CHAPTER 2
CONIC SECTIONS

2.1 *Introduction*

A particle moving under the influence of an inverse square force moves in an orbit that is a conic section; that is to say an ellipse, a parabola or a hyperbola. We shall prove this from dynamical principles in a later chapter. In this chapter we review the geometry of the conic sections. We start off, however, with a brief review (eight equation-packed pages) of the geometry of the straight line.

2.2 *The Straight Line*

It might be thought that there is rather a limited amount that could be written about the geometry of a straight line. We can manage a few equations here, however, (there are 35 in this section on the Straight Line) and we shall return for more on the subject in Chapter 4.

Most readers will be familiar with the equation

$$y = mx + c \qquad 2.2.1$$

for a straight line. The slope (or gradient) of the line, which is the tangent of the angle that it makes with the x-axis, is m, and the intercept on the y-axis is c. There are various other forms that may be of use, such as

$$\frac{x}{x_0} + \frac{y}{y_0} = 1 \qquad 2.2.2$$

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1} \qquad 2.2.3$$

which can also be written

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0 \qquad 2.2.4$$

$$x \cos\theta + y \sin\theta = p \qquad 2.2.5$$

The four forms are illustrated in figure II.1.

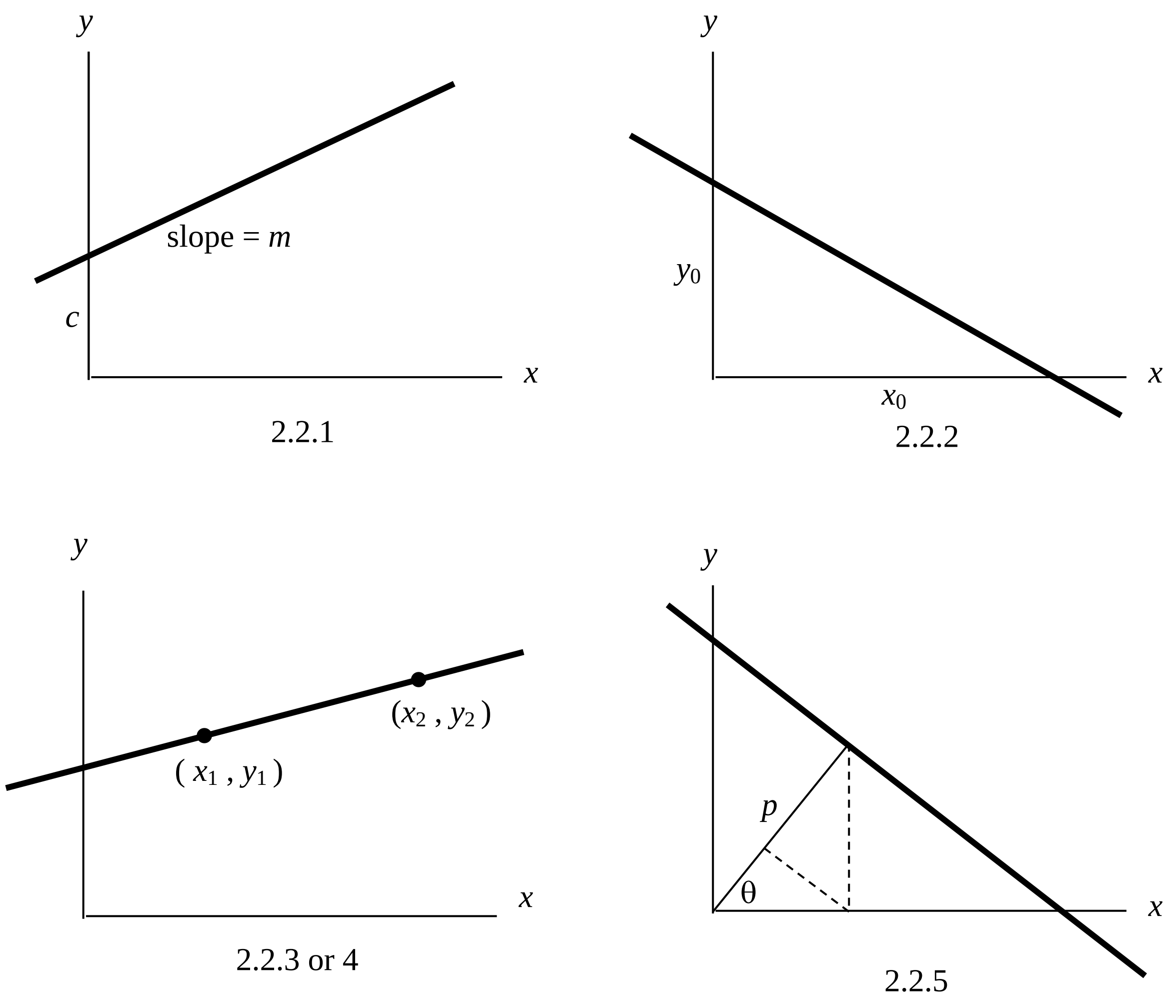


FIGURE II.1

A straight line can also be written in the form

$$Ax + By + C = 0. \tag{2.2.6}$$

If $C = 0$, the line passes through the origin. If $C \neq 0$, no information is lost, and some arithmetic and algebra are saved, if we divide the equation by C and re-write it in the form

$$ax + by = 1. \tag{2.2.7}$$

Let P (x , y) be a point on the line and let P_0 (x_0 , y_0) be a point in the plane not necessarily on the line. It is of interest to find the perpendicular distance between P_0 and the line. Let S be the square of the distance between P_0 and P. Then

$$S = (x - x_0)^2 + (y - y_0)^2 \qquad 2.2.8$$

We can express this in terms of the single variable x by substitution for y from equation 2.2.7. Differentiation of S with respect to x will then show that S is least for

$$x = \frac{a + b(bx_0 - ay_0)}{a^2 + b^2} \qquad 2.2.9$$

The corresponding value for y, found from equations 2.2.7 and 2.2.9, is

$$y = \frac{b + a(ay_0 - bx_0)}{a^2 + b^2}. \qquad 2.2.10$$

The point P described by equations 2.2.9 and 2.2.10 is the closest point to P_0 on the line. The perpendicular distance of P from the line is $p = \sqrt{S}$ or

$$p = \frac{1 - ax_0 - by_0}{\sqrt{a^2 + b^2}}. \qquad 2.2.11$$

This is positive if P_0 is on the same side of the line as the origin, and negative if it is on the opposite side. If the perpendicular distances of two points from the line, as calculated from equation 2.2.11, are of opposite signs, they are on opposite sides of the line. If p = 0, or indeed if the numerator of 2.2.11 is zero, the point P_0 (x_0 , y_0) is, of course, on the line.

Let A(x_1 , y_1), B(x_2 , y_2) and C(x_3 , y_3) be three points in the plane. What is the area of the triangle ABC? One way to answer this is suggested by figure II.2.

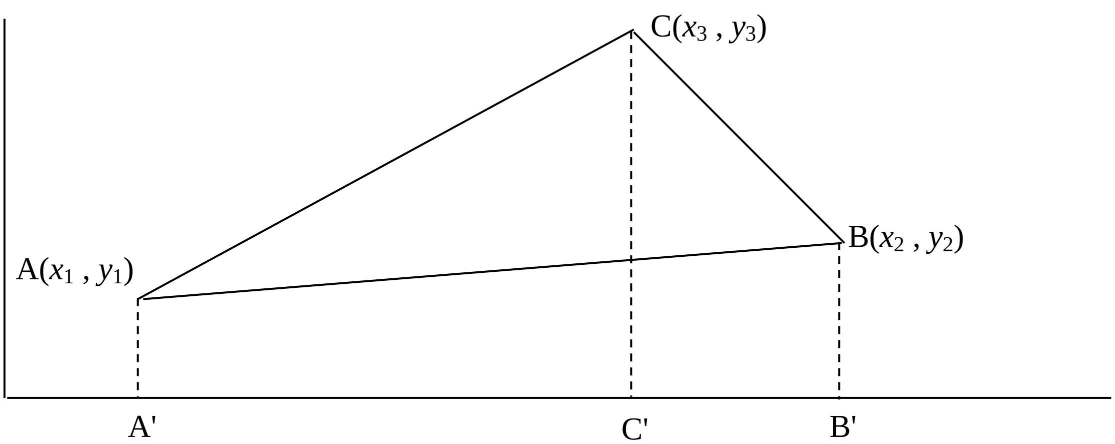


FIGURE II.2

We see that

area of triangle ABC = area of trapezium A'ACC' (see comment*)
+ area of trapezium C'CBB'
– area of trapezium A'ABB'.

$$= \tfrac{1}{2}(x_3 - x_1)(y_3 + y_1) + \tfrac{1}{2}(x_2 - x_3)(y_2 + y_3) - \tfrac{1}{2}(x_2 - x_1)(y_2 + y_1)$$

$$= \tfrac{1}{2}\left[x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)\right]$$

$$= \tfrac{1}{2}\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix} \qquad 2.2.12$$

* Since writing this section I have become aware of a difference in U.S./British usages of the word "trapezium". Apparently in British usage, "trapezium" means a quadrilateral with two parallel sides. In U.S. usage, a trapezium means a quadrilateral with no parallel sides, while a quadrilateral with two parallel sides is a "trapezoid". As with many words, either British or U.S. usages may be heard in Canada. In the above derivation, I intended the British usage. What is to be learned from this is that we must always take care to make ourselves clearly understood when using such ambiguous words, and not to assume that the reader will interpret them the way we intend.

The reader might like to work through an alternative method, using results that we have obtained earlier. The same result will be obtained. In case the algebra proves a little tedious, it may be found easier to work through a numerical example, such as: calculate the area of the triangle ABC, where A, B, C are the points (2,3), (7,4), (5,6) respectively. In the second method, we note that the area of a triangle is $\frac{1}{2}$ × base × height. Thus, if we can find the length of the side BC, and the perpendicular distance of A from BC, we can do it. The first is easy:

$$(\mathrm{BC})^2 = (x_3 - x_2)^2 + (y_3 - y_2)^2. \qquad 2.2.13$$

To find the second, we can easily write down the equation to the line BC from equation 2.2.3, and then re-write it in the form 2.2.7. Then equation 2.2.11 enables us to find the perpendicular distance of A from BC, and the rest is easy.

If the determinant in equation 2.2.12 is zero, the area of the triangle is zero. This means that the three points are collinear.

The angle between two lines

$$y = m_1 x + c_1 \qquad 2.2.14$$

and

$$y = m_2 x + c_2 \qquad 2.2.15$$

is easily found by recalling that the angles that they make with the x-axis are $\tan^{-1} m_1$ and $\tan^{-1} m_2$ together with the elementary trigonometry formula $\tan(A-B) = (\tan A - \tan B)/(1+\tan A \tan B)$. It is then clear that the tangent of the angle between the two lines is

$$\frac{m_2 - m_1}{1 + m_1 m_2}. \qquad 2.2.16$$

The two lines are at right angles to each other if

$$m_1 m_2 = -1 \qquad 2.2.17$$

The line that bisects the angle between the lines is the locus of points that are equidistant from the two lines. For example, consider the two lines

$$-2x + 5y = 1 \qquad 2.2.18$$

$$30x - 10y = 1 \qquad 2.2.19$$

Making use of equation 2.2.11, we see that a point (x, y) is equidistant from these two lines if

$$\frac{1+2x-5y}{\sqrt{29}} = \pm\frac{1-30x+10y}{\sqrt{1000}}. \qquad 2.2.20$$

The significance of the $\pm$ will become apparent shortly. The + and – choices result, respectively, in

$$-8.568x + 8.079y = 1 \qquad 2.2.21$$

and

$$2.656x + 2.817y = 1. \qquad 2.2.22$$

The two continuous lines in figure II.3 are the lines 2.2.18 and 2.2.19. There are two bisectors, represented by equations 2.2.21 and 2.2.22, shown as dotted lines in the figure, and they are at right angles to each other. The choice of the + sign in equation 2.2.20 (which in this case results in equation 2.2.21, the bisector in figure II.3 with the positive slope) gives the bisector of the sector that contains the origin.

An equation of the form

$$ax^2 + 2hxy + by^2 = 0 \qquad 2.2.23$$

can be factored into two linear factors with no constant term, and it therefore represents two lines intersecting at the origin. It is left as an exercise to determine the angles that the two lines make with each other and with the x axis, and to show that the lines

$$x^2 + \left(\frac{a-b}{h}\right)xy - y^2 = 0 \qquad 2.2.24$$

are the bisectors of 2.2.23 and are perpendicular to each other.

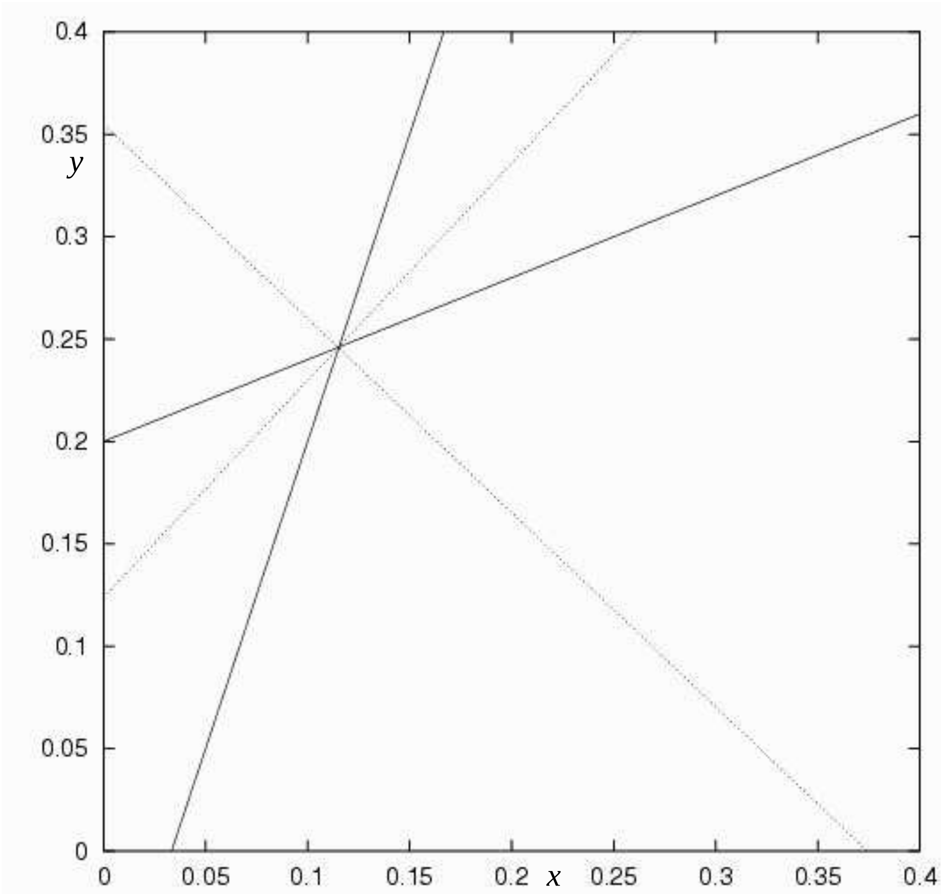


FIGURE II.3

Given the equations to three straight lines, can we find the area of the triangle bounded by them? To find a general algebraic expression might be a bit tedious, though the reader might like to try it, but a numerical example is straightforward. For example, consider the lines

$$x - 5y + 12 = 0, \qquad 2.2.25$$

$$3x + 4y - 9 = 0, \qquad 2.2.26$$

$$3x - y - 3 = 0. \qquad 2.2.27$$

By solving the equations in pairs, it is soon found that they intersect at the points (–0.15789, 2.36842), (1.4, 1.2) and (1.92857, 2.78571). Application of equation 2.2.12 then gives the area as 1.544. The triangle is drawn in figure II.4. Measure any side and the corresponding height with a ruler and see if the area is indeed about 1.54.

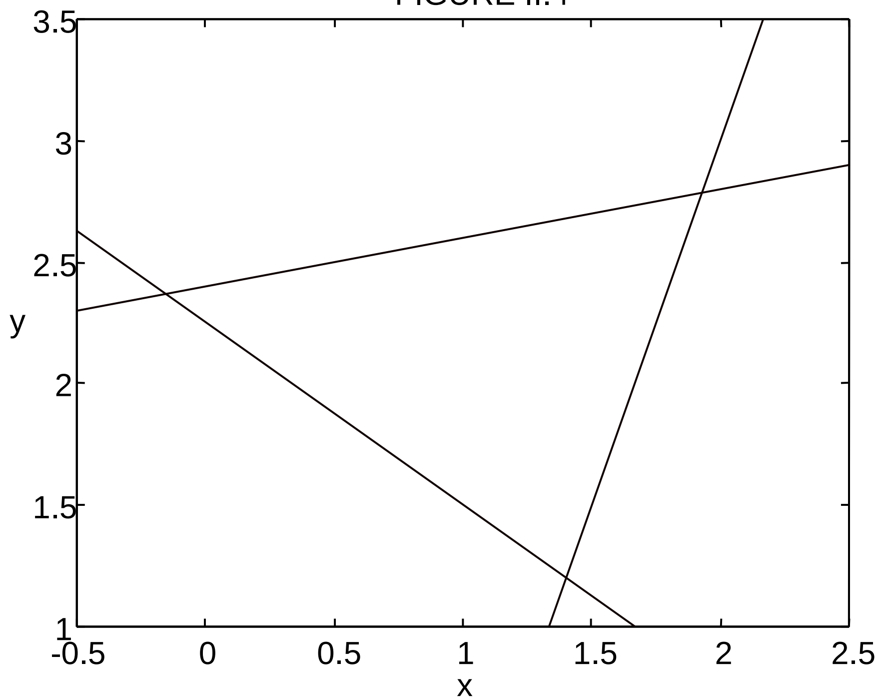
But now consider the three lines

$$x - 5y + 12 = 0, \qquad 2.2.28$$

$$3x + 4y - 9 = 0, \qquad 2.2.29$$

$$3x + 23y - 54 = 0. \qquad 2.2.30$$

FIGURE II.4



By solving the equations in pairs, it will be found that all three lines intersect at the same point (please do this), and the area of the triangle is, of course, zero. Any one of these equations is, in fact, a linear combination of the other two. You should draw these three lines accurately on graph paper (or by computer). In general, if three lines are

$$A_1x + B_1y + C_1 = 0 \qquad 2.2.31$$

$$A_2x + B_2y + C_2 = 0 \qquad 2.2.32$$

$$A_3x + B_3y + C_3 = 0 \qquad 2.2.33$$

they will be concurrent at a single point if

$$\begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} = 0 \,. \qquad 2.2.34$$

Thus the determinant in equation 2.2.12 provides a test of whether three points are collinear, and the determinant in equation 2.2.34 provides a test of whether three lines are concurrent.

Finally - at least for the present chapter - there may be rare occasion to write the equation of a straight line in polar coordinates. It should be evident from figure II.5 that the equations

$$r = p\csc(\theta - \alpha) \quad \text{or} \quad r = p\csc(\alpha - \theta) \qquad 2.2.35$$

describe a straight line passing at a distance p from the pole and making an angle α with the initial line. If $p = 0$, the polar equation is merely $\theta = \alpha$.

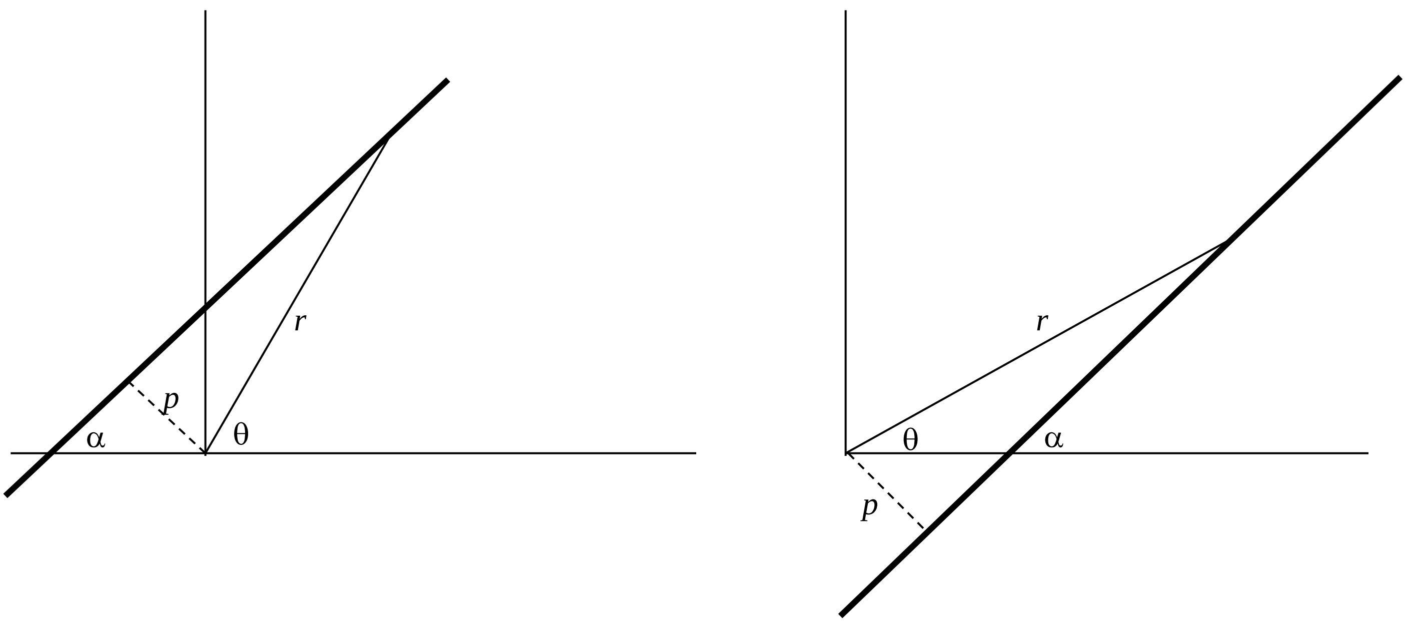


FIGURE II.5

2.3 *The Ellipse*

An ellipse is a figure that can be drawn by sticking two pins in a sheet of paper, tying a length of string to the pins, stretching the string taut with a pencil, and drawing the figure that results. During this process, the sum of the two distances from pencil to one pin and from pencil to the other pin remains constant and equal to the length of the string. This method of drawing an ellipse provides us with a formal definition, which we shall adopt in this chapter, of an ellipse, namely:

An ellipse is the locus of a point that moves such that the sum of its distances from two fixed points called the foci is constant (see figure II.6).

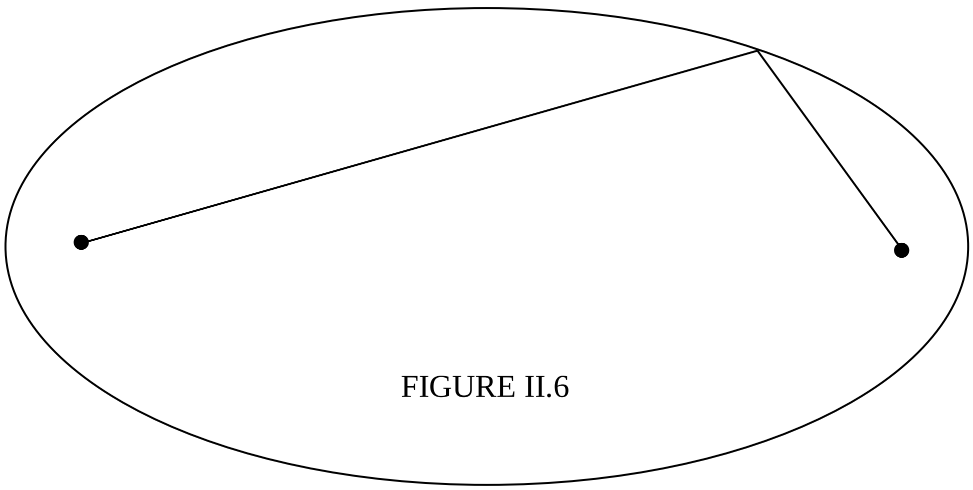


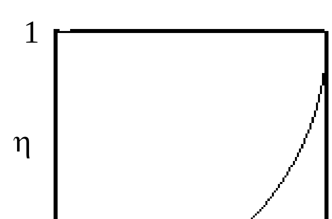
FIGURE II.6

We shall call the sum of these two distances (i.e the length of the string) 2a. The ratio of the distance between the foci to length of the string is called the *eccentricity* e of the ellipse, so that the distance between the foci is 2ae, and e is a number between 0 and 1.

The longest axis of the ellipse is its major axis, and a little bit of thought will show that its length is equal to the length of the string; that is, 2a. The shortest axis is the minor axis, and its length is usually denoted by 2b. The eccentricity is related to the ratio b/a in a manner that we shall shortly discuss.

The ratio $\eta = (a - b)/a$ is called the *ellipticity* if the ellipse. It is merely an alternative measure of the noncircularity. It is related to the eccentricity, and we shall obtain that relation shortly, too. Until then, Figure II.7 shows pictorially the relation between the two.

FIGURE II.7



We shall use our definition of an ellipse to obtain its equation in rectangular coordinates. We shall place the two foci on the x-axis at coordinates ($-ae$, 0) and (ae, 0) (see figure II.8).

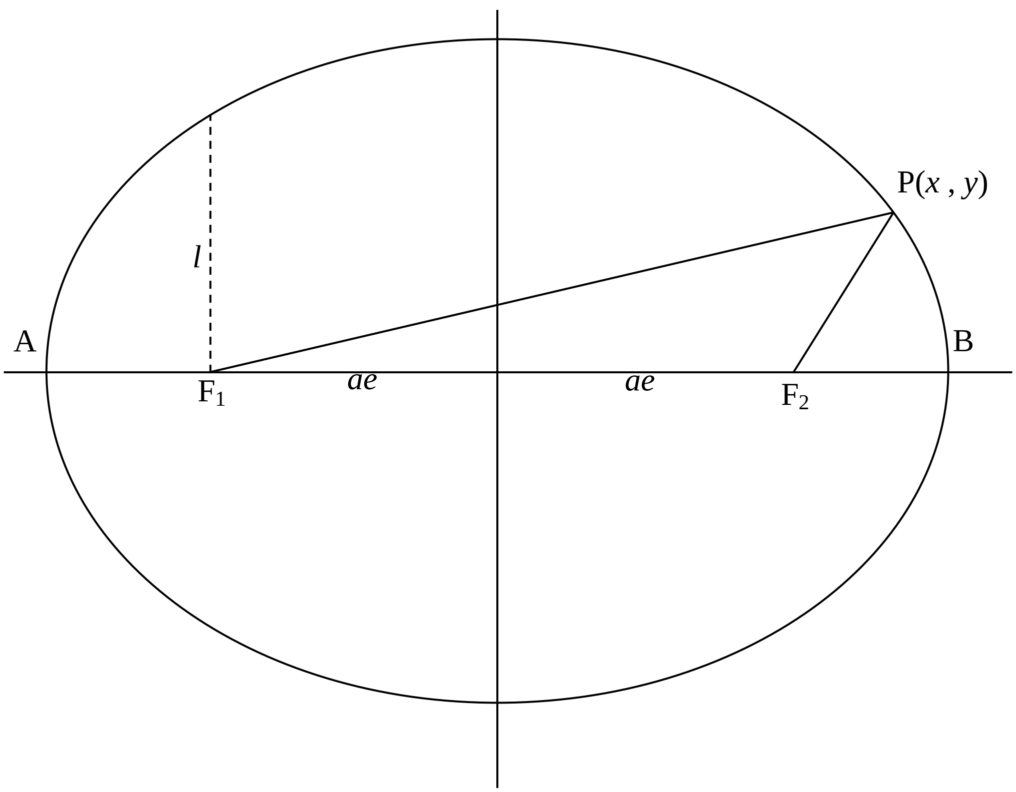


FIGURE II.8

The definition requires that $PF_1 + PF_2 = 2a$. That is:

$$\left[(x+ae)^2+y^2\right]^{\frac{1}{2}}+\left[(x-ae)^2+y^2\right]^{\frac{1}{2}}=2a, \qquad 2.3.1$$

and this is the equation to the ellipse. The reader should be able, after a little bit of slightly awkward algebra, to show that this can be written more conveniently as

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1. \qquad 2.3.2$$

By putting x = 0, it is seen that the ellipse intersects the y-axis at $\pm a\sqrt{1-e^2}$ and therefore that $a\sqrt{1-e^2}$ is equal to the semi minor axis b. Thus we have the familiar equation to the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad 2.3.3$$

as well as the important relation between a, b and e:

$$b^2 = a^2 (1 - e^2) \qquad 2.3.4$$

The reader can also now derive the relation between ellipticity η and eccentricity e:

$$\eta = 1 - \sqrt{(1-e^2)}. \qquad 2.3.5$$

This can also be written

$$e^2 = \sqrt{\eta(2-\eta)} \qquad 2.3.6$$

or

$$e^2 + (\eta - 1)^2 = 1. \qquad 2.3.7$$

This shows, incidentally, that the graph of η versus e, which we have drawn in figure II.7, is part of a circle of radius 1 centred at e = 0, η = 1.

In figures II.9 I have drawn ellipses of eccentricities 0.1 to 0.9 in steps of 0.1, and in figure II.10 I have drawn ellipses of ellipticities 0.1 to 0.9 in steps of 0.1. You may find that ellipticity gives you a better idea than eccentricity of the noncircularity of an ellipse. For an exercise, you should draw in the positions of the foci of each of these ellipses, and decide whether eccentricity or ellipticity gives you a better idea of the "ex-centricity" of the foci. Note that the eccentricities of the orbits of Mars and Mercury are, respectively, about 0.1 and 0.2 (these are the most eccentric of the planetary orbits except for comet-like Pluto), and it is difficult for the eye to see that they depart at all from circles - although, when the foci are drawn, it is obvious that the foci are "ex-centric".

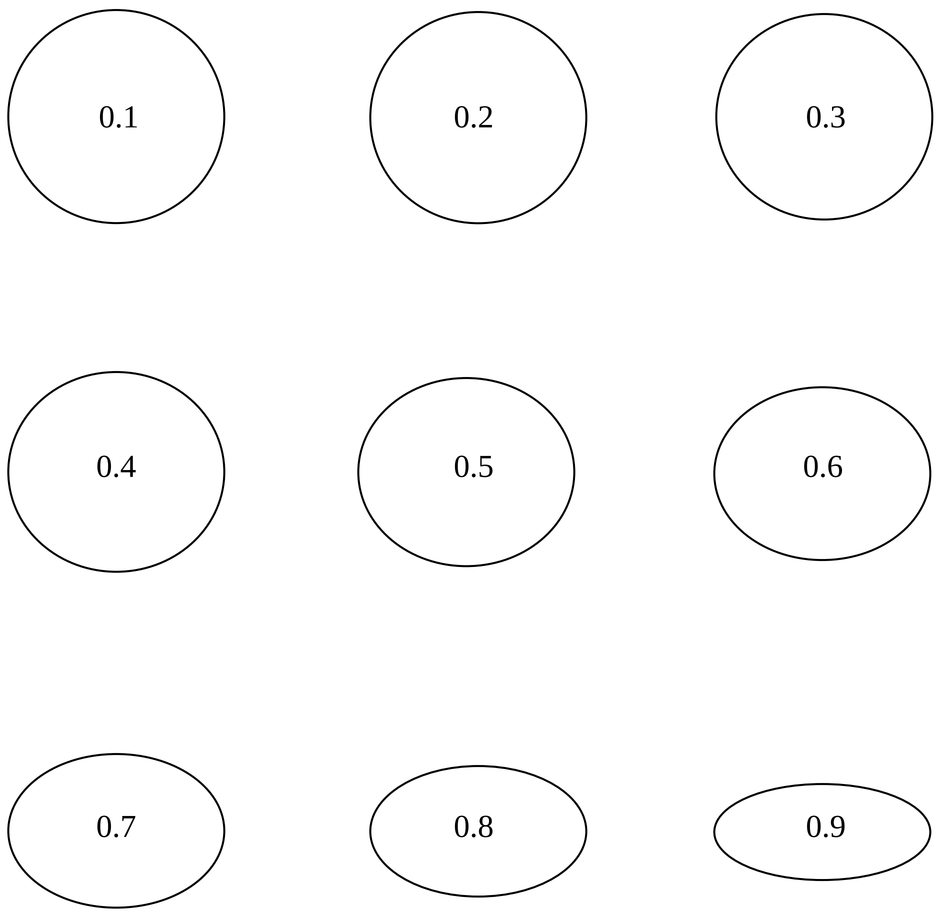


FIGURE II.9
The number inside each ellipse is its eccentricity.

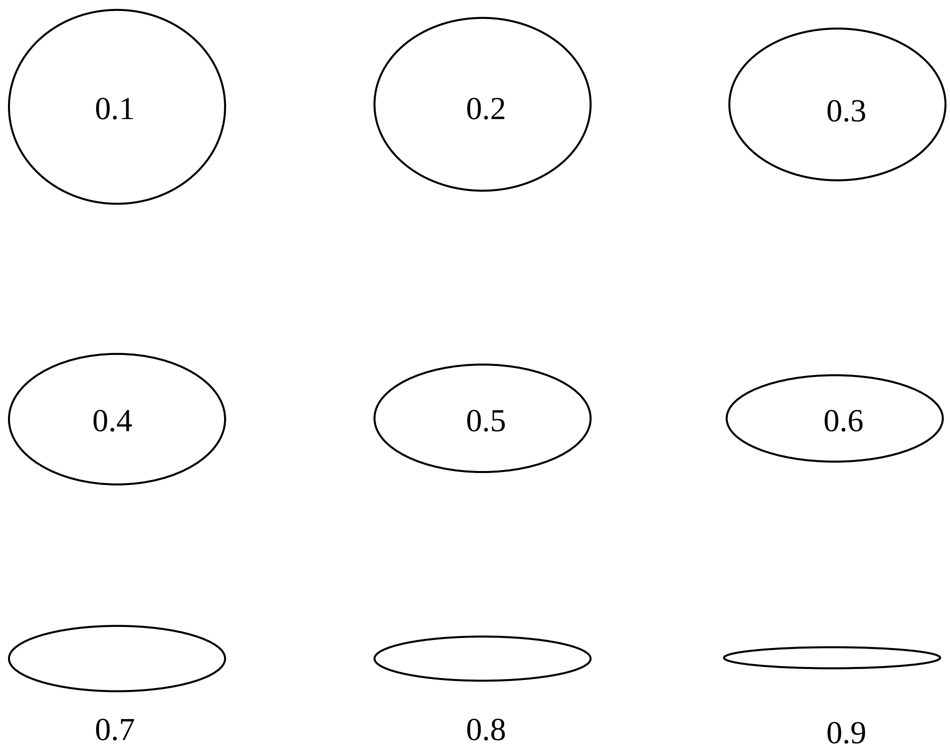
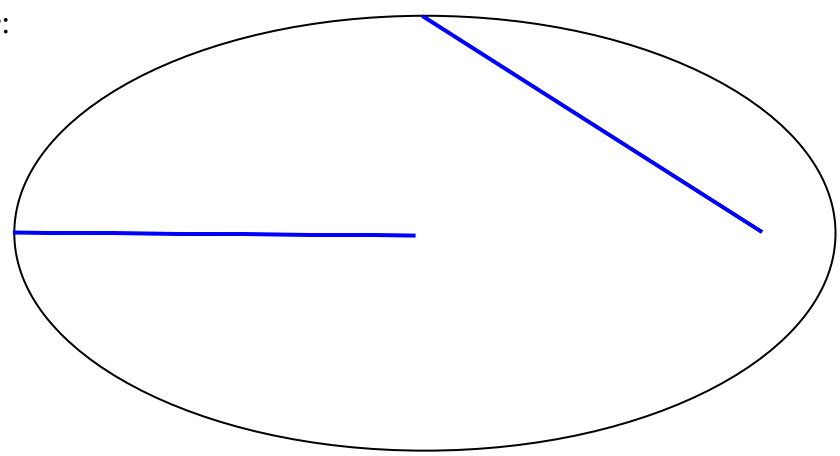


FIGURE II.10
The figure inside or below each ellipse is its ellipticity.

Short exercise:



Show that the distance between an end of the minor axis and a focus is equal to the length of the semi major axis.

In the theory of planetary orbits, the Sun will be at one focus. Let us suppose it to be at F_2 (see figure II.8). In that case the distance F_2 B is the perihelion distance q, and is equal to

$$q = a(1-e). \tag{2.3.8}$$

The distance F_2 A is the aphelion distance Q (pronounced ap-helion by some and affelion by others – and both have defensible positions), and it is equal to

$$Q = a(1+e). \tag{2.3.9}$$

A line parallel to the minor axis and passing through a focus is called a *latus rectum* (plural: *latera recta*). The length of a semi latus rectum is commonly denoted by l (sometimes by p). Its length is obtained by putting $x = ae$ in the equation to the ellipse, and it will be readily found that

$$l = a(1-e^2). \tag{2.3.10}$$

The length of the semi latus rectum is an important quantity in orbit theory. It will be found, for example, that the energy of a planet is closely related to the semi major axis a of its orbit, while its angular momentum is closely related to the semi latus rectum.

The circle whose diameter is the major axis of the ellipse is called the *eccentric circle* or, preferably, the *auxiliary circle* (figure II.11). Its equation is

$$x^2 + y^2 = a^2. \tag{2.3.11}$$

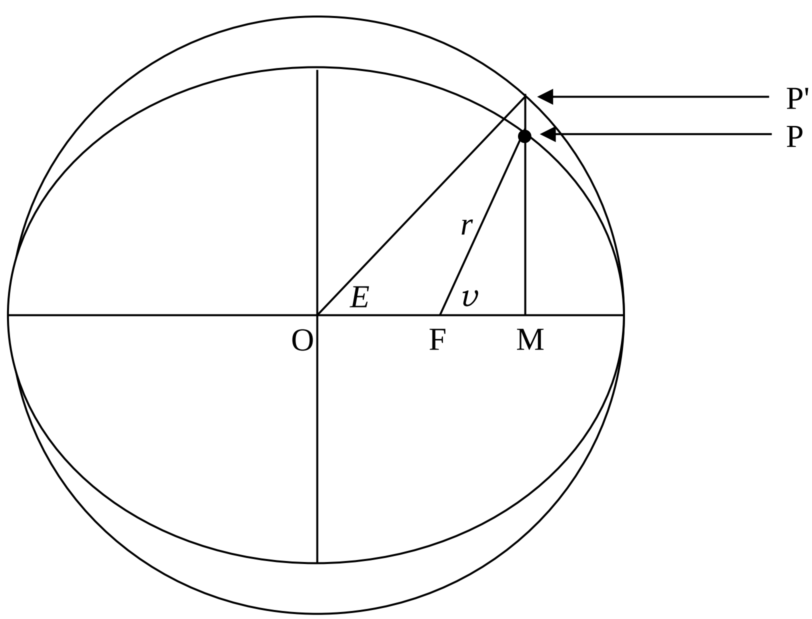


FIGURE II.11

In orbit theory the angle υ (denoted by f by some authors) is called the *true anomaly* of a planet in its orbit. The angle E is called the *eccentric anomaly*, and it is important to find a relation between them.

We first note that, if the eccentric anomaly is E, the abscissas of P' and of P are each $a\cos E$. The ordinate of P' is $a\sin E$. By putting $x = a\cos E$ in the equation to the ellipse, we immediately find that the ordinate of P is $b \sin E$. Several deductions follow. One is that any point whose abscissa and ordinate are of the form

$$x = a\cos E, \qquad y = b\sin E \qquad 2.3.12$$

is on an ellipse of semi major axis a and semi minor axis b. These two equations can be regarded as parametric equations to the ellipse. They can be used to describe an ellipse just as readily as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad 2.3.13$$

and indeed this equation is the E-eliminant of the parametric equations.

The ratio PM/P'M for any line perpendicular to the major axis is b/a. Consequently the area of the ellipse is b/a times the area of the auxiliary circle; and since the area of the auxiliary circle is πa^2, it follows that the area of the ellipse is πab.

In figure II.11, the distance r is called the *radius vector* (plural *radii vectores*), and from the theorem of Pythagoras its length is given by

$$r^2 = b^2 \sin^2 E + a^2 (\cos E - e)^2. \qquad 2.3.14$$

On substituting $1 - \cos^2 E$ for $\sin^2 E$ and $a^2 (1 - e^2)$ for b^2, we soon find that

$$r = a(1 - e \cos E) \qquad 2.3.15$$

It then follows immediately that the desired relation between υ and E is

$$\cos\upsilon = \frac{\cos E - e}{1 - e\cos E}. \qquad 2.3.16$$

From trigonometric identities, this can also be written

$$\sin\upsilon = \frac{\sqrt{1-e^2}\,\sin E}{1 - e\cos E} \qquad 2.3.17a$$

or $$\tan\upsilon = \frac{\sqrt{1-e^2}\,\sin E}{\cos E - e} \qquad 2.3.17\text{b}$$

or $$\tan\tfrac{1}{2}\upsilon = \sqrt{\frac{1+e}{1-e}}\tan\tfrac{1}{2}E. \qquad 2.3.17\text{c}$$

The inverse formulas may also be useful:

$$\cos E = \frac{e + \cos\upsilon}{1 + e\cos\upsilon} \qquad 2.3.17\text{d}$$

$$\sin E = \frac{\sin\upsilon\sqrt{1-e^2}}{1 + e\cos\upsilon} \qquad 2.3.17\text{e}$$

$$\tan E = \frac{\sin\upsilon\sqrt{1-e^2}}{e + \cos\upsilon} \qquad 2.3.17\text{f}$$

$$\tan\tfrac{1}{2}E = \sqrt{\frac{1-e}{1+e}}\tan\tfrac{1}{2}\upsilon. \qquad 2.3.17\text{g}$$

There are a number of miscellaneous geometric properties of an ellipse, some, but not necessarily all, of which may prove to be of use in orbital calculations. We describe some of them in what follows.

Tangents to an ellipse.

Find where the straight line $y = mx + c$ intersects the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \qquad 2.3.18$$

The answer to this question is to be found by substituting $mx + c$ for y in the equation to the ellipse. After some rearrangement, a quadratic equation in x results:

$$\left(a^2m^2 + b^2\right)x^2 + 2a^2cmx + a^2\left(c^2 - b^2\right) = 0. \qquad 2.3.19$$

If this equation has two real roots, the roots are the x-coordinates of the two points where the line intersects the ellipse. If it has no real roots, the line misses the ellipse. If it has two coincident real roots, the line is tangent to the ellipse. The condition for this is found by setting the discriminant of the quadratic equation to zero, from which it is found that

$$c^2 = a^2m^2 + b^2. \qquad 2.3.20$$

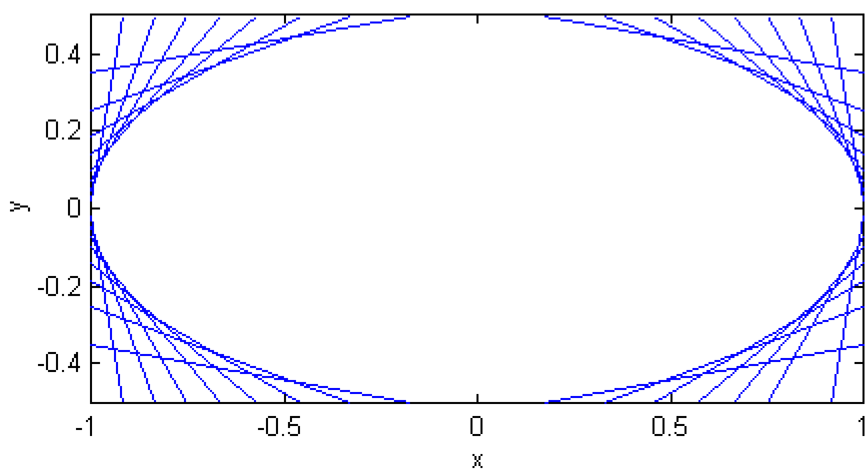
Thus a straight line of the form

$$y = mx \pm \sqrt{a^2m^2 + b^2} \qquad 2.3.21$$

is tangent to the ellipse.

Figure II.12 shows several such lines, for $a = 2b$ and slopes ($\tan^{-1} m$) of 0° to 180° in steps of 10°

FIGURE II.12



Director Circle.

The equation we have just derived for a tangent to the ellipse can be rearranged to read

$$m^2\left(a^2 - x^2\right) + 2mx + b^2 - y^2 = 0. \qquad 2.3.22$$

Now the product of the slopes of two lines that are at right angles to each other is -1 (equation 2.2.17). Therefore, if we replace m in the above equation by $-1/m$ we shall obtain another tangent to the ellipse, at right angles to the first one. The equation to this second tangent becomes (after multiplication throughout by m)

$$m^2\left(b^2 - y^2\right) - 2mx + a^2 - x^2 = 0. \qquad 2.3.23$$

If we eliminate m from these two equations, we shall obtain an equation in x and y that describes the point where the two perpendicular tangents meet; that is, the equation that will describe a curve that is the locus of the point of intersection of two perpendicular tangents. It turns out that this curve is a circle of radius $\sqrt{a^2 + b^2}$, and it is called the *director circle.*

It is easier than it might first appear to eliminate m from the equations. We merely have to add the equations 2.3.22 and 2.3.23:

$$m^2\left(a^2+b^2-x^2-y^2\right)+a^2+b^2-x^2-y^2=0. \qquad 2.3.24$$

For real *m*, this can only be if

$$x^2+y^2=a^2+b^2, \qquad 2.3.25$$

which is the required locus of the director circle of radius $\sqrt{a^2+b^2}$. It is illustrated in figure II.13.

We shall now derive an equation to the line that is tangent to the ellipse at the point (x_1, y_1).

Let $(x_1, y_1) = (a\cos E_1, b\sin E_1)$ and $(x_2, y_2) = (a\cos E_2, b\sin E_2)$ be two points on the ellipse.

The line joining these two points is

$$\frac{y-b\sin E_1}{x-a\cos E_1}=\frac{b\left(\sin E_2-\sin E_1\right)}{a\left(\cos E_2-\cos E_1\right)}$$

$$=\frac{2b\cos\frac{1}{2}\left(E_2+E_1\right)\sin\frac{1}{2}\left(E_2-E_1\right)}{-2a\sin\frac{1}{2}\left(E_2+E_1\right)\sin\frac{1}{2}\left(E_2-E_1\right)}=-\frac{b\cos\frac{1}{2}\left(E_2+E_1\right)}{a\sin\frac{1}{2}\left(E_2+E_1\right)}. \qquad 2.3.26$$

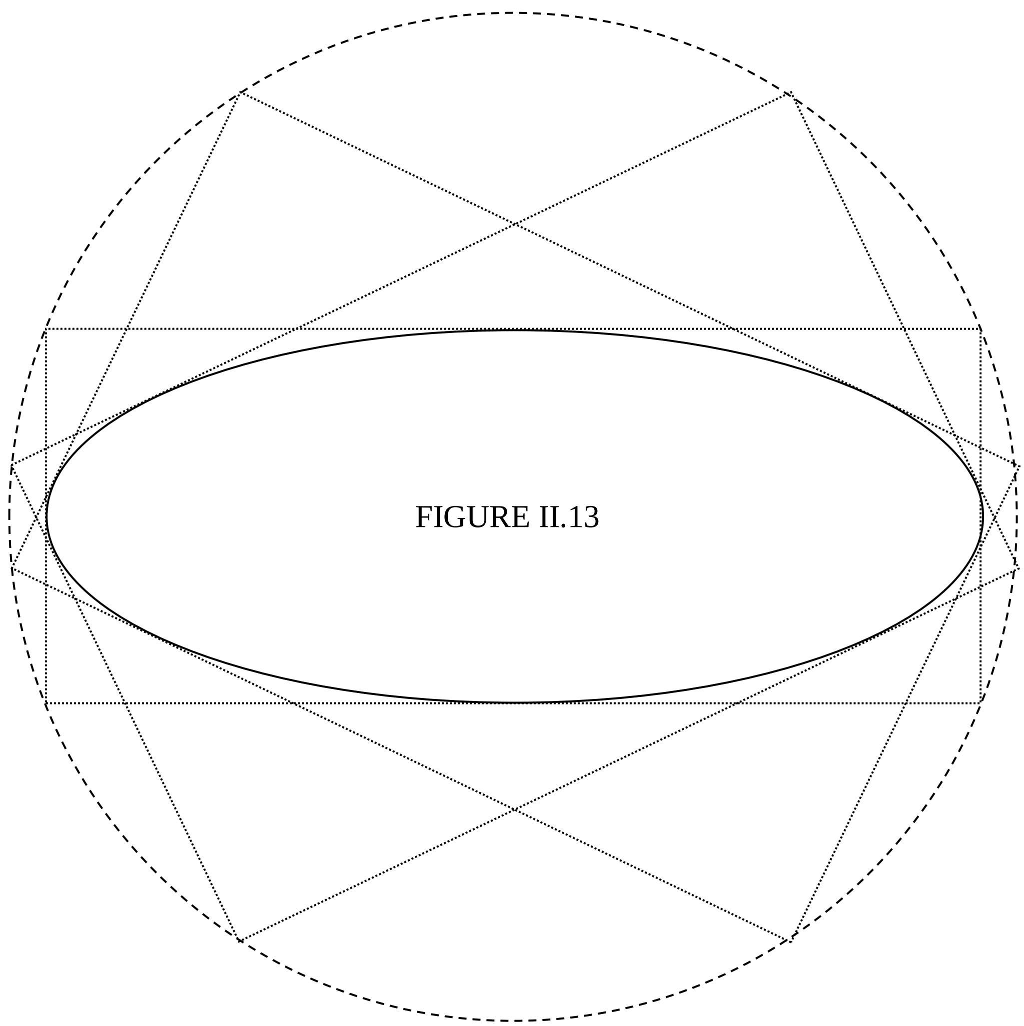


FIGURE II.13

Now let E_2 approach E_1, eventually coinciding with it. The resulting equation

$$\frac{y - b\sin E}{x - a\cos E} = -\frac{b\cos E}{a\sin E}, \qquad 2.3.27$$

in which we no longer distinguish between E_1 and E_2, is the equation of the straight line that is tangent to the ellipse at (a cos E , b sin E). This can be written

$$\frac{x\cos E}{a} + \frac{y\sin E}{b} = 1 \qquad 2.3.28$$

or, in terms of (x_1 , y_1),

$$\frac{x_1 x}{a^2} + \frac{y_1 y}{b^2} = 1, \qquad 2.3.29$$

which is the tangent to the ellipse at (x_1 , y_1).

An interesting property of a tangent to an ellipse, the proof of which I leave to the reader, is that F_1P and F_2P make equal angles with the tangent at P. If the inside of the ellipse were a reflecting mirror and a point source of light were to be placed at F_1 , it would be imaged at F_2 . (Have a look at figure II.6 or II.8.) This has had an interesting medical application. A patient has a kidney stone. The patient is asked to lie in an elliptical bath, with the kidney stone at F_2 . A small explosion is detonated at F_1; the explosive sound wave emanating from F_1 is focused as an implosion at F_2 and the kidney stone at F_2 is shattered. Don't try this at home.

Directrices.

The two lines $x = \pm a/e$ are called the *directrices* (singular *directrix*) of the ellipse (figure II.14).

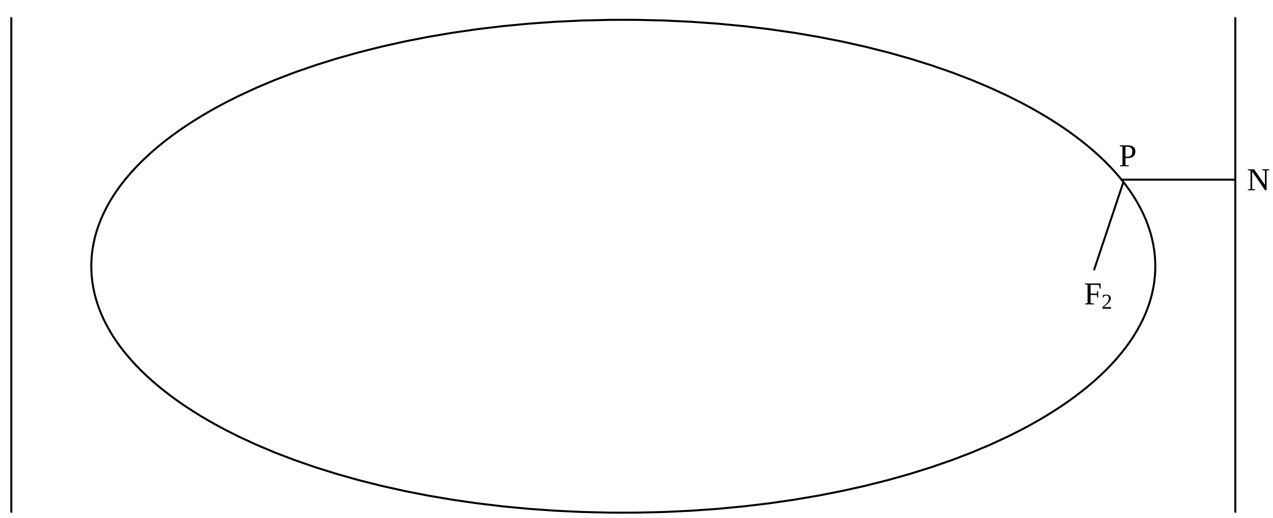


FIGURE II.14

The ellipse has the property that, for any point P on the ellipse, the ratio of the distance PF_2 to a focus to the distance PN to a directrix is constant and is equal to the eccentricity of the ellipse. Indeed, this property is sometimes used as the definition of an ellipse, and all the equations and properties that we have so far derived can be deduced from such a definition. We, however, adopted a different definition, and the focus-directrix property must be derived. This is straightforward, for, (recalling that the abcissa of F_2 is *ae*) we see from figure II.14 that the square of the desired ratio is

$$\frac{(x-ae)^2+y^2}{(a/e-x)^2}. \qquad 2.3.30$$

On substitution of $$b^2\left(1-\left(\tfrac{x}{a}\right)^2\right)=a^2\left(1-e^2\right)\left(1-\left(\tfrac{x}{a}\right)^2\right)=\left(1-e^2\right)\left(a^2-x^2\right) \qquad 2.3.31$$

for y^2, the above expression is seen to reduce to e^2.

Another interesting property of the focus and directrix, although a property probably with not much application to orbit theory, is that if the tangent to an ellipse at a point P intersects the directrix at Q, then P and Q subtend a right angle at the focus. (See figure II.15).

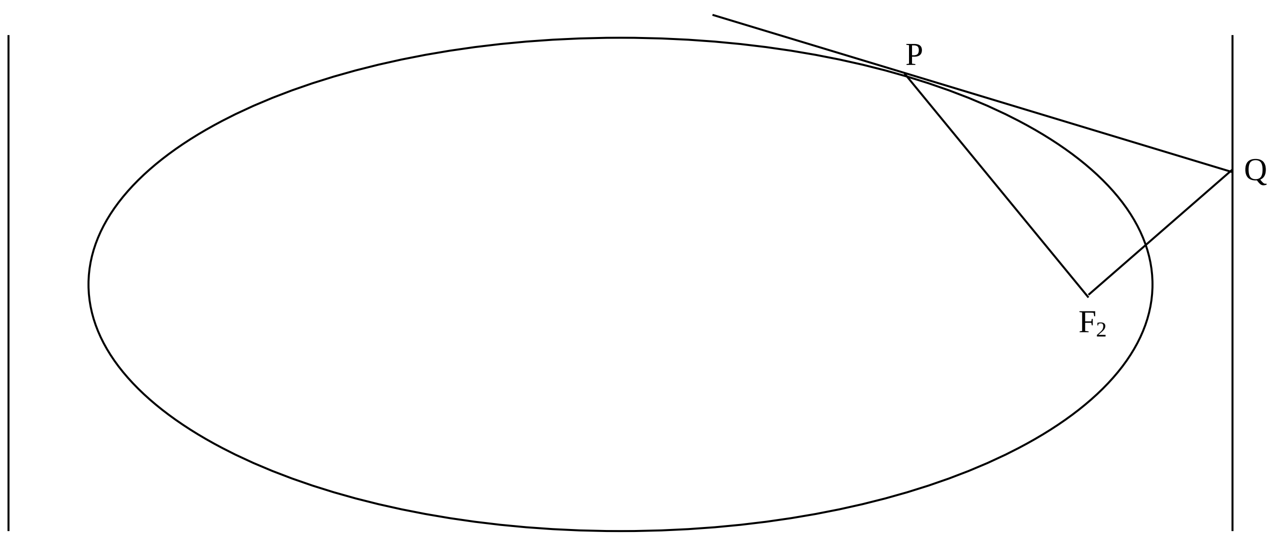


FIGURE II.15

Thus the tangent at P = (x_1 , y_1) is

$$\frac{x_1x}{a^2}+\frac{y_1y}{b^2}=1 \qquad 2.3.32$$

and it is straightforward to show that it intersects the directrix $x = a/e$ at the point

$$\left(\frac{a}{e} \quad , \quad \frac{b^2}{y_1}\left(1-\frac{x_1}{ae}\right)\right).$$

The coordinates of the focus F_2 are $(ae, 0)$. The slope of the line PF_2 is $(x_1 - ae)/y_1$ and the slope of the line QF_2 is

$$\frac{\frac{b^2}{y_1}\left(1-\frac{x_1}{ae}\right)}{\frac{a}{e} - ae}.$$

It is easy to show that the product of these two slopes is -1, and hence that PF_2 and QF_2 are at right angles.

Conjugate Diameters.

The left hand of figure II.16 shows a circle and two perpendicular diameters. The right hand figure shows what the circle would look like when viewed at some oblique angle. The circle has become an ellipse, and the diameters are no longer perpendicular. The diameters are called *conjugate diameters* of the ellipse. One is conjugate to the other, and the other is conjugate to the one. They have the property - or the definition - that each bisects all chords parallel to the other, because this property of bisection, which is obviously held by the perpendicular diameters of the circle, is unaltered in projection.

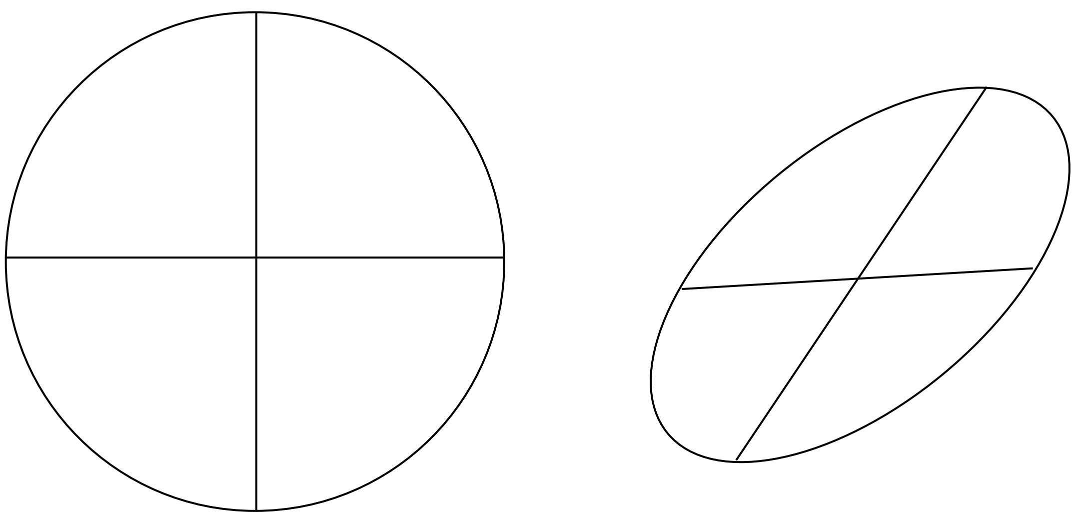


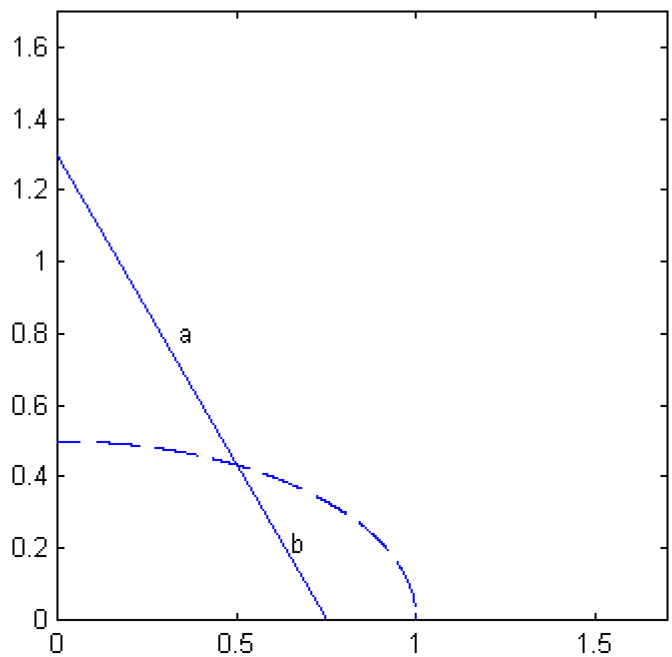
FIGURE II.16

It is easy to draw two conjugate diameters of an ellipse of eccentricity e either by making use of this last-mentioned property or by noting that that the product of the slopes of two conjugate diameters is $e^2 - 1$. The proof of this is left for the enjoyment of the reader.

A Ladder Problem.

No book on elementary applied mathematics is complete without a ladder problem. A ladder of length $a + b$ leans against a smooth vertical wall and a smooth horizontal floor. A particular rung is at a distance a from the top of the ladder and b from the bottom of the ladder. Show that, when the ladder slips, the rung describes an ellipse. (This result will suggest another way of drawing an ellipse.) See figure II.17.

FIGURE II.17



If you have not done this problem after one minute, here is a hint. Let the angle that the ladder makes with the floor at any time be E. That is the end of the hint.

The reader may be aware that some of the geometrical properties that we have discussed in the last few paragraphs are more of recreational interest and may not have much direct application in the theory of orbits. In the next subsection we return to properties and equations that are very relevant to orbital theory - perhaps the most important of all for the orbit computer to understand.

Polar Equation to the Ellipse.

We shall obtain the equation in polar coordinates to an ellipse whose focus is the pole of the polar coordinates and whose major axis is the initial line ($\theta = 0^\circ$) of the polar coordinates. In figure II.18 we have indicated the angle θ of polar coordinates, and it may occur to the reader that we have previously used the symbol υ for this angle and called it the true anomaly. Indeed at present, υ and θ are identical, but a little later we shall distinguish between them.

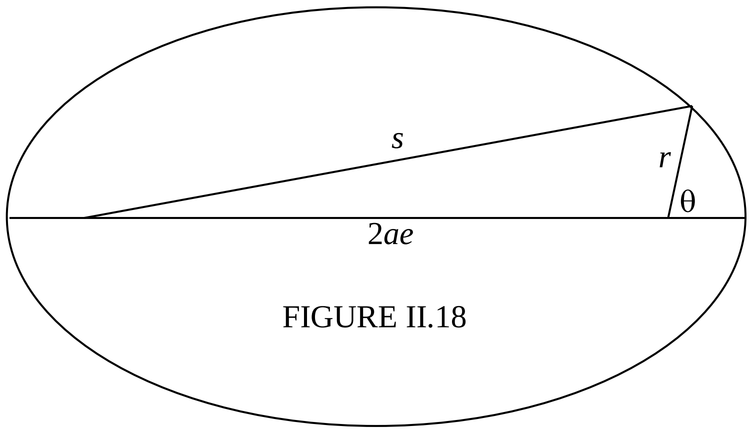


FIGURE II.18

From our definition of the ellipse, $s = 2a - r$, and so

$$s^2 = 4a^2 - 4ar + r^2. \tag{2.3.33}$$

From the cosine formula for a plane triangle,

$$s^2 = 4a^2e^2 + r^2 + 4aer\cos\theta. \tag{2.3.34}$$

On equating these expressions we soon obtain

$$a(1 - e^2) = r(1 + e\cos\theta). \tag{2.3.35}$$

The left hand side is equal to the semi latus rectum l, and so we arrive at the polar equation to the ellipse, focus as pole, major axis as initial line:

$$r = \frac{l}{1 + e\cos\theta}. \tag{2.3.36}$$

If the major axis is inclined at an angle ω to the initial line (figure II.19), the equation becomes

$$r = \frac{l}{1 + e\cos(\theta - \omega)} = \frac{l}{1 + e\cos\upsilon}. \tag{2.3.37}$$

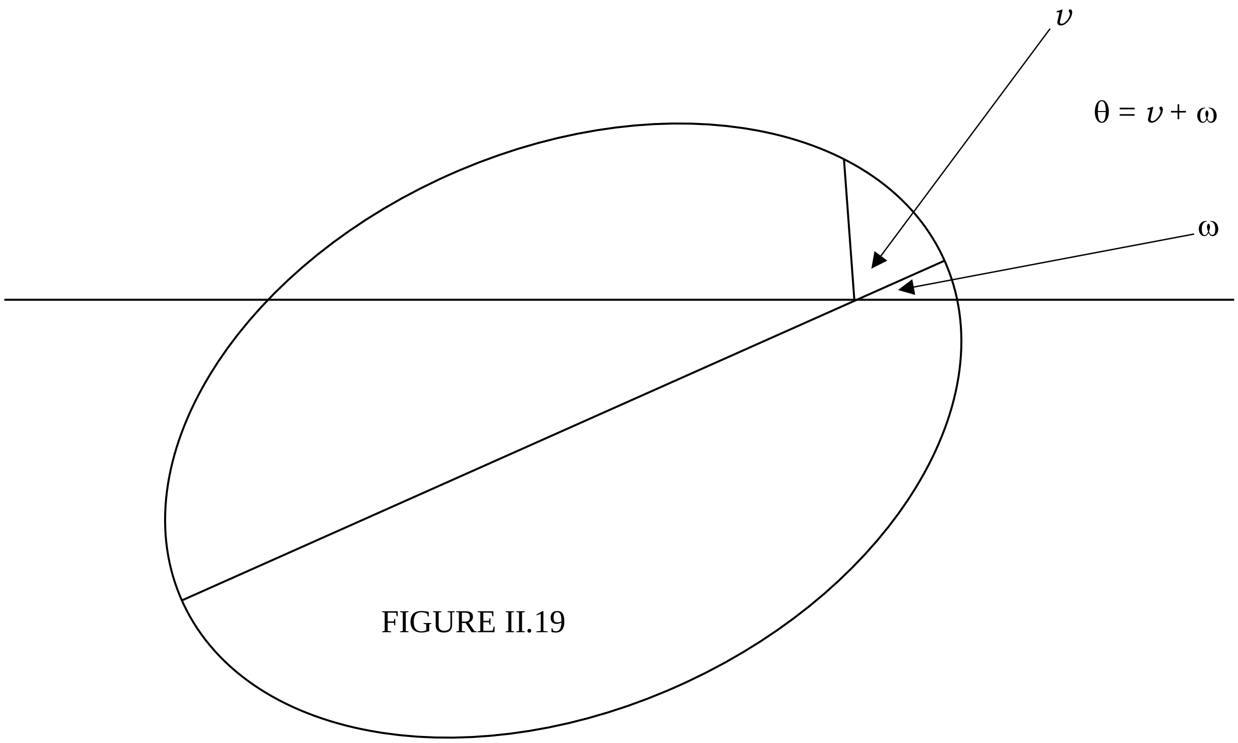


FIGURE II.19

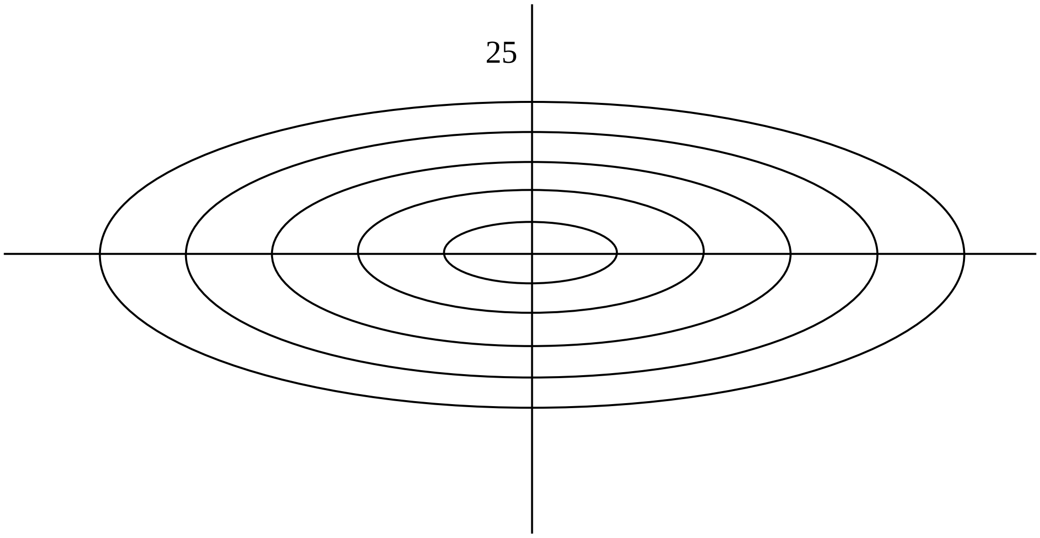
The distinction between θ and υ is now evident. θ is the angle of polar coordinates, ω is the angle between the major axis and the initial line (ω will be referred to in orbital theory as the "argument of perihelion"), and υ, the true anomaly, is the angle between the radius vector and the initial line.

Similar and confocal ellipses

A family of *similar ellipses* is not the same as a family of *confocal ellipses*. Here is a family of similar ellipses. The equation to the family is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = k^2. \tag{2.3.38}$$

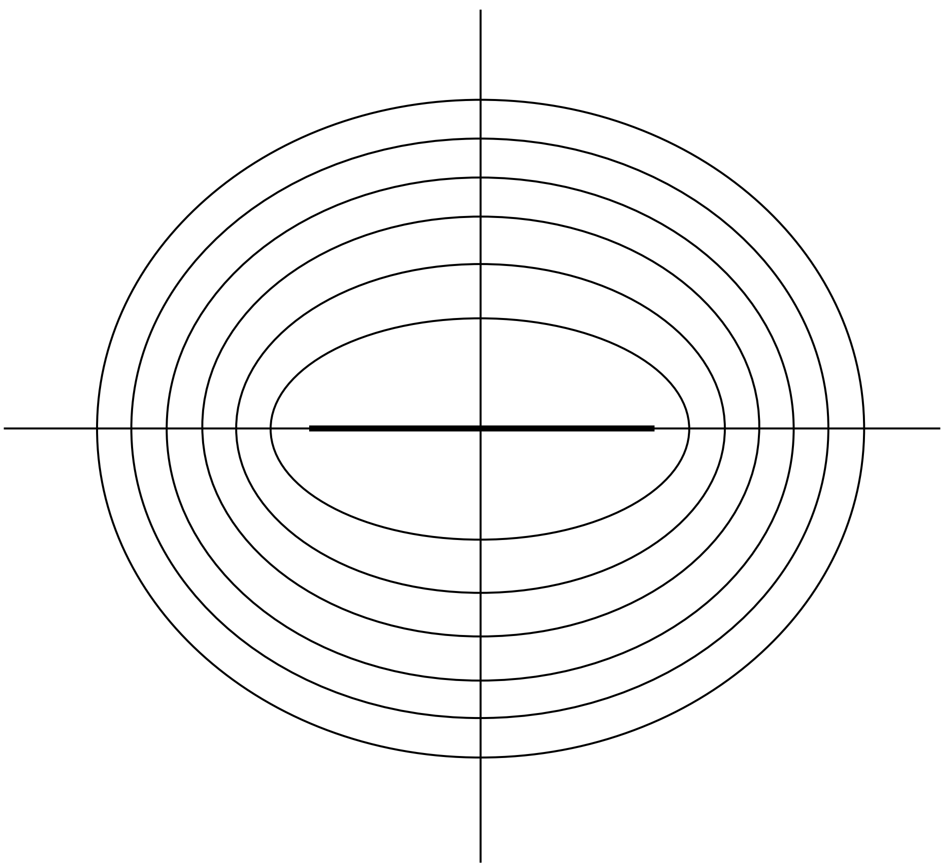
The parameter k is a dimensionless number, and the lengths of the semi major and semi minor axes are ka and kb respectively.



Here is a family of confocal ellipses. The foci are at the end of the bar. The equation to the family is

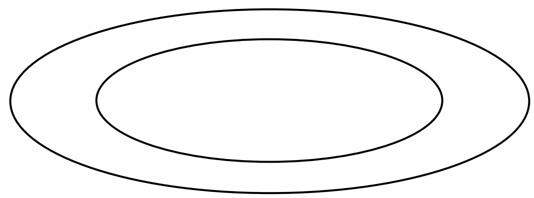
$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - f^2} = 1.$$

The constant distance between the two foci is $2f$. The parameter a is the semi major axis of each ellipse.

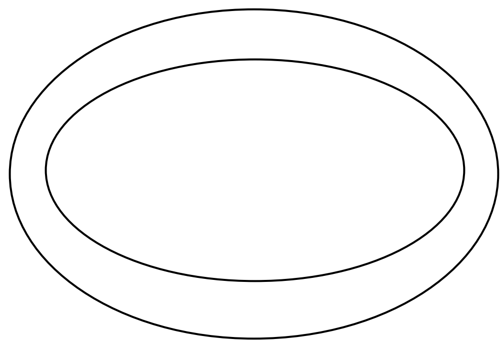


Annuli drawn between similar ellipses, or between confocal ellipses, are not of uniform thickness:

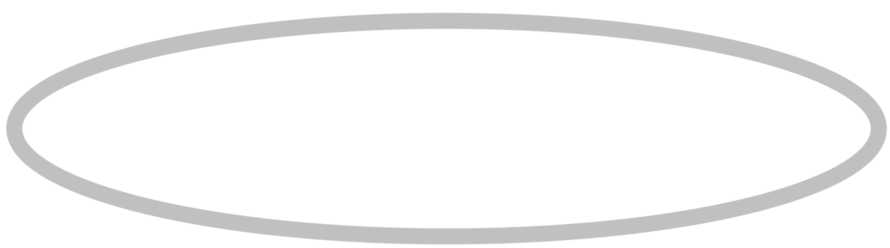
Similar



Confocal



I don’t believe that you can draw an annulus of constant thickness in which both the inner and outer edges are true ellipses. I think I can probably prove this, if I can find the time. The figure here:



looks like a thick elliptical annulus of uniform thickness, but I suspect that neither the inner nor the outer edge is a true ellipse. A rugger ball is presumably of uniform thickness, but I think that one or both of the inner and outer surfaces are not exactly prolate spheroids. (A North American “football” is too pointy to be called a spheroid.)

2.4 *The Parabola*

The equation y = constant times x represents a straight line. The equation y = constant times x^2, on the other hand, represents a *parabola*. And so, of course, does the equation y^2 = constant times x. For convenience in introductory accounts the constant in the latter case is often written as $4a$. In an astronomical context, however, the constant is often written as $4q$, which is what we shall do here. Thus, we shall write the equation as $y^2 = 4qx$. For example, in the case of a comet moving around the Sun in a parabolic orbit, q is the perihelion distance of the comet.

The equation

$$y = 4qx \qquad 2.4.0\text{i}$$

represents a parabola with its vertex at the origin of coordinates, and its axis coincident with the x-axis.

The point F (q, 0), which is on the axis of the parabola and at a distance q from the vertex, is called the *focus* of the parabola. The line perpendicular to the axis and passing through the focus is called the *latus rectum* of the parabola. It should be easy to see from equation 2.4.0i that the y coordinates of the ends of the latus rectum (where $x = q$) are at $y = \pm 2q$, and consequently the length of the latus rectum is $4q$. Half of this length is the *semi latus rectum*. It is usually given the symbol l¸ and of course $l = 2q$.

A parabola has only one focus, and consequently has only one latus rectum. An ellipse or a hyperbola, has two *foci*, and consequently two *latera recta*.

The word "parabola" is not Latin, and it does not take a Latin plural "parabolae". It is derived from Greek, but is regarded as an established English word, and therefore takes the regular English plural "parabolas". The same is true of "hyperbola" and "hyperbolas".

The equation $x^2 + y^2 = a^2$ represents a circle of radius a. All circles, or course, have the same shape, but they have different sizes, or radii. Likewise the equation $y^2 = 4qx$ represents a parabolla of latus rectum $4q$. All parabola (of course ?!) have the same shape, but they have different sizes, or latera recta. I draw three parabolas below, with q = 1, 2 and three, and I leave it to the reader to think about whther or not he agrees with me that they all have the same shape, or, if they di, in what sense are they the same shape?

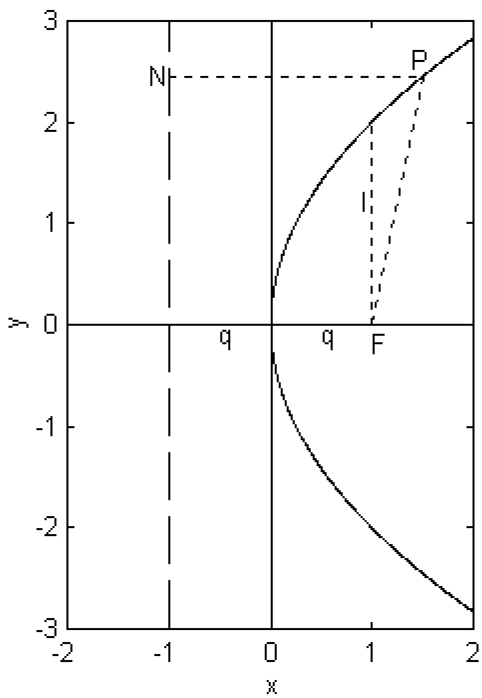
It is of interest to calculate the distance of a point P $(x, \sqrt{4qx})$ on a parabola to its focus F $(q, 0)$. The square of this distance is $d^2 = (x - q)^2 + 4qx = (x - q)^2$. That is $d = x + q$. That is to say, the distance PF is independent of y and is equal to the distance of P from the line $x = -q$. This line is called the *directrix* of the parabola.

Indeed some writers like to *define* a parabola as the locus of a point such that its distances from a fixed poin called the focus is equation to its distance from a fixed straight line called the directrix. In what follows we do that, and use that property to show that the equation that follows is $y = 4qx$; that is to say the reverse of what we have just done.

We define a parabola as the locus of a point that moves such that its distance from a fixed straight line called the *directrix* is equal to its distance from a fixed point called the *focus*. Unlike the ellipse, a parabola has only one focus and one directrix. However, comparison of this definition with the focus - directrix property of the ellipse (which can also be used to define the ellipse) shows that the parabola can be regarded as a limiting form of an ellipse with eccentricity equal to unity.

We shall find the equation to a parabola whose directrix is the line $y = -q$ and whose focus is the point $(q , 0)$. Figure II.20 shows the parabola. F is the focus and O is the origin of the coordinate system. The vertex of the parabola is at the origin. In an orbital context, for example, the orbit of a comet moving around the Sun in parabolic orbit, the Sun would be at the focus F, and the distance between vertex and focus would be the perihelion distance, for which the symbol q is traditionally used in orbit theory.

FIGURE II.20



From figure II.20, it is evident that the definition of the parabola (PF = PN) requires that

$$(x-q)^2 + y^2 = (x+q)^2, \tag{2.4.1}$$

from which

$$y^2 = 4qx, \qquad 2.4.2$$

which is the equation to the parabola. The reader should also try sketching the parabolas $y^2 = -4qx, \quad x^2 = 4qy, \quad x^2 = -4qy, \quad (y-2)^2 = 4q(x-3)$.

The line parallel to the y-axis and passing through the focus is the *latus rectum*. Substitution of $x = q$ into $y^2 = 4qx$ shows that the latus rectum intersects the parabola at the two points $(q\ ,\ \pm 2q)$, and that the length l of the semi latus rectum is $2q$.

The equations

$$x = qt^2\ , \quad y = 2qt \qquad 2.4.3$$

are the parametric equations to the parabola, for $y^2 = 4qx$ results from the elimination of t between them. In other words, if t is any variable, then any point that satisfies these two equations lies on the parabola.

Most readers will know that if a particle is moving with constant speed in one direction and constant acceleration at right angles to that direction, as with a ball projected in a uniform gravitational field or an electron moving in a uniform electric field, the path is a parabola. In the constant speed direction the distance is proportional to the time, and in the constant acceleration direction, the distance is proportional to the square of the time, and hence the path is a parabola.

Returning to the focus-directrix property, it will have occurred to the reader, that, give any straight line, and a point not on the line, it should be possible to construct a parabola for which the line is its directrix and the point it its focus. For example, construct the parabola which has the line $y = \frac{1}{2}x - \frac{1}{4}$ as directrix and the point $(4\ , 3)$ as focus.

Solution: The distance d between a point $(x\ , y)$ and the point $(4\ , 3)$ is given by

$$d^2 = (x - 4)^2 + (y - 3)^2 = x^2 + y^2 - 8x - 6y + 25$$

The line can be written as $2x - 4y = 1$. On referring to equation 2.2.11, we find that the distance p of the point $(4\ , 3)$ from the line is given by

$$20p^2 = (1 - 2x + 4y)^2$$

On equating these two distances, we soon find that the equation to the required parabola is

$$16x^2 + 16xy + 4y^2 - 156x - 128y + 499 = 0.$$

Tangents to a Parabola.

Where does the straight line $y = mx + c$ intersect the parabola $y^2 = 4qx$? The answer is found by substituting $mx + c$ for y to obtain, after rearrangement,

$$m^2x^2 + 2(mc - 2q)x + c^2 = 0. \qquad 2.4.4$$

The line is tangent if the discriminant is zero, which leads to

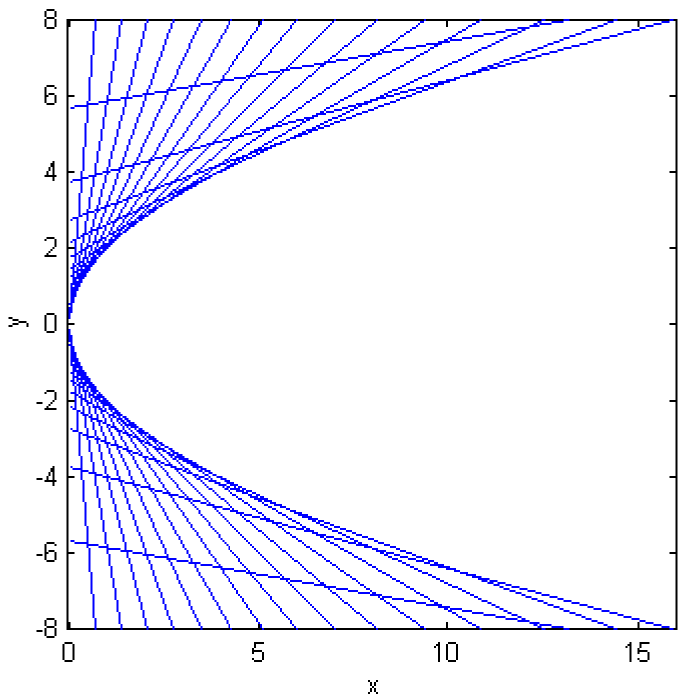
$$c = q/m. \qquad 2.4.5$$

Thus a straight line of the form

$$y = mx + q/m \qquad 2.4.6$$

is tangent to the parabola. Figure II.22 (sorry - there is no figure II.21!) illustrates this for several lines, the slopes of each differing by 5° from the next.

FIGURE II.22



We shall now derive an equation to the line that is tangent to the parabola at the point (x_1, y_1).

Let $(x_1, y_1) = (qt_1^2, 2qt_1)$ be a point on the parabola, and
Let $(x_2, y_2) = (qt_2^2, 2qt_2)$ be another point on the parabola.

The line joining these two points is

$$\frac{y - 2qt_1}{x - qt_1^2} = \frac{2q(t_2 - t_1)}{q\left(t_2^2 - t_1^2\right)} = \frac{2}{t_2 + t_1}. \qquad 2.4.7$$

Now let t_2 approach t_1, eventually coinciding with it. Putting $t_1 = t_2 = t$ in the last equation results, after simplification, in

$$ty = x + qt^2, \qquad 2.4.8$$

being the equation to the tangent at $(qt^2, 2qt)$.

Multiply by $2q$:

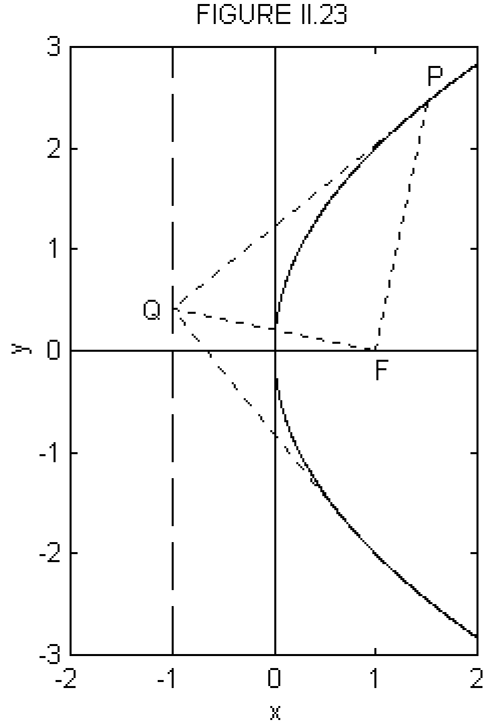
$$2qty = 2q(x + qt^2) \qquad 2.4.9$$

and it is seen that the equation to the tangent at (x_1, y_1) is

$$y_1 y = 2q\,(x_1 + x). \qquad 2.4.10$$

There are a number of interesting geometric properties, some of which are given here. For example, if a tangent to the parabola at a point P meets the directrix at Q, then, just as for the ellipse, P and Q subtend a right angle at the focus (figure II.23). The proof is similar to that given for the ellipse, and is left for the reader.

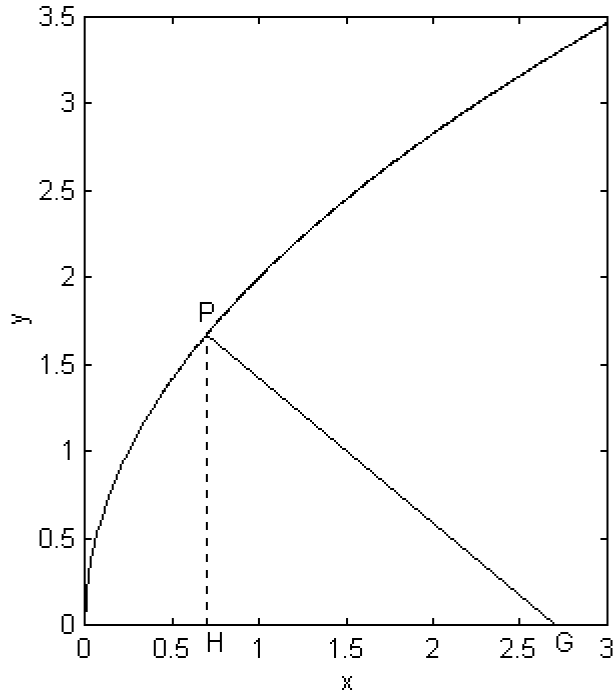
FIGURE II.23



The reader will recall that perpendicular tangents to an ellipse meet on the director circle. The analogous theorem *vis-à-vis* the parabola is that perpendicular tangents meet on the directrix. This is also illustrated in figure II.23. The theorem is not specially important in orbit theory, and the proof is also left to the reader.

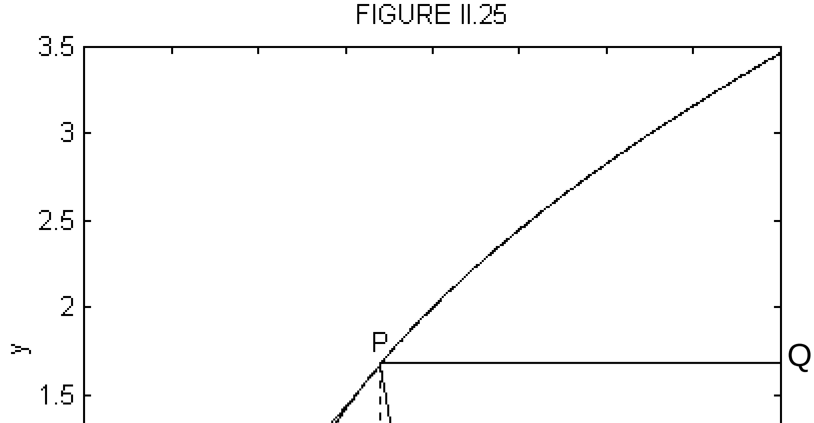
Let PG be the normal to the parabola at point P, meeting the axis at G (figure II.24). We shall call the length GH the subnormal. A curious property is that the length of GH is always equal to l, the length of the semi latus rectum (which in figure II.24 is of length 2 – i.e. the ordinate where x = 1), irrespective of the position of P. This proof again is left to the reader.

FIGURE II.24



The following two geometrical properties, while not having immediate applications to orbit theory, certainly have applications to astronomy.

FIGURE II.25



The tangent at P makes an angle α with the x-axis, and PF makes an angle β with the x-axis (figure II.25). We shall show that $\beta = 2\alpha$ and deduce an interesting consequence.

The equation to the tangent (see equation 2.4.8) is $ty = x + qt^2$, which shows that

$$\tan\alpha = 1/t. \qquad 2.4.11$$

The coordinates of P and F are, respectively, $\left(qt^2,\ 2qt\right)$ and $\left(q\ ,\ 0\right)$, and so, from the triangle PFH, we find.

$$\tan\beta = \frac{2t}{t^2 - 1}. \qquad 2.4.12$$

Let $\tau = 1/t$, then $\tan\alpha = \tau$ and $\tan\beta = 2\tau/(1 - \tau^2)$, which shows that $\beta = 2\alpha$.

This also shows that triangle JFP is isosceles, with the angles at J and P each being α. This can also be shown as follows.

From the equation $ty = x + qt^2$, we see that J is the point $(-qt^2\ , 0)$, so that JF = $q\,(t^2 + 1)$.

From triangle PFH, we see that

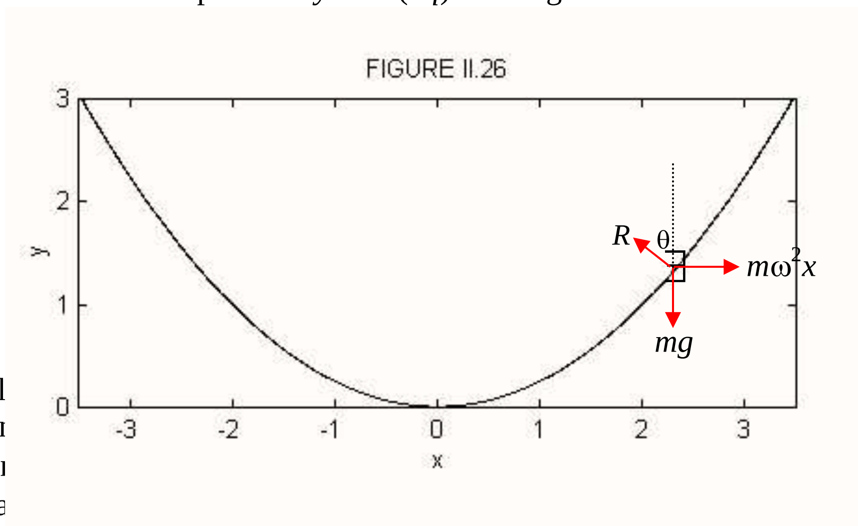
$$(\mathrm{PF})^2 = 4q^2t^2 + q^2\left(t^2 - 1\right)^2 - q^2\left(t^2 + 1\right)^2. \qquad 2.4.13$$

Therefore

$$\mathrm{PF} = \mathrm{JF}. \qquad 2.4.14$$

Either way, since the triangle JPF is isosceles, it follows that QP and PF make the same angle α to the tangent. If the parabola is a cross section of a telescopic mirror, any ray of light coming in parallel to the axis will be focussed at F, so that a paraboloidal mirror, used on-axis, does not suffer from spherical aberration. (This property holds, of course, only for light parallel to the axis of the paraboloid, so that a paraboloidal mirror, without some sort of correction, gives good images over only a narrow field of view.)

Now consider what happens when you stir a cup of tea. The surface takes up a shape that looks as though it might resemble the parabola $y = x^2/(4q)$ - see figure II.26:



Suppose the l the surface is in equilibrium (i e centrifugal force $m\omega^2 x$ and the h the vertical (and the tangent ma

$$\tan\theta = \frac{\omega^2 x}{g}. \qquad 2.4.15$$

But the slope of the parabola $y = x^2/(4q)$ is $x/(2q)$, so that the surface is indeed a parabola with semi latus rectum $2q = g/\omega^2$.

This phenomenon has been used in Canada to make a successful large telescope (diameter 6 m) in which the mirror is a spinning disc of mercury that takes up a perfectly paraboloidal shape. Another example is the spin casting method that has been successfully used for the production of large, solid glass paraboloidal telescope mirrors. In this process, the furnace is rotated about a vertical axis while the molten glass cools and eventually solidifies into the required paraboloidal effect.

Exercise. The 6.5 metre diameter mirrors for the twin Magellan telescopes at Las Campañas, Chile, have a focal ratio *f*/1.25. They were made by the technique of spin casting at The University of Arizona's Mirror Laboratory. At what speed would the furnace have had to be rotated in order to achieve the desired focal ratio? (Answer = 7.4 rpm.) Notice that *f*/1.25 is quite a deep paraboloid. If this mirror had been made by traditional grinding from a solid disc, what volume of material would have had to be removed to make the desired paraboloid? (Answer - a whopping 5.4 cubic metres, or about 12 tons!)

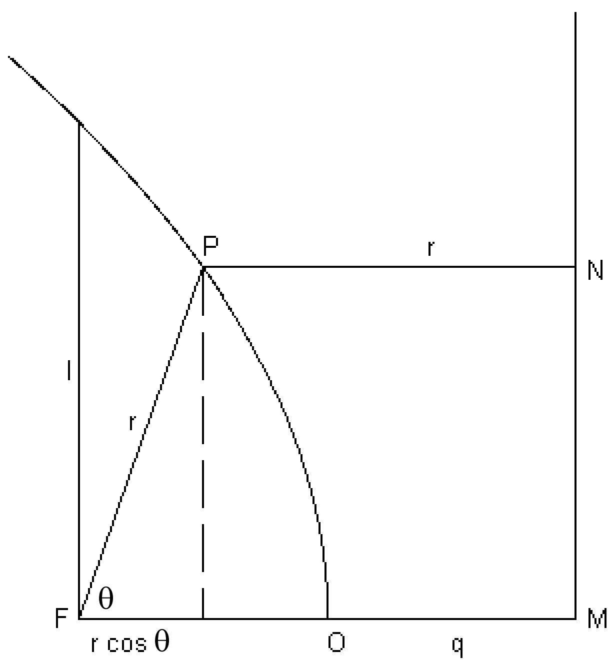
Polar equation to the Parabola.

As with the ellipse, we choose the focus as pole and the axis of the parabola as initial line. We shall orient the parabola so that the vertex is toward the right, as in figure II.27.

We recall the focus-directrix property, FP = PN. Also, from the definition of the directrix, FO=OM = q, so that FM = $2q = l$, the length of the semi latus rectum. It is therefore immediately evident from figure II.27 that $r\cos\theta + r = 2q = l$, so that the polar equation to the parabola is

$$r = \frac{l}{1+\cos\theta}. \tag{2.4.16}$$

FIGURE II.27



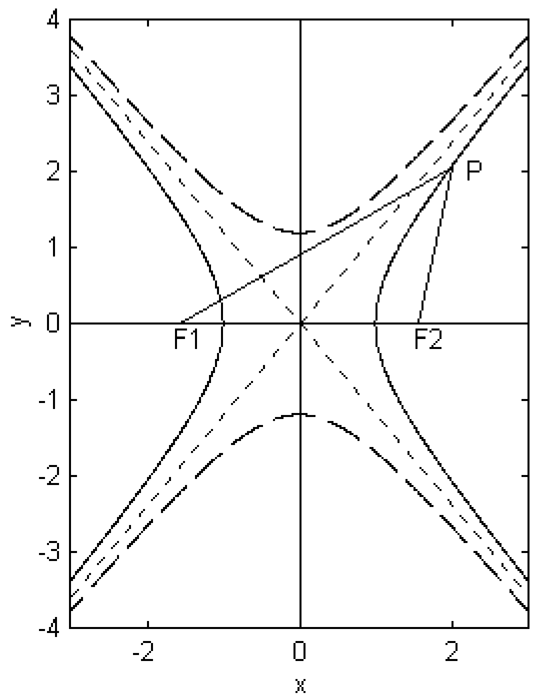
This is the same as the polar equation to the ellipse (equation 2.3.36), with $e = 1$ for the parabola. I have given different derivations for the ellipse and for the parabola; the reader might like to interchange the two approaches and develop equation 2.3.36 in the same manner as we have developed equation 2.4.16.

When we discuss the hyperbola, I shall ask you to show that its polar equation is also the same as 2.3.36. In other words, equation 2.3.36 is the equation to a *conic section*, and it represents an ellipse, parabola or hyperbola according to whether $e < 1$, $e = 1$ or $e > 1$.

2.5 *The Hyperbola*

A hyperbola is the locus of a point that moves such that the difference between its distances from two fixed points called the foci is constant. We shall call the difference between these two distances $2a$ and the distance between the foci $2ae$, where e is the eccentricity of the hyperbola, and is a number greater than 1. See figure II.28.

FIGURE II.28



For example, in a Young's double-slit interference experiment, the mth bright fringe is located at a point on the screen such that the path difference for the rays from the two slits is m wavelengths. As the screen is moved forward or backwards, this relation continues to hold for the mth bright fringe, whose locus between the slits and the screen is therefore a hyperbola. The "Decca" system of radar navigation, first used at the D-Day landings in the Second World War and decommissioned only as late as 2000 on account of being rendered obsolete by the "GPS" (Global Positioning Satellite) system, depended on this property of the hyperbola. (Since writing this, part of the Decca system has been re-commissioned as a back-up in case of problems with GPS.) Two radar transmitters some distance apart would simultaneously transmit radar pulses. A ship would receive the two signals separated by a short time interval, depending on the difference between the distances from the ship to the two transmitters. This placed the ship on a particular hyperbola. The ship would also listen in to another pair of transmitters, and this would place the ship on a second hyperbola. This then placed the ship at one of the four points where the two hyperbolas intersected. It would usually be obvious which of the four points was the correct one, but any ambiguity could be resolved by the signals from a third pair of transmitters.

In figure II.28, the coordinates of F_1 and F_2 are, respectively, $(-ae\ ,\ 0)$ and $(ae\ ,\ 0)$.
The condition $PF_1 - PF_2 = 2a$ requires that

$$\left[(x+ae)^2+y^2\right]^{1/2} - \left[(x-ae)^2+y^2\right]^{1/2} = 2a, \qquad 2.5.1$$

and this is the equation to the hyperbola. After some arrangement, this can be written

$$\frac{x^2}{a^2} - \frac{y^2}{a^2\left(e^2-1\right)} = 1, \qquad 2.5.2$$

which is a more familiar form for the equation to the hyperbola. Let us define a length b by

$$b^2 = a^2\,(e^2-1). \qquad 2.5.3$$

The equation then becomes

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad 2.5.4$$

which is the most familiar form for the equation to a hyperbola.

Problem. When a meteor streaks across the sky, it can be tracked by radar. The radar instrumentation can determine the range (distance) of the meteoroid as a function of time. Show that, if the meteoroid is moving at constant speed (a questionable assumption, because it must be decelerating, but perhaps we can assume the decrease in speed is negligible during the course of the observation), and if the range r is plotted against the time, the graph will be a hyperbola. Show also that, if r^2 is plotted against t, the graph will be a parabola of the form

$$r^2 = at^2 + bt + c,$$

where $a = V^2$, $b = -2V^2t_0$, $c = V^2t_0^2 + r_0^2$, V = speed of the meteoroid, t_0 = time of closest approach, r_0 = distance of closest approach.

Radar observation of a meteor yields the following range-time data:

t (s)	r (km)	
0.0	101.4	*
0.1	103.0	
0.2	105.8	
0.3	107.8	
0.4	111.1	
0.5	112.6	
0.6	116.7	
0.7	119.3	
0.8	123.8	*
0.9	126.4	
1.0	130.6	
1.1	133.3	
1.2	138.1	
1.3	141.3	*

Assume that the velocity of the meteor is constant.

Determine
i. The time of closest approach (to 0.01 s)
ii. The distance of closest approach (to 0.1 km)
iii. The speed (to 1.0 km s^{-1})

If you wish, just use the three asterisked data to determine a, b and c. If you are more energetic, use *all* the data, and determine a, b and c by least squares, and the probable errors of V, t_0 and r_0.

The distance between the two vertices of the hyperbola is its transverse axis, and the length of the semi transverse axis is a – but what is the geometric meaning of the length b? This is discussed below in the next subsection (on the conjugate hyperbola) and again in a later section on the impact parameter.

The lines perpendicular to the x-axis and passing through the foci are the two *latera recta*. Since the foci are at $(\pm ae, 0)$, the points where the latera recta intersect the hyperbola can be found by putting $x = ae$ into the equation to the hyperbola, and it is then found that the length l of a semi latus rectum is

$$l = a(e^2 - 1). \qquad 2.5.5$$

The Conjugate Hyperbola.

The equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1 \qquad 2.5.6$$

is the equation to the conjugate hyperbola. The conjugate hyperbola is drawn dashed in figure II.28, and it is seen that the geometric meaning of b is that it is the length of the semi transverse axis of the conjugate hyperbola. It is a simple matter to show that the eccentricity of the conjugate hyperbola is $e/\sqrt{e^2-1}$.

The Asymptotes.

The lines
$$y = \pm\frac{bx}{a} \qquad 2.5.7$$

are the *asymptotes* of the hyperbola. This can also be written

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0. \qquad 2.5.8$$

Thus
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = c \qquad 2.5.9$$

is the hyperbola, the asymptotes, or the conjugate hyperbola, if $c = +1$, 0 or -1 respectively. The asymptotes are drawn as dotted lines in figure II.28.

The semi angle ψ between the asymptotes is given by

$$\tan\psi = b/a. \qquad 2.5.10$$

Easy Exercise

If the eccentricity of a hyperbola is e, show that the eccentricity of its conjugate is $\dfrac{e}{\sqrt{e^2-1}}$.

Corollary No one will be surprised to note that this implies that, if the eccentricities of a hyperbola and its conjugate are equal, then each is equal to $\sqrt{2}$.

The Directrices.

The lines $y = \pm a/e$ are the directrices, and, as with the ellipse (and with a similar proof), the hyperbola has the property that the ratio of the distance PF_2 to a focus to the distance PN to the directrix is constant and is equal to the eccentricity of the hyperbola. This ratio (i.e. the eccentricity) is less than one for the ellipse, equal to one for the parabola, and greater than one for the hyperbola. It is not a property that will be of great importance for our purposes, but is worth mentioning because it is a property that is sometimes used to define a hyperbola. I leave it to the reader to draw the directrices in their correct positions in figure II.28.

Parametric Equations to the Hyperbola.

The reader will recall that the point (a cos E, b sin E) is on the ellipse $(x^2/a^2) + \left(y^2/b^2\right) = 1$ and that this is evident because this equation is the E-eliminant of $x = a$ cos E and $y = b$ sin E. The angle E has a geometric interpretation as the eccentric anomaly. Likewise, recalling the relation $\cosh^2\phi - \sinh^2\phi = 1$, it will be evident that $(x^2/a^2) - \left(y^2/b^2\right) = 1$ can also be obtained as the ϕ–eliminant of the equations

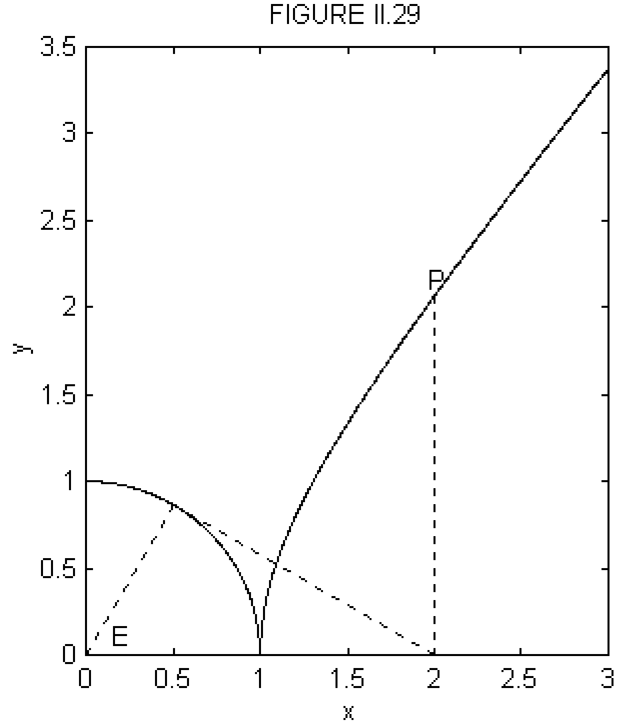
$$x = a\cosh\phi, \qquad y = b\sinh\phi \qquad 2.5.11$$

These two equations are therefore the parametric equations to the hyperbola, and any point satisfying these two equations lies on the hyperbola. The variable ϕ is not an angle, and has no geometric interpretation analogous to the eccentric anomaly of an ellipse. The equations

$$x = a\sec E\ , \quad y = b\tan E \qquad 2.5.12$$

can also be used as parametric equations to the hyperbola, on account of the trigonometric identity $1 + \tan^2 E = \sec^2 E$. In that case, the angle E does have a geometric interpretation (albeit not a particularly interesting one) in relation to the *auxiliary circle*, which is the circle of radius a centred at the origin. The meaning of the angle should be evident from figure II.29, in which E is the eccentric angle corresponding to the point P.

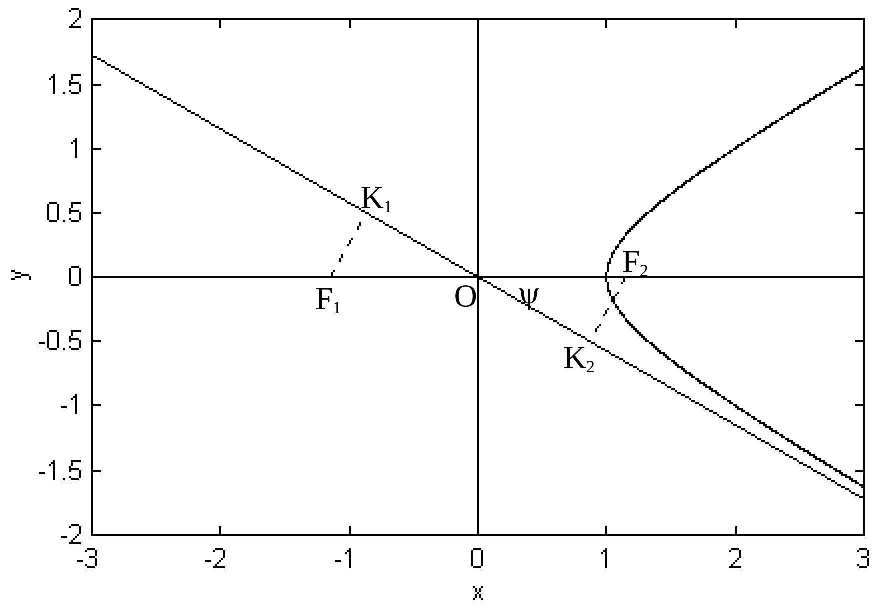
FIGURE II.29



Impact Parameter.

A particle travelling very fast under the action of an inverse square attractive force (such as an interstellar meteoroid or comet - if there are such things - passing by the Sun, or an electron in the vicinity of a positively charged atomic nucleus) will move in a hyperbolic path. We prove this in a later chapter, as well as discussing the necessary speed. We may imagine the particle initially approaching from a great distance along the asymptote at the bottom right hand corner of figure II.30. As it approaches the focus, it no longer moves along the asymptote but along an arm of the hyperbola.

FIGURE II.30



The distance $K_2 F_2$, which is the distance by which the particle would have missed F_2 in the absence of an attractive force, is commonly called the *impact parameter*. Likewise, if the force had been a repulsive force (for example, suppose the moving particle were a positively charged particle and there were a centre of repulsion at F_1, F_1K_1 would be the impact parameter. Clearly, F_1K_1 and F_2K_2 are equal in length. The symbol that is often used in scattering theory, whether in celestial mechanics or in particle physics, is b - but is this b the same b that goes into the equation to the hyperbola and which is equal to the semi major axis of the conjugate hyperbola?

$OF_2 = ae$, and therefore $K_2F_2 = ae \sin \psi$. This, in conjunction with $\tan \psi = b/a$ and $b^2 = a^2 (e^2 - 1)$, will soon show that the impact parameter is indeed the same b that we are familiar with, and that b is therefore a very suitable symbol to use for impact parameter.

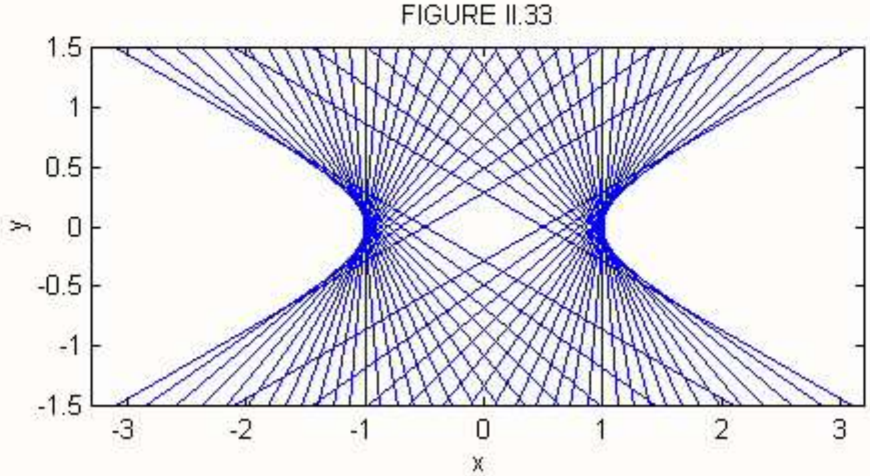
Tangents to the Hyperbola.

Using the same arguments as for the ellipse, the reader should easily find that lines of the form

$$y = mx \pm \sqrt{a^2m^2 - b^2} \qquad 2.5.13$$

are tangent to the hyperbola. This is illustrated in figure II.33 for a hyperbola with $b = a/2$, with tangents drawn with slopes 30° to 150° in steps of 5°. (The asymptotes have ψ = 26° 34'.) (Sorry, but there are no figures II.31 or II.32 - computer problems!)

FIGURE II.33



Likewise, from similar arguments used for the ellipse, the tangent to the hyperbola at the point (x ,y) is found to be

$$\frac{x_1x}{a^2} - \frac{y_1y}{b^2} = 1. \qquad 2.5.14$$

Director Circle.

As for the ellipse, and with a similar derivation, the locus of the points of intersection of perpendicular tangents is a circle, called the director circle, which is of radius $\sqrt{(a^2 - b^2)}$. This is not of particular importance for our purposes, but the reader who is interested might like to prove this by the same method as was done for the director circle of the ellipse, and might like to try drawing the circle and some tangents. If $b > a$, that is to say if $\psi > 45°$ and the angle between the asymptotes is greater than 90° , the director circle is not real and it is of course not possible to draw perpendicular tangents.

Rectangular Hyperbola.

If the angle between the asymptotes is 90° , the hyperbola is called a *rectangular hyperbola*. For such a hyperbola, $b = a$, the eccentricity is $\sqrt{2}$, the director circle is a point, namely the origin, and perpendicular tangents can be drawn only from the asymptotes.

The equation to a rectangular hyperbola is

$$x^2 - y^2 = a^2 \qquad 2.5.15$$

and the asymptotes are at 45° to the x axis.

Let Ox', Oy' be a set of axes at 45° to the x axis. (That is to say, they are the asymptotes of the rectangular hyperbola.) Then the equation to the rectangular hyperbola referred to its asymptotes as coordinate axes is found by the substitutions

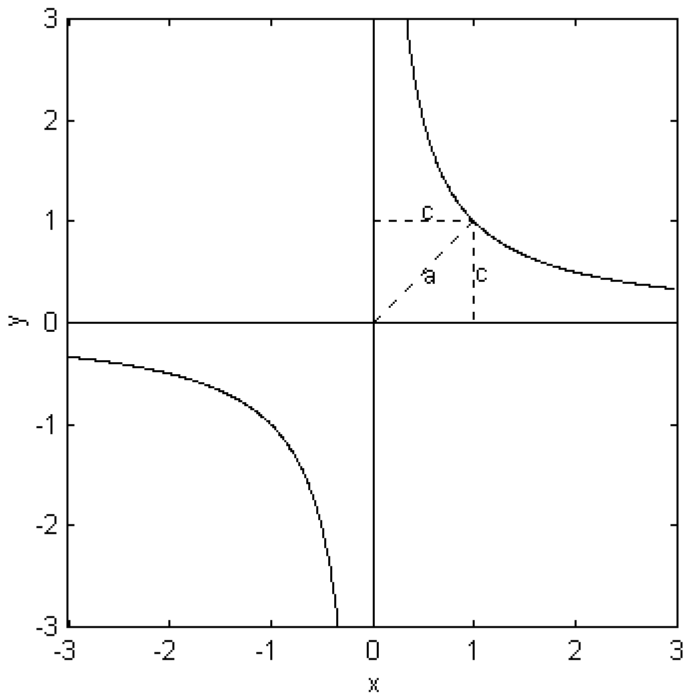
$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos 45^\circ & \sin 45^\circ \\ -\sin 45^\circ & \cos 45^\circ \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \qquad 2.5.16$$

into $x^2 - y^2 = a^2$. This results in the equation

$$x'y' = \tfrac{1}{2}a^2 = c^2, \quad \text{where} \quad c = a/\sqrt{2}, \qquad 2.5.17$$

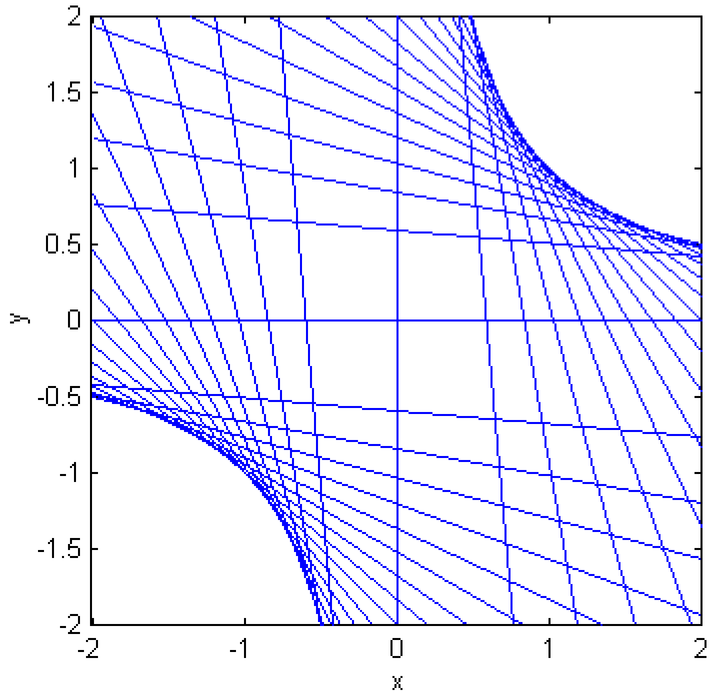
for the equation to the rectangular hyperbola referred to its asymptotes as coordinate axes. The geometric interpretation of c is shown in figure II.32, which is drawn for $c = 1$, and we have called the coordinate axes x and y. The length of the semi transverse axis is $c\sqrt{2}$.

FIGURE II.32



The simple equation $y = 1/x$ is a rectangular hyperbola and indeed it is this equation that is shown in figure II.32.

It is left to the reader to show that the parametric equations to the rectangular hyperbola $xy = c^2$ (we have dropped the primes) are $x = ct$, $y = c/t$, that lines of the form $y = mx \pm 2c\sqrt{-m}$ are tangent to $xy = c^2$ (figure II.35, drawn with slopes from 90° to 180° in steps of 5°), and that the tangent at $(x_1\ , y_1)$ is $x_1 y + y_1 x = 2c$.



Equation of a hyperbola referred to its asymptotes as axes of coordinates.

We have shown that the equation to a *rectangular* hyperbola referred to its asymptotes as axes of coordinates is $x'y' = \frac{1}{2}a^2 = c^2$. In fact the equation $x'y' = c^2$ is the equation to *any* hyperbola (centred at (0 , 0)), not necessarily rectangular, when referred to its asymptotes as axes of coordinates, where $c^2 = \frac{1}{4}(a^2 + b^2)$. In the figure below I have drawn a hyperbola and a point on the hyperbola whose coordinates with respect to the horizontal and vertical axes are (x, y), and whose coordinates with respect to the asymptotes are (x', y'). I have shown the distances x and y

with blue dashed lines, and the distances x' and y' with red dashed lines. The semiangle between the asymptotes is ψ.

The equation to the hyperbola referred to the horizontal and vertical axes is

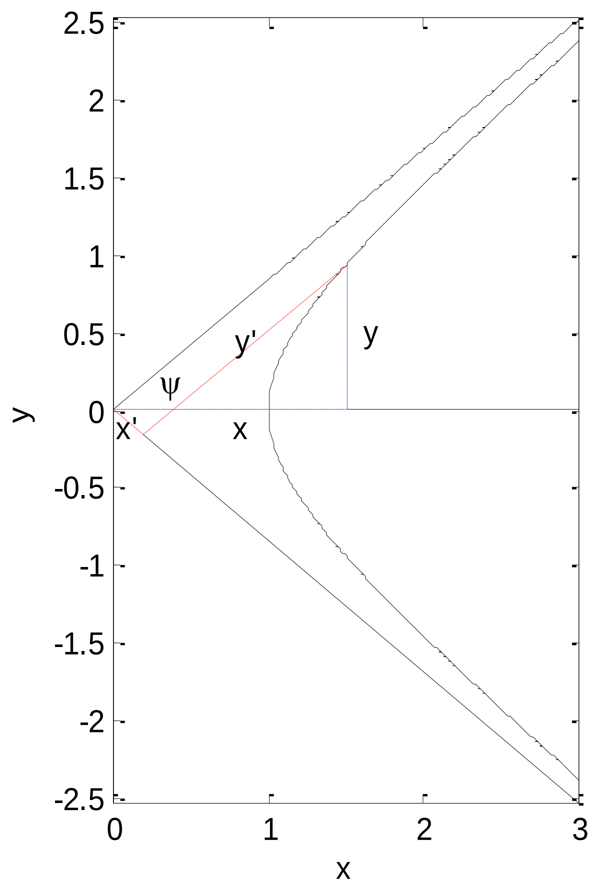
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \qquad 2.5.18$$

From the drawing, we see that

$$x = (x' + y')\cos\psi, \qquad y = (y' - x')\sin\psi. \qquad 2.5.19\text{a,b}$$

If we substitute these into equation 2.5.18, and also make use of the relation tan $\psi = b/a$ (equation 2.5.10), we arrive at the equation to the hyperbola referred to the asymptotes as axes of coordinates:

$$x'y' = \tfrac{1}{4}(a^2 + b^2) = c^2. \qquad 2.5.20$$



Polar equation to the hyperbola.

We found the polar equations to the ellipse and the parabola in different ways. Now go back and look at both methods and use either (or both) to show that the polar equation to the hyperbola (focus as pole) is

$$r = \frac{l}{1 + e\cos\theta}. \qquad 2.5.21$$

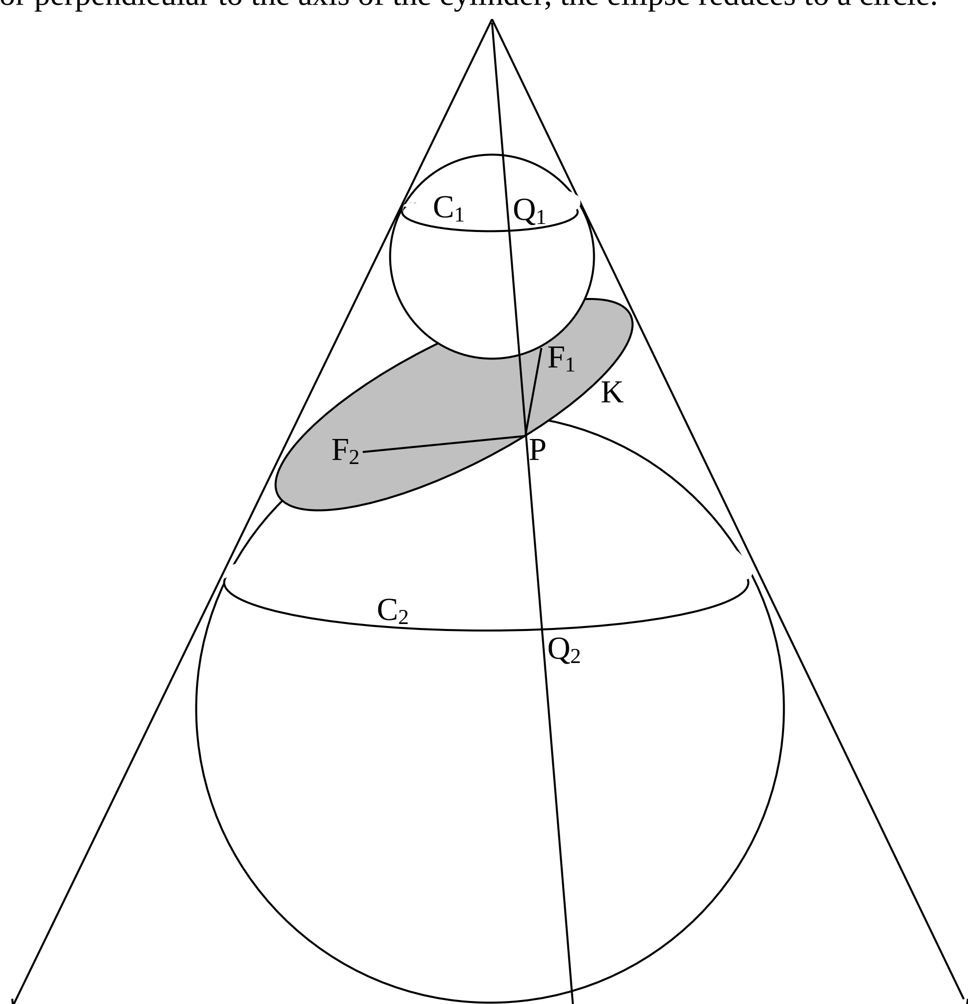
This is the polar equation to any conic section - which one being determined solely by the value of *e*. You should also ask yourself what is represented by the equation

$$r = \frac{l}{1 - e\cos\theta}. \qquad 2.5.22$$

Try sketching it for different values of *e*.

2.6 *Conic Sections*.

We have so far defined an ellipse, a parabola and a hyperbola without any reference to a cone. Many readers will know that a plane section of a cone is either an ellipse, a parabola or a hyperbola, depending on whether the angle that the plane makes with the base of the cone is less than, equal to or greater than the angle that the generator of the cone makes with its base. However, given the definitions of the ellipse, parabola and hyperbola that we have given, proof is required that they are in fact conic sections. It might also be mentioned at this point that a plane section of a circular cylinder is also an ellipse. Also, of course, if the plane is parallel with the base of the cone, or perpendicular to the axis of the cylinder, the ellipse reduces to a circle.



A simple and remarkable proof can be given in the classical Euclidean "Given. Required. Construction. Proof. Q.E.D." style.

Given: A cone and a plane such that the angle that the plane makes with the base of the cone is less than the angle that the generator of the cone makes with its base, and the plane cuts the cone in a closed curve K. Figure II.36.

Required: To prove that K is an ellipse.

Construction: Construct a sphere above the plane, which touches the cone internally in the circle C_1 and the plane at the point F_1. Construct also a sphere below the plane, which touches the cone internally in the circle C_2 and the plane at the point F_2.

Join a point P on the curve K to F_1 and to F_2.
Draw the generator that passes through the point P and which intersects C_1 at Q_1 and C_2 at Q_2.

Proof: $PF_1 = PQ_1$ (Tangents to a sphere from an external point.)

$PF_2 = PQ_2$ (Tangents to a sphere from an external point.)

$$\therefore \; PF_1 + PF_2 = PQ_1 + PQ_2 = Q_1 Q_2$$

and Q_1Q_2 is independent of the position of P, since it is the distance between the circles C_1 and C_2 measured along a generator.

$\therefore$ K is an ellipse and F_1 and F_2 are its foci. Q.E.D.

A similar argument will show that a plane section of a cylinder is also an ellipse.

The reader can also devise drawings that will show that a plane section of a cone parallel to a generator is a parabola, and that a plane steeper than a generator cuts the cone in a hyperbola. The drawings are easiest to do with paper, pencil, compass and ruler, and will require some ingenuity. While I have seen the above proof for an ellipse in several books, I have not seen the corresponding proofs for a parabola and a hyperbola, but they can indeed be done, and the reader should find it an interesting challenge. If the reader can use a computer to make the drawings and can do better than my poor effort with figure II.36, s/he is pretty good with a computer, which is a sign of a misspent youth.

2.7 *The General Conic Section*

The equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad 2.7.1$$

represents an ellipse whose major axis is along the x axis and whose centre is at the origin of coordinates. But what if its centre is not at the origin, and if the major axis is at some skew angle to the x axis? What will be the equation that represents such an ellipse? Figure II.37.

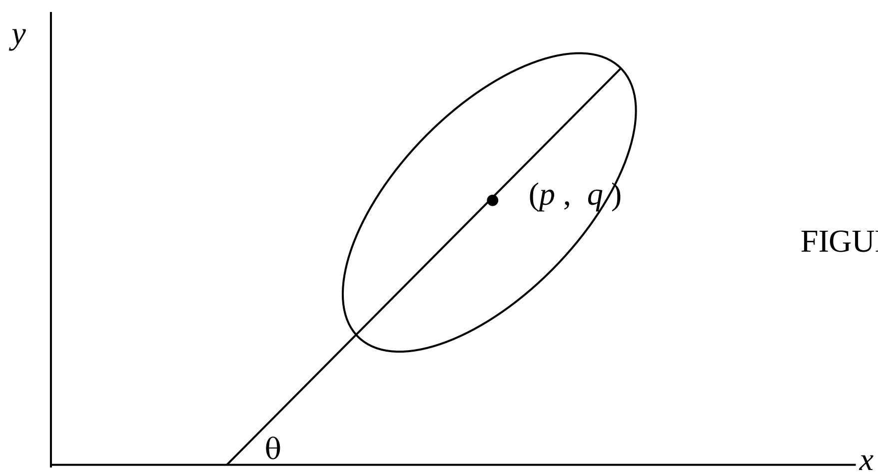


FIGURE II.37

If the centre is translated from the origin to the point (p, q), the equation that represents the ellipse will be found by replacing x by $x - p$ and y by $y - q$. If the major axis is inclined at an angle θ to the x axis, the equation that represents the ellipse will be found by replacing x by $x\cos\theta + y\sin\theta$ and y by $-x \sin \theta + y \cos \theta$. In any case, if the ellipse is translated or rotated or both, x and y will each be replaced by linear expressions in x and y, and the resulting equation will have at most terms in x^2, y^2, xy, x, y and a constant. The same is true of a parabola or a hyperbola. Thus, any of these three curves will be represented by an equation of the form

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0. \qquad 2.7.2$$

(The coefficients a and b are not the semi major and semi minor axes.) The apparently random notation for the coefficients arises because these figures are plane sections of three-dimensional surfaces (the ellipsoid, paraboloid and hyperboloid) which are described by terms involving the coordinate z as well as x and y. The customary notation for these three-dimensional surfaces is very systematic, but when the terms in z are removed for the two- dimensional case, the apparently random notation a, b, c, f, g, h remains. In any case, the above equation can be divided through by the constant term without loss of generality, so that the equation to an ellipse, parabola or hyperbola can be written, if preferred, as

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0. \qquad 2.7.3$$

Is the converse true? That is, does an equation of this form always necessarily represent an ellipse, parabola or hyperbola?

Not quite. For example,

$$6x^2 + xy - y^2 - 17x - y + 12 = 0 \qquad 2.7.4$$

represents two straight lines (it can be factored into two linear terms - try it), while

$$2x^2 - 4xy + 4y^2 - 4x + 4 = 0 \qquad 2.7.5$$

is satisfied only by a single point. (Find it.)

However, a plane section of a cone can be two lines or a single point, so perhaps we can now ask whether the general second degree equation must always represent a conic section. The answer is: close, but not quite.

For example,

$$4x^2 + 12xy + 9y^2 + 14x + 21y + 6 = 0 \qquad 2.7.6$$

represents two parallel straight lines, while

$$x^2 + y^2 + 3x + 4y + 15 = 0 \qquad 2.7.7$$

cannot be satisfied by any real (x,y).

However, a plane can intersect a *cylinder* in two parallel straight lines, or a single straight line, or not at all. Therefore, if we stretch the definition of a cone somewhat to include a cylinder as a special limiting case, then we can say that the general second degree equation in x and y does indeed always represent a conic section.

Is there any means by which one can tell by a glance at a particular second degree equation, for example

$$8x^2 + 10xy - 3y^2 - 2x - 4y - 2 = 0, \qquad 2.7.8$$

what type of conic section is represented? The answer is yes, and this one happens to be a hyperbola. The discrimination is made by examining the properties of the determinant

$$\Delta = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} \qquad 2.7.9$$

I have devised a table after the design of the dichotomous tables commonly used by taxonomists in biology, in which the user is confronted by a couplet (or sometimes triplet) of alternatives, and is then directed to the next part of the table. I shall spare the reader the derivation of the table; instead, I shall describe its use.

In the table, I have used the symbol $\bar{a}$ to mean the cofactor of *a* in the determinant, $\bar{h}$ the cofactor of *h*, $\bar{g}$ the cofactor of *g*, etc. Explicitly, these are

$$\bar{a} = bc - f^2, \qquad 2.7.10$$
$$\bar{b} = ca - g^2, \qquad 2.7.11$$
$$\bar{c} = ab - h^2, \qquad 2.7.12$$
$$\bar{f} = gh - af, \qquad 2.7.13$$
$$\bar{g} = hf - bg \qquad 2.7.14$$

and

$$\bar{h} = fg - ch. \qquad 2.7.15$$

The first column labels the choices that the user is asked to make. At the beginning, there are two choices to make, 1 and 1' The second column says what these choices are, and the fourth column says where to go next. Thus, if the determinant is zero, go to 2; otherwise, go to 5. If there is an asterisk in column 4, you are finished. Column 3 says what sort of a conic section you have arrived at, and column 5 gives an example.

No matter what type the conic section is, the coordinates of its centre are $\left(\bar{g}/\bar{c},\ \bar{f}/\bar{c}\right)$ and the angle θ that its major or transverse axis makes with the *x* axis is given by

$$\tan 2\theta = \frac{2h}{a-b}. \qquad 2.7.16$$

Thus if *x* is first replaced with $x + \bar{g}/\bar{c}$ and *y* with $y + \bar{f}/\bar{c}$, and then the new *x* is replaced with $x\cos\theta - y\sin\theta$ and the new *y* with $x\sin\theta + y\cos\theta$, the equation will take the familiar form of a conic section with its major or transverse axis coincident with the *x* axis and its centre at the origin. Any of its properties, such as the eccentricity, can then be deduced from the familiar equations. You should try this with equation 2.7.8.

Key to the Conic Sections

1	Δ=0		2	
1'	Δ≠0		5	
2	$\bar{c} > 0$	Point	*	$x^2 - 2xy + 2y^2 - 2x + 2 = 0$
2'	$\bar{c} = 0$		3	
2''	$\bar{c} < 0$	Two nonparallel straight lines	4	
3	$\bar{h} = 0$	Straight line	*	$4x^2 + 4xy + y^2 + 12x + 6y + 9 = 0$
3'	$\bar{h} \neq 0$	Two parallel straight lines	*	$4x^2 + 12xy + 9y^2 + 14x + 21y + 6 = 0$
4	$a + b = 0$	Two perpendicular straight lines	*	$6x^2 + 5xy - 6y^2 + x + 8y - 2 = 0$
4'	$a + b \neq 0$	Two straight lines, neither parallel nor perpendicular	*	$6x^2 - xy - y^2 + 34x + 13y - 12 = 0$
5	$\bar{c} > 0$		6	
5'	$\bar{c} = 0$	Parabola	*	$9x^2 - 12xy + 4y^2 - 18x - 101y + 19 = 0$
5''	$\bar{c} < 0$		8	
6	$a\Delta > 0$	Nothing	*	$x^2 + y^2 + 3x + 4y + 15 = 0$
6'	$a\Delta < 0$		7	
7	a=b and h=0	Circle	*	$x^2 + y^2 - 6x - 8y + 9 = 0$
7'	Not so	Ellipse	*	$14x^2 - 4xy + 11y^2 - 44x - 58y + 71 = 0$
8	a+b=0	Rectangular hyperbola	*	$7x^2 - 48xy - 7y^2 + 10x - 28y + 100 = 0$
8'	a+b≠0	Hyperbola (not Rectangular)	*	$8x^2 + 10xy - 3y^2 - 2x - 4y - 2 = 0$

When faced with a general second degree equation in x and y, I often find it convenient right at the start to calculate the values of the cofactors from equations 2.7.10 – 2.7.15.

Here is an exercise that you might like to try. Show that the ellipse $ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0$ is contained within the rectangle whose sides are

$$x = \frac{\bar{g} \pm \sqrt{\bar{g}^2 - \bar{a}\bar{c}}}{\bar{c}} \qquad 2.7.18$$

$$y = \frac{\bar{f} \pm \sqrt{\bar{f}^2 - \bar{b}\bar{c}}}{\bar{c}} \qquad 2.7.19$$

In other words, these four lines are the vertical and horizontal tangents to the ellipse.

This is probably not of much use in celestial mechanics, but it will probably be useful in studying Lissajous ellipses, or the Stokes parameters of polarized light. It is also useful in programming a computer to draw, for example, the ellipse $14x^2 - 4xy + 11y^2 - 44x - 58y + 71 = 0$. To do this, you will probably want to start at some value of x and calculate the two corresponding values of y, and then move to another value of x. But at which value of x should you start? Equation 2.7.18 will tell you.

But what do equations 2.7.18 and 2.7.19 mean if the conic section equation $ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0$ is not an ellipse? They are still useful if the conic section is a hyperbola. Equations 2.7.18 and 2.7.19 are still vertical and horizontal tangents - but in this case the hyperbola is entirely outside the limits imposed by these tangents. If the axes of the hyperbola are horizontal and vertical, one or other of equations 2.7.18 and 2.7.19 will fail.

If the conic section is a parabola, equations 2.7.18 and 2.7.19 are not useful, because $\bar{c} = 0$. There is only one horizontal tangent and only one vertical tangent. They are given by

$$x = \frac{\bar{a}}{2\bar{g}} \qquad 2.7.20$$

and

$$y = \frac{\bar{b}}{2\bar{f}}\,. \qquad 2.7.21$$

If the axis of the parabola is horizontal or vertical, one or other of equations 2.7.20 and 2.7.21 will fail.

If the second degree equation represents one or two straight lines, or a point, or nothing, I imagine that all of equations 2.7.18 – 2.7.21 will fail - unless perhaps the equation represents horizontal or vertical lines. I haven't looked into this; perhaps the reader would like to do so.

Here is a problem that you might like to try. The equation $8x^2 + 10xy - 3y^2 - 2x - 4y - 2 = 0$ represents a hyperbola. What are the equations to its axes, to its asymptotes, and to its conjugate hyperbola? Or, more generally, if $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ represents a hyperbola, what are the equations to its axes, to its asymptotes, and to its conjugate hyperbola?

Before starting, one point worth noting is that the original hyperbola, its asymptotes, and the conjugate hyperbola) have the same centre, which means that *g* and *f* are the same for each, and they have the same axes, which means that *a*, *h*, and *b* are the same for each. *They differ only in the constant term.*

If you do the first problem, $8x^2 + 10xy - 3y^2 - 2x - 4y - 2 = 0$, there will be a fair amount of numerical work to do. When I did it I didn't use either pencil and paper or a hand calculator. Rather I sat in front of a computer doing the numerical calculations with a Fortran statement for every stage of the calculation. I don't think I could have done it otherwise without making lots of mistakes. *The very first thing I did* was to work out the cofactors $\bar{a}, \bar{h}, \bar{b}, \bar{g}, \bar{f}, \bar{c}$ and store them in the computer, and also the coordinates of the centre (x_0, y_0) of the hyperbola, which are given by $x_0 = \bar{g}/\bar{c}$, $y_0 = \bar{f}/\bar{c}$.

Whether you do the particular numerical problem, or the more general algebraic one, I suggest that you proceed as follows. First, refer the hyperbola to a set of coordinates x', y' whose origin coincides with the axes of the hyperbola. This is done by replacing x with $x' + x_0$ and y with $y' + y_0$. This will result in an equation of the form $ax'^2 + 2hx'y' + by'^2 + c' = 0$. The coefficients of the quadratic terms will be unchanged, the linear terms will have vanished, and the constant term will have changed. At this stage I got, for the numerical example, $8x'^2 + 10x'y' - 3y'^2 - 1.8163 = 0$.

Now refer the hyperbola to a set of coordinates x'', y'' whose axes are parallel to the axes of the hyperbola. This is achieved by replacing x' with $x''\cos\theta - y''\sin\theta$ and y' with $x''\sin\theta + y''\cos\theta$, where $\tan 2\theta = 2h/(a-b)$. There will be a small problem here, because this gives two values of θ differing by 90º, and you'll want to decide which one you want. In any case, the result will be an equation of the form $a''x''^2 + b''y''^2 + c' = 0$, in which a'' and b'' are of opposite sign. Furthermore, if you happen to understand the meaning of the noise "The trace of a matrix is invariant under an orthogonal transformation", you'll be able to check for arithmetic mistakes by noting that $a'' + b'' = a + b$. If this is not so, you have made a mistake. Also, the constant term should be unaltered by the rotation (note the single prime on the *c*). At this stage, I got $9.933x''^2 - 4.933y''^2 - 1.8163 = 0$. (All of this was done with Fortran statements on the computer - no actual calculation or writing done by me - and the numbers were stored in the computer to many significant figures).

In any case this equation can be written in the familiar form $\frac{x''^2}{A^2} - \frac{y''^2}{B^2} = 1$, which in this case I made to be $\frac{x''^2}{0.4283^2} - \frac{y''^2}{0.6088^2} = 1$. We are now on familiar ground. The axes of the hyperbola are $x'' = 0$ and $y'' = 0$, the asymptotes are $\frac{x''^2}{A^2} - \frac{y''^2}{B^2} = 0$ and the conjugate hyperbola is $\frac{x''^2}{A^2} - \frac{y''^2}{B^2} = -1$.

Now, starting from $\frac{x''^2}{A^2} - \frac{y''^2}{B^2} = 0$ for the asymptotes, or from $\frac{x''^2}{A^2} - \frac{y''^2}{B^2} = -1$ for the conjugate hyperbola, we reverse the process. We go to the single-primed coordinates by replacing x'' with $x'\cos\theta + y'\sin\theta$ and y'' with $-x'\sin\theta + y'\cos\theta$, and then to the original coordinates by replacing x' with $x - x_0$ and y' with $y - y_0$.

This is what I find:

Original hyperbola: $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$

Conjugate hyperbola: $ax^2 + 2hxy + by^2 + 2gx + 2fy + c_{\text{conj}} = 0$,
where $c_{\text{conj}} = -(2g\bar{g} + 2f\bar{f} + c\bar{c})/\bar{c} = -(2gx_0 + 2fy_0 + c)$.

Asymptotes: $ax^2 + 2hxy + by^2 + 2gx + 2fy + c_{\text{asymp}} = 0$,
where c_{asymp} can be written in any of the following equivalent forms:

$$c_{\text{asymp}} = +(a\bar{g}^2 + 2h\bar{g}\bar{f} + b\bar{f}^2)/\bar{c}^2 = ax_0^2 + 2hx_0y_0 + by_0^2 = -(g\bar{g} + f\bar{f})/\bar{c}.$$

[The last of these three forms can be derived very quickly by recalling that a condition for a general second degree equation in x and y to represent two straight lines is that the determinant Δ should be zero. A glance at this determinant will show that this implies that $g\bar{g} + f\bar{f} + c\bar{c} = 0$.]

Axes of hyperbolas: $(y - x\tan\theta - y_0 + x_0\tan\theta)(y + x\cot\theta - y_0 - x_0\cot\theta) = 0$,
where $\tan 2\theta = 2h/(a - b)$.

Example:

Original hyperbola: $8x^2 + 10xy - 3y^2 - 2x - 4y - 2 = 0$

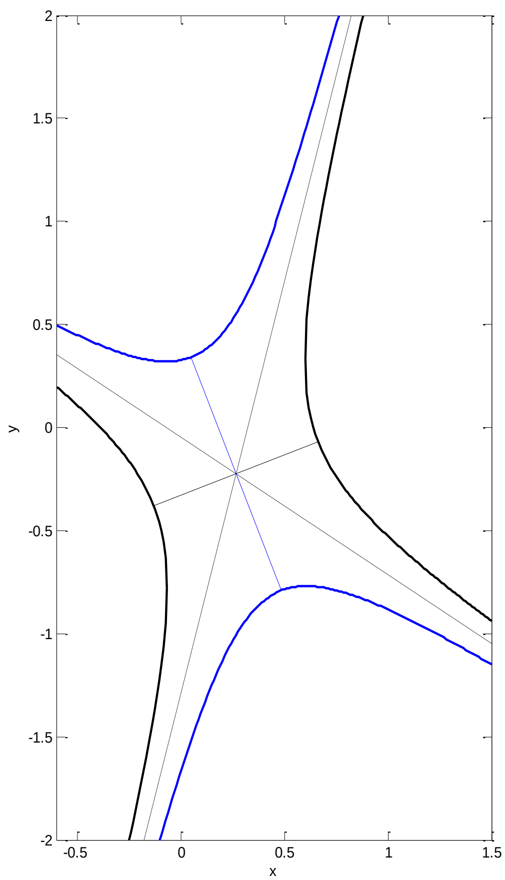
Conjugate hyperbola: $8x^2 + 10xy - 3y^2 - 2x - 4y + \frac{80}{49} = 0$

Asymptotes: $8x^2 + 10xy - 3y^2 - 2x - 4y - \frac{9}{49} = 0$,

which can also be written $(4x - y - \frac{9}{7})(2x + 3y + \frac{1}{7}) = 0$

Axes of hyperbolas: $(y - 0.3866x + 0.3275)(y + 2.5866x - 0.4613)$.

These are shown in the figure below - the original hyperbola in black, the conjugate in blue.



The centre is at (0.26531, -0.22449).

The slopes of the two asymptotes are 4 and $-\frac{2}{3}$. From equation 2.2.16 we find that the tangent of the angle between the asymptotes is $\tan 2\psi = \frac{14}{5}$, so that 2ψ = 70º.3, and the angle between the asymptote and the major axis of the original hyperbola is 54º.8, or tan ψ = 1.419. This is equal (see equations 2.5.3 and 2.5.10) to $\sqrt{e^2 - 1}$, so the eccentricity of the original hyperbola is 1.735.

From Section 2.2, shortly equation 2,5,6, we soon find that the eccentricity of the conjugate hyperbola is $\csc\psi = 1.223$.

An interesting question occurs to me. We have found that, if $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ is a hyperbola, then the equations to the conjugate hyperbola and the asymptotes are of a similar form, namely

$ax^2 + 2hxy + by^2 + 2gx + 2fy + c_{\text{conj}} = 0$ and $ax^2 + 2hxy + by^2 + 2gx + 2fy + c_{\text{asymp}} = 0$,

and we found expressions for c_{conj} and c_{asymp}. But what if $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ is not a hyperbola? What if it is an ellipse? What do the other equations represent, given that an ellipse has neither a conjugate nor asymptotes?

For example, $14x^2 - 4xy + 11y^2 - 44x - 58y + 71 = 0$ is an ellipse. What are $14x^2 - 4xy + 11y^2 - 44x - 58y + 191 = 0$ and $14x^2 - 4xy + 11y^2 - 44x - 58y + 131 = 0$?

I used the key on page 47, and it told me that the first of these equations is satisfied by no real points, which I suppose is the equation's way of telling me that there is no such thing as the conjugate to an ellipse. The second equation was supposed to be the "asymptotes", but the key shows me that the equation is satisfied by just one real point, namely (2 , 3), which coincides with the centre of the original ellipse. I didn't expect that. Should I have done so?

2.8 *Fitting a Conic Section Through Five Points.*

FIGURE II.38

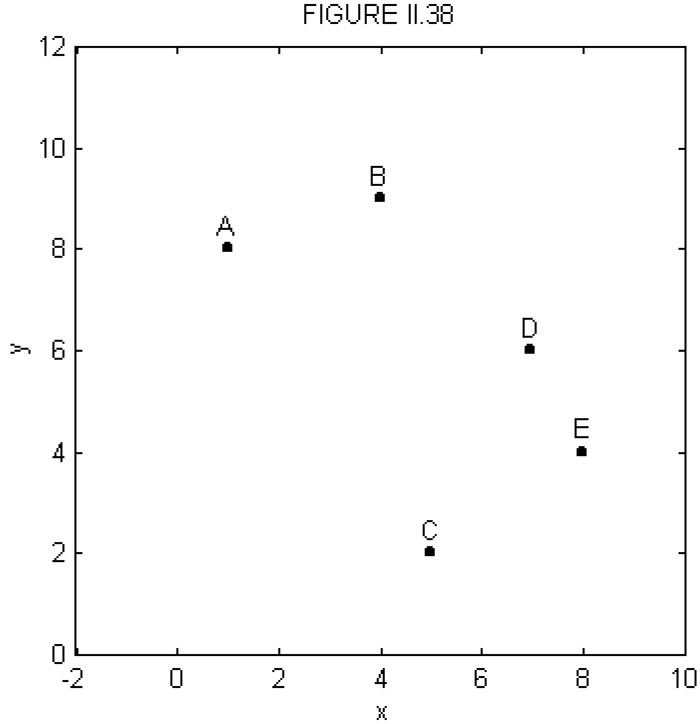


Figure II.38 shows the five points A(1,8), B(4,9), C(5,2), D(7,6), E(8,4). Problem: Draw a conic section through the five points.

The first thing to notice is that, since a conic section is of the form

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0 \qquad 2.8.1$$

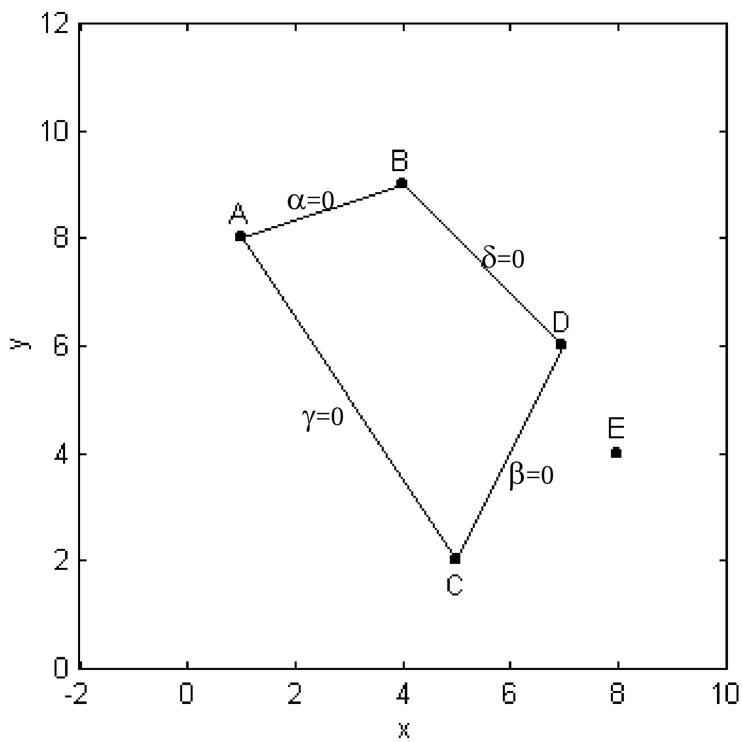
(remember that there is no loss of generality by taking the constant term to be 1), five points are necessary and sufficient to define a conic section uniquely. One and only one conic section can be drawn through these five points. We merely have to determine the five coefficients. The most direct (but not the fastest or most efficient) way to do this is to substitute each of the (x, y) pairs into the equation in turn, thus obtaining five linear equations in the five coefficients.

There is a better way.

We write down the equations for the straight lines AB, CD, AC and BD. Let us call these equations $\alpha = 0$, $\beta = 0$, $\gamma = 0$ and $\delta = 0$ respectively (figure II.39).

Then $\alpha\beta = 0$ is the equation that represents the two straight lines AB and CD, and $\gamma\delta = 0$ is the equation that represents the two straight lines AC and BD. The equation $\alpha\beta + \lambda\gamma\delta = 0$, where λ is an arbitrary constant, is a second degree equation that represents any conic section that passes through the points A, B, C and D. By inserting the coordinates of E in this equation, we can find the value of λ that forces the equation to go through all five points. This model of unclarity will become clear on following an actual calculation for the five points of the present example.

FIGURE II.39



The four straight lines are

$$\alpha = 0: \qquad x - 3y + 23 = 0$$

$$\beta = 0: \qquad 2x - y - 8 = 0$$

$$\gamma = 0: \qquad 3x + 2y - 19 = 0$$

$$\delta = 0: \qquad x + y - 13 = 0$$

The two pairs of lines are

$$\alpha\beta = 0: \qquad 2x^2 - 7xy + 3y^2 + 38x + y - 184 = 0$$

$$\gamma\delta = 0: \qquad 3x^2 + 5xy + 2y^2 - 58x - 45y + 247 = 0$$

and the family of conic sections that passes through A, B, C and D is

$\alpha\beta + \lambda\gamma\delta = 0$:

$$(2+3\lambda)x^2 - (7-5\lambda)xy + (3+2\lambda)y^2 + (38-58\lambda)x + (1-45\lambda)y - 184 + 247\lambda = 0.$$

Now substitute $x = 8$, $y = 4$ to force the conic section to pass through the point E. This results in the value

$$\lambda = \tfrac{76}{13}.$$

The equation to the conic section passing through all five points is therefore

$$508x^2 + 578xy + 382y^2 - 7828x - 6814y + 32760 = 0$$

We can, if desired, divide this equation by 2 (since all coefficients are even), or by 32760 (to make the constant term equal to 1) but, to make the analysis that is to follow easier, I choose to leave the equation in the above form, so that the constants f, g and h remain integers.

The constants have the values

$$a = 508, \; b = 382, \; c = 32760, \; f = -3407, \; g = -3914, \; h = 289$$

and the cofactors have the values

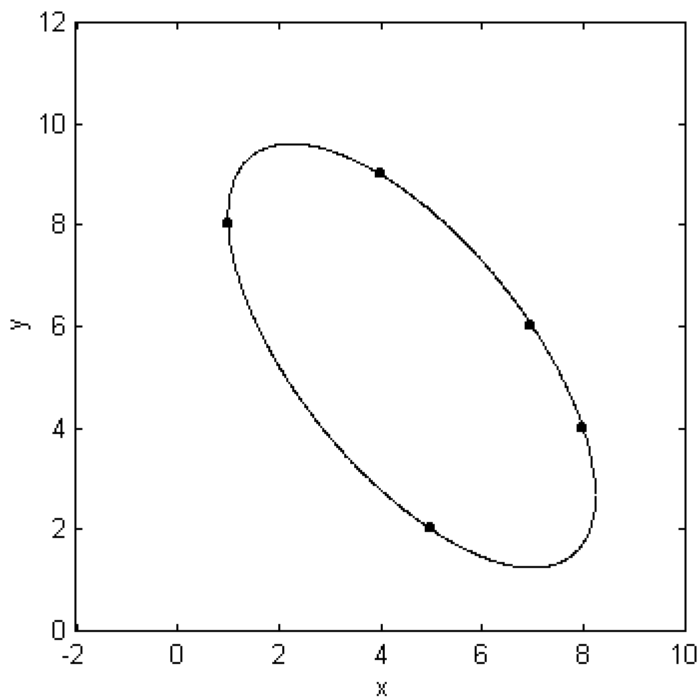
$\bar{a}$ = 906 671, $\bar{b}$ = 1 322 684, $\bar{c}$ = 110 535

$\bar{f}$ = 599 610, $\bar{g}$ = 510 525, $\bar{h}$ = 3 867 358

Let us consult the dichotomous table. The value of the determinant is $\Delta = a\bar{a} + h\bar{h} + g\bar{g}$ (or $h\bar{h} + b\bar{b} + f\bar{f}$, or $g\bar{g} + f\bar{f} + c\bar{c}$; try all three sums to check for arithmetic mistakes). It comes to $\Delta = -419\,939\,520$, so we proceed to option 5. $\bar{c} > 0$, so we proceed to option 6. a and Δ have opposite signs, so we proceed to 7. a does not equal b, nor is h equal to zero. Therefore we have an ellipse. It is drawn in figure II.40.

The centre of the ellipse is at (4.619, 5.425), and its major axis is inclined at an angle 128$^{\text{o}}$ 51' to the x-axis. If we now substitute x + 4.619 for x and y + 5.425 for y, and then substitute x cos 128$^{\text{o}}$ 51' + y sin 128^0 51' for the new value of x and $-x$ sin 128$^{\text{o}}$ 51' + y cos 128$^{\text{o}}$ 51' for the new value of y, the equation will assume its the familiar form for an ellipse referred to its axes as coordinate axes and its centre as origin.

FIGURE II.40



2.9 *Fitting a Conic Section Through n Points*

What is the best ellipse passing near to the following 16 points?

(1,50) (11,58) (20,63) (30,60)
(42,59) (48,52) (54,46) (61,42)
(61,19) (45,12) (35,10) (25,13)
(17,17) (14,22) (5,29) (3,43)

This is answered by substituting each point (x , y) in turn in the equation

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0, \qquad 2.9.1$$

thus obtaining 16 equations in the coefficients a, h, b, g, f. (The constant term can be taken to be unity.) These are the equations of condition. The five normal equations can then be set up and solved to give those values for the coefficients that will result in the sum of the squares of the residuals being least, and it is in that sense that the "best" ellipse results. The details of the method are given in the chapter on numerical methods. The actual solution for the points given above is left as an exercise for the energetic.

It might be thought that we are now well on the way to doing some real orbital theory. After all, suppose that we have several positions of a planet in orbit around the Sun, or several positions of the secondary component of a visual binary star with respect to its primary component; we can now fit an ellipse through these positions. However, in a real orbital situation we have some additional information as well as an additional constraint. The additional information is that, for each position, we also have a time. The constraint is that the orbit that we deduce must obey Kepler's second law of planetary motion - namely, that the radius vector sweeps out equal areas in equal times. We shall have to await Part II before we get around actually to computing orbits.