

APPENDIX A: The ellipse

August 15, 1997

Because of its importance in both approximating the earth's shape and describing satellite orbits, an informal discussion of the ellipse is presented in this appendix. The earth's rotation causes a mass redistribution such that the equatorial radius is larger than the polar radius resulting in an ellipsoidal shape. The other important application of the ellipse to modern geodesy arises from the fact that planetary and artificial earth satellite orbits are elliptical. The mathematics of the ellipse are reviewed to provide all the definitions important in geodesy. Some definitions for satellite orbit mechanics are provided where confusion with geodetic notation often occurs.

The historical usage of ellipse terminology has been developed in several different fields, resulting in multiple ways to define the ellipse. This has led to a confusion of symbols, or notation. In some cases the same notation is used for different quantities. Equally confusing, the same notation is sometimes used whether the axes origin is located at either the center of the ellipse or at one focus of the ellipse. The relationships between the notations of standard mathematic textbooks, of geodesy, and of satellite applications are provided in this appendix. Within geodesy, the notation sometimes varies, and this too is noted. This presentation of the different notations is to assist the user to identify the context, and to enable the user to be able to shift between these contexts.

There is no "official" ellipse definition since it can be defined in so many ways. Some of these definitions are illustrated in this appendix without a rigorous development of the mathematics. A common notation is used in all the examples in order to illustrate the connections between the different ways in which the ellipse may be formed and defined.

I. Ellipse symbols

This table is a compilation of the symbols used for the parameters important for defining the ellipse. The symbol most commonly used, or best related to geodesy, for each parameter is listed in the left hand column. These are the symbols used in the examples of this appendix. Other frequently used symbols are included in the right hand column. In

addition, there are double entries for the symbols used to signify different parameters. These parameters are explained in this appendix.

Table A-1
Ellipse Terminology

Symbol	Parameter	Other symbols
A	Point of apogee	
a	semimajor axis	
b	semiminor axis	
c	half focal separation	ae, ϵ
E	Eccentric, Parametric, or reduced angle or eccentric anomaly	e, t, u, β
e	(first) eccentricity	ϵ
e'	Second eccentricity	ϵ'
F	foci	
f	(first) flattening (or ellipticity)	
f'	Second flattening	
M	Mean anomaly	
P	Point of perigee	
P(x, y)	Points on the ellipse	Q, many
p	Semilatus rectum	
R	Radial distance from focus	r
R _M	Radius of curvature in meridian direction	M
R _N	radius of curvature in prime vertical	N, v, R _v
r	radial distance from center	
S	Distance from focus to ellipse	
α	angular eccentricity	
ϵ	Linear eccentricity	E
Θ	True anomaly	f, θ , v, ψ
θ	central or geocentric angle	ϕ'
ϕ	Geodetic latitude	

II. Ellipse components and definitions

A. Conceptual ellipse

A simple way to illustrate the ellipse is to picture a piece of string with each end fastened to fixed points called focus points or foci (Figure A1). The string length is arbitrarily set to $2a$. If a pencil is used to pull the string tight and is then moved around the foci, the resulting shape will be an ellipse. The length of string remains constant at $2a$, but the distance (S_1 and S_2) from the pencil to each focus will change at each point. The foci are located at F_1 and F_2 , and A and B are two arbitrary points on the ellipse. All the points on the closed curve defined by the ellipse are represented by the set of x and y points, $P(x,y)$.

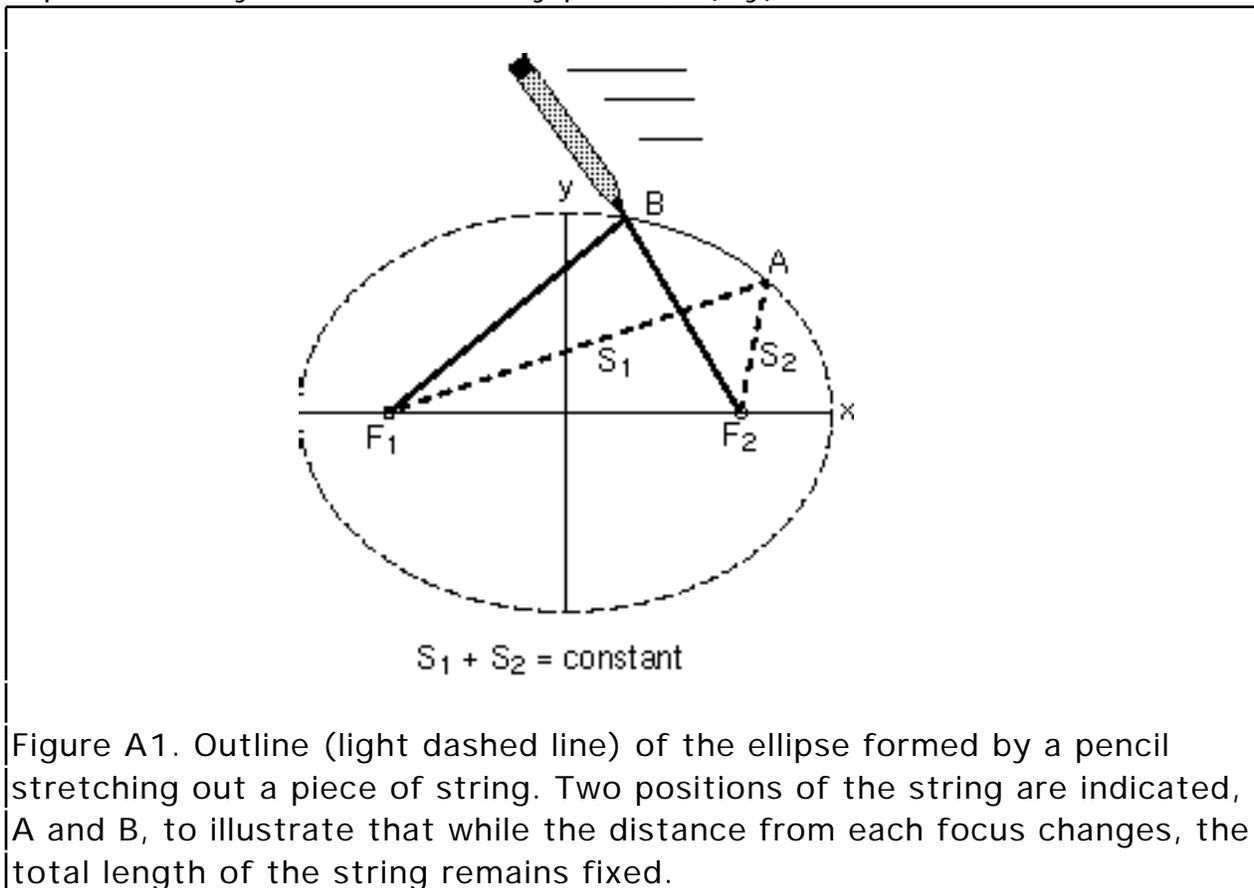


Figure A1. Outline (light dashed line) of the ellipse formed by a pencil stretching out a piece of string. Two positions of the string are indicated, A and B, to illustrate that while the distance from each focus changes, the total length of the string remains fixed.

The line connecting the foci defines an axis of symmetry, the major axis,

for the ellipse. In this appendix, the foci and major axis will always be located on the x axis. The perpendicular line passing through the mid-point between the foci is also an axis of symmetry. This line defines the minor axis. The intersection of the two axes is the center of the ellipse. The coordinate systems used to define the ellipse will be located either at the ellipse center or at one of the foci. The convention is that the distances from the origin to the foci are $\pm c$ and are referred to as the half focal separation (Figures A2 and A3).

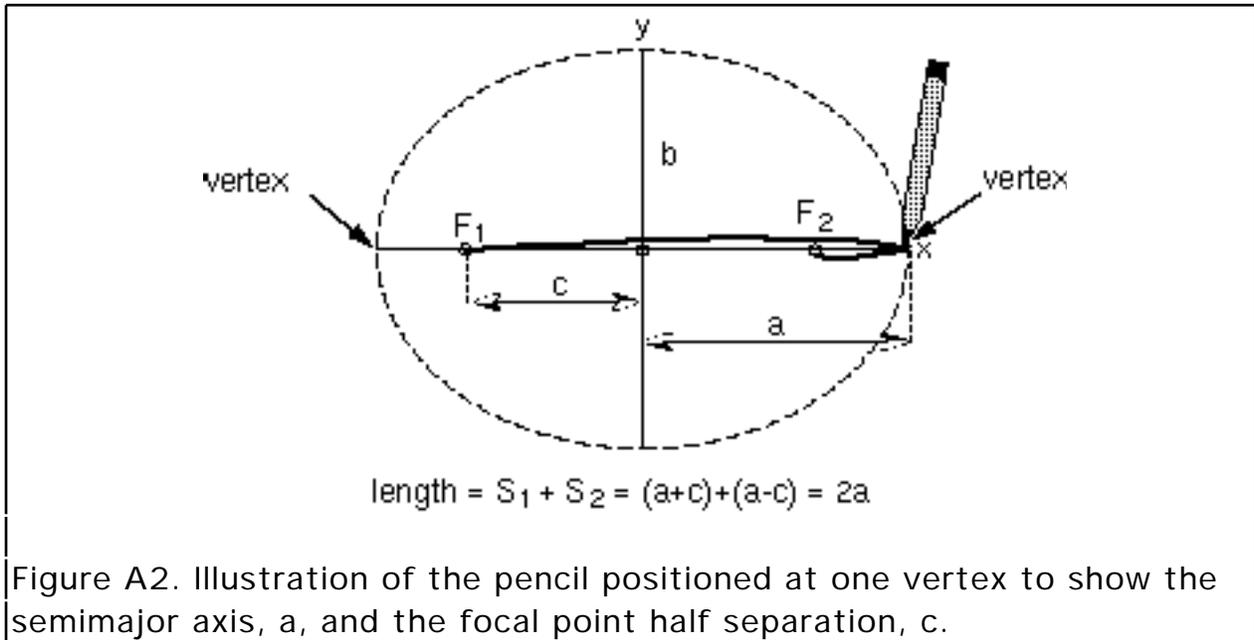


Figure A2. Illustration of the pencil positioned at one vertex to show the semimajor axis, a , and the focal point half separation, c .

When the point P is located on the x axis, $P(\pm x, 0)$, (Figure A2), the two string segments will lie on top of one another (note that in the diagram the segments are drawn curved so that they can be observed). This point of intersection of the ellipse with the x axis is called the vertex. The distance between the origin and one vertex is half the length of the string. With the string length given as $2a$ the distances along the x axis from the origin to the vertices are $\pm a$. The distance from the center of the ellipse to the vertex is called the semimajor axis of length a .

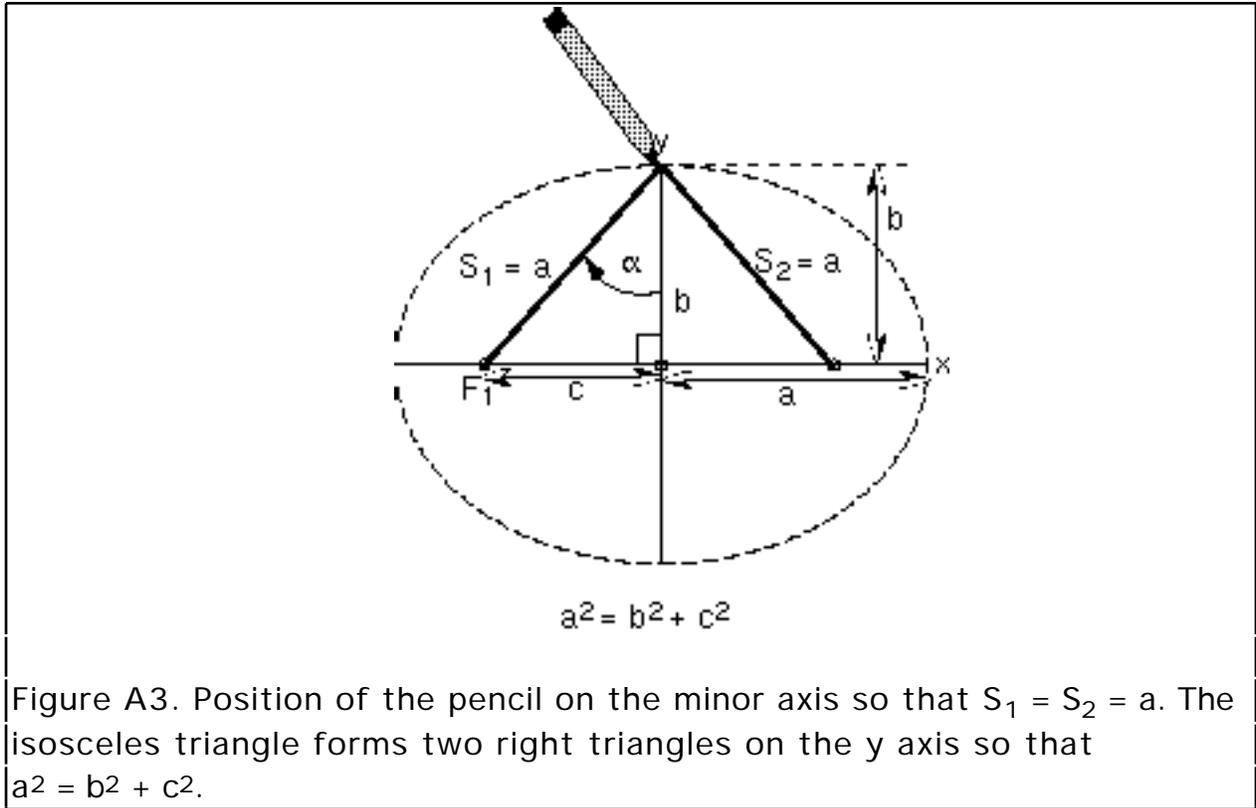


Figure A3. Position of the pencil on the minor axis so that $S_1 = S_2 = a$. The isosceles triangle forms two right triangles on the y axis so that $a^2 = b^2 + c^2$.

When the ellipse intersects the y axis, $P(0, \pm y)$, the two segments of the string are equal, forming an isosceles triangle (Figure A3). This triangle is divided into two similar right triangles by the y axis. The distance from the ellipse center to the ellipse is called the semiminor axis of length b . (Note that the hypotenuse of each triangle is equal in length to the semimajor axis.) By the Pythagorean theorem,

$$a^2 = b^2 + c^2. \tag{A1}$$

1. Ellipse axes terminology

The foci are always located on the major axis and a , b , and c are used to represent the semimajor axis, the semiminor axis, and the half focal separation (or the distance from the center of the ellipse to one foci) (Figure A2). The center of the ellipse is the point of intersection of its two axes of symmetry.

2. Eccentricity

One way to specify the shape of an ellipse is given by the eccentricity, e . The ellipse eccentricity, e , can vary between 0 and 1. An eccentricity of 0 means the foci coincide, and the shape will be a circle of radius a . At the other limit, $e=1$, is a line $2a$ in length passing through the foci. The usual geodetic definition is.

$$e^2 = 1 - \frac{b^2}{a^2} \quad (\text{A2})$$

Other forms of the eccentricity common for geodetic applications include

$$\begin{aligned} e^2 &= \frac{a^2 - b^2}{a^2}, \\ e &= \frac{\sqrt{a^2 - b^2}}{a}, \\ e^2 &= \frac{c^2}{a^2}. \end{aligned} \quad (\text{A3})$$

Other common forms of this relationship are:

$$\begin{aligned} \frac{b^2}{a^2} &= 1 - e^2 & b^2 &= a^2(1 - e^2) \\ \frac{b}{a} &= (1 - e^2)^{\frac{1}{2}} & b &= a \sqrt{1 - e^2} \\ e &= \frac{c}{a} & c &= a e \\ c &= \sqrt{a^2 - b^2} \end{aligned} \quad (\text{A4})$$

The angle α in Figure A3 is referred to as the angular eccentricity since

$$\sin \alpha = \frac{c}{a} = e. \quad (\text{A5})$$

(NOTE: in different texts ε appears in one of three ways, either as the half focal separation, c , as the eccentricity, e , and as the linear

eccentricity, $\epsilon = \sqrt{a^2 - b^2}$.)

3. Ellipse flatness

Ellipse shape is also expressed by the flatness, f . In geodesy the shape of the ellipsoid (ellipsoid of rotation) that represents earth models is usually specified by the flatness. The flatness is computed as:

$$f = 1 - \frac{b}{a},$$

or

$$f = \frac{a-b}{a}.$$

(A6)

The relationship between e and f is

$$e^2 = 2f - f^2,$$
$$f = 1 - \sqrt{1 - e^2}.$$

(A7)

4. Second eccentricity and second flatness

The eccentricity and flatness, e and f , are both defined by a ratio with the semimajor axis. These are also referred to the first eccentricity and the first flatness. Analogous quantities defined as the ratio to the semiminor axis are referred to as the second eccentricity, e' , and the second flatness, f' ,

$$e'^2 = \frac{a^2 - b^2}{b^2},$$
$$f' = \frac{a}{b} - 1.$$

(A8)

Other forms for the second eccentricity are:

$$\frac{a^2}{b^2} = e'^2 + 1,$$
$$\frac{a}{b} = (1 + e'^2)^{\frac{1}{2}}$$

(A9)

5. Specifying an ellipse

The shape and size of an ellipse can be specified by any pair combination of a or b with c , e , e' , f or f' . Different applications use different sets. The common combinations are:

- semimajor and semiminor axes (a , b),
- semimajor and eccentricity (a , e),
- semimajor and flatness (a , f).

6. The directrix

The directrix is a straight line perpendicular to the major axis. The unique property of the directrix is that the horizontal distance from a point P on the ellipse to the directrix is proportional to the distance from the closest focus to that point (see the left hand side of figure A4). The constant of proportionality is e . Since the ellipse has two foci, the ellipse has two directrices and they are located $\pm(a/e)$ from the ellipse center (Figure A4).

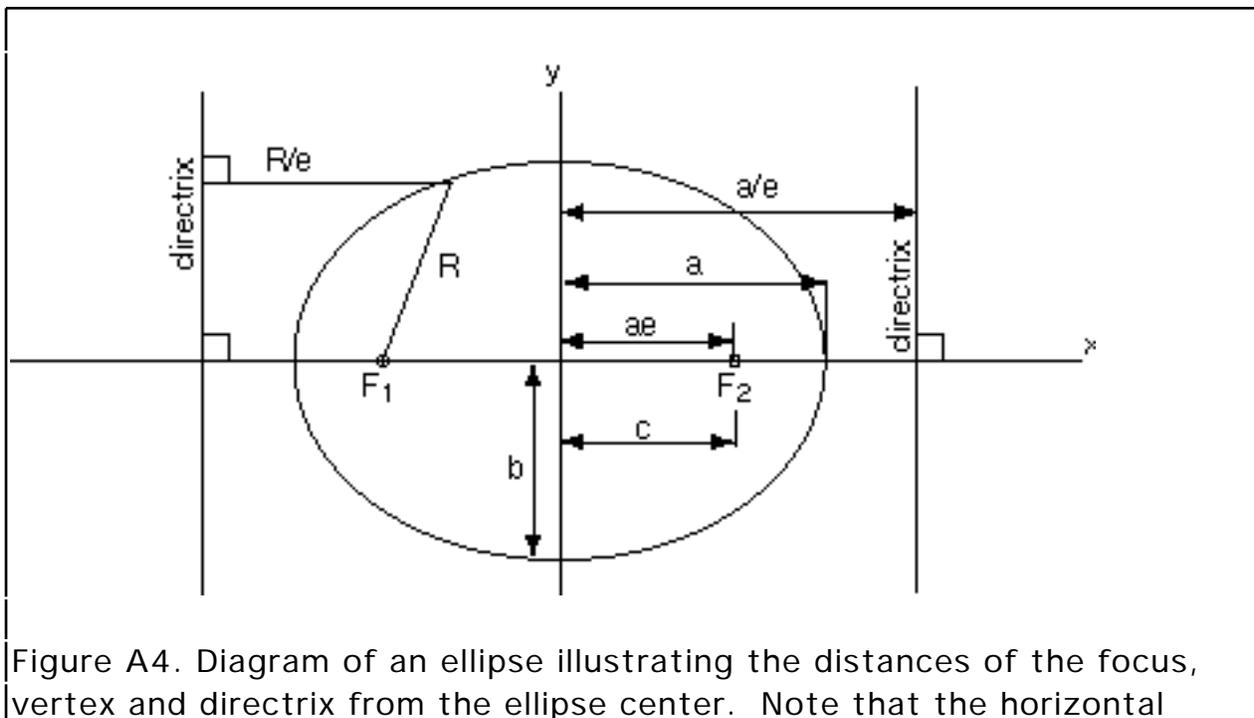


Figure A4. Diagram of an ellipse illustrating the distances of the focus, vertex and directrix from the ellipse center. Note that the horizontal

distance to the directrix from the ellipse is proportional by e to the distance from that point to the closest focus.

III. The mathematical ellipse

A mathematical definition of the ellipse is the locus of points $P(x,y)$ whose sum of distances from two fixed points, the foci, is constant. When the foci coincide, the ellipse is a circle, and as e is increased, the distance between the foci increases and the shape becomes more elongated, or squashed, until in the limit it is a straight line.

In specifying an ellipse mathematically, it is important to know the location chosen for the origin. There are two common conventions, the origin at either the center or at one focus of the ellipse. This section presents the equations for both the origin at the ellipse center and the origin at one focus. For consistency, all focus centered equations have the right hand focus (F_2 of figure A4) as the origin.

A. The ellipse equation

1. Cartesian coordinates

a. Centered origin

The equation for the ellipse in Cartesian coordinates with the origin at the ellipse center is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (\text{A10})$$

The constants a and b are the semimajor and semiminor axes.

b. Focus origin

When using the origin at the focus the Cartesian form of the equation is

$$\frac{(x-c)^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (\text{A11})$$

The offset c is the distance from the origin to a focus. Other formulas can be derived using the relationships between a , b , and c (Equation A1).

2. Ellipse in polar coordinates

When using polar coordinates, the ellipse can be specified with the origin at either the ellipse center or the origin at one focus. The reader should be aware that r and θ are regularly used to define a point on the ellipse for both coordinate systems. When measured from the ellipse center, θ is the central angle and r is the distance from the ellipse center. When θ is measured at one focus it is called the true anomaly and r is the distance from the focus to the point on P . To distinguish the two, upper case symbols Θ and R will be used in equations centered on the focus. Many applications fail to make clear the distinction between the true anomaly and central angle and the different distances represented by r .

a. Ellipse centered origin

The ellipse can be drawn as a distance r from the ellipse center where the length of r depends on the central angle θ (Figure A5).

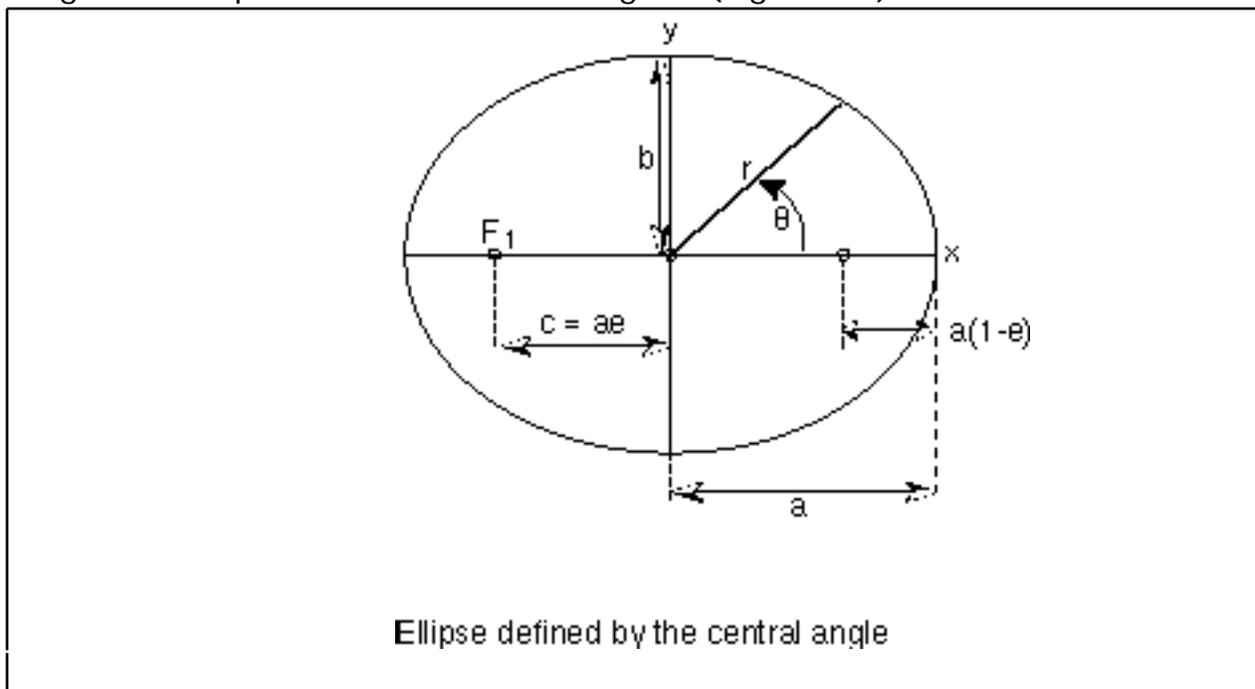


Figure A5. The ellipse defined by the central angle θ and the radius r . The angle θ is measured counter clockwise from the semimajor axis.

Using the central angle, θ , the length of r is determined by any of the following:

$$\begin{aligned}
 r^2 &= \frac{b^2}{1 - e^2 \cos^2 \theta}, \\
 r^2 &= \frac{a^2}{1 - e^2 \sin^2 \theta}, \\
 r^2 &= \frac{a^2(1 - e^2)}{1 - e^2 \cos^2 \theta}, \\
 r^2 &= \frac{a^2 b^2}{a^2 \sin^2 \theta + b^2 \cos^2 \theta}.
 \end{aligned}
 \tag{A12}$$

One can check equations like A12 by evaluating r for θ equals 0° and 90° . For θ equals 0° , r is on the x axis and r^2 equals a^2 . For θ equals 90° , r is on the y axis and r^2 equals b^2 .

The conversion between Cartesian and polar coordinates is obtained from

$$\begin{aligned}
 x &= r \cos \theta, \\
 y &= r \sin \theta.
 \end{aligned}
 \tag{A13}$$

b. Ellipse centered at one focus

Figure A6 presents the definitions used to locate a point on the ellipse when measured from a focus. The origin is set at one focus and the vertex closest to the origin is called the point of perigee for earth satellites. The vertex farthest to the origin is then called the apogee. The angle Θ , the true anomaly, is measured at the focus, moving counter clockwise from perigee. In terms of these variables the radius from the focus is given by:

$$R = \frac{a(1 - e^2)}{(1 + e \cos \Theta)}.
 \tag{A14}$$

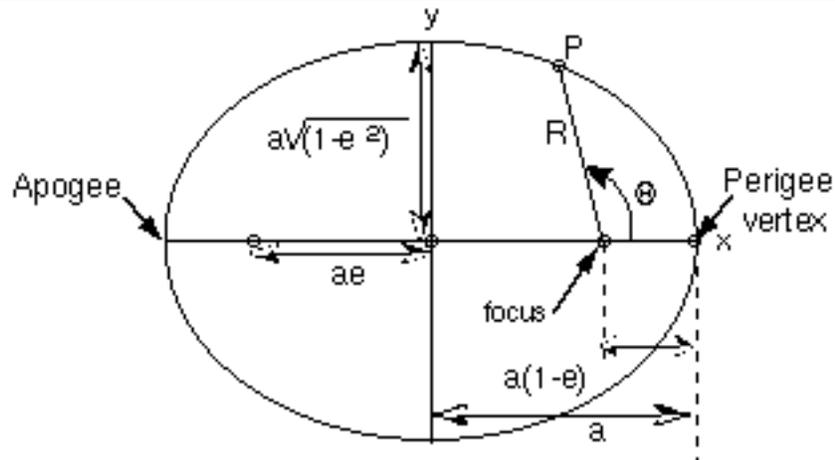


Figure A6. Ellipse defined by the true anomaly Θ , measured from the focus, and the radial distance R .

i. The semilatus rectum

When the true anomaly, Θ , is 90° , the radius, R , is called the semilatus rectum, p . The semilatus rectum is the line parallel to the minor axis from the focus to the ellipse (Figure A7).

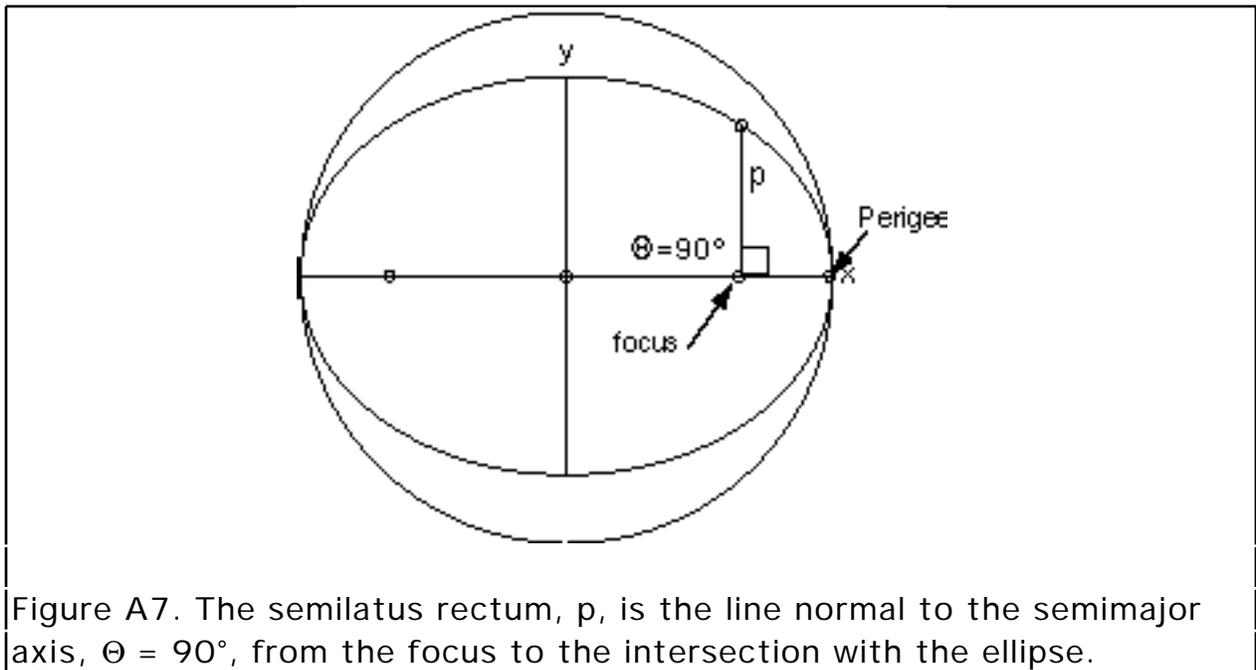
$$\begin{aligned}
 p &= a(1 - e^2), \\
 p &= \frac{b^2}{a}.
 \end{aligned}
 \tag{A15}$$

Equations relating p to a , b , and c are:

$$\begin{aligned}
 a^2 &= \frac{4e^2 p^2}{(1-e^2)^2}, \\
 p &= \left[\frac{1-e^2}{2e} \right] a, \\
 b^2 &= \frac{4e^2 p^2}{(1-e^2)}, \\
 p &= \left[\frac{(1-e^2)^{\frac{1}{2}}}{2e} \right] b, \\
 c &= \frac{2e^2 p}{(1-e^2)}.
 \end{aligned}
 \tag{A16}$$

The radial distance from the foci, R , is given by:

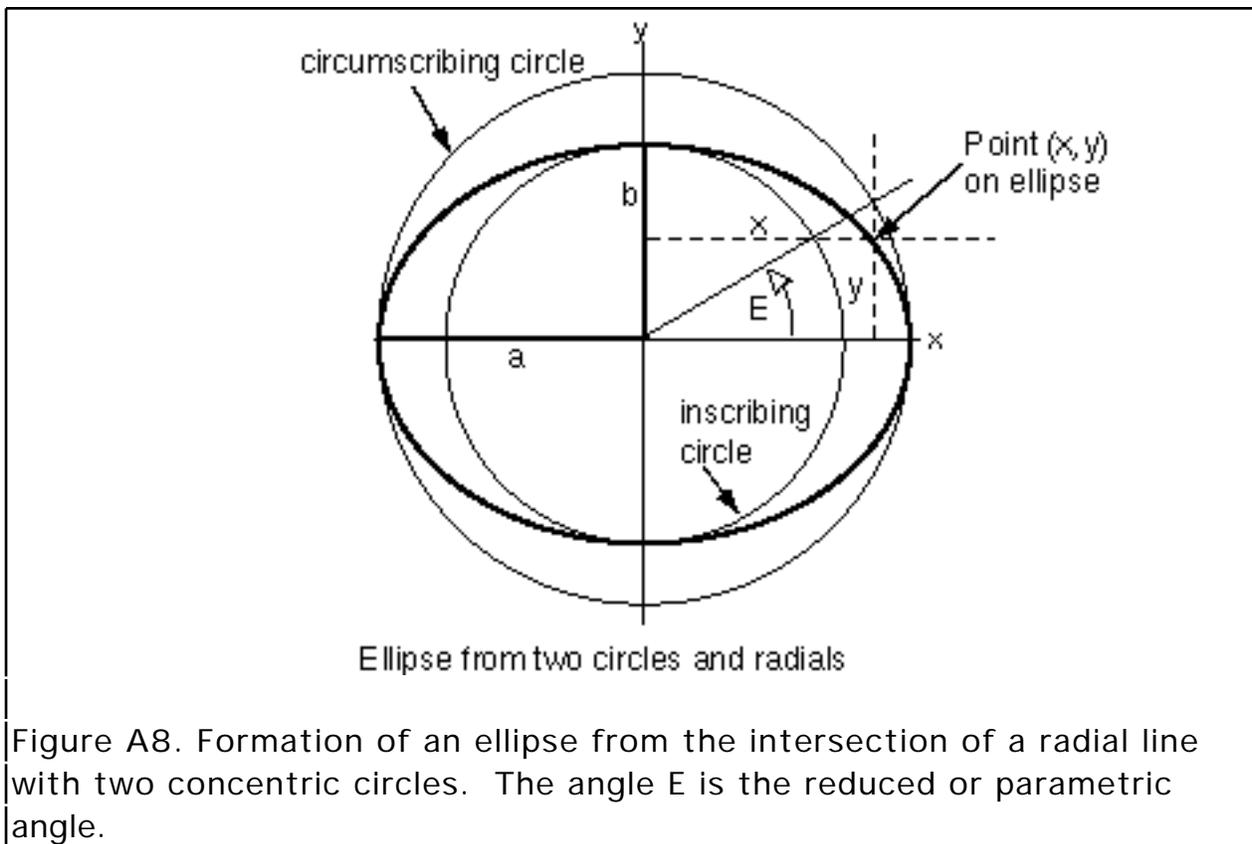
$$\begin{aligned}
 R &= e(2p + R \cos \Theta), \\
 R &= \frac{2ep}{1 - \cos \Theta}.
 \end{aligned}
 \tag{A17}$$



B. Ellipse from circles and radiating lines

1. Ellipse from the intersection of two concentric circles with radiating lines

An important way to construct an ellipse is illustrated in Figure A8. Two concentric circles of radii, a and b , define the ellipse. The radius of the inscribed circle, b , defines the minor axis and the radius of the circumscribed circle, a , defines the major axis. Next, a radial line is drawn from the center at angle E . The angle E is called the eccentric anomaly in satellite work and the reduced latitude in geodesy. It is also called the parametric angle. Often the development of the ellipse is given only showing the circumscribing circle.



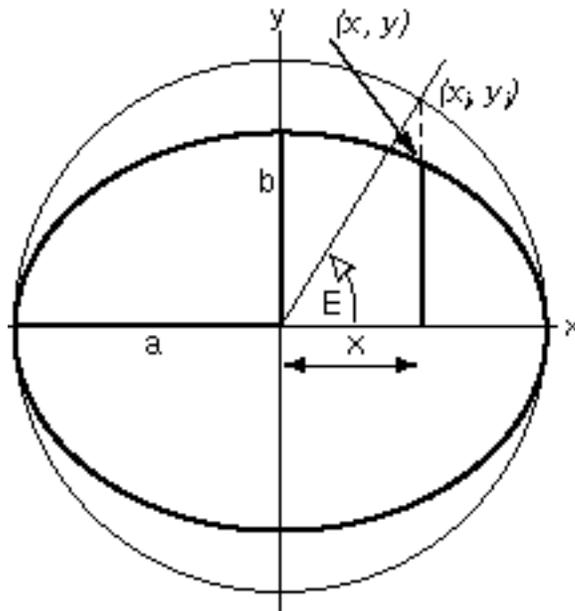
The points that the radial line intersect with the two circles give the x and y coordinates. The y coordinate is taken from the intersection with the smaller inscribed circle of radius b . The x coordinate is taken from the intersection with the larger circumscribing circle of radius a . The x and y coordinates of the ellipse are given by:

$$\begin{aligned}
 x &= a \cos E, \\
 y &= b \sin E, \\
 y &= a(1 - e^2)^{\frac{1}{2}} \sin E
 \end{aligned}
 \tag{A18}$$

2. One-way reduction of a circle

An alternate way to define an ellipse from a circumscribed circle and radial lines is a one-way reduction of a circle, a kind of foreshortening (Figure A9). Radial lines of angle E are drawn to the circumscribed circle. The point of intersection is (x_i, y_i) . The x value is found as in the previous example. The y_i value of this intersection is scaled by b/a to give the y value of the ellipse.

$$\begin{aligned}
 x_i &= a \cos E \\
 x &= x_i \\
 y_i &= a \sin E \\
 y &= \frac{b}{a} y_i \\
 y &= b \sin E
 \end{aligned}
 \tag{A18a}$$



Ellipse from one-way reduction

Figure A9. Ellipse from one way reduction of lines normal to the semimajor axis. The orthogonal lines from the intersection of the radial at angle E on the circle, dashed line, are reduced by the constant b/a , heavy line.

C. Conic section

An ellipse is also formed by the intersection of a right circular cone and a plane inclined less steeply than the side of the cone (Figure A10). When the plane does not pass through the base the shape of the intersection is an ellipse. The eccentricity is determined by the steepness of the cone and the angle of intersection between the cone and the intersecting plane. When the intersecting plane is parallel to the cone base, the intersecting line is a circle.

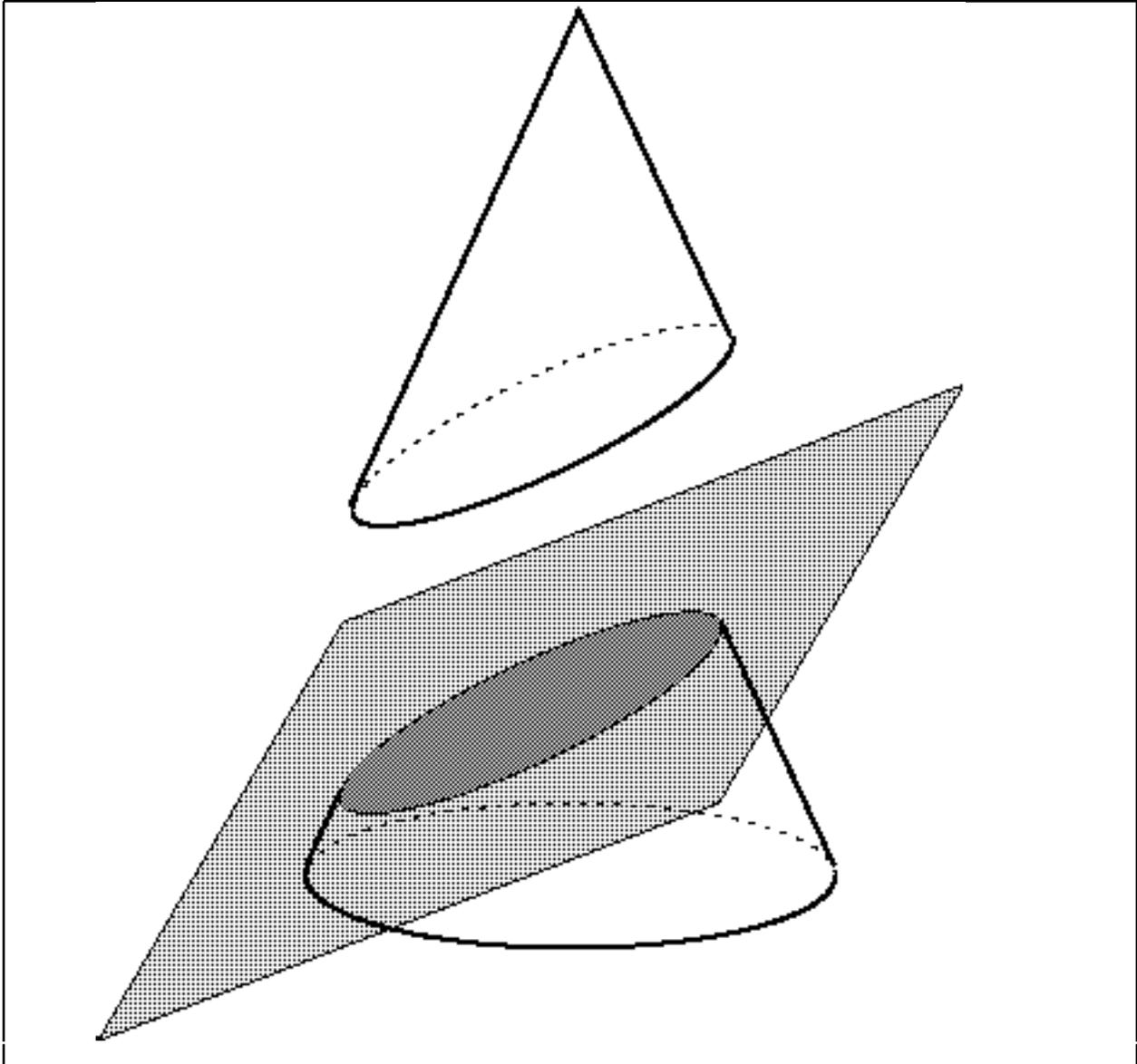


Figure A10. The ellipse formed by the intersection of a plane, light grey, with a right circular cone. This exploded view shows the ellipse of intersection in dark grey.

The circle, ellipse, parabola, and hyperbola are call conic sections because they can be generated in this manner.

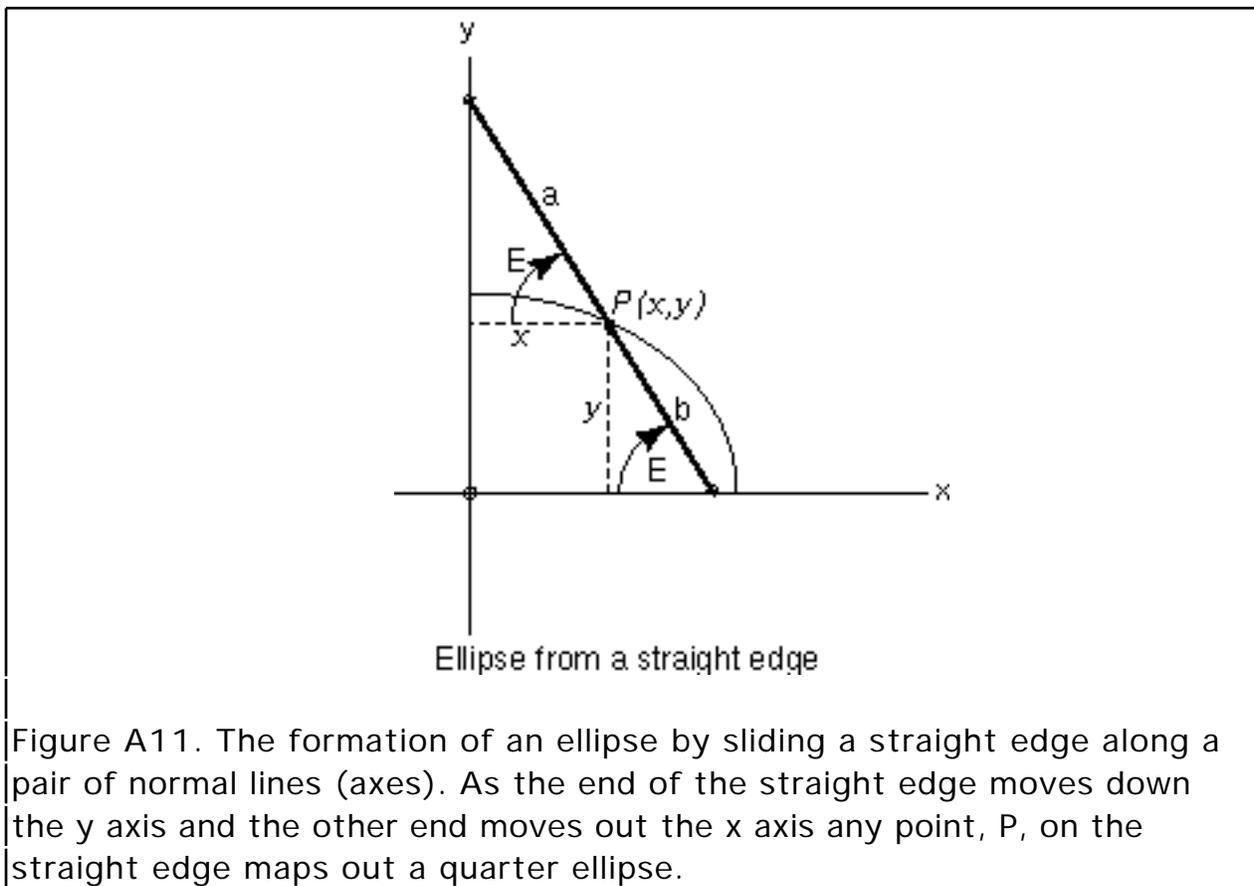
D. Ellipse from a straight edge

A straight edge of length $a+b$ can be used to construct $1/4$ of an ellipse.

Let the ends be attached to, but able to slide along the axes (Figure A11). The point P is a distance a from the end of the straight edge on the y axis. As the one end of the straight edge slides down the y axis from $y = a+b$ to $y = 0$, the location of point P will map out the curve of the ellipse. The angle E between the straight edge and the x axis is used to define the x and y coordinates. This angle has the same magnitude as the eccentric anomaly.

$$\begin{aligned} x &= a \cos E, \\ y &= b \sin E. \end{aligned} \tag{A19}$$

The full ellipse is created by repeating the exercise in all four quadrants.



E. Tangent to the ellipse and radius of curvature

Two important aspects of the ellipse needed for geodesy are the tangent to the ellipse and the radius of curvature. For the earth the tangent will be (approximately) the local horizontal plane. The radius of curvature of the ellipse is one of the “effective radii” of the earth needed to convert angular differences to linear distances. (Both “effective radii” are described in the ellipsoid appendix.)

Figure A12 illustrates the tangent line to the ellipse at a point P. The line perpendicular to the point of tangency is also drawn and is labeled PQ. The distance from P to Q, the point of intersection with the y axis, defines the radius of curvature, R_N . Note that R_N does not intersect the y axis at the origin and it forms an angle ϕ with the x axis. In geodesy applications ϕ is called the geodetic latitude and is the latitude found on maps. (In geodesy applications θ , the geocentric angle, is signified by ϕ' and called the geocentric latitude.)

The centered Cartesian coordinates in terms of ϕ are given by

$$\begin{aligned} x &= \frac{a \cos \phi}{(1 - e^2 \sin^2 \phi)^{\frac{1}{2}}}, \\ y &= \frac{a(1 - e^2) \sin \phi}{(1 - e^2 \sin^2 \phi)^{\frac{1}{2}}}. \end{aligned} \tag{A20}$$

And the radius of curvature, R_N , is obtained from

$$R_N = \frac{a}{(1 - e^2 \sin^2 \phi)^{\frac{1}{2}}}. \tag{A21}$$

The Cartesian coordinates from R_N are obtained by

$$\begin{aligned} x &= R_N \cos \phi, \\ y &= R_N (1 - e^2) \sin \phi \end{aligned} \tag{A22}$$

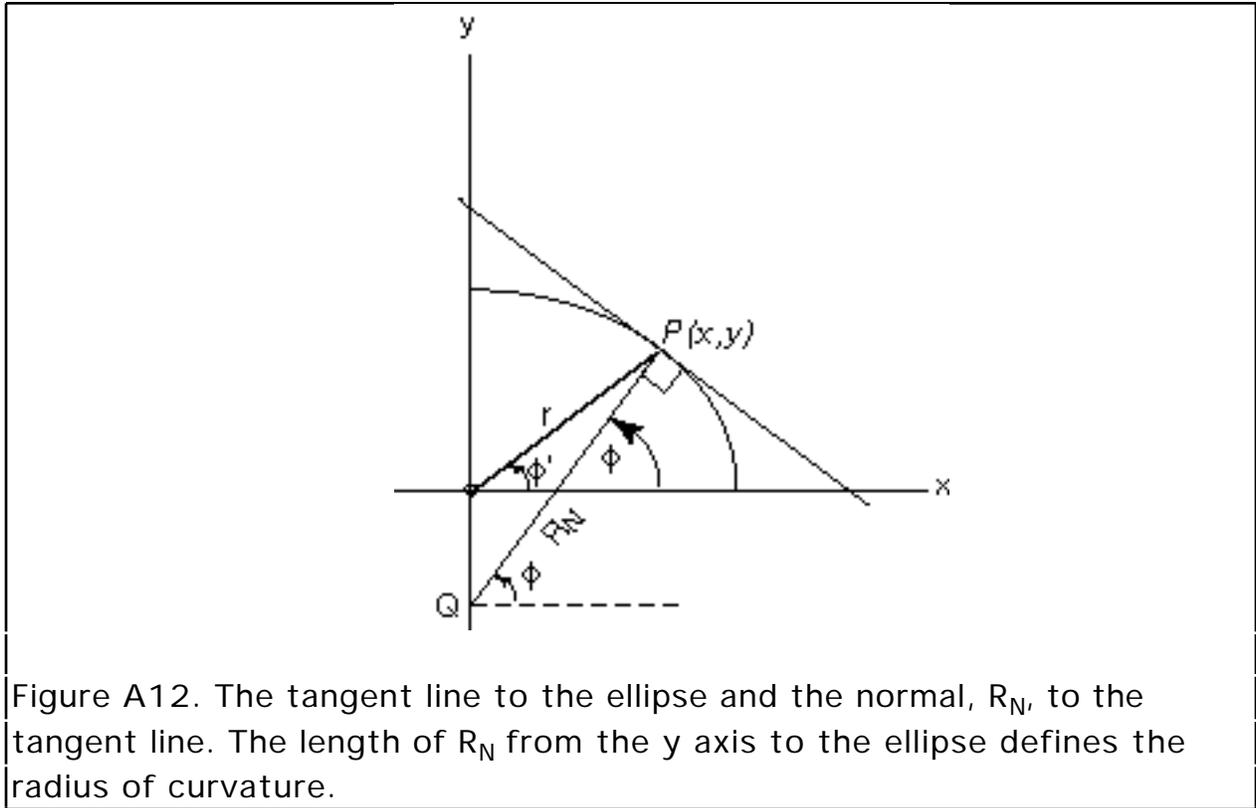


Figure A12. The tangent line to the ellipse and the normal, R_N , to the tangent line. The length of R_N from the y axis to the ellipse defines the radius of curvature.

IV. Coordinate conversions

It is often necessary to convert the point on the ellipse to either a different coordinate centered system or to convert between the different polar angles. The conversion equations are presented in this section.

A. Cartesian conversion between centered origin and focus origin

The transformation of the Cartesian coordinate systems with the origins at the center and a focal point are accomplished by moving the origin of the x axis. When using the focus centered origin, only the x axis is an axis of symmetry. Using the following symbols for

- centered origin coordinates (x, y)
- focal point origin coordinates (x', y')

$$\begin{aligned}
 x' &= x - c, \\
 x' &= x - ae, \\
 y' &= y
 \end{aligned}
 \tag{A23}$$

B. Summary of angles

A very busy diagram with most of the lines and angles discussed in this appendix is shown in Figure A13. Subsets of this figure are shown in Figures A14 and A15.

Figure A13 summarizes the four polar angles, E , ϕ , θ , Θ , and the two radii, R , r , plus the radius of curvature R_N . A summary of the different equations for converting between the four polar measurements, and cartesian coordinates, concludes this appendix.

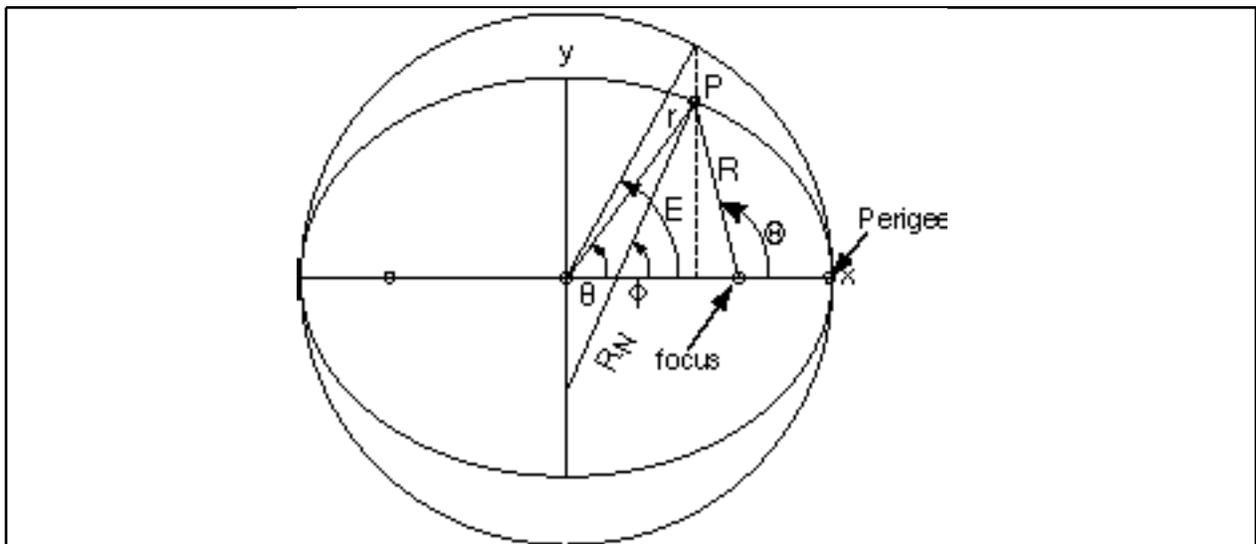


Figure A13. Summary diagram showing the differences between r , R , R_N and E , θ , Θ , and ϕ .

Figure A14 illustrates the three center origin polar measurements. The various conversions for the cartesian location of the ellipse to the different measurement angles are provided in the following tables.

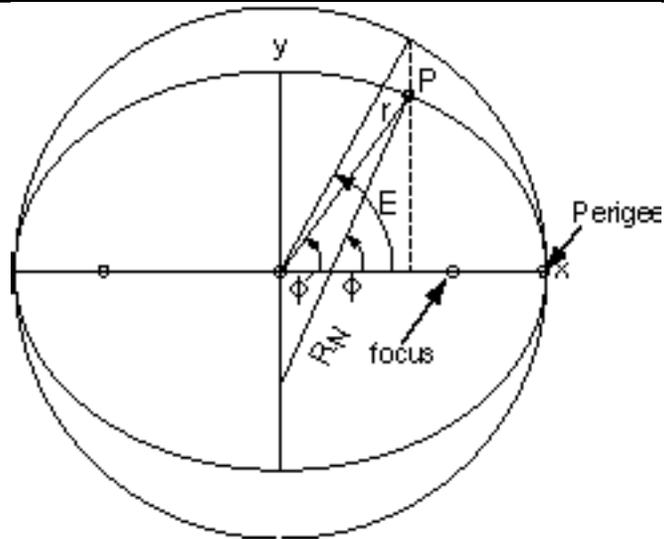


Figure A14. Summary diagram illustrating the differences between r , R_N and E , ϕ , and ϕ' .

Table A-2.

Conversion of cartesian coordinates to the different polar coordinates.

Cartesian	geocentric (centric) $\phi' (= \theta)$	eccentric E	Geodetic ϕ
$x =$	$r \cos \phi'$	$= a \cos E$	$= R_N \cos \phi$ $= \frac{a \cos \phi}{\sqrt{1 - a^2 \sin^2 \phi}}$ $= \frac{a^2 \cos \phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}$

$y =$	$r \sin \phi'$	$= b \sin E$	$= R_N(1 - e^2)\sin \phi$ $= \frac{b}{a} \frac{b \sin \phi}{\sqrt{1 - e^2 \sin^2 \phi}}$ $= \frac{b^2 \sin \phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}$
$\frac{y}{x} =$	$\tan \phi'$	$= \frac{b}{a} \tan E$	$= \frac{b^2}{a^2} \tan \phi$

Table A-3.

Conversion of the radial distance from the ellipse center for the three center origin angles.

r^2	$= \frac{a^2}{1 - e^2 \sin^2 \phi'}$	$= a^2(1 - e^2 \sin^2 E)$	$= R_N[\cos^2 \phi + \frac{b^4}{a^4} \sin^2 \phi]$
	$= \frac{b^2}{1 - e^2 \cos^2 \phi'}$	$= a^2 \cos^2 E + b^2 \sin^2 E$	$= \frac{a^4 \cos^2 \phi + b^4 \sin^2 \phi}{a^2 \cos^2 \phi + b^2 \sin^2 \phi}$
	$= \frac{a^2 b^2}{a^2 \sin^2 \phi' + b^2 \cos^2 \phi'}$		

Table A-4.

Center origin angle conversions

$\cos E$	$= \frac{\cos \phi}{\sqrt{1 - e^2 \sin^2 \phi}}$	$= \frac{R_N}{a} \cos \phi$
$\sin E$	$= \frac{b}{a} \frac{\sin \phi}{\sqrt{1 - e^2 \sin^2 \phi}}$	$= \frac{R_N b}{a^2} \sin \phi$

$\cos \phi'$	$= \frac{a}{r} \frac{\cos \phi}{\sqrt{1 - e^2 \sin^2 \phi}}$	$= \frac{R_N}{r} \cos \phi$
$\sin \phi'$	$= \frac{b^2}{a r} \frac{\sin \phi}{\sqrt{1 - e^2 \sin^2 \phi}}$	$= \frac{R_N b^2}{r a^2} \sin \phi$
		$= \frac{R_N}{r} (1 - e^2) \sin \phi$
$\cos \phi'$	$= \frac{a}{r} \cos E$	
$\sin \phi'$	$= \frac{b}{r} \sin E$	

4. Conversion between the focus origin and the eccentric anomaly

Unlike the center origin angles, E and θ , which are always in the same quadrant, Θ can be in a different quadrant (as drawn in Figure A15). When the point on P lies between the center y axis and the latus rectum, Θ will be in a different quadrant than the center angles. The usual angle conversion procedure is to find $\cos \Theta$ and $\sin \Theta$ from E (Table A-4) and then use a four quadrant arc tangent to find Θ .

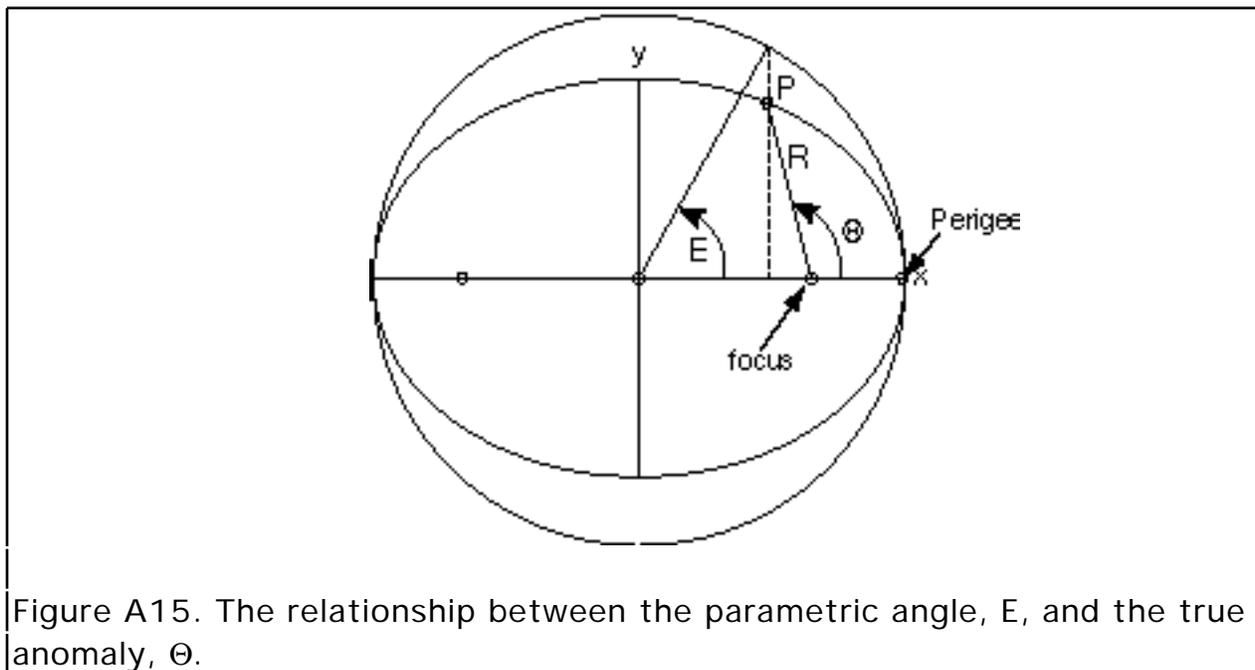


Figure A15. The relationship between the parametric angle, E , and the true anomaly, Θ .

Figure A15 illustrates the measurement location for the eccentric anomaly, E, and the true anomaly, Θ . Below are given the conversion equations for transforming from one angle to another.

Table A-5.

Conversion between eccentric anomaly and true anomaly.

x'	$= R \cos \Theta$	$= a(\cos E - e)$
y'	$= R \sin \Theta$	$= b \sin E$
	$\cos \Theta$	$= \frac{\cos E - e}{1 - e \cos E}$ $= \frac{a \cos E - e}{a(1 - e \cos E)}$
	$\sin \Theta$	$= \frac{\sqrt{1 - e^2}}{1 - e \cos E} \sin E$ $= \frac{b}{a} \frac{\sin E}{1 - e \cos E}$
R	$= \frac{a(1 - e^2)}{1 - e \cos \Theta}$ $= \frac{b^2}{a} \frac{1}{1 - e \cos \Theta}$	$= a(1 - e \cos E)$