Newton and the Transmutation of Force

Tristan Needham

1. INTRODUCTION. The year 1687 witnessed an event that was to dramatically alter the course of science, indeed, of western civilization—the publication of Sir Isaac Newton's *Philosophiae Naturalis Principia Mathematica* [**1**], better known simply as the *Principia*. Ten years ago, having been intrigued by the mathematical glimpses contained in Westfall's magnificent biography [**2**], I took up in earnest the study of Newton's great work, and the experience was as dazzling as anything one might hope to encounter on the road to Damascus.

The primary purpose of the paper is to present a new geometric way of understanding a beautiful but little-known fact concerning the transmutation of central force fields by means of complex mappings. A second purpose, however, is to swell the Newtonian congregation! To this end, in this introduction we shall supply five very elementary examples (suitable for the classroom) of the power of the *Principia*'s methods. This should also help to eliminate the potential 'culture shock' presented by the chosen methodology of the rest of the paper. First, though, a few general remarks on the *Principia* are perhaps in order. Since no mention will be made in this section of the principal problem to be analyzed, the impatient reader is welcome to jump to the next section where the work itself is begun.

As illustrated in [**3**], one motivation for the study of the *Principia* is that *unknown results* of importance to modern mathematics may thereby be revealed. For example, in his splendid little book on seventeenth century science [**4**], V. I. Arnol'd devotes an entire chapter to the *Principia*'s astonishing but neglected Lemma XXVIII ([**1**], p. 110). As he explains, this is in fact a brilliant topological proof of a result on the transcendence of Abelian integrals!

No less exciting than Newton's results, is the *method* he used to obtain them. Arnol'd puts it well: referring to Prop. LXX ([**1**], p. 193), "This sample of Newton's argument shows how it is possible to solve problems from potential theory without analysis, without knowing the theory of harmonic functions, or the fundamental solution of the Laplace equation, or the simple and double layer potentials. Similar arguments, preceding the rise of analysis, often occur in papers of those times and turn out to be very powerful" ([**4**], p. 27). The expression "preceding the rise of analysis" is rather curious in this context, for Newton had himself invented the calculus some twenty years before constructing the geometric argument to which Arnol'd refers. This observation leads us to the following famous 'paradox.'

Why did the inventor of the calculus not use it in the *Principia*? The answer—according to a persistent and pernicious myth—is that he *did* use calculus to make his great discoveries, but then disguised the fact (for Machiavellian reasons) by translating his arguments into geometry. In the hope that the wide readership of the *Monthly* may help to put this three-century old story to rest, let us invoke the authority of Westfall ([**2**], p. 424):

> No such papers demonstrating propositions of the *Principia* in a form different from that published have ever been found, except for a few in which he later set a couple of propositions over into analytical terms. The problem vanishes, however, when we view the *Principia* against the background of Newton's mathematical development in the years immediately preceding. Around 1680, the study of ancient geometry led Newton to a revulsion from the inelegant demonstrations of modern analysis.

We now turn to the actual nature of Newton's geometrical calculus. Ten years ago, immediately upon seeing what Newton was up to, I began to try my own hand at using his method. As the examples I worked out at that time are much more simple-minded than those in the *Principia*, let me now use four of them (plus one by Arnol'd) to illustrate Newton's approach. I make no claim to originality in what follows; on the contrary, I feel certain that such examples were well known to Newton himself.

(1) Let $T = \tan\theta$ and suppose we wish to explain why $dT/d\theta = 1 + T^2$. FIG. 1 illustrates the increase ΔT $(= cr)$ in T that results from an increase of $\Delta\theta$ in θ. With a as center, draw the circle through c, and let the tangent to this circle at c cut ar in s, thereby producing the equal angles cab and rcs. The desired result follows almost instantly upon observing the behavior of the triangle crs in the limit as $\Delta\theta \rightsquigarrow 0$. Already in FIG. 1 it is hard to distinguish between cs and the arc $L \cdot \Delta\theta$, and indeed at the very end of the shrinking process their "last ratio" (as Newton would say) will be unity. Thus,

$$\frac{cr}{cs} \rightsquigarrow \frac{1}{L}\left(\frac{dT}{d\theta}\right).$$

But in the shrinking process it is also clear that the triangle crs becomes similar to acb, and hence an alternative expression to the above is

$$\frac{cr}{cs} \rightsquigarrow \frac{L}{1}.$$

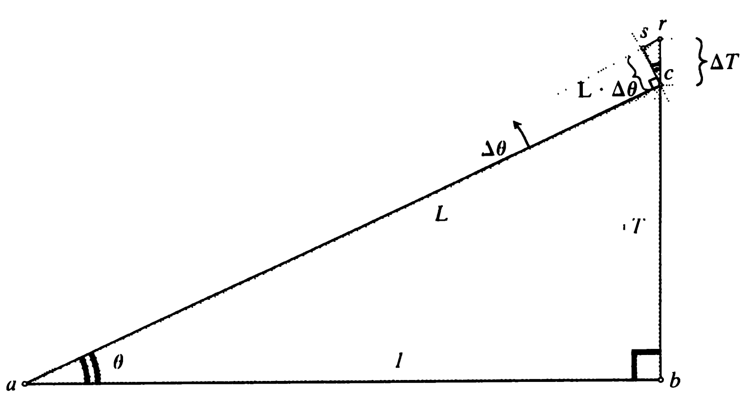

Figure 1.

Combining these two views of the limit process, we conclude that

$$\frac{dT}{d\theta} = L^2 = 1 + T^2.$$

In the firm belief that the most persuasive advocate of the Newtonian approach will be you yourself, we present the remaining examples as problems.

(2) Reconsider Fig. 1. Instead of looking at triangles with constant unit base, consider those of constant unit hypotenuse. By drawing the diagram for the new case, deduce that $(\sin\theta)' = \cos\theta$ and $(\cos\theta)' = -\sin\theta$. Note the economy with which one diagram yields both results.

(3) In the xy-plane, consider the family of triangles bounded by the coordinate axes and a variable line L through the fixed point $Q(a,b)$. See Fig. 2(A). By drawing the change in area resulting from a tiny rotation of L about Q, find the minimum area of such a triangle. Note the generality of the Newtonian reasoning. For example, consider the shaded area in Fig. 2(B) that is cut off from the ellipse by L as it rotates about the arbitrary interior point Q. Deduce that the area will be minimum [maximum, upon rotation of L by π] when L is parallel to the tangent at P.

(4) Consider Fig. 2(C). A basic property of ellipses that we shall need later in the paper is that the sum of the focal distances (F_1P and F_2P) is constant, and hence equal to the major axis. By drawing a diagram of the changes in the focal distances that result when P is minutely displaced along the ellipse, deduce the 'reflection property': light emitted from F_1 is reflected to F_2. In like manner, deduce the corresponding reflection properties of the other two conic sections.

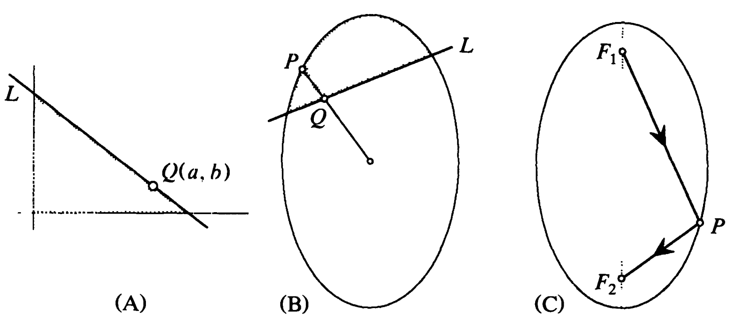

Figure 2.

(5) Arnol'd [**4**], p. 28: "Here is an example of a problem that people like Barrow, Newton and Huygens would have solved in a few minutes and which present-day mathematicians are not, in my opinion, capable of solving quickly [The only exception I know—G. Faltings—proves the rule.]: to calculate

$$\lim_{x\to 0} \frac{\sin(\tan x) - \tan(\sin x)}{\arcsin(\arctan x) - \arctan(\arcsin x)}."$$

We shall add a hint that Arnol'd does not supply: think what the graphs of the four functions must look like (*roughly*) near the origin. The answer is given on p. 108 of [**4**].

Certainly these problems were not chosen at random, but I trust that a point has nevertheless been made. Having been sensitized to the possibility of such solutions, you will no doubt find many other examples of your own whenever you next have occasion to teach calculus.

Although I derived much pleasure and knowledge from the application of the *Principia*'s ideas to problems of ordinary calculus, the real reward came with the gradual realization that these ideas could be applied, very naturally, to *complex* analysis, yielding an approach that I hold to be considerably more elementary and intuitive than the conventional one. These ideas have now taken on the definite form of a forthcoming book from Oxford University Press, entitled *Visual Complex Analysis*. The following is essentially an expanded extract from that work.

2. THE TALE OF THE POUND NOTE. In commemoration of the three-hundredth anniversary of the publication of the *Principia* in 1687, the Bank of England brought out the new pound note shown in FIG. 3. The diagram that it bears is indeed fitting, for it is a faithful reproduction of that found in Prop. XI ([**1**], p. 56), and it represents Newton's solution to the problem that catalyzed his entire endeavor in the *Principia*: to mathematically link the observed motion of the planets—which Kepler had found to be elliptical—to a solar attraction diminishing as the square of the distance.

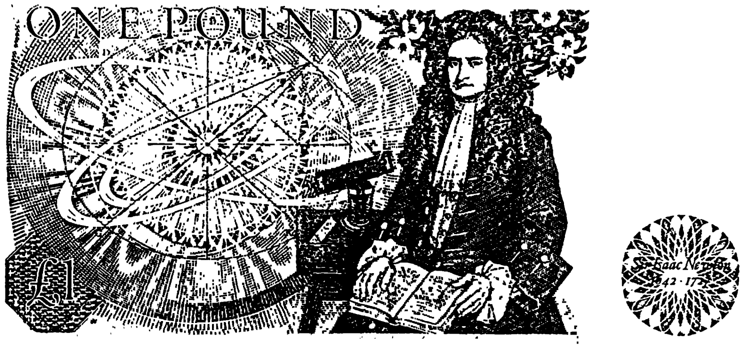

Figure 3.

Unhappy (presumably) with the baldness of Newton's figure, the decision was taken to embellish the pound note with what is clearly intended to represent the sun and some attendant bodies in orbit. But in doing so an extraordinary blunder was committed, for instead of being at the focus of the ellipse, this sun sits squarely at its *center*! Here our tale begins, for a closer examination of the *Principia* will reveal that there exists a point of view from which this blunder ceases to be a blunder at all.

Early in his investigations, Newton realized—even leaving aside their physical importance—that two special power laws enjoyed a degree of mathematical elegance not shared by the others: force decreasing as the square of the distance, and force increasing directly as the distance. He further observed that these "two principal cases" (as he called them) exhibited striking similarities. For example, only for these two laws are the orbits conic sections. This fact inspired Newton to go further.

Immediately following the direct proof employing FIG. 3, he gave an ingenious alternative argument in which he showed that the inverse square law for the force directed to the sun at the focus is a *consequence* of the linear force law that would be required if the elliptical orbit were instead caused by a fictitious attracting body at the center. In the light of this fact, the artist's embellishment no longer seems out of place!

Newton couched his argument in terms of a more general result which we will state now and prove later. Consider FIG. 4. A force field F_{OLD} emanating from A holds P in a given orbit. Newton asks the question, if the center of attraction is moved from A to any other place B, then into what new force field F_{NEW} must F_{OLD} be transmuted in order that the orbit of P remain the same? Draw AQ parallel to BP and meeting the tangent at P in Q. Newton's answer (Cor. III, Prop. VII) is

$$F_{\text{NEW}} \propto \left(\frac{AQ^3}{BP^2 \cdot AP}\right) F_{\text{OLD}}. \tag{1}$$

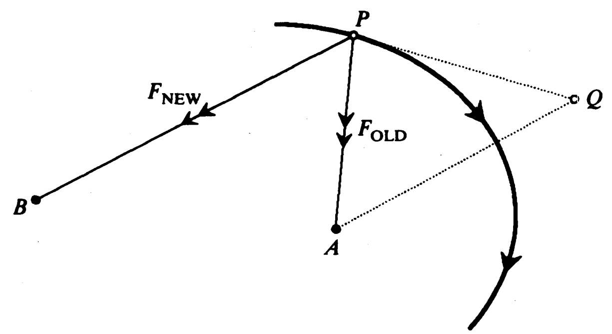

Figure 4.

Let us now follow Newton's beautiful application of this result to the transmutation of a linear force field into an inverse square field. Consider FIG. 5, in which the body at P is shown orbiting the sun at the focus B, and in which SC and RA

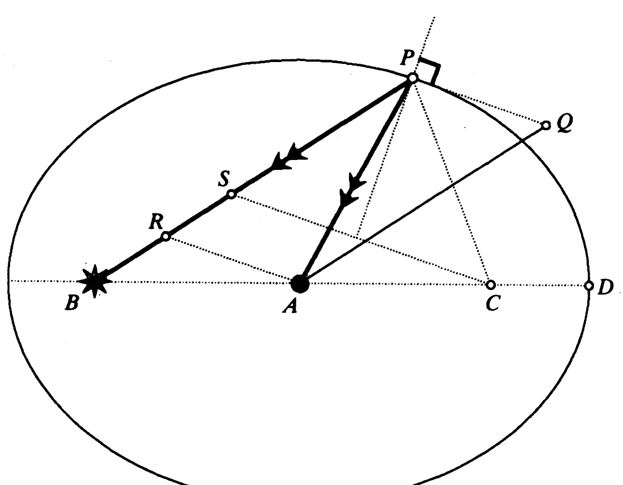

Figure 5.

have been drawn parallel to the tangent PQ. We are now to think of the gravitational field emanating from the focus as arising out of the field that a fictitious body at the center would need to possess in order to hold P in its given elliptical orbit:

$$F_{\text{FOCUS}} \propto \left(\frac{AQ^3}{BP^2 \cdot AP}\right) F_{\text{CENTER}}. \tag{2}$$

Notice two things: $BA = CA$ implies $BR = SR$; the reflection property established in the introduction says that BP and CP make equal angles with the normal at P, and thus $PS = CP$. It follows that

$$AQ = PR = \frac{BP + PS}{2} = \frac{BP + CP}{2} = AD,$$

and since AD is constant, (2) therefore becomes

$$F_{\text{FOCUS}} \propto \left(\frac{1}{BP^2 \cdot AP}\right) F_{\text{CENTER}}.$$

But Newton has already established in Prop. X that $F_{\text{CENTER}} \propto AP$, and he thereby concludes that the gravitational field of the sun is

$$F_{\text{FOCUS}} \propto \frac{1}{BP^2}.$$

3. ENTER THE COMPLEX PLANE. Consider a particle of unit mass located at the point z in the complex plane, and subject to a force $|z|$ directed towards the origin. The differential equation governing its motion will therefore be $\ddot{z} = -z$, the general solution of which is

$$z = pe^{it} + qe^{-it}, \tag{3}$$

where p and q may, without any real loss of generality, be taken as real and satisfying $p > q$. As illustrated on the left of FIG. 6, this is elliptical motion with the attracting point at the center, and with foci at $\pm 2\sqrt{pq}$. These facts are

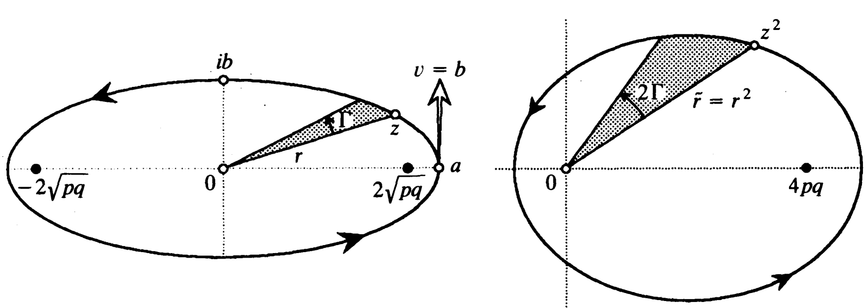

Figure 6.

perhaps more readily verified upon re-expressing (3) as $z = a\cos t + ib\sin t$, where $a = p + q$ and $b = p - q$. Of course each of these numbers has a double significance: a is both the semimajor axis and the point of launch; b is both the semiminor axis and the speed of launch.

The possibility of a connection with Newton's idea of transmuting the linear field into the gravitational one appears from the following surprising geometric fact. If we apply the mapping $z \rightsquigarrow z^2$ to an origin-centered ellipse, then the image is not some strange ugly shape, as one might expect, but rather another *perfect ellipse*; furthermore, this ellipse automatically has one focus at the *origin*. See FIG. 6. Before exploring the implications, let us verify this fact: squaring (3),

$$z \rightsquigarrow z^2 = \left(pe^{it} + qe^{-it}\right)^2 = p^2e^{i2t} + q^2e^{-i2t} + 2pq.$$

The first two terms correspond to an origin-centered ellipse with foci at $\pm 2pq$; the last term therefore translates the left-hand focus to the origin.

Compare this purely geometric fact with Newton's dynamical argument. While leaving the orbit fixed, Newton moves the attracting point from the center to the focus, and deduces that the force law is transmuted from linear to inverse square; while leaving the attracting point fixed at the origin, $z \rightsquigarrow z^2$ transforms an orbit of the linear field into an orbit of the inverse square field. But we are only in a position to make the latter statement because of our prior knowledge of what the orbits in the two fields look like. *Is there instead some* a priori *reason why $z \rightsquigarrow z^2$ should map orbits of the linear field to orbits of the gravitational field?*

That there is indeed such a reason was apparently first discovered by K. Bohlin in 1911 ([**5**]); we shall therefore refer to this result as Bohlin's theorem. For a particularly clear account of the conventional explanation, see [**4**], p. 95. For more on the history of the idea, as well as applications to celestial mechanics, see [**6**], [**7**], and [**8**].

Granted that the origin-centered elliptical orbits for the linear field are easily derived (as above), Bohlin's theorem now allows us to view the geometric fate of these orbits under $z \rightsquigarrow z^2$ as a novel *explanation* of the elliptical motion of planets about the sun as focus. It is strange, though, that the only gravitational orbits we have managed to explain in this way are the ellipses; where are the hyperbolic orbits?

To resolve this, we must generalize Bohlin's result. In Section 6 we will show that gravitational orbits arise not only as the images of those in a linear field that is attractive, but also of those in a linear field that is *repulsive*. For the moment, though, let us simply use our prior knowledge of gravitational orbits (i.e. cheat) to empirically confirm that this result will indeed supply the missing hyperbolic orbits.

The differential equation for the repulsive linear field is $\ddot{z} = z$, the two basic solutions of which are $e^{\pm t}$. The (essentially) general solution can then be obtained by superposing conjugate amounts of these two:

$$z = \lambda e^t + \bar{\lambda} e^{-t}. \tag{4}$$

As illustrated on the left of Fig. 7, these are hyperbolic orbits with center (i.e. intersection of asymptotes) at the origin, and with foci at $\pm 2|\lambda|$. This may seem more familiar upon rewriting (4) as $z = a\cosh t + ib\sinh t$, where $a = \lambda + \bar{\lambda}$, and $ib = \lambda - \bar{\lambda}$. Both these numbers have the same physical significance as in the elliptical case: a = launch point; b = launch speed.

According to the asserted extension of Bohlin's result, we should seek the gravitational hyperbolas amongst the images under $z \rightsquigarrow z^2$ of the orbits in the

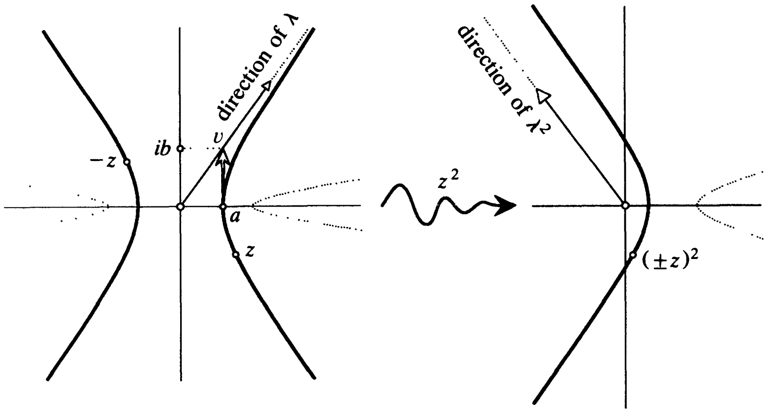

Figure 7.

repulsive linear field. What are these images?

$$z \rightsquigarrow z^2 = \left(\lambda e^t + \bar{\lambda} e^{-t}\right)^2 = \lambda^2 e^{2t} + \bar{\lambda}^2 e^{-2t} + 2\lambda\bar{\lambda}. \tag{5}$$

The first two terms correspond to an origin-centered hyperbola with foci at $\pm 2|\lambda^2|$; the last term therefore translates the left-hand focus to the origin, apparently as hoped for.

We say 'apparently' because FIG. 7 presents a problem. While the solid hyperbola does indeed map to a gravitational orbit about the origin as attracting center, the dotted one does *not*. A possible interpretation of the dotted image would be that it represents a gravitational orbit about the *other* focus. This will not do, however, because we wish to think of the physical cause of the orbits as residing permanently at the origin, the other focus being devoid of physical influence. From this point of view it is clear that the dotted orbit must correspond to a *repulsive* field emanating from the origin. In the next section we will verify that this repulsive field is in fact inverse square, just like the attractive field that produces the solid orbit.

Next we shall use the conserved total energy E of the orbiting particle to characterize the distinction between those hyperbolas in the repulsive linear field that map to attractive orbits, and those that map to repulsive orbits.

If the particle's speed is v, then since we shall always use a particle of unit mass, the kinetic energy contribution has the definite value $\frac{1}{2}v^2$, while the potential energy contribution is only defined up to a constant. The constant is fixed by arbitrarily assigning zero potential energy to some point in the plane. We shall make a very natural choice and assign zero energy to a point where the field vanishes: for a power law that diminishes with distance, at infinity; for a power law that increases with distance, at the origin.

In the repulsive linear field the total energy is then $E = (1/2)(v^2 - r^2)$, where $r = |z|$. Because E is constant, we may evaluate it at any point of the hyperbolic orbit. At launch, $v = b$ and $r = a$, so

$$E = \tfrac{1}{2}(b^2 - a^2).$$

But FIG. 7 informs us that the images that are attracted to the origin correspond to values of λ for which λ^2 (the asymptotic direction of the image) points to the left. Since $4\lambda^2 = (a^2 - b^2) + i2ab$, we deduce that the image is attracted or repelled according as E is positive or negative. This characterization is equally applicable to the attractive linear field, for in that case $E = \frac{1}{2}(b^2 + a^2)$ is always positive, while the images are always attracted.

Returning to the repulsive field, we also anticipate that when $E = 0$ $(a^2 = b^2)$, the image is neither attracted nor repelled, and is thus a straight line. Using (5), the reader may readily verify the truth of this conjecture.

Granted Bohlin's result and its extension, we have now explained the two principal kinds of motion in a gravitational field, and it only remains to explain the orbit which cannot decide if it's elliptical or hyperbolic: the androgynous parabola. Since ellipses and hyperbolas arise as images of orbits in linear force fields that are attractive and repulsive, respectively, we anticipate that the parabola will arise in the transitional case of zero force. The reader may now verify that the rectilinear orbits that arise in the absence of force do indeed map to parabolas with foci at the origin.

To end this section we will show that the effects of $z \rightsquigarrow z^2$ on both ellipses and hyperbolas are actually two equivalent facets of a single phenomenon. We shall argue that the result for ellipses implies the result for hyperbolas; the truth of the converse will then be apparent.

FIG. 8 shows an origin-centered hyperbola $\mathscr{H}$ whose image under z^2 is sought. To find this image, draw the family of ellipses that is confocal to $\mathscr{H}$, and let p be an intersection of $\mathscr{H}$ with one of these ellipses $\mathscr{E}$. Now a ray of light emitted from one of the foci towards p will be reflected directly towards the other focus by $\mathscr{E}$, and directly away from it by $\mathscr{H}$. But the reflected beam of light from a rotating mirror itself rotates twice as fast as the mirror, and thus to turn the beam through two right angles (as here) we must turn the mirror through one right angle. Thus $\mathscr{H}$ is perpendicular to $\mathscr{E}$, and hence to the entire family of ellipses. Conversely, it is clear that any orthogonal trajectory through the ellipses must be a confocal hyperbola. Now apply $z \rightsquigarrow z^2$. By the assumed result, the family of ellipses is mapped to another confocal family of ellipses, one focus being at the origin. But because $z \rightsquigarrow z^2$ is *conformal*, the image of $\mathscr{H}$ must be an orthogonal trajectory through the new family of ellipses. Q.E.D.

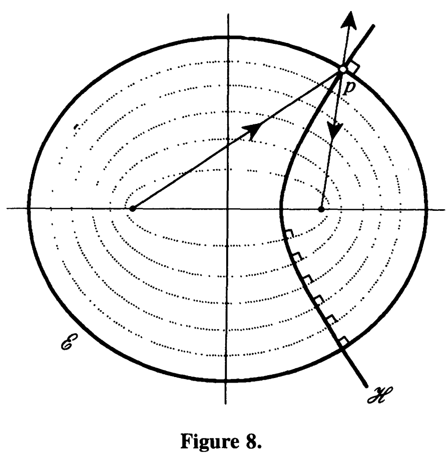

Figure 8.

4. NEWTON'S GEOMETRY OF FORCE. The second fundamental physical constant associated with a central orbit is its angular momentum h. If r is the distance of the particle from the center of force, and ω its angular speed about this center, then $h = r^2\omega$. The geometric interpretation of this quantity is that it's simply twice the "areal speed" $\mathscr{A}=$ (the rate at which the radius vector sweeps out area).

There is a complication associated with Bohlin's theorem that arises from the following basic theorem of Newton [Props. I & II]: $\mathscr{A}$, and therefore h, will remain constant if and only if the force field is central. Suppose that we watch a film of a particle orbiting in the linear field along the left-hand ellipse of FIG. 6, while its image travels round the right-hand ellipse. Although the complete path traced by the image is indeed a genuine gravitational orbit, the way in which the image moves in *time* is physically impossible! For, letting tilde indicate a quantity associated with the image under $z \rightsquigarrow z^2$, we find that

$$\widetilde{\mathscr{A}} = \tfrac{1}{2}\tilde{r}^2\tilde{\omega} = \tfrac{1}{2}(r^2)^2(2\omega) = 2r^2\mathscr{A},$$

thereby making it impossible for both $\mathscr{A}$ and $\widetilde{\mathscr{A}}$ to remain constant. If $\mathscr{A}$ is held constant [physical motion for preimage] while we make our film, then when we play it back, the projector must be continually speeded up and slowed down (in proportion to r^2) to make the image appear to sweep out area at a constant rate.

In light of this fact it is curious that the conventional explanation of Bohlin's result characterizes a force field by means of a *temporal* differential equation. To overcome the above difficulty, it becomes necessary to introduce a fictitious time coordinate into the image orbit (cf. our film projector trick) in such a way that area is swept out at a constant rate with respect to it. It therefore seems natural to seek an alternative explanation that avoids explicit mention of time, and instead directly addresses the geometry of the orbits. Before embarking on the details of this new explanation, let us outline the strategy that will be followed.

In the absence of force, a particle will move in a straight line; *bending* is therefore the manifestation of force, and this can be quantified in terms of the curvature of the orbit. In this section we derive the formula that relates the force to the curvature, and give three interesting examples of its use. In the next section we discover how the curvature of an orbit is transformed under an analytic mapping. Finally, having understood both the transformation of curvature and its relationship to force, in Section 6 we deduce the transmutation of force that is effected by an analytic mapping.

The relationship between force and curvature could be derived rapidly by appealing to standard elementary results in dynamics. However, continuing in the spirit of the introduction, we choose instead to illustrate the *Principia's* method by deriving it from scratch, using little more than similar triangles.

In FIG. 9(A), which shows a particle orbiting in a central force field emanating from C, we have drawn QS and CY parallel to the normal at P, and QR parallel to the radius CP. Our problem is to determine from the shape of the orbit the force F that is acting on the particle when it is at the point P, and ultimately to express the result in terms of the curvature κ. Initially, we shall simply follow Newton.

After a short time Δt the particle will have moved along the orbit from P to Q, and the radius will have swept out an area $(1/2)h\,\Delta t$. The net motion from P to Q can be thought of as compounded of the force-free motion PR along the tangent, together with the force-induced deflection QR in the direction of the force acting at P. This view of events becomes increasingly accurate as we shorten Δt, for as Q

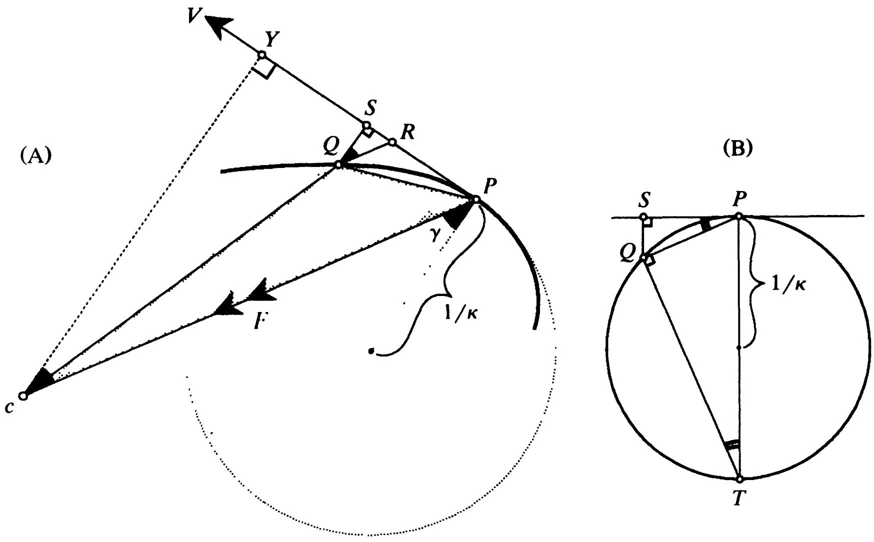

Figure 9.

is taken closer and closer to P, the force (including its direction) acting in the arc PQ becomes more and more nearly constant, and coincides with that at P.

Before pursuing this idea, and in the interests of brevity, let us adopt the following notation and form of words as being equivalent to the statement

$$\text{“}\lim_{Q \to P}\left(\frac{X}{Y}\right) = 1\text{”:}$$

$$\text{“}X \asymp Y\text{”} \Leftrightarrow \text{“}X \text{ and } Y \text{ become equal as } Q \text{ coalesces with } P\text{”.}$$

The basic theorems on limits inform us that " $\asymp$ " inherits many of the properties of " = ". This simple notational device merely paraphrases the idea which Newton expresses verbally in the Scholia on pages 35 and 37 of the *Principia*. It allows us to draw on the intuitive power of his infinitesimal geometry while continuing to pay lip-service to the tyrannical ϵ-δ legacy of Cauchy and Weierstrass.

Returning to the investigation, and recalling that under the action of a constant force, *distance* = (*half*) · (*acceleration*) · (*time squared*), we deduce that the force at P is given by

$$F \asymp \left[\frac{2 \cdot QR}{(\Delta t)^2}\right]. \tag{6}$$

We may now render this formula purely geometrical by using the area law to eliminate time:

$$\begin{aligned} \tfrac{1}{2}h\,\Delta t &= (\text{area of sector } CPQ) \\ &\asymp (\text{area of shaded triangle } CPQ) \\ &\asymp \tfrac{1}{2}(PQ \cdot CY), \end{aligned}$$

where the last 'equality' follows from observing the evolution as $Q \rightsquigarrow P$ of the altitude from C to the base PQ. Substituting for Δt in (6), we obtain one of the formulae [Cor. II., Prop. VI] that lie at the heart of Newton's geometric investigation of force in the *Principia*:

$$F \asymp h^2\left[\frac{2 \cdot QR}{PQ^2 \cdot CY^2}\right].$$

In order to re-express Newton's formula in terms of curvature, consider FIG. 9(B). From the similar triangles PSQ and TQP, we find that

$$\frac{PQ}{PT} = \frac{QS}{PQ} \quad \Rightarrow \quad \kappa = \left(\frac{2 \cdot QS}{PQ^2}\right).$$

This formula may also be applied to the non-circular orbit in FIG. 9(A), for as Q coalesces with P, it will yield the curvature of the orbit at P. Next, observe that we may re-express QR as $(QS \cdot \sec\gamma)$, and CY as $(CP \cdot \cos\gamma)$, where γ is the angle between the normal and the radius at P. Thus

$$F \asymp h^2\left(\frac{2 \cdot QS}{PQ^2}\right)\left(\frac{\sec^3\gamma}{CP^2}\right),$$

and writing r for CP, we arrive at our destination [cf. Prop. VII]:

$$\boxed{F = h^2\left(\frac{\kappa \sec^3\gamma}{r^2}\right).} \tag{7}$$

In this formula, γ is to be taken as acute so that $\sec\gamma$ is always positive; in the next section we shall introduce a sign for κ so as to distinguish between attractive and repulsive forces.

Before describing some applications of this formula, let us note another result that we shall need later. If v is the speed of the particle at P, then [Cor. I, Prop. I]

$$v \asymp \left(\frac{PQ}{\Delta t}\right) \asymp \left(\frac{h}{CY}\right),$$

and therefore

$$v = h\left(\frac{\sec\gamma}{r}\right). \tag{8}$$

As our first application of (7), let us ask and answer the question, what force field can yield a circular orbit that passes through the center of attraction? See FIG. 10(A). Since $\sec\gamma = (2\rho/r)$ and $\kappa = (1/\rho)$, we immediately find [Cor. I,

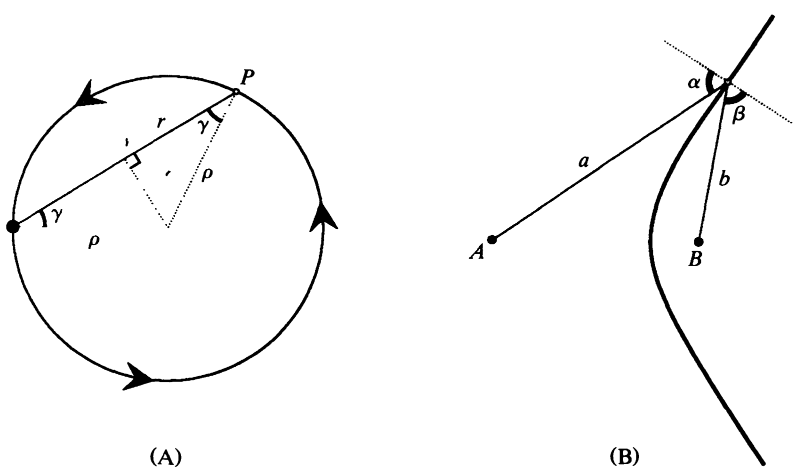

Figure 10.

$$F = \left(\frac{8h^2\rho^2}{r^5}\right).$$

The significance of the $(1/r^5)$ field, and this particular orbit within it, will be discussed in Section 6.

For our second example, turn to Fig. 10(B). Granted that such a hyperbola is a gravitational orbit about the focus B as attracting center, you may now verify (using $\alpha = \beta$) our previous claim that if the orbit is instead due to a repulsive field emanating from the other focus A, then this field must also be inverse square.

As the last of our three examples, let us verify Newton's "transmutation formula" (1). Reconsider FIG. 4. Let fall the perpendicular AR from A onto the tangent PRQ, and let α and β be the angles PAR and QAR respectively. Then

$$AP \cdot \cos\alpha = AR = AQ \cdot \cos\beta \quad \Rightarrow \quad \frac{\sec\beta}{\sec\alpha} = \frac{AQ}{AP}.$$

Using (7) to write down the ratio of the new force to the old one, the result (1) is now readily verified.

5. ANALYTIC TRANSFORMATION OF CURVATURE. Suppose that an analytic mapping $f(z)$ acts on an orbit of curvature κ, yielding an image orbit of curvature $\tilde{\kappa}$. What is $\tilde{\kappa}$ in terms of κ and $f(z)$? This is the question that we must answer if we are to understand the relationship between the force fields that hold the preimage and image in their respective orbits. While we could certainly answer this question by calculation, we choose instead to present a novel solution that is entirely geometric.

The means by which the curvature of an orbit will be determined is illustrated in FIG. 11(A), where a particle is seen orbiting in a counterclockwise circle. As shown, ξ and ζ are two successive chords of equal length in the direction of motion, and ϵ is the angle of turning from ξ to ζ. Because the angle ϕ that each of these chords subtends at the center must equal the turning angle ϵ,

$$\frac{1}{\kappa} \cdot \epsilon = (\text{the arc } PQ) \asymp |\xi|.$$

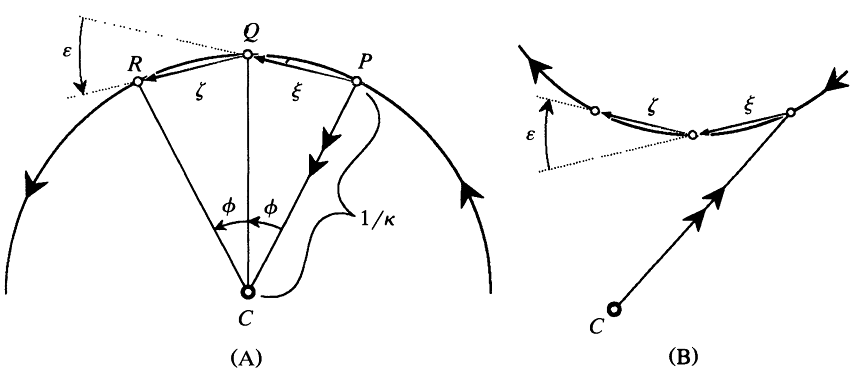

Figure 11.

The curvature can therefore be expressed as

$$\kappa \asymp \frac{\epsilon}{|\xi|}. \tag{9}$$

It is clear that this construction can immediately be applied to a general non-circular orbit: as the equal lengths of ξ and ζ diminish, and as Q and R coalesce with P, we obtain the curvature of the orbit at P.

Before continuing, let us introduce some interrelated conventions in order to systematically distinguish between attractive and repulsive forces. Although time is essentially to be suppressed in what follows, it is still helpful to picture the particle as moving along the orbit, and our first convention is to insist that the sense of this motion be *counterclockwise* about the center of force.

With this convention in place, (9) now attributes a definite sign to κ. For example, if the center of force is at C in both FIG. 11(A) and FIG. 11(B), then κ is positive in the first, and negative in the second; we also notice that the responsible forces are attractive and repulsive, respectively. Thus, by using this signed curvature in (7), we may identify a positive F with attraction, and a negative F with repulsion.

We now return to the original problem, which is both depicted and solved in FIG. 12. The left-hand figure shows the preimage orbit of curvature κ, and the top figure shows its image under $f(z)$. Just as ξ connects P and Q, so $\tilde{\xi}$ connects the image points $\tilde{P}$ and $\tilde{Q}$, and we shall say that $\tilde{\xi}$ is the "image" of ξ. Likewise, just

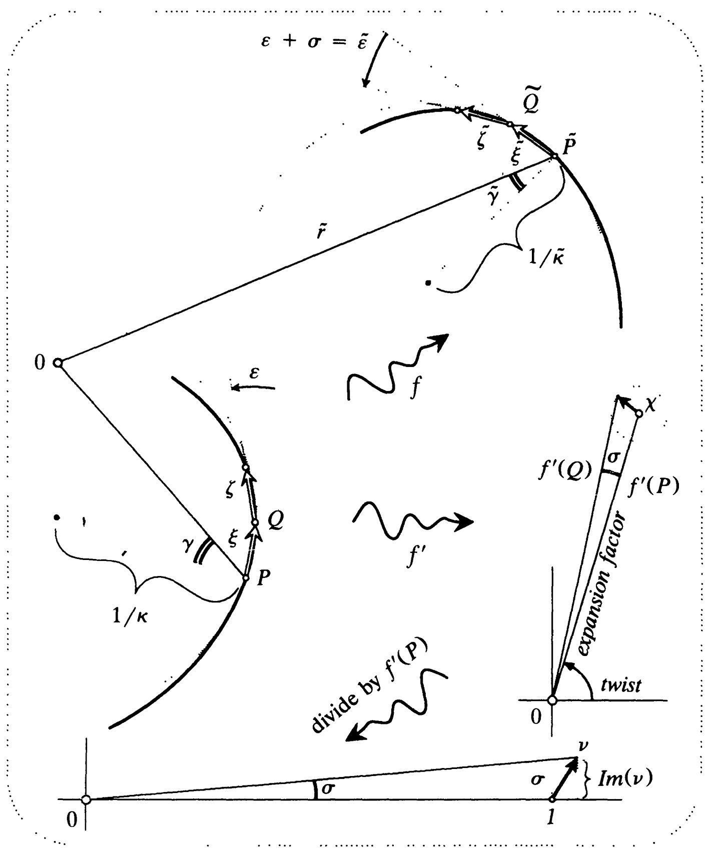

Figure 12.

as ξ and ζ yield the curvature of the preimage, so their images yield $\tilde{\kappa}$:

$$\tilde{\kappa} \asymp \frac{\tilde{\epsilon}}{|\tilde{\xi}|}, \tag{10}$$

and our problem therefore reduces to finding $\tilde{\epsilon}$ and $|\tilde{\xi}|$.

The distinguishing characteristic of an analytic function is that its local effect is simply to expand and twist. Thus if we imagine a microscopic circle centered at P, its image under such a mapping will not be a tiny ellipse (as it would be in general), rather it will be another tiny *circle* centered at $\tilde{P}$. As illustrated in the right-hand figure, the information of the expansion factor and the twist that carry the former circle into the latter is stored in the single complex number $f'(P) = (\textit{expansion factor})e^{i(\textit{twist})}$. Part of the problem is therefore easily solved:

$$|\tilde{\xi}| \asymp (\textit{expansion factor}) \cdot |\xi| = |f'(P)| \cdot |\xi|. \tag{11}$$

The more interesting and difficult part of the problem is to find $\tilde{\epsilon}$. If ξ and ζ both underwent precisely the same twist, then the turning angle $\tilde{\epsilon}$ for the images would equal the original angle ϵ. However, the twist at Q will differ very slightly, say by σ, from that at P. Thus

$$\tilde{\epsilon} = \epsilon + (\textit{extra twist}) = \epsilon + \sigma. \tag{12}$$

This angle σ (which we must find) is illustrated in the right-hand figure.

To find σ, we appeal to the astonishing fact that if a mapping sends tiny circles to tiny circles, then *so does its derivative*: in more conventional language, an analytic mapping is infinitely differentiable. Thus $f'(z)$ maps a tiny circle centered at P to a tiny circle [dotted] centered at $f'(P)$, and the expansion and twist that carries the former to the latter is encoded as $f''(P)$: if (as shown) χ is the image of ξ under $f'(z)$, then

$$\chi \asymp f''(P) \cdot \xi.$$

Knowing χ, we are now very close to finding the extra twist σ, for we observe that it is the angle at the origin in the triangle of the right-hand figure, the sides of which are the known quantities $f'(P)$ and χ. It is easier to obtain an expression for σ if we first rotate this triangle to the real axis. This rotation is achieved quite naturally (see the bottom figure) by dividing by $f'(P)$; the sides of the triangle now become 1 and $\nu = [\chi/f'(P)]$. Because σ equals the almost vertical arc through 1, we see from the figure that

$$\sigma = \text{arc} \asymp \text{Im}(\nu) = \text{Im}\left[\frac{\chi}{f'(P)}\right] \asymp \text{Im}\left[\frac{f''(P) \cdot \xi}{f'(P)}\right].$$

Thus, from (10), (11), and (12), and taking evaluation at P as understood, we obtain

$$\tilde{\kappa} \asymp \frac{\left(\text{Im}\left[\dfrac{f'' \cdot \xi}{f'}\right] + \epsilon\right)}{|f'| \cdot |\xi|}.$$

Finally, using (9), and writing $\hat{\xi}$ for the unit tangent at P, we arrive at our transformation formula:

$$\boxed{\tilde{\kappa} = \frac{1}{|f'|}\left(\text{Im}\left[\frac{f'' \cdot \hat{\xi}}{f'}\right] + \kappa\right).} \tag{13}$$

While I doubt that this interesting formula can be new, I confess that I have been unable to find it elsewhere.

Let us immediately do an example. With $f(z) = e^z$, and writing $z = x + iy$, we find that

$$\tilde{\kappa} = e^{-x}(\sin \phi + \kappa),$$

where ϕ is the angle that the tangent makes with the horizontal. FIG. 13 shows three line-segments on the left, and their images under $z \rightsquigarrow e^z$ on the right. The reader is encouraged to verify the accord between this figure and our formula.

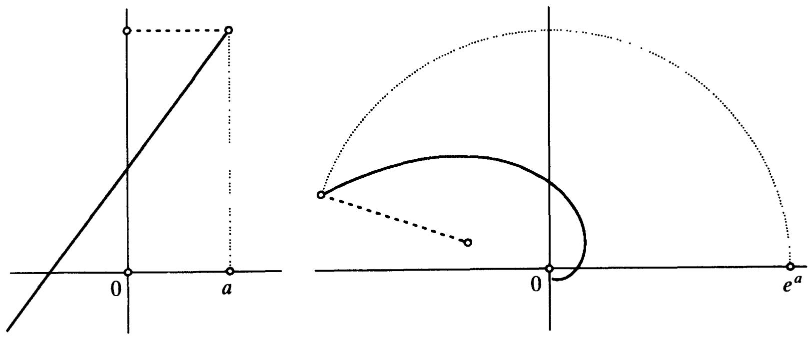

Figure 13.

In order to understand Bohlin's theorem and its generalizations, we need to know the transformation law for the power mappings $f(z) = z^m$. Substituting this into (13), we obtain

$$\tilde{\kappa} = \frac{1}{|m|r^{m-1}}\left[(m-1)\mathrm{Im}\left(\frac{\hat{\xi}}{z}\right) + \kappa\right] = \frac{1}{|m|r^{m-1}}\left[(m-1)\frac{\sin\psi}{r} + \kappa\right],$$

where ψ is the angle from z to the directed unit tangent $\hat{\xi}$.

With the origin as center of force, our convention of counterclockwise motion is now equivalent to $0 < \psi < \pi$, and so $\sin \psi$ is positive. Using this fact, we may rewrite the above result in its final form:

$$\tilde{\kappa} = \frac{1}{mr^{m-1}}\left[(m-1)\frac{\cos\gamma}{r} + \kappa\right], \tag{14}$$

where γ is the acute angle, familiar from the force formula (7), between the normal and the radius.

Notice that in the denominator of this formula we have written m instead of $|m|$, which would be the strict deduction from (13). Thus the formula agrees with the strict deduction when m is positive, and is opposite to it when m is negative. The appropriateness of this choice stems from the convention of counterclockwise motion. With $m > 0$, conventional motion for the preimage results in conventional motion for the image, as depicted in FIG. 12. However, with $m < 0$, the sense of the motion is *reversed* by the mapping, and the value of $\tilde{\kappa}$ furnished by (13) is therefore opposite to that required by our convention.

As a simple example of this, consider the case of inversion, $z \rightsquigarrow (1/z)$. A vertical line is mapped to a circle through the origin having *positive* curvature according to our convention. To check (14) we simply put $m = -1$ and $\kappa = 0$:

$$\tilde{\kappa} = +2r\cos\gamma.$$

Thus the formula yields the correct sign, and a picture quickly reveals that the actual value is also correct.

6. THE TRANSMUTATION OF FORCE. At last we are in a position to find out the effect of an analytic mapping on force itself. Applying $z \rightsquigarrow f(z)$,

$$F = h^2\left(\frac{\kappa\sec^3\gamma}{r^2}\right) \rightsquigarrow \tilde{F} = \tilde{h}^2\left(\frac{\tilde{\kappa}\sec^3\tilde{\gamma}}{\tilde{r}^2}\right),$$

where $\tilde{\kappa}$ is known from (13), and where $\tilde{r} = |f(P)|$ is the distance to the image, as illustrated in Fig. 12. Note that a general conformal mapping will *not* preserve the angle γ. However, if we restrict ourselves to the *power mappings* $f(z) = z^m$, then the ray through P is mapped to the ray through $\tilde{P}$, and therefore $\tilde{\gamma} = \gamma$.

As our first example of the above idea, let us consider $f(z) = z^2$, for this should lead us to Bohlin's result. Putting $m = 2$ in (14), and then substituting for the original force and speed from (7) and (8), we deduce that the force responsible for the image orbit is

$$\tilde{F} = \tilde{h}^2 \cdot \frac{\frac{1}{2}\left[\dfrac{\cos\gamma}{r^2} + \dfrac{\kappa}{r}\right]\sec^3\gamma}{\tilde{r}^2} = \left(\frac{\tilde{h}}{h}\right)^2 \frac{\left[\frac{1}{2}v^2 + \frac{1}{2}rF\right]}{\tilde{r}^2}.$$

Even if F is a simple power law, generally this $\tilde{F}$ will not be. However, if and *only* if the original force field is the attractive or repulsive *linear* one ($F = \pm r$), the numerator in the above expression magically becomes the constant total energy E of the particle in the original field:

$$\tilde{F} = \left(\frac{\tilde{h}}{h}\right)^2 \frac{E}{\tilde{r}^2}\ !$$

The image, therefore, moves in a field that is *inverse-square*. Furthermore, if the original orbit has positive energy then its image is attracted to 0, while if it has negative energy then its image is repelled by 0. We have therefore successfully explained all of our empirical findings in Section 3.

Before describing the general result (due to Arnol'd, [**3**]) on "dual" pairs of power laws, we shall treat one further special case that arises from the following interesting question. Might there exist a force field that is *self*-dual in the sense that the images of orbits within it (under a complex power mapping) would simply be new orbits in the *same* field?

If this were possible then clearly the responsible mapping would have to be *self-inverse*. Discarding the identity mapping as trivial, the sole possiblity is therefore inversion: $z \rightsquigarrow (1/z)$. Putting $m = -1$ into (14), and noting that $\tilde{r} = (1/r)$, we find that

$$\tilde{F} = +4H^2\frac{\left[\frac{1}{2}v^2 - \frac{1}{4}rF\right]}{\tilde{r}^5},$$

where H stands for the ratio of the angular momenta, $(\tilde{h}/h)$. In order for the numerator to become the total energy, the potential energy must be $-\frac{1}{4}rF$, and therefore the original force field must be the attractive or repulsive inverse fifth power: $F = \pm(1/r^5)$. As anticipated, the force acting on the image is then *also* inverse-fifth, and furthermore it is attractive or repulsive according as the original energy is positive or negative.

We previously found an orbit of this field in FIG. 10(A); its very special character can be seen from the fact that under inversion it maps to a force-free straight line. For pictures of more general orbits, as well as their mathematical classification, the interested reader may consult [**10**].

We turn now to the general case. Suppose that the mapping $z \rightsquigarrow z^m$ acts on orbits in a force field F. Recalling that $\tilde{r} = r^m$, we find that

$$\tilde{F} = \frac{2(m-1)}{m} H^2 \left[\tfrac{1}{2} v^2 + \frac{rF}{(2m-2)} \right] \tilde{r}^{[(2/m)-3]}. \tag{15}$$

In order to obtain the total energy in this expression, the potential energy must be $[rF/(2m-2)]$. Only for a power law is this energy proportional to rF, and, in greater detail, for $F = \pm r^A$ it equals $[rF/(A+1)]$. We deduce that the mapping that effects the transmutation must be related to the original force field by $(2m-2) = (A+1)$. If $\tilde{A}$ stands for the exponent in the image power law, so that $\tilde{A} = [(2/m) - 3]$, we may then summarize our findings as follows.

Associated with each power law (exponent A) there is precisely one power law (exponent $\tilde{A}$) that is "dual" in the sense that orbits of the former are mapped to orbits of the latter by $z \rightsquigarrow z^m$, and the relationships between the forces and the mapping are:

$$(A+3)(\tilde{A}+3) = 4 \qquad \text{and} \qquad m = \frac{(A+3)}{2}. \tag{16}$$

Observing the coefficient $[(m-1)/m]$ in (15), we further conclude that (in general) positive energy orbits in either the attractive or repulsive field $F = \pm r^A$ map to attractive orbits in the dual field, while negative energy orbits map to repulsive ones. However, if $-3 < A < -1$ (e.g. gravity) then these roles are reversed. In all cases, zero energy orbits map to force-free rectilinear orbits.

Letting $[A, \tilde{A}]$ stand for a pair of dual force laws, we see from (16) that amongst *integer* exponents (and excluding $z \rightsquigarrow z$ as trivial) there are only three such pairs:

$$[1, -2] \qquad [-4, -7] \qquad [-5, -5]. \tag{17}$$

Of these only one is new to us, namely, $[-4, -7]$; the mapping in this case is $z \rightsquigarrow (1/\sqrt{z})$.

7. CONCLUDING REMARKS. **(I)** In his book, Arnol'd alludes to, but does not state, a connection between (16) and a result of Newton's [Prop. XLV]. While I cannot be sure of what *Arnol'd* had in mind, let me at least point out *a* connection.

Now it so happened that the solar system formed in such a way that the planetary and lunar orbits are almost circular. For this reason Newton undertook the investigation of almost circular orbits for a general power law $F = r^A$. He found that the angle Φ between successive turning points of r (aphelion to perihilion, or vice versa) is given by the formula

$$\Phi = \frac{\pi}{\sqrt{A+3}}.$$

With $A = 1$ or -2, this value of Φ is exact (as you may verify in FIG. 6) even for highly non-circular orbits.

Next we observe that turning points may be identified by the property $\gamma = 0$. If $z \rightsquigarrow z^m$ then $\tilde{\gamma} = \gamma$, and so a turning point is mapped to a turning point, and no new turning points are created in the process. We deduce that the angular separation of turning points on the image curve is $\widetilde{\Phi} = m\Phi$. But if the existence of duality is assumed [of course this is only possible in hindsight] then this image curve will be an orbit for some new power law, say with exponent $\tilde{A}$, and therefore

$$\widetilde{\Phi} = \frac{\pi}{\sqrt{\tilde{A} + 3}}.$$

Equating these two expressions for $\widetilde{\Phi}$, we recover half of the information contained in (16) but expressed in a different form:

$$m^2 = \left(\frac{A + 3}{\tilde{A} + 3}\right).$$

(II) With the exception of the pair $[1, -2]$, the integer power laws ($A = -4, -5, -7$) that are singled out in (17) are rather mysterious. Their physical significance is unknown to me, and that in itself is strange, for the music of mathematics is seldom played without an accompanying echo being heard in Nature. At the purely mathematical level, however, there is clearly more here than first meets the eye. For example, consider the following question (see [**9**]): for which integer power laws that diminish with distance is the polar equation of the orbit expressible in terms of elliptic functions? It turns out that there are precisely three. They are: -4, -5, and -7!

ACKNOWLEDGMENTS. First I should like to thank Dr. Stanley Nel for his enthusiastic reception of the main idea, and for his helpful comments on the first draft. I also thank Dr. Subrahmanyan Chandrasekhar for helpful observations on the Newtonian portion of the paper. Lastly, I thank Dr. James Finch for his patient and very able assistance in overcoming various problems with $\mathrm{T_EX}$, and in particular for his figuring out how to import my diagrams, which I had created using "CorelDRAW".

REFERENCES

1. *Newton's Principia*, U. of California Press, 1934.
2. R. Westfall, *Never at Rest*, Cambridge U. Press, 1980.
3. V. I. Arnol'd and V. A. Vasil'ev, Newton's Principia read 300 years later, *Notices Amer. Math. Soc.*, 36 (1989) 1148–1154.
4. V. I. Arnol'd, *Huygens & Barrow, Newton & Hooke*, Birkhäuser Verlag, 1990.
5. K. Bohlin, Note sur le problème des deux corps et sur une intégration nouvelle dans le problème des trois corps, *Bull. Astr.*, 28 (1911) 113–119.
6. K. R. Meyer, The Geometry of Harmonic Oscillators, this *Monthly*, 97 (1990) 447–465.
7. D. G. Saari, A visit to the Newtonian *N*-body problem via elementary complex variables, this *Monthly* 97 (1990) 105–119.
8. V. Szebehely, *Theory of Orbits*, Acad. Press, New York, 1967. Chapter 3.
9. H. Goldstein, *Classical Mechanics*, second Ed., Addison-Wesley, 1980. Chapter 3.
10. W. D. MacMillan, *Statics and the Dynamics of a Particle*, McGraw-Hill, 1927. Chapter XII, Section V. Also by the same author, see *Amer. J. Math.*, Vol. XXX, 282–306.

Mathematics Department
University of San Francisco
San Francisco, CA 94117-1080

Note added in proof: We attributed the general duality law (16) to Arnol'd, but it was in fact discovered by Edward Kasner in or before 1909. It appears in his *Differential-Geometric Aspects of Dynamics*, published by the A.M.S. in 1913.