Curves from Motion, Motion from Curves

Rida T. Farouki

Abstract. Geometry and kinematics have been intimately connected in their historical evolution and, although it is currently less fashionable, the further development of such connections is crucial to many computer–aided design and manufacturing applications. In this survey, we explore a variety of classical and modern problems that illustrate how simple rules of motion produce interesting curves and, conversely, the computational problems of generating motions with prescribed paths and speeds. These encompass the geometry of trajectories under centripetal forces; the transformation of rotary motion into motion along general curves by mechanisms; real–time curve interpolators for digital motion control; and the description of spatial motions that involve variations of both position and orientation. Such case studies illustrate some of the intellectual appeal, and practical importance, of a sustained dialog between the study of curves and of motions.

§1. Preamble

Our intent in this paper is to survey the intricate web of historical connections between geometry and kinematics, a theme that has played a key role in the development of mechanics and analysis. In contemplating this theme, we are obviously confronted by a profusion of interesting and fruitful topics — and we are thus obliged to adopt a rather anecdotal approach.

Apart from its intrinsic interest, we choose this subject with the hope of promoting greater synergy between modern–day problems of geometric design and motion control. Modern CAD systems are mainly concerned with creating "static" geometrical descriptions of artifacts, but the processes by which these artifacts are actually fabricated often involve complicated motions of a tool — e.g., a cutter in a milling machine, or a wire electrode in electrical discharge machining — relative to a workpiece. Compared to the sophistication of CAD models, current methods for motion planning in manufacturing processes are often rather crude and naive. Thus, there is much scope for securing greater precision and reliability, through the use of advanced mathematical methods, in the relatively undeveloped field of manufacturing geometry.

The symbiosis between geometry and kinematics has deep historical roots. Newton, in his *Quadrature of Curves* (1676), aptly characterizes it as follows:

Curve and Surface Design: Saint-Malo 1999

Pierre-Jean Laurent, Paul Sablonnière, and Larry L. Schumaker (eds.), pp. 63–90.

ISBN 0-8265-1356-5.

> *Lines (curves) are described, and thereby generated, not by the apposition of parts but by the continued motion of points ... These geneses really take place in the nature of things, and are daily seen in the motion of bodies.*

However, this has not always been a happy union. Insofar as it embodies both *spatial* and *temporal* information, kinematics subsumes geometry. To upgrade a *curve* into a *motion* requires the ability to *rectify* (or measure arc lengths of) curves. As a basic philosophical tenet, Descartes held this to be impossible — see §3 below — and he sought to banish all curves whose definitions explicitly or implicitly assume rectifications from the "rigorous" domain of geometry to the nascent (and less–exact) science of mechanics. Although, in modern times, the philosophical/existential problem of rectification is no longer troublesome, we must still address the *computational* difficulties it entails (see §3).

The antithesis of Descartes' attempt to divorce geometry from kinematics would ultimately find its logical expression, in the context of the special theory of relativity, with Minkowski's introduction [42] of the concept of "space–time" as the most natural setting for the study of physical phenomena:

> *Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.*

The recent introduction of the Minkowski metric of space–time into problems of geometric design reveals a remarkable confluence of ideas concerning medial axis transforms, Pythagorean hodographs, envelopes, and offset curves [43].

In this survey we shall attempt, through a series of anecdotal sketches, to promote greater interest in the relationship between geometry and kinematics, and its application to CAD/CAM problems. We commence in §2 and §3 with a brief review of the manner in which curves may be defined, and the problems that measurement of arc length incurs. Perhaps the simplest motion is that of a particle experiencing a force toward a fixed center, of magnitude proportional to a power of the radial distance r. As is well–known, Newton showed that an r^{-2} force of gravity explains the conic form of planetary and cometary orbits. We shall see in §4, however, that this is just one aspect of a more profound theory of *motion under centripetal forces* in Newton's *Principia*.

Mechanisms such as gears, cams, and linkages are used to transform forces and motions in machinery. In §5 we discuss the four–bar linkage, a mechanism that directly transforms rotary motion into motion on a general curved path. CNC machines offer a more flexible approach to motion generation, based on sophisticated servo–systems that drive linear or rotary axes in a coordinated manner. In §6 we discuss the problem of *real-time interpolators*, which must accurately and efficiently interpret the path and speed information to generate "reference point" data required by the digital control algorithm.

The preceding examples are concerned with motion in Euclidean spaces. A motion that involves not just positional but also *orientational* coordinates (such as in 5–axis machining) may be regarded as the motion of a point in a higher–dimensional, non–Euclidean (soma) space. Some subtle problems that arise with such motions are discussed in §7. Finally, §8 offers some concluding thoughts on our theme of connections between geometry and kinematics.

§2. Curves and Motions

Analytic geometry has its origins in the computational investigation of curves specified by suitable coordinate equations. There are basically two ways to define a plane curve in terms of Cartesian coordinates (x, y). We may select a *predicate function*, that indicates whether or not each point in the plane lies on the curve — this is typically a bivariate polynomial f in the coordinates, and the curve is the locus of points on which the polynomial vanishes:

$$f(x, y) = 0. \tag{1}$$

On the other hand, we may choose a pair of *generating functions*

$$x(t),\ y(t) \tag{2}$$

that produce an ordered sequence of curve points when evaluated at successive values of a continuous "auxiliary variable" or parameter t.

Whereas the *implicit* description (1) is essentially "static," the *parametric* form (2) offers a more "dynamic" characterization of curves — it embodies the suggestion of motion along a curve, incurred by steady increase of the curve parameter t. It is a mistake, however, to invest too much hope in the capacity of parametric curves to adequately describe motions. Motion specification is concerned as much with the instants in time at which a body assumes given positions along a path, and corresponding *velocities* and *accelerations*, as with the path *geometry*. A motion is really a geometrical locus in Minkowski space, with one temporal and one or more spatial dimensions.

Of course, we can always *interpret* the parameter t as *time*, and equations (2) then completely specify *a* motion. However, if we wish to use only "simple" (polynomial or rational) functions, such motions are mathematical curiosities: except in trivial cases, they are neither solutions to appropriate equations of motion, nor do they represent motions of practical interest that we may wish to impose on a given locus. To emphasize that the curve parameter generally lacks any geometrical or temporal significance, we henceforth use the Greek character ξ to denote it, and we explicitly reserve t for *time*.

Connections between the study of curves and motions is a recurrent theme in the history of science and technology. At the inception of analytic geometry, motions offered an intuitive means to construct and analyze loci of increasing sophistication: see, for example, the remarkably diverse historical applications of the roulettes generated by the rolling motions of lines and circles (cycloids, circle involutes, epicycloids and hypocycloids, epitrochoids and hypotrochoids) discussed in [21]. Conversely, the conics of the ancient Greeks make a rather surprising appearance in the solution to the premier problem of dynamics: the determination of planetary orbits. In the modern computer era, the problem of producing a desired motion along a given path is central to real–time control of manufacturing, inspection, robotic, and other devices.

§3. Towards an Impossible Ideal

Since speed on a curved path is the rate of change of distance with time, the problem of rectification — i.e., the measurement of arc length — is evidently critical to the description of motion. This problem, however, has been fraught with computational difficulty since Descartes founded analytic geometry in an appendix *La géométrie* [15] to the *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences* (1637). He asserts that:

> *Geometry should not include lines (curves) that are like strings, in that they are sometimes straight and sometimes curved, since the ratios between straight and curved lines are not known, and I believe cannot be discovered by human minds, and therefore no conclusion based upon such ratios can be accepted as rigorous and exact.*

Nevertheless, Descartes' dictum began to crumble almost immediately after its enunciation, amid a flurry of counter–examples.

For example, Galileo [26] realized that, when a body is dropped into a hole drilled through the center of a static Earth, it executes linear simple harmonic motion across the full Earth diameter under the influence of gravity:

> ... *if the terrestrial globe were perforated through the center, a cannon ball descending through the hole would have acquired at the center such an impetus from its speed that it would pass beyond the center and be driven upward through as much space as it had fallen, its velocity beyond the center always diminishing with losses equal to the increments acquired in the descent* ...

On a rotating Earth, however, the body will have an initial tangential velocity, and the nature of its motion in the hypothetical case of "permeable" matter (which exerts gravitational forces but does not impede motion) is not obvious. Galileo's pupil, Evangelista Torricelli (1608–1647), conjectured that the path would be a logarithmic spiral about the Earth's center, described by

$$r = a\,\mathrm{e}^{k\theta} \tag{3}$$

in polar coordinates (also known as an "equi–angular" spiral, since the tangent makes a fixed angle, $\cot^{-1} k$, with the radius vector). Isaac Newton re–iterated this conjecture [2,60] in a letter dated November 28, 1679 to Robert Hooke, who criticized it during a Royal Society meeting the following December 11. As we shall see in §4 below, Torricelli and Newton were quite wrong: the path is actually — as intuitively argued by Hooke — an ellipse.

During his investigations, however, Torricelli discovered a rectification of the spiral (3) in 1645 through the Archimedean "method of exhaustion" — he showed that, for $-\infty < \theta \le 0$, the arc length equals the length of the tangent at $\theta = 0$ extended to the y–axis [9] — namely, $\sqrt{1+k^{-2}}\,a$ (see Figure 1). This is a truly remarkable result, since the curve must execute an infinite number of gyrations about the origin before terminating there!

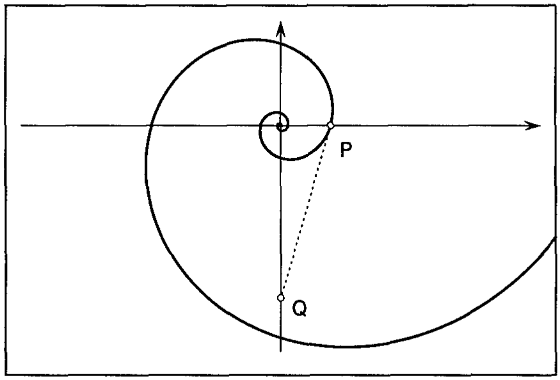

Fig. 1. The arc length of (3) for $\theta \leq 0$ equals the length PQ on the tangent line.

The logarithmic spiral was also known to Thomas Harriot (1560–1621) as the projection of a rhumb line on the Earth's surface — i.e., the path traced by a ship that sails in a fixed compass direction — onto the equatorial plane [55]. Jakob Bernoulli was so fascinated by the self-similarity of this curve under coordinated rotations and dilatations about the origin, that he arranged to have it engraved on his tombstone with the caption *Eadem mutata resurgo* [3] — "Though changed I shall arise the same."

Subsequently, another curve was rectified by Gilles Personne de Roberval (1602–1675) and Christopher Wren (1632–1723) — namely, the cycloid

$$x(\theta) = a(\theta - \sin\theta), \quad y(\theta) = a(1 - \cos\theta) \tag{4}$$

traced by a fixed point on a circle of radius a that rolls without slipping on a straight line (see Figure 2). They showed that a single "arch" ($0 \leq \theta < 2\pi$) of this curve has length $8a$. Although it has now fallen into obscurity, the cycloid was a virtual "proving ground" for novel mathematical ideas and methods in the mid–17th century: it caught the attention of all the leading scientists, and prompted international competitions and acrimonious controversies. See [21] for a discussion of its tautochrone and brachistochrone properties.

To Descartes, however, the rectification of curves such as the spiral (3) and the cycloid (4) was suspect — they are not true "geometrical" (i.e., *algebraic*) but rather "mechanical" (i.e., *transcendental*) curves. By introducing angular variables, their definitions essentially *presuppose* a rectification (of the circle). Nonetheless, it was not long before even an algebraic curve, under the scrutiny of William Neil (1637–1670), Hendrick van Heuraet (1633–1660), and Pierre de Fermat (1601–1665), succumbed to rectification — the cuspidal cubic

$$x(\xi) = \xi^2, \quad y(\xi) = k\,\xi^3 \tag{5}$$

known as the "semicubical parabola." Its arc length s, measured from $\xi = 0$, is an *algebraic* function of the parameter:

$$s(\xi) = \frac{(9k^2\xi^2 + 4)^{3/2} - 8}{27k^2}.$$

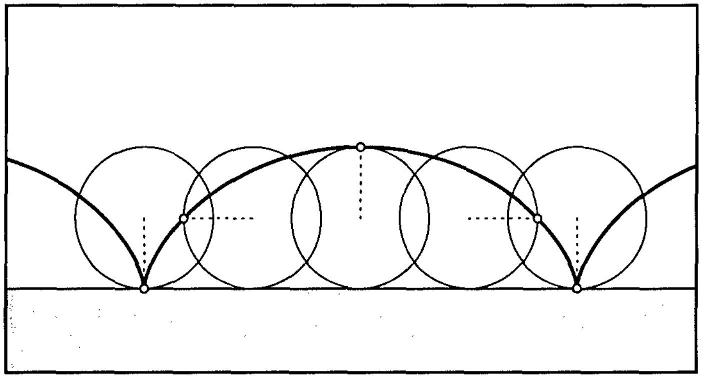

Fig. 2. Cycloid: the length of an arch is eight times the radius of the rolling circle.

Ironically, van Heuraet — an associate of Huygens — published his results in an appendix to van Schooten's 1659 Latin version of Descartes, *Geometria a Renato Des Cartes.* Neil's results also appeared in 1659, in the *Tractatus duo, prior de cycloide, posterior de cissoide* published by John Wallis, and Fermat's work followed in 1660 in *De linearum curvarum cum lineis rectis comparatione dissertatio geometrica* — an appendix to a treatise by de Lalouvère (this was the only publication by Fermat to appear during his lifetime).

Christiaan Huygens (1629–1695), in his *Horologium oscillatorium* of 1673, gave a historical account [34] of these rectifications that provoked arguments over the priority he attributed to van Heuraet and Wren for their discovery — see Chapter 8 of [31]. This dispute reflects the philosophical importance of the rectification problem, which had been held impossible through long tradition that originated with Aristotle, was reinforced in the 11th century by Ibn Rushd (Averroes), and culminated in Descartes' dogmatic assertion. Huygens' theory of evolutes and involutes, employed in his design of an isochronous pendulum clock [21], offered profound new insight into this age–old problem. The cubic (5) was recognized as the evolute (locus of centers of curvature) of a parabola, while the cycloid (4) has an identical (displaced) cycloid as its evolute.

All these results preceded a formal development of the calculus. Whereas the latter resolved existential issues concerning arc lengths by defining them, for a (sufficiently smooth) parametric curve $(x(\xi), y(\xi))$, through the integral

$$s(\xi) \,=\, \int_0^\xi \sqrt{x'^2(\tau)+y'^2(\tau)}\,\mathrm{d}\tau\,, \tag{6}$$

there remained the awkward fact that this does not admit analytic reduction except in trivial or exceptional cases, such as the cubic (5).

In fact, with the emergence of differential geometry, it became customary to assume $s \equiv \xi$ — i.e., the integrand in (6) is precisely unity — although this *natural* or *arc-length* parameterization has only a hypothetical existence: it is fundamentally incompatible with curves (except straight lines) parameterized by "simple" functions. This fact is obvious for polynomial curves, but its proof for *rational* curves is subtle [23], involving Pythagorean triples of polynomials, partial fraction decompositions, and the calculus of residues.

An offshoot to this proof was the introduction of Pythagorean hodograph (PH) *curves*, whose hodograph components satisfy the condition

$$x'^2(\xi) + y'^2(\xi) \equiv \sigma^2(\xi)$$

for some polynomial $\sigma(\xi)$, and are thus [22] of the form

$$x'(\xi) = u^2(\xi) - v^2(\xi), \quad y'(\xi) = 2\,u(\xi)v(\xi), \quad \sigma(\xi) = u^2(\xi) + v^2(\xi)$$

where $u(\xi)$, $v(\xi)$ are relatively prime polynomials. For PH curves, the integral (6) evidently reduces to a *polynomial* function of the parameter ξ. This fact proves especially propitious in the formulation of *real–time CNC interpolators* for digital motion control applications (see §6 below).

Venturing beyond PH curves, one may seek to encompass a broader class of loci by allowing more complicated arc–length functions. Suppose we allow s to be an *algebraic function* of the parameter ξ — i.e., there exists a bivariate polynomial $F(\cdot,\cdot)$ such that (6) satisfies

$$F(s(\xi),\xi) = 0\,.$$

An algebraic function cannot, in general, be described by a simple closed–form expression. Nevertheless, one can show [53] that (6) is algebraic if and only if there exists a polynomial $h(\xi)$ such that

$$[\,x'^2(\xi) + y'^2(\xi)\,]\,h(\xi) \equiv h'^2(\xi)\,.$$

As an immediate consequence, if the function (6) is algebraic, it must have the simple form $s(\xi) = 2\sqrt{h(\xi)} + \text{constant}$ (note that the PH curves are subsumed as special instances, corresponding to $h = \frac{1}{4}\,[\,\int u^2 + v^2\,\mathrm{d}\xi\,]^2$). The cubic (5) is the simplest (non–PH) example of these *algebraically–rectifiable curves*, with $h(\xi) = (9k^2\xi^2 + 4)^3/2916k^4$ — indeed, it is the *unique* cubic with this property. See [53] for details on algebraically–rectifiable quartics and quintics.

Since arc–length parameterization by rational functions is fundamentally impossible, it seems natural to ask "how close" we can approach this elusive ideal. Consider, for example, a degree–n polynomial curve $\boldsymbol{r}(\xi)$: a parameter transformation $\xi \in [\,0,1\,] \rightarrow \tau \in [\,0,1\,]$ of the form

$$\xi = \frac{(1-\alpha)\tau}{\alpha(1-\tau) + (1-\alpha)\tau}$$

gives a rational representation of the same degree, and offers a single degree of freedom, α, to control the "parameter flow" over the curve. Using the integral

$$I = \int_0^1 (|\boldsymbol{r}'(\tau)| - 1)^2\,\mathrm{d}\tau \tag{7}$$

as a measure of "closeness" to arc–length parameterization (for which $I = 0$), the value of α that minimizes (7) can be found [18] as the unique root on $(0,1)$ of a quadratic equation; see also [36]. However, this optimal parameterization offers limited scope for improvement, since we fix the curve degree n.

Another approach, based on the polynomial arc–length functions $s(\xi)$ of PH curves, employs the Legendre series to compute a convergent sequence of (constrained) polynomial approximations $\xi_1(s), \xi_2(s), \ldots$ to the *inverse* of this function [19], such that

$$\lim_{k \to \infty} \xi_k(s(\xi)) \equiv 1 \qquad \text{for } \xi \in [\,0,1\,],$$

given the normalization $s \in [\,0,1\,]$. The coefficients of $\xi_k(s)$ can be determined through closed–form reduction of certain integrals. For sufficiently high k, the re–parameterized version $\boldsymbol{r}_k(s) = \boldsymbol{r}(\xi_k(s))$ comes arbitrarily close to the exact arc–length parameterization, although it is formally of degree kn.

§4. Curves from Motion I. Centripetal Forces

By integrating the description of the forces and laws of motion that govern a physical system, the science of dynamics provides a rigorous and quantitative approach to analyzing motions. Perhaps more so than in any other branch of science [12], the theoretical canonization and empirical triumph of dynamics are the fruits of a single pre–eminent mind: Sir Isaac Newton.

Perhaps the simplest (non–trivial) problem of dynamics is that of motion under a centripetal force — i.e., a force always directed toward or away from a fixed center, whose magnitude depends only on distance r from that center. The term *centripetal* — "seeking the center" — was introduced [60] by Newton in his *De motu corporum in gyrum* of 1684 (in recognition of the fact that, to overcome the inertial tendency of a body to move in a straight line, circular motion requires a steady force directed toward a fixed center):

> *I call that, by which a body is impelled or attracted toward some point which is regarded as a center, centripetal force.*

The basic questions concerning centripetal forces are: *what kinds of orbit arise from different dependencies of the force on* r — and, conversely, *knowing the type of orbit, can we deduce the dependence of the force on* r?

In the late 17th century the context for interest in such questions was, of course, the search for an explanation of Kepler's (empirical) laws of planetary motion — namely: (i) the orbits of the planets are ellipses, with the sun at one focus; (ii) the radial line between the sun and a planet sweeps out area at a uniform rate; and (iii) the squares of the orbital periods are proportional to the cubes of the mean distances of planets from the sun. As we now know, these are direct consequences of an *inverse–square* (r^{-2}) gravitational force. Newton discovered this at an early stage in his career, but remained characteristically secretive about it. It is Edmond Halley who deserves credit for coaxing Newton into disseminating his arguments and, ultimately, codifying dynamics through publication of the *Principia*. According [13] to Abraham de Moivre:

> *In 1684 D^r Halley came to visit him at Cambridge, after they had been some time together the D^r asked him what he thought the Curve would be that would be described by the Planets supposing the force of attraction towards the Sun to be reciprocal to the square of their distance from it. S^r Isaac replied immediately it would be an Ellipsis, the D^r struck with joy & amazement asked him how he knew it, why saith he, I have calculated it, whereupon D^r Halley asked him for his calculation without any further delay, S^r Isaac looked among his papers but could not find it, but he promised to renew it, & then send it him ...*

In fact Halley, as Clerk to the Royal Society, printed the *Principia* at his own expense; the Society's funds had been depleted by the production of a *Historia Piscium* (*History of Fishes*) that failed to become a best–seller. Subsequently, Halley's salary was paid entirely in copies of this *Historia Piscium* [13].

The inverse–square nature of gravitational force, now common knowledge, was established by Newton as the rational explanation for Kepler's laws. It is not widely known, however, that the *Principia* thoroughly analyzes a variety of power–law (r^n) centripetal forces, and shows that different integer exponents n yield circular, conic, spiral, and other orbits in an often surprising manner. The profundity of these results — and their anticipation of an elegant theory of *dual centripetal forces* due to K. Bohlin [7] and E. Kasner [38] — has recently been emphasized by Arnol'd [2], Chandrasekhar [10], and Needham [45].

As reflected in the title of Chandrasekhar's recent book [10], the obscurity of such interesting results in the *Principia* is due to Newton's exclusive reliance on forbiddingly Euclidean argumentations. The modern reader, equipped with predominantly analytic/algebraic skills, is usually reduced to a state of dismay and bewilderment upon a first encounter with the *Principia.* Needham [45] argues convincingly that the re–discovery of Newton's "geometrical calculus" is a very rewarding endeavor, and Arnol'd [2] gives an example of the type of problem — the transcendental nature of certain area integrals — that seems "obvious" to Newtonian thinking, but not to modern modes of thought.

To elucidate connections between the *Principia*'s results and the Bohlin–Kasner theory of dual centripetal forces, it is convenient [2,44] to adopt the complex–number representation $z = x + \mathrm{i}\,y = r\,\mathrm{e}^{\mathrm{i}\theta}$ of the Euclidean plane. Under a centripetal force proportional to the n–th power of distance ¿from the origin, the equation of motion is then

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -\,k\,|z|^n\,\frac{z}{|z|}\,. \tag{8}$$

Here, the centripetal force is attractive or repulsive according to whether k is positive or negative. Hooke's law, for example, corresponds to $n = +1$, while $n = -2$ represents a Coulomb (gravitational or electrostatic) force.

There are two "constants of motion" associated with solutions $z(t)$ to the differential equation (8) — the angular momentum and energy,

$$L = \mathrm{Im}\left(\bar{z}\,\frac{\mathrm{d}z}{\mathrm{d}t}\right) = r^2\,\frac{\mathrm{d}\theta}{\mathrm{d}t} \qquad \text{and} \qquad E = \frac{1}{2}\left|\frac{\mathrm{d}z}{\mathrm{d}t}\right|^2 + k\,\frac{|z|^{n+1}}{n+1}\,. \tag{9}$$

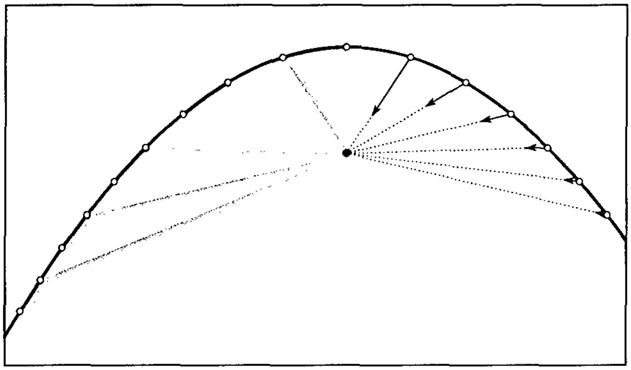

Fig. 3. Constancy of angular momentum L under an r^{-2} centripetal force.

Although we may formally replace $|z|^{n+1}/(n+1)$ by $\ln|z|$ when $n=-1$, this case is usually excluded. For $n \leq -2$, the "potential energy" component of E is naturally negative, and tends to zero as $|z| \rightarrow \infty$. Conversely, for $n \geq 0$, the potential energy is naturally positive and vanishes when $|z|=0$. The function $\ln|z|$, however, diverges for both $|z| \rightarrow 0$ and $|z| \rightarrow \infty$; it does not represent a satisfactory scale–free potential energy with a natural reference value.

The polar form of L given in (9) serves as a reminder of its geometrical interpretation, namely, the rate at which the position vector sweeps out area in the orbit — this constancy of "areal velocity" is expressed by Kepler's second law. Section II of Book I of the *Principia* is devoted to "the determination of centripetal forces," and Newton's immediate concern [46] is to prove:

PROPOSITION I. THEOREM I

The areas which revolving bodies describe by radii drawn to an immovable centre of force do lie in the same immovable plane, and are proportional to the times in which they are described.

PROPOSITION II. THEOREM II

Every body that moves in any curved line described in a plane, and by a radius drawn to a point either immovable, or moving forwards with an uniform rectilinear motion, describes about that point areas proportional to the times, is urged by a centripetal force directed to that point.

In other words, Newton first shows that "centripetal force $\Longleftrightarrow$ $L=$ constant" (this principle is illustrated in Figure 3, for the case of a Coulomb r^{-2} force).

Consider now the conformal map $\boldsymbol{z} \rightarrow \boldsymbol{w}$ of the plane given by

$$\boldsymbol{w} = \boldsymbol{z}^{\alpha}, \tag{10}$$

under which the orbit $\boldsymbol{z}(t)$ determined by equation (8) is transformed into an orbit $\boldsymbol{w}(\tau)$, where t and τ denote times on these orbits corresponding to an angular position θ about the origins of the $\boldsymbol{z}$ and $\boldsymbol{w}$ planes. Writing $\boldsymbol{z} = r\,\mathrm{e}^{\mathrm{i}\theta}$ and $\boldsymbol{w} = \rho\,\mathrm{e}^{\mathrm{i}\theta}$, we stipulate that these orbits have equal angular momentum

$$L = r^2\,\frac{\mathrm{d}\theta}{\mathrm{d}t} = \rho^2\,\frac{\mathrm{d}\theta}{\mathrm{d}\tau} = \text{constant}, \tag{11}$$

and we ask: *is the orbit $\boldsymbol{w}(\tau)$ also the solution to an equation of motion*

$$\frac{\mathrm{d}^2\boldsymbol{w}}{\mathrm{d}\tau^2} = -\,\kappa\,|\boldsymbol{w}|^m\,\frac{\boldsymbol{w}}{|\boldsymbol{w}|} \tag{12}$$

under a power-law centripetal force — and if so, how are the exponents n, m, α related? It transpires that this problem has an elegant and *unique* solution: corresponding force–law exponents n and m are related by

$$(n+3)(m+3) = 4, \tag{13}$$

and the exponent α of the map (10) is given in terms of them by

$$\alpha = \frac{n+3}{2} = \frac{2}{m+3}. \tag{14}$$

To derive equations (12)–(14), we first note from (10) and (11) that derivatives with respect to t and τ are related by

$$\frac{\mathrm{d}}{\mathrm{d}\tau} = |\boldsymbol{z}|^{2(1-\alpha)}\,\frac{\mathrm{d}}{\mathrm{d}t}. \tag{15}$$

Now by applying (15) to $\boldsymbol{w} = \boldsymbol{z}^\alpha$ twice, and invoking (8), we obtain

$$\frac{\mathrm{d}^2\boldsymbol{w}}{\mathrm{d}\tau^2} = 2\,\alpha\,(1-\alpha)\left[\frac{1}{2}\left|\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t}\right|^2 + k\,\frac{|\boldsymbol{z}|^{n+1}}{2(\alpha-1)}\right]|\boldsymbol{z}|^{2-3\alpha}\,\frac{\boldsymbol{z}^\alpha}{|\boldsymbol{z}^\alpha|}.$$

We observe that by choosing $2(\alpha-1) = n+1$, i.e., $\alpha = (n+3)/2$ as in (14), the expression in brackets coincides with the energy constant E in (9). With this choice, substitution from (10) gives

$$\frac{\mathrm{d}^2\boldsymbol{w}}{\mathrm{d}\tau^2} = -\,2\,\alpha\,(\alpha-1)\,E\,|\boldsymbol{w}|^{(2-3\alpha)/\alpha}\,\frac{\boldsymbol{w}}{|\boldsymbol{w}|}.$$

This is of the desired form (12), with $\kappa = 2\alpha(\alpha-1)E$ and $m = (2-3\alpha)/\alpha$, i.e., $\alpha = 2/(m+3)$ as in (14). Finally, equation (13) follows from the individual relations in (14), between n and α, and m and α, derived above.

Equation (13) describes a hyperbolic relation between the force exponents n and m. For each n, except -3, the orbit $\boldsymbol{z}(t)$ determined by (8) is mapped by (10) into a dual orbit $\boldsymbol{w}(\tau)$ of equal angular momentum, determined by (12).

Dual orbits corresponding to *integer* force exponents are of special interest — they are (with $n \geq m$) as follows:

$$\begin{array}{llll} \text{(a)} & n=+1, & m=-2, & z \to z^2, \\ \text{(b)} & n=-1, & m=-1, & z \to z, \\ \text{(c)} & n=-4, & m=-7, & z \to z^{-1/2}, \\ \text{(d)} & n=-5, & m=-5, & z \to z^{-1}. \end{array}$$

Cases (b) and (d) identify "self–dual" forces (as noted above, however, case (b) is usually excluded on physical grounds). Case (a) reveals the beautiful result that an orbit $\boldsymbol{z}(t)$ under a linear Hooke's–law force is mapped by $\boldsymbol{w} = \boldsymbol{z}^2$ to an orbit $\boldsymbol{w}(\tau)$ under an inverse–square Coulomb force. Indeed, Newton shows that the Hooke and Coulomb forms are the *only* centripetal forces that admit conic orbits. In this regard, see [44] for an interesting anecdote concerning the £1 note issued to commemorate the *Principia*'s 300th anniversary.

To investigate the geometry of orbits under power–law centripetal forces, it is convenient to employ polar coordinates (r, θ). Using the fact that $r^2\dot{\theta} = L$, we can write the scalar components of the equation of motion (8) as

$$\ddot{r} - L^2 r^{-3} + k r^n = 0 \qquad \text{and} \qquad 2\dot{r}\dot{\theta} + r\ddot{\theta} = 0\,, \tag{16}$$

where dots denote time derivatives. To eliminate the time variable, and obtain a purely geometrical description of the orbit, we set $u = 1/r$ and note that

$$\frac{\mathrm{d}}{\mathrm{d}t} = \dot{\theta}\,\frac{\mathrm{d}}{\mathrm{d}\theta} = Lu^2\,\frac{\mathrm{d}}{\mathrm{d}\theta}\,.$$

With $\beta = k/L^2$, the first of equations (16) can then be transformed [10,57] to

$$u'' + u - \beta\, u^{-(n+2)} = 0\,, \tag{17}$$

where primes denote derivatives with respect to θ. By solving this differential equation, we obtain polar–coordinate expressions $r(\theta) = 1/u(\theta)$ describing the shapes of orbits. One may verify the functional form and geometrical nature of some representative solutions known to Newton:

$$\begin{array}{lll} n = +1: & r(\theta) = (\beta\sin^2\theta + \cos^2\theta)^{-1/2} & \text{ellipse, center at origin;} \\ n = -2: & r(\theta) = (\beta + \cos\theta)^{-1} & \text{ellipse, focus at origin;} \\ n = -3: & r(\theta) = \exp\left(\sqrt{\beta - 1}\,\theta\right) & \text{logarithmic spiral;} \\ n = -5: & r(\theta) = \sqrt{\beta/2}\,\cos\theta & \text{circle through origin.} \end{array}$$

The cases $n = 1$ and -3 are treated in PROPOSITION X. PROBLEM V and PROPOSITION IX. PROBLEM IV of the *Principia*, Book I. The case

$n = -5$ is pathological, since the orbit passes *through* the center of force! In PROPOSITION V. PROBLEM II, Newton actually treats a generalization of the $n = -5$ case — he shows that a particle $\boldsymbol{p}$ will execute a circular orbit if it is attracted to any center $\boldsymbol{c}$ by a force proportional to $r^{-2}\ell^{-3}$, where r is the distance of $\boldsymbol{p}$ from $\boldsymbol{c}$, and ℓ is the length of the chord containing $\boldsymbol{p}$ and $\boldsymbol{c}$. When $\boldsymbol{c}$ lies *on* the circle, we have $\ell = r$, and hence an r^{-5} force.

In Section III of Book I, Newton is concerned with "the motion of bodies in eccentric conic sections." He treats the case $n = -2$ in PROPOSITION XI. PROBLEM VI, and also discusses parabolic and hyperbolic orbits. Kepler's third law is also derived, in PROPOSITION XV. THEOREM VII.

Incidentally, the case $n = 1$ provides the correct solution to the problem of motion in the gravity of a permeable rotating Earth, considered by Torricelli (see §3). If the Earth is a homogeneous sphere of mass M and radius R, the gravitational force at distance r from the center is equal to

$$\frac{GM(r/R)^3}{r^2},$$

G being the gravitational constant — i.e., it is proportional to r. Thus, the path is an ellipse, and not the logarithmic spiral suggested by Torricelli (which requires an r^{-3} force). Actually, it is a very *shallow* ellipse — the minor axis is smaller than the major axis R by the dimensionless factor $\sqrt{GM/\omega^2 R^3} \approx 291$, where $\omega = 2\pi$ rads/day (this factor is the ratio of the orbital velocity at $r = R$ to the tangential velocity ωR at the equator due to the Earth's rotation).

A shallow ellipse is an obvious perturbation to Galileo's simple harmonic motion through a non-rotating permeable Earth, and in retrospect Torricelli's conjectured spiral trajectory — revived by Newton in 1679 — may seem rather naive. Newton soon redeemed himself, however, through the publication of his *Principia* in 1686, which contains the correct solution as part of a remarkably comprehensive theory of orbital motions under centripetal forces.

It is a sobering experience, for the modern reader, to pierce the *Principia*'s veil of geometrical argumentations, and appreciate its profound insights. To contemporaries, Newton's creation was a virtually miraculous event — Halley composed an ode to preface the *Principia*, extolling "the illustrious man" and his work, "a signal distinction of our time and race" [46]:

Matters that vexed the minds of ancient seers,
And for our learned doctors often led
To loud and vain contention, now are seen
In reason's light, the clouds of ignorance
Dispelled at last by science ...

However, Newton was not exempt from the sarcasm of critics, such as the poet Alexander Pope [51], who were often more eloquent in their converse views:

Superior beings, when of late they saw
A mortal man unfold all Nature's law,
Admired such wisdom in an Earthly shape,
And showed a NEWTON as we show an ape.

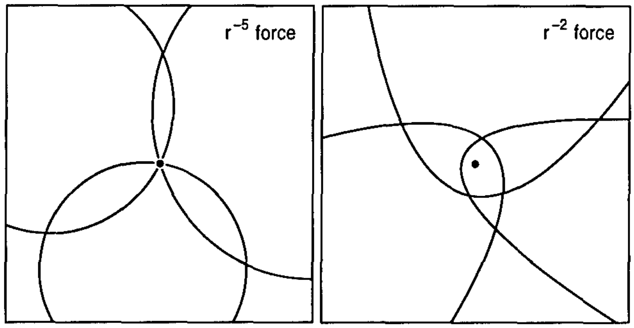

Fig. 4. Geometry of zero–energy orbits under r^{-5} and r^{-2} centripetal forces.

To conclude, we note that the singular orbit $r(\theta) = \sqrt{\beta/2}\,\cos\theta$, identified above for an r^{-5} force, corresponds to the case where the orbital energy

$$E = \frac{\dot{r}^2}{2} + \frac{L^2}{2r^2} - \frac{k}{4r^4}$$

is zero: it is the analog of zero–energy parabolic orbits under an r^{-2} force (see Figure 4). As $r \to 0$ ($t \to \pi k/8L^3$), the positive kinetic energy and negative potential energy both become infinite in a manner such as to maintain $E = 0$. The angular momentum L is also conserved, in a limiting sense, as $r \to 0$.

For orbits with $E \neq 0$, equation (17) can be integrated to obtain

$$\sqrt{2}\int_{1/r_0}^{1/r} \frac{\mathrm{d}u}{\sqrt{\beta u^4 - 2u^2 + \gamma}} = \theta$$

where $\gamma = E/L^2$ and $r = r_0$ for $\theta = 0$. A further reduction, giving r explicitly in terms of θ, is possible upon introducing Jacobian elliptic functions [41], but we shall not pursue it here. A special case is a circular orbit $u = u_0 = 1/r_0$, of energy $E = \frac{1}{4}(2L^2 - ku_0^2)u_0^2$, but this is highly unstable — any perturbation will cause r to rapidly decay to 0 or grow to ∞ (see Figure 5, comparing orbits under r^{-5} and r^{-2} forces perturbed by introducing an initial negative/positive radial velocity corresponding to a 10^{-3} fractional change in E).

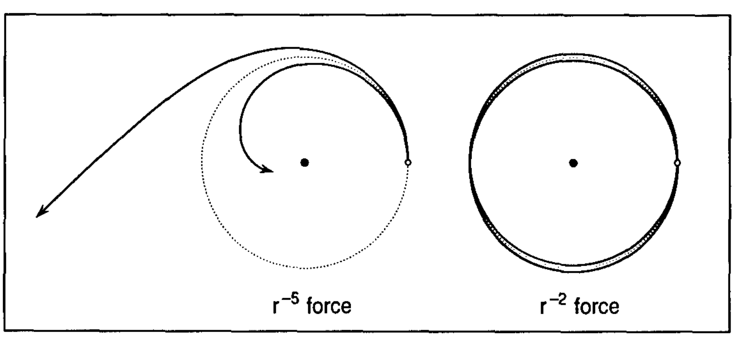

Fig. 5. Relative stability of circular orbits under r^{-5} and r^{-2} centripetal forces.

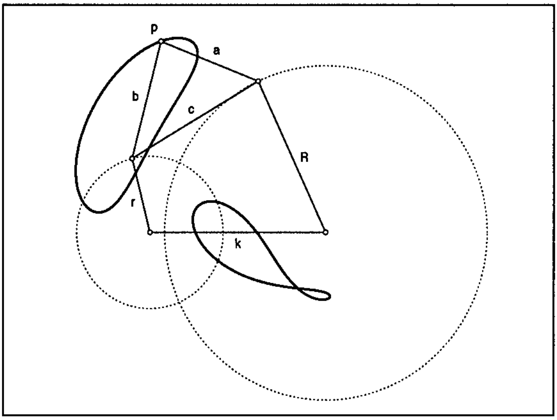

Fig. 6. Geometrical dimensions of a planar four–bar linkage.

In fact, only the $n = +1$ and -2 force laws admit *stable* periodic orbits, in the sense that perturbations to them always produce similar, "neighboring" closed orbits (a proof is given in Appendix A of Goldstein [28]).

§5. Curves from Motion II. Four–bar Linkage

A mechanism is a device that is designed to transform "input" motions and forces, from a given power source, into "output" motions and forces — better suited for use in some practical application. A mechanism typically comprises several rigid members connected by joints that allow certain types of relative motion. According to the Kempe theorem [39], mechanisms that employ only revolute and prismatic joints can (in principle) be designed to produce motion along *any* plane algebraic curve. We consider here the coupler curves of planar four–bar linkages, which serve to transform a rotational input motion into an output motion along some general curved trajectory.

(In early studies, such mechanisms were called *three*–bar linkages, since the "ground" link was not counted. It is now customary to include it, to give a closed kinematic chain. The idea of kinematic chains was introduced by the German engineer Franz Reuleaux, in his *Theoretische Kinematik* of 1875).

Figure 6 shows the geometrical configuration of a four–bar linkage. Such mechanisms are found in diverse contexts (windshield wipers, electric shavers, cranes, etc.). Historically, the most famous example was the "parallel motion" mechanism devised by James Watt (1736–1819) for his double–acting steam engine of 1782. In Figure 6, the link of length k is held fixed, while links of length r and R pivot about its two endpoints. These links are connected by a further link of length c, whose ends are thus constrained to lie on circles of radii r and R centered on $(0, 0)$ and $(k, 0)$. A point at a fixed position relative to the link of length c thus traces a locus, called the coupler curve, when the links of length r and R rotate (see Figure 7). We identify a specific point $\boldsymbol{p}$ by taking it to be the apex of a triangle of sides a and b, with base c.

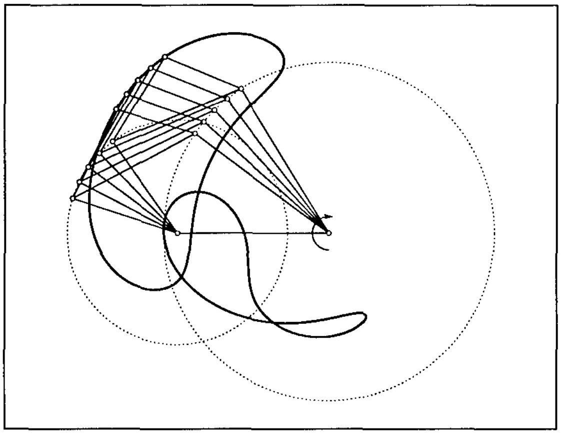

Fig. 7. Generation of a coupler curve by a four–bar linkage.

According to the Grashof theorem [29], the sum of lengths of the shortest and longest links should not exceed the sum of lengths of the other two links, if there is to be continuous relative rotation between two links. This condition is satisfied by the configuration shown in Figure 6, a *crank–rocker* mechanism: the link of length r (the crank) completes full revolutions, while that of length R (the rocker) oscillates through partial revolutions; the link of length c is the coupler. Other possible mechanisms — the *double-crank* or *double-rocker* — are obtained by varying the lengths k, r, R, and c of the links.

Four–bar linkages are capable of generating a rich variety of curved paths: the Hrones–Nelson "atlas" [32], for example, illustrates over *seven thousand* different forms of the coupler curve! These paths are all (parts of) an algebraic curve of degree 6, that depends on six parameters — the dimensions k, r, R, a, b, and c. Its equation can be succinctly expressed [4] in the form

$$f(x,y) \,=\, u^2(x,y) \,+\, v^2(x,y) \,-\, w^2(x,y) \,=\, 0 \tag{18}$$

where

$$\begin{aligned} u(x,y) \,&=\, a\,[\,(x-k)\cos\gamma + y\sin\gamma\,]\,(x^2+y^2+b^2-r^2) \\ &\quad -\, bx\,[\,(x-k)^2+y^2+a^2-R^2\,]\,, \\ v(x,y) \,&=\, a\,[\,(x-k)\sin\gamma - y\cos\gamma\,]\,(x^2+y^2+b^2-r^2) \\ &\quad +\, by\,[\,(x-k)^2+y^2+a^2-R^2\,]\,, \\ w(x,y) \,&=\, 2ab\,[\,x(x-k)\sin\gamma + y^2\sin\gamma - ky\cos\gamma\,]\,, \end{aligned} \tag{19}$$

and we have introduced the angle $\gamma = \cos^{-1}(a^2+b^2-c^2)/2ab$ in lieu of c. The curve defined by (18) and (19) has an ordinary triple point at each of the two circular points at infinity, and three affine double points (two of which may be complex conjugates) that always lie [4] on the circle

$$x^2 \,-\, kx \,+\, y^2 \,-\, ky\cot\gamma \,=\, 0\,, \tag{20}$$

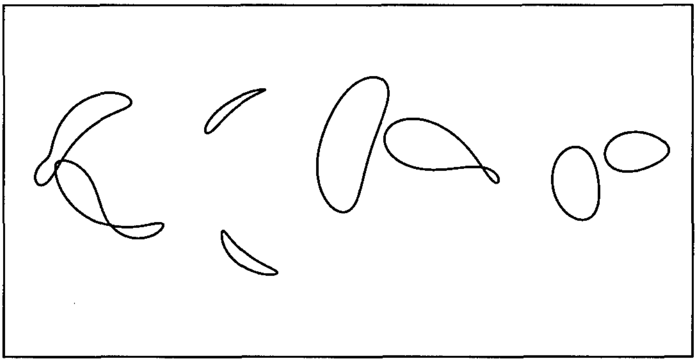

Fig. 8. Coupler curves for various values of the parameters k, r, R, a, b, c.

illustrated in the example in Figure 9 below. The coupler curve falls one short of the maximum of 10 double points that an algebraic curve of degree 6 may have, and is therefore of genus 1 — i.e., it is an *elliptic* curve.

For a crank–rocker, the curve defined by (18) and (19) always comprises two real loops. The physical mechanism traces just one of them: to trace the other loop, the initial configuration of the linkage must be changed. Further examples of crank–rocker coupler curves are shown in Figure 8. For double–crank and double–rocker mechanisms, the coupler curve also has two loops, but in the latter case the mechanism cannot trace either loop entirely. Four–bar linkages that do *not* satisfy the Grashof condition exhibit single–loop coupler curves, which may self–intersect (as in the Figure 9 example — note that the mechanism cannot trace the entire curve). Equations (18) and (19) encompass all these forms for suitable choices of k, r, R, a, b, and c.

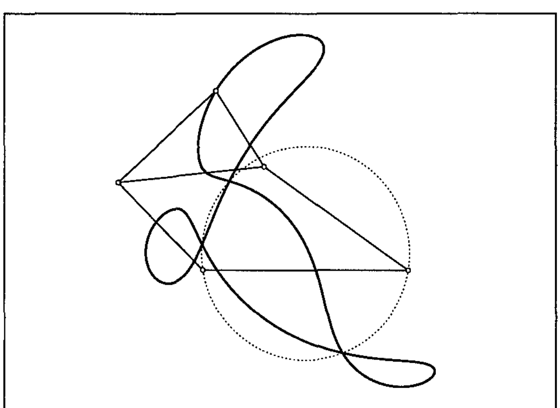

Fig. 9. Single–loop coupler curve for a non–Grashof mechanism.

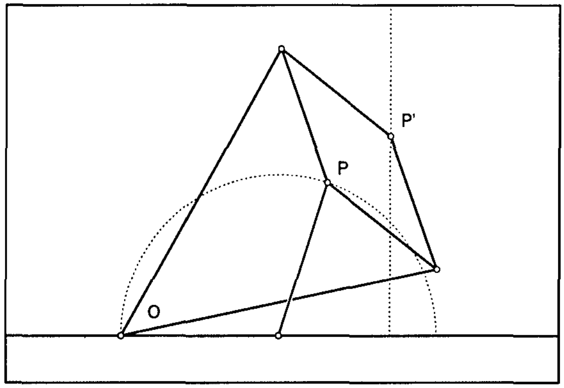

Fig. 10. Peaucellier mechanism: circular motion of P yields linear motion of P'.

A remarkable property of four–bar linkage coupler curves is expressed by the Roberts–Chebyshev theorem [50] — a given coupler curve may actually be traced by *three different four–bar mechanisms* (which are said to be cognates with respect to the given curve). This is not merely a mathematical curiosity: once a path is realized through a specific mechanism, one of its cognates may be found to produce better force transmission characteristics.

The design of a mechanism for a given path may be based on consultation of an "atlas" of coupler curves [32], or use of numerical methods to find the linkage dimensions that will give a locus interpolating prescribed points [6,58]. Allowing for freedoms in the choice of coordinate system, the general coupler curve defined by (18) and (19) can be made to interpolate nine points, but determining the mechanism parameters involves solving a formidable system of non–linear algebraic equations. Wampler et al. [58] have shown, for example, that (counting cognates) the nine–point problem has 4326 solutions, many of which may correspond to complex values for the link dimensions, or interpolate the discrete points on incompatible portions of the coupler curve.

As previously noted, one of the first applications of four–bar linkages was to transform the reciprocating linear motion of a piston into rotary motion of a shaft. Watt, Chebyshev, Roberts, and others proposed approximate solutions to this problem, but linkages that offer *exact* transformations between linear and circular motion were not known until 1864, when a captain in the French army, A. Peaucellier, devised the mechanism shown in Figure 10, comprising four links of length a and two of length b ($> a$). If the point P is constrained to move on a circle passing through the pivot O, the point P' traces a straight line, such that O, P, P' always remain collinear, and the distances r and r' of P and P' from O satisfy $rr' = b^2 - a^2$. Thus, P and P' are images of each other under *inversion* in the circle with center O and radius $\sqrt{b^2 - a^2}$.

Further details on the geometrical and kinematical properties of coupler curves may be found in standard texts [4,16,30,33] on kinematics — see also [47] for an interesting history of coupler curve synthesis methods.

§6. Motion from Curves I. Multi–axis CNC Machines

We have described above how interesting curves can arise from motions under specified kinematical or dynamical constraints. Computer numerical control (CNC) technology is concerned with the converse problem — i.e., the physical realization (by cutting tools, robot arms, sensors, etc.) of motions specified by geometrical paths and given speeds or feedrates along these paths.

To produce a desired motion, a CNC machine must drive each of its axes in an independent but coordinated manner. The controller algorithm employs digital representations of space and time: the unit of time, or sampling interval (typically $\sim$ 1 millisecond), is defined by a "clock" within the algorithm, while the *basic length unit* (BLU, typically $\sim$ 10 microns), or spatial resolution of the machine, is determined by position encoders mounted on its axes.

Within each sampling interval Δt, the controller must compare the actual position of each axis (as measured by the encoders) with the intended position (computed from the specified paths and feedrates by a real–time interpolator). The discrepancy between the actual and desired positions is used to generate control signals for the machine drives, ensuring that the specified paths/speeds are accurately realized. The discrete positions on a curved path $\boldsymbol{r}(\xi)$ computed by the interpolator are known as reference points — they are identified by the sequence $\xi_0, \xi_1, \xi_2, \ldots$ of parameter values satisfying $\xi_0 = 0$ and

$$\int_{\xi_{k-1}}^{\xi_k} \frac{\sigma \, \mathrm{d}\xi}{V} = \Delta t \qquad \text{for } k = 1, 2, \ldots \,, \tag{21}$$

where $\sigma(\xi) = |\boldsymbol{r}'(\xi)|$ is the parametric speed of the curve and V is the (constant or variable) feedrate. Since the integral does not ordinarily have a closed–form reduction, and the unknowns are *limits of integration*, equation (21) is difficult to solve accurately and efficiently — even if V is constant.

Because of this computational difficulty, it is customary to employ simple (piecewise–linear/circular) "G code" approximations to curved tool paths [1]. Compared to its electromechanical hardware sophistication, the part program data that drives a CNC machine is embarrassingly crude. Some authors [11,61] have proposed to drive CNC machines along general curved paths by invoking approximate solutions to (21), based on the Taylor–series expansion

$$\xi_k = \xi_{k-1} + \frac{V}{\sigma} \Delta t + \frac{V}{\sigma^2} \left(V' - \frac{\boldsymbol{r}' \cdot \boldsymbol{r}''}{\sigma^2} V \right) \frac{(\Delta t)^2}{2} + \cdots \,, \tag{22}$$

where primes indicate derivatives with respect to the curve parameter ξ, and it is understood that σ, $\boldsymbol{r}'$, $\boldsymbol{r}''$, V, V', etc., are evaluated at ξ_{k-1}.

The extension of (22) to cubic and higher–order terms incurs complicated coefficients, and is thus ill–suited to real–time computation. Truncation errors are inevitable with this approach (most implementations, in fact, retain only the linear term). Note also that, for a non–constant feedrate, the variation of V cannot be usefully specified as a function of ξ. It must be given in terms of

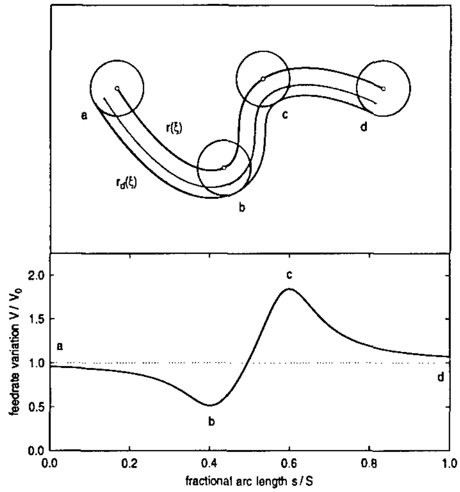

Fig. 11. Curvature dependent feedrate (24) for a constant material removal rate.

a physically meaningful variable, such as time t, arc length s, or curvature κ — in terms these variables, the derivative of V in (22) may be written as

$$V' = \frac{\sigma}{V}\frac{\mathrm{d}V}{\mathrm{d}t} = \sigma\frac{\mathrm{d}V}{\mathrm{d}s} = \sigma\frac{\mathrm{d}\kappa}{\mathrm{d}s}\frac{\mathrm{d}V}{\mathrm{d}\kappa}\,.$$

The Pythagorean–hodograph curves offer an elegant and rigorous solution to this dilemma [24]. For PH curves, the integral (21) admits a simple analytic reduction — not only for constant feedrate, but also varying feedrates specified in a number of useful ways, such as:

(a) any function $V(t)$ of *time* with known indefinite integral;
(b) a linear or quadratic polynomial $V(s)$ in the *arc length*;
(c) simple rational expressions $V(\kappa)$ in the local *curvature*;
(d) constant feedrate V along an *offset* to a specified curve.

In (almost) all these cases, the interpolation equation (21) reduces to the form

$$s(\xi_k) = F(\ldots,\xi_{k-1}), \tag{23}$$

where $s(\xi)$ is the polynomial arc–length function of the PH curve, and F is a known elementary function of the parameters describing the feedrate variation and the preceding reference point ξ_{k-1}. Since $s(\xi)$ is a *monotone* polynomial, equation (23) has a unique real root for the value of ξ_k, which may be obtained to machine precision by a few Newton–Raphson iterations starting from ξ_{k-1}.

The ability to perform real–time interpolation with continuously varying feedrates is extremely useful in a variety of practical problems. For example, the curvature–dependent feedrate

$$V(\kappa) = \frac{V_0}{1+\kappa(d-\frac{1}{2}\delta)} \tag{24}$$

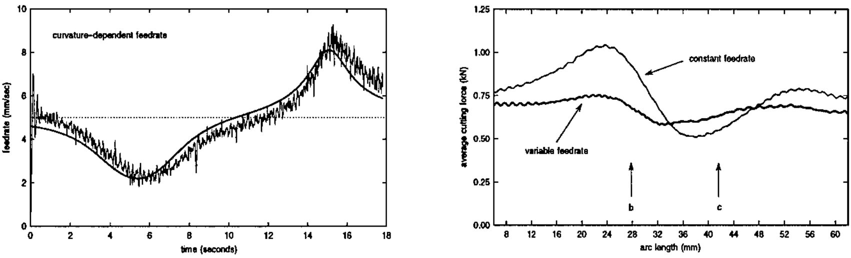

Fig. 12. Measured feedrate (left) and cutting force (right) using the function (24).

can suppress machining force variations [20] when a fixed depth of cut δ is to be removed along a curved path $\mathbf{r}(\xi)$ with a tool of radius d (see Figure 11: the material removal rate is higher at the "concave" location b than the "convex" location c if a constant feedrate V_0 is employed). Figure 12 shows the measured feedrate and (time–averaged) cutting force, as obtained from an experimental implementation of the PH curve interpolator for (24) on a CNC mill.

An important use of *time*–dependent feedrates $V(t)$ is the specification of smooth accelerations and decelerations along curved paths. This is especially important for high–speed machining, in which the dynamical issues of starting and stopping high–speed motions on curved paths become a serious concern. With G code part programs, for example, acceleration/deceleration intervals may span many short linear/circular segments, and thus require a cumbersome real–time block "look ahead" capability for their implementation.

Using PH curves, on the other hand, it is easy to specify a smooth feed acceleration from feedrate $V = 0$ for $t < 0$ up to a desired constant feedrate V_m for $t > T$ along a curve $\mathbf{r}(\xi)$. If $\tau = t/T$ is the "normalized" time during the acceleration interval $t \in [\,0, T\,]$, we use the polynomial feedrate function

$$V(\tau) \,=\, \sum_{k=0}^{n} V_k \binom{n}{k} (1-\tau)^{n-k} \tau^k \tag{25}$$

of *odd* degree n, with $V_0 = \cdots = V_{(n-1)/2} = 0$ and $V_{(n+1)/2} = \cdots = V_n = V_m$. This gives $C^{(n-1)/2}$ continuity with $V = 0$ for $t < 0$ and $V = V_m$ for $t > T$ — in particular, $n = 3$ and $n = 5$ yield C^1 and C^2 feedrate variations.

The interpolation equation for such time–dependent feedrate functions on PH curves is remarkably simple — it is precisely of the form (23), with the right–hand side being simply the integral $F(\tau)$ of (25), a polynomial of degree $n+1$. Only the right–hand side "constant" changes on using higher degrees n, and the incremental cost of evaluating this constant, in each sampling interval Δt, is insignificant for any "reasonable" choice of the degree n.

There are many other possibilites for specifying and optimizing feedrates along PH curves. For example, curved paths are realized on CNC machines by coordinated motions of independently–powered axes, and the chosen feedrates and feed accelerations should not impose demands on the motors of each axis that exceed their torque or power capacity. For PH curves, a thorough analysis of this problem is possible [25], allowing an *a priori* determination of safe fixed feedrates and minimum feed acceleration intervals for a given path geometry.

§7. Motion from Curves II. Quaternion Methods

We have only been concerned, thus far, with the motion of *points* in *Euclidean spaces.* The problems of motion in non–Euclidean spaces, or motion of bodies of finite extent (involving changes of position *and* orientation), are much more challenging — they arise frequently in animation, robotics, 5–axis machining, dynamics, and many other applications. Thus, to conclude, we briefly consider some basic problems in the use of quaternions [5,8,52] to formulate spatial rigid body motions as time–parameterized loci in non–Euclidean spaces.

This subject has seen intense interest [17,27,35,37,40,48,54,59] — see also the extensive bibliography in [52] — in recent years, and substantial progress in "motion design" has been made. Nevertheless, the fundamental problems we encounter with purely translational (point) motion carry over to and, indeed, become much deeper in the context of general spatial motions.

The revival of quaternions in motion–design applications helps remedy a steady historical decline of interest in them. The introduction of quaternions by Sir William Hamilton (1805–1865) predates (and subsumes) development of the "ordinary" vector analysis in $\mathbb{R}^3$ by James Clerk Maxwell (1831–1879), Josiah Willard Gibbs (1839–1903), and Oliver Heaviside (1850–1925), who — along with later generations of physicists — considered quaternions to be an unduly cumbersome medium for describing the laws of nature [14].

A general displacement in Euclidean 3–space $\mathbb{R}^3$ can be interpreted as a screw displacement — i.e., a rotation about a fixed axis and a translation along that axis [8]. Six parameters are required to describe such displacements. Let $\boldsymbol{P} = (X, Y, Z)^T$ and $\boldsymbol{p} = (x, y, z)^T$ be point coordinates in a "fixed" frame Σ and a "movable" frame σ. The spatial displacement carrying Σ into σ may be described by an orthogonal rotation matrix $\boldsymbol{M}$ and a translation vector $\boldsymbol{d}$:

$$\boldsymbol{P} = \boldsymbol{M}\boldsymbol{p} + \boldsymbol{d}\,.$$

The rotation matrix can be expressed in terms of Euler parameters

$$c_0 = \cos\tfrac{1}{2}\phi\,, \quad c_1 = \lambda\sin\tfrac{1}{2}\phi\,, \quad c_2 = \mu\sin\tfrac{1}{2}\phi\,, \quad c_3 = \nu\sin\tfrac{1}{2}\phi,$$

where (λ, μ, ν) are direction cosines of the axis and ϕ is the rotation angle, as

$$\boldsymbol{M} = \begin{bmatrix} c_0^2+c_1^2-c_2^2-c_3^2 & 2(c_1c_2-c_0c_3) & 2(c_1c_3+c_0c_2) \\ 2(c_2c_1+c_0c_3) & c_0^2-c_1^2+c_2^2-c_3^2 & 2(c_2c_3-c_0c_1) \\ 2(c_3c_1-c_0c_2) & 2(c_3c_2+c_0c_1) & c_0^2-c_1^2-c_2^2+c_3^2 \end{bmatrix}.$$

Note that, since the Euler parameters satisfy the normalization condition

$$c_0^2 + c_1^2 + c_2^2 + c_3^2 = 1\,, \tag{26}$$

only three are independent. In lieu of the translation vector $\boldsymbol{d} = (d_1, d_2, d_3)$, Study [56] introduced four new parameters:

$$\begin{aligned} c_0' &= (c_1 d_1 + c_2 d_2 + c_3 d_3)/2\,, \\ c_1' &= (-c_0 d_1 - c_3 d_2 + c_2 d_3)/2\,, \\ c_2' &= (c_3 d_1 - c_0 d_2 - c_1 d_3)/2\,, \\ c_3' &= (-c_2 d_1 + c_1 d_2 - c_0 d_3)/2\,, \end{aligned}$$

which, by definition, satisfy the constraint

$$c_0 c_0' + c_1 c_1' + c_2 c_2' + c_3 c_3' = 0\,. \tag{27}$$

The set of all displacements in $\mathbb{R}^3$ can then be regarded as points, or soma, in a 6–dimensional space spanned by eight coordinates $(c_0, c_1, c_2, c_3, c_0', c_1', c_2', c_3')$ subject to the two algebraic constraints (26) and (27).

A compact and elegant algebraic description of these soma is obtained [8] by combining the eight coordinates into a dual quaternion of the form

$$\mathcal{C} = c_0 + \varepsilon c_0' + \mathrm{i}\,(c_1 + \varepsilon c_1') + \mathrm{j}\,(c_2 + \varepsilon c_2') + \mathrm{k}\,(c_3 + \varepsilon c_3')\,. \tag{28}$$

Here, the quaternion basis elements satisfy the multiplication rules

$$\mathrm{i}\,\mathrm{j} = \mathrm{k}\,, \quad \mathrm{j}\,\mathrm{k} = \mathrm{i}\,, \quad \mathrm{k}\,\mathrm{i} = \mathrm{j}\,, \quad \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$$

(so that multiplication is non–commutative: $\mathrm{j}\,\mathrm{i} = -\,\mathrm{i}\,\mathrm{j}$, etc). The components of the quaternion (28) are "dual numbers" of the form $x + \varepsilon\, x'$ for real x, x' — the dual basis element ε satisfies $\varepsilon^2 = 0$ $(\neq \varepsilon)$. The relations (26), (27) and $\varepsilon^2 = 0$ ensure that (28) is a *unit* dual quaternion: its components satisfy

$$(c_0 + \varepsilon c_0')^2 + (c_1 + \varepsilon c_1')^2 + (c_2 + \varepsilon c_2')^2 + (c_3 + \varepsilon c_3')^2 = 1\,.$$

For a spatial displacement specified by the unit dual quaternion (28), we can extract the geometrical parameters as follows:

$$\boldsymbol{d} = 2\,(c_0'\boldsymbol{c} - c_0\boldsymbol{c}' - \boldsymbol{c} \times \boldsymbol{c}')\,, \quad \phi = 2\,\cos^{-1} c_0\,, \quad (\lambda, \mu, \nu) = \frac{\boldsymbol{c}}{\sin\frac{1}{2}\phi}\,,$$

where $\boldsymbol{c} = (c_1, c_2, c_3)$, $\boldsymbol{c}' = (c_1', c_2', c_3')$, $0 \leq \cos^{-1} c_0 \leq \pi$, and $\times$ is the familiar vector cross product in $\mathbb{R}^3$. Note also that the translation distance is

$$d = |\boldsymbol{d}| = 2\sqrt{c_0'^2 + c_1'^2 + c_2'^2 + c_3'^2}\,.$$

When $c_0' = c_1' = c_2' = c_3' = 0$ the dual quaternion (28) specifies a pure rotation (the non–commutativity of quaternion multiplication reflects the importance of the *order* in which spatial rotations are executed — see Figure 13).

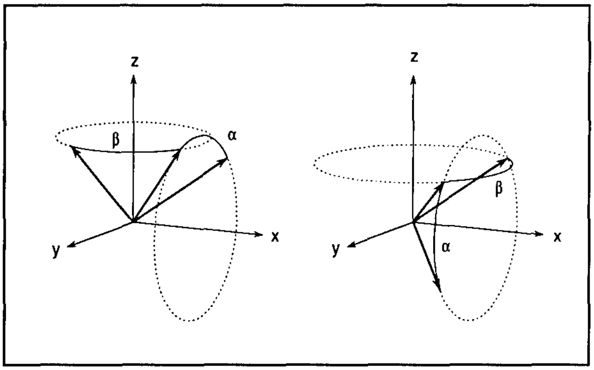

Fig. 13. Effect of rotations $R_x(\alpha)R_z(\beta)$ and $R_z(\beta)R_x(\alpha)$ applied to a vector $\boldsymbol{v}$.

The principal advantage of the form (28) is that the outcome of successive displacements, $\mathcal{A}$ followed by $\mathcal{B}$, corresponds to the ordered product $\mathcal{C} = \mathcal{B}\mathcal{A}$ of their dual quaternion representations. With

$$\begin{aligned}\mathcal{A} &= a_0 + \varepsilon a_0' \,+\, \mathrm{i}\,(a_1 + \varepsilon a_1') \,+\, \mathrm{j}\,(a_2 + \varepsilon a_2') \,+\, \mathrm{k}\,(a_3 + \varepsilon a_3')\,,\\ \mathcal{B} &= b_0 + \varepsilon b_0' \,+\, \mathrm{i}\,(b_1 + \varepsilon b_1') \,+\, \mathrm{j}\,(b_2 + \varepsilon b_2') \,+\, \mathrm{k}\,(b_3 + \varepsilon b_3')\,,\end{aligned}$$

the elements of $\mathcal{C}$ are homogeneous quadratic forms in those of $\mathcal{A}$ and $\mathcal{B}$:

$$\begin{aligned}c_0 &= a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3\,,\\ c_1 &= a_1b_0 + a_0b_1 + a_3b_2 - a_2b_3\,,\\ c_2 &= a_2b_0 + a_0b_2 + a_1b_3 - a_3b_1\,,\\ c_3 &= a_3b_0 + a_0b_3 + a_2b_1 - a_1b_2\,,\end{aligned}$$

$$\begin{aligned}c_0' &= a_0b_0' + a_0'b_0 - a_1b_1' - a_1'b_1 - a_2b_2' - a_2'b_2 - a_3b_3' - a_3'b_3\,,\\ c_1' &= a_1b_0' + a_1'b_0 + a_0b_1' + a_0'b_1 + a_3b_2' + a_3'b_2 - a_2b_3' - a_2'b_3\,,\\ c_2' &= a_2b_0' + a_2'b_0 + a_0b_2' + a_0'b_2 + a_1b_3' + a_1'b_3 - a_3b_1' - a_3'b_1\,,\\ c_3' &= a_3b_0' + a_3'b_0 + a_0b_3' + a_0'b_3 + a_2b_1' + a_2'b_1 - a_1b_2' - a_1'b_2\,.\end{aligned}$$

Now a general rigid body motion, involving both translation and rotation, corresponds to a locus of points in soma space, where a value of the time t is associated with each point. It is tempting to invoke a unit dual quaternion $\mathcal{C}(t)$ parameterized explicitly by "simple" — i.e., (piecewise) polynomial or rational — functions of time, and most motion design schemes rely upon this model. Typically, a sequence $\mathcal{C}_1,\ldots,\mathcal{C}_N$ of displacements at times $t_1,\ldots,t_N$ are specified, and one seeks a "smooth" motion interpolating them.

In §2 we emphasized that motion specification is as much concerned with velocities and accelerations — determined by the precise nature of the time parameterization — as with the path geometry, and this precept also holds for spatial rigid body motions. An *ad hoc* or "indirect" time parameterization of a quaternion locus $\mathcal{C}(t)$ can incur linear/angular velocities or accelerations that are undesirable or, at least, only determinable *a posteriori* — after the motion has been specified, rather than being an integral part of its specification.

For rational point motion, the difficulty in simultaneously specifying both path geometry *and* speed along the path arises from the fact that curves do not, in general, admit rational arc length representations. As indicated in §6, the use of special (PH) curves can resolve this problem. For rational rigid body motions, however, the difficulty is not just one of computation, but also the more fundamental issue of how we characterize "distance travelled" in terms of both the translational and orientational components. Should they be treated together, or separately? In other words, can we introduce a suitable metric for soma space that allows us to define "arc length" along a unit dual quaternion locus $C(t)$, and thus formulate methods to specify both positions/orientations *and* linear/angular speeds for rational spatial motions?

Ravani and Roth [49] have proposed, by analogy with elliptic geometry, a metric that yields a *dual number* value for distances in soma space. However, the use of this metric in motion design, or of alternate (real–valued) functions that exhibit the usual properties of metrics, remains to be explored.

§8. Closure

With its opening AXIOMS, or LAWS OF MOTION, the *Principia* establishes uniform motion as the natural state of a free body. The forces that act upon bodies incur deviations from uniform motion, in a deterministic though subtle manner that reveals appealing and useful connections between geometry and kinematics. Although, in this rather brief and eclectic survey, we have offered only a few illustrative anecdotes on this theme, we hope they have stirred the interest of some inquisitive readers, and have thus helped to promote further theoretical developments and practical applications.

References

1. Standard RS–274–D, Interchangeable variable block data format for positioning, contouring, and contouring/positioning numerically controlled machines, Electronic Industries Association, Washington, D.C., 1979.
2. Arnol'd, V. I., *Huygens and Barrow, Newton and Hooke* (translated by E. J. F. Primrose), Birkhäuser, Basel, 1990.
3. Bell, E. T., *Men of Mathematics*, Simon and Schuster, New York, 1937.
4. Beyer, R., *The Kinematic Synthesis of Mechanisms* (translated by H. Kuenzel), McGraw Hill, New York, 1963.
5. Blaschke, W., *Kinematik und Quaternionen*, Deutscher Verlag der Wissenschaften, Berlin, 1960.
6. Blechschmidt, J. L. and J. J. Uicker, Linkage synthesis using algebraic curves, ASME J. Mech. Trans. Auto. Design **108** (1986), 543–548.
7. Bohlin, K., Note sur le problème des deux corps et sur une intégration nouvelle dans le problème des trois corps, Bulletin Astronomique **28** (1911), 113–119.

8. Bottema, O. and B. Roth, *Theoretical Kinematics*, North Holland, 1979.
9. Boyer, C. B. and U. C. Merzbach, *A History of Mathematics*, Wiley, New York, 1991.
10. Chandrasekhar, S. *Newton's Principia for the Common Reader*, Oxford Univ. Press, 1995.
11. Chou, J–J. and D. C. H. Yang, Command axis generation for three–axis CNC machining, ASME J. Eng. Indus. **113** (August 1991), 305–310.
12. Cohen, I. B., *Introduction to Newton's Principia*, Harvard Univ. Press, 1971.
13. Cook, A., *Edmond Halley: Charting the Heavens and the Seas*, Oxford Univ. Press, 1998.
14. Crowe, M. J., *A History of Vector Analysis*, Dover, New York, 1994.
15. Descartes, R., *The Geometry of René Descartes* (translated from French and Latin by D. E. Smith and M. L. Latham), Dover, New York, 1954.
16. Erdman, A. G. and G. N. Sandor, *Mechanism Design* Vols. 1 and 2, Prentice–Hall, Englewood Cliffs, NJ, 1984.
17. Fang, Y. C., C. C. Hsieh, M. J. Kim, J. J. Chang, and T. C. Woo, Real time motion fairing with unit quaternions, Comput. Aided Design **30** (1998), 191–198.
18. Farouki, R. T., Optimal parameterizations, Comput. Aided Geom. Design **14** (1997), 153–168.
19. Farouki, R. T., Convergent inversion approximations for polynomials in Bernstein form, Comput. Aided Geom. Design **17** (2000), 179–196.
20. Farouki, R. T., J. Manjunathaiah, D. Nicholas, G–F. Yuan, and S. Jee, Variable–feedrate CNC interpolators for constant material removal rates along Pythagorean–hodograph curves, Comput. Aided Design **30** (1998), 631–640.
21. Farouki, R. T. and J. Rampersad, Cycles upon cycles: an anecdotal history of higher curves in science and engineering, in *Mathematical Methods for Curves and Surfaces II* (M. Daehlen, T. Lyche, and L. L. Schumaker, eds.), Vanderbilt Univ. Press, Nashville TN (1998), 95–116.
22. Farouki, R. T. and T. Sakkalis, Pythagorean hodographs, IBM J. Res. Develop. **34** (1990), 736–752.
23. Farouki, R. T. and T. Sakkalis, Real rational curves are not "unit speed," Comput. Aided Geom. Design **8** (1991), 151–157.
24. Farouki, R. T. and S. Shah, Real–time CNC interpolators for Pythagorean hodograph curves, Comput. Aided Geom. Design **13** (1996), 583–600.
25. Farouki, R. T., Y–F. Tsai, and C. R. Wilson, Feedrate and feed acceleration constraints along curved tool paths, Comput. Aided Geom. Design (1999), submitted.
26. Galilei, G. (translated by S. Drake), *Dialogue Concerning the Two Chief World Systems* (2nd ed.), Univ. of California Press, 1967.

27. Ge, Q. J. and B. Ravani, Geometric construction of Bézier motions, ASME J. Mech. Design **116** (1995), 749–755.

28. Goldstein, H., *Classical Mechanics* (2nd ed.), Addison–Wesley, Reading MA, 1980.

29. Grashof, F., *Theoretische Maschinenlehre* (3 vols.), Voss, Leipzig, 1883.

30. Hartenberg, R. S. and J. Denavit, *Kinematic Synthesis of Linkages*, McGraw Hill, New York, 1964.

31. Hofmann, J. E., *Leibniz in Paris 1672–1676: His Growth to Mathematical Maturity*, Cambridge Univ. Press, 1974.

32. Hrones, J. A. and G. L. Nelson, *Analysis of the Four–Bar Linkage: Its Application to the Synthesis of Mechanisms*, The Technology Press of MIT, Cambridge, MA, and Wiley, New York, 1951.

33. Hunt, K. R., *Kinematic Geometry of Mechanisms*, Oxford Univ. Press, 1990.

34. Huygens, C., *Horologium Oscillatorium sive De Motu Pendulorum ad Horologia Aptato Demonstrationes Geometricæ*, Paris, 1673; translated into English by R. J. Blackwell as *The Pendulum Clock, or Geometrical Demonstrations Concerning the Motion of Pendula as Applied to Clocks*, Iowa State Univ. Press, Ames, IA, 1986.

35. Jüttler, B., Visualization of moving objects using dual quaternion curves, Computers and Graphics **18** (1994), 315–326.

36. Jüttler, B., A vegetarian approach to optimal parameterizations, Comput. Aided Geom. Design **14** (1997), 887–890.

37. Jüttler, B. and Wagner M., Computer aided design with spatial rational B–spline motions, ASME J. Mech. Design **118** (1996), 193–201.

38. Kasner, E., *Differential–Geometric Aspects of Dynamics*, American Math. Society, New York, 1913.

39. Kempe, A. B., On a general method of describing plane curves of the n–th degree by linkwork, Proc. London Math. Soc. **7** (1876), 213–216.

40. Kim, M.–S. and K.–W. Nam, Interpolating solid orientations with circular blending quaternion curves, Comput. Aided Design **27** (1995), 385–398.

41. Lawden, D. F., *Elliptic Functions and Applications*, Springer, New York, 1989.

42. Minkowski, H., Space and Time, translation of an address delivered at the 80th Assembly of German Natural Scientists and Physicians, Sept. 21, 1908, Köln, in *The Principle of Relativity*, Dover, 1952.

43. Moon, H. P., Minkowski Pythagorean hodographs, Comput. Aided Geom. Design **16** (1999), 739–753.

44. Needham, T., Newton and the transmutation of force, American Math. Monthly **100** (1993), 119–137.

45. Needham, T., *Visual Complex Analysis*, Oxford Univ. Press, 1997.

46. Newton, I., *Philosophiæ Naturalis Principia Mathematica*, 1686; Florian Cajori's revision of Andrew Motte's translation, published as *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World*, Vols. 1 & 2, Univ. of California Press, Berkeley, 1962.

47. Nolle, H., Linkage coupler curve synthesis: A historical review I. Developments up to 1875 & II. Developments after 1875, Mech. Mach. Theory **9** (1974), 147–168 & 325–348.

48. Park, F. C. and B. Ravani, Smooth invariant interpolation of rotations, ACM Trans. Graphics **16** (1997), 277–295.

49. Ravani, B. and B. Roth, Mappings of spatial kinematics, ASME J. Mech. Transm. Auto. Design **106** (1984), 341–347.

50. Roberts, S., Three-bar motion in plane space, Proc. London Math. Soc. **7** (1875), 14–23.

51. Rogers, P. (ed.), *Alexander Pope: A Critical Edition of the Major Works*, Oxford Univ. Press, 1993.

52. Röschel, O., Rational motion design — a survey, Comput. Aided Design **30** (1998), 169–178.

53. Sakkalis, T. and R. Farouki, Algebraically rectifiable parametric curves, Comput. Aided Geom. Design **10** (1993), 551–569.

54. Shoemake, K., Animating rotation with quaternion curves, Computer Graphics **19** (1985), 245–254.

55. Stillwell, J., *Mathematics and Its History*, Springer, New York, 1989.

56. Study, E., *Die Geometrie der Dynamen*, Leipzig, 1903.

57. Symon, K. R., *Mechanics* (3rd ed.), Addison–Wesley, Reading MA, 1971.

58. Wampler, C. W., A. P. Morgan, and A. J. Sommese, Complete solution of the nine–point path synthesis problem for four–bar linkages, *Mechanism Synthesis and Analysis* DE–Vol. 25 (Proceedings of 1990 ASME Design Technical Conference, Chicago), 361–368.

59. Wang, W. and B. Joe, Orientation interpolation in quaternion space using spherical biarcs, Proceedings, Graphics Interface '93, 24–32.

60. Westfall, R. S., *Force in Newton's Physics: The Science of Dynamics in the Seventeenth Century*, American Elsevier, New York, 1971.

61. Yang, D. C. H. and T. Kong, Parametric interpolator versus linear interpolator for precision CNC machining, Comput. Aided Design **26** (1994), 225–234.

Rida T. Farouki
Department of Mechanical and Aeronautical Engineering,
University of California, Davis, CA 95616, USA.
farouki@ucdavis.edu