Chapter 4 Rank, Inverse and Determinants

4.1 Row and Column Spaces and The Rank

Let A be an $m \times n$ matrix partitioned into its rows:

$$A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}$$

_

_

where $\alpha_i \in \mathbf{F}^{1 \times n}$, i = 1, ..., m. The span of the rows of A is a subspace of $\mathbf{F}^{1 \times n}$, and is called the **row space** of A, denoted rs (A):

 $\operatorname{rs}(A) = \operatorname{span}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m) \subset \mathbf{F}^{1 \times n}$

If $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is linearly independent then it is a basis for rs (A). Otherwise, it can be reduced to a basis by means of elementary row operations. From the discussion in Section 1.4 and Section 3.2.1 it is clear that if

$$R = \begin{bmatrix} \boldsymbol{\rho}_1 \\ \vdots \\ \boldsymbol{\rho}_r \\ O \end{bmatrix}$$

is the reduced row echelon form of A then

- a) rs (A) = rs (R)b) { ρ_1, \dots, ρ_r } is a basis for rs (A)
- c) dim (rs(A)) = r

Thus the row rank of a matrix defined in Section 1.4 is the dimension of its row space. Now let us partition A into its columns:

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$$

where $\mathbf{a}_j \in \mathbf{F}^{m \times 1}$, j = 1, ..., n. The span of the columns of A, which is a subspace of $\mathbf{F}^{m \times 1}$, is called the **column space** of A, denoted cs (A):

$$\operatorname{cs}(A) = \operatorname{span}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \subset \mathbf{F}^{m \times 1}$$

Again, if $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ is linearly independent then it is a basis for cs(A). Otherwise, it can be reduced to a basis by means of elementary column operations on A (see Exercise 1.32). If

 $C = [\mathbf{c}_1 \cdots \mathbf{c}_\rho \ O]$

is the reduced column echelon form of A, where ρ (the number of nonzero columns of C) is the column rank of A, then

- a) $\operatorname{cs}(A) = \operatorname{cs}(C)$
- b) $\{\mathbf{c}_1, \dots, \mathbf{c}_{\rho}\}$ a basis for $\operatorname{cs}(A)$
- c) dim $(cs(A)) = \rho$

It is interesting to examine the relation between the row rank and the column rank of a matrix. Consider the reduced row echelon form of A, and rename the basic columns of A as $\mathbf{b}_1, \ldots, \mathbf{b}_r$ and the non-basic columns as $\mathbf{g}_1, \ldots, \mathbf{g}_{\nu}$, where $\nu = n - r$. Then with the notation of Section 1.5

$$\begin{bmatrix} B & G \end{bmatrix} \xrightarrow{\text{e.r.o.}} \begin{bmatrix} I_r & H \\ O & O \end{bmatrix}$$
(4.1)

or equivalently

$$\begin{bmatrix} B & \mathbf{g}_j \end{bmatrix} \xrightarrow{\text{e.r.o.}} \begin{bmatrix} I_r & \mathbf{h}_j \\ O & \mathbf{0} \end{bmatrix}, \quad j = 1, \dots, \nu$$

$$(4.2)$$

where B and G are $m \times r$ and $m \times \nu$ submatrices of A consisting of its basic and non-basic columns respectively. (4.2) implies that each of the $m \times r$ systems

$$B\mathbf{u} = \mathbf{g}_j, \quad j = 1, \dots, \nu$$

is consistent and has a solution $\mathbf{u} = \mathbf{h}_j$, that is,

$$\mathbf{g}_i = B\mathbf{h}_i, \quad j = 1, \dots, \nu$$

This shows that every non-basic column can be written as a linear combination of the basic columns. In other words,

$$\mathbf{g}_j \in \operatorname{cs}(B), \quad j = 1, \dots, \nu$$

Thus

$$\operatorname{cs}(A) = \operatorname{cs}[B \ G] = \operatorname{cs}(B)$$

Moreover, since r(B) = r, the only solution of the homogeneous system $B\mathbf{u} = \mathbf{0}$ is the trivial solution $\mathbf{u} = \mathbf{0}$. This shows that columns of B are also linearly independent, and therefore, form a basis for cs (A). Since B has r columns, we have

R1.
$$\rho(A) = r(A)$$

The common value of the row and column ranks of A is simply called the **rank** of A. Thus the row and column spaces of a given matrix, which are subspaces of different vector spaces, have the same dimension r, which is the maximum number of linearly independent rows and also the maximum number of linearly independent columns of that matrix.

Recall that the image of an $m \times n$ matrix A is defined as

$$\operatorname{im}(A) = \{ \mathbf{y} \, | \, \mathbf{y} = A\mathbf{x}, \, \mathbf{x} \in \mathbf{F}^{n \times 1} \}$$

Since for any \mathbf{x} , $A\mathbf{x}$ is a linear combination of the columns of A (coefficients being the components of \mathbf{x}), it follows that

 $\operatorname{im}(A) = \operatorname{cs}(A)$

That explains why we use the same term "rank" for both the dimension of the image of a linear transformation and the dimension of the column space of a matrix that defines a linear transformation.

Example 4.1

Let us find bases for the row and column spaces of the matrix

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

The reduced row echelon form of A is obtained as

$$A \xrightarrow{\text{e.r.o.}} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R$$

from which we conclude that r = 2, that rows 1 and 2 of R form a basis for the row space of A, and that the basic columns 1 and 3 of A form a basis for the column space of A. Let us verify these conclusions.

Any $\mathbf{x} \in \operatorname{rs}(A)$ is of the form

$$\mathbf{x} = c_1 \begin{bmatrix} 1 & 1 & -1 & 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 & 3 & -2 & 5 \end{bmatrix} + c_3 \begin{bmatrix} 2 & 2 & -1 & 3 \end{bmatrix}$$

= $(c_1 + 3c_2 + 2c_3) \begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix} + (-c_1 - 2c_2 - c_3) \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}$

Thus rows 1 and 2 of R span the row space of A. Since rows 1 and 2 of R are also linearly independent, they form a basis for rs (A).

Any $\mathbf{y} \in \operatorname{cs}(A)$ is of the form

$$\mathbf{y} = c_1 \begin{bmatrix} 1\\3\\2 \end{bmatrix} + c_2 \begin{bmatrix} 1\\3\\2 \end{bmatrix} + c_3 \begin{bmatrix} -1\\-2\\-1 \end{bmatrix} + c_4 \begin{bmatrix} 2\\5\\3 \end{bmatrix}$$

$$= (c_1 + c_2 + c_4) \begin{bmatrix} 1\\3\\2 \end{bmatrix} + (c_3 - c_4) \begin{bmatrix} -1\\-2\\-1 \end{bmatrix}$$
(4.3)

so that columns 1 and 3 of A (the basic columns) span the column space of A. It is easy to verify that they are also linearly independent, and therefore, form a basis for cs(A).

We can also find a basis for cs(A) by considering its reduced column echelon form, which is obtained by the sequence elementary column operations described below.

$\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$	$ \begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	$\begin{bmatrix} 2\\5\\3 \end{bmatrix}$	$-C_1 + C_2 \rightarrow C_2$ $C_1 + C_3 \rightarrow C_3$ $-2C_1 + C_4 \rightarrow C_4$	ſ	$\frac{1}{3}$	0 0 0	0 1 1	$\begin{bmatrix} 0\\ -1\\ -1 \end{bmatrix}$
	2 -1	പ	\longrightarrow	L	2	0	T	-1]

$C_2 \longleftrightarrow C_3$	Г	1	0	0	0]
$C_1 - 3C_2 \rightarrow C_1$		0	1	0	0
$C_2 + C_4 \rightarrow C_4$	-	1	1	0	0

Thus the nonzero columns 1 and 2 of the reduced column echelon form of A form a basis for its column space. This can also be verified by observing that a typical vector in the column space given in (4.3) can be expressed as

$$\mathbf{y} = c_1 \begin{bmatrix} 1\\3\\2 \end{bmatrix} + c_2 \begin{bmatrix} 1\\3\\2 \end{bmatrix} + c_3 \begin{bmatrix} -1\\-2\\-1 \end{bmatrix} + c_4 \begin{bmatrix} 2\\5\\3 \end{bmatrix}$$
$$= (c_1 + c_2 - c_3 + 2c_4) \begin{bmatrix} 1\\0\\-1 \end{bmatrix} + (3c_1 + 3c_2 - 2c_3 + 5c_4) \begin{bmatrix} 0\\1\\1 \end{bmatrix}$$

A square matrix of order k is called **nonsingular** if r(A) = k, and **singular** if r(A) < k. Let A be an $m \times n$ matrix, and consider the $m \times r$ submatrix B in (4.1) that consists of the basic columns of A, where r = r(A). Since r(B) = r(A) = r, dim (rs(B)) = r. Let C be the $r \times r$ submatrix of B consisting of the basic rows of B (the rows corresponding to the pivot elements in the reduced column echelon form of B). Then r(C) = r(B) = r, so that C is a nonsingular submatrix of A. This shows that if r(A) = r, then A contains an $r \times r$ nonsingular submatrix. Suppose $r < \min\{m, n\}$, and consider any $k \times k$ submatrix of A with k > r. Since any k columns of A are linearly dependent, so are the columns of this submatrix, and therefore, it must be singular. We thus conclude that

R2. the rank of a matrix is the order of its largest nonsingular submatrix.

Example 4.2

Since the matrix A in Example 4.1 has rank 2, it must have a nonsingular submatrix of order 2, and all square submatrices of order 3 must be singular.

Indeed, the 2×2 submatrix

$$\left[\begin{array}{rrrr}1 & -1\\3 & -2\end{array}\right]$$

consisting of first and second rows and first and third columns is nonsingular as can easily be shown by observing that its reduced row echelon form is I_2 .

Now consider the 3×3 submatrix consisting of columns 1, 2, 3. Since the first two columns of this submatrix are identical, subtracting one from the other produces a zero column, showing immediately that the submatrix has rank less than 2, that is, it is singular. The same is true for the submatrix consisting of columns 1, 2, 4. The 3×3 submatrices consisting of columns 1, 3, 4 and of columns 2, 3, 4 are identical, and have the reduced column echelon forms

Γ	1	0	0]	
	0	1	0	
L	-1	1	0]	

and therefore, they are also singular. Thus all 3×3 submatrices formed by picking any three columns out of four are singular.

Let C = AB. If $\mathbf{y} \in \operatorname{cs}(C)$ then $\mathbf{y} = C\mathbf{x} = A(B\mathbf{x}) \in \operatorname{cs}(A)$ for some \mathbf{x} , so that $\operatorname{cs}(C) \subset \operatorname{cs}(A)$. Hence $r(C) \leq r(A)$. Similarly, $\operatorname{rs}(C) \subset \operatorname{rs}(B)$, and therefore, $r(C) \leq r(B)$. As a result, we have

R3. $r(AB) \le \min \{ r(A), r(B) \}$

Example 4.3

Let C = AB, where

	г 1	1 1	0 7		-1	1	0
	1	1 -1	2		0	-1	-1
A =	3	3 -2	5 ,	B =	Ő	0	1
	2	2 - 1	3		0	0	1
	L 2		<u> </u>		1	0	1

We established in Example 4.1 that r(A) = 2. Also, from

$$B \xrightarrow{\text{e.c.o.}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$

we get r(B) = 3. Hence we must have $r(C) \le \min\{2,3\}$, that is, r(C) = 0, 1 or 2. Indeed, computing C as

$$C = \left[\begin{array}{rrrr} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{array} \right]$$

we find r(C) = 1.

Example 4.4

Computing the rank of a matrix may pose numerical difficulties, similar to those encountered when dealing with ill-conditioned systems, when some rows or columns are nearly linearly dependent.

Consider the matrix

	0.9502	0.2312	0.7189
A =	0.6067	0.4859	0.1208
	0.8913	0.7621	0.1292

A calculator that operates with 4-digit floating point arithmetic computes the reduced row echelon form of the matrix using Gaussian elimination as

	1.0000	0.0000	0.9999]
R =	0.0000	1.0000	-0.9998
	0.0000	0.0000	0.0000

and therefore, its row rank as r = 2. On the other hand, the same calculator computes the reduced column echelon form of A as C = I, and therefore, its column rank as $\rho = 3$. Apparently, Gaussian elimination is not reliable in computation of the rank.

The fact is that A is nonsingular, and therefore, has rank r = 3. (The reader can verify this by using MATLAB's built-in function **rank**). However, the matrix

 $\tilde{A} = \left[\begin{array}{cccc} 0.9502 & 0.2312 & 0.7190 \\ 0.6067 & 0.4859 & 0.1208 \\ 0.8913 & 0.7621 & 0.1292 \end{array} \right]$

which differs from A only in the fourth decimal digit of the element in the (1,3) position has rank $\tilde{r} = 2$. (The third column of \tilde{A} is the difference of the first two; hence \tilde{A} has only two linearly independent columns.) Thus although A has rank r = 3, it is very close to a matrix with rank $\tilde{r} = 2$. Whether A should be viewed as having rank two or rank three depends on the numerical accuracy desired in the particular application it appears.

4.2 Inverse

In Section 3.3.2 we stated the following facts concerning a linear transformation \mathcal{A} : $\mathbf{X} \to \mathbf{Y}$ without proof:

a) If \mathcal{A} is one-to-one (that is, ker $(\mathcal{A}) = \{\mathbf{0}\}$) then it has a left inverse $\hat{\mathcal{A}}_L$: $\mathbf{Y} \to \mathbf{X}$, not necessarily unique, such that

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

b) If \mathcal{A} is onto (that is, im $(\mathcal{A}) = \mathbf{Y}$) then it has a right inverse $\hat{\mathcal{A}}_R : \mathbf{Y} \to \mathbf{X}$, not necessarily unique, such that

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

c) If \mathcal{A} is both one-to-one and onto then it has a unique inverse $\hat{\mathcal{A}} : \mathbf{Y} \to \mathbf{X}$ such that

 $\hat{\mathcal{A}}(\mathcal{A}(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{X}$ and $\mathcal{A}(\hat{\mathcal{A}}(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in \mathbf{Y}$

In this section, we will prove these statements for a linear transformation defined by a matrix. To be precise, we first define left inverse, right inverse, and (two-sided) inverse of a matrix:

- a) A matrix \hat{A}_L that satisfies $\hat{A}_L A = I$ is called a **left inverse** of A.
- b) A matrix \hat{A}_R that satisfies $A\hat{A}_R = I$ is called a **right inverse** of A.
- c) A matrix \hat{A} that satisfies $\hat{A}A = A\hat{A} = I$ is called an **inverse** of A.

It is a simple exercise to show that for $A \in \mathbf{F}^{m \times n}$, ker $(A) = \{\mathbf{0}\}$ if and only if r(A) = n and im $(\mathcal{A}) = \mathbf{Y}$ if and only if r(A) = m. With this observation, we state facts (a)-(c) above and few additional facts as a theorem, whose proof will be given in the next subsection.

Theorem 4.1 Let $A \in \mathbf{F}^{m \times n}$. Then

- a) A has a left inverse \hat{A}_L if and only if r(A) = n.
- b) A has a right inverse \hat{A}_R if and only if r(A) = m.
- c) A has an inverse \hat{A} if and only if r(A) = m = n, that is, A is square and nonsingular.
- d) If r(A) = m = n then \hat{A}_L , \hat{A}_R and \hat{A} are unique and $\hat{A}_L = \hat{A}_R = \hat{A}$.

Note that since $r(A) \leq \min \{m, n\}$, A can have a left inverse only when $n \leq m$ and a right inverse only when $m \leq n$.

Assuming that parts (a)-(c) of Theorem 4.1 are true, part (d) can be proved by a simple argument: If r(A) = m = n then A has an inverse \hat{A} , which is certainly also a left inverse. Suppose that A has another left inverse \hat{A}_L that satisfies $\hat{A}_L A = I$. Then postmultiplying both sides with \hat{A} we obtain

$$\hat{A} = (\hat{A}_L A)\hat{A} = \hat{A}_L (A\hat{A}) = \hat{A}_L I = \hat{A}_L$$

contradicting the assumption. It can similarly be shown that \hat{A} is the only right inverse of A.

Because of the fact stated in Theorem 4.1(c), a nonsingular matrix is also called **invertible**. It is customary to denote the unique inverse of a square, nonsingular matrix A by A^{-1} .

4.2.1 Elementary Matrices

A matrix obtained from the identity matrix by a single elementary row or column operation is called an **elementary matrix**. Corresponding to the three types of elementary operations there are three types of elementary matrices. For example,

	0	0	1	0			1	0	0	0			1	0	0	0	
F	0	1	0	0		\mathbf{F}	0	c	0	0		F	0	1	0	0	
$L_1 =$	1	0	0	0	,	$L_2 =$	0	0	1	0	,	$L_3 =$	0	0	1	0	
	0	0	0	1			0	0	0	1			0	2	0	1	

are 4×4 elementary matrices of Type I, Type II, and Type III, respectively. E_1 is obtained from I_4 by interchanging the first and the third rows (or columns), E_2 by multiplying the second row (or column) by the scalar $c \neq 0$, and E_3 by adding 2 times the second row to the forth row (or 2 times the fourth column to the second column).

It is left to the reader to show that an elementary row operation on an $m \times n$ matrix A can be represented by premultiplying A with the corresponding $m \times m$ elementary matrix. For example, if

$$A = \begin{bmatrix} 1 & 0 & 3 & -1 \\ -2 & 1 & -4 & 3 \\ 3 & -2 & -1 & 0 \end{bmatrix} \xrightarrow{2R_1 + R_2 \to R_2} \begin{bmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & 2 & 1 \\ 3 & -2 & -1 & 0 \end{bmatrix} = B$$

then B = EA, where

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
(4.4)

is an elementary matrix obtained from I by the same elementary row operation. Similarly, an elementary column operation on an $m \times n$ matrix A can be represented by postmultiplying A with the corresponding $n \times n$ elementary matrix. Moreover, if E is an $n \times n$ elementary matrix that represents an elementary operation on the rows of a square matrix of order n, then E^t is also an elementary matrix that represents the same operation on the corresponding columns of A.

Let E be an elementary matrix that represents an elementary row operation on I, and let \hat{E} represent the inverse operation. Then clearly,

$$\hat{E}E = I$$

On the other hand, the same E can also be considered as representing an elementary column operation on I, and \hat{E} the inverse operation. Then

 $E\hat{E} = I$

As a result \hat{E} is the unique inverse of E, that is,

$$E^{-1} = \hat{E}$$

Moreover, E^{-1} is also an elementary matrix of the same type as E. For example, the inverse of the elementary matrix E in (4.4) is

$$E^{-1} = \left[\begin{array}{rrrr} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right]$$

which represents the inverse operation operation of adding -2 times the first row to the second row.

If E_1, E_2, \ldots, E_k are elementary matrices of order n, the product

$$E_s = E_k \cdots E_2 E_s$$

represents a sequence of elementary row operations on I_n . Then the product

$$\hat{E}_s = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$

represents a sequence of elementary row operations that undo the operations represented by E_s , so that $\hat{E}_s E_s = I$. Similarly, $E_s \hat{E}_s = I$. We thus conclude that $\hat{E}_s = E_s^{-1}$, that is,

$$(E_k \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$
(4.5)

An elementary matrix of Type I is also called an **elementary permutation** matrix for the obvious reason that it permutes (reorders) the rows or columns of the matrix that it multiplies. The reader can show that if P is an elementary permutation matrix then

$$P^{-1} = P^t$$

Let $P_s = P_k \cdots P_2 P_1$, where P_1, P_2, \ldots, P_k are elementary permutation matrices. Since a permutation followed by another permutation is also a permutation, we can conveniently call P_s a **permutation matrix**.¹ Note that a permutation matrix contains a single 1 in every row and column. The inverse of such a permutation matrix can be found by means of (4.5) to be

$$P_s^{-1} = P_1^{-1} P_2^{-2} \cdots P_k^{-1} = P_1^t P_2^t \cdots P_k^t = (P_k \cdots P_2 P_1)^t = P_s^t$$

For example,

0	0	1	0	$ ^{-1}$	0	0	1	0		0	0	0	1]
0	1	0	0		0	1	0	0		0	1	0	0
0	0	0	1	=	0	0	0	1	=	1	0	0	0
1	0	0	0		1	0	0	0		0	0	1	0

¹Permutations are discussed in Section 4.4.

4.2.2 Left, Right and Two-Sided Inverses

If $A \in \mathbf{F}^{m \times n}$ has a left inverse $\hat{A}_L \in \mathbf{F}^{n \times m}$ so that $\hat{A}_L A = I_n$ then

$$r(A_L A) = n \le r(A) \le \min\{m, n\}$$

and we must have r(A) = n. This proves the necessity part of Theorem 4.1(a). (As a byproduct, we also find that $r(\hat{A}_L) = n$.) Conversely, if r(A) = n then

$$E_q \cdots E_2 E_1 A = QA = R = \begin{bmatrix} I_n \\ O \end{bmatrix}$$
(4.6)

for some elementary matrices E_1, \ldots, E_q , where R is the reduced row echelon form of A, and Q represents the sequence of elementary row operations used to transform A into R. Partitioning rows of Q in accordance with the partitioning of R, (4.6) can be written as

$$\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} A = \begin{bmatrix} Q_1 A \\ Q_2 A \end{bmatrix} = \begin{bmatrix} I_n \\ O \end{bmatrix}$$
(4.7)

from which we observe that

$$\hat{A}_L = Q_1 = R^t Q$$

is a left inverse of A. Thus we not only prove the sufficiency part of Theorem 4.1(a), but also give a method to construct a left inverse when the sufficiency condition is satisfied.

Example 4.5

Consider the matrix

$$A = \left[\begin{array}{rrr} 1 & 1 \\ 2 & 3 \\ 0 & 1 \end{array} \right]$$

in Example 3.37. The elementary row operations that transform A into its reduced row echelon form can be summarized as

$\left[\begin{array}{c}1\\0\\0\end{array}\right]$	-1 1 0	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \\ \end{bmatrix}$	$ \begin{array}{cccc} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{array} $	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \\ \end{bmatrix}$	$\begin{array}{c}1\\-2\\0\end{array}$	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1\\ 3\\ 1 \end{bmatrix}$
=	$\left[\begin{array}{c}3\\-2\\2\end{array}\right]$	-1 1 -1	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$	=	$\left[\begin{array}{c}1\\0\\0\end{array}\right]$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	

A left inverse of A is thus obtained as

$$\hat{A}_L = \left[\begin{array}{rrr} 3 & -1 & 0 \\ -2 & 1 & 0 \end{array} \right]$$

Note that \hat{A}_L above is different from both of the left inverses mentioned in Example 3.37.

The proof of part (b) of Theorem 4.1 follows similar lines: If $A \in \mathbf{F}^{m \times n}$ has a right inverse $\hat{A}_R \in \mathbf{F}^{n \times m}$ so that $A\hat{A}_R = I_m$ then from

$$r(A\hat{A}_R) = m \le r(A) \le \min\{m, n\}$$

we get r(A) = m. (We also have $r(\hat{A}_R) = m$.) On the other hand, if r(A) = m then

$$AE_1E_2\cdots E_p = AP = C = \begin{bmatrix} I_m & O \end{bmatrix}$$

$$\tag{4.8}$$

where C is the reduced column echelon form of A, and P represents the sequence of elementary column operations used to transform A into C. Partitioning columns of P in accordance with the partitioning of C, (4.8) can be written as

$$A[P_1 \quad P_2] = [AP_1 \quad AP_2] = [I_m \quad O]$$

$$(4.9)$$

from which a right inverse of A is obtained as

$$\hat{A}_R = P_1 = PC^t$$

Example 4.6

The matrix

$$B = \left[\begin{array}{rrr} 1 & 2 & 1 \\ 0 & 1 & 2 \end{array} \right]$$

considered in Example 3.38 can be transformed into its reduced column echelon form as

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

from which a right inverse of B is obtained as

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

Finally, part (c) of Theorem 4.1 follows from parts (a) and (b): If A has an inverse \hat{A} then it is also a left inverse and a right inverse so that r(A) = m = n, that is, A is square and nonsingular. Conversely, if r(A) = m = n then the row echelon form of A is I_n so that (4.6) reduces to

$$E_q \cdots E_2 E_1 A = Q A = I_n \tag{4.10}$$

Thus Q is a left inverse of A. Premultiplying both sides of (4.10) with the product $E_1^{-1}E_2^{-1}\cdots E_q^{-1}$, we obtain²

$$A = E_1^{-1} E_2^{-2} \cdots E_q^{-1}$$

²This expression shows that every nonsingular matrix can be expressed as a product of elementary matrices.

which implies that

$$AQ = E_1^{-1} E_2^{-2} \cdots E_q^{-1} E_q \cdots E_2 E_1 = I$$

that is, Q is also a right inverse of A. Hence if r(A) = m = n then A has an inverse

$$A^{-1} = Q$$

This completes the proof of Theorem 4.1.

MATLAB provides a built-in command inv to compute the unique inverse of a square, nonsingular matrix.

From (4.10) it follows that

$$E_k \cdots E_2 E_1 [A \ I] = [I \ Q] = [I \ A^{-1}]$$

The expression above provides a convenient method to find the inverse of a nonsingular matrix by means of elementary operations as illustrated by the following example.

Example 4.7

Show that the matrix

$$A = \left[\begin{array}{rrrr} 1 & -1 & 0 \\ 2 & -1 & 2 \\ 3 & 0 & 5 \end{array} \right]$$

is nonsingular, and then find its inverse.

We form the augmented matrix $[A \ I]$, and perform elementary row operations to reduce A into its reduced row echelon form.

$\begin{bmatrix} 1 & -1 \\ 2 & -1 \\ 3 & 0 \end{bmatrix}$	$\begin{array}{c cccc} 0 & & 1 \\ 2 & & 0 \\ 5 & & 0 \end{array}$	$egin{array}{c} 0 \ 1 \ 0 \end{array}$	$\begin{bmatrix} 0\\0\\1 \end{bmatrix} \rightarrow \begin{bmatrix} 1\\0\\0 \end{bmatrix}$	$-1 \\ 1 \\ 3$	$egin{array}{ccc} 0 & \ 2 & \ 5 & \ \end{array}$	$1 \\ -2 \\ -3$	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
			$\rightarrow \begin{bmatrix} 1\\0\\0 \end{bmatrix}$	$egin{array}{c} 0 \ 1 \ 0 \end{array}$	$\begin{array}{ccc} 2 & \\ 2 & \\ -1 & \end{array}$	$-1 \\ -2 \\ 3$	$1 \\ 1 \\ -3$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
			$\rightarrow \begin{bmatrix} 1\\0\\0 \end{bmatrix}$	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	$egin{array}{ccc} 0 & \ 0 & \ 1 & \end{array}$	$5 \\ 4 \\ -3$	$-5 \\ -5 \\ 3$	$\begin{bmatrix} 2\\2\\-1 \end{bmatrix}$

Since the reduced row echelon form of A is I, it is nonsingular, and

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

MATLAB gives the same answer. The reader can also verify that $A^{-1}A = AA^{-1} = I$.

The following properties of inverse are easy to show, and are left to the reader.

I1. $I^{-1} = I$.

I2. If A is nonsingular then so is A^h , and $(A^h)^{-1} = (A^{-1})^h$.

I3. If A_1, A_2, \ldots, A_k are nonsingular matrices of order *n* then their product is also nonsingular, and

$$(A_k \cdots A_2 A_1)^{-1} = A_1^{-1} A_2^{-2} \cdots A_k^{-1}$$

Note that the third property above is a generalization of (4.5) stated for a product of elementary matrices to a product of arbitrary nonsingular matrices.

Some special matrices have special inverses. For example, a diagonal matrix

$$D = \operatorname{diag} \left[d_1, d_2, \dots, d_n \right]$$

is nonsingular if and only if $d_i \neq 0$ for all *i*, in which case

$$D^{-1} = \text{diag} [1/d_1, 1/d_2, \dots, 1/d_n]$$

Some other similar results are left to the reader as exercise.

.

4.2.3 Generalized Inverse

~

If \hat{A}_G is a left inverse or a right inverse or a two-sided inverse of A, then it certainly satisfies both of the relations

$$\hat{AA_G}A = A, \quad \hat{A}_G A \hat{A}_G = \hat{A}_G \tag{4.11}$$

If none of the rank conditions of Theorem 4.1 holds, then A does not have a left or a right inverse, nor a two-sided inverse. However, it may still be possible to construct a matrix \hat{A}_G that satisfies the above relations. Such a matrix, if it exists, is called a **generalized inverse** of A.

Let $A \in \mathbf{F}^{m \times n}$ with r(A) = r, and let

~

$$QA = \left[\begin{array}{c} Q_1\\ Q_2 \end{array}\right]A = \left[\begin{array}{c} R_1\\ O \end{array}\right] = R$$

where R is the reduced row echelon form of A and Q represents the sequence of elementary row operations that transform A into R. Since $r(R) = r(R_1) = r$

$$RP = \begin{bmatrix} R_1 \\ O \end{bmatrix} \begin{bmatrix} P_1 & P_2 \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = N$$

where N is the reduced column echelon form of R, and P represents the sequence of elementary column operations that transform R into N. The matrix

$$N = QAP = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} A \begin{bmatrix} P_1 & P_2 \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$
(4.12)

is called the **normal form** of A^{3} .

Let

$$\hat{A}_G = PN^t Q = P_1 Q_1$$

³Of course, the normal form can also be obtained by first obtaining the reduced column echelon form C of A and then finding the reduced row echelon form of C.

Noting that

$$A = Q^{-1} N P^{-1} \tag{4.13}$$

 $NN^tN = N$, and $N^tNN^t = N^t$, straightforward multiplications give

$$\begin{aligned} A\hat{A}_G A &= (Q^{-1}NP^{-1})(PN^tQ)(Q^{-1}NP^{-1}) \\ &= Q^{-1}NN^tNP^{-1} = Q^{-1}NP^{-1} = A \\ \hat{A}_G A\hat{A}_G &= (PN^tQ)A(PN^tQ) \\ &= PN^tNN^tQ = PN^tQ = \hat{A}_G \end{aligned}$$

Hence \hat{A}_G is a generalized inverse of A.

The MATLAB command pinv(A) computes a special generalized inverse of A, which reduces to a left inverse when r = n and to a right inverse when r = m.

Example 4.8

The matrix

$$A = \left[\begin{array}{rrr} 1.0 & -0.8 & 0.6 \\ -0.5 & 0.4 & -0.3 \end{array} \right]$$

can be reduced to its normal form by means of elementary operations that are summarized as

$$\left[\begin{array}{rrrr} 1.0 & 0.0 \\ 0.5 & 1.0 \end{array}\right] \left[\begin{array}{rrrr} 1.0 & -0.8 & 0.6 \\ -0.5 & 0.4 & -0.3 \end{array}\right] \left[\begin{array}{rrrr} 1.0 & 0.8 & -0.6 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{array}\right] = \left[\begin{array}{rrrr} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

A generalized inverse of A is then obtained as

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

The normal form of A can also be obtained by a different sequence of elementary operations:

$$\left[\begin{array}{ccc} 0.0 & -2.0 \\ 1.0 & 2.0 \end{array}\right] \left[\begin{array}{ccc} 1.0 & -0.8 & 0.6 \\ -0.5 & 0.4 & -0.3 \end{array}\right] \left[\begin{array}{cccc} 1.0 & -0.8 & 0.6 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{array}\right] = \left[\begin{array}{cccc} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

which result in a different generalized inverse

$$\hat{A}_G = \begin{bmatrix} 1\\0\\0 \end{bmatrix} \begin{bmatrix} 0&-2\\0&-2 \end{bmatrix} = \begin{bmatrix} 0&-2\\0&0\\0&0 \end{bmatrix}$$

The MATLAB command pinv computes yet another generalized inverse

$$\hat{A}_G = \left[\begin{array}{ccc} 0.4000 & -0.2000 \\ -0.3200 & 0.1600 \\ 0.2400 & -0.1200 \end{array} \right]$$

* 4.3 Equivalence and Similarity

Recall from Section 3.2.1 that if Q is the matrix of transition from a basis \mathbf{R} of an n-dimensional vector space \mathbf{X} to another basis \mathbf{R}' , and if P is the matrix of transition from \mathbf{R}' to \mathbf{R} , then $QP = PQ = I_n$. This shows that Q and P are both nonsingular, and are inverses of each other. Conversely, any nonsingular $n \times n$ matrix Q can be viewed as a matrix of transition from a basis \mathbf{R} to another basis \mathbf{R}' in some n-dimensional vector space \mathbf{X} , say $\mathbf{F}^{n\times 1}$. In other words, if $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)$ is a basis for \mathbf{X} , and \mathbf{r}'_i are defined as

$$\mathbf{r}'_{j} = \sum_{i=1}^{n} p_{ij}\mathbf{r}_{i}, \quad j = 1, \dots, n$$

$$(4.14)$$

where $P = [p_{ij}]$ is nonsingular, then $\mathbf{R}' = (\mathbf{r}'_1, \ldots, \mathbf{r}'_n)$ is also a basis for **X** (see Exercise 4.21). Moreover, the matrix of transition from **R** to **R**' is precisely $Q = P^{-1}$.

Also recall from Section 3.3.1 that if a linear transformation $\mathcal{A} : \mathbf{X} \to \mathbf{Y}$ has a matrix representation A with respect to a pair of bases (\mathbf{R}, \mathbf{S}) and a representation A' with respect to another pair $(\mathbf{R}', \mathbf{S}')$, then

$$A' = Q_m A P_n \tag{4.15}$$

where Q_m is the nonsingular matrix of transition from **S** to **S'** in **Y**, and P_n is the nonsingular matrix of transition from **R'** to **R** in **X**. (Subscripts *n* and *m* refer to the dimensions of **X** and **Y**.) In particular, if $\mathcal{A} : \mathbf{F}^{n \times 1} \to \mathbf{F}^{m \times 1}$ is a linear transformation defined by an $m \times n$ matrix \mathcal{A} , then its representation with respect to the canonical bases ($\mathbf{E}^n, \mathbf{E}^m$) of $\mathbf{F}^{n \times 1}$ and $\mathbf{F}^{m \times 1}$ is the \mathcal{A} matrix itself (see Example 3.33). Thus if \mathcal{A}' is an $m \times n$ matrix that is related to \mathcal{A} as in (4.15) then it represents the same linear transformation with respect to a different pair of bases, which are uniquely defined by the matrices P_n and $P_m = Q_m^{-1}$.

From the discussion in Section 4.1 we observe that two $m \times n$ matrices A' and Aare row equivalent if they are related as $A' = Q_m A$, where Q_m is an $m \times m$ nonsingular matrix that stands for the elementary row operations performed on A to obtain A'. Thus all row equivalent $m \times n$ matrices represent the same linear transformation with respect to a fixed basis in $\mathbf{F}^{n\times 1}$ and different bases in $\mathbf{F}^{m\times 1}$. Their common (unique) reduced row echelon form can be considered as a canonical form that represents the equivalence class formed by these row equivalent matrices. Similarly, two $m \times n$ matrices A' and A are column equivalent if they are related as $A' = AP_n$, where P_n is an $n \times n$ nonsingular matrix that stands for the elementary column operations performed on A to obtain A'. All column equivalent $m \times n$ matrices represent the same linear transformation with respect to a fixed basis in $\mathbf{F}^{m \times 1}$ and different bases in $\mathbf{F}^{n \times 1}$. Their common (unique) reduced column echelon form is a canonical form that represents the equivalence class formed by column equivalent matrices. Combining the two types of equivalence, we call A and A' equivalent if they are related as in (4.15) for some nonsingular matrices Q_m and P_n . Thus equivalent matrices represent the same linear transformation with respect to different bases, and their common (unique) normal form is a canonical form that represents the equivalence class formed by equivalent matrices.

Example 4.9

The class of matrices that are row equivalent to

$$A = \left[\begin{array}{rrrr} 1 & 2 & 1 & 4 \\ 2 & 4 & 1 & 5 \\ 3 & 6 & 2 & 9 \end{array} \right]$$

is represented by its reduced row echelon form

$$R = \left[\begin{array}{rrrr} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

and the class of matrices that are column equivalent to ${\cal A}$ are represented by its column echelon form

$$C = \left[\begin{array}{rrrr} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{array} \right]$$

The reduced column echelon form of ${\cal R}$

$$N = \left[\begin{array}{rrrr} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

which is also the reduced row echelon form of C, represents all matrices that are equivalent to A, that is, all 4×3 matrices with rank r = 2.

When a square matrix A of order n is viewed as the representation of a linear transformation from an n-dimensional vector space \mathbf{X} into another n-dimensional vector space \mathbf{Y} , then by choosing suitable bases for \mathbf{X} and \mathbf{Y} , A can be transformed into its normal form N as in (4.12). However, if it is viewed as a linear operator on $\mathbf{F}^{n\times 1}$, that is, as a linear transformation from $\mathbf{F}^{n\times 1}$ into itself, then it is natural to use the same basis in both its domain and codomain. In this case, the equivalence relation in (4.12) becomes

$$A' = P^{-1}AP \tag{4.16}$$

Two square matrices related as in (4.16) are called **similar**. Thus similarity is a special case of equivalence. We will discuss similarity transformations in detail in the next chapter.

4.4 LU Decomposition

Some applications require solving an $n \times n$ system of equations

$$A\mathbf{x} = \mathbf{b} \tag{4.17}$$

for several values of **b**. Since the elementary row operations involved in reducing A into a row echelon form are independent of **b**, it would be a waste of time to repeat the same operations for each new value of **b**.

LU decomposition is an algorithm, based on Gaussian Elimination, for factoring a nonsingular matrix ${\cal A}$ into a product

$$A = L U \tag{4.18}$$

where L is a lower triangular matrix with unity diagonal elements and U is an upper triangular matrix.

With A factored as in (4.18), (4.17) is written as

$$LU\mathbf{x} = \mathbf{b} \tag{4.19}$$

Defining $\mathbf{z} = U\mathbf{x}$, the last equation is decomposed into two $n \times n$ systems

$$L \mathbf{z} = \mathbf{b}$$

$$U \mathbf{x} = \mathbf{z}$$
(4.20)

Since L is lower triangular, for any given **b** the first system in (4.20) can easily be solved for **z** by means of forward substitutions. Once **z** is obtained, the second system in (4.20), whose coefficient matrix is upper triangular, can be solved by means of backward substitutions to obtain a solution for **x**. If (4.17) is to be solved for a different **b**, all we have to do is to solve the two systems in (4.20) using simple forward and backward substitutions.

Example 4.10

Obtain the LU decomposition of

$$A = \left[\begin{array}{rrrr} 2 & -2 & 1 \\ -4 & 3 & -3 \\ 6 & -8 & 4 \end{array} \right]$$

Let $A_1 = A$. The Gaussian Elimination algorithm applied to the first column of A_1 yields

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

which we write in compact form as $L_1A_1 = A_2$.

Now the algorithm applied to the second column of A_2 yields

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

or $L_2 A_2 = U$.

Thus $L_2L_1A = U$, and therefore, $A = L_1^{-1}L_2^{-1}U = LU$ where

$$L = L_1^{-1}L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}$$

From the example above we observe that the first column elements of the matrix L, which are the first column elements of L_1^{-1} , can be obtained directly from the first column elements of A_1 as

$$l_{i1} = a_{i1}^{(1)} / a_{11}^{(1)}, \quad i = 2, \dots, n$$

Similarly, the second column elements of L are those that appear in L_2^{-1} , and can be obtained from the second column elements of A_2 as

$$l_{i2} = a_{i2}^{(2)} / a_{22}^{(2)}, \quad i = 3, \dots, n$$

and so on. We also observe that L and U can be stored on the original matrix A, L on the lower left half of A and U on the upper right half and on the diagonal (since the diagonal elements of L are all 1, they need not be stored). These observations lead to the basic LU decomposition algorithm given in Table 4.1, which overwrites A with L and U.

Table 4.1: Basic LU Decomposition Algorithm

1. For j = 1 : n - 12. For i = j + 1 : n3. $\mu_{ij} = a_{ij}/a_{jj}$ 4. $a_{ij} \leftarrow \mu_{ij}$ 5. For q = j + 1 : n6. $a_{iq} \leftarrow a_{iq} - \mu_{ij}a_{jq}$ 7. End 8. End 9. End

Clearly, the algorithm requires that $a_{jj} \neq 0$ at every step. If $a_{jj} = 0$ at any step, then to continue the reduction the *j*th row of A must be interchanged with a row below it to bring a nonzero element to the pivot position. Even if $a_{jj} \neq 0$, for reasons of numerical accuracy, the pivot element is chosen to be the largest element in magnitude among $\{a_{pj} : p \geq j\}$.⁴ Since row interchanges can conveniently be represented by premultiplying A with a permutation matrix P, LU decomposition of A with row interchanges is equivalent to basic LU factorization of PA. Rather than using a permutation matrix P to keep track of the row interchanges, a permutation list I serves the same purpose. With row interchanges, the LU factorization algorithm is modified as in Table 4.2.

MATLAB provides the build in function lu for obtaining the LU decomposition. The command [L,U,P]=lu(A) returns the matrices involved.

⁴This is known as partial pivoting.

Table 4.2: LU Decomposition with Partial Pivoting

I = n1. 2. For j = 1 : n - 1Find $p \ge j$ such that $|a_{pj}| = \max\{ |a_{ij}| : i \ge j \}$ 3. 4. Interchange row j of A with row pInterchange jth element of I with the pth element 5.For i = j + 1 : n6. $\mu_{ij} = a_{ij}/a_{jj}$ 7. 8. $a_{ij} \leftarrow \mu_{ij}$ 9. For q = j + 1 : n10. $a_{iq} \leftarrow a_{iq} - \mu_{ij} a_{jq}$ 11. End 12.End 13.End

Example 4.11

Obtain the LU factorization of

	Γ 1	0	2	2
4	-2	-4	2	0
$A \equiv$	4	8	0	4
	2	8	-2	6

with partial pivoting.

The steps of the algorithm are summarized below.

$$\begin{split} j &= 1: \ p = 3, \quad \mathbf{I} \to \{3, 2, 1, 4\} \\ A &\longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ -2 & -4 & 2 & 0 \\ 1 & 0 & 2 & 2 \\ 2 & 8 & -2 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ -1/2 & 0 & 2 & 2 \\ 1/4 & -2 & 2 & 1 \\ 1/2 & 4 & -2 & 4 \end{bmatrix} \\ j &= 2: \ p = 4, \quad \mathbf{I} \to \{3, 4, 1, 2\} \\ A &\longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ 1/2 & 4 & -2 & 4 \\ 1/4 & -2 & 2 & 1 \\ -1/2 & 0 & 2 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ 1/2 & 4 & -2 & 4 \\ 1/4 & -1/2 & 1 & 3 \\ -1/2 & 0 & 2 & 2 \end{bmatrix} \end{split}$$

$$j = 3: p = 4, \quad \mathbf{I} \to \{3, 4, 2, 1\}$$

$$A \longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ 1/2 & 4 & -2 & 4 \\ -1/2 & 0 & 2 & 2 \\ 1/4 & -1/2 & 1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 8 & 0 & 4 \\ 1/2 & 4 & -2 & 4 \\ -1/2 & 0 & 2 & 2 \\ 1/4 & -1/2 & 1/2 & 2 \end{bmatrix}$$

Thus

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ -1/2 & 0 & 1 & 0 \\ 1/4 & -1/2 & 1/2 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 4 & 8 & 0 & 4 \\ 0 & 4 & -2 & 4 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

and

 $P = \left[\begin{array}{rrrr} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right]$

The reader can verify that

$$LU = PA$$

4.5 Determinant of a Square Matrix

4.5.1 Permutations

A sequence of integers $\mathbf{J}_n = (j_1, j_2, \ldots, j_n)$ in which each integer from 1 to *n* appears only once is called a **permutation** of $(1, 2, \ldots, n)$. The sequence $(1, 2, \ldots, n)$ is called the **natural order** of the integers form 1 to *n*. There are *n*! permutations of *n* integers including the natural order.

In every permutation other than the natural order there is at least one integer which is followed by one or more smaller integers. The total number of integers that follow a larger integer is called the number of **inversions** in a permutation. For example, the permutation (4, 6, 1, 5, 2, 3) contains nine inversions since 4 is followed by 1, 2, and 3; 6 is followed by 1, 5, 2 and 3; and 5 is followed by 2 and 3. The **sign** of a permutation \mathbf{J}_n is defined as $s(\mathbf{J}_n) = (-1)^k$, where k is the total number of inversions. That is, a permutation has a positive sign if it contains an even number of inversions, and a negative sign if it contains an odd number of inversions.

The interchange of any two integers in a permutation is called a **transposition**. A transposition involving adjacent integers is an adjacent transposition. If the adjacent integers j_p and j_{p+1} of a permutation \mathbf{J}_n are interchanged, then the total number of inversions is either increased or decreased by exactly one depending on whether $j_p < j_{p+1}$ or $j_p > j_{p+1}$. Thus an adjacent transposition changes the sign of a permutation. Now consider the transposition of any two integers j_p and j_q with p < q. This can be achieved by first placing j_p between j_{q-1} and j_q by means of q - p - 1 forward adjacent transpositions, and then placing j_q between j_{p-1} and j_{p+1} by means of q-p backward adjacent transpositions. Thus transposition of j_p and j_q requires a total of 2q - 2p - 1 adjacent transpositions. Since 2q - 2p - 1 is an odd number, we conclude that any transposition changes the sign of a permutation.

Finally, we note that if a permutation has a total of k inversions, then it can be reduced to the natural order by k adjacent transpositions. To show this, suppose that in the permutation, n is followed by i_n smaller integers, n-1 by i_{n-1} smaller integers, etc. Then the total number of inversions is $k = i_n + i_{n-1} + \cdots + i_2$. The integer n can be brought into its natural position by i_n adjacent transpositions, at each step interchanging n with the next integer. Then, n-1 can be put into its natural position by i_{n-1} adjacent transpositions, etc. Thus the permutation can be put into natural order by $i_n + i_{n-1} + \cdots + i_2 = k$ adjacent transpositions.

4.5.2 Determinants

Let A be a square matrix of order n. The scalar associated with A

$$\det A = \sum_{\mathbf{J}_n = (j_1, \dots, j_n)} s(\mathbf{J}_n) a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$
(4.21)

where the sum is taken over all n! permutations of $(1, \ldots, n)$, is called the **determinant** of A. Thus the determinant of a 2×2 matrix is

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{22}$$

and the determinant of a 3×3 matrix is

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{array}{c} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}$$
(4.22)

Consider a typical product term $a_{1j_1}a_{2j_2}\cdots a_{nj_n}$ in (4.21). Reordering the elements so that the column indices appear in natural order, we obtain a product term $a_{i_11}a_{i_22}\cdots a_{i_nn}$, where $\mathbf{I}_n = (i_1,\ldots,i_n)$ is another permutation of the integers $(1,\ldots,n)$. Clearly, $s(\mathbf{J}_n) = s(\mathbf{I}_n)$ as the same transpositions are involved in putting \mathbf{J}_n into natural order and the natural order into \mathbf{I}_n . Also, to each product term $a_{1j_1}a_{2j_2}\cdots a_{nj_n}$ in (4.21) there corresponds a unique product term $a_{i_11}a_{i_22}\cdots a_{i_nn}$. This shows that the determinant of A can also be expressed as

$$\det A = \sum_{\mathbf{I}_n = (i_1, \dots, i_n)} s(\mathbf{I}_n) a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}$$
(4.23)

where the sum is again over all n! permutations. The expressions in (4.21) and (4.23) are called the **row expansion** and the **column expansion** of det A.

MATLAB function det computes the determinant of a square matrix.

The following properties of determinants follow from the definition.

- D1. det $A^t = \det A$.
- D2. If B is obtained from A by a Type I elementary row (column) operation, then $\det B = -\det A$.
- D3. If any two rows (columns) of A are identical, then det A = 0.
- D4. If B is obtained from A by multiplying a row (column) by a scalar c (Type II elementary operation), then det $B = c \cdot \det A$. As a consequence, if A contains a zero row (column), then det A = 0.

D5. If a row of A is expressed as the sum of two rows as

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}'_p + \boldsymbol{\alpha}''_p \\ \vdots \\ \boldsymbol{\alpha}_n \end{bmatrix}$$

then $\det A = \det A' + \det A''$, where

$$A' = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha'_p \\ \vdots \\ \alpha_n \end{bmatrix}, \quad A'' = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha''_p \\ \vdots \\ \alpha_n \end{bmatrix}$$

A corresponding property holds for columns of A.

D6. If B is obtained from A by a Type III elementary operation, then $\det B = \det A$.

To prove property D1, we note that if $A^t = B = [b_{ij}]$ so that $b_{ij} = a_{ji}$ for all (i, j) then

$$\det A^t = \sum_{\mathbf{J}_n} s(\mathbf{J}_n) b_{1j_1} b_{2j_2} \cdots b_{nj_n}$$
$$= \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} = \det A$$

To prove property D2, suppose B is obtained by interchanging the $p{\rm th}$ and $q{\rm th}$ rows of A. Then

$$\det B = \sum_{\mathbf{I}_n} s(i_1, \dots, i_p, \dots, i_q, \dots, i_n) b_{i_1 1} \cdots b_{i_p p} \cdots b_{i_q q} \cdots b_{i_n n}$$
$$= \sum_{\mathbf{I}_n} -s(i_1, \dots, i_q, \dots, i_p, \dots, i_n) a_{i_1 1} \cdots a_{i_q p} \cdots a_{i_p q} \cdots a_{i_n n}$$
$$= -\det A$$

The same property also holds if B is obtained from A by interchanging any two columns, because then B^t will be obtained from A^t by interchanging the corresponding rows so that det $B = \det B^t = -\det A^t = -\det A$. In fact, because of property D1, any result about the determinant of a matrix involving its rows is also valid for its columns, and need not be proved separately.

To prove D3, let B be obtained from A by interchanging the identical rows. Then B = A, so that det $B = \det A$. However, by property D2, we also have det $B = -\det A$. So, det A = 0.

To prove D4, let B be obtained from A by multiplying pth row by a scalar c. If $A = [a_{ij}]$ and $B = [b_{ij}]$, then

$$b_{ij} = \begin{cases} a_{ij}, & i \neq p \\ ca_{pj}, & i = p \end{cases}$$

 $\quad \text{and} \quad$

$$\det B = \sum_{\mathbf{J}_n} s(\mathbf{J}_n) b_{1j_1} \cdots b_{pj_p} \cdots b_{nj_n}$$
$$= \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots ca_{pj_p} \cdots a_{nj_n}$$
$$= c \cdot \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots a_{pj_p} \cdots a_{nj_n} = c \cdot \det A$$

Property D5 follows from

$$\det A = \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots (a'_{pj_p} + a''_{pj_p}) \cdots a_{nj_n}$$
$$= \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots a'_{pj_p} \cdots a_{nj_n} + \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots a''_{pj_p} \cdots a_{nj_n}$$
$$= \det A' + \det A''$$

Finally, to prove D6, let B be obtained from A by adding c times row p to row q, that is,

$$B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \vdots \\ \alpha_q + c\alpha_p \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \vdots \\ \alpha_q \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \vdots \\ c\alpha_p \\ \vdots \\ c\alpha_p \\ \vdots \\ \alpha_n \end{bmatrix} = A + C$$

Then, by D5 and D3, $\det B = \det A + \det C = \det A$.

Using the definition and the properties above, we can find explicit expressions for the determinants of some special matrices. For example, if A is a block upper triangular matrix of the form

$$A = \begin{bmatrix} B & \mathbf{c} \\ \mathbf{0} & a \end{bmatrix}$$
(4.24)

then since $a_{nj_n} = 0$ when $j_n \neq n$, (4.21) reduces to

$$\det A = \sum_{\mathbf{J}_n} s(\mathbf{J}_n) a_{1j_1} \cdots a_{n-1,j_{n-1}} a_{nj_n}$$
$$= \sum_{\mathbf{J}_{n-1}} s(\mathbf{J}_{n-1}) a_{1j_1} \cdots a_{n-1,j_{n-1}} a_{nn}$$

$$= a \cdot \sum_{\mathbf{J}_{n-1}} s(\mathbf{J}_{n-1}) b_{1j_1} \cdots b_{n-1,j_{n-1}}$$
$$= a \cdot \det B$$

Similarly, if

$$A = \left[\begin{array}{cc} a & \mathbf{0} \\ \boldsymbol{\gamma} & B \end{array} \right]$$

then $\det A = a \cdot \det B$.

Using the last result repeatedly we observe that if ${\cal A}$ is a lower triangular matrix then

$$\det \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = a_{11} \cdot \det \begin{bmatrix} a_{22} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{bmatrix}$$
$$\vdots$$
$$= a_{11}a_{22}\cdots a_{nn}$$

Obviously, the same is true for an upper triangular matrix. An immediate consequence of this result is that

 $\det(\operatorname{diag}[d_1, d_2, \ldots, d_n]) = d_1 d_2 \cdots d_n$

As a special case, we have

 $\det I = 1$

4.5.3 Laplace Expansion of Determinants

Consider the determinant of the 3×3 matrix in (4.22). Grouping the product terms on the right-hand side of the expression, we can write the determinant as

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$= a_{11}\det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12}\det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13}\det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

In the above expression, each determinant multiplying a first row element a_{1j} is precisely the determinant of a 2 × 2 submatrix of A which is obtained by deleting the first row and the *j*th column of A. A different grouping of the product terms in (4.22) would result in a similar expression. We now generalize this observation to matrices of arbitrary order.

Let $A = [a_{ij}]$ be a square matrix of order n. The determinant of the $k \times k$ submatrix of A obtained by deleting any n - k rows and any n - k columns of A is

called a **minor** of A. Let A_{ij} denote the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the *i*th row and the *j*th column of A, and let the corresponding minor be denoted by $m_{ij}^A = \det A_{ij}$. The signed minor $(-1)^{i+j}m_{ij}^A$ is called the **cofactor** of the element a_{ij} . We then have

D7. For any fixed $1 \le i \le n$,

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} m_{ij}^{A}$$
(4.25)

Alternatively, for any fixed $1 \le j \le n$,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} m_{ij}^{A}$$
(4.26)

The expressions in (4.25) and (4.26) are called the **Laplace expansion** of det A with respect to the *i*th row and the *j*th column.

To prove (4.25) let us first consider the special case of i = n, and express the last row of A as the sum of n rows, the *j*th one of which contains all 0's except a_{nj} at its *j*th position. Then, by repeated use of property D5, we can express det A as

$$\det A = \det A_1 + \dots + \det A_j + \dots + \det A_n$$

where

$$A_{j} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,j} & \cdots & a_{n-1,n} \\ 0 & \cdots & a_{nj} & \cdots & 0 \end{bmatrix}$$

Let B_j be obtained from A_j by moving the *j*th column to the last position by means of n - j adjacent transpositions, so that

$$\det A_j = (-1)^{n-j} \cdot \det B_j = (-1)^{n+j} \cdot \det B_j, \quad j = 1, \dots, n$$

We also observe that

$$B_j = \left[\begin{array}{cc} A_{nj} & \mathbf{b}_j \\ \mathbf{0} & a_{nj} \end{array} \right]$$

where $\mathbf{b}_{j} = \text{col} [a_{1j}, ..., a_{n-1,j}]$, so that

$$\det B_j = a_{nj} \cdot \det A_{nj} = a_{nj} m_{nj}^A$$

Hence

$$\det A = \sum_{j=1}^{n} \det A_j = \sum_{j=1}^{n} (-1)^{n+j} a_{nj} m_{nj}^A$$

This establishes (4.25) for i = n.

Now, for any fixed *i*, let *B* be the matrix obtained from *A* by moving the *i*th row to the *n*th position by means of n - i adjacent transpositions, so that det B =

 $(-1)^{n-i} \det A$. Also, for any j, $b_{nj} = a_{ij}$, and the submatrix B_{nj} is the same as the submatrix A_{ij} . Thus

$$\det A = (-1)^{i-n} \det B = (-1)^{i-n} \sum_{j=1}^{n} (-1)^{n+j} b_{nj} m_{nj}^B = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} m_{ij}^A$$

which proves (4.25) for an arbitrary *i*. Finally, (4.26) follows from (4.25) on using property D1.

Example 4.12

Find the determinant of

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

Since the second column contains most zeros, we prefer to expand det A with respect to the second column, because we do not need to calculate the cofactors of zero elements. Thus

 $\det A$

$$= (-1)^{2+2} \cdot 1. \det \begin{bmatrix} 3 & -1 & 2\\ 2 & 0 & 4\\ -4 & 1 & 1 \end{bmatrix} + (-1)^{3+2} \cdot 2. \det \begin{bmatrix} 3 & -1 & 2\\ -1 & 3 & 0\\ -4 & 1 & 1 \end{bmatrix}$$
$$= \{(-1)^{1+2} \cdot (-1) \cdot \det \begin{bmatrix} 2 & 4\\ -4 & 1 \end{bmatrix} + (-1)^{3+2} \cdot 1. \det \begin{bmatrix} 3 & 2\\ 2 & 4 \end{bmatrix} \}$$
$$-2 \cdot \{(-1)^{1+3} \cdot 2. \det \begin{bmatrix} -1 & 3\\ -4 & 1 \end{bmatrix} + (-1)^{3+3} \cdot 1. \det \begin{bmatrix} 3 & -1\\ -1 & 3 \end{bmatrix} \}$$
$$= \{(2+16) - (12-4)\} - 2 \cdot \{2 \cdot (-1+12) + (9-1)\} = -50$$

Example 4.13

Laplace expansion becomes even more convenient when coupled with elementary operations of Type III as we illustrate below.

$$\det \begin{bmatrix} i & 0 & 1+i & i \\ 0 & i & 1-i & -1 \\ i & -i & -1 & 1 \\ -1 & 0 & 2+i & i \end{bmatrix} = \det \begin{bmatrix} i & 0 & 1+i & i \\ 0 & i & 1-i & -1 \\ i & 0 & -i & 0 \\ -1 & 0 & 2+i & i \end{bmatrix}$$
$$= i \cdot \det \begin{bmatrix} i & 1+i & i \\ i & -i & 0 \\ -1 & 2+i & i \end{bmatrix}$$
$$= i \cdot \det \begin{bmatrix} i & 1+i & i \\ i & -i & 0 \\ -1-i & 1 & 0 \end{bmatrix}$$
$$= i^2 \cdot \det \begin{bmatrix} i & -i \\ -1-i & 1 \end{bmatrix}$$
$$= -1$$

From properties D2, D4 and D6 we observe that for an elementary matrix

$$\det E = \begin{cases} -1, & \text{if } E \text{ is a Type I elementary matrix} \\ c, & \text{if } E \text{ is a Type II elementary matrix} \\ 1, & \text{if } E \text{ is a Type III elementary matrix} \end{cases}$$

This observation allows us to conclude that if E is an elementary matrix then

a) det $E \neq 0$

b) det(EA) = det(AE) = (det E)(det A)

Let $R = E_k \cdots E_1 A$ be the reduced row echelon form of A, so that det $R = (\det E_k) \cdots (\det E_1)(\det A)$. If A is singular then det R = 0 (as it contains one or more zero rows), and we must have det A = 0. If, on the other hand, A is nonsingular, then R = I so that det R = 1, and therefore, det $A \neq 0$. We thus obtain the following result.

D8. A is nonsingular if and only if det $A \neq 0$.

Now consider a product AB. If A is singular then so is AB, and we have $\det AB = 0$. If A is nonsingular then we can represent it as a product of elementary matrices as $A = E_k \cdots E_1$. Then $AB = E_k \cdots E_1 B$ so that $\det AB = (\det E_k) \cdots (\det E_1)(\det B) = (\det A)(\det B)$. In either case, we have

D9. det(AB) = (det A)(det B)

An immediate consequence of Property D9 is that for a nonsingular matrix A

 $\det A^{-1} = (\det A)^{-1}$

4.5.4 Cramer's Rule and a Formula for A^{-1}

Let A be a nonsingular matrix of order n. The Cramer's rule states that the unique solution of the system

$$A\mathbf{x} = \mathbf{b} \tag{4.27}$$

is given by

$$x_j = \frac{\det B_j}{\det A} , \quad j = 1, 2, \dots, n$$

where B_j is obtained by replacing the *j*th column of A with **b**, that is,

$$B_{j} = \begin{bmatrix} a_{11} & \dots & a_{1,j-1} & b_{1} & a_{1,j+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,j-1} & b_{2} & a_{2,j+1} & \dots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,j-1} & b_{n} & a_{n,j+1} & \dots & a_{nn} \end{bmatrix}$$

Expanding det B_j with respect to the *j*th column we get

$$x_j = \frac{1}{\det A} \sum_{p=1}^n (-1)^{p+j} b_p m_{pj}^A, \quad j = 1, 2, \dots, n$$
(4.28)

That x_j 's given by (4.28) satisfy the system can be verified by substitution and using the properties of determinants as follow.

$$\sum_{j=1}^{n} a_{ij} x_j = \frac{1}{\det A} \sum_{j=1}^{n} a_{ij} \left[\sum_{p=1}^{n} (-1)^{p+j} b_p m_{pj}^A \right]$$
$$= \frac{1}{\det A} \sum_{p=1}^{n} b_p \left[\sum_{j=1}^{n} (-1)^{p+j} a_{ij} m_{pj}^A \right]$$
$$= \frac{1}{\det A} \sum_{p=1}^{n} \delta_{ip} (\det A) b_p$$
$$= b_i, \quad i = 1, 2, \dots, n$$

where we used the symbol

 $\delta_{ip} = \begin{cases} 1, & p = i \\ 0, & p \neq i \end{cases}$

to express the fact that

$$\sum_{i=1}^{n} (-1)^{p+j} a_{ij} m_{pj}^{A} = \begin{cases} \det A, & p = i \\ 0, & p \neq i \end{cases}$$

Example 4.14

Cramer's rule applied to a 2×2 system

$$\left[\begin{array}{cc}a_{11}&a_{12}\\a_{21}&a_{22}\end{array}\right]\left[\begin{array}{c}x_1\\x_2\end{array}\right] = \left[\begin{array}{c}b_1\\b_2\end{array}\right]$$

gives

$$x_{1} = \frac{1}{\det A} \det \begin{bmatrix} b_{1} & a_{12} \\ b_{2} & a_{22} \end{bmatrix} = \frac{a_{22}b_{1} - a_{12}b_{2}}{a_{11}a_{22} - a_{12}a_{21}}$$
$$x_{2} = \frac{1}{\det A} \det \begin{bmatrix} a_{11} & b_{1} \\ a_{21} & b_{2} \end{bmatrix} = \frac{a_{11}b_{2} - a_{21}b_{1}}{a_{11}a_{22} - a_{12}a_{21}}$$

or in matrix form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Since the unique solution of the system $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$, the matrix on the right-hand side of the last expression in Example 4.14 must be the inverse of the coefficient matrix. This example shows that the Cramer's rule can be used to obtain a formula for the inverse of a 2 × 2 nonsingular matrix in terms of determinants. Let us generalize this result to higher order matrices.

Let the inverse of an $n \times n$ nonsingular matrix $A = [a_{ij}]$ be expressed in terms of its columns as

 $A^{-1} = \begin{bmatrix} \hat{\mathbf{a}}_1 & \hat{\mathbf{a}}_2 & \cdots & \hat{\mathbf{a}}_n \end{bmatrix}$

Since

$$I = [\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n] = AA^{-1} = [A\hat{\mathbf{a}}_1 \ A\hat{\mathbf{a}}_2 \ \cdots \ A\hat{\mathbf{a}}_n]$$

each $\hat{\mathbf{a}}_j$ is the unique solution of the system $A\mathbf{x} = \mathbf{e}_j$. Thus if

 $\hat{\mathbf{a}}_j = \operatorname{col}\left[\hat{a}_{1j},\ldots,\hat{a}_{nj}\right]$

then by Cramer's rule we have

$$\hat{a}_{ij} = \frac{1}{\det A} \sum_{p=1}^{n} (-1)^{p+i} \delta_{pj} m_{pi}^{A}$$
$$= \frac{1}{\det A} (-1)^{i+j} m_{ji}^{A}, \quad i, j = 1, \dots, n$$

We thus obtain the formula

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A \tag{4.29}$$

for the inverse of A, where

adj
$$A = [(-1)^{i+j} m_{ij}^A]^t$$

is called the **adjugate** of A. Note that adj A is the transpose of a matrix consisting of the cofactors of A.

Example 4.15

The determinant and the minors of the matrix

$$A = \left[\begin{array}{rrrr} 1 & 2 & -1 \\ -1 & 0 & 3 \\ 1 & 2 & 0 \end{array} \right]$$

are found as

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

and

$$m_{11} = \det \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix} = -6 \qquad m_{12} = \det \begin{bmatrix} -1 & 3 \\ 1 & 0 \end{bmatrix} = -3$$
$$m_{13} = \det \begin{bmatrix} -1 & 0 \\ 1 & 2 \end{bmatrix} = -2 \qquad m_{21} = \det \begin{bmatrix} 2 & -1 \\ 2 & 0 \end{bmatrix} = 2$$
$$m_{22} = \det \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} = 1 \qquad m_{23} = \det \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = 0$$
$$m_{31} = \det \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix} = 6 \qquad m_{32} = \det \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix} = 2$$
$$m_{33} = \det \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} = 2$$

Then

$$A^{-1} = \frac{1}{2} \begin{bmatrix} -6 & 3 & -2 \\ -2 & 1 & 0 \\ 6 & -2 & 2 \end{bmatrix}^{t} = \frac{1}{2} \begin{bmatrix} -6 & -2 & 6 \\ 3 & 1 & -2 \\ -2 & 0 & 2 \end{bmatrix}$$

In practice, formula (4.29) is seldom used to calculate the inverse of a matrix. Gaussian Elimination (as in Example 4.7) is preferred for reasons of efficiency and numerical accuracy.

4.6 Exercises

- 1. Find bases for the row and the column spaces of the coefficient matrices in Exercise 1.21.
- 2. Prove that if two $m \times n$ matrices R_1 and R_2 in reduced row echelon form are row equivalent, then $R_1 = R_2$. Explain how this result implies that the reduced row echelon form of a matrix is unique. Hint: Since rs $(R_1) = rs(R_2)$, $r_1 = r_2$. Also, column indices of the leading entries of R_1 and R_2 must be the same.
- 3. Use MATLAB command rank(A) to find the rank of matrices in Exercise 1.21. Do the results agree with the results you obtained in Exercise 4.1?
- 4. A famous example of ill-conditioned matrices are Hilbert matrices.⁵ A Hilbert matrix of order n is defined as

$$H_n = \left[\frac{1}{i+j-1}\right]_{n \times n}$$

Thus

$$H_2 = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \text{ and } H_3 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

It is known that Hilbert matrices are nonsingular, that is, $r(H_n) = n$.

- (a) Use MATLAB command **rref** to find reduced row echelon forms and ranks of H_n for n = 10, 11, 12, 13.
- (b) Use MATLAB command rank to find ranks of H_n for n = 10, 11, 12, 13. Apparently, MATLAB does not use the reduced row echelon form to compute the rank of a matrix.⁶
- 5. Show that an elementary row (column) operation on an $m \times n$ matrix A is equivalent to premultiplying (postmultiplying) A with the corresponding elementary matrix.
- 6. Show that if E is an elementary matrix which represents an elementary operation on the rows of a square matrix A, then E^t represents the same operation on the corresponding columns of A.
- 7. Write down the inverses of the following elementary matrices.

	Γ1	0	3			Γ1	0	0 -	
$E_1 =$	0	1	0	,	$E_2 =$	0	5	0	
	0	0	1			0	0	1	

⁵An application involving Hilbert matrices is considered in Exercise 7.33.

 $^{^6\}mathrm{We}$ will consider the algorithm used by MATLAB in Chapter 8.

- 8. Show that if P is a permutation matrix, then $P^{-1} = P^t$.
- 9. (a) Referring to (4.7), show that $\hat{A}_L = Q_1 + XQ_2$ is a left inverse of A for arbitrary choice of X.
 - (b) Express each of the left inverses of A considered in Example 3.37 as above.
- 10. (a) Referring to (4.9), show that $\hat{A}_R = P_1 + P_2 Y$ is a right inverse of A for arbitrary choice of Y.
 - (b) Express the right inverse of B considered in Example 3.38 as above.
- 11. Let $A \in \mathbf{F}^{m \times n}$. Show that
 - (a) $\ker(A) = \{\mathbf{0}\}$ if and only if r(A) = n.
 - (b) im $(A) = \mathbf{F}^{m \times 1}$ if and only if r(A) = m.
- 12. (a) Find inverses of the following matrices by using Gaussian Elimination.
 - (b) Use MATLAB command inv to compute the inverses of the same matrices.

	г 9	9	1 7		1+i	2i	-1	-2i
4		2 1 1	1	B =	1-2i	-1-i	1+i	1
A =			1,		1-2i	-1	1+i	1-i
	L 2	1	-1 J		2i	-1	-1 - i	1+i

- 13. Write the matrices in Exercise 4.12 as products of elementary matrices.
- 14. Find the inverses of

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{n \times n}$$

by inspection.

15. Execute the following MATLAB commands and comment on the result.

A=hilb(10); B=inv(A); C=A*B

- 16. Use MATLAB command **pinv** to find a left inverse of the A matrix in Example 4.5 and a right inverse of the B matrix in Example 4.6. Verify that $\hat{A}_L A = I$ and $B\hat{B}_R = I$.
- 17. Find two different generalized inverses of the matrix

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

- 18. (a) Referring to (4.12), show that $\hat{A}_G = (P_1 + P_2 Y)(Q_1 + XQ_2)$ is a generalized inverse of A for arbitrary choices of X and Y.
 - (b) Express the generalized inverse computed by MATLAB in Example 4.8 as above.

$$A = \left[\begin{array}{cc} A_{11} & O \\ A_{21} & A_{22} \end{array} \right]$$

where A_{11} and A_{22} are square submatrices.

- (a) Show that A is nonsingular if and only if A_{11} and A_{22} are both nonsingular.
- (b) Assuming A is nonsingular, find A^{-1} in terms of A_{11}^{-1} , A_{22}^{-1} and A_{21} .
- 20. Let $A \in \mathbf{F}^{m \times n}$ and $B \in \mathbf{F}^{n \times m}$ be such that $I_m + AB$ is nonsingular.
 - (a) Show that $I_n + BA$ is also nonsingular. Hint: If $I_n + BA$ is singular, then $(I_n + BA)\mathbf{c} = \mathbf{0}$ for some $\mathbf{c} \neq \mathbf{0}$. Premultiply both sides with A.
 - (b) Show that $(I_m + AB)^{-1}A = A(I_n + BA)^{-1}$.
 - (c) Verify (b) for

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

- 21. Show that if $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ is an ordered basis for \mathbf{X} and \mathbf{r}'_j are as defined in (4.14), where $P = [p_{ij}]$ is nonsingular, then $\mathbf{R}' = (\mathbf{r}'_1, \dots, \mathbf{r}'_n)$ is also a basis for \mathbf{X} . Hint: Show that \mathbf{R}' is linearly independent.
- 22. (a) Obtain the normal forms of the coefficient matrices in Exercise 1.21.
 - (b) Verify your results by using the MATLAB commands

```
R=rref(A); % Reduced row echelon form of A
N=rref(R')'; % Reduced column echelon form of R
```

- 23. Find bases for $\mathbf{R}^{4\times 1}$ and $\mathbf{R}^{3\times 1}$ with respect to which the matrix A in Example 4.9 has the representation R, C or N.
- 24. Let A be an $m \times n$ matrix with r(A) = r. Show that it can be expressed as A = BC, where B is an $m \times r$ matrix with full column rank (that is, r(B) = r) and C is an $r \times n$ matrix with full row rank (that is, r(C) = r). Hint: Partition Q^{-1} and P^{-1} in (4.13) suitably.
- 25. Obtain the LU decompositions of the matrices in Exercises 4.12 and 4.14
 - (a) without row interchanges
 - (b) with arbitrary row interchanges.
- 26. Use MATLAB command lu to find the LU decompositions of the matrices in Exercises 4.12 and 4.14.
- 27. Obtain an LU decomposition of

$$A = \begin{bmatrix} 2 & -1 & 1 & 3 \\ -2 & 1 & -1 & -1 \\ 0 & 1 & 1 & -2 \\ 4 & -3 & 3 & 8 \end{bmatrix}$$

28. Execute the following MATLAB commands and comment on the result.

```
A=hilb(5);
[L,U,P]=lu(A);
C=PA-L*U
```

29. Explain why permuting the rows of A does not cause any difficulty in the use of LU decomposition in solving linear systems. Hint: (4.19) becomes

 $LU\mathbf{x} = P\mathbf{b}$

and accordingly, the first equation in (4.20) has to be slightly modified.

30. Find determinants of the following matrices.

A =	$\begin{bmatrix} 1\\ -2\\ 2\\ 0 \end{bmatrix}$	$\begin{array}{c} 0 \\ 0 \\ 1 \\ 1 \end{array}$	$\begin{array}{c}2\\-4\\4\\3\end{array}$	3 -5 0 4	,	<i>B</i> =		 $\begin{array}{c} 0 \\ 1 \\ \vdots \\ 0 \end{array}$	$\begin{array}{c} 1\\ 0\\ \vdots\\ 0\end{array}$	
I		T	0	1 -	1		L 1	 0	0]	$n \times n$

- 31. (a) Use MATLAB command det to find the determinant of the A matrix in Exercise 4.30.
 - (b) Use MATLAB command **rand** to generate several random matrices (of various orders), and compute their determinants using the **det** command. Observe that all the matrices you generated randomly are nonsingular.
 - (c) Generate Hilbert matrices of order n = 2,..., 20, using the MATLAB command hilb, and compute their determinants using the MATLAB command det. Comment on the result.
- 32. Show that for the block lower triangular matrix A in Exercise 4.19

 $\det A = (\det A_{11})(\det A_{22})$

Hint: The result is obvious if A_{11} is singular. If A_{11} is nonsingular, let $A_{11} = E_1 \cdots E_k$, where E_j 's are elementary matrices.

33. Let $\mathbf{p}, \mathbf{q} \in \mathbf{R}^{3 \times 1}$ be fixed linearly independent vectors. Show that the set of all $\mathbf{x} \in \mathbf{R}^{3 \times 1}$ for which

 $\det \left[\mathbf{x} \ \mathbf{p} \ \mathbf{q} \right] = 0$

is a subspace of $\mathbf{R}^{3 \times 1}$, and find a basis for it.

34. Show that the equation

$$\det \begin{bmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{bmatrix} = 0$$

describes a straight line through the points (x_1, y_1) and (x_2, y_2) in the xy plane.

35. (a) Let $\mathbf{a} \in \mathbf{R}^{n \times 1}$ and $q \in \mathbf{R}$ be given. Obtain a linear system in $\mathbf{x} \in \mathbf{R}^{n \times 1}$ whose solutions are exactly the same as the solutions of

$$\det \begin{bmatrix} 1 & \mathbf{a}^T \\ \mathbf{x} & I_n \end{bmatrix} = q$$

(b) Find all solutions of the above equation for $\mathbf{a} = \operatorname{col}[1, 2, 3]$ and q = 0.

36. Let

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ \vdots & \vdots & & \vdots \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{bmatrix}$$

Use induction on n to show that

det
$$V = \prod_{i=2}^{n} \prod_{j=1}^{i-1} (r_i - r_j)$$

V is called a **Vandermonde's matrix**.

37. Solve the following linear system of equations by using the Cramer's rule.

1	2	3]	Γ	x_1	1	Γ	1	٦
2	6	6		x_2	=		0	
1	2	1	L	x_3		L	3	

38. Use formula (4.29) to calculate inverses of A in Exercise 4.12 and C in Exercise 4.14.