53

General Mathematical and Physical Concepts

Edward M. Mikhail
Purdue University

53.1 Coordinate Systems

Two-Dimensional Systems

Figure 53.1 depicts two commonly used coordinate systems, one polar (r, θ) and the other *Cartesian or rectangular* (x_1, x_2). A point p can be located either by the angle θ, measured from the reference direction x_1, and range r from the reference point 0, or by its two distances from two perpendicular axes, x_1, x_2. The relationships between the two systems are given by

$$x_1 = r\cos\theta$$

$$x_2 = r\sin\theta \tag{53.1}$$

$$r = (x_1^2 + x_2^2)^{1/2}$$

$$\theta = \tan^{-1}(x_2/x_1) \tag{53.2}$$

Three-Dimensional Coordinate Systems

Figure 53.2 shows two systems of three-dimensional coordinates: *spherical* (α, β, r) and Cartesian or rectangular (x_1, x_2, x_3). The Cartesian system depicted in Fig. 53.2 is *right-handed*, since a right-threaded screw rotated by an angle less than 90° from $+x_1$ to $+x_2$ would advance in the direction of $+x_3$. The relations between these two systems are as follows:

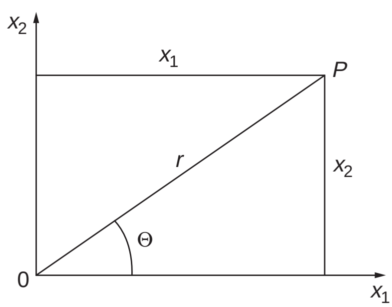

FIGURE 53.1 Two-dimensional coordinate systems.

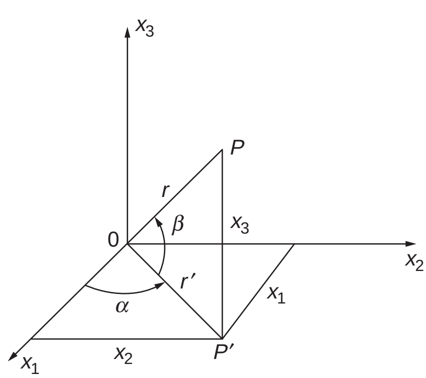

FIGURE 53.2 Three-dimensional right-handed Cartesian coordinate system.

$$r' = r\cos\beta$$

$$x_1 = r'\cos\alpha$$

$$x_2 = r'\sin\alpha$$

$$x_1 = r\cos\alpha\cos\beta$$

$$x_2 = r\sin\alpha\cos\beta \tag{53.3}$$

$$x_3 = r\sin\beta$$

$$r = (x_1^2 + x_2^2 + x_3^2)^{1/2}$$

$$\alpha = \tan^{-1}(x_2/x_1) \tag{53.4}$$

$$\beta = \sin^{-1}(x_3/r)$$

One example of spherical coordinates consists of latitude ϕ, longitude λ, and earth radius R, when the earth is considered as a sphere. A more accurate representation of the earth is an ellipsoid of revolution about its minor (shorter) axis. Coordinates referring to the earth ellipsoid, and other figures, are given in Chapter 55, "Geodesy."

Examples of rectangular systems include the geocentric and local space rectangular (LSR) coordinate systems. The geocentric system has its origin at the center of the earth, *Z* through the North Pole, *X* through the point of zero longitude, and the *Y* completing a right-handed system. The LSR varies with location, with its *XY* plane either tangent or secant to the ellipsoid and passing through a selected point λ_0, ϕ_0, and its *Z* along the local zenith.

53.2 Plane Geometry

Distance between Two Points

The distance between two points P_1 and P_2, with coordinates X_1, Y_1 and X_2, Y_2, is given by

$$L_{12} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{53.5}$$

Slope of a Line between Two Points

The slope of any line in a plane is the tangent of the angle it makes with the *X* axis. The slope angle is the counterclockwise angle from the +*X* axis to the line in a *specified direction*. If the slope angle for the line P_1P_2 is θ_{12}, which is <90°, and the slope angle for the line P_2P_1 is θ_{21}, which is >180°, then

$$\theta_{21} = \theta_{12} + 180° \tag{53.6}$$

The slope m_{12} of the line P_1P_2 is

$$m_{12} = \tan\theta_{12} = \frac{Y_2 - Y_1}{X_2 - X_1} \tag{53.7}$$

and that for the line P_2P_1 is

$$m_{21} = \tan\theta_{21} = \frac{Y_1 - Y_2}{X_1 - X_2} \tag{53.8}$$

Note that $m_{21} = \tan \theta_{21} = \tan(\theta_{12} + 180°) = \tan \theta_{12}$.

Azimuth

Azimuth is a clockwise angle with a magnitude between 0 and 360˚. It is measured either from North (or the $+Y$ axis) or from South (or the $-Y$ axis). Thus, for any line, α_N and α_S designate azimuths from North and from South, respectively. One azimuth angle is obtained from the other by simply adding 180˚ and dropping 360˚ whenever the sum exceeds 360˚. Thus

$$\begin{aligned} \alpha_S &= \alpha_N + 180° \\ \alpha_N &= \alpha_S + 180° \end{aligned} \tag{53.9}$$

Bearing

Bearing is another form of expressing the direction of a line in surveying. It is always an *acute* angle, with a magnitude between 0 and 90˚, and is a positive quantity. The bearing is the angle the line makes with either N (for North) or S (for South). The quadrant is indicated by specifying whether the angle is on the east or west side of the meridian (Y axis). Thus, the bearing angle is preceded by either N or S and succeeded by either E or W. If β designates a bearing, Table 53.1 shows how to convert bearing to azimuth. Table 53.2 shows how to convert azimuth to bearing.

TABLE 53.1 Conversion of Bearing to Azimuth

Bearing	α_N, deg	α_S, deg
$N\,\beta°\,E$	β	$180+\beta$
$S\,\beta°\,E$	$180-\beta$	$360-\beta$
$S\,\beta°\,W$	$\beta+180$	β
$N\,\beta°\,W$	$360-\beta$	$180-\beta$

Source: Tables 53.1 and 53.2 are from Anderson, J.M. and Mikhail, E.M. *Introduction to Surveying*, McGrawHill, Inc., New York, NY, 1985, p. 665. With permission.

TABLE 53.2 Conversion of Azimuth to Bearing

α_N, deg	α_S, deg	β			Bearing
0–90	180–270	α_N	or	$\alpha_S - 180°$	$N\,\beta°\,E$
90–180	270–360	$180° - \alpha_N$	or	$360° - \alpha_S$	$S\,\beta°\,E$
180–270	0–90	$\alpha_N - 180°$	or	α_S	$S\,\beta°\,W$
270–360	90–180	$360° - \alpha_N$	or	$180° - \alpha_S$	$N\,\beta°\,W$

Source: Tables 53.1 and 53.2 are from Anderson, J.M. and Mikhail, E.M. *Introduction to Surveying*, McGrawHill, Inc., New York, NY, 1985, p. 665. With permission.

Relation between Azimuth and Slope Angle

Considering the commonly used azimuth from North, α_N, the slope angle θ is obtained from

$$\theta = 90° - \alpha_N \tag{53.10}$$

Internal and External Division of a Line Segment

Let points *I* and *E* divide a line segment in the proportion s_1/s_2, where *I* is an internal and *E* is an external point. If (x_1, y_1) and (x_2, y_2) are the coordinates of the ends of the line segment, the coordinates of *I* and *E* are given by

$$\begin{aligned} X_1 &= \frac{s_1 X_2 + s_2 X_1}{s_1 + s_2} \\ Y_1 &= \frac{s_1 Y_2 + s_2 Y_1}{s_1 + s_2} \end{aligned} \tag{53.11}$$

$$\begin{aligned} X_E &= \frac{s_1 X_2 - s_2 X_1}{s_1 - s_2} \\ y_E &= \frac{s_1 Y_2 - s_2 Y_1}{s_1 - s_2} \end{aligned} \tag{53.12}$$

Equation of a Line with Known *X* and *Y* Intercepts

$$\begin{aligned} \frac{X}{X_0} + \frac{Y}{Y_0} &= 1 \\ Y_0 X + X_0 Y &= X_0 Y_0 \end{aligned} \tag{53.13}$$

where X_0 and Y_0 are the *X* and *Y* intercepts, respectively.

General Equation of a Line

$$aX + bY + c = 0 \tag{53.14}$$

Since two points define a line, only two of the three coefficients *a*, *b*, *c* are independent. This can be shown by dividing by *c*, or

$$\frac{a}{c}X + \frac{b}{c}Y + 1 = 0 \quad \text{or} \quad a'X + b'Y + 1 = 0$$

The two intercepts are obtained from Eq. (53.14) as

$$X \text{ intercept} = -\frac{c}{a} \tag{53.15}$$

$$Y \text{ intercept} = -\frac{c}{b}$$

and the slope of the line is given by

$$m = \tan\theta = -\frac{a}{b} \tag{53.16}$$

Lines Parallel to the Axes

The equation of a line parallel to the X axis is simply

$$Y = k_1 \tag{53.17}$$

where k_1 is the distance of the line from the X axis. Similarly, the equation of a line parallel to the Y axis is

$$X = k_2 \tag{53.18}$$

where k_2 is the distance between the Y axis and the line.

Equation of a Line with Given Slope and Y Intercept

$$Y = mX + k \tag{53.19}$$

where m is the slope and k is the Y intercept.

Equation of a Line with a Given Slope Passing through a Given Point

$$Y - Y_p = m(X - X_p) \tag{53.20}$$

where m is the slope and X_p, Y_p are the coordinates of the point. In terms of azimuth α, the equation becomes

$$Y - Y_p = (X - X_p)\cot\alpha \tag{53.21}$$

Equation of a Line Joining Two Points

The equation of a line passing through two points P_1 and P_2 with coordinates X_1, Y_1 and X_2, Y_2 is given by

$$\frac{Y - Y_1}{X - X_1} = \frac{Y_2 - Y_1}{X_2 - X_1} \tag{53.22}$$

Equation of a Line with Given Length and Slope of the Perpendicular from Origin

$$X \cos\theta + Y \sin\theta = p \tag{53.23}$$

where p is the length of the perpendicular from the origin to the line, and θ is the angle it makes with the X axis.

Perpendicular Distance from the Origin to a Line

From the general form of the equation of a line, Eq. (53.14), the length of the perpendicular from the origin to the line is given by

$$p = \left|\frac{c}{\sqrt{a^2 + b^2}}\right| \tag{53.24}$$

Perpendicular Distance from a Point to a Line

$$s = \left|\frac{aX_1 + bY_1 + c}{\sqrt{a^2 + b^2}}\right| \tag{53.25}$$

where X_1, Y_1 are the coordinates of the point, and the line is given by the general equation $aX + bY + c = 0$.

Equation of a Line through a Point and Parallel to Another Line

$$X \cos\theta + Y \sin\theta = p + s \tag{53.26}$$

in which $X \cos \theta + Y \sin \theta = p$ is the equation of the given line and s is the perpendicular distance between the two lines. The value of s cannot be taken as its absolute value; its proper sign must be determined. This is done by realizing that the value of the left-hand side of Eq. (53.14) will always be positive for all points on one side of the line and negative for all points on the other side. (The value is of course zero for points falling on the line.) Thus,

$$\frac{a}{\sqrt{a^2 + b^2}}X + \frac{b}{\sqrt{a^2 + b^2}}Y = -\frac{c}{\sqrt{a^2 + b^2}} + \frac{aX_1 + bY_1 + c}{\sqrt{a^2 + b^2}}$$

which when clearing fractions becomes

$$aX + bY - (aX_1 + bY_1) = 0 \tag{53.27}$$

This is the equation sought, in which a, b belong to the given line and X_1, Y_1 are the coordinates of the given point.

Equation of a Line through a Point and Perpendicular to Another Line

The given line has the general equation $aX + bY + c = 0$, and the given point P has coordinates X_p, Y_p. The slope of the given line is

$$m = -\frac{a}{b}$$

The slope of the line perpendicular to the given line is b/a. Thus,

$$Y - Y_p = \frac{b}{a}(X - X_p)$$

or

$$bX - aY - (bX_p - aY_p) = 0 \tag{53.28}$$

Angle between Two Lines

The angle γ, between two lines is given by

$$\gamma = \theta_2 - \theta_1 \tag{53.29}$$

where θ_1 and θ_2 are the slope angles of the two lines. Then

$$\tan\gamma = \tan(\theta_2 - \theta_1)$$

or

$$\tan\gamma = \frac{\tan\theta_2 - \tan\theta_1}{1 + \tan\theta_1 \tan\theta_2}$$

With the line slopes $m_1 = \tan\theta_1$ and $m_2 = \tan\theta_2$,

$$\tan\gamma = \frac{m_2 - m_1}{1 + m_1 m_2} \tag{53.30}$$

If $\tan\gamma = 0$ or $m_1 = m_2$, the two lines are parallel. On the other hand, if $\tan\gamma = \infty$ or $m_1 m_2 = -1$, the two lines are perpendicular to each other.

Point of Intersection of Two Lines

If the point of intersection of the two lines is q, its coordinates X_q, Y_q satisfy their equations. Then, to get X_q, Y_q we simultaneously solve the two equations of the lines:

$$\begin{aligned} X_q &= \frac{c_1 b_2 - c_2 b_1}{a_1 b_2 - a_2 b_1} \\ Y_q &= \frac{c_1 a_2 - c_2 a_1}{a_1 b_2 - a_2 b_1} \end{aligned} \tag{53.31}$$

Equation of a Circle

Given a circle of radius r and center coordinates X_c, Y_c, its equation is

$$(X - X_c)^2 + (Y - Y_c)^2 = r^2 \tag{53.32}$$

If the circle's center is the origin of the coordinate system, $X_c = Y_c = 0$, its equation reduces to

$$X^2 + Y^2 = r^2 \tag{53.33}$$

Equation (53.32) may be expanded to the form

$$X^2 + Y^2 + 2dX + 2eY + f = 0 \tag{53.34}$$

which represents the general form of the equation of a circle. It contains three coefficients, d, e, f, which represent three geometric elements such as the radius and the two coordinates of its center.

Intersection of a Line and a Circle

In general, a straight line intersects a circle in two points. Given the equations of a circle and a line,

$$X^2 + Y^2 + 2dX + 2eY + f = 0$$

$$Y = mX + k$$

we can eliminate Y and get a general *quadratic* equation in X,

$$AX^2 + BX + C = 0$$

Its two roots are in general given by

$$X_1 = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad \text{and} \quad X_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

for each of which a value for r is obtained from the equation of the line.

In addition to the case of two points of intersection, two other situations are possible, depending upon the quantity under the radical, $B^2 - 4AC$. If this quantity is zero, or $B^2 = 4AC$, then $X_1 = X_2$ and the line is tangent to the circle at one point. In the second situation, $B^2 < 4AC$, the quantity is negative, which means that one cannot take the square root. (This is usually referred to as the imaginary solution to the quadratic equation.) In this case, the line misses the circle and no intersection takes place.

Areas

Circle.

$$A = \pi r^2 = \frac{\pi d^2}{4} \tag{53.35}$$

where r and d are the radius and diameter of the circle, respectively.

Sector.

$$A = \frac{1}{2} ra = \frac{\pi r^2 \theta°}{360°} \tag{53.36}$$

where r is the radius, a is the arc length, and θ is the angle at the center in degrees.

Segment (Less Than a Semicircle).

$$A = \frac{1}{2} ra - \frac{1}{2} r^2 \sin\theta \tag{53.37}$$

or

$$A = \frac{\pi r^2 \theta°}{360°} - \frac{1}{2} r^2 \sin\theta \tag{53.38}$$

where r, a, and θ are as defined above. If the chord length c is given, then

$$\sin\frac{\theta}{2} = \frac{c}{2r} \tag{53.39}$$

Triangle. For a *general* triangle, the area T is given by

$$T = \frac{1}{2} bh \qquad (h \text{ is perpendicular to } b) \tag{53.40}$$

or

$$T = \frac{1}{2} bc \sin A \tag{53.41}$$

or

$$T = \sqrt{s(s-a)(s-b)(s-c)} \tag{53.42}$$

with

$$s = \frac{1}{2}(a+b+c) \tag{53.43}$$

where *a*, *b*, *c* are the lengths of the sides, *h* is the height of the triangle (which is perpendicular to the base *b*), and *A*, *B*, *C* are the interior angles opposite to the side lengths *a*, *b*, *c*, respectively. Other useful relations for a plane triangle are

$$A + B + C = 180° \tag{53.44a}$$

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c} \tag{53.44b}$$

$$a^2 = b^2 + c^2 - 2bc\cos A \tag{53.44c}$$

$$b^2 = c^2 + a^2 - 2ac\cos B \tag{53.44d}$$

$$c^2 = a^2 + b^2 - 2ab\cos C \tag{53.44e}$$

$$\tan\frac{A}{2} = \frac{1}{s-a}\sqrt{\frac{(s-a)(s-b)(s-c)}{s}} \tag{53.44f}$$

$$\cos\frac{A}{2} = \sqrt{\frac{s(s-a)}{bc}} \tag{53.44g}$$

$$\cos\frac{B}{2} = \sqrt{\frac{s(s-b)}{ca}} \tag{53.44h}$$

$$\cos\frac{C}{2} = \sqrt{\frac{s(s-c)}{ab}} \tag{53.44i}$$

$$\frac{a-b}{a+b} = \frac{\tan(A-B)/2}{\tan(A+B)/2} \tag{53.44j}$$

$$h = \frac{2}{b}\sqrt{s(s-a)(s-b)(s-c)} \tag{53.44k}$$

For an *equilateral* triangle, where sides $a = b = c$ and angles $A = B = C = 60°$, then the area *T* becomes

$$T = \frac{a^2\sqrt{3}}{4} \qquad \left(with\ h = \frac{a\sqrt{3}}{2}\right) \tag{53.45}$$

Square.

$$A = a^2 \qquad (53.46$$

where *a* is the side length.

Rectangle.

$$A = ab \tag{53.47}$$

where a and b are its width and length.

Parallelogram. Let a and b be the sides, h the altitude upon side b, C the acute angle, and A the area. Then

$$A = bh = ab\sin C \tag{53.48}$$

Trapezoid. If b_1 and b_2 are the parallel sides, and h is the altitude between them, then the area A is

$$A = \frac{1}{2}h(b_1 + b_2) \tag{53.49}$$

For an *isosceles* trapezoid, if a is the length of one of the two nonparallel sides, and C is the acute angle between a and b_2, then

$$A = \frac{1}{2}a(b_1 + b_2)\sin C \tag{53.50}$$

53.3 Three-Dimensional Geometry

Distance between Two Points

The distance between two points (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) is

$$d_{12} = [(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2]^{1/2} \tag{53.51}$$

Equation of a Plane

The general equation of a plane is given by

$$AX + BY + CZ + D = 0 \tag{53.52}$$

Only three of the four coefficients A, B, C, D are independent, since we can divide by D and get

$$\frac{A}{D}X + \frac{B}{D}Y + \frac{C}{D} + 1 = 0$$

or

$$EX + FY + GZ + 1 = 0 \tag{53.53}$$

Three noncollinear points determine a plane by writing three linear equations and solving them for E, F, G. When D in Eq. (53.52) is zero, or when no 1 is in Eq. (53.53), the plane passes through the origin. (For vector representation of a plane, see "Planes and Lines" in Section 53.4.)

Equation of a Straight Line

Since a straight line is the intersection of two planes, and since a plane is expressed by one linear equation, a straight line in three-dimensional space is represented by two linear equations. The two equations of a line passing through two points (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) are given by

$$\frac{X - X_1}{X_2 - X_1} = \frac{Y - Y_1}{Y_2 - Y_1} = \frac{Z - Z_1}{Z_2 - Z_1} \tag{53.54}$$

For vector representation of a straight line, see "Planes and Lines" in the following section.

Equation of a Sphere

If (X_c, Y_c, Z_c) represents the center of the sphere and R is its radius, its equation is given by

$$(X - X_c)^2 + (Y - Y_c)^2 + (Z - Z_c)^2 = R^2 \tag{53.55}$$

53.4 Vector Algebra

Definitions

A *vector* is an entity which has a magnitude and direction. In two- and three-dimensional spaces, it is a directed line segment from one point to another. The projections of the vector on the x_1, x_2, and x_3 axes are a_1, a_2, and a_3 and are called the vector components. It is represented by a column:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

The *length* of the vector is designated by $|\mathbf{a}|$ and is given by

$$|\mathbf{a}| = (a_1^2 + a_2^2 + a_3^2)^{1/2} \tag{53.56}$$

A vector's *direction* is given either by the angles α, β, γ it makes with the axes or by their cosines. The latter are called *direction cosines* and are given by

$$\cos\alpha = \frac{a_1}{|\mathbf{a}|} \qquad \cos\beta = \frac{a_2}{|\mathbf{a}|} \qquad \cos\gamma = \frac{a_3}{|\mathbf{a}|} \tag{53.57}$$

It is evident that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \tag{53.58}$$

Generalizing a vector to n dimensions, we write

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

Vector Operations

Equality.

$$\mathbf{a} = \mathbf{b} \qquad \text{when } a_1 = b_1, a_2 = b_2, \cdots, a_n = b_n$$

Addition/Subtraction.

$$\mathbf{c} = \mathbf{a} \pm \mathbf{b} \qquad \text{or} \qquad c_1 = a_1 \pm b_1, c_2 = a_2 \pm b_2, \cdots, c_n = a_n \pm b_n$$

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$$

Multiplication by a Scalar. A *scalar* is a quantity which has magnitude but no direction, such as mass, temperature, time, etc., and will be designated by a lowercase Greek letter.

$$\lambda\mathbf{a} = \begin{bmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{bmatrix}$$

$$\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a} = \mu(\lambda\mathbf{a}) \tag{53.59}$$

$$\begin{aligned} (\lambda + \mu)\mathbf{a} &= \lambda\mathbf{a} + \mu\mathbf{a} \\ \lambda(\mathbf{a} + \mathbf{b}) &= \lambda\mathbf{a} + \lambda\mathbf{b} \\ |\lambda\mathbf{a}| &= \lambda|\mathbf{a}| \end{aligned}$$

Any vector a is reduced to a unit vector $\mathbf{a}^\circ$ when dividing its components by its length, which is a scalar, or $\mathbf{a}^\circ = \mathbf{a}/|\mathbf{a}|$. The components of $\mathbf{a}^\circ$ are the direction cosines of **a**. Unit vectors along the coordinate axes are called base or basis vectors and are given by

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{53.60}$$

Any vector in 3-space is uniquely expressed as

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \tag{53.61}$$

The right-handed system introduced in Section 53.1 can be generalized for three vectors **a**, **b**, **c**. If they are not coplanar, and they have the same initial point, then they are said to form a right-handed system if a right-threaded screw rotated through an angle *less than 180°* from **a** to **b** would advance in the direction c.

Vector Products

Dot (or Scalar) Product.

$$\mathbf{a} \cdot \mathbf{b} = \sum_{p=1}^{n} a_p b_p = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \tag{53.62}$$

This is also called the *inner product*. It is a scalar and has the following properties:

$$
\begin{aligned}
\mathbf{a}\cdot\mathbf{b} &= \mathbf{b}\cdot\mathbf{a}\\
\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) &= \mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}\\
\lambda(\mathbf{a}\cdot\mathbf{b}) &= (\lambda\mathbf{a})\cdot\mathbf{b} = \mathbf{a}\cdot(\lambda\mathbf{b}) = (\mathbf{a}\cdot\mathbf{b})\lambda\\
\mathbf{i}\cdot\mathbf{i} &= \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1\\
\mathbf{i}\cdot\mathbf{j} &= \mathbf{j}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{i} = 0
\end{aligned} \tag{53.63}
$$

The dot product of a vector with itself is equal to the square of its length, or

$$\mathbf{a}\cdot\mathbf{a} = a_1^2 + a_1^2 + \cdots a_n^2 = |\mathbf{a}|^2 \tag{53.64}$$

If θ is the angle between two vectors **a** and **b** (in two- or three-dimensional space), it can be shown that

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta \tag{53.65}$$

It follows that if **a** is perpendicular to **b**, then $\mathbf{a}\cdot\mathbf{b} = 0$.

Cross (or Vector) Product. $\mathbf{a}\times\mathbf{b}$ (read "**a** cross **b**") is another vector **c**, which is perpendicular to both **a** and **b** and in a direction such that **a**, **b**, **c** (in this order) form a right-handed system. The length of **c** is given by

$$|\mathbf{c}| = |\mathbf{a}\times\mathbf{b}| = |\mathbf{a}||\mathbf{b}|\sin\theta \tag{53.66}$$

where θ is the angle between **a** and **b**. This quantity is the area of the parallelogram determined by **a** and **b**. If $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, then c is given by the determinant

$$\mathbf{c} = \mathbf{a}\times\mathbf{b} = \begin{bmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3}\\ b_1 & b_2 & b_3\end{bmatrix} \tag{53.67}$$

It has the following properties

$$
\begin{aligned}
\mathbf{a}\times\mathbf{b} &= -(\mathbf{b}\times\mathbf{a})\\
\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c} \qquad \text{(observing the order)}\\
\mathbf{a}\cdot(\mathbf{a}\times\mathbf{c}) &= 0\\
|\mathbf{a}\times\mathbf{b}|^2 &= |\mathbf{a}|^2|\mathbf{b}|^2-(\mathbf{a}\cdot\mathbf{b})^2\\
\mathbf{i}\times\mathbf{i} &= \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0\\
\mathbf{i}\times\mathbf{j} &= \mathbf{k};\ \mathbf{j}\times\mathbf{k} = \mathbf{i};\ \mathbf{k}\times\mathbf{i} = \mathbf{j}
\end{aligned} \tag{53.68}
$$

For two nonzero vectors, if $\mathbf{a}\times\mathbf{b} = \mathbf{0}$, then **a** and **b** are parallel.

Scalar Triple Product.

$$\mathbf{a}\times\mathbf{b}\cdot\mathbf{c} = \begin{vmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix} \tag{53.69}$$

is a scalar which is equal to the volume of the parallelepiped determined by **a**, **b**, **c.** If it is zero, then the three vectors are coplanar. It has the following properties:

$$
\begin{aligned}
\mathbf{a}\times\mathbf{b}\cdot\mathbf{c} &= \mathbf{b}\times\mathbf{c}\cdot\mathbf{a} = \mathbf{c}\times\mathbf{a}\cdot\mathbf{b}\\
\mathbf{a}\times\mathbf{b}\cdot\mathbf{c} &= \mathbf{a}\cdot\mathbf{b}\times\mathbf{c}
\end{aligned}
$$

Planes and Lines

If $\mathbf{p_0}$ is a given point in a plane, $\mathbf{n}$ is a nonzero vector normal to the plane, and $\mathbf{p}$ is any point in the plane, then the equation of the plane takes the form

$$(\mathbf{p} - \mathbf{p_0}) \cdot \mathbf{n} = 0 \qquad \text{or} \qquad \mathbf{p} \cdot \mathbf{n} - \mathbf{p_0} \cdot \mathbf{n} = 0 \tag{53.70}$$

Let $\mathbf{n} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$, $\mathbf{p_0} = X_0\mathbf{i} + Y_0\mathbf{j} + Z_0\mathbf{k}$, and $\mathbf{p} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$. Then Eq. (53.70) becomes

$$A(X - X_0) + B(Y - Y_0) + C(Z - Z_0) = 0 \tag{53.71}$$

or

$$AX + BY + CZ + D = 0$$

where D = – $(AX_0 + BY_0 + CZ_0)$. Two planes are parallel when they have a common normal vector n, and are perpendicular when their normals are, or $\mathbf{n_1} \cdot \mathbf{n_2} = 0$.

If $\mathbf{p_0}$ represents a given point on a line, $\mathbf{p}$ any other point on the line, and $\mathbf{v}$ is a given nonzero vector parallel to the line, then

$$\mathbf{p} = \mathbf{p_0} + \lambda\mathbf{v} \tag{53.72}$$

is an equation of the line. In component form, it yields three scalar equations describing the parametric form (λ is the *running parameter*):

$$\begin{aligned} X &= X_0 + \lambda v_x \\ Y &= Y_0 + \lambda v_y \\ Z &= Z_0 + \lambda v_z \end{aligned} \tag{53.73}$$

If λ is eliminated, one gets the usual two-equation form of a straight line in space; see Eq. (53.54).

53.5 Matrix Algebra

Definition

A *matrix* is a group of numbers or scalar functions collected in two-dimensional (rectangular) array. A matrix is designated by a boldface capital Roman letter. Thus, an $m \times n$ matrix can be symbolically written as

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

Types of Matrices

Square Matrix. This is a matrix in which the number of rows equals the number of columns. In this case, $\underset{m,m}{A}$ is a square matrix of order m. The *principal* (or *main*) diagonal of a square matrix is composed of all elements a_{ij} for which $i = j$.

Row Matrix.

$$\underset{1,n}{\mathbf{a}} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

Column Matrix or Vector.

$$\underset{m,1}{\mathbf{b}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Diagonal Matrix.

$$\mathbf{D} = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mm} \end{bmatrix}$$

That is, $d_{ij} = 0$ for all $i \neq j$.

Scalar Matrix.

$$\mathbf{A} = \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a \end{bmatrix}$$

$$a_{ij} = 0 \qquad \text{for all} \quad i \neq j$$
$$a_{ij} = a \qquad \text{for all} \quad i = j$$

Identity or Unit Matrix.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$a_{ij} = 0 \qquad \text{for all} \quad i \neq j$$
$$a_{ij} = 1 \qquad \text{for all} \quad i = j$$

Null Matrix. A null or zero matrix is a matrix whose elements are all zero. It is denoted by a boldface zero, **0**.

Upper Triangular Matrix.

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ 0 & u_{22} & & u_{2m} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & u_{mm} \end{bmatrix}$$

with $u_{ij} = 0$ for $i > j$.

Lower Triangular Matrix.

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mm} \end{bmatrix}$$

where $l_{ij} = 0$ for $i > j$.

Basic Matrix Operations

Two matrices **A** and **B** are equal if they are of the same dimensions and each element $a_{ij} = b_{ij}$ for all i and j. The sum of two matrices **A** and **B** is possible only if they are of equal dimensions, and the elements of the resulting matrix **C** are $c_{ij} = a_{ij} + b_{ij}$ for all i, j. The following relations apply to addition (and subtraction) of matrices:

$$\begin{aligned} \mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A} \\ \mathbf{A}+(\mathbf{B}+\mathbf{C}) &= (\mathbf{A}+\mathbf{B})+\mathbf{C} = \mathbf{A}+\mathbf{B}+\mathbf{C} \\ \mathbf{A}+(-\mathbf{A}) &= \mathbf{0} \end{aligned} \tag{53.74}$$

with **0** being the zero or null matrix, and **–A** is the matrix composed of $-a_{ij}$ as elements.

Multiplication of a matrix by a scalar α results in another $\mathbf{B} = \alpha\mathbf{A}$ whose elements are $b_{ij} = \alpha a_{ij}$ for all i and j.

The following relations hold for scalar multiplication (λ, μ are scalars):

$$\begin{aligned} \lambda(\mathbf{A}+\mathbf{B}) &= \lambda\mathbf{A}+\lambda\mathbf{B} \\ (\lambda+\mu)\mathbf{A} &= \lambda\mathbf{A}+\mu\mathbf{A} \\ \lambda(\mathbf{AB})+(\lambda\mathbf{A})\mathbf{B} &= \mathbf{A}(\lambda\mathbf{B}) \\ \lambda(\mu\mathbf{A}) &= (\lambda\mu)\mathbf{A} \end{aligned} \tag{53.75}$$

The product of two matrices is another matrix. The two matrices must be *conformable for multiplication*, i.e., the number of columns of the first matrix must equal the number of rows of the second matrix. Thus, if **A** is an $m \times q$ matrix and **B** is a $q \times n$ matrix, the product **AB**, *in that order*, is another matrix **C** with m rows (as in **A**) and n columns (as in **B**). Each element c_{ij} in **C** is obtained by multiplying each one of the q elements in the ith row in **A** by the corresponding element in the jth column in **B** and adding. Algebraically, this is written as

$$c_{ij} = a_{il}b_{1j} + a_{i2}b_{2j} + \cdots + a_{iq}b_{qj} = \sum_{k=1}^{q} a_{ik}b_{kj} \tag{53.76}$$

To illustrate matrix multiplication:

$$\underset{2,1}{\mathbf{C}} = \underset{2,3}{\mathbf{A}} \; \underset{3,1}{\mathbf{B}} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} (1\times 1)+(0\times 5)+(2\times 3) \\ (2\times 1)+(1\times 5)+(0\times 3) \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

Matrix multiplication is not commutative; that is, in general $\mathbf{FG} \neq \mathbf{GF}$, even if the dimensions of the matrices allow multiplication in both directions (e.g., $m \times n$ and $n \times m$, or square matrices). The following relationships regarding matrix multiplication hold:

$$\begin{aligned}
\mathbf{AI} &= \mathbf{IA} = \mathbf{A} && \text{in which } \mathbf{I} \text{ is the unit or identity matrix}\\
\mathbf{AB} &\neq \mathbf{BA} &&\\
\mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C} = \mathbf{ABC} && \text{(associative law)}\\
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB} + \mathbf{AC} && \text{(distributive laws)}\\
(\mathbf{A}+\mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} && \text{(distributive laws)}
\end{aligned} \tag{53.77}$$

$\mathbf{AB} = \mathbf{0}$ is possible without either **A** or **B** equaling **0**. Also, $\mathbf{AB} = \mathbf{AC}$ does not imply $\mathbf{B} = \mathbf{C}$.

The transpose of the $m \times n$ matrix **A** is an $n \times m$ matrix formed from **A** by interchanging rows and columns such that the *i*th row of **A** becomes the *i*th column of the transposed matrix. We denote the transpose of **A** by $\mathbf{A}^T$. If $\mathbf{B} = \mathbf{A}^T$, it follows that $b_{ij} = a_{ji}$ for all *i* and *j*. The following relationships apply to the transpose of a matrix:

$$\begin{aligned}
(\mathbf{A}+\mathbf{B})^T &= \mathbf{A}^T + \mathbf{B}^T &&\\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T && \text{(note reverse order)}\\
(\alpha\mathbf{A})^T &= \alpha\mathbf{A}^T &&\\
(\mathbf{A}^T)^T &= \mathbf{A} &&
\end{aligned} \tag{53.78}$$

A square matrix **A** is *symmetric* if $\mathbf{A}^T = \mathbf{A}$. Diagonal, scalar, and identity matrices are symmetric, since each is equal to its transpose. For any matrix **A** (not necessarily square), both $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric. If **B** is a symmetric matrix of suitable dimensions, then for any matrix **A**, both $\mathbf{ABA}^T$ and $\mathbf{A}^T\mathbf{BA}$ are also symmetric.

If **a** is a column matrix (or vector), then $\mathbf{a}^T\mathbf{a}$ is a positive scalar which is equal to the sum of the squares of its elements; for example, the square of the vector's length.

A square matrix **A** is *skew-symmetric* if $\mathbf{A}^T = -\mathbf{A}$ and $a_{ij} = -a_{ji}$ for all *i,j*. For any square matrix **A**, the matrix $(\mathbf{A} + \mathbf{A}^T)$ is symmetric and $(\mathbf{A} - \mathbf{A}^T)$ is skew-symmetric.

The *trace of a square* matrix is the scalar which is equal to the sum of its main diagonal elements. It is denoted by tr (**A**); thus $\text{tr}\underset{n,n}{(\mathbf{A})} = a_{11} + a_{22} + \cdots a_{nn}$. The following are properties of the trace:

$$\begin{aligned}
\text{tr}(\mathbf{A}) &= \text{tr}(\mathbf{A}^T) &&\\
\text{tr}(\lambda\mathbf{A}) &= \lambda\text{tr}(\mathbf{A}^T) &&\\
\text{tr}(\mathbf{A}+\mathbf{B}) &= \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) &&\\
\text{tr}(\mathbf{AB}) &= \text{tr}(\mathbf{BA}) &&\\
\text{tr}(\mathbf{FAF}^{-1}) &= \text{tr}(\mathbf{A}) && (\mathbf{F} \text{ nonsingular matrix})
\end{aligned} \tag{53.79}$$

Matrix Inverse

Division of matrices is not defined. Instead, the *inverse of a square* matrix **A**, if it exists, is the unique matrix $\mathbf{A}^{-1}$ with the following property:

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{53.80}$$

where **I** is the identity matrix.

The properties of the inverse are

$$\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \ \text{(note reverse order)}\\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A}\\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T\\
(\lambda\mathbf{A})^{-1} &= \frac{1}{\lambda}\mathbf{A}^{-1}
\end{aligned} \tag{53.81}$$

A square matrix which has an inverse is called *nonsingular*, whereas a matrix which does not have an inverse is called *singular*.

It was stated previously that **AB** can equal **0** without either $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$. If, however, either **A** or **B** is nonsingular, then the other matrix must be a null matrix. Hence, the product of two nonsingular matrices cannot be a null or zero matrix.

Associated with each *square matrix* **A** is a unique scalar called the *determinant* of **A**. It is denoted either by det **A** or by $|\mathbf{A}|$. Thus, for

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

the determinant is expressed as

$$|\mathbf{A}| = \begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix}$$

The determinant of order n (for an $n \times n$ square matrix) can be defined in terms of determinants of order $n - 1$ and less. The determinant of a 1×1 matrix is defined as the value of that one element, i.e., for $\mathbf{A} = [a_{11}]$, $|\mathbf{A}| = \det \mathbf{A} = a_{11}$.

If **A** is an $n \infty n$ matrix, and one row and one column of **A** are deleted, the resulting matrix is an $(n - 1) \times (\mathbf{n} - 1)$ *submatrix* of **A**. The determinant of such a submatrix is called a *minor* of **A**, and it is designated by m_{ij}, where i and j correspond to the deleted row and column, respectively. More specifically, m_{ij} is known as the *minor of the element* a_{ij} in **A**. Thus, each element of **A** has a minor.

The *cofactor* c_{ij} of an element a_{ij} is defined as

$$c_{ij} = (-1)^{i+j} m_{ij} \tag{53.82}$$

The determinant of an $n \times n$ matrix **A** can now be defined as

$$|\mathbf{A}| = a_{11}c_{11} + a_{12}c_{12} + \cdots + a_{1n}c_{1n} \tag{53.83}$$

which states that the determinant of **A** is the sum of the products of the elements of the first row of **A** and their corresponding cofactors. (It is equally possible to define $|\mathbf{A}|$ in terms of any other row or column, but for simplicity we used the first row.) On the basis of this definition, the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

has cofactors $c_{11} = |a_{22}| = a_{22}$ and $c_{12} = -|a_{21}| = -a_{21}$, and the determinant of **A** is

$$|\mathbf{A}| = a_{11}c_{11} + a_{12}c_{12} = a_{11}a_{22} - a_{12}a_{21}$$

The *cofactor matrix* **C** of a matrix **A** is the square matrix of the same order as **A** in which each element a_{ij} is replaced by its cofactor c_{ij}.

The *adjoint matrix* of **A**, denoted by adj **A**, is the transpose of its cofactor matrix, i.e.,

$$\operatorname{adj} \mathbf{A} = \mathbf{C}^{\mathrm{T}} \tag{53.84}$$

It can be shown that

$$\mathbf{A}(\operatorname{adj} \mathbf{A}) = (\operatorname{adj} \mathbf{A})\mathbf{A} = |\mathbf{A}|\mathbf{I} \tag{53.85}$$

Comparison of Eqs. (53.80) and (53.85) leads directly to a procedure for evaluating the inverse from the adjoint matrix, namely,

$$\mathbf{A}^{-1} = \frac{\text{adj }\mathbf{A}}{|\mathbf{A}|} \tag{53.86}$$

It is easy to show that for a 2 × 2 matrix, the adjoint matrix is simply

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

A square matrix is called orthogonal if its inverse is equal to its transpose, or $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$. Thus, a matrix **M** is orthogonal when

$$\mathbf{M}^{\mathrm{T}}\mathbf{M} = \mathbf{M}\mathbf{M}^{\mathrm{T}} = \mathbf{I} \tag{53.87}$$

The columns of an orthogonal matrix are mutually orthogonal vectors of unit length. Also,

$$|\mathbf{M}| = \pm 1 \tag{53.88}$$

when $|\mathbf{M}| = +1$, **M** is called "proper orthogonal"; otherwise it is termed "improper orthogonal." The product of two orthogonal matrices is also an orthogonal matrix.

Matrix Inverse by Partitioning

Let **A** be an $n \times n$ square nonsingular matrix whose inverse is to be evaluated. We *partition* **A** in the form

$$A = \begin{array}{c} \begin{matrix} s & m \end{matrix} \\ \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \end{array} \begin{matrix} s \\ m \end{matrix}$$

where $\mathbf{A}_{11}$ is $s \times s$, $\mathbf{A}_{12}$ is $s \times m$, $\mathbf{A}_{21}$ is $m \times s$, $\mathbf{A}_{22}$ is $m \times m$, and $m + s = n$. The inverse $\mathbf{A}^{-1}$ exists, and we shall denote it, in the correspondingly partitioned form, by

$$\mathbf{A}^{-1} = \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

From the basic definition of an inverse we have $\mathbf{AA}^{-1} = \mathbf{AB} = \mathbf{I}$, or in the partitioned form,

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

which, when multiplied out, leads to four matrix equations in the four $\mathbf{B}_{ij}$ submatrices as unknowns, the solution of which, when $\mathbf{A}_{11}^{-1}$ exists, is given by

$$\begin{aligned} \mathbf{B}_{11} &= \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{B}_{21} &= -\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\ \mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{aligned} \tag{53.89}$$

Alternatively, when $\mathbf{A}_{22}^{-1}$ exists, the solution is

$$\begin{aligned} \mathbf{B}_{11} &= [\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}]^{-1} \\ \mathbf{B}_{12} &= -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{B}_{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11} \\ \mathbf{B}_{22} &= \mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{12} \end{aligned} \tag{53.90}$$

If $\mathbf{A}$ is originally a symmetric matrix, then $\mathbf{A}_{21} = \mathbf{A}_{12}^{\mathrm{T}}$ correspondingly $\mathbf{B}_{21} = \mathbf{B}_{21}^{\mathrm{T}}$.

The *rank* of a matrix is the order of the largest nonzero determinant that can be formed from the elements of the matrix by appropriate deletion of rows or columns (or both). Thus, a matrix is said to be of *rank m* if and only if it has *at least one nonsingular submatrix of order m*, but has no nonsingular submatrix of order more than *m*. A nonsingular matrix of order *n* has a rank *n*. A matrix with zero rank has elements that must all be zero.

The inverse $\mathbf{A}^{-1}$ is defined only for square matrices, and it exists when the rank of $\mathbf{A}$ is equal to its order. A more general inverse may be defined for rectangular matrices with arbitrary rank. It is called the generalized inverse, denoted by $\mathbf{A}^{-}$, and satisfies the relation

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A} \tag{53.91}$$

This condition is not sufficient to define a unique $\mathbf{A}^{-}$. Additional conditions may be imposed on $\mathbf{A}^{-}$, such as

$$\begin{aligned} \mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} &= \mathbf{A}^{-} \\ (\mathbf{A}\mathbf{A}^{-})^{\mathrm{T}} &= \mathbf{A}\mathbf{A}^{-} \\ (\mathbf{A}^{-}\mathbf{A})^{\mathrm{T}} &= \mathbf{A}^{-}\mathbf{A} \end{aligned} \tag{53.92}$$

If we impose all four conditions in Eqs. (53.91) and (53.92), the inverse is called the *pseudo inverse* or the Moore–Penrose inverse, and is denoted by $\mathbf{A}^{+}$.

The Eigenvalue Problem

For a square matrix $\mathbf{A}$ of order *n*, we seek a nonzero vector $\mathbf{x}$ and a scalar λ such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{53.93}$$

which is called the "eigenvalue problem." A solution λ_0 and $\mathbf{x}_0$ to this problem is called an *eigenvalue* (proper value, characteristic value) and the corresponding *eigenvector* (proper vector, characteristic vector) of the matrix $\mathbf{A}$. An eigenvector, if one exists, can be determined only to a scalar multiplication, for if λ_0, $\mathbf{x}_0$ satisfy Eq. (53.93), then λ_0, $\alpha\mathbf{x}_0$, where α is an arbitrary scalar, will also.

Equation (53.93) can be rewritten as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0 \tag{53.94}$$

which represents a set of homogenous linear equations. For a nontrivial solution to this set the following condition must be satisfied:

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{53.95}$$

Equation (53.95) represents a real polynomial equation of degree *n*:

$$b_n(-\lambda)^n + b_{n-1}(-\lambda)^{n-1} + \cdots + b_0 = 0 \tag{53.96}$$

where

$$\begin{aligned}
b_n &= 1 \\
b_{n-1} &= a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii} = \text{tr}(\mathbf{A}) = \text{trace of } \mathbf{A} \\
&\vdots \\
b_{n-r} &= \text{sum of all principal minors of order } r \text{ of } \mathbf{A} \\
&\vdots \\
b_0 &= |\mathbf{A}| = \text{determinant of } \mathbf{A}
\end{aligned} \tag{53.97}$$

Equation (53.96) is called the *characteristic equation* of **A**, or the *eigenvalue equation*. The matrix $(\mathbf{A} - \lambda\mathbf{I})$ is called the *characteristic matrix*. There are n roots for Eq. (53.96), counting multiplicity. These are the n eigenvalues of **A**, $\lambda_1, \lambda_2, \ldots, \lambda_n$. For an eigenvalue λ_i, we solve the set of (homogeneous) linear equations $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x} = 0$ to determine the components of the corresponding eigenvector $\mathbf{x}_i$. In general, λ_i and $\mathbf{x}_i$ are either real or complex numbers and vectors, respectively.

If the matrix A is *symmetric*, then:

1. The eigenvalues are real.
2. The eigenvectors are all mutually orthogonal; that is,

$$\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j = \mathbf{x}_j^{\mathrm{T}}\mathbf{x}_i = 0$$

Bilinear and Quadratic Forms

If **A** is a square matrix of order n and **x** and **y** are two arbitrary n vectors, then the scalar

$$u = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{y} \tag{53.98}$$

is called a *bilinear form*. If, however, the matrix **A** is also *symmetric*, then

$$v = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \tag{53.99}$$

is called a *quadratic form* with the kernel **A**.

The matrix **A** is called *positive definite* if $v > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and we write $\mathbf{A} > \mathbf{0}$. If $v \geq 0$ for all **x** and there exists a nonzero vector **x** for which equality holds, we say **A** is *positive semidefinite* (or *nonnegative definite*) and write $\mathbf{A} \geq \mathbf{0}$. There are corresponding definitions for *negative definite* (or *nonpositive definite*). If there exist vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $\mathbf{x}_1^{\mathrm{T}}\mathbf{A}\mathbf{x}_1 > 0$ and $\mathbf{x}_2^{\mathrm{T}}\mathbf{A}\mathbf{x}_2 < 0$, we say **A** is *indefinite*.

For a positive definite matrix **A** it is necessary and sufficient that

$$a_{11} > 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \cdots, \quad |\mathbf{A}| > 0$$

A quadratic form represents, in general, a conic section of some kind. Considering the two-dimensional case for simplicity, we write

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = b \qquad \text{with } \mathbf{A} \text{ symmetric} \tag{53.100}$$

or

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = b$$

which is the equation of an ellipse.

53.6 Coordinate Transformations

Linear Transformations

A general *linear transformation* of a vector **x** to another vector **y** takes the form

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{t} \tag{53.101}$$

Each element of the **y** vector is a linear combination of the elements of **x** plus a translation or shift represented by an element of the **t** vector. The matrix **M** is called the *transformation matrix*, which is in general rectangular, and **t** is called the translation vector. For our use we restrict **M** to being square nonsingular; thus, the inverse relation exists, or

$$\mathbf{x} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{t}) \tag{53.102}$$

in which case it is called *affine transformation*. Although both Eqs. (53.101) and (53.102) apply to higher-dimension vectors, we will limit our discussions, without loss of generality, to the more practical two- and three-dimensional spaces, where the elements of the transformations can be depicted geometrically.

Two-Dimensional Linear Transformations

There are six *elementary* transformations, each representing a single effect, which are geometrically represented in Fig. 53.3. Initially, four vectors (1,3) (1,5), (3,3) (3,5) representing the corners of a square (solid lines in Fig. 53.3) are referred to the x_1, x_2 coordinate system. Each of the six elementary transformations operates on the square, and the resulting y_1, y_2 coordinates are plotted to show the effect on the location, orientation, size, and shape of the square after the transformation (dashed lines in Fig. 53.3). In displaying the effects of the transformations, we either display the new figure (dashed lines) in the same coordinate system, or we change the coordinate system. It is easier for the student to visualize these transformations if the new figure is drawn without changing the coordinate system, which we did in Fig. 53.3. However, as we discuss each elementary transformation, we will comment on the second interpretation when appropriate.

1. **Translation**

$$\mathbf{y} = \mathbf{x} + \mathbf{t} \qquad \text{where } \mathbf{M} = \mathbf{I} \tag{53.103}$$

The square is shifted 3 units in x_1 direction and 1 unit in x_2 direction, as shown in Fig. 53.3(a). Alternatively, the solid square remains and the coordinate axes shifted (in the opposite direction and shown in dashed lines).

2. **Uniform Scale**

$$\mathbf{y} = \mathbf{M}\mathbf{x} \qquad \mathbf{M} = \mathbf{U} = \begin{bmatrix} u & 0 \\ 0 & u \end{bmatrix} = u\mathbf{I} \tag{53.104}$$

The (dotted) square is enlarged by the uniform scale u (= 1.5 in Fig. 53.3(b)), which results from all four point coordinate pairs multiplied by u. Alternatively, the solid square is referred to the same coordinate system, except that the units along the axes are now $1/u$ of the original units.

3. **Rotation**

$$\mathbf{y} = \mathbf{M}\mathbf{x} \quad \mathbf{M} = \mathbf{R} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \tag{53.105}$$

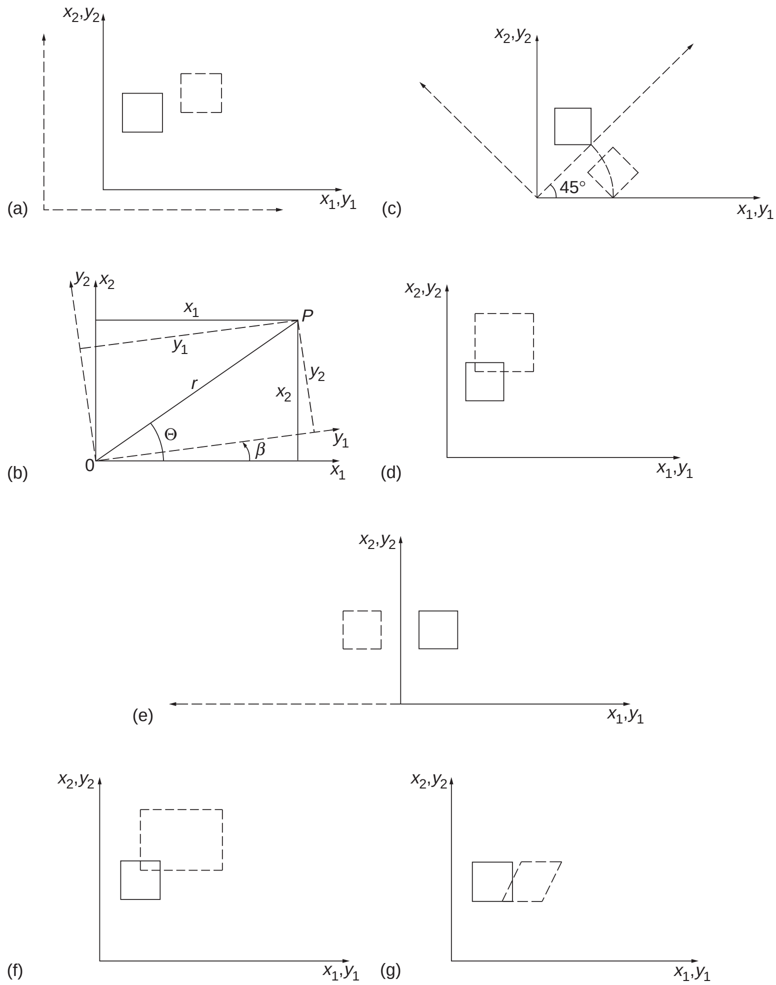

FIGURE 53.3 (a) Translation. (b) Uniform scale. (c) Rotation. (d) Rotation of a two-dimensional coordinate system. (e) Reflection. (f) Stretch (nonuniform scale). (g) Skew (nonperpendicularity of axes).

The square retains its shape, but is rotated through β about the origin of the coordinate system. In Fig. 53.3(c), the coordinate system is also rotated (45°). The elements of **R** are derived from Fig. 53.3(d) as follows:

$$\begin{aligned} y_1 &= r\cos(\theta-\beta) = r\cos\theta\cos\beta + r\sin\theta\sin\beta \\ y_2 &= r\sin(\theta-\beta) = r\sin\theta\cos\beta - r\cos\theta\sin\beta \end{aligned}$$

or

$$\begin{aligned} y_1 &= x_1\cos\beta + x_2\sin\beta \\ y_2 &= -x_1\sin\beta + x_2\cos\beta \end{aligned}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{53.106}$$

The matrix **R** is proper orthogonal, $\mathbf{R}^{-1} = \mathbf{R}^T$ and $|\mathbf{R}| = +1$. Rotation matrices do not change the length of the vector, i.e., $|\mathbf{x}| = |\mathbf{y}|$. Considering the square of the vector length,

$$\mathbf{y}^T\mathbf{y} = (\mathbf{Mx})^T\mathbf{Mx} = \mathbf{x}^T\mathbf{M}^T\mathbf{Mx}^T = \mathbf{x}^T\mathbf{x}$$

or

$$\mathbf{x}^T(\mathbf{M}^T\mathbf{M} - \mathbf{I})\mathbf{x} = \mathbf{0}$$

which for a nontrivial solution means that $\mathbf{M}^T\mathbf{M} = \mathbf{I}$, thus showing that **M** is an orthogonal matrix.

4. **Reflection**

$$\mathbf{y} = \mathbf{Mx} \qquad \mathbf{M} = \mathbf{F} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Figure 53.3(e) shows reflection of the x_1 axis (i.e., about the x_2 axis). **F** is improper orthogonal, $\mathbf{F}^{-1} = \mathbf{F}^T = \mathbf{F}$ and $|\mathbf{F}| = -1$.

5. **Stretch (Two Scale Factors)**

$$\mathbf{y} = \mathbf{Mx} \qquad \mathbf{M} = \mathbf{S} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \tag{53.107}$$

The square is transformed into a rectangle as shown in Fig. 53.3(f), in which

$$\mathbf{S} = \begin{bmatrix} 2 & 0 \\ 0 & 1.5 \end{bmatrix}$$

6. **Skew (Shear)**

$$\mathbf{y} = \mathbf{Mx} \qquad \mathbf{M} = \mathbf{K} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \tag{53.108}$$

The square is transformed into a parallelogram as shown in Fig. 53.3(g), where

$$\mathbf{K} = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$$

From these elementary transformations several affine transformations may be constructed using various sequences. The following are two of the commonly used transformations in photogrammetry.

Four-Parameter Transformation.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} u & 0 \\ 0 & u \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{53.109a}$$

or

$$\begin{aligned} y_1 &= ux_1\cos\beta + ux_2\sin\beta + t_1 \\ y_2 &= -ux_1\sin\beta + ux_2\cos\beta + t_2 \end{aligned} \tag{53.109b}$$

or

$$\begin{aligned} y_1 &= ax_1 + bx_2 + c \\ y_2 &= -bx_1 + ax_2 + d \end{aligned} \tag{53.109c}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \tag{53.109d}$$

The inverse transformation is given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{u} \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} y_1 - c \\ y_2 - d \end{bmatrix} \tag{53.109e}$$

or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{a^2 + b^2} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} y_1 - c \\ y_2 - d \end{bmatrix} \tag{53.109f}$$

This transformation has four parameters: a uniform scale, a rotation, and two translations. It is a conformal transformation.

Six-Parameter Transformation.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{53.110a}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c \\ f \end{bmatrix} \tag{53.110b}$$

The six parameters of this transformation consist of two scales, one skew factor (lack of perpendicularity of the axes), one rotation, and two shifts. The inverse transformation is given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{ae - bd} \begin{bmatrix} e & -b \\ -d & a \end{bmatrix} \begin{bmatrix} y_1 - c \\ y_2 - f \end{bmatrix} \tag{53.110c}$$

Three-Dimensional Linear Transformations

As in the two-dimensional case, affine transformation in three dimensions can be factored out in several elementary transformations: translation, uniform scale, nonuniform scale, rotations, reflections, etc. Consideration, however, is limited to the seven-parameter transformation, which is composed of a uniform scale change, three translations, and three rotations.

We first consider rotations in three-dimensional space.

Rotations of a Three-Dimensional Coordinate System. There are three elementary rotations, one about each of the three axes. They are frequently performed in sequence one after the other. A set of three of these is as follows, where $\mathbf{x}$ is the original system, $\mathbf{x}'$ is once rotated, and $\mathbf{x}''$ is twice rotated:

1. β_1 about x_1 axis, positive rotation advances $+x_2$ to $+x_3$
2. β_2 about x'_2 axis, positive rotation advances $+x'_3$ to $+x'_1$
3. β_3 about x''_3 axis, positive rotation advances $+x''_1$ to $+x''_2$

Each of the three elementary rotations is represented in matrix form by

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta_1 & \sin\beta_1 \\ 0 & -\sin\beta_1 & \cos\beta_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{M}_{\beta_1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{53.111a}$$

where x_1, x_2, x_3 are the coordinates before rotation and x'_1, x'_2, x'_3 are the coordinates after rotation. Similarly, rotations of $+\beta_2$ about the x'_2 axis and $+\beta_3$ about the x''_3 axis are given by

$$\begin{bmatrix} x''_1 \\ x''_2 \\ x''_3 \end{bmatrix} = \begin{bmatrix} \cos\beta_2 & 0 & -\sin\beta_2 \\ 0 & 1 & 0 \\ \sin\beta_2 & 0 & \cos\beta_1 \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \mathbf{M}_{\beta_2} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{53.111b}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x'''_1 \\ x'''_2 \\ x'''_3 \end{bmatrix} = \begin{bmatrix} \cos\beta_3 & \sin\beta_3 & 0 \\ -\sin\beta_3 & \cos\beta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x''_1 \\ x''_2 \\ x''_3 \end{bmatrix} = \mathbf{M}_{\beta_3} \begin{bmatrix} x''_1 \\ x''_2 \\ x''_3 \end{bmatrix} \tag{53.111c}$$

The three rotations in Eq. (53.111) are often referred to as *elementary* rotations, since they may be used to construct any required set of sequential rotations. By successive substitution, the total rotation matrix is obtained:

$$\mathbf{y} = \mathbf{x}''' = \mathbf{M}_{\beta_3}\mathbf{M}_{\beta_2}\mathbf{M}_{\beta_1}\mathbf{x} = \mathbf{M}\mathbf{x} \tag{53.112}$$

in which **M** is now a function of the three rotation angles β_1, β_2, β_3. The most commonly used set of sequential rotations (in photogrammetry) is given the symbols ω, ϕ, κ where $\omega \equiv \beta_1$; $\phi \equiv \beta_2$; $\kappa \equiv \beta_3$. In this case, the matrix **M** which rotates the object coordinate system (X, Y, Z) parallel to the photo coordinate system (x, y, z) is given by

$$\mathbf{M} = \begin{bmatrix} \cos\phi\cos\kappa & \cos\omega\sin\kappa + \sin\omega\sin\phi\cos\kappa & \sin\omega\sin\kappa - \cos\omega\sin\phi\cos\kappa \\ -\cos\phi\sin\kappa & \cos\omega\cos\kappa - \sin\omega\sin\phi\sin\kappa & \sin\omega\cos\kappa + \cos\omega\sin\phi\sin\kappa \\ \sin\phi & -\sin\omega\cos\phi & \cos\omega\cos\phi \end{bmatrix} \tag{53.113}$$

in which ω is about the X axis, ϕ is about the once-rotated Y axis, and κ is about the twice-rotated Z axis. The matrix **M** is orthogonal, since $\mathbf{M}_\omega$, $\mathbf{M}_\phi$, and $\mathbf{M}_\kappa$ are each orthogonal.

Seven-Parameter Transformation. This transformation contains seven parameters: a uniform scale change u, three rotations β_1, β_2, β_3, and three translations t_1, t_2, t_3. It takes the general form

$$\mathbf{y} = u\mathbf{M}\mathbf{x} + \mathbf{t} \tag{53.114}$$

The orthogonal matrix **M** is a function of only three independent parameters, in this case the angles β_1, β_2, β_3. This transformation is useful for different applications, such as absolute orientation, model connection, etc.

The orthogonal matrix **M** may be constructed by other methods besides sequential rotations. Two such methods follow.

Constructing M by One Rotation about a Line. This is also often referred to as the *solid body rotation*. Given a three-dimensional object in two different orientations, there exists a line in space about which the object may be rotated by a finite angle to change it from one orientation to the other. If the said line has λ, μ, ν as direction cosines and the angle of rotation is designated by α, the rotation matrix is given by

$$\mathbf{M} = \left[\begin{array}{c|c|c} \lambda^2(1-\cos\alpha)+\cos\alpha & \lambda\mu(1-\cos\alpha)-\nu\sin\alpha & \lambda\nu(1-\cos\alpha)+\mu\sin\alpha \\ \lambda\mu(1-\cos\alpha)+\nu\sin\alpha & \mu^2(1-\cos\alpha)+\cos\alpha & \mu\nu(1-\cos\alpha)-\lambda\sin\alpha \\ \lambda\nu(1-\cos\alpha)-\mu\sin\alpha & \mu\nu(1-\cos\alpha)+\lambda\sin\alpha & \nu^2(1-\cos\alpha)+\cos\alpha \end{array}\right] \tag{53.115}$$

A Purely Algebraic Derivation of M. The following skew-symmetric matrix contains only three parameters a, b, c:

$$\mathbf{S} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \tag{53.116a}$$

An orthogonal matrix **M** can be obtained from **M** using

$$\mathbf{M} = (\mathbf{I}+\mathbf{S})(\mathbf{I}-\mathbf{S})^{-1} = (\mathbf{I}-\mathbf{S})^{-1}(\mathbf{I}+\mathbf{S}) \tag{53.116b}$$

in which **M** is the identity matrix. Then

$$\begin{aligned} \mathbf{M} &= (\mathbf{I}-\mathbf{S})^{-1}(\mathbf{I}+\mathbf{S}) \\ &= \frac{1}{1+a^2+b^2+c^2}\begin{bmatrix} 1+a^2-b^2-c^2 & 2ab-2c & 2ac+2b \\ 2ab+2c & 1-a^2+b^2-c^2 & 2bc-2a \\ 2ac-2b & 2bc+2a & 1-a^2-b^2+c^2 \end{bmatrix} \end{aligned} \tag{53.117}$$

Nonlinear Transformations

In addition to the linear transformations discussed so far, we use nonlinear transformations both in two and three dimensions. In two dimensions we have the following two transformations:

Eight-Parameter Transformation. The equations

$$\begin{aligned} y_1 &= \frac{a_1x_1+b_1x_2+c_1}{a_0x_1+b_0x_2+1} \\ y_2 &= \frac{a_2x_1+b_2x_2+c_2}{a_0x_1+b_0x_2+1} \end{aligned} \tag{53.118a}$$

represent the projective transformation from the **x** to the **y** coordinate systems, with the eight transformation parameters being a_0, b_0, a_1, …, c_2. Its inverse is given by

$$\begin{aligned} x_1 &= \frac{(c_1-y_1)(b_0y_2-b_2)-(c_2-y_2)(b_0y_1-b_1)}{(a_0y_1-a_1)(b_0y_2-b_2)-(a_2y_2-a_2)(b_0y_1-b_1)} \\ x_2 &= \frac{(a_0y_1-a_1)(c_2-y_2)-(a_0y_2-a_2)(c_1-y_1)}{(a_0y_1-a_1)(b_0y_2-b_2)-(a_0y_2-a_2)(b_0y_1-b_1)} \end{aligned} \tag{53.118b}$$

These equations describe the central projectivity between two planes.

Two-Dimensional General Polynomials.

$$\begin{aligned} y_1 &= a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 + \cdots \\ y_2 &= b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2 + \cdots \end{aligned} \tag{53.119a}$$

These polynomials can obviously be extended to higher powers in x_1, x_2. A special case of these is the conformal form given in the following section.

Two-Dimensional Conformal Polynomials. The conformal property preserves the angles between intersecting lines after the transformation. If we impose the two conditions

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2} \qquad \text{and} \qquad \frac{\partial y_1}{\partial x_2} = -\frac{\partial y_2}{\partial x_1} \tag{53.119b}$$

on the general polynomials in Eq. (53.119a), we get

$$\begin{aligned} y_1 &= A_0 + A_1x_1 + A_2x_2 + A_3(x_1^2 - x_2^2) + A_4(2x_1x_2) + \cdots \\ y_2 &= B_0 - A_2x_1 + A_1x_2 - A_4(x_1^2 - x_2^2) + A_3(2x_1x_2) + \cdots \end{aligned} \tag{53.119c}$$

Note that the first three terms after the equal signs are the same as those in the four-parameter transformation given in Eq. (53.109c). Equation (53.119c) can also be derived using complex numbers by writing

$$(y_1 + y_2i) = (a_0 + b_0i) + (a_1 + b_1i)(x_1 + x_2i) + (a_3 + b_3i)(x_1 + x_2i)^2 + \cdots$$

in which $i = \sqrt{-1}$. Expanding and equating y_1 to the real part and y_2 to the imaginary part (multiplier of i) on the right-hand side leads to Eq. (53.119c).

Three-dimensional General Polynomials.

$$\begin{aligned} y_1 &= a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_2^2 + a_6x_1x_2 + a_7x_2x_3 + a_8x_1x_3 + \cdots \\ y_2 &= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1^2 + b_5x_2^2 + b_6x_1x_2 + b_7x_2x_3 + b_8x_1x_3 + \cdots \\ y_3 &= c_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_1^2 + c_5x_2^2 + c_6x_1x_2 + c_7x_2x_3 + c_8x_1x_3 + \cdots \end{aligned} \tag{53.120a}$$

We can extend these polynomials to higher order. Unlike the two-dimensional case, conformal transformation does not exist in three dimensions beyond the first-order (or linear) case given by the seven-parameter transformation, Eq. (53.114). A close approximation, which exists for only second-degree terms, is derived by imposing conditions similar to those in Eq. (53.119b) on every pair of coordinates in Eq. (53.120a). This makes the projections of the 3-space onto each of the three planes conformal. Thus, imposing the following on the general polynomials in Eq. (53.120a)

$$\begin{aligned} &\frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2} = \frac{\partial y_3}{\partial x_3} \\ &\frac{\partial y_1}{\partial x_2} = -\frac{\partial y_2}{\partial x_1}; \quad \frac{\partial y_2}{\partial x_3} = -\frac{\partial y_3}{\partial x_2}; \quad \frac{\partial y_1}{\partial x_3} = -\frac{\partial y_3}{\partial x_1} \end{aligned} \tag{53.120b}$$

leads to

$$\begin{aligned} y_1 &= A_0 + Ax_1 + Bx_2 - Cx_3 + E(x_1^2 - x_2^2 - x_3^2) + 0 + 2Gx_3x_1 + 2Fx_1x_2 + \cdots \\ y_2 &= B_0 - Bx_1 + Ax_2 + Dx_3 + F(-x_1^2 + x_2^2 - x_3^2) + 2Gx_2x_3 + 0 + 2Ex_1x_2 + \cdots \\ y_3 &= C_0 + Cx_1 - Dx_2 + Ax_3 + G(-x_1^2 - x_2^2 + x_3^2) + 2Fx_2x_3 + 2Ex_3x_1 + 0 + \cdots \end{aligned} \tag{53.120c}$$

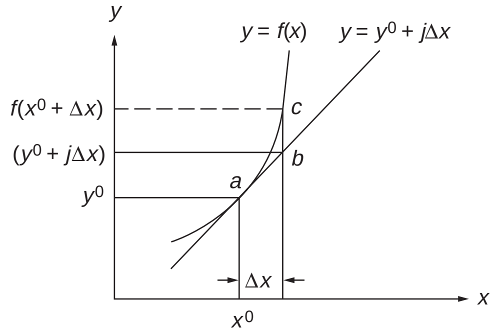

FIGURE 53.4 Linearization.

53.7 Linearization of Nonlinear Functions

Frequently, the equations expressing the geometric and physical conditions of a problem are nonlinear, which makes their direct solution difficult and uneconomical. We linearize these equations using series expansion, usually Taylor's series, which in general is given by the following for $y = f(x)$:

$$y = f(x^0) + \left.\frac{df}{dx}\right|_{x^0} \Delta x + \frac{1}{2!}\left.\frac{d^2y}{dx^2}\right|_{x^0} (\Delta x)^2 + \cdots + \frac{1}{n!}\left.\frac{d^ny}{dx^n}\right|_{x^0} (\Delta x)^n + \cdots \tag{53.121}$$

This gives the value of y at $(x^0 + \Delta x)$, given the value of the function $f(x^0)$ at x^0. Equation (53.121) includes still higher-order terms, and therefore we usually drop the second- and higher-order terms and use the approximation

$$y \approx f(x^0) + \left.\frac{dy}{dx}\right|_{x^0} \Delta x \approx y^0 + j\Delta x \tag{53.122}$$

with obvious correspondence in terms.

The technique of linearization is demonstrated in Fig. 53.4. The curve represents the original nonlinear function $f(x)$, whereas the straight line represents the linearized form, Eq. (53.122).

That line is tangent to the curve at the given point a, (x^0, y^0). When Δx is given (or evaluated), the value of the function would be approximated by point b, whose ordinate is $(y^0 + j\Delta x)$, and the exact value from the nonlinear function is point c, with ordinate $f(x^0 + \Delta x)$. The error arising from using the linear form is the line segment bc.

One Function of Two Variables

$$\begin{aligned} y &= f(x_1, x_2) \\ &= f(x_1^0, x_2^0) + \left.\frac{\partial y}{\partial x_1}\right|_{x_1^0, x_2^0} \Delta x_1 + \left.\frac{\partial y}{\partial x_2}\right|_{x_1^0, x_2^0} \Delta x_2 \\ &\quad + \frac{1}{2!}\left.\frac{\partial^2 y}{\partial x_1^2}\right|_{x_1^0, x_2^0} (\Delta x_1)^2 + \frac{1}{2!}\left.\frac{\partial^2 y}{\partial x_2^2}\right|_{x_1^0, x_2^0} (\Delta x_2)^2 \\ &\quad + \left.\frac{\partial y}{\partial x_1}\right|_{x_1^0, x_2^0} \left.\frac{\partial y}{\partial x_2}\right|_{x_1^0, x_2^0} (\Delta x_1)(\Delta x_2) + \cdots \end{aligned} \tag{53.123}$$

For the linearized form, Eq. (53.123) is truncated to

$$y = y^0 + j_1\Delta x_1 + j_2\Delta x_2 \tag{53.124}$$

where

$$y^0 = f(x_1^0, x_2^0) \qquad j_1 = \left.\frac{\partial y}{\partial x_1}\right|_{x_1^0, x_2^0} \qquad j_2 = \left.\frac{\partial y}{\partial x_2}\right|_{x_1^0, x_2^0}$$

Equation (53.124) can be rewritten in matrix form as

$$y = y^0 + [\,j_1 \quad j_2\,]\begin{bmatrix}\Delta x_1 \\ \Delta x_2\end{bmatrix}$$

or

$$y = y^0 + \mathbf{J}_{yx}\Delta\mathbf{x} \tag{53.125}$$

where

$$\mathbf{J}_{yx} = \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix}\dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2}\end{bmatrix}$$

is the Jacobian of y with respect to **x**.

Two Functions of One Variable

$$\begin{aligned} y_1 &= f_1(x) \approx y_1^0 + j_1\Delta x \\ y_2 &= f_2(x) \approx y_2^0 + j_2\Delta x \end{aligned} \tag{53.126}$$

or

$$\mathbf{y} = \mathbf{y}^0 + \mathbf{J}_{yx}\Delta x$$

with

$$\begin{aligned} y_1^0 &= f_1(x^0) \\ y_2^0 &= f_2(x^0) \\ \mathbf{J}_{yx} &= [j_1 \quad j_2]^{\mathrm{T}} = \begin{bmatrix}\left.\dfrac{dy_1}{dx}\right|_{x^0} & \left.\dfrac{dy_2}{dx}\right|_{x^0}\end{bmatrix}^{\mathrm{T}} \end{aligned}$$

Two Functions of Two Variables Each

$$\begin{aligned} y_1 &= f_1(x_1, x_2) \approx y_1^0 + j_{11}\Delta x_1 + j_{12}\Delta x_2 \\ y_2 &= f_2(x_1, x_2) \approx y_2^0 + j_{21}\Delta x_1 + j_{22}\Delta x_2 \end{aligned} \tag{53.127a}$$

or

$$\begin{bmatrix}y_1 \\ y_2\end{bmatrix} \approx \begin{bmatrix}y_1^0 \\ y_2^0\end{bmatrix} + \begin{bmatrix}j_{11} & j_{12} \\ j_{21} & j_{22}\end{bmatrix}\begin{bmatrix}\Delta x_1 \\ \Delta x_2\end{bmatrix} \tag{53.127b}$$

or

$$\mathbf{y} = \mathbf{y}^0 + \mathbf{J}_{yx}\Delta\mathbf{x} \tag{53.127c}$$

where

$$\mathbf{y}^0 = \begin{bmatrix} y_1^0 \\ y_2^0 \end{bmatrix} = \begin{bmatrix} f_1(x_1^0, x_2^0) \\ f_2(x_1^0, x_2^0) \end{bmatrix}$$

and

$$\mathbf{J}_{xy} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix}$$

evaluated at x_1^0, x_2^0.

General Case of *m* Functions of *n* Variables

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_n) \\ y_2 &= (x_1, x_2, \ldots, x_n) \\ &\vdots \\ y_m &= f_m(x_1, x_2, \ldots, x_n) \end{aligned} \tag{53.128a}$$

With the auxiliaries,

$$\mathbf{y}^0 = \begin{bmatrix} y_1^0 \\ y_2^2 \\ \vdots \\ y_m^0 \end{bmatrix} = \begin{bmatrix} f_1(x_1^0, x_2^0, \ldots, x_n^0) \\ f_2(x_1^0, x_2^0, \ldots, x_n^0) \\ \vdots \quad \vdots \\ f_m(x_1^0, x_2^0, \ldots, x_n^0) \end{bmatrix}$$

$$\mathbf{J}_{yx} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \text{ evaluated at } \mathbf{x}^0$$

$$\Delta\mathbf{x} = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{bmatrix}$$

the linearized form of Eq. (53.128a) becomes

$$\mathbf{y} \approx \mathbf{y}^0 + \mathbf{J}_{yx}\Delta\mathbf{x} \tag{53.128b}$$

which represents the general form, with $\mathbf{y}$, $\mathbf{y}^0$ being $m \times 1$ vectors, $\mathbf{J}$ an $m \times n$ Jacobian matrix, and $\Delta\mathbf{x}$ an $n \times 1$ vector. Equations (53.122), (53.125), and (53.127c) are special cases of Eq. (53.128b).

Differentiation of a Determinant

The partial derivative of a $p \times p$ determinant with respect to a scalar is composed of the sum of p determinants, each having the elements of only one row or one column replaced by their derivatives. Thus, given the determinant $d = |\mathbf{D}_1\ \mathbf{D}_2 \cdots\ \mathbf{D}_p|$ in which $\mathbf{D}_i, i = 1, 2 \cdots p$ represents its p columns, then

$$\frac{\partial d}{\partial x} = \left| \frac{\partial \mathbf{D}_1}{\partial x} \mathbf{D}_2 \cdots\ \mathbf{D}_p \right| + \left| \mathbf{D}_1 \frac{\partial \mathbf{D}_2}{\partial x} \cdots\ \mathbf{D}_\mathbf{P} \right| + \cdots + \left| \mathbf{D}_1 \mathbf{D}_2 \cdots\ \frac{\partial \mathbf{D}_p}{\partial x} \right| \tag{53.129}$$

An expression similar to Eq. (53.129) can be written in which rows instead of the columns of d are partially differentiated.

Differentiation of a Quotient

The partial derivative of $g = U/W$ with respect to a variable x is given by

$$\frac{\partial g}{\partial x} = \frac{1}{W}\left[\frac{\partial U}{\partial x} - \frac{U}{W}\frac{\partial W}{\partial x} \right] \tag{53.130}$$

Both U and W can be general functions, including determinants, of several variables.

53.8 Map Projections

Map projection is concerned with the theory and techniques of proper representation of the curved earth surface on the plane of a map. When the map is of such large scale as to represent a very limited area, the earth curvature is insignificant, and field survey measurements can be directly represented on the map. On the other hand, as the surface area of the earth gets larger, this curvature becomes significant and must be dealt with. The earth is an ellipsoid and is also sometimes approximated by a sphere; neither of these surfaces can accurately be developed into a plane. Therefore, all map projection methods must by necessity contain some distortion. Various methods are selected to fit best the shape of the area to be mapped and to minimize the effects of particular distortions.

Locations on the earth are represented by meridians of longitude, λ, and parallels of latitude, ϕ. On the map these are represented by scaled linear distances X, Y, using the dimensions of the earth ellipsoid and selected criteria which the specific map projection must satisfy. These are obtained from transformation equations taking the general functional form of

$$\begin{aligned} X &= f_x(\lambda, \phi) \\ Y &= f_y(\lambda, \phi) \end{aligned} \tag{53.131}$$

Although all modern map projections are performed by computer programs, several are based on geometric projection of the earth onto one of three surfaces: a plane, a cylinder, or a cone. It is clear that the cylinder and cone are chosen because they can be developed into a plane — that of the map. When a plane is used, it is tangent to the earth's surface at a point and the projection center is either the center of the earth, as in the *gnomonic* projection shown in Fig. 53.5(a), or the point diametrically opposite to the tangent point, as in the *stereographic* projection shown in Fig. 53.5(b). If the projection lines are perpendicular to the plane, we have an *orthographic* projection, Fig. 53.5(c).

A cone is usually selected with its axis coincident with the earth's polar axis. It may be tangent to the earth at one small circle, called *standard parallel*, or intersect it in two standard parallels. When developed, the scale will be true (i.e., without any distortion or error) at the standard parallels; see Fig. 53.6. Polyconic projections use a series of frustums of cones, each from a separate cone.

Like a cone, a cylinder may be selected to be tangent to the earth or secant to it. It may be *regular*, with its axis being the polar axis as in Fig. 53.7(a); *transverse*, with one tangent meridian as in Fig. 53.7(b); *secant*, with two meridians of intersection; or *oblique* cylindrical, shown in Fig. 53.7(c).

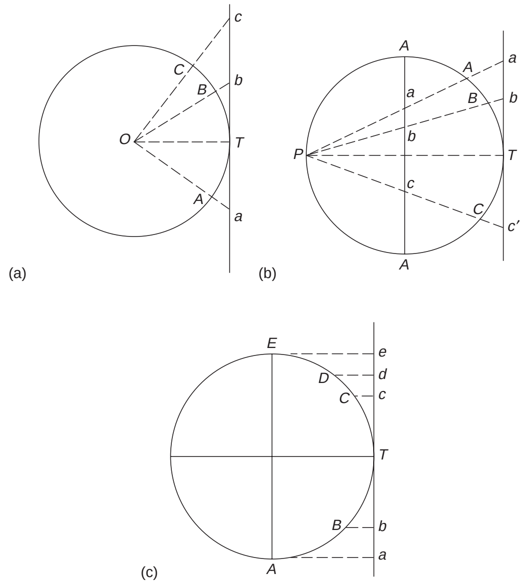

FIGURE 53.5 (a) Gnomonic projection. (b) Stereographic projection. (c) Orthographic projection.

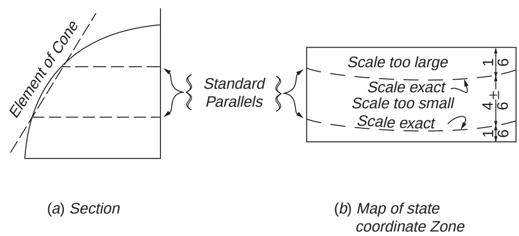

FIGURE 53.6 Lambert conformal conic projection. (Source: Davis, R. E., Foote, F. S., Anderson, J. M., and Mikhail, E. M. 1981. *Surveying: Theory and Practice,* 6th ed., p. 570. McGraw-Hill, New York. With permission.)

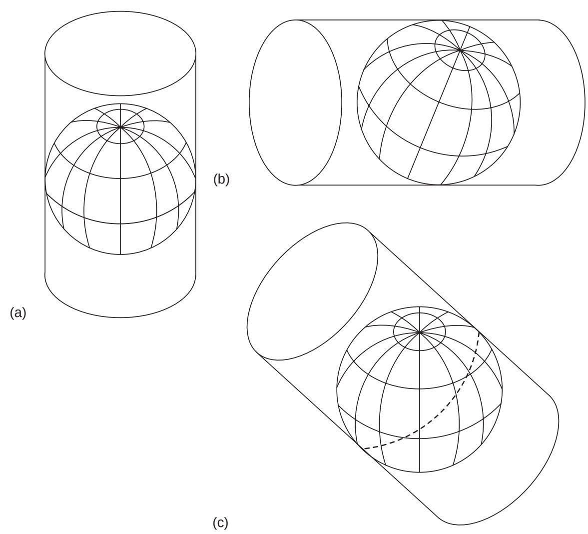

FIGURE 53.7 (a) Vertical cylinder. (b) Transverse horizontal cylinder. (c) Oblique cylinder.

Equation (53.131) will produce a *perfect* map — without any distortions — if it satisfies *all* of the following conditions:

1. All distances and areas have correct relative magnitudes.
2. All azimuths and angles are correctly represented on the map.
3. All great circles are shown on the map as straight lines.
4. Geodetic longitudes and latitudes are correctly shown on the map.

No one map projection can satisfy *all* these conditions. However, each class can satisfy some selected conditions. The following are four classes:

1. **Conformal** or **orthomorphic projection** results in a map showing the correct angle between any pair of short intersecting lines, thus making small areas appear in correct *shape*. As the scale varies from point to point, the shapes of larger areas are incorrect.
2. An **equal-area projection** results in a map showing all areas in proper relative *size*, although these areas may be much out of shape and the map may have other defects.
3. In an **equidistant projection** distances are correctly represented from one central point to other points on the map.
4. In an **azimuthal projection** the map shows the correct *direction* or azimuth of any point relative to one central point.

For conformal mapping, a new latitude, ψ, called the isometric latitude, is used in place of ϕ, where

$$\psi = \ln\left[\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2}\tan\left(\frac{\pi}{4} + \frac{\phi}{2}\right)\right] \tag{53.132}$$

in which $e^2 = (a^2 - b^2)/a^2$, with a, b being the semimajor and semiminor axes of the earth ellipsoid, respectively. Then, Eq. (53.131) is replaced by

$$\begin{aligned} X &= f_1(\lambda, \psi) \\ Y &= f_2(\lambda, \psi) \end{aligned} \tag{53.133}$$

In order for the mapping in Eq. (53.133) to be conformal, the following Cauchy–Riemann conditions must be satisfied:

$$\frac{\partial X}{\partial \lambda} = \frac{\partial Y}{\partial \psi} \quad \text{and} \quad \frac{\partial X}{\partial \psi} = -\frac{\partial Y}{\partial \lambda} \tag{53.134}$$

Two commonly used conformal projections are the Lambert conformal conic projection and the transverse Mercator projection. A figure of the former is shown in Fig. 53.6, where the projection cone intersects the ellipsoid in two standard parallels. It is very widely used in the U.S., particularly as a State Plane Coordinate System for those states, or zones thereof, with greater east–west extent than north–south. The transverse Mercator projection is shown in Fig. 53.8, where the cylinder is either tangent or secant to the ellipsoid. When it is tangent, the scale at the central meridian is 1:1. But when it is not, the scale at the central meridian is less than 1:1, as shown in Fig. 53.8. The central meridian is the origin of the map X coordinate, while the origin of the map Y coordinate is the equator. This projection is used as a State Plane Coordinate System for states with greater north–south extent.

An extensively used map projection system is the *Universal* Transverse Mercator, or UTM, schematically shown in Fig. 53.9. It is in 6° wide zones, with the scale at each central meridian of a zone being 0.9996. A false easting for each central meridian is 500,000 m. A transverse Mercator projection with 3° wide zones is possible, where the scale at the central meridian is improved to 0.9999.

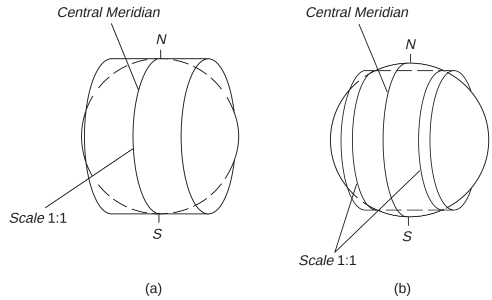

FIGURE 53.8 Transverse Mercator projection. (a) Cylinder with one standard line. (b) Cylinder with two standard lines.

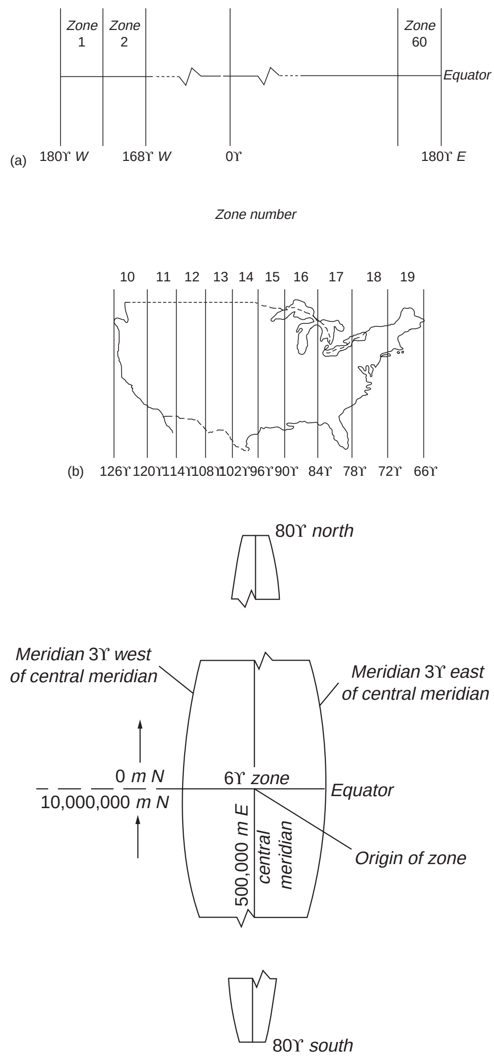

FIGURE 53.9 (a) Universal Transverse Mercator zones. (b) UTM zones in the United States. (c) *X* and *Y* coordinates of the origin of a UTM grid zone. (*Source:* U.S. Army Field Manual, *Map Reading*, FM 21-26.)

53.9 Observational Data Adjustment

Mathematical Model for Adjustment

In surveying engineering, measurements are rarely used directly as the required information. They are frequently used in subsequent operations to derive other quantities, often computationally, such as directions, lengths, relative positions, areas, shapes, and volumes. The relationships applied in the computational effort are the mathematical representations of the geometric and physical conditions of the problem, which, together with the quality of the measurements, are called the *mathematical model*.

This mathematical model is composed of two parts: a functional model and a stochastic model. The *functional model* is the part which describes the geometric or physical characteristics of the survey problem and the resulting mathematical relationships. The *stochastic model* is the part of the mathematical model that describes the statistical properties of all the elements involved in the functional model. It designates which parts are the observables, which are constants, and which are unknown parameters to be estimated in the **adjustment**. It also provides the information necessary to properly describe the quality of the observations to be used in the adjustment.

As a simple example, consider the size and shape of a plane triangle. While the shape depends only on angles, its size requires at least one side. Therefore, this model has three angular elements, the interior angles, and three linear (or distance) elements, the triangle sides. Two angles and one side will be the minimum number of measurements necessary to uniquely fix the triangle. If more measurements than these three are obtained, *redundancy* will exist, thus leading to inconsistency, which is resolved through an adjustment technique. Once the number of measurements is decided upon, the required set of independent condition equations can be written to express the functional model. The stochastic model will denote those elements (of the total of six) which are observed, and the quality of the observations.

The *a priori* quality of an observed angle or distance is usually expressed by a *standard deviation*, σ, or its square, the *variance*, σ^2. Correlation between observations is represented by the *covariance*. Thus, for two observables, or *random variables*, say $\bar{x}$ and $\bar{y}$, the variances, σ_x^2 and σ_y^2, and the covariance, σ_{xy}, are collected in a single square symmetric matrix called the variance-covariance matrix, or simply the covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \tag{53.135}$$

where the variances are along the main diagonal and the covariance off the diagonal. The concept of the covariance matrix can be extended to the multidimensional case by considering n random variables $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ and writing

$$\Sigma_{xx} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix} \tag{53.136}$$

which is an $n \times n$ square symmetric matrix.

Often in practice, the variances and covariances are not known in absolute terms but only to a scale factor. The scale factor is given the symbol σ_0^2 and is termed the *reference variance*, although other names, such as "variance factor" and "variance associated with weight unity," have also been used. The square root σ_0, of σ_0^2, is called the "reference standard deviation" and was classically known as the "standard error of unit weight." The relative variances and covariances are called *cofactors* and are given by

$$\mathbf{q}_{ii} = \frac{\sigma_i^2}{\sigma_0^2} \qquad \text{and} \qquad \mathbf{q}_{ij} = \frac{\sigma_{ij}}{\sigma_0^2} \tag{53.137}$$

Collecting the cofactors in a square symmetric matrix produces the *cofactor matrix* **Q**, with the obvious relationship with the covariance matrix.

$$\mathbf{Q} = \frac{1}{\sigma_0^2}\Sigma \tag{53.138}$$

When **Q** is nonsingular, its inverse is called the *weight matrix* and designated by **W**; thus,

$$\mathbf{W} = \mathbf{Q}^{-1} = \sigma_0^2\Sigma^{-1} \tag{53.139}$$

If σ_0^2 is equal to 1, or, in other words, if the covariance matrix is known, the weight matrix becomes its inverse.

Design/Preanalysis

Engineering *design* is more frequently known as *preanalysis* in surveying engineering. It refers to the task of determining the observations to be made and their required accuracy so that the required accuracy of the final product is met. This is usually done by an iterative procedure, often using interactive graphics, and applying the established mathematical model of the problem. The physical limitations of the project, such as visibility and accessibility problems, are first imposed on the design. Next, what is considered to be an adequate set of measurements, with suitable accuracy estimates, is input in a design program to estimate the required unknown parameters and their expected accuracy. The overall accuracy of the estimated parameters is reflected by the posterior covariance matrix Σ. (This is why preanalysis is equivalent to *covariance analysis* used in the mathematical literature.) From Σ, individual confidence measures, such as **error ellipses** and **ellipsoids**, are computed and compared to the design requirements. If they are too large, additional observations or measurements of increased quality are attempted and the process repeated. On the other hand, if they are too small (i.e., too good), reduced observations or observations of decreased quality (using less expensive equipment/techniques) are attempted instead. The procedure is iterated until an optimum design results.

Data Acquisition

The results of preanalysis provide the information required to set up the specifications for data acquisition, particularly the quantities to be measured and their required accuracy. This leads to deciding on equipment to be used and observational techniques to be applied. It is important that the selected instruments be properly calibrated and in good working order, and that the procedures specified be rigorously followed. Good field practices must be followed to minimize blunders. In fact, all field activities are carefully planned and monitored so that the following operation, preprocessing of data, can be effectively carried out to yield the most suitable data entering the adjustment.

Data Preprocessing

Preprocessing of survey data involves the elimination of *blunders* and the correction for all known systematic errors. The resulting preprocessed measurements should have essentially nothing but random errors before they are used in the adjustment. Any uncorrected systematic errors known to still exist in the measurements must then be modeled mathematically and accounted for during the adjustment.

Because of the significant cost of the field operations, it is becoming more of a requirement to perform preprocessing in the field so as to minimize the cost of any remeasurement. Progressively more sophisticated preprocessing programs are available for use with portable computers in conjunction with electronic data collectors.

There are several techniques for checking the observations which apply to the different types of surveys, such as triangulation, trilateration, traverse, leveling, etc. These techniques depend on the shape of the network and various minimum combinations of observations.

Data Adjustment

When redundant measurements exist, an *adjustment* of the observed data becomes necessary in order to resolve the inconsistency between the observations and the model. As an illustration, consider the size and shape of a plane triangle in which the three interior angles A, B, C and two sides a, b (opposite to A and B, respectively) are measured. Suppose that we are interested in the length of the side c. We obviously have more information than needed, since any one side, a or b, and any two angles suffice to solve the triangle and hence determine c. However, due to random measurement errors, every combination selected would be expected to lead to a slightly different value for c. A choice from all possible combinations would be essentially arbitrary. More important is the fact that one should take advantage of using *all* the available information in one operation, in which each measurement contributes relative to its role in the mathematical model and commensurate with its quality (variance) relative to the quality of all other measurements. One of the most commonly used techniques of adjustment of redundant survey data is the method of **least squares**.

Least Squares Adjustment

Although least squares is an estimation procedure, it has been traditionally referred to in surveying as an *adjustment* technique. This stems from the fact that after the adjustment the original observations are replaced by a set of *adjusted* observations that *are* consistent with the model. Thus, after least squares adjustment the five observations in the triangle example are replaced by a consistent set, $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{a}$, $\hat{b}$, in that *any* minimum combination of these would yield the *same* triangle solution. Each adjusted observation is the sum of the original measurement and a *residual*, v, which is calculated in the adjustment.

The method of least squares is based on the observational residuals. If all the residuals in a given set of n observations, **1** are denoted by the vector $\boldsymbol{v}$, and the weight matrix associated with these observations is **W**, the least squares criterion is given by

$$\phi = \boldsymbol{v}^{\mathrm{T}}\mathbf{W}\boldsymbol{v} \rightarrow \text{minimum} \qquad (53.140)$$

Note that ϕ is a scalar, for which a minimum is obtained by equating to zero its partial derivatives with respect to v. In Eq. (53.140) the weight matrix of the observations, **W**, may be full, implying that the observations are correlated. If the observations are uncorrelated, **W** will be a diagonal matrix, and the criterion simplifies to

$$\phi = \sum_{i=1}^{n} w_i v_i^2 = w_1 v_1^2 + w_2 v_2^2 + \cdots w_n v_n^2 \rightarrow \text{minimum} \qquad (53.141)$$

which says that the sum of the weighted squares of the residuals is a minimum. Another and simpler case involves observations which are uncorrelated and of equal weight (precision), for which $\mathbf{W} = \mathbf{I}$, and ϕ becomes

$$\phi = \sum_{i=1}^{n} v_1^2 = v_i^2 + v_2^2 + \ldots v_n^2 \rightarrow \text{minimum} \qquad (53.142)$$

The case covered by Eq. (53.142) is the oldest and may have accounted for the name "least squares," since it seeks the "least" sum of the squares of the residuals.

If we denote by n_0 the minimum number of independent variables needed to determine the selected model uniquely, then the *redundancy*, r, is given by

$$r = n - n_0 \tag{53.143}$$

As illustrations, consider the following examples.

1. The shape of a plane triangle is uniquely determined by a minimum of two interior angles, or $n_0 = 2$. If three interior angles are measured, then, with $n = 3$, redundancy is $r = 1$.
2. The size and shape of a plane triangle require a minimum of three observations, at least one of which is the length of one side, or $n_0 = 3$. If three interior angles and two side lengths are available, then with $n = 5$ the redundancy is $r = 2$.

After the redundancy r is determined, the adjustment proceeds by writing equations that relate the model variables in order to reflect the existing redundancy. Such equations will be referred to either as *condition equations* or simply as *conditions*. The number of independent conditions, c, will be equal to r if only observational variables and constants are involved. In many situations, however, additional unknown variables, called *parameters*, are carried in the adjustment. In such a case, if the number of unknown parameters is u, then a total of

$$c = r + u \tag{53.144}$$

independent condition equations in terms of both the n observations and u parameters must be written. In order for the parameter to be functionally independent, number, u, should not exceed the minimum number of variables, n_0, necessary to specify the model. Hence, the following relation must be satisfied:

$$0 \le u \le n_0 \tag{53.145}$$

Similarly, for the formulated condition equations to be independent, their number, c, should not be larger than the total number of observations n. Hence,

$$r \le c \le n \tag{53.146}$$

Techniques of Least Squares

Although there is only one least squares criterion, there are several techniques by which least squares may be applied. Regardless of which technique is applied, the results of an adjustment of a given set of measurements associated with a specified model *must* be the same. The choice of a technique, therefore, is mostly a matter of convenience and computational economy. The first technique is called *adjustment of observations only*. The condition equations take the form

$$\underset{r,n}{\mathbf{A}}\ \underset{n,1}{\mathbf{v}} = \underset{r,1}{\mathbf{f}} \tag{53.147}$$

For linear adjustment problems the vector $\mathbf{f}$ is given by

$$\mathbf{f} = \mathbf{d} = \mathbf{A1} \tag{53.148}$$

in which $\mathbf{d}$ is a vector of numerical constants and $\mathbf{1}$ is the vector of given numerical values of the measurements.

Let the cofactor matrix of the observations be denoted by $\mathbf{Q}(= \mathbf{W}^{-1})$. Then

$$\mathbf{k} = (\mathbf{AQA}^{\mathrm{T}})^{-1}\mathbf{f} = \mathbf{Q}_e^{-1}\mathbf{f} = \mathbf{W}_e\mathbf{f} \tag{53.149}$$

$$\mathbf{v} = \mathbf{QA}^{\mathrm{T}}\mathbf{k} \tag{53.150}$$

$$\hat{\mathbf{l}} = \mathbf{l} + \mathbf{v} \tag{53.151}$$

Error propagation:

$$\mathbf{Q}_{vv} = \mathbf{QA}^{\mathrm{T}}\mathbf{W}_e\mathbf{AQ} \tag{53.152}$$

$$\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{Q} - \mathbf{Q}_{vv} \tag{53.153}$$

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^{\mathrm{T}}\mathbf{Wv}}{r} \tag{53.154}$$

The second technique is called *adjustment of indirect observations*. The condition equations are of the general (nonlinear) form:

$$\mathbf{l} + \mathbf{v} + \mathbf{F}(\mathbf{x}) = 0 \tag{53.155}$$

When linearized at approximations x^0 for the $u = n_0$ parameters x, they become

$$\mathbf{v} + \mathbf{B}\boldsymbol{\Delta} = \mathbf{f} \tag{53.156}$$

where Δ is a vector of unknown parameter corrections, $\mathbf{B} = \left.\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right|_{x^0}$ is an n by u coefficient matrix, and $\mathbf{f} = -\mathbf{F}(\mathbf{x}^0) - 1$. With $\mathbf{W} = \mathbf{Q}^{-1}(= \Sigma^{-1}$ if $\sigma_0 = 1)$ as the weight matrix of the observations, then

$$\mathbf{N} = \mathbf{B}^{\mathrm{T}}\mathbf{WB} \tag{53.157}$$

$$\mathbf{t} = \mathbf{B}^{\mathrm{T}}\mathbf{Wf} \tag{53.158}$$

$$\Delta = \mathbf{N}^{-1}\mathbf{t} \tag{53.159}$$

$$\hat{\mathbf{x}} = \mathbf{x}^0 + \boldsymbol{\Sigma\Delta} \qquad \text{(iterate)} \tag{53.160}$$

Error propagation:

$$\mathbf{Q}_{\hat{x}\hat{x}} = \mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1} \tag{53.161}$$

$$\mathbf{Q}_{vv} = \mathbf{BN}^{-1}\mathbf{B}^{\mathrm{T}} \tag{53.162}$$

$$\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{Q} - \mathbf{Q}_{vv} \tag{53.163}$$

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^{\mathrm{T}}\mathbf{Wv}}{r} \tag{53.164}$$

In the preceding technique, the largest number allowed for the parameters, $u = n_0$, must be carried in the adjustment so that each condition equation contains one and only one observation. In many applications, it is more economical to carry in the adjustment only the parameters of interest, which are fewer in number: $u < n_o$. This technique is called *combined adjustment of observations and parameters*. The general (nonlinear) condition equations are of the form

$$\mathbf{F}(\mathbf{1}, \mathbf{x}) = \mathbf{0} \tag{53.165}$$

which in linearized form becomes

$$\mathbf{Av} + \mathbf{B\Delta} = \mathbf{f} \tag{53.166}$$

where

$$\mathbf{A} = \left.\frac{\partial \mathbf{F}}{\partial \mathbf{l}}\right|_{l^0, x^0} \qquad \mathbf{B} = \left.\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right|_{l^0, x^0}$$

and

$$\mathbf{f} = -[\mathbf{F}(\mathbf{l}^0, \mathbf{x}^0) + \mathbf{A}(\mathbf{l} - \mathbf{l}^0)] \tag{53.167}$$

Then

$$\mathbf{Q}_e = \mathbf{AQA}^{\mathrm{T}} \tag{53.168}$$

$$\mathbf{W}_e = \mathbf{Q}_e^{-1} = (\mathbf{AQA}^{\mathrm{T}})^{-1} \tag{53.169}$$

$$\mathbf{N} = \mathbf{B}^{\mathrm{T}}\mathbf{W}_e\mathbf{B} \tag{53.170}$$

$$\mathbf{t} = \mathbf{B}^{\mathrm{T}}\mathbf{W}_e\mathbf{f} \tag{53.171}$$

$$\Delta = \mathbf{N}^{-1}\mathbf{t} \tag{53.172}$$

$$\hat{\mathbf{x}} = \mathbf{x}^0 + \Sigma\Delta \qquad \text{(iterate)} \tag{53.173}$$

$$\mathbf{v} = \mathbf{QA}^{\mathrm{T}}\mathbf{W}_e(\mathbf{f} - \mathbf{B\Delta}) \tag{53.174}$$

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^{\mathrm{T}}\mathbf{Wv}}{r} \tag{53.175}$$

Error propagation:

$$\mathbf{Q}_{\hat{x}\hat{x}} = \mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1} \tag{53.176}$$

$$\mathbf{Q}_{w} = \mathbf{QA}^{T}(\mathbf{W}_e - \mathbf{W}_e\mathbf{BQ}_{\Delta\Delta}\mathbf{B}^{\mathrm{T}}\mathbf{W}_e)\mathbf{AQ} \tag{53.177}$$

$$\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{Q} - \mathbf{Q}_{vv} \tag{53.178}$$

The linear set $\mathbf{N\Delta} = \mathbf{t}$ is called the *normal equations*.

In the three techniques above, the parameters $\mathbf{x}$ are functionally independent. In many surveying engineering applications, functions may exist between the parameters in the adjustment. Examples include points on geometric forms (lines, planes, surfaces), known distances, angles, etc., to be fixed in the adjustment. These functions are called *constraint equations* or simply *constraints*, to distinguish them from conditions. They are characterized by not containing any observations. The technique when constraints exist is called *adjustment with functional constraints*; of course, the condition equations to be combined with the constraints can take any of the three situations discussed. It is more general to consider the combined adjustment technique, thus:

$$\mathbf{F}(\mathbf{1}, \mathbf{x}) = \mathbf{0} \tag{53.179}$$

$$\mathbf{G}(\mathbf{x}) = \mathbf{0} \tag{53.180}$$

are the general (nonlinear) conditions and constraints. The linearized form is

$$\mathbf{Av} + \mathbf{B}\boldsymbol{\Delta} = \mathbf{f} \tag{53.181}$$

$$C\boldsymbol{\Delta} = g \tag{53.182}$$

where $\mathbf{C} = \left.\frac{\partial \mathbf{G}}{\partial \mathbf{x}}\right|_{\mathbf{x}^0}$ and $\mathbf{g} = -\mathbf{G}(\mathbf{x}^0)$

$$\mathbf{N} = \mathbf{B}^{\mathrm{T}}\mathbf{W}_e\mathbf{B} \qquad \mathbf{t} = \mathbf{B}^{\mathrm{T}}\mathbf{W}_e\mathbf{f} \tag{53.183}$$

$$\mathbf{M} = \mathbf{C}\mathbf{N}^{-1}\mathbf{C}^{\mathrm{T}} \tag{53.184}$$

$$\mathbf{k}_c = \mathbf{M}^{-1}(\mathbf{g} - \mathbf{C}\mathbf{N}^{-1}\mathbf{t}) \tag{53.185}$$

$$\boldsymbol{\Delta} = \mathbf{N}^{-1}(\mathbf{t} + \mathbf{C}^{\mathrm{T}}\mathbf{k}_c) \tag{53.186}$$

$$\mathbf{x} = \mathbf{x}^0 + \boldsymbol{\Sigma\Delta} \qquad \text{(iterate)} \tag{53.187}$$

$$\mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1}(\mathbf{I} - \mathbf{C}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{C}\mathbf{N}^{-1}) \tag{53.188}$$

The quantities v, $\hat{\sigma}_0^2$, $\mathbf{Q}_{vv}$, and $\mathbf{Q}_{ll}$ are as given by Eqs. (53.174), (53.175), (53.177), and (53.178), respectively.

Another technique is to perform the adjustment sequentially. For a fixed number of parameters, both the inverse of the normal equations coefficient matrix **N** and the constant terms vector **t** are sequentially updated due to either addition or deletion of a set of condition equations. The relations for *sequential adjustment* are as follows:

$$\mathbf{A}_i\mathbf{v}_i + \mathbf{B}_i\Delta = \mathbf{f}_i \quad \text{with} \quad \mathbf{Q}_i \tag{53.189}$$

are the conditions to be added or subtracted. Let $\mathbf{N}_{I-1}$, $\mathbf{t}_{I-1}$ and $\mathbf{N}_I$, $\mathbf{t}_I$ designate the matrices *before* and *after*, respectively, incorporating the effects of Eq. (53.189). Then

$$\mathbf{N}_I^{-1} = \mathbf{N}_{I-1}^{-1}[\mathbf{I} \mp \mathbf{B}_i^{\mathrm{T}}(\mathbf{Q}_{e_i} \pm \mathbf{B}_i\mathbf{N}_{I-1}^{-1}\mathbf{B}_i^{\mathrm{T}})^{-1}\mathbf{B}_i\mathbf{N}_{I-1}^{-1}] \tag{53.190}$$

$$\mathbf{t}_I = \mathbf{t}_{I-1} \pm \mathbf{B}_i^{\mathrm{T}}\mathbf{W}_{e_i}\mathbf{f}_i \tag{53.191}$$

in which $\mathbf{Q}_{e_i} = \mathbf{A}_i\mathbf{Q}_i\mathbf{A}_i^{\mathrm{T}}$ and $\mathbf{W}_{e_i} = \mathbf{Q}_{e_i}^{-1}$. In Eqs. (53.190) and (53.191) the *upper signs* refer to condition *addition* and the *lower signs* to condition *deletion*.

Finally, a very flexible technique is used in which all the variables in the *mathematical model are considered as observables*. This requires *a priori* estimates for all model variables, but more importantly, estimates of their *a priori* weights. The *a priori* weights provide the mechanism for effectively distinguishing between different groups of variables in the model. Very large weights leave a variable essentially as a constant in the adjustment, while very small or even zero weight leaves it as an unknown parameter that can freely adjust. This is referred to as the *unified least squares* technique and applies to *all* of the techniques presented above. As an example, we consider the combined adjustment of observations and parameters with the linear (or linearized) conditions: $\mathbf{Av} + \mathbf{B}\Delta = \mathbf{f}$. Let $\mathbf{x}$ be the prior estimates of the parameters and $\mathbf{W}_{xx}$ its corresponding prior weight matrix. The solution given by Eqs. (53.168) to (53.178) essentially apply, except that

$$\Delta = (\mathbf{N} + \mathbf{W}_{xx})^{-1}(\mathbf{t} - \mathbf{W}_{xx}\mathbf{f}_x) \tag{53.192}$$

$$\mathbf{f}_x = \mathbf{x}^0 - \mathbf{x} \tag{53.193}$$

$$\mathbf{Q}_{\Delta\Delta} = (\mathbf{N} + \mathbf{W}_{xx})^{-1} \tag{53.194}$$

Assessment of Adjustment Results

After least squares adjustment it is quite important in surveying to analyze the results and provide a statement regarding the quality of the estimates. This operation is often referred to as *postadjustment analysis*, which applies various statistical techniques.

Test on Reference Variance

The first test is on the estimated reference variance, $\hat{\sigma}_0^2$. Let the *a priori* reference variance be σ_0^2; let r be the degrees of freedom (redundancy) in the adjustment, and assume that the residuals v_i are normally distributed. The statistic $r\sigma_0^2/\sigma_0$ has a χ^2 distribution with r degrees of freedom. The two-tailed $100(1 - \alpha)$ confidence region for σ_0^2 is given by

$$(r\hat{\sigma}_0^2/\chi^2_{r,\,\alpha/2}) < \sigma_0^2 < (r\sigma_0^2/\chi^2_{r,\,1-\alpha/2}) \tag{53.195}$$

If σ_0^2 is incorrect or the mathematical model used is improper or incomplete (does not adequately account for systematic errors), then σ_0^2 will fall outside this interval.

Test for Blunders or Outliers

If v_i is the ith residual and σ_{v_i} is its standard deviation, then

$$\bar{v}_i = v_i/\sigma_{v_i} \tag{53.196}$$

is called the *standardized residual*. Frequently, the effort involved in computing $\mathbf{\Sigma}_{vv}$ is quite extensive, and therefore an approximate estimate of σ_{v_i} may be obtained from

$$\hat{\sigma}_{v_i} = [(n-u)/n]^{1/2}\hat{\sigma}_0\sigma_{1_i}/\sigma_0 \tag{53.197}$$

in which n is the number of observations, u is the number of parameters (thus $n - u = r$, the redundancy), and σ_{1_i} is the *a priori* standard deviation of observation l_i. When σ_0^2 is known, $\bar{v}_i$ has a probability density function, or pdf, $N(0, \sigma_{v_i^2})$ and

$$\bar{v}_i = \left|\frac{v_i}{\sigma_{v_i}}\right| < N_{1-\alpha/2} \tag{53.198}$$

If σ_0^2 is not known, then

$$\bar{v}_i = \left|v_i/\hat{\sigma}_{v_i}\right| < \tau_{r,\,1-\alpha/2} \tag{53.199}$$

in which $\hat{\sigma}_{v_i}$ is computed from Eq. (53.197), and τ_r has a Tau pdf with r degrees of freedom. If r is large, as in surveying, photogrammetric, or geodetic nets with extensive observations, τ_r may be replaced by Student t_r pdf or even normal pdf.

Confidence Region for Estimated Parameters

The covariance matrix for the parameters as evaluated from the least squares is given by (see, for example, Eq. (53.161))

$$\Sigma_{\hat{x}\hat{x}} = \sigma_0^2 N^{-1} \tag{53.200}$$

It can be shown that a region of constant probability is bounded by a u-dimensional hyperellipsoid centered at $\hat{\mathbf{x}}$, if the parameters are assumed to have a multivariate normal pdf. The function

$$k^2 = (\mathbf{x} - \hat{\mathbf{x}})^T \sum_{\hat{x}\hat{x}}^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \tag{53.201}$$

describes the hyperellipsoid. The quadratic k^2 has a χ_u^2 distribution, with the probability for a point estimate being

$$P(\chi_u^2 < k^2) = 1 - \alpha$$

For the two-dimensional case (error ellipses), typical values are

p	0.394	0.500	0.900	0.950	0.990
k	1.000	1.177	2.146	2.447	3.035

and for the three-dimensional case, they are

P	0.199	0.500	0.900	0.950	0.990
k	1.000	1.538	2.500	2.700	3.368

when $k = 1$, we usually call it the *standard* region, *standard error ellipse* (for 2-D), or *standard error ellipsoid* (for 3-D). The standard regions for several dimensions are

Dimension	1	2	3	4	5	6
P	0.683	0.394	0.199	0.090	0.037	0.014

Given Σ, for example, for a point in a plane, the semimajor axis, a, and semiminor axis, b, of the standard error ellipse are computed from the eigenvalues and eigenvectors (see the discussion of "Basic Matrix Operations," in Section 53.5). If one is interested in the 90% confidence region (i.e., significance level of $\alpha = 0.10$), the a, b are multiplied by 2.146. For the standard regions, there is a 0.683 probability that an adjusted point falls in a one-dimensional interval, a 0.394 probability that it falls inside the standard error ellipse, and only a 0.199 probability that it falls within the standard error ellipsoid.

Applications in Surveying Engineering

Level Net

Let l_{ij} represent the observed difference in elevation between two points whose (unknown) elevations are x_i and x_j. If l_{ij} is from point i to point j, then the condition equation is given by

$$x_i + l_{ij} + v_{ij} - x_j = 0$$

or

$$v_{ij} + x_i - x_j = -l_{ij} \tag{53.202}$$

This condition equation is in the form of adjustment of indirect observations. At least one benchmark (a point of known elevation) is needed for any given level net.

Traverse

There are two conditions, one for a measured angle α_i, and one for a measured distance d_{ij}. If α_i is at station i from the line $i - 1$ to i clockwise to the line from i to $i + 1$, then

$$\alpha_i = A_{i,\, i+1} - A_{i,\, i-1}$$

or

$$v_i + \tan^{-1}\left(\frac{x_{i-1} - x_i}{y_{i-1} - y_i}\right) - \tan^{-1}\left(\frac{x_{i+1} - x_i}{y_{i+1} - y_i}\right) = -\alpha_i \tag{53.203}$$

where **A** represents the azimuth and $(x, y)_{i-1,\, i,\, i+1}$ are the coordinates of the three points involved. The distance condition is given by

$$v_{ij} - [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2} = -d_{ij} \tag{53.204}$$

Again, this is in the form of adjustment of indirect observations. If any of the points involved is a fixed point (i.e., with known coordinates), its coordinates are not carried as unknown parameters in the adjustment.

Trilateration

This is the operation in which *only* distances are measured in the network. Therefore, the only condition used is that given by Eq. (53.204).

Triangulation

This is the operation in which chains of triangles are connected together. The fundamental measurement is the angle, and the adjustment unit is the triangle. The single condition for one triangle is

$$\sum l_i - \pi = 0 \tag{53.205}$$

where l_i represents all the measured angles inside a single triangle. For a quadrilateral, there is another condition called the *side condition*, the composition of which can be found in the literature.

Defining Terms

Adjustment of observations — The mathematical technique used to resolve the inconsistency between the measurements collected and the underlying mathematical model when *redundancy* exists, or when the measurements exceed the minimum necessary to uniquely define the model.

Azimuthal projection — A map projection that yields a map where correct *direction* or *azimuth* of any point relative to one central point is shown.

Conformal or orthomorphic projection — A map projection technique that preserves *angles* between short intersecting lines, thus making small areas appear in their correct *shape* on the map. Scale varies from point to point, and thus larger areas are incorrect.

Equal-area projection — A map projection technique that results in a map showing all areas in proper relative *size*; however, they may be distorted in *shape*.

Equidistant projection — A map projection technique in which distances are correctly represented from one central point to other points on the map.

Error ellipses and ellipsoids — Confidence regions about estimated survey points in two dimensions (ellipses) or three dimensions (ellipsoids) for specified probabilities.

Error propagation — The general technique of determining the covariance matrix of a set of quantities, which are estimated from functions of another set of quantities of known values and covariance matrix.

Least squares adjustment — The most common adjustment technique used in surveying engineering; it is based on minimizing the sum of the weighted squares of the observational residuals.

Map projection — The theory and techniques of proper representation of the curved earth surface on the plane of a map.

References

Anderson, J.M. and Mikhail, E.M. 1985. *Introduction to Surveying.* McGraw-Hill, New York.

Davis, R.E. et al. 1981. *Surveying: Theory and Practice,* 6th ed. McGraw-Hill, New York.

Krakiwsky, E.J., ed. 1983. *Papers for the CIS Adjustment and Analysis Seminars*. The Canadian Institute of Surveying, Ottawa.

Mikhail, E.M. 1976. *Observations and Least Squares*. University Press of America, Lanham, MD.

Mikhail, E.M. and Gracie, G. 1981. *Analysis and Adjustment of Survey Measurements*. Van Nostrand Reinhold, New York.

Vanicek, P. and Krakiwsky, E.J. 1982. *Geodesy: The Concepts*. North-Holland, Amsterdam.

Further Information

Articles on advances in the general field, and particularly in observational data adjustment, can be found in the following periodicals:

Bulletin Géodésique, published by Springer International

Manuscripta Geodetica, published by Springer International

Photogrammetric Engineering and Remote Sensing, published by the American Society for Photogrammetry and Remote Sensing, Bethesda, MD

Surveying and Land Information Systems, published by the American Congress on Surveying and Mapping, Bethesda, MD

The Photogrammetric Record, published by The Photogrammetric Society, London, England

Photogrammetria, Journal of the International Society for Photogrammetry and Remote Sensing, published by Elsevier, Amsterdam, The Netherlands

Geomatica, Journal of the Canadian Institute of Geomatics, Ottawa, Canada