TRI PP 95 3
Feb 1995

Elementary derivation of Kepler's Laws

Erich Vogt

Physics Department, University of British Columbia, Vancouver, Canada
and
TRIUMF, 4004 Wesbrook Mall, Vancouver, Canada, V6T 2A3

Abstract

A simple derivation of all three so-called Kepler Laws is presented in which the orbits, bound and unbound, follow directly and immediately from conservation of energy and angular momentum. The intent is to make this crowning achievement of Newtonian Mechanics easily accessible to students in introductory physics courses. The method is also extended to simplify the derivation of the Rutherford Scattering Law.

(Submitted to the American Journal of Physics)

I. INTRODUCTION

The so-called Kepler's Laws of planetary motion have been of central interest for Newtonian Mechanics ever since the appearance of Newton's *Principia* [1]. They are discussed in most introductory textbooks of physics [2,3] and continue to be a subject of lively interest in the pages of the American Journal of Physics [4]. This interest is not surprising because the understanding of planetary motion has been one of the oldest challenges in many human cultures and continues to excite the sense of wonder among young scientists today.

The purpose of the present article is to give a new elementary derivation of all three of the Kepler Laws intended to make their physics accessible to first year university students taking introductory mechanics. I have used this derivation in my own introductory classes for more than a decade and find that it, and the many associated problems, are a highlight of the introduction which I give to physics. In contrast, most first-year textbooks give a description of Kepler's Laws but apparently regard their derivation as too difficult. Perhaps the derivation given here can then fill an important gap.

The elementary proof, given in the next section, follows directly, in a few easy steps, from conservation of energy and angular momentum which, in turn, follow from $F = ma$ and the central nature of the universal gravitational force, $F = GmM/r^2$. These conservation laws, on which we build, are usually covered thoroughly, and often even elegantly, in first year textbooks.

In succeeding sections, beyond the proof, we provide further discussion of bound elliptic orbits and extend the treatment to the unbound Kepler orbits and to the Rutherford Scattering Law.

II. ELEMENTARY PROOF OF KEPLER'S LAWS

a) Kepler's First Law (The Law of Orbits): All planets move in elliptical orbits having the Sun at one focus.

For a planet of mass m in a bound orbit(negative total energy E), around the sun of mass M, we have the constant total energy, E

$$E \equiv mv^2/2 - GMm/r \ , \tag{1}$$

where r is the distance of the planet from the sun and v its velocity. $(-E/m)$ is a positive constant of the motion. Because the force is central we also have conserved angular momentum, ℓ

$$\ell \equiv mvh \ , \tag{2}$$

where $h(\equiv r\,sin\phi$, with ϕ the angle between $\mathbf{v}$ and $\mathbf{r}$) is the perpendicular distance from the planet's instantaneous velocity vector to the sun (see Fig. 1). From the definition of h we have $h \leq r$. (ℓ/m) is also a positive constant of the motion. Using Eq. (2) in Eq. (1) we obtain

$$\frac{[(\ell/m)^2/2(-E/m)]}{h^2} - \frac{[GM/(-E/m)]}{r} = -1 \ . \tag{3}$$

An 'orbit' equation is, in general, a relationship between two independent variables. Equation (3) is an orbit equation connecting two independent coördinates, r & h. What kind of an orbit? It is exactly an ellipse because the standard ellipse equation, transformed into the two variables r & h is (see Sec. 3)

$$\frac{b^2}{h^2} - \frac{2a}{r} = -1 \ (h/ \leq r) \ , \tag{4}$$

where a is the semi-major axis of the ellipse and b is the semi-minor axis. The equality of Eq. (3) and Eq. (4) not only completes the proof of Kepler's First Law, but also immediately gives the orbit parameters, a & b, in terms of the constants of the motion, $(-E/m)$ and (ℓ/m)

$$a = GM/2(-E/m) \ , \ b = (\ell/m)/(-2E/m)^{\frac{1}{2}} \ . \tag{5}$$

b) Kepler's Second Law (The Law of Areas): A line joining planet to the Sun sweeps out equal areas in equal times.

This law is the only one of the three commonly proved in introductory physics textbooks. Referring to Fig. 1, the time derivative of the area, A, swept out is

$$dA/dt = \frac{1}{2}vh = \ell/2m \ , \tag{6}$$

which is constant. Thus this law is directly associated with the conservation of angular momentum.

c) Kepler's Third Law (The Law of Periods): The square of the period of any planet about the Sun is proportional to the cube of the planet's mean distance from the Sun.

Using the second law the period, T, of the planet must be equal to the total area of the ellipse, divided by the constant, dA/dt. The total area of an ellipse is πab. Therefore,

$$T = 2\pi ab/(\ell/m) = (2\pi/\sqrt{GM})a^{\frac{3}{2}}$$
$$\text{or } T^2 = (4\pi^2/GM)a^3 \ . \tag{7}$$

Interpreting the semi-major axis as the mean distance from the sun, the result Eq. (7) proves the third law. The constant, $(4\pi^2/GM)$, which applies to all planets is, about, $3.0 \times 10^{-34} \, y^2/m^3$.

III. SOME PROPERTIES OF ELLIPSES

Referring to Fig. 1, the familiar equations for the ellipse relate to the coördinates x & y,

$$x^2/a^2 + y^2/b^2 = 1 \ , \tag{8}$$

or, alternatively, to the polar coördinates r & θ,

$$r = a(1-\epsilon^2)/(1+\epsilon\cos\theta) , \tag{9}$$

where the eccentricity, ϵ, is defined by $c \equiv (a^2-b^2)^{\frac{1}{2}} \equiv \epsilon a$, with c being the distance from the center of the ellipse to its focus.

Next we provide the derivation, beginning with Eq. (7) of the unfamiliar form of the ellipse, Eq. (4), required for the proof of Sec. 2. The coördinate r, shown on Fig. 1, is defined by

$$r \equiv [y^2+(x-c)^2]^{\frac{1}{2}} . \tag{10}$$

We substitute for y^2 from Eq. (7),

$$y^2 = b^2(1-x^2/a^2) = (1/a^2)(a^2-c^2)(a^2-x^2) ,$$

to obtain

$$r = a^{-1}(a^2-cx) , \tag{11}$$

which is also an equation for the ellipse in terms of the coördinates r & x.

To find h in terms of x (or y) we start with the general formula for the perpendicular distance, h, from an arbitrary point (x_1, y_1) to an arbitrary straight line, $y = y'x + y_0$, where $y'(\equiv dy/dx)$ is the slope of the line and y_0 is its y intercept. The formula is

$$h = (1+y'^2)^{-\frac{1}{2}}(y_0-y_1+y'x_1) . \tag{12}$$

Here $(x_1, y_1) \equiv (c, 0)$ and the y-intercept is $y_0 = y - y'x$. The straight-line (see Fig. 1) is the tangent to the ellipse at the position (x, y) of the planet so that

$$y' \equiv dy/dx = -(x/y)(b^2/a^2) , \tag{13}$$

which follows directly from Eq. (7). Substituting into the square of Eq. (12) for y', y_0, y_1 and x_1, we obtain

$$\begin{aligned} h^2 &= (1+x^2b^4/y^2a^4)^{-1}(y+x^2b^2/ya^2-xcb^2/ya^2)^2 \\ &= (y^2a^4+x^2b^4)^{-1}(y^2a^2+b^2x^2-b^2xc)^2 \\ &= b^2[a^4(1-x^2/a^2)+x^2(a^2-c^2)]^{-1}[a^2(1-x^2/a^2)+x^2-xc]^2 \\ &= b^2(a^4-x^2c^2)^{-1}(a^2-xc)^2 \\ &= b^2(a^2+xc)^{-1}(a^2-xc) , \end{aligned} \tag{14}$$

which, again, is an equation for the ellipse in terms of the coördinates h & x.

Solving Eq. (10) for x yields

$$x = (a^2/c)(1-r/a) , \tag{15}$$

and similarly, solving Eq. (13) for x yields

$$x = (a^2/c)(1-h^2/b^2)/(1+h^2b^2) . \tag{16}$$

Finally, equating Eq. (13) and Eq. (14) yields the desired ellipse Eq. (4).

From a teaching point of view the present approach to the Kepler problem works so well because the physics content of the proofs is all contained, as in Sec. 2 above, in a few simple statements pertaining to the conservation of energy and angular momentum. The complications, such as they are, occur only in the derivation of the unfamiliar form of the ellipse equation. Even this derivation involves straightforward mathematics which, in my experience, is familiar to students taking introductory physics courses. In order not to deflect from the physics interest it is my practise to present the derivation given in this section as a handout intended to help those students who wish to know more about ellipses and who want to relate the unfamiliar form of Eq. (4) to something that they know.

It is interesting to plot the ellipse pairs of coördinates other than the usual x & y of Fig. 2. For example, the x & r ellipse of Eq. (11) or Eq. (15) is a straight line for which the elliptical motion lies between the maximum and minimum values of r, that is, between $(1-\epsilon)a$ and $(1+\epsilon)a$. Similarly, for the x & h equation the ellipse lies between the limits $(1-\epsilon)a \leq h \leq (1+\epsilon)h$, or correspondingly, $-a \leq x \leq a$. The h & r ellipse, Eq. (4), which is of importance to us in this paper, is shown on Fig. 2, for two values of the eccentricity, $\epsilon = 0$ and $\epsilon = 0.8660$. The latter corresponds to a choice of $b = a/2$, which was also the choice for the ellipse of Fig. 2. By the definition of h, the only values of h which can have any physical meaning are those for which $h \leq r$. Thus the elliptical motion takes place in the half quadrant for which, $h \leq r$, that is, below the dashed line of Fig. 2. For $\epsilon = 0$ we have a circle and, indeed, on Fig. 2, the only physical point is $h = r = a$.

IV. UNBOUND KEPLER ORBITS

It is well known that when the total energy E is a positive constant the orbit of the Kepler problem is hyperbolic. This fact is often stated in introductory physics texts. We now prove it by the same simple methods used for elliptic orbits above.

For positive E we rewrite Eq. (3) as

$$\frac{[(\ell/m)^2/2(E/m)]}{h^2} - \frac{[GM/(E/m)]}{r} = 1 . \tag{17}$$

For this orbit equation we note that the relevant hyperbola, shown on Fig. 3, obeys the equation

$$\frac{b^2}{h^2} - \frac{2a}{r} = 1 , \tag{18}$$

which can be contrasted with the ellipse, Eq. (4). Again, the equality of Eq. (17) and Eq. (18) proves that the orbit is hyperbolic and gives the orbit parameters, a & b, to be

$$a = GM/2(E/m) , \quad b = (\ell/m)/(2E/m)^{\frac{1}{2}} . \tag{19}$$

To complete the proof we must derive the hyperbola Eq. (18), in terms of the coördinates r & h, from the usual equation of a hyperbola, in terms of the coördinates x & y:

$$x^2/a^2 - y^2/b^2 = 1 . \tag{20}$$

In the (x, y) plane of Fig. 3 the hyperbola lies between two asymptotes whose directions are determined by the choice of b & a. (For the hyperbola illustrated in Fig. 3 we have chosen $b = a/2$, as we did for the ellipse of Fig. 2 above.) There are, of course, two equal hyperbolas, both of which satisfy Eq. (20). We have drawn only the left-side hyperbola on Fig. 3, and not its image mirrored in the y-axis.

This choice between the two hyperbolas is arbitrary but the choice for a system of hyperbola plus focus is not. When we refer to the physics we place the center-of-force at the focus: $(x, y) = (-c, 0)$ or $(+c, 0)$ with $c^2 \equiv a^2 + b^2$. For the left hyperbola shown on Fig. 3 the choice of $(-c, 0)$ as the focus corresponds to gravitational attraction - the orbit is "pulled around" the center-of-force. Placing the focus at the other position, $(+c, 0)$, would correspond to the unphysical orbit of antigravity - with the orbit "pushed away" from the center-of-force. Although this latter case has no relevance to the Kepler problem it does provide the orbit for Rutherford scattering (see Sec. 6, below). The orbits for the left-side hyperbola, with the two possible foci both obey the same hyperbola equation in the x & y coördinates, that is, Eq. (20), but for the r & h coördinates the two cases obey different equations. We are interested here in the case of the Kepler problem, with the left-side hyperbola taken together with the left-side focus.

To begin the proof of Eq. (18) we note that from Eq. (20) we have

$$y = b(x^2/a^2 - 1)^{\frac{1}{2}} , \tag{21}$$

$$y' \equiv dy/dx = (b^2/a^2)(x/y) , \tag{22}$$

and the y-intercept for the tangent line at point P is

$$y_0 = y - y'x . \tag{23}$$

Choosing the left-side focus, $(x_1, y_1) = (-c, 0)$ we find at once, in place of Eq. (10)

$$\begin{aligned} r &\equiv [y^2 + (x + c)^2]^{\frac{1}{2}} \\ &= -a^{-1}(a^2 + xc) . \end{aligned} \tag{24}$$

Similarly we find, using Eq. (12)

$$h^2 = -b^2(a^2 + xc)/(a^2 - xc) , \tag{25}$$

which can be compared to Eq. (13) for the ellipse. Combining Eq. (24) and Eq. (25) yields the desired hyperbola, Eq. (18).

V. ACCESSIBLE ORBIT PROBLEMS

With the simple relationship between the orbit parameters and the constants of the motion, derived above, a myriad of interesting problems can immediately be tackled by the students. The point is that any two pieces of information pertaining to a, b, c, ℓ/m, E/m, T etc., completely specify the orbit. A few examples are:

i) If you fire a cannonball horizontally at the North Pole with an initial velocity of $v = 0.98 \times 10^4$ m/s, sketch the orbit and find the orbit parameters. (Assume the earth is spherically symmetric and neglect air friction.) Find the period of the motion. The solution for this problem is an elliptical orbit whose major axis lies along the earth's axis and is greater than the earth radius.

ii) If your friendly computer "Hal" launches you from your spaceship into outerspace at a distance from the Sun of 3.1×10^{11} m, with a speed of 8.2×10^{4} m/s and a direction of motion such that your perpendicular distance is 1.86×10^{11} m from the Sun, find out what will be your distance of closest approach to the sun. Also, find out if your orbit is bound or unbound.

iii) If a comet were to strike the earth in such a way that its orbital velocity instantly increased by 10% but the direction of the velocity remained unchanged by the collision, find the effect on the earth's orbit (originally assumed to be circular) and its period.

Further, a great deal of celestial mechanics becomes accessible and transparent.

VI. THE RUTHERFORD SCATTERING LAW

In spite of its importance the Rutherford Scattering Law is not a subject normally covered in introductory physics textbooks, perhaps because the concepts of cross sections are usually optional or omitted. Indeed, the concept of a differential cross section, needed for the Rutherford Law is quite sophisticated for a first year course. However, because of my own personal predilections in physics I like to say more about atomic and nuclear cross sections in my introductory class than is the standard fare. When I then also give a full treatment, as above, of the Kepler Laws it is very tempting to go further and derive the Rutherford Scattering Law. The connection between the crowning achievement of Newtonian Mechanics and the foundations of modern subatomic physics is very compelling. This is an approach to which I was led by the PSSC courses of two decades ago and which has been admirably presented by French in his textbook, Newtonian Mechanics [3].

The purpose of this section is to show that the derivation of the Rutherford Law benefits fully from the simplification of the Kepler Laws introduced above. If the Coulomb force is $F = kQ_1Q_2/r^2$, in an obvious notation, then the total energy, E, of an alpha particle in its orbit around a gold nucleus is given by

$$E = mv^2/2 + kQ_1Q_2/r \ . \tag{26}$$

Using the conserved angular momentum, $\ell \equiv mvh$, we find, instead of Eq. (3)

$$\frac{\ell^2/2mE}{h^2} + \frac{kQ_1Q_2/E}{r} = 1 \ . \tag{27}$$

This orbit is that for the Kepler problem with a repulsive force. We have the left-side hyperbola with the right-side focus, as shown on Fig. 3, where the coördinates r' & h' are also indicated. In terms of r' & h' the hyperbola equation is

$$\frac{b^2}{(h')^2} + \frac{2a}{r'} = 1 \ , \tag{28}$$

which is to be compared with the ellipse Eq. (4) and Eq. (18) which pertains to the hyperbola-focus system for attractive forces. Again, comparing Eq. (28) and Eq. (27) gives us the orbit parameters:

$$a = kQ_1Q_2/2E \ , \quad b = (\ell^2/2mE)^{\frac{1}{2}} \ . \tag{29}$$

The proof of Eq. (28) follows closely the derivation of Sec. 4 above. Equations (21), (22) and (23) still apply. However, with the choice of $(x_1, y_1) = (+c, 0)$ we find

$$r' = a^{-1}[a^2 + xc] , \tag{30}$$

$$\text{and} \quad (h')^2 = -b^2(a^2 - xc)/(a^2 + xc) , \tag{31}$$

which are to be compared with the corresponding variables of Sec. 3 and Sec. 4. Combining Eq. (30) and Eq. (31) now yields Eq. (28).

When the alpha particle is far from the gold nucleus the potential energy vanishes and therefore $E = T_\alpha$, the initial kinetic energy of the alpha particle. The orbit parameter b is in fact, from Fig. 3, the usual "impact parameter". From the geometry of Fig. 3 we see that

$$b = a \cot(\theta/2) , \tag{32}$$

where θ is the scattering angle, and then

$$db/d\theta = (a/2)\sin^{-2}(\theta/2) . \tag{33}$$

To complete the derivation of the Rutherford Scattering Law we need to introduce the definitions pertaining to the differential scattering cross section, $d\sigma/d\Omega$. Here we follow the conventional treatment whose elements are the following. The partial cross section element, $d\sigma$, is defined to be proportional to the fraction of alpha particles whose impact parameters lie in between b & $b + db$

$$d\sigma \equiv 2\pi\, b db , \tag{34}$$

$d\Omega$ is the area on the unit sphere between θ and $\theta + d\theta$

$$d\Omega \equiv 2\pi \sin\theta d\theta = 4\pi \sin(\theta/2)\cos(\theta/2) d\theta . \tag{35}$$

Therefore, using Eq. (20)

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{b}{2\sin(\theta/2)\cos(\theta/2)} \frac{db}{d\theta} \\ &= a^2/[4\sin^4(\theta/2)] = [k Q_1 Q_2 / 4T_\alpha \sin^2(\theta/2)]^2 , \end{aligned} \tag{36}$$

which is the Rutherford Scattering Law.

VII. CONCLUSION

A derivation is given of Kepler's Laws in which the physics is easy and immediate. Any complications reside in the mathematics of ellipses and even these are well within the grasp of the students who have, in the past decade or more, come into my introductory physics class. Therefore, the derivations presented here fulfilled their intent of making the whole Keplerian problem easily accessible to physics students in first year. The intellectual payoff is large for the effort involved and that is the essence of introducing physics to willing students.

VIII. REFERENCES

[1] Newton, I., "Principia" (1687). F. Cajori's revision of A. Motte's translation (1929), University of California Press, Berkeley, California (1960).

[2] Examples are: Resnick, R., Halliday, D., and Krane, K.S. "Physics" 4th Edition, John Wiley and Sons, New York, N.Y. (1992) or Benson, H., "University Physics", John Wiley and Sons, New York, N.Y. (1991).

[3] French, A.P. "Newtonian Mechanics", W.W. Norton and Co., New York, N.Y. (1971). This excellent textbook gives a very thorough analysis of the whole Kepler problem including many interesting historical facts.

[4] Sivardière, J., Am. J. Phys., 56, 132–135 (1988) which contains many earlier references. See also W. Hauser, Am. J. Phys., 53, 905–907 (1985). Almost every volume of Am. J. Phys. contains several articles pertaining to the Kepler problem.

FIGURE CAPTIONS

Fig. 1. The geometry for the bound elliptical orbit of a planet at point P around the Sun at the focus. The ellipse parameters (a, b and c) are shown as well as three alternative pairs of coördinates: x & y, r & θ, r & h, where h is the perpendicular distance from the focus to the tangent at point P. The ellipse shown has $b = a/2$.

Fig. 2. The elliptical orbit, Eq. (4), in terms of the coördinates r & h for two different values of the eccentricity: $\epsilon = 0$ (a circle) and $\epsilon = 0.8660$ (the ellipse of Fig. 1 for which $b = a/2$). Since $h \leq r$ the elliptical motion pertains to that part of the curve which lies in the lower half of the quadrant, that is, below the dashed line of $r = h$.

Fig. 3. The geometry of the hyperbola pertaining to unbound Kepler orbits of planets and to the orbit of alpha particles in Rutherford scattering. The hyperbola (heavy line) is confined between two asymptotes. For the unbound Kepler orbit the Sun is at the focus $(x, y) \equiv (-c, 0)$ and the orbit parameters a, b & c are indicated. As in Fig. 1, above, we have chosen $b = a/s$. For the planet at point P its distance, r, from the Sun as well as the perpendicular distance, h, to its tangent line are also shown. For alpha-nucleus scattering the same hyperbola applies but the nucleus is at the other focus, $(x, y) \equiv (c, 0)$ for which the alpha-nucleus distance is r' and the perpendicular distance to the tangent line is h'. r' & h' are indicated on the figure as well as the scattering angle θ.

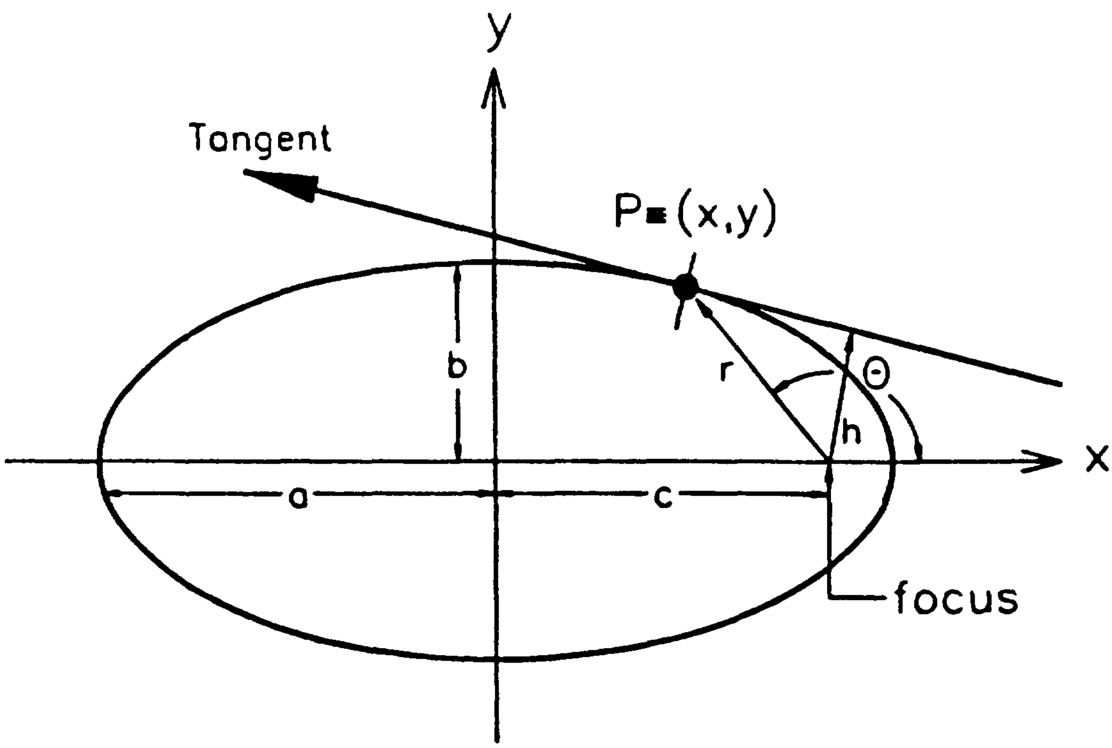
y
Tangent
P≡(x,y)
b
r
Θ
h
x
a
c
focus

Fig. 1

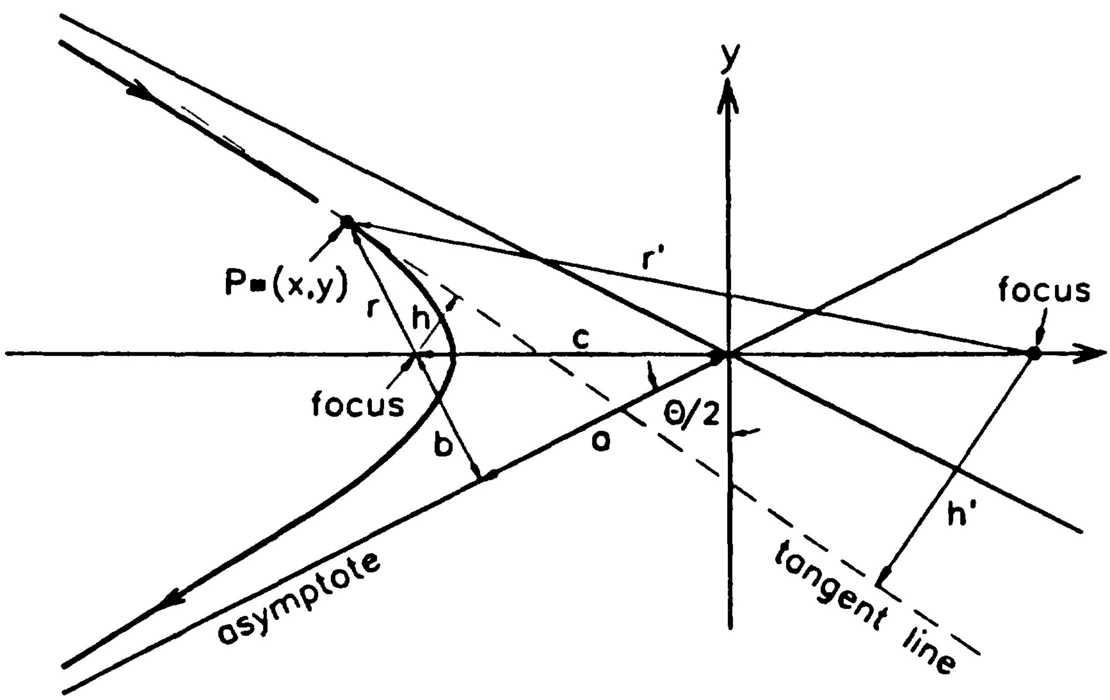
y
r'
P≡(x,y)
focus
r
h
c
focus
Θ/2
b
a
h'
asymptote
tangent line

Fig. 3

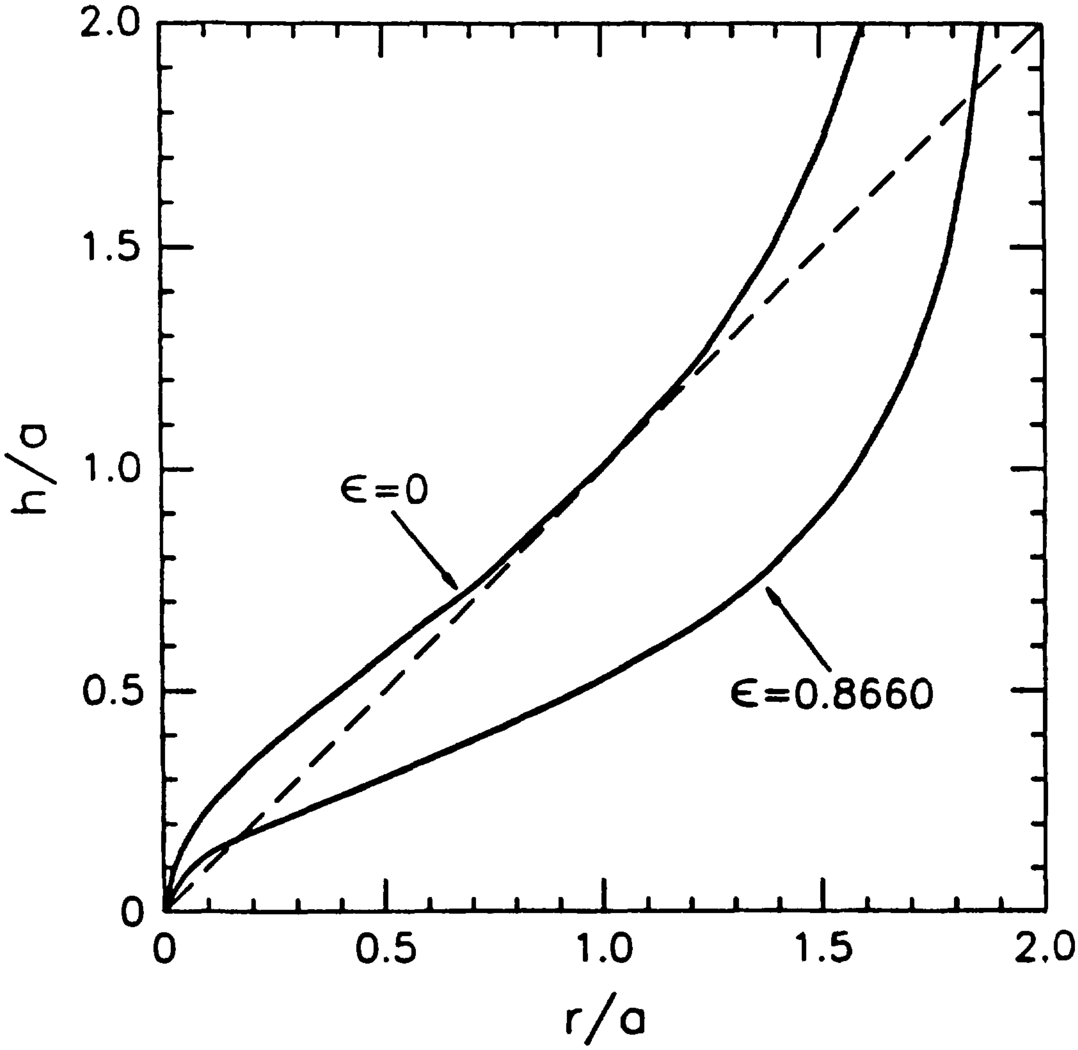

Fig. 2