Appendix C
Vector and matrix algebra

Concepts
- Scalars
- Vectors, rows and columns, matrices
- Adding and subtracting vectors and matrices
- Multiplying them by scalars
- Products of vectors and matrices, scalar and dot products
- Systems of linear equations, linear substitution
- Transposition
- Unit vectors and identity matrices
- Gauss and Gauss–Jordan elimination
- Invertible and singular matrices, inverses
- Determinants

This appendix summarizes the elementary linear algebra used in this book. Much of it is simple vector and matrix algebra that you can learn from the summary itself, particularly if you devise and work through enough two- and three-dimensional examples as you read it. Some of the techniques summarized here require you to solve systems of linear equations by methods covered in school mathematics and commonly subsumed under the title *Gauss elimination*. There are no examples here to lead you through a full review of elimination, so if you need that, you should consult a standard linear algebra text.[1] Almost all the linear algebra used in the book is two- or three- dimensional, so there's

[1] For example, Larson and Edwards 1991, chapter 1.

little need for the full multidimensional apparatus, particularly for determinants. However, many simple techniques, even in three dimensions, are best explained by the general theory summarized here.

In the context of vector and matrix algebra, numbers are often called *scalars*. For the material in this appendix, the scalars could be any complex numbers, or you could restrict them to real numbers. Applications in this book only need real scalars.

Vectors

An n-tuple (pair, triple, quadruple, ...) of scalars can be written as a horizontal row or vertical column. A column is called a *vector*. In this book, a vector is denoted by an uppercase letter; in this appendix it's in the range O to Z. Its entries are identified by the corresponding lowercase letter, with subscripts. The row with the same entries is indicated by a superscript t. For example, consider

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad\qquad X^t = [x_1, \ldots, x_n].$$

You can also use a superscript t to convert a row back to the corresponding column, so that $X^{tt} = X$ for any vector X. Occasionally it's useful to consider a scalar as a column or row with a single entry.

In analytic geometry it's convenient to use columns of coordinates for points. Coefficients of linear equations are usually arranged in rows. For points, that convention tends to waste page space. This book uses the compact notation $<x_1, x_2, x_3>$ to stand for the column $[x_1, x_2, x_3]^t$.

You can *add* two vectors with the same number of entries:

$$X + Y = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Vectors satisfy *commutative* and *associative* laws for addition:

$$X + Y = Y + X \qquad X + (Y + Z) = (X + Y) + Z.$$

Therefore, as in scalar algebra, you can rearrange repeated sums at will and omit many parentheses.

The *zero* vector and the *negative* of a vector are defined by the equations

$$O = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad\qquad -X = -\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -x_1 \\ \vdots \\ -x_n \end{bmatrix}.$$

Clearly,

$$-O = O \qquad X + O = X$$
$$-(-X) = X \qquad X + (-X) = O.$$

You can regard vector *subtraction* as composition of negation and addition. For example, $X - Y = X + (-Y)$, and you can rewrite the last equation displayed above as $X - X = O$. You should state and verify appropriate manipulation rules.

You can *multiply* a vector by a scalar:

$$Xs = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} s = \begin{bmatrix} x_1 s \\ \vdots \\ x_n s \end{bmatrix}.$$

This product is also written sX.[2] You should verify these manipulation rules:

$$X1 = X \qquad X0 = O \qquad X(-s) = -(Xs) = (-X)s$$
$$X(-1) = -X \qquad Ot = O$$

$$(Xr)s = X(rs) \qquad (\textit{associative law})$$

$$X(r + s) = Xr + Xs \qquad (\textit{distributive laws})$$
$$(X + Y)s = Xs + Ys.$$

Similarly, you can *add* and *subtract* rows X^t and Y^t with the same number of entries, and define the *zero* row and the *negative* of a row. The *product* of a scalar and a row is

$$sX^t = s[x_1, \ldots, x_n] = [sx_1, \ldots, sx_n].$$

These rules are useful:

$$X^t \pm Y^t = (X \pm Y)^t \quad -(X^t) = (-X)^t \qquad s(X^t) = (sX)^t.$$

Finally, you can multiply a row by a vector with the same number of entries to get their *scalar product*:

$$X^t Y = [x_1, \ldots, x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + \cdots + x_n y_n.$$

A notational variant used in analytic geometry is the *dot* product: $X \cdot Y = X^t Y$. (Don't omit the dot. Vector entries are often point coordinates, and

[2] Xs is more closely compatible with matrix multiplication notation, discussed later. Each form has advantages, so this book uses both.

the juxtaposition XY usually signifies the *distance* between X and Y.) With a little algebra you can verify the following manipulation rules:

$$O^tX = 0 = X^tO \qquad (sX^t)Y = s(X^tY) = X^t(Ys)$$
$$X^tY = Y^tX \qquad (-X^t)Y = -(X^tY) = X^t(-Y)$$
$$(X^t + Y^t)Z = X^tZ + Y^tZ \qquad (\textit{distributive laws})$$
$$X^t(Y + Z) = X^tY + X^tZ.$$

Matrices

An $m \times n$ *matrix* is a rectangular array of mn scalars in m rows and n columns. In this book, a matrix is denoted by an uppercase letter; in this appendix it's in the range A to O. Its entries are identified by the corresponding lowercase letter, with double subscripts:

$$A = \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}}_{n \text{ columns}} \left.\vphantom{\begin{bmatrix} a \\ \vdots \\ a \end{bmatrix}}\right] m \text{ rows.}$$

A is called *square* when $m = n$. The a_{ij} with $i = j$ are called *diagonal* entries. $m \times 1$ and $1 \times n$ matrices are columns and rows with m and n entries, and 1×1 matrices are handled like scalars.

You can *add* or *subtract* $m \times n$ matrices by adding or subtracting corresponding entries, just as you add or subtract columns and rows. A matrix whose entries are all zeros is called a *zero* matrix, and denoted by O. You can also define the *negative* of a matrix, and the *product* sA of a scalar s and a matrix A. Manipulation rules analogous to those mentioned earlier for vectors and rows hold for matrices as well; check them yourself.

You can *multiply* an $m \times n$ matrix A by a vector X with n entries; their product AX is the vector with m entries, the products of the rows of A by X:

$$AX = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}.$$

You can verify the following manipulation rules:

$$OX = O = AO \qquad (sA)X = (AX)s = A(Xs)$$
$$(-A)X = -(AX) = A(-X)$$

$$(A+B)X = AX + BX \qquad (\textit{distributive laws})$$
$$A(X+Y) = AX + AY.$$

The definition of the product of a matrix by a column was motivated by the notation for a system of m linear equations in n unknowns x_1 to x_n; you can write $AX = R$ as an abbreviation for the system

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = r_1 \\ \quad\vdots \qquad\qquad\quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = r_n \end{cases}.$$

Similarly, you can *multiply* a row X^t with m entries by an $m \times n$ matrix A; their product X^tA is the row with n entries, the products of X^t by the columns of A:

$$X^tA = [x_1, \ldots, x_m] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$
$$= [x_1a_{11} + \cdots + x_ma_{m1}, \ldots, x_1a_{1n} + \cdots + x_ma_{mn}].$$

Similar manipulation rules hold. Further, you can check the *associative law*

$$X^t(AY) = (X^tA)Y.$$

You can *multiply* an $l \times m$ matrix A by an $m \times n$ matrix B. Their product AB is an $l \times n$ matrix that you can describe two ways. Its columns are the products of A by the columns of B, and its rows are the products of the rows of A by B:

$$AB = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{l1} & \cdots & a_{lm} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1m}b_{m1} & \cdots & a_{11}b_{1n} + \cdots + a_{1m}b_{mn} \\ \vdots & & \vdots \\ a_{l1}b_{11} + \cdots + a_{lm}b_{m1} & \cdots & a_{l1}b_{1n} + \cdots + a_{lm}b_{mn} \end{bmatrix}.$$

The i,kth entry of AB is thus $a_{i1}b_{1k} + \cdots + a_{im}b_{mk}$. You can check these manipulation rules:

$$AO = O = OB \qquad (sA)B = s(AB) = A(sB)$$
$$(-A)C = -(AC) = A(-C)$$

$$(A+B)C = AC + BC \qquad (\textit{distributive laws})$$
$$A(C+D) = AC + AD.$$

The definition of the product of two matrices was motivated by the formulas for linear substitution; from

$$\begin{cases} z_1 = a_{11}y_1 + \cdots + a_{1m}y_m \\ \vdots \qquad\qquad \vdots \\ z_l = a_{l1}y_1 + \cdots + a_{lm}y_m \end{cases} \qquad \begin{cases} y_1 = b_{11}x_1 + \cdots + b_{1n}x_n \\ \vdots \qquad\qquad \vdots \\ y_m = b_{m1}x_1 + \cdots + b_{mn}x_n \end{cases}$$

you can derive

$$\begin{cases} z_1 = (a_{11}b_{11} + \cdots + a_{1m}b_{m1})x_1 + \cdots + (a_{11}b_{1n} + \cdots + a_{1m}b_{mn})x_n \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ z_l = (a_{l1}b_{11} + \cdots + a_{lm}b_{m1})x_1 + \cdots + (a_{l1}b_{1n} + \cdots + a_{lm}b_{mn})x_n \ . \end{cases}$$

That is, from $Z = AY$ and $Y = BX$ you can derive $Z = (AB)X$. In short, $A(BX) = (AB)X$. From this rule, you can deduce the *general associative law*:

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

Proof: jth column of $A(BC) = A(j$th column of $BC)$
$= A(B(j$th column of $C))$
$= (AB)(j$th column of $C)$
$= j$th column of $(AB)C$. ♦

The *commutative* law $AB = BA$ *doesn't* generally hold. For example,

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

This example also shows that the product of nonzero matrices can be O.

Every $m \times n$ matrix A has a *transpose* A^t, the $n \times m$ matrix whose j,ith entry is the i,jth entry of A:

$$A^t = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}^t = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix}.$$

The following manipulation rules hold:

$$A^{tt} = A \qquad O^t = O$$
$$(A + B)^t = A^t + B^t \qquad (sA)^t = s(A^t).$$

The transpose of a vector is a row, and vice-versa, so this notation is consistent with the earlier use of the superscript t. If A is an $l \times m$ matrix and B is an $m \times n$ matrix, then

$$(AB)^t = B^t A^t.$$

Proof: j,ith entry of $(AB)^t = i,j$th entry of AB
$= ($ ith row of $A)(j$th column of $B)$
$= ($ jth column of $B)^t(i$th row of $A)^t$
$= ($ jth row of $B)(i$th column of $A)$
$= j,i$th entry of B^tA^t. ♦

Consider vectors with n entries. Those of the jth *unit* vector U^j are all 0 except the jth, which is 1. For any row X^t with n entries, X^tU^j is the jth entry of X^t. For any $m \times n$ matrix A, AU^j is the jth column of A. For example,

$$U^1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad X^t U^1 = [x_1, \ldots, x_n] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = x_1.$$

$$AU^1 = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}.$$

The $n \times n$ matrix I whose jth column is the jth unit vector is called an *identity* matrix. Its only nonzero entries are the diagonal entries 1. Clearly, $I^t = I$. For any $m \times n$ matrix A, $AI = A$. *Proof*:

jth column of $AI = A(j$th column of $I)$
$= AU^j = j$th column of A. ♦

In particular, for any row X^t with n entries, $X^tI = X^t$.

Similarly, you may consider rows with m entries. The *unit* rows $(U^i)^t$ are the rows of the $m \times m$ identity matrix I. You can verify that for any column X with m entries, $(U^i)^tX$ is the ith entry of X. For any $m \times n$ matrix A, $(U^i)^tA$ is the ith row of A. This yields $IA = A$ for any $m \times n$ matrix A. In particular, $IX = X$ for any column X of length m.

Gauss elimination

The most common algorithm for solving a linear system $AX = R$ is called *Gauss elimination*. Its basic strategy is to replace the original system step by step with equivalent simpler ones until you can analyze the resulting system easily. Two systems are called *equivalent* if they have the same sets of solution vectors X. You need only two types of operations to produce the simpler systems:

I. interchange two equations,
II. subtract from one equation a scalar multiple of a different equation.

Obviously, type (I) operations don't change the set of solution vectors; they produce equivalent systems. If you perform a type (II) operation, subtracting s times the ith equation from the jth, then any solution of the original system clearly satisfies the modified one. On the other hand, you can reconstruct the original from the modified system by subtracting $(-s)$ times its ith row from its jth, so any solution of the modified system satisfies the original—the systems are equivalent.

The simpler systems ultimately produced by Gauss elimination have matrices A of special forms. A linear system and its matrix A are called *upper triangular* if $a_{ij} = 0$ whenever $i > j$, and *diagonal* if $a_{ij} = 0$ whenever $i \neq j$.

The first steps of Gauss elimination, called its *downward pass*, are type (I) and (II) operations that convert the original $m \times n$ system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = r_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = r_2 \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = r_m \end{cases}$$

into an equivalent upper triangular system:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \cdots + a_{1n}x_n = r_1 \\ \qquad\quad a_{22}x_2 + \cdots + a_{2m}x_m + \cdots + a_{2n}x_n = r_2 \\ \qquad\qquad \ddots \qquad\qquad\qquad\qquad \vdots \\ \qquad\qquad\qquad a_{mm}x_m + \cdots + a_{mn}x_n = r_m \end{cases} \quad \text{if } m \le n, \text{ or}$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = r_1 \\ \qquad\quad a_{22}x_2 + \cdots + a_{2n}x_n = r_2 \\ \qquad\qquad \ddots \qquad\qquad \vdots \\ \qquad\qquad\qquad a_{nn}x_n = r_n \\ \qquad\qquad\qquad\quad 0 = r_{n+1} \\ \qquad\qquad\qquad\qquad \vdots \\ \qquad\qquad\qquad\quad 0 = r_m \end{cases} \quad \text{if } m > n.$$

The algorithm considers in turn the diagonal coefficients a_{11} to $a_{m-1\,m-1}$, called *pivots*. If a pivot is zero, search downward for a nonzero coefficient; if you find one, interchange rows—a type (I) operation—to make it the new pivot. If not, then proceed to the next equation. Use a nonzero pivot a_{kk} with type (II) operations to eliminate the x_k terms from all equations after the kth. This process clearly produces an equivalent upper triangular system.

You can apply the downward pass to *any* linear system. In this book, it's used mostly with *square* systems, where $m = n$. Until the last heading of this appendix—*Gauss–Jordan elimination*—assume that's the case.

If the downward pass yields a square upper triangular matrix with no zero pivot, the original system and its matrix are called *nonsingular*. This property is independent of the right-hand sides of the equations; it depends only on the original matrix A. In the nonsingular case, you can perform more type (II) operations—constituting the *upward pass*—to convert *any* system $AX = R$ to an equivalent *diagonal* system:

$$\begin{cases} a_{11}x_1 & & & = r_1 \\ & a_{22}x_2 & & = r_2 \\ & & \ddots & \vdots \\ & & a_{nn}x_n & = r_n \; . \end{cases}$$

This system clearly has the unique solution

$$X = \langle r_1/a_{11}, r_2/a_{22}, \ldots, r_n/a_{nn} \rangle .$$

Given any $n \times p$ matrix C, you can repeat the process p times to solve equation $AB = C$ for the unknown $n \times p$ matrix B. If you solve the linear systems $AX = j$th column of C for $j = 1$ to p and assemble the solutions X as the corresponding columns of B, then $AB = C$. *Proof*:

$$\begin{aligned} j\text{th column of } AB &= A(j\text{th column of } B) \\ &= A(\text{solution } X \text{ of } AX = j\text{th column of } C) \\ &= j\text{th column of } C. \; \blacklozenge \end{aligned}$$

On the other hand, if A is *singular*—the downward pass yields an upper triangular matrix with a zero pivot—then you can construct a nonzero solution of the *homogeneous* system $AX = O$. For example, the system

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 = 0 \\ \phantom{2x_1 + {}} 0x_2 + 5x_3 = 0 \\ \phantom{2x_1 + 3x_2 + {}} 6x_3 = 0 \end{cases}$$

has solution $X = \langle -1.5s, s, 0 \rangle$ for *any* values of s. In general, proceed back up the diagonal, solving the system as though you were performing the upward pass. When you encounter a zero pivot, give the corresponding X entry an arbitrary value—the parameter s in this example. Use a distinct parameter for each zero pivot.

The previous two paragraphs are crucial for the theory of matrix inverses, hence they're worth recapitulation. If an $n \times n$ matrix A is nonsingular—the downward pass yields an upper triangular matrix with no zero pivot—then for every $n \times p$ matrix C, the equation $AB = C$ has a unique solution B. But if A is singular, then at least one such equation—in particular, $AX = O$ —has multiple solutions.

Matrix inverses

A matrix A is called *invertible* if there's a matrix B such that $AB = I = BA$. Clearly, invertible matrices must be square. A zero matrix O isn't invertible, because $OB = O \neq I$ for any B. Also, some nonzero square matrices aren't invertible. For example, for every 2×2 matrix B,

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ b_{21} & b_{22} \end{bmatrix} \neq I,$$

hence the leftmost matrix in this display isn't invertible. When there exists B such that $AB = I = BA$, it's unique; if also $AC = I = CA$, then $B = BI = B(AC) = (BA)C = IC = C$. Thus an invertible matrix A has a unique *inverse* A^{-1} such that

$$AA^{-1} = I = A^{-1}A.$$

Clearly, I is invertible and $I^{-1} = I$.

The inverse and transpose of an invertible matrix are invertible, and any product of invertible matrices is invertible:

$$(A^{-1})^{-1} = A \qquad (A^t)^{-1} = (A^{-1})^t \qquad (AB)^{-1} = B^{-1}A^{-1}.$$

Proof: The first result follows from the equations $AA^{-1} = I = A^{-1}A$; the second, from $A^t(A^{-1})^t = (A^{-1}A)^t = I^t = I$ and $(A^{-1})^tA^t = (AA^{-1})^t = I^t = I$. The third follows from $(AB)(B^{-1}A^{-1}) = ((AB)B^{-1})A^{-1} = (A(BB^{-1}))A^{-1} = (AI)A^{-1} = AA^{-1} = I$ and equation $(B^{-1}A^{-1})(AB) = I$, which you can check. ♦

The main result about inverses is that a square matrix A is invertible if and only if it's nonsingular. *Proof*: If A is nonsingular, use Gauss elimination to solve equation $AB = I$. To show that also $BA = I$, the first step is to verify that A^t is nonsingular. Were that not so, you could find $X \neq O$ such that $A^tX = O$, as mentioned under the previous heading. But then $X = I^tX = (AB)^tX = B^tA^tX = B^tO = O$ —contradiction! Thus A^t must be nonsingular, and you can solve equation $A^tC = I$. That entails

$$\begin{aligned} BA &= I^tBA^{tt} = (A^tC)^tBA^{tt} = C^tA^{tt}BA^{tt} = C^tABA^{tt} = C^tIA^{tt} \\ &= C^tA^{tt} = (A^tC)^t = I^t = I. \end{aligned}$$

Thus B is the inverse of A. Conversely, if A has an inverse B, then A must be nonsingular, for otherwise you could find $X \neq O$ with $AX = O$, which would imply $O = BAX = IX = X$ —contradiction! ♦

Determinants

The *determinant* of an $n \times n$ matrix A is

$$\det A = \textstyle\sum_\varphi a_{1,\varphi(1)} a_{2,\varphi(2)} \cdots a_{n,\varphi(n)} \operatorname{sign} \varphi,$$

where the sum ranges over all $n!$ permutations φ of $\{1, \ldots, n\}$, and $\operatorname{sign}\varphi = \pm 1$ depending on whether φ is even or odd. In each term of the sum there's one factor from each row and one from each column. For example, the permutation $1,2,3 \to 3,2,1$ is odd because it just transposes 1 and 3, so it corresponds to the term $a_{13}a_{22}a_{31}(-1)$ in the determinant sum for a 3×3 matrix A. For the theory of permutations, consult a standard algebra text.[3]

Usually you don't need the full apparatus of the theory of permutations. Most of the determinants you'll meet in this book are 2×2 or 3×3, and for them it's enough to write out the sums in full. For the 2×2 case there are two permutations of $\{1,2\}$, and

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

Clearly, the determinant of a 2×2 matrix is zero if and only if one row or column is a scalar multiple of the other.

For the 3×3 case, there are six permutations of $\{1,2,3\}$ and

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

Figure C.1 shows a handy scheme for remembering this equation. The indicated diagonals in the diagram, with their signs, contain the factors of the terms in the determinant sum.

The most important properties of determinants are closely tied to the linear system techniques summarized under the previous two headings.

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

Figure C.1 Evaluating a 3×3 determinant

First, the determinant of an upper triangular matrix is the product of its diagonal entries; in particular, identity matrices have determinant 1. *Proof*

[3] For example, Mostow, Sampson, and Meyer 1963, section 10.3.

Each term in the determinant sum except $a_{11}a_{22}\cdots a_{nn}$ contains at least one factor a_{ij} with $i > j$, and that factor must be zero. ♦

Next, if B results from a square matrix A by interchanging two rows, then $\det B = -\det A$. *Proof*: Each term in the sum for $\det B$ corresponds to a term with the opposite sign in the sum for $\det A$. ♦

A square matrix A with two equal rows has determinant zero. *Proof*: Interchanging them reverses the sign of the determinant but doesn't change the matrix. ♦

If all rows of square matrices A, B, and C are alike except that the ith row of A is the sum of the corresponding rows of B and C, then $\det A = \det B + \det C$. *Proof*: Each term in the sum for $\det A$ is the sum of the corresponding terms of $\det B$ and $\det C$. ♦

A square matrix A with a row of zeros has determinant zero. *Proof*: By the previous paragraph, $\det A = \det A + \det A$. ♦

If B results from a square matrix A by multiplying the ith row of A by a scalar s, then $\det B = s \det A$. *Proof*: Each term in the sum for $\det B$ is s times the corresponding term of $\det A$. ♦

If B results from a square matrix A by subtracting s times its ith row from its jth, then $\det B = \det A$. *Proof*: Construct C from A by replacing its jth row by $(-s)$ times its ith row, so that $\det B = \det A + \det C$. Construct D from A by replacing its jth row by its ith, so that $\det C = -s \det D$. Then D has two equal rows, so $\det D = 0$, hence $\det C = 0$, hence $\det B = \det A$. ♦

A square matrix A has determinant zero if and only if it's singular. *Proof*: By the previous discussion, $\det A$ is $(-1)^k$ times the product of the diagonal entries of the matrix that results from A through the downward pass of Gauss elimination; k is the number of row interchanges required in that process. ♦

An $n \times n$ matrix A has the same determinant as its transpose. *Proof*: $\det A^t$ is the sum of the terms $a_{\varphi(1),1}\cdots a_{\varphi(n),n} \operatorname{sign}\varphi$ for all the permutations φ of $\{1,\dots,n\}$. You can rearrange each term's factors and write it in the form $a_{1,\chi(1)}\cdots a_{n,\chi(n)} \operatorname{sign}\varphi = a_{1,\chi(1)}\cdots a_{n,\chi(n)} \operatorname{sign}\chi$, where $\chi = \varphi^{-1}$. Since the permutations φ correspond one-to-one with their inverses χ, $\det A^t = \sum_\chi a_{1,\chi(1)}\cdots a_{n,\chi(n)} \operatorname{sign}\chi = \det A$. ♦

By the previous paragraph, you can formulate in terms of columns some of the earlier results that relate determinants to properties of their rows. The next sequence of results leads slowly to the important equation $\det AB = \det A \det B$. Its proof uses some special matrices.

A *type* (I) *elementary matrix* E^{ij} results from the $n \times n$ identity matrix by interchanging its ith and jth rows, where $i \neq j$. Clearly, $\det E^{ij} = -1$. You can check that interchanging the ith and jth rows of any $n \times n$ matrix A yields the matrix $E^{ij}A$. Thus $\det E^{ij}A = \det E^{ij} \det A$ and $E^{ij}E^{ij} = I$, so E^{ij} is its own inverse.

A *type* (II) *elementary matrix* $E^{ij,c}$ results from the $n \times n$ identity matrix by subtracting c times its ith row from its jth, where $i \neq j$. Clearly, $\det E^{ij,c} = 1$. You can check that subtracting c times the ith row from the jth of any $n \times n$ matrix A yields the matrix $E^{ij,c}A$. Thus $\det E^{ij,c}A = \det E^{ij,c} \det A$ and $E^{ij,-c}E^{ij,c} = I$, so $(E^{ij,c})^{-1} = E^{ij,-c}$, another type (II) elementary matrix.

If D and A are $n \times n$ matrices and D is diagonal, then $\det DA = \det D \det A$. *Proof*: Each row of DA is the product of the corresponding row of A and diagonal entry of D. ♦

If A and B are $n \times n$ matrices, then $\det AB = \det A \det B$. *Proof*: If AB has an inverse X, then $A(BX) = (AB)X = I$, so A has an inverse. Thus if A is singular, so is AB, and $\det AB = 0 = \det A \det B$. Now suppose A is invertible. Execute the downward pass of Gauss elimination on A, performing type (I) and type (II) operations until you get an upper triangular matrix U. Each operation corresponds to left multiplication by an elementary matrix, so $E_k E_{k-1} \cdots E_2 E_1 A = U$ for some elementary matrices E_1 to E_k. The diagonal of U contains no zero, so you can perform more type (II) operations until you get a diagonal matrix D. Thus $E_l E_{l-1} \cdots E_{k+2} E_{k+1} U = D$ for some more elementary matrices E_{k+1} to E_l. This yields

$$E_l E_{l-1} \cdots E_k \cdots E_2 E_1 A = D \qquad A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} \cdots E_{l-1}^{-1} E_l^{-1} D.$$

These inverses are all elementary matrices and

$$\begin{aligned}
\det AB &= \det(E_1^{-1} E_2^{-1} \cdots E_k^{-1} \cdots E_{l-1}^{-1} E_l^{-1} DB) \\
&= \det E_1^{-1} \det E_2^{-1} \cdots \det E_k^{-1} \cdots \det E_{l-1}^{-1} \det E_l^{-1} \det D \det B \\
&= \det(E_1^{-1} E_2^{-1} \cdots E_k^{-1} \cdots E_{l-1}^{-1} E_l^{-1} D) \det B \\
&= \det A \det B. \ ♦
\end{aligned}$$

Determinants play important roles in analytic geometry. In three dimensions, these involve the *cross product* of two vectors. For a detailed description, see section 5.6.

Gauss–Jordan elimination

To solve a linear system $AX = B$ whose $m \times n$ matrix A isn't square, first perform the downward pass of Gauss elimination, converting it to an upper triangular system as described earlier. Instead of the upward pass, however, complete the process described next, called *Gauss–Jordan elimination*.

If $m > n$, then the last $m - n$ equations have the form $0 = r_k$ with $m < k \leq n$. If any of these $r_k \neq 0$, the system has no solution. Otherwise, you can ignore those last equations, and the system is equivalent to an $m \times m$ upper triangular system. Proceed as described earlier.

If $m < n$, use further type (II) operations to eliminate all x_k terms above diagonal entries with coefficients $a_{kk} \neq 0$. If any equation in the resulting equivalent system has the form $0 = r_k$ with $r_k \neq 0$, the original system has no solution. Otherwise, ignoring equations of the form $0 = 0$, proceed backward through the system as in the upward pass of Gauss elimination. When you encounter a zero pivot, assign the corresponding X entry a parameter representing an arbitrary value. Use a different parameter for each zero pivot.