and β in terms of a and b.

2.10.7. Show that a necessary and sufficient condition that

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

is positive definite is that $\alpha > \sqrt{2}$. Hint: perform the LU factorization of A.

2.10.8. Perform the LU factorization of $A = A(2 \times 2)$ to prove Lemma 2.22.

2.10.9. Verify the LU factorization

$$T = \begin{bmatrix} a_1 & 1 & & \\ 1 & a_2 & 1 & \\ & 1 & a_3 & 1 \\ & & 1 & a_4 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1/u_1 & 1 & & \\ & 1/u_2 & 1 & \\ & & 1/u_3 & 1 \end{bmatrix} \begin{bmatrix} u_1 & 1 & & \\ & u_2 & 1 & \\ & & u_3 & 1 \\ & & & u_4 \end{bmatrix} = LU$$

for this, supposedly nonsingular, tridiagonal matrix T. Compute L^{-1} and U^{-1}, and use $T^{-1} = U^{-1}L^{-1}$ to show that the symmetric inverse may be written in terms of the components of two vectors $x = [x_1\ x_2\ x_3\ x_4]^T, y = [y_1\ y_2\ y_3\ y_4]^T$ in the form

$$T^{-1} = \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 & x_1y_4 \\ x_1y_2 & x_2y_2 & x_2y_3 & x_2y_4 \\ x_1y_3 & x_2y_3 & x_3y_3 & x_3y_4 \\ x_1y_4 & x_2y_4 & x_3y_4 & x_4y_4 \end{bmatrix}.$$

Notice that one may assume $x_1 = 1$, and that vectors x and y are obtained from the solution of $x_1 Ty = e_1$ and $y_4 Tx = e_4$. Extend the result to $T = T(n \times n)$.

2.10.10. Let P and Q be symmetric and positive definite, and

$$R = \begin{bmatrix} P & B^T \\ B & Q \end{bmatrix}.$$

Show that if

$$S = \begin{bmatrix} I & \mathrm{O} \\ -Q^{-1}B & I \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} I & -P^{-1}B^T \\ \mathrm{O} & I \end{bmatrix}$$

then

$$S^T R S = \begin{bmatrix} P - B^T Q^{-1} B & \\ & Q \end{bmatrix} \quad \text{and} \quad T^T R T = \begin{bmatrix} P & \\ & Q - BP^{-1}B^T \end{bmatrix}.$$

Deduce that R is symmetric and positive definite if and only if $P - B^T Q^{-1} B$ is symmetric and positive definite, or $Q - BP^{-1}B^T$ is symmetric and positive definite.

2.10.11. Let

$$A = \begin{bmatrix} B & a \\ a^T & \alpha \end{bmatrix}$$

be symmetric and positive definite, and let

$$P^T = \begin{bmatrix} I & o \\ -x^T & 1 \end{bmatrix}$$

be with x such that $Bx = a$. Show that

$$P^T A P = \begin{bmatrix} B & o \\ o^T & \beta \end{bmatrix}, \ \beta > 0.$$

2.10.12. Show that if A is positive definite, then

$$A = \begin{bmatrix} \alpha^2 & a^T \\ a & B \end{bmatrix} = \begin{bmatrix} \alpha & o^T \\ \alpha^{-1} a & I \end{bmatrix} \begin{bmatrix} 1 & o^T \\ o & B - \alpha^{-2} a a^T \end{bmatrix} \begin{bmatrix} \alpha & \alpha^{-1} a^T \\ o & I \end{bmatrix}.$$

2.10.13. Write out the block triangular factorization

$$A = \begin{bmatrix} B & A^T \\ A & C \end{bmatrix} = \begin{bmatrix} L & \\ X & I \end{bmatrix} \begin{bmatrix} I & \\ & Y \end{bmatrix} \begin{bmatrix} L^T & X^T \\ & I \end{bmatrix},$$

where $B = LL^T$. Find blocks X and Y in terms of block A, B, C.

2.10.14. Suppose that $A = LL^T$ and let A' be a principle submatrix of A. Show that $A' = L'L'^T$, where L' is the corresponding submatrix of L.

2.10.15. Let symmetric and positive definite A be factored as $A = LL^T$. Show that $L_{ii}^2 \leq A_{ii}, i = 1, 2, \ldots, n$. When does equality hold?

2.11 The determinant of a matrix

We know by now that for every square matrix A there exists a permutation matrix P, a lower-triangular matrix L with unit diagonal entries, and an upper-triangular matrix U so that $PA = LU$. Matrix A is nonsingular only if upper-triangular matrix U is of type 1, and hence the product of the diagonal entries of U constitutes a scalar function of matrix A, a determinant, that determines by being zero or not whether A is singular or not.

We first conveniently assume that A is factorable, and introduce the

Definition. *Let $A = A(n \times n)$ be factorable as $A = LU$ where L is lower-triangular with $L_{ii} = 1$, and U is upper-triangular. The product*

$$\det(A) = U_{11}U_{22} \cdots U_{nn} \tag{2.137}$$

is the determinant of A.

If A is nonsingular, then we may write the LU factorization as $A = LDU,\ L_{ii} = U_{ii} = 1$, and $\det(A) = D_{11}D_{22} \cdots D_{nn}$.

Theorem 2.38. $\det(A)$ *is unique.*

Proof. If A is nonsingular, then the LU factorization of A is unique, and if A is singular, then $\det(A) = 0$ for any factorization. End of proof.

Theorem 2.39. $\det(A) = 0$ *if and only if A is singular.*

Proof. $\det(A) = 0$ if and only if the upper-triangular U is of type 0. End of proof.

Theorem 2.40. $\det(A) = \det(A^T)$.

Proof. If A is singular, the equation reduces to $0 = 0$, so assume that A is nonsingular. Then $A = LDU$ and $\det(A) = D_{11}D_{22}\cdots D_{nn}$. On the other hand $A^T = U^T D L^T = L'DU'$, ${L_{ii}}' = {U_{ii}}' = 1$, and also $\det(A^T) = D_{11}D_{22}\cdots D_{nn}$. End of proof.

Lemma 2.41. *Let U and U' be upper-triangular, and L and L' lower-triangular. Then*

1. $\det(U) = U_{11}U_{22}\cdots U_{nn}$.
2. $\det(L) = L_{11}L_{22}\cdots L_{nn}$.
3. $\det(UU') = \det(U)\det(U')$.
4. $\det(LL') = \det(L)\det(L')$.
5. $\det(U^{-1}) = 1/\det(U)$.
6. $\det(L^{-1}) = 1/\det(L)$.
7. $\det(AU) = \det(A)\det(U)$.
8. $\det(LA) = \det(L)\det(A)$

Proof.

1. Obvious.

2. $\det(L) = \det(L^T)$.

3. and 4. $(UU')_{ii} = U_{ii}{U_{ii}}', (LL')_{ii} = L_{ii}{L_{ii}}'$.

5. and 6. $(U^{-1})_{ii} = 1/U_{ii},\ (L^{-1})_{ii} = 1/L_{ii}$.

7. and 8. If $A = L'U'$, then $AU = L'U'U$ and $LA = LL'U'$.

End of proof.

Before turning to the question of the influence of row and column permutations on the determinant, and the extension of the determinant definition to matrices that do not admit an LU factorization because of a zero pivot, we find it instructive to observe the following example. Permutation matrix P below can not be LU factored but modified matrix $P'(\epsilon)$ can by virtue of a nonzero pivot ϵ:

$$P = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, \text{ and } P'(\epsilon) = \begin{bmatrix} \epsilon & 1 \\ 1 & \end{bmatrix} = \begin{bmatrix} 1 & \\ \epsilon^{-1} & 1 \end{bmatrix} \begin{bmatrix} \epsilon & 1 \\ & -\epsilon^{-1} \end{bmatrix}. \tag{2.138}$$

But $\det(P'(\epsilon)) = -1$ for *any* ϵ no matter how small.

Theorem 2.42. *Let P be a permutation matrix affecting a succession of p row interchanges, and P' a permutation matrix affecting a succession of p' column interchanges. Then*

$$\det(PAP') = (-1)^{p+p'} \det(A). \tag{2.139}$$

Proof. Because $\det(A) = \det(A^T)$, and $(PA)^T = A^T P^T$ it is enough that we prove the theorem for the permutation of rows only. Also, the proof is done if we prove that for one swap involving only two rows $\det(PA) = -\det(A)$, and for this we need consider only successive rows since the interchange of rows i and $i+k$ can be achieved through a succession of $2k-1$ interchanges of adjacent rows. For instance

$$\begin{array}{ccccccccccccccc} & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 \\ i=2 & 2 & & 5 & & 3 & & 3 & & 3 & & 3 & & 5 \\ & 3 & \rightarrow & 3 & = & 2 & \rightarrow & 4 & \rightarrow & 4 & \rightarrow & 5 & \rightarrow & 3 \\ & 4 & & 4 & & 4 & & 2 & & 5 & & 4 & & 4 \\ i+k=5 & 5 & & 2 & & 5 & & 5 & & 2 & & 2 & & 2 \\ & k=3 & & & & & & & & & & & & \end{array} . \tag{2.140}$$

Moreover, because the effect of the row interchanges is not felt in the factorization until the pivot reaches down to the first displaced row, we may assume that the interchanged rows are 1 and 2. Finally, in view of statements 7. and 8. of lemma 2.41 it is enough that we prove that $\det(PL) = -1$ if $L_{ii} = 1$. Indeed,

$$\begin{bmatrix} 1 & & \\ L_{21} & 1 & \\ L_{31} & L_{32} & 1 \end{bmatrix} \rightarrow \begin{bmatrix} L_{21} & 1 & \\ 1 & & \\ L_{31} & L_{32} & 1 \end{bmatrix} \rightarrow \begin{bmatrix} L_{21} & 1 & \\ & -L_{21}^{-1} & \\ & & 1 \end{bmatrix} .$$

End of proof.

Now we use Theorem 2.42 to extend the definition of the determinant to any square matrix as $\det(A) = (-1)^{p+p'} \det(PAP')$ for any permutation of rows and columns.

Before proceeding with more theorems on the determinant we introduce permutation matrix

$$P = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix} , \ P^2 = I \tag{2.143}$$

that has this remarkable property: that if L is lower-triangular, then $PLP = U$ is upper-triangular, and if U is upper-triangular, then $PUP = L$ is lower-triangular.

The next theorem introduces two fundamental properties of determinants that by themselves make determinants worth considering.

Theorem 2.44. *Let A and B be square matrices. Then*

1. $\det(A^{-1}) = 1/\det(A)$.

2. $\det(AB) = \det(A)\det(B)$.

Proof.

1. Let P be the permutation matrix of eq.(2.00).If $A = LDU,\ L_{ii} = U_{ii} = 1$, then $A^{-1} = U^{-1}D^{-1}L^{-1}$ and

$$\det(A^{-1}) = \det(PA^{-1}P) = \det((PU^{-1}P)(PD^{-1}P)(PL^{-1}P)).$$

Now, $PU^{-1}P$ is lower-triangular, $PL^{-1}P$ is upper-triangular, and

$$\det(A^{-1}) = \det(PD^{-1}P) = \det(D^{-1}).$$

2.If one of the matrices is singular, then the equation reduces to $0 = 0$, so we assume that both A and B are nonsingular. Now

$$\det(AB) = \det(LDUL'D'U') \tag{2.145}$$

where $L_{ii} = {L_{ii}}' = U_{ii} = {U_{ii}}' = 1$. Hence by Lemma 2.41

$$\det(AB) = \det(DUL'D'). \tag{2.146}$$

Since D and D' are diagonal matrices of type 1, there are upper-and lower-triangular matrices U' and L'', ${U_{ii}}' = {L_{ii}}'' = 1$, so that $DU = U'D$ and $L'D' = D'L''$, and

$$\begin{aligned} \det(AB) &= \det((PU'P)(PDD'P)(PL''P)) \\ &= \det(PDD'P) = \det(DD') = \det(D)\det(D') = \det(A)\det(B). \end{aligned} \tag{2.147}$$

End of proof.

Notice that

$$P = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \tag{2.000}$$

and hence by the formality of $\det(AB) = \det(A)\det(B),\ \det(P) = -1$.

The *numerical* LU factorization of PA furnishes $\det(A)$. But writing $\det(A)$ *algebraically* in terms of the entries of A becomes an immense task unless drastic notational abbreviations are undertaken.

For $A = A(2 \times 2)$ the LU factorization is with

$$U = \begin{bmatrix} A_{11} & A_{12} \\ 0 & (A_{11}A_{22} - A_{12}A_{21})/A_{11} \end{bmatrix} \tag{2.148}$$

and

$$\det(A) = A_{11}A_{22} - A_{12}A_{21} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \tag{2.149}$$

which is our determinant of sec.1.2. The algebraic LU factorization of $A = A(3 \times 3)$ yields

$$U = \begin{bmatrix} \Delta_1 & A_{12} & A_{13} \\ & \Delta_2/\Delta_1 & (A_{23}A_{11} - A_{21}A_{13})/\Delta_1 \\ & & \Delta_3/\Delta_2 \end{bmatrix} \tag{2.150}$$

with

$$\Delta_1 = A_{11} \ , \ \Delta_2 = \begin{vmatrix} A_{11} & A_{22} \\ A_{21} & A_{22} \end{vmatrix} \ , \ \Delta_3 = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \tag{2.151}$$

and

$$\begin{aligned} \det(A) = \Delta_3 = {} & A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} + A_{12}A_{23}A_{31} \\ & - A_{12}A_{21}A_{33} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}. \end{aligned} \tag{2.152}$$

Notice that

$$\Delta_3 = \sum_1^6 \pm A_{1i}A_{2j}A_{3k} \tag{2.153}$$

in which indices ijk run over all six permutations 123, 132, 231, 213, 312, 321, and in which the plus sign is taken for an even number of permutations, and the minus sign for an odd. Equation (2.153) is often taken as the starting point for the theory of determinants, and from it all their properties are deduced.

Cramer's rule, if not practical, is at least historically interesting. To prove it we write $x = A^{-1}f$ in the inflated form $X_i = A^{-1}F_i, \ i = 1, 2, \ldots, n$, where the columns of F_i are the columns of A except for the ith which is f; where the columns of X_i are the columns of I, except for the ith which is x,

$$[e_1 \ e_2 \ldots x \ldots e_n] = A^{-1}[a_1 \ a_2 \ldots f \ldots a_n]. \tag{2.154}$$

It is easily shown that $\det(X_j) = x_j$, and hence by the rules that $\det(AB) = \det(A)$ $\det(B)$ and $\det(A^{-1}) = 1/\det(A)$,

$$x_i = \det(A^{-1}F_i) = \det(F_i)/\det(A) \tag{2.155}$$

which is Cramer's rule.

The theoretical artifice of putting a minute nonzero value for a truly zero pivot in order to guarantee an LU factorization raises the question of the dependence of $\det(A)$ on the coefficients of matrix A. Can small changes in A_{ij} result in an explosive change in $\det(A)$, or does $\det(A)$ remain barely affected?

Theorem 2.45. $\det(A)$ *depends continuously upon the coefficients of* A.

Proof. Suppose that nonsingular matrix A is slightly changed by the addition of E into $A+E$. Then $\det(A+E) = \det(A(I+A^{-1}E)) = \det(A)\det(I+A^{-1}E)$. As the coefficients of E become ever smaller, the pivots of $I+A^{-1}E$ become all nearly 1, and the off-diagonal entries much smaller than 1, and hence $\det(I+A^{-1}E)$ cannot be far from 1.

In case A is singular we write $A+E = G(H+G^{-1}E)$, where H is the Hermite form of A. The 1 pivots of H are only slightly changed by the addition of the small entries of $G^{-1}E$, while the originally zero pivots of H are small and remain so in forward elimination, and $\det(H+G^{-1}E)$ is small. End of proof.

The continuous dependence of $\det(A)$ upon A_{ij} is obvious in eq. (2.153).

The continuous dependence of $\det(A)$ on A_{ij} does not mean that small changes in the coefficients of A may not cause *practically* large changes in $\det(A)$, as we have discussed in sec.1.2 and demonstrated numerically in eq. (1.19).

Exercises

2.11.1. Compute, without row or column interchanges, $\det(A)$

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

2.11.2. Show that if

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

then rank $(D)=$ rank(A) +rank(B). Also that $\det(D) = \det(A)\det(B)$.

2.11.3. Prove that

$$\det \begin{bmatrix} A & C \\ & B \end{bmatrix} = \det(A)\det(B)$$

and that

$$|\det \begin{bmatrix} C & A \\ B & \end{bmatrix}| = |\det(A)\det(B)|.$$

2.11.4. Prove that

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(D - CA^{-1}B)\det(A) = \det(A - BD^{-1}C)\det(D).$$

2.11.5. Prove that if $CD^T + DC^T = O$, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD^T + BC^T)$$

provided D is nonsingular.

2.11.6. Prove that if $CD^T + DC^T = O$, then

$$|\det \begin{bmatrix} A & B \\ C & D \end{bmatrix}| = |\det(AD^T + BC^T)|.$$

2.11.7. Let $A = LL^T$, and show that

$$\max_{i,j} |L_{ij}| \leq \max_i \sqrt{A_{ii}}.$$

2.11.8. Prove that if $A = A(n \times n)$ is symmetric and positive definite, then

$$\det(A) \leq A_{11}A_{22}\ldots A_{nn}.$$

Show that equality occurs if and only if A is diagonal.

2.11.9. Prove that if the entries of both A and A^{-1} are integers,then $\det(A) = \pm 1$.

2.11.10. Show that the entries of diagonal matrix $I' = I'(2 \times 2), I'_{ii} = \pm 1$ can be arranged so that for any $A(2 \times 2)$, $\det(A + I') \neq 0$. Generalize to higher-order matrices.

2.11.11. Prove that if $A = A(n \times n) = -A^T$, then for odd n $\det(A) = 0$.

2.11.12. Prove that a necessary and sufficient condition for $A = A^T$ to be positive definite is that $\det(S) > 0$ for any diagonal submatrix S of A.

2.12.13. Let $B = B(n \times n) = [b_1\ b_2 \quad b_n]$ be nonsingular so that $A = B^T B$ is symmetric positive definite. Show that

$$\det{}^2(B) \leq ({b_1}^T b_1)({b_2}^T b_2) \ldots ({b_n}^T b_n).$$

When does equality hold?

2.12.14. Show that $\det(I + uv^T) = 1 + u^T v$. Find a similar expression for $\det(A + uv^T)$ where A is nonsingular.

2.12.15. Let $e = [1\ 1 \ldots 1]^T$, and $E = ee^T$. Prove that

$$\det(A + \alpha E) = \det(A)(1 + \alpha e^T A^{-1} e).$$

2.12.16. Show that elementary operations exist to the effect that

$$\begin{bmatrix} A & x \\ y^T & 1 \end{bmatrix} \rightarrow \begin{bmatrix} A - xy^T & o \\ o^T & 1 \end{bmatrix}.$$

Write the factorization

$$\begin{bmatrix} A & x \\ y^T & 1 \end{bmatrix} = \begin{bmatrix} L & \\ v^T & 1 \end{bmatrix} \begin{bmatrix} U & u \\ & \alpha \end{bmatrix}$$

and compute u, v, α, assuming that A is nonsingular. Use all this to prove that

$$\det(A - xy^T) = \det(A)(1 - y^T A^{-1} x).$$

2.11.17. Let $A = [a_1\ a_2\ a_3] = A(t)$. Show that

$$\dot{\det} A = \det[\dot{a}_1\ a_2\ a_3] + \det[a_1\ \dot{a}_2\ a_3] + \det[a_1\ a_2\ \dot{a}_3]$$

where overdot means differentiation with respect to parameter t.

2.11.18. The *trace*, $\mathrm{tr}(A)$, of the square matrix A is

$$\mathrm{tr}(A) = A_{11} + A_{22} + A_{33} + \cdots + A_{nn}.$$

Show that:

1. tr$(A+B)$ =tr(A)+tr(B).
2. tr$(\alpha A) = \alpha$tr(A).
3. tr(AB) =tr(BA).
4. tr$(B^{-1}AB)$ =tr(A).
5. $\mathrm{tr}(A^TA) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^2$.

2.11.19. For 2×2 A and B show that

$$\mathrm{tr}(AB) - \mathrm{tr}(A)\mathrm{tr}(B) = \det(A) + \det(B) - \det(A+B).$$

2.11.20. If matrices A and B share a certain property, if AB and BA possess this same property, if A^{-1} and B^{-1} possess the property, and if I possesses the property, then all such matrices form a *nonsingular multiplicative group*.

Which of the following matrices form a multiplicative group?

1. Diagonal with $D_{ii} \neq 0$.
2. Nonsingular symmetric.
3. Lower (upper) triangular of type 1.
4. Lower (upper) triangular with unit diagonal.
5. Possessing the property that $A^{-1} = A^T$.
6. Nonsingular that commute with a given matrix.
7. $A = I + aa^T$.
8. $A = I + ab^T$.
9. $A = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$.

2.12 Variational formulations

Here we digress slightly from the spirit of linear algebra to that of analysis.

We look upon

$$\phi(x) = \frac{1}{2}x^T Ax - x^T f \ , \ A = A^T \tag{2.156}$$

as a scalar function of the n variables $x_1, x_2, \ldots, x_n$. It is, in fact, a quadratic function. The gradient *grad* ϕ of $\phi(x)$ is defined as being the vector

$$grad\ \phi = \left[\frac{\partial\phi}{\partial x_1}\ \frac{\partial\phi}{\partial x_2}\ \cdots\ \frac{\partial\phi}{\partial x_n}\right]^T, \tag{2.157}$$

and by

$$\phi(x) = \frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_i x_j - \sum_i x_i f_i \tag{2.158}$$

we establish that

$$grad\ \phi = Ax - f \tag{2.159}$$

provided that $A = A^T$.

Theorem 2.46. *If A is symmetric and positive semidefinite, then the following statements are equivalent:*

1. $As = f$.
2. $\phi(s) = \min_x \phi(x)$.

Proof. Assume that $As = f$ is consistent. Then

$$\phi(x) = \frac{1}{2}(x-s)^T A(x-s) - \frac{1}{2}s^T As \tag{2.160}$$

and since $(x-s)^T A(x-s) \geq 0$, $\phi(s) = -\frac{1}{2}s^T As$ are minima of $\phi(x)$. Assume that s minimizes $\phi(x)$. A necessary condition for this is that

$$grad\ \phi = Ax - f = o. \tag{2.161}$$

Every vector that minimizes $\phi(x)$ is a solution of $Ax = f$. Because A is positive semidefinite, and in view of eq. (2.158), all extremum points of $\phi(x)$ are minimum points. End of proof.

Exercises

2.12.1. The matrix

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

is positive semidefinite. Find the minima of $\phi(x) = \frac{1}{2}x^TAx$.

2.12.2. For given vector a find the extrema of $(a^Tx)^2/x^Tx$.

2.13 On least squares

If positive definiteness is such a desirable property in matrices, why not do the following: Let

$$Ax = f \tag{2.162}$$

be a linear system with a nonsingular coefficient matrix A. Premultiplication by A^T produces the equivalent system

$$A^TAx = A^Tf \tag{2.163}$$

with a positive definite and symmetric A^TA.

One reason for not forming A^TA and A^Tf is that for large systems it is cumbersome and expensive. But there is a deeper numerical reason why we should avoid it. Consider the simple example of

$$A = \begin{bmatrix} 1 & 0.999 \\ 1 & 1 \end{bmatrix}\ ,\ EA = \begin{bmatrix} 1 & 0.999 \\ 0 & 0.001 \end{bmatrix}, \tag{2.164}$$

where E is an elementary operations matrix, and

$$A^TA = \begin{bmatrix} 2 & 1.999 \\ 1.999 & 1.998001 \end{bmatrix}\ ,\ E(A^TA) = \begin{bmatrix} 2 & 1.999 \\ 0 & 1.998001 - 1.9980005 \end{bmatrix} \tag{2.165}$$

and if the arithmetic keeps only 6 digits, A^TA is practically singular.

If the system of linear equations $Ax = f$ is *inconsistent*, then no x can be found such that

$$r = Ax - f = o. \tag{2.166}$$

However, we might settle for an x that *minimizes* $\phi(x) = \frac{1}{2}r^Tr$. If an x does exist such that $r = o$, then the minimization will find it.

For r in eq. (2.166)

$$\phi(x) = \frac{1}{2}x^TA^TAx - x^TA^Tf + \frac{1}{2}f^Tf. \tag{2.167}$$

A necessary condition for an extremum of $\phi(x)$ is that

$$\begin{aligned} grad\ \phi(x) &= A^TAx - A^Tf = o \\ &= A^T(Ax - f) = o \\ &= A^Tr = o. \end{aligned} \tag{2.168}$$

We shall not be greatly concerned with this *least squares* problem, but we still want to prove that system (2.163) *is soluble for any* $A = A(m \times n)$ matrix.

First we prove

Lemma 2.47. *For any* $A = A(m \times n)$ *and* $x \neq o, A^TAx = o$ *if and only if* $Ax = o$.

Proof. Of course if $Ax = o$ $x \neq o$, then $A^TAx = o$. To prove the converse, namely that $A^TAx = o$ only if $Ax = o$, assume that $y = Ax \neq o$ but $A^TAx = A^Ty = o$. Then $x^TA^TAx = y^Ty = 0$, which is an absurdity if $y \neq o$. Hence our assumption is wrong and the proof is done.

Theorem 2.48. *For any* $A = A(m \times n)$ *matrix, and any right-hand side vector* f, *system*

$$A^TAx = A^Tf \tag{2.169}$$

is soluble.

Proof. By Theorem 1.19 the system is soluble if and only if for any solution y of $A^TAy = o$, $y^TA^Tf = f^TAy = 0$. By Lemma 2.47 this holds true. End of proof.

exercises

2.13.1. Find the least squares solution of

$$2x - 3y = 1$$
$$2x - 3y = 2.$$

2.13.2. Write the least squares solution to

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

2.13.3. Find the least squares solution to $x = 0, y = 0, x + y = 1$, and give it a geometrical interpretation. Do the same for $-x + y = 1, x - y = 1, x + y = 0$. Also, $x = 1, y = 2, y = -2x + 2$.

Answers

Section 2.2

2.2.1. $\alpha = 0$ if $x \neq o$, $\alpha = \alpha$ if $x = o$.

2.2.2. No.

2.2.3. $x = y = o$.

2.2.4. $a^T b = b^T a = 0$,

$$ab^T = \begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix}, \quad ba^T = \begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix},$$

no.

2.2.5. c, c, 12, 12, no.

2.2.6.

$$ab^T = \begin{bmatrix} a_1b_1 & a_1b_2 \\ a_2b_1 & a_2b_2 \end{bmatrix}, \quad pq^T = \begin{bmatrix} p_1q_1 & p_1q_2 \\ p_2q_1 & p_2q_2 \end{bmatrix}, \quad b^T p = b_1p_1 + b_2p_2.$$

2.2.8. $(1 \times 1),\ (n \times n),\ (1 \times 1),\ (n \times n),\ (n \times 1),\ (1 \times n)$.

2.2.9. No.

2.2.10.

$$ab^T = \begin{bmatrix} 2 & 1 & -1 \\ -2 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}, \quad (ab^T)^2 = a(b^T a)b^T = O.$$

2.2.11.

$$(ab^T)^6 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}.$$

2.2.12. $27aa^T + 216bb^T$.

2.2.13. $(ab^T)^2 = ab^T ab^T = (b^T a)(ab^T)$.

2.2.14. $\alpha_1 = \alpha_4 = 1 + \gamma^2, \alpha_2 = \alpha_3 = 2\gamma$.

Section 2.3

2.3.1. $-2I$.

2.3.2. No.

2.3.3 $\alpha = 2$.

2.3.5.

$$AB = \begin{bmatrix} 2 & 3 \\ 4 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 3 & -2 & 3 \\ 1 & -3 & 1 \\ 2 & -1 & 2 \end{bmatrix}.$$

2.3.17. $\alpha_0 = L_{11}L_{22},\ \alpha_1 = -(L_{11} + L_{22})$.

2.3.18. $\alpha_0 = 7,\ \alpha_1 = -5$.

2.3.19. $A^8 = 987A - 377I$.

2.3.20. $\alpha\beta = 6$.

2.3.25.

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix}, \alpha^2 + \beta\gamma = 0.$$

2.3.32. $a^T b = b^T a = \alpha$.

2.3.33. $\alpha = 2$, no real α.

2.3.37. $\delta = 1 - \alpha, \beta\gamma = \alpha(1 - \alpha)$.

2.3.40. $v^T u = u^T v = -2$.

2.3.41. $\alpha = 0$.

2.3.42. $\alpha^2 = 4,\ \beta(2\alpha + \beta) = 0$.

2.3.43.

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix} \quad \alpha^2 + \beta\gamma = 1,\ \text{or } A = \begin{bmatrix} \pm 1 & \\ & \pm 1 \end{bmatrix}.$$

2.3.44. $X = \alpha I + \beta A$, for any α, β.

2.3.46.

$$X = \begin{bmatrix} \alpha & \beta & \gamma \\ & \alpha & \beta \\ & & \alpha \end{bmatrix}$$

for any α, β, γ.

2.3.47. $\beta = -2/3$.

2.3.48. $\alpha = \beta = -1/2$.

2.3.49. $\alpha = -1$, also $\alpha = 1 - 1/\beta$ if $A = \beta I$.

2.3.52.

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

for any α, β.

2.3.56.

$$P = \begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix}, \ P^3 = I.$$

2.3.58.

$$P = \begin{bmatrix} & 1 & \\ & & 1 \\ & & \end{bmatrix}, \ P^3 = O.$$

2.3.61. $c_1 = -2\alpha, c_2 = \alpha^2 - \beta^2$.

Section 2.4

2.4.5.

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

2.4.6.

$$X = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}, \ \text{no } X.$$

2.4.16.

$$B = \begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix}, \quad \alpha^2 + \beta\gamma = 1.$$

Section 2.5

2.5.1. $A^2 = I$.

2.5.3. $X = I, Y = -I$.

2.5.4. $X = -A^T$.

Section 2.6

2.6.1. The rank is 3.

2.6.6. $P_1 = [e_1\ e_2\ e_3], P_2 = [e_1\ e_3\ e_2], P_3 = [e_2\ e_1\ e_3], P_4 = [e_2\ e_3\ e_1], P_5 = [e_3\ e_1\ e_2], P_6 = [e_3\ e_2\ e_1],\ \alpha_1 = 2, \alpha_2 = 3, \alpha_3 = 1, \alpha_4 = 3$.

Section 2.7

2.7.1.

$$AB = I,\ B = \frac{1}{3}\begin{bmatrix} 3\alpha & 3\beta \\ 1 & -1 \\ 1-3\alpha & 2-3\beta \end{bmatrix}.$$

No right inverse for A'.

2.7.2.

$$B = \frac{1}{3}\begin{bmatrix} 3\alpha & 1 & 2-3\alpha \\ 3\beta & -1 & 1-3\beta \end{bmatrix}.$$

2.7.3. That A has a left inverse.

2.7.4.

$$X = \frac{1}{3}\begin{bmatrix} 3\alpha & 1 & 2-3\alpha \\ 3\beta & -1 & 1-3\beta \end{bmatrix};\ X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix},\ X_{11} + X_{12} + X_{21} + X_{22} + X_{31} + X_{32} = 1.$$

Section 2.8

2.8.1. $X = A^{-1}B, X = BA^{-1}$.

2.8.2. $X = (A - BF)^{-1}(C - BG), Y = G - FX$.

2.8.5. $BA = 2I$. 2.8.6. $\beta = -\alpha/(1+\alpha)$. For $\alpha = -1$

2.8.7. $\alpha = -1/(1 + a^T b)$.

2.8.11. $\alpha = 1/(1-\kappa), \beta = -\alpha$.

2.8.12. $C^{-1} = I - AB$.

2.8.15.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

2.8.19. $(\gamma - \beta)(\alpha - \gamma) \neq 0$.

2.8.20.

$$P^{-1} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ 1 & -2 & 1 & & \\ -1 & 3 & -3 & 1 & \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}.$$

2.8.21.

$$A^{-1} = \begin{bmatrix} -3 & -4 & 2 \\ -5 & -5 & 3 \\ 2 & 2 & -1 \end{bmatrix}, \ B^{-1} = \begin{bmatrix} 2 & -3 & -4 \\ 3 & -5 & -5 \\ -1 & 2 & 2 \end{bmatrix}, \ T^{-1} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 2 & -2 & -2 \\ -1 & -2 & 1 & 1 \\ -1 & -2 & 1 & 2 \end{bmatrix}.$$

2.8.22.

$$A^{-1} = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 5 & 4 & 3 & 2 & 1 \\ 4 & 4 & 4 & 3 & 2 & 1 \\ 3 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

2.8.23.

$$A^{-1} = \frac{1}{\alpha} \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ \alpha & -1 & -1 & -1 & -1 \end{bmatrix}, B^{-1} = \frac{1}{\alpha} \begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & 1 & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} 1 & & & & \\ -\alpha & 1 & & & \\ \alpha^2 & -\alpha & 1 & & \\ -\alpha^3 & \alpha^2 & -\alpha & 1 & \\ \alpha^4 & -\alpha^3 & \alpha^2 & -\alpha & 1 \end{bmatrix}.$$

2.8.24.

$$A^{-1} = \frac{1}{\alpha - 1} \begin{bmatrix} 0 & \alpha & -1 \\ \alpha & -\alpha - 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}, \ \alpha \neq 1.$$

2.8.25. $\alpha = 2; \alpha = 0; \alpha^2 = \sqrt{2} - 1$

2.8.27. $\lambda = \pm 1$.

2.8.28. $\lambda = 0, \lambda = -3$.

2.8.39.

$$A^{-1} = \frac{1}{1-\alpha^2} \begin{bmatrix} B^{-1} & -\alpha I \\ -\alpha I & B \end{bmatrix}.$$

2.8.41.

$$A^{-1} = \begin{bmatrix} I & o \\ -a^T & 1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} S^{-1} & o \\ -a^T S^{-1} & 1 \end{bmatrix}, \ C^{-1} = \begin{bmatrix} I + ba^T & -b \\ -a^T & 1 \end{bmatrix}.$$

2.8.71. $\alpha = v^T A u, \gamma = -1/(1 + v^T A^{-1} u)$.

Section 2.9

2.9.1. $\alpha = \pm 1$.

2.9.3. $\alpha > 1; \alpha^2 < 1$.

2.9.4. $B_{11} = 2, B_{22} = -1, B_{12} = 1 + \alpha, B_{21} = 1 - \alpha$.

2.9.10.

$$A = \begin{bmatrix} \pm 2 & \\ & \pm 3 \end{bmatrix}, A = \pm \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}.$$

2.9.11. $X_{11} = X_{22} = \pm 1, X_{12} = X_{21} = \mp 1$.

Section 2.10

2.10.1.

$$A = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & -4 & 2 \\ & -1 & -3 \\ & & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & & \\ -1/2 & 1 & \\ & -2/7 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & \\ & 7/2 & -1 \\ & & 12/7 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}.$$

Matrices B and C are positive definite.

2.10.2.

$$A = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 1 & -2 \\ & 3 & 1 \\ & & 0 \end{bmatrix},\ B = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ 1 & \alpha & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ & 0 & 1 \\ & & \beta \end{bmatrix},\ \alpha + \beta = 2.$$

2.10.3.

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

2.10.5.

$$A = \frac{1}{2} \begin{bmatrix} 0 & 2 \\ 3 & -5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 4 & 1 \\ & 2 & -1 \end{bmatrix}$$

2.10.13. $X = AL^{-T}, Y = C - AB^{-1}A^T$.