When action is not least

C. G. Gray[a)]
Guelph-Waterloo Physics Institute and Department of Physics, University of Guelph, Guelph, Ontario N1G2W1, Canada

Edwin F. Taylor[b)]
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 17 February 2006; accepted 26 January 2007)

We examine the nature of the stationary character of the Hamilton action S for a space-time trajectory (worldline) $x(t)$ of a single particle moving in one dimension with a general time-dependent potential energy function $U(x,t)$. We show that the action is a local minimum for sufficiently short worldlines for all potentials and for worldlines of any length in some potentials. For long enough worldlines in most time-independent potentials $U(x)$, the action is a saddle point, that is, a minimum with respect to some nearby alternative curves and a maximum with respect to others. The action is never a true maximum, that is, it is never greater along the actual worldline than along every nearby alternative curve. We illustrate these results for the harmonic oscillator, two different nonlinear oscillators, and a scattering system. We also briefly discuss two-dimensional examples, the Maupertuis action, and newer action principles.

[DOI: 10.1119/1.2710480]

I. INTRODUCTION

Several authors[1–12] have simplified and elaborated the action principle and recommended that it be introduced earlier into the physics curriculum. Their work allows us to see in outline how to empower students early in their studies with the fundamental yet simple extensions of Newton's principles of motion made by Maupertuis, Euler, Lagrange, Jacobi, Hamilton, and others. The simplicity of the action principle derives from its use of a scalar energy and time. Its transparency comes from the use of numerical[11] and analytical[12] methods of varying a trial worldline to find a stationary value of the action, skirting not only the equations of motion but also the advanced formalism of Lagrange and others characteristic of upper level mechanics texts. The goal of this paper is to discuss the conditions for which the stationary value of the action for an actual worldline is a minimum or a saddle point.

For single-particle motion in one dimension (1D), the Hamilton action S is defined as the integral along an actual or trial space-time trajectory (worldline) $x(t)$ connecting two specified events P and R,

$$S=\int_P^R L(x,\dot{x},t)dt, \tag{1}$$

where L is the Lagrangian, x is the position, t is the time, and $P(x_P,t_P)$ and $R(x_R,t_R)$ are fixed initial and final space-time events. A dot, as in $\dot{x}$, indicates the time derivative. The Lagrangian $L(x,\dot{x},t)$ depends on t implicitly through $x(t)$ and may also depend on t explicitly, for example, through a time-dependent potential. For simplicity we use Cartesian coordinates throughout, but the methods and conclusions apply for generalized coordinates.

The Hamilton action principle compares the numerical value of the action S along the actual worldline to its value along every adjacent curve (trial worldline) anchored to the same initial and final events. These alternative curves are arbitrary as long as they are piecewise smooth, have the same end events, and are adjacent to (near to) a worldline that the particle will indeed follow. For example, the adjacent curves need not conserve the total energy. The Hamilton action principle says that with respect to all nearby curves the action along the actual worldline is stationary, that is, it has zero variation to first order; formally we write $\delta S=0$. Whether or not this stationary value of the action is a local minimum is determined by examining $\delta^2 S$ and higher order variations of the action with respect to the nearby curves, as we will discuss in this paper.

Misconceptions concerning the stationary nature of the action abound in the literature. Even Lagrange wrote that the value of the action can be maximum,[13] a common error[14] of which the authors of this paper have been guilty.[12,15] Other authors use *extremum* or *extremal*,[16] which incorrectly includes a maximum and formally fails to include a saddle point. [Mathematicians often use the (correct) term *critical* instead of stationary, but because the former term has other meanings in physics we use the latter.] A similar error mars treatments of Fermat's principle of optics, which is erroneously said to allow the travel time of a light ray between two points to be a maximum.[17]

The present paper has three primary purposes: First it describes conditions under which the action is a minimum and different conditions under which it is a saddle point. These conditions involve second variations of the action. Some pioneers of the second variation theory of the calculus of variations are Legendre, Jacobi, Weierstrass, Kelvin and Tait, Mayer, and Culverwell.[18] Although inspired by the early work of Culverwell,[21] our derivation of these conditions is new, simpler, and more rigorous; it is also simpler than modern treatments.[24,25] Second, this paper explains the results with qualitative heuristic descriptions of how a particle responds to space-varying forces derived from the potential in which it moves. (Those who prefer immediate immersion in the formalism can begin with Sec. IV.) Third, it clarifies these results and illustrates the variety of their consequences by applying them to the harmonic oscillator, two nonlinear oscillators, and a scattering system. Criteria used to decide the nature of the stationary value of the action are also useful for other purposes in classical and semiclassical mechanics,[28] but are not discussed in this paper.

Appendix A adapts the results to the important Maupertuis action W. Appendix B gives examples of both Hamilton and Maupertuis action for two-dimensional motion. Appendix C discusses open questions on the stationary nature of action for some newer action principles.

II. KINETIC FOCUS

This section introduces the concept of *kinetic focus* due to Jacobi,[22] which plays a central role in determining the nature of the stationary action. We start with an analogous example taken verbatim from Whittaker,[36] an analysis of the relative length along different paths. Whittaker employs the Maupertuis action principle (discussed in our Appendix A), which requires fixed total energy along trial paths, not fixed travel time as with the Hamilton action principle. In force-free systems the value of the Maupertuis action is proportional to the path length. The term kinetic focus is defined formally later in this section. Figures 1(a) and 1(b) illustrate this example. Whittaker wrote that "A simple example illustrative of the results obtained in this article is furnished by the motion of a particle confined to a smooth sphere under no forces. The trajectories are great-circles on the sphere and the [Maupertuis] action taken along any path (whether actual or trial) is proportional to the length of the path. The kinetic focus of any point A is the diametrically opposite point A' on the sphere, because any two great circles through A intersect again (for the first time) at A'. The theorems of this article amount, therefore in this case to the statement that an arc of a great circle joining any two points A and B on the sphere is the shortest distance from A to B when (and only when) the point A' diametrically opposite to A does not lie on the arc, that is, when the arc in question is less than half a great-circle."

The elaboration of this analogy is discussed in the captions of Figs. 1(a) and 1(b) using equilibrium lengths of a rubber band on a slippery spherical surface.

For a contrasting example, we apply a similar analysis to free-particle motion in a flat plane. In this case the length of the straight path connecting two points is a minimum no matter how far apart the endpoints. A rubber band stretched between the endpoints on a slippery surface will always snap back when deflected in any manner and released. An alternative straight path that deviates slightly in direction at the initial point A continues to diverge and does not cross the original path again. Therefore no kinetic focus of the initial point A exists for the original path.

Note that on both the sphere and the flat plane there is no path of true maximum length between any two separated points. The length of any path can be increased by adding wiggles.

How do we find the kinetic focus? In Fig. 1(a) we place the terminal point C at different points along a great circle path between A and A'. When C lies between A and A', every nearby alternative path such as AEC is not a true path (a path of minimum length), because it does not lie along a great circle. When terminal point C reaches A', there is suddenly more than one alternative great circle path connecting A and A'. (In this special case an infinite number of alternative great circle paths connect A and A'.) Any alternative great circle path between A and A' can be moved sideways to coalesce with the original path ABA'. The kinetic focus is defined by the existence of this coalescing alternative true path. As the final point C moves away from the initial point A, the kinetic focus A' is defined as the earliest terminal point at which two true paths can coalesce.

The term kinetic focus in mechanics derives from an analogy[37] to the focus in optics, that point A' at which rays

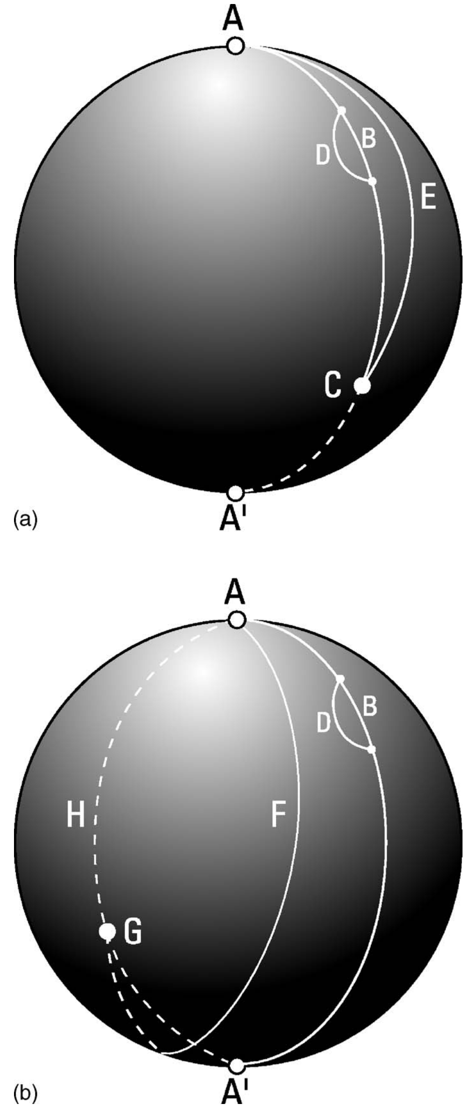

Fig. 1. (a) On a sphere the great circle line ABC starting from the north pole at A is the shortest distance between two points as long as it does not reach the south pole at A'. On a slippery sphere a rubber band stretched between A and C will snap back if displaced either locally, as at D, or by pulling the entire line aside, as along AEC. The point A' is called the *antipode* of A or in general the *kinetic focus* of A. We say that if a great circle path terminates before the kinetic focus of its initial point, the length of the great circle path is a minimum. (b) If the great circle $ABA'G$ passes through antipode A' of the initial point A, then the resulting line has a minimum length only when compared with some alternative lines. For example on a slippery sphere a rubber band stretched along this path will still snap back from local distortion, as at D. However, if the entire rubber band is pulled to one side, as along AFG, then it will not snap back, but rather slide over to the portion AHG of a great circle down the backside of the sphere. With respect to paths like AFG, the length of the great circle line $ABA'G$ is a maximum. With respect to all possible variations we say that the length of path $ABA'G$ is a saddle point. If a great circle path terminates beyond the kinetic focus of its initial point, the length of the great circle path is a saddle point.

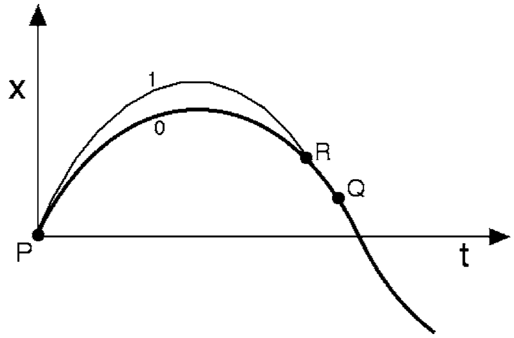

Fig. 2. From the common initial event P we draw a true worldline 0 and a second true worldline 1 that terminates at some event R on the original worldline 0. The event nearest to P at which worldline 1 coalesces with worldline 0 is the kinetic focus Q.

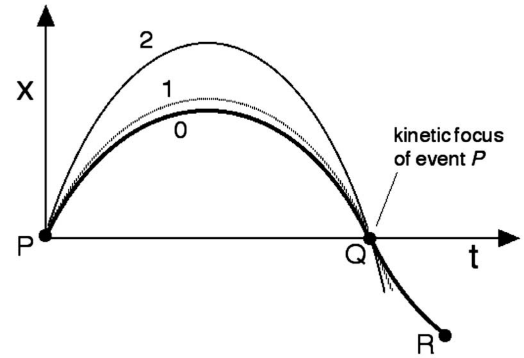

Fig. 3. Several true harmonic oscillator worldlines with initial event $P=(0,0)$ and initial velocity $v_0>0$. Starting at the initial fixed event P at the origin, all worldlines pass through the same event Q. That is, Q is the kinetic focus for all worldlines of the family starting at the initial event $P=(0,0)$. Worldlines 1 and 0 differ infinitesimally; worldlines 2 and 0 differ by a finite amount. This oscillator is discussed in detail in Sec. VIII.

emitted from an initial point A converge under some conditions, such as interception by a converging lens.

This paper deals mainly with the action principle for the Hamilton action S, which determines worldlines in space-time for fixed end-events (that is, end-positions and travel time) rather than the action principle for the Maupertuis action W (see Appendix A), which determines spatial orbits (as well as space-time worldlines) for fixed end-positions and total energy. The kinetic focus for the Hamilton action has a use similar to that for the examples of the Maupertuis action in Figs. 1(a) and 1(b). We will show that a worldline has a minimum action S if it terminates before reaching the kinetic focus of its initial event. In contrast, a worldline that terminates beyond the kinetic focus of the initial event P has an action that is a saddle point.

We use the label P for the initial event on the worldline 0 (see Fig. 2), Q for the kinetic focus of P on the worldline, and R for a fixed but arbitrary event on the worldline that terminates on worldline 0 and also terminates another true worldline (#1 in Fig. 2) connecting P to R. For the Hamilton action S our definition of the kinetic focus of a worldline is the following. The kinetic focus Q of an earlier event P on a true worldline is the event closest to P at which a second true worldline, with slightly different velocity at P, intersects the first worldline, in the limit for which the two worldlines coalesce as their initial velocities at P are made equal.

The kinetic focus is central to the understanding of the stationary nature of the action S, but its definition may seem obscure. To preview the consequences of this definition, we briefly discuss some examples that we will discuss later in the paper. Figure 3 shows the true worldlines of the harmonic oscillator, whose potential energy has the form $U(x)=\left(\frac{1}{2}\right)kx^2$. The harmonic oscillator is the single 1D case of the definition of space-time kinetic focus with the following exceptional characteristic: every worldline originating at P in Fig. 3 passes through the same recrossing point. The 2D spatial paths on the sphere in Figs. 1(a) and 1(b) show the same characteristic: Every great circle path starting at A passes through the antipode at A'. In both cases we can find the kinetic focus without taking the limit for which the velocities at the initial point are equal and the two worldlines coalesce, but we can take that limit. For the harmonic oscillator this limit occurs when the amplitudes are made equal. The harmonic oscillator will turn out to be the only exception to many of the rules for the action.

A more typical case is the quartic oscillator (see Fig. 4), which is described by $U(x)=Cx^4$. In this case alternative worldlines starting from the initial event P can cross anywhere along the original worldline (some crossing events are indicated by little squares in Fig. 4). When the alternative worldline coalesces with the original worldline, the crossing point has reached the kinetic focus Q.

Another typical case is the piecewise-linear oscillator shown in Fig. 5. This oscillator has the potential energy $U(x)=C|x|$. For the piecewise-linear oscillator, as for the quartic oscillator, alternative worldlines starting from P can cross at various events along the original worldline. Note that for the piecewise-linear oscillator an alternative worldline that crosses the original worldline before its kinetic focus lies below the original worldline instead of above it (as for the quartic oscillator). We can equally well use an alternative worldline that crosses from below or one that crosses from above to define the coalescing worldline and kinetic focus.

Notice the gray line labeled *caustic* in Figs. 4 and 5, and also in Fig. 6, which shows the worldlines for a repulsive

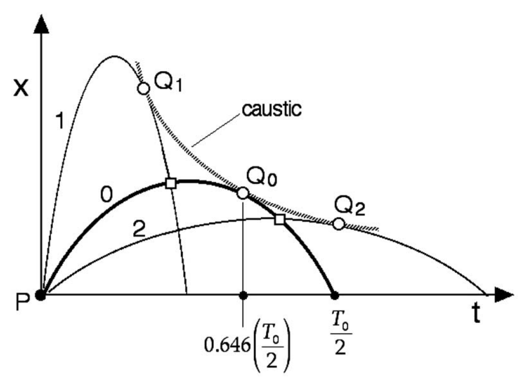

Fig. 4. Schematic space-time diagram of a family of true worldlines for the quartic oscillator $[U(x)=Cx^4]$ starting at $P=(0,0)$ with $v_0>0$. The kinetic focus occurs at a fraction 0.646 of the half-period $T_0/2$, illustrated here for worldline 0. The kinetic foci of all worldlines of this family lie along the heavy gray line, the caustic. Squares indicate recrossing events of worldline 0 with the other two worldlines. This oscillator is discussed in detail in Sec. IX.

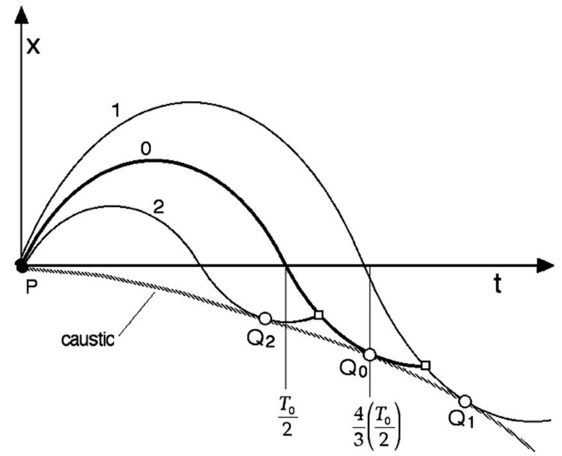

Fig. 5. Schematic space-time diagram of a family of true worldlines for a piecewise-linear oscillator $[U(x)=C|x|]$, with initial event $P=(0,0)$ and initial velocity $v_0>0$. The kinetic focus Q_0 of worldline 0 occurs at $\frac{4}{3}$ of its half-period $T_0/2$. Similarly, small circles Q_1 and Q_2 are the kinetic foci of worldlines 1 and 2, respectively. The heavy gray curve is the caustic, the locus of all kinetic foci of different worldlines of this family (originating at the origin with positive initial velocity). Squares indicate events at which the other worldlines recross worldline 0. This oscillator is discussed in detail in Sec. IX.

potential. The caustic is the line along which the kinetic foci lie for a particular family of worldlines (such as the family of worldlines that start from P with positive initial velocity in Figs. 4 and 5). A caustic is also an envelope to which all worldlines of a given family are tangent. The caustics in Figs. 5 and 6 are space-time caustics, envelopes for space-time trajectories (worldlines). Figure 7 shows a purely spatial caustic/envelope for a family of parabolic paths (orbits) in a linear gravitational potential. The word caustic is derived from optics[37] (along with the word *focus*). When a cup of coffee is illuminated at an angle, a bright curved line with a cusp appears on the surface of the coffee (see Fig. 8). Each point on this spatial optical caustic or ray envelope is the focus of light rays reflected from a small portion of the circular inner surface of the cup.

In Figs. 4–7 the caustic for a family of worldlines (or paths) represents a limit for those worldlines (or paths). No

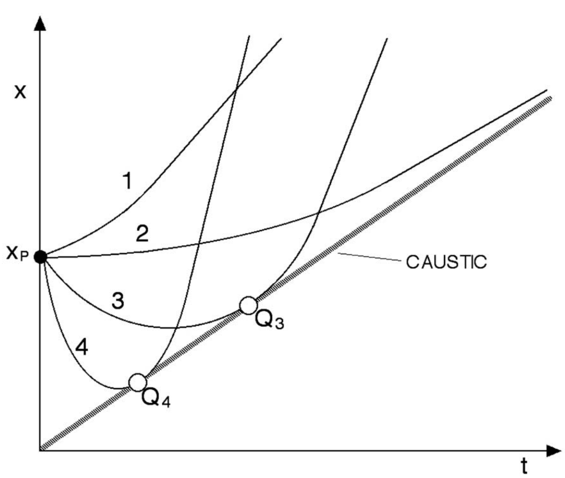

Fig. 6. Schematic space-time diagram for the repulsive inverse square potential $[U(x)=C/x^2]$, with a family of worldlines starting at $P(x_P,0)$ with various initial velocities. Intersections are events where two worldlines cross. The heavy gray straight line $x_Q=(\gamma/x_P)t_Q$, where $\gamma=(2C/m)^{1/2}$, is the caustic, the locus of kinetic foci Q (open circles) and envelope of the indirect worldlines. Worldline 2, with zero initial velocity, is asymptotic to the caustic, with kinetic focus Q at infinite space and time coordinates. The caustic divides space-time: each final event above the caustic can be reached by two worldlines of this family of worldlines, each final event on the caustic by one worldline of the family, and each final event below the caustic by no worldline of the family. This system is discussed in detail in Sec. X.

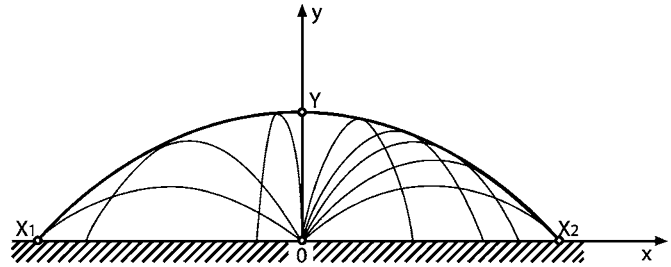

Fig. 7. For the Maupertuis action, the heavy line envelope (the "parabola of safety") is the locus of spatial kinetic foci (x_Q,y_Q), or spatial caustic, of the family of parabolic orbits of energy E originating from the origin $O(x_P,y_P)=(0,0)$ with various directions of the initial velocity $\mathbf{v}_0$. The potential is $U(x,y)=mgy$. The horizontal and vertical axes are x and y, respectively, and the caustic/envelope equation is $y=v_0^2/2g-gx^2/2v_0^2$, found by Johann Bernoulli in 1692. The caustic divides space. Each final point (x_R,y_R) inside the caustic can be reached from initial point (x_P,y_P) by two orbits of the family, each final point on the caustic by one orbit of the family, and each point outside the caustic by no orbit of the family. Y is the vertex (highest reachable point $y=v_0^2/2g$) of the caustic and X_1, X_2 denote the maximum range points $(x=\pm v_0^2/g)$. This system is discussed in detail in Appendix B. (Figure adapted from Ref. 43.)

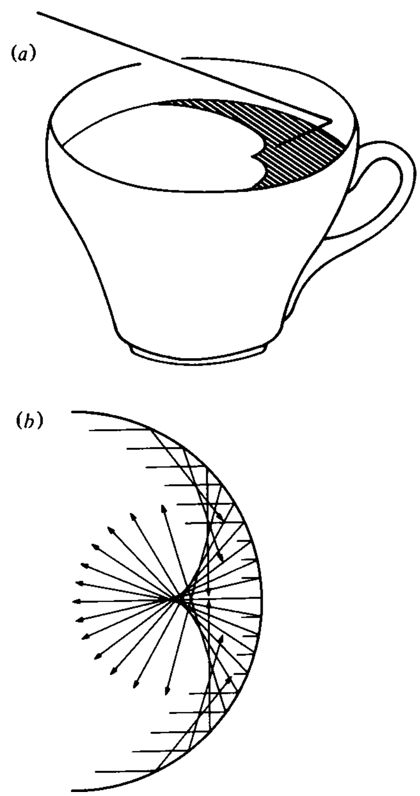

Fig. 8. The coffee-cup optical caustic. The caustic shape in panel (b) (a nephroid) was derived by Johann Bernoulli in 1692 (Ref. 44).

worldline of that family exists for final events outside the caustic. At least one worldline can pass through any event inside the caustic. Exactly one worldline can pass through an event on the caustic, and this event is the worldline's kinetic focus. This observation is consistent with the definition of the kinetic focus as an event at which two separate worldlines coalesce.

At the kinetic focus the worldline is tangent to the caustic. When two curves touch but do not cross and have equal slope at the point where they touch, the curves are said to *osculate* or *kiss*, which leads to a summary preview of the results of this paper:

> When a worldline terminates before it kisses the caustic, the action is minimum; when the worldline terminates after it kisses the caustic, the action is a saddle point.

One consequence is that when we use a computer to plot a family of worldlines (by whatever means), we can eyeball the envelope/caustic and locate the kinetic focus of each worldline visually.

This summary covers every case but one, because when no kinetic focus exists, there is no caustic so a worldline of any length has the minimum action. The one case not covered by this rule is the harmonic oscillator. For the harmonic oscillator and also for the sphere geodesics of Fig. 1, the caustic collapses to a single point at the kinetic focus. In this case there is no caustic curve; only one caustic point (a focal point) exists. A corresponding optical case is a concave reflecting parabolic surface of revolution illuminated with incoming light rays parallel to its axis; the optical caustic collapses into a single point at the focus (focal point) of the parabolic mirror. When the optical caustic reduces to a point for a lens or mirror system, the resulting images have minimum distortion (minimum aberration).

For the quartic and the piecewise-linear oscillators (and the harmonic oscillator) subsequent crossing points exist at which two worldlines can coalesce. We have defined the kinetic focus as the first of these, the one nearest the initial event P. The procedure for locating the subsequent kinetic foci is identical to that for locating the first one and is discussed briefly in the examples of Secs. VIII and IX. For 1D potentials subsequent kinetic foci[45] exist for the bound worldlines but not for the scattering worldlines, for example, those in Fig. 6. We shall not be concerned with subsequent kinetic foci; when we refer to the kinetic focus we mean the first one, as we have defined it. We shall show in what follows that for a few potentials [for example, $U(x)=C$ and $U(x)=Cx$] kinetic foci do not exist, because true worldlines beginning at a common initial event P do not cross again.

The definition of kinetic focus in terms of coalescing worldlines provides a practical way to find the kinetic focus. In Fig. 2 note the slopes of nearby worldlines 1 and 0 at the initial event P. The initial slope of curve 1 is only slightly different from that of worldline 0; as that difference approaches zero, the crossing event approaches the kinetic focus Q. The slope of a worldline at any point measures the velocity of the particle at that point. This coalescence of two worldliness as their initial velocities approach each other leads to a method for finding the kinetic focus: Launch an identical second particle from event P (simultaneously with the original launch) but with a slightly different initial velocity, that is, with a slightly different slope of the worldline. Worldline number 1 is also a true worldline. In the limiting case of a vanishing difference in the initial velocities at event P (vanishing angle between the initial slopes), the two worldlines will cross again, and the two particles collide at the kinetic focus event Q.

We can convert this practical (actually heuristic) idea into an analytical method, often easily applied when we have an analytic expression for the worldline. Let the original worldline be described by the function $x(t,v_0)$, where v_0 is the initial velocity. Then the second worldline is the same function with incrementally increased initial velocity $x(t,v_0+\Delta v_0)$. We form the expansion in Δv_0

$$x(t,v_0+\Delta v_0) = x(t,v_0) + \frac{\partial x}{\partial v_0}\Delta v_0 + \mathcal{O}(\Delta v_0^2). \tag{2}$$

At an intersection point R we have $x(t_R,v_0+\Delta v_0)=x(t_R,v_0)$. For intersection point R near Q we therefore have

$$\frac{\partial x}{\partial v_0}\Delta v_0 + \mathcal{O}(\Delta v_0^2) = 0, \tag{3}$$

which implies that for $R\rightarrow Q$ when $\Delta v_0 \rightarrow 0$, we have

$$\frac{\partial x}{\partial v_0} = 0. \tag{4}$$

Equation (4) is an analytic condition for the incrementally different worldline that crosses the original worldline at the kinetic focus, and hence it locates the kinetic focus Q. Sections VIII and IX discuss applications of this method.

III. WHY WORLDLINES CROSS

Section II defines the kinetic focus in terms of recrossing worldlines. The burden of this paper is to show that when a kinetic focus exists, the action along a worldline is a minimum if it terminates before the kinetic focus Q of the initial event P, whereas the action is a saddle point when the worldline terminates beyond the kinetic focus. In this section we consider only actual worldlines and describe qualitatively why two worldlines originating at the same initial event cross again at a later event. We also examine the special initial conditions at P under which the coalescing worldlines determine the position of the kinetic focus Q. The key parameter turns out to be the second spatial derivative $U'' \equiv \partial^2 U/\partial x^2$ of the potential energy function U. Sufficiently long worldlines can cross again only if they traverse a space in which $U''>0$. For simplicity we restrict our discussion here to time-independent potentials $U(x)$, but continue to use partial derivatives of U with respect to x to remind ourselves of this restriction. Features that can arise for time-dependent potentials $U(x,t)$ are discussed in Sec. XI.

Think of two identical particles that leave initial event P with different velocities and hence different slopes of their space-time worldlines, so that their worldlines diverge. The following description is valid whether the difference in the initial slopes is small or large. Figures 3–5 illustrate the following narrative. At every event on its worldline each particle experiences the force $F=-U'=-\partial U/\partial x$ evaluated at that location. For a short time after the two particles leave P they are at essentially the same displacement x, so they feel nearly the same force $-U'$. Hence the space-time curvature of the two worldlines (the acceleration) is nearly the same. Therefore the two worldlines will initially curve in concert while their initial relative velocity carries them apart; at the beginning their worldlines steadily diverge from one another. As time goes by, this divergence carries one particle, call it II, into a region in which the second spatial derivative U'' is (let us say) positive. Then particle II feels more force than particle I (but still in the same x direction as the force on it). As a result the worldline of II will head back toward particle I, leading to converging worldlines. As the two particles draw near again, they are once more in a region of almost equal U' and therefore experience nearly equal acceleration, so their relative velocity remains nearly constant until the worldlines intersect, at which event the two particles collide.

Note the crucial role played by the positive value of U'' in the relative space-time curvatures of worldlines I and II necessary for them to recross. Suppose instead that $U''<0$. Then as II moves away from I, it enters a region of smaller slope U' and hence smaller force than that on particle I. Hence the two worldlines will diverge even more than they did originally; the more they separate, the greater will be their rate of divergence. As long as both particles move in a region where $U''<0$, the two worldlines will never recross. (If $U''=0$, the two worldlines continue indefinitely to diverge at the initial rate.)

As a special case let the relative velocity of the two particles at launch be only incrementally different from one another (Fig. 2) for motion in potentials with $U''>0$, and let this difference of initial velocities approach zero. In this limit the particles will by definition collide at the kinetic focus Q of the initial event P. It may seem strange that an incremental relative velocity at P results in a recrossing at Q at a significant distance along the worldline from P. We might think that as this difference in slope increases from zero, the recrossing event would start at P and move smoothly away from it along worldline I, not "snap" all the way to Q. The source of the snap lies in the first and second spatial derivatives of U. When both particles start from the same initial event, the first derivative at essentially the same displacement leads to nearly the same force $-U'$ on both particles, so that any difference in the initial velocity, no matter how small, continues increasing the separation. It is only with greater relative displacement over time that the difference in these forces, quantified by $U''>0$, deflects the two worldlines back toward one another, leading to eventual recrossing. No alternative true worldline starting at P and with negligibly different initial velocity crosses the original worldline earlier than its kinetic focus (though widely divergent worldlines may cross sooner, as shown in Figs. 4 and 5). One consequence of this result is that a worldline terminating before its kinetic focus has minimum action, as shown analytically in Sec. VI.

In other words, potentials with $U''(x)>0$ are stabilizing, that is, they bring together neighboring trajectories that initially slightly diverge. Potentials with $U''(x)<0$ are destabilizing, that is, they push further apart neighboring trajectories that initially slightly diverge. It is thus not surprising that trajectory stability is closely related to the character of the stationary trajectory action (saddle point or minimum).[23,29–31]

Planetary orbits also exhibit crossing points distant from the location of a disturbance; an incremental change in the velocity at one point in the orbit leads to initial and continued divergence of the two orbits which, for certain potential functions, reverses to bring them together again at a distant point. This later crossing point is defined as the kinetic focus for the Maupertuis action W applied to spatial orbits (see Appendix B). This reconvergence has important consequences for the stability of orbits and the continuing survival of life on Earth as our planet experiences small nudges from the solar wind, meteor impacts, and shifting gravitational forces from other planets.

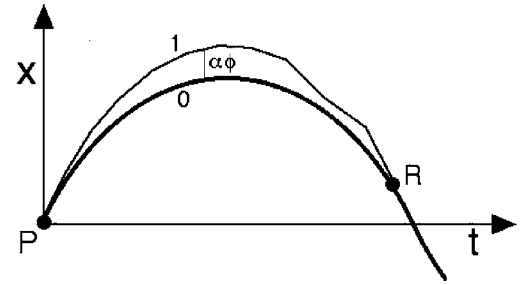

Fig. 9. An original true worldline, labeled 0, starts at initial event P. We draw an arbitrary adjacent curve, labeled 1, anchored at two ends on P and a later event R on the original worldline. The variational function $\alpha\phi$ is chosen to vanish at the two ends P and R.

IV. VARIATION OF ACTION FOR AN ADJACENT CURVE

> "…another feature in classical mechanics that seemed to be taboo in the discussion of the variational principle of classical mechanics by physicists: the second variation…" Martin Gutzwiller[46]

The action principle says that the worldline that a particle follows between two given fixed events P and R has a stationary action with respect to every possible alternative adjacent curve between those two events (Fig. 9). Thus the action principle employs not only actual worldlines, but also freely imagined and constructed curves adjacent to the original worldline, curves that are not necessarily worldlines themselves. In this paper the word *worldline* (or for emphasis *true worldline*) refers to a space-time trajectory that a particle might follow in a given potential. The word *curve* means an arbitrarily constructed trajectory that may or may not be a worldline. To study the action we need curves as well as worldlines. (In the literature the terms actual, true, and real trajectory are used synonymously with our term worldline; the terms virtual and trial trajectory are used for our use of the term curve.)

In this section we investigate the variational characteristics of the action S of a worldline in order to determine whether S is a local minimum or a saddle point with respect to arbitrary nearby curves between the same fixed events. In Fig. 9 a true worldline labeled 0 and described by the function $x_0(t)$ starts at initial event P. We construct a closely adjacent arbitrary curve, labeled 1 and described by the function $x(t)$, which starts at the same initial event P and terminates at a later event R on the original worldline. To compare the action along $P0R$ on the worldline $x_0(t)$ with the action along $P1R$ on the arbitrary adjacent curve $x(t)$, let

$$x(t)=x_0(t)+\alpha\phi(t), \tag{5}$$

and take the time derivative,

$$\dot{x}(t) = \dot{x}_0(t) + \alpha\dot{\phi}(t). \tag{6}$$

In Eqs. (5) and (6) α is a real numerical constant of small absolute value and $\phi(t)$ is an arbitrary real function of time that goes to zero at both P and R. The action principle says that action in Eq. (1) along $x_0(t)$ is stationary with respect to the action along $x(t)$ for small values of α. To simplify the analysis we restrict $\phi(t)$ to be a continuous function with at most a finite number of discontinuities of the first derivative; that is, all curves $x(t)$ are assumed to be at least piecewise smooth.[47] Within this limitation $x(t)$ represents all possible curves adjacent to $x_0(t)$, not only any actual nearby worldlines. From Eqs. (5) and (6) the Lagrangian $L(x,\dot{x},t)$ can be regarded as a function of α and hence expanded in powers of α for small α,

$$L = L_0 + \alpha\frac{dL}{d\alpha} + \frac{\alpha^2}{2}\frac{d^2L}{d\alpha^2} + \frac{\alpha^3}{6}\frac{d^3L}{d\alpha^3} + \cdots , \tag{7}$$

where $L_0 = L(x_0,\dot{x}_0,t)$ and the derivatives are evaluated at $\alpha = 0$, that is, along the original worldline $x_0(t)$. We apply Eqs. (5) and (6) to the first derivative in Eq. (7),

$$\frac{dL}{d\alpha} = \frac{\partial L}{\partial x}\frac{dx}{d\alpha} + \frac{\partial L}{\partial \dot{x}}\frac{d\dot{x}}{d\alpha} = \phi\frac{\partial L}{\partial x} + \dot{\phi}\frac{\partial L}{\partial \dot{x}}, \tag{8}$$

so that we can write in operator form

$$\frac{d}{d\alpha} = \phi\frac{\partial}{\partial x} + \dot{\phi}\frac{\partial}{\partial \dot{x}}, \tag{9}$$

and apply it twice in succession to yield

$$\frac{d^2L}{d\alpha^2} = \left(\phi\frac{\partial}{\partial x} + \dot{\phi}\frac{\partial}{\partial \dot{x}}\right)\left(\phi\frac{\partial L}{\partial x} + \dot{\phi}\frac{\partial L}{\partial \dot{x}}\right)$$
$$= \phi^2\frac{\partial^2 L}{\partial x^2} + 2\phi\dot{\phi}\frac{\partial^2 L}{\partial x\partial \dot{x}} + \dot{\phi}^2\frac{\partial^2 L}{\partial \dot{x}^2}. \tag{10}$$

We consider the most common case in which the Lagrangian L is equal to the difference between the kinetic and potential energy:

$$L = K - U = \tfrac{1}{2}m\dot{x}^2 - U(x,t), \tag{11}$$

where U may be time dependent. Then L has the partial derivatives

$$\frac{\partial L}{\partial x} = -\frac{\partial U}{\partial x}, \quad \frac{\partial L}{\partial \dot{x}} = m\dot{x}, \tag{12}$$

$$\frac{\partial^2 L}{\partial x^2} = -\frac{\partial^2 U}{\partial x^2}, \quad \frac{\partial^2 L}{\partial x\partial \dot{x}} = 0, \quad \frac{\partial^2 L}{\partial \dot{x}^2} = m, \tag{13}$$

$$\frac{\partial^3 L}{\partial x^3} = -\frac{\partial^3 U}{\partial x^3}, \quad \frac{\partial^3 L}{\partial \dot{x}^3} = 0. \tag{14}$$

Hence the second α derivative of L reduces to

$$\frac{d^2L}{d\alpha^2} = -\phi^2\frac{\partial^2 U}{\partial x^2} + m\dot{\phi}^2. \tag{15}$$

We apply Eq. (9) to Eq. (15) to obtain the third derivative of L for this case:

$$\frac{d^3L}{d\alpha^3} = -\phi^3\frac{\partial^3 U}{\partial x^3}. \tag{16}$$

The expansion (7) for L defined by Eq. (11) now becomes

$$L = L_0 + \alpha\left(-\phi\frac{\partial U}{\partial x} + m\dot{x}_0\dot{\phi}\right) + \frac{\alpha^2}{2}\left(-\phi^2\frac{\partial^2 U}{\partial x^2} + m\dot{\phi}^2\right)$$
$$+ \frac{\alpha^3}{6}\left(-\phi^3\frac{\partial^3 U}{\partial x^3}\right) + \cdots , \tag{17}$$

where the U derivatives are evaluated along $x_0(t)$. If we substitute Eq. (17) into the action integral (1), we obtain an expansion of S in powers of α:

$$S = S_0 + \delta S_0 + \delta^2 S_0 + \delta^3 S_0 + \cdots . \tag{18}$$

The standard result of the action principle[48] is that along a true worldline the action is stationary; that is, the term δS_0 in Eq. (18), called the first order variation (or simply the first variation), is zero for all variations around an actual worldline $x_0(t)$. [This condition is necessary and sufficient for the validity of Lagrange's equation of motion for $x_0(t)$.] We will need the higher-order variations $\delta^2 S_0$ and $\delta^3 S_0$ for an actual worldline. From Eqs. (17) and (18) they take the forms

$$\delta^2 S = \frac{\alpha^2}{2}\int_P^R \left(-\phi^2\frac{\partial^2 U}{\partial x^2} + m\dot{\phi}^2\right) dt, \tag{19}$$

and

$$\delta^3 S = -\frac{\alpha^3}{6}\int_P^R \phi^3\frac{\partial^3 U}{\partial x^3}\, dt, \tag{20}$$

where the derivatives of U are evaluated along $x_0(t)$. In Eqs. (19) and (20) and for most of what follows we use the compact standard notations $\delta^2 S = \delta^2 S_0$ and $\delta^3 S = \delta^3 S_0$ (as well as $\delta S = \delta S_0$).

In the remainder of this article we use Eqs. (19) and (20) to determine when the action is greater or less for a particular adjacent curve than for the original worldline, paying primary attention to the second order variation $\delta^2 S$. When the action is greater for all adjacent curves than for the worldline, $\delta^2 S > 0$ and the action along the worldline is a true minimum. The phrase "for all adjacent curves" means that the value of $\delta^2 S$ in Eq. (19) is positive for all possible variations $\alpha\phi(t)$. Equation (19) shows immediately that when $\partial^2 U/\partial x^2$ is zero or negative along the entire worldline, then the integrand is everywhere positive, leading to $\delta^2 S > 0$. Hence, if $\partial^2 U/\partial x^2 \leqslant 0$ everywhere, a worldline of any length has minimum action. This result was previewed in the qualitative argument of Sec. III.

The outcome is more complicated when $\partial^2 U/\partial x^2$ is neither zero nor negative everywhere along the worldline. We show in Sec. V that even in this case we have $\delta^2 S > 0$ for sufficiently short worldlines, so that the action is still a minimum. Later sections show that "sufficiently short" means a worldline terminated before the kinetic focus. For a worldline terminated beyond the kinetic focus, the action is smaller ($\delta^2 S < 0$) for at least one adjacent curve and greater ($\delta^2 S > 0$) for all other adjacent curves, a condition called a *saddle point* in the action. When the action is a saddle point, the value of $\delta^2 S$ in Eq. (19) is negative for at least one variation $\alpha\phi(t)$ and positive for all other variations $\alpha\phi(t)$.

When $\delta^2 S=0$ for one or more adjacent curves, as happens at a kinetic focus,[49] we need to examine the higher-order variations to see whether $S-S_0$ is positive, negative, or zero for these particular adjacent curves.

There is no worldline whose action is a true maximum, that is, for which $\delta^2 S<0$ or more generally for which $S-S_0<0$ for every adjacent curve. The following intuitive proof by contradiction was given briefly by Jacobi[22] and in more detail by Morin[50] for the Lagrangian $L=K-U$ with K positive as in Eq. (11). Consider an actual worldline for which it is claimed that S in Eq. (1) is a true maximum. Now modify this worldline by adding wiggles somewhere in the middle. These wiggles are to be of very high frequency and very small amplitude so that they increase the kinetic energy K compared to that along the original worldline with only a small change in the corresponding potential energy U. The Lagrangian $L=K-U$ for the region of wiggles is larger for the new curve and so is the overall time integral S. The new worldline has greater action than the original worldline, which we claimed to have maximum action. Therefore S cannot be a true maximum for any actual worldline.

V. WHEN THE ACTION IS A MINIMUM

We now employ the formalism of Sec. IV to analyze the action along a worldline that begins at initial event P and terminates at various final events R that lie along the worldline farther and farther from P. In this section we show that the action is a minimum for a sufficiently short worldline PR in all potentials, and we give a rough estimate of what sufficiently short means. (We showed in Sec. IV that the action is a minimum for all worldlines in some potentials.) In Sec. VI we show that sufficiently short means before the terminal event reaches the event R at which $\delta^2 S$ first vanishes for a particular, unique variation. We also will show that this R is Q, the kinetic focus of the worldline. In Sec. VII we show that conversely $\delta^2 S$ must vanish at the kinetic focus, and that when final event R is beyond Q, the action along PR is a saddle point.

In considering different locations of the terminal event R along the worldline, it is important to recognize that the set of incremental functions ϕ that go to zero at P and at R will be different for each terminal position R. Particular functions may have similar forms for all R; for example, assuming $t_P=0$ for simplicity, we might have $\phi=A(t/t_R)(1-t/t_R)$ or $\phi=A\sin(\pi t/t_R)$. However, ϕ need not be so restricted; the only restrictions are that ϕ go to zero at both P and R and be piecewise smooth. Statements about the value of $\delta^2 S$ for each different terminal event R are taken to be true for all possible ϕ for the particular R that satisfy these conditions.

For a sufficiently short worldline the action is always a minimum compared with that of adjacent curves, as mentioned in Sec. III. The formalism developed in Sec. IV confirms this result as follows. (Here we follow and elaborate Whittaker,[36] apart from a qualification given in the following.) We rewrite Eq. (19) using $U''(x)$:

$$\delta^2 S = -\frac{\alpha^2}{2}\int_P^R \phi^2 U''\,dt + \frac{\alpha^2}{2}\int_P^R m\dot{\phi}^2\,dt. \tag{21}$$

Because $\phi=0$ at P, we can write

$$\phi(t)=\int_P^t \dot{\phi}(t')dt' \le (t-t_P)\dot{\phi}_{\max} < T\dot{\phi}_{\max}, \tag{22}$$

where $T=t_R-t_P$, and $\dot{\phi}_{\max}$ is the maximum value between P and R. With this substitution the magnitude of the first integral in Eq. (21) for $\delta^2 S$ can be bounded:

$$\left|\int_P^R(-\phi^2 U'')dt\right| < T^3\dot{\phi}_{\max}^2|U''_{\max}|. \tag{23}$$

The second integral in Eq. (21) can be rewritten as

$$\int_P^R m\dot{\phi}^2\,dt = mT\langle\dot{\phi}^2\rangle, \tag{24}$$

where $\langle\dot{\phi}^2\rangle$ is the mean square of $\dot{\phi}$ over the time interval T. Compare Eqs. (23) and (24) and note that $\dot{\phi}_{\max}^2$ and $\langle\dot{\phi}^2\rangle$ have the same order of magnitude for all values of T; the reader can check the special case $\phi(t)=A\sin(n\pi(t-t_P)/T)$, where n is any nonzero integer. Here we assume for simplicity that $\phi(t)$ is smooth and is nonzero for all times t in the range T except possibly at discrete points; a similar argument can be given if this condition is violated. Also note that $|U''_{\max}|$ will not increase as R becomes closer to P. Thus if the range is sufficiently small, the most important term in $\delta^2 S$ is the one that contains $\dot{\phi}$. In this limit Eq. (21) reduces to

$$\delta^2 S \to \frac{1}{2}\alpha^2 m\int_P^R \dot{\phi}^2 dt$$
$$> 0 \text{ for sufficiently short worldlines}. \tag{25}$$

This quantity is positive because m, α^2, and $\dot{\phi}^2$ are all positive.[51] Therefore $\delta^2 S$ adds to the action, which demonstrates that the action is always a true minimum along a sufficiently short worldline. We shall use this result repeatedly in the remainder of this paper.

We can give a rough estimate[55] of the largest possible value of T such that $\delta^2 S>0$ for all variations (the exact value is given in Sec. VI). If we use Eqs. (23) and (24) in Eq. (21), we see that

$$\delta^2 S > \frac{\alpha^2}{2}[mT\langle\dot{\phi}^2\rangle - \dot{\phi}_{\max}^2|U''_{\max}|T^3], \tag{26}$$

so that $\delta^2 S>0$ if

$$mT\langle\dot{\phi}^2\rangle > \dot{\phi}_{\max}^2|U''_{\max}|T^3, \tag{27}$$

or

$$T < \frac{\dot{\phi}_{\rm rms}}{\dot{\phi}_{\max}}\frac{T_0}{2\pi}, \tag{28}$$

where $T_0/2\pi=(m/|U''_{\max}|)^{1/2}$ and $\dot{\phi}_{\rm rms}\equiv\langle\dot{\phi}^2\rangle^{1/2}$ is the root-mean-square value of $\dot{\phi}$.

For the harmonic oscillator T_0 is exactly equal to the period, and for a general oscillator T_0 is a time of the order of the period. Assume for simplicity that $\phi(t)$ is smooth and is nonvanishing over the whole range T with exceptions only at discrete points, for example, $\phi(t)=A\sin(n\pi(t-t_P)/T)$; a similar argument can be constructed if this condition is violated. The ratio $\dot{\phi}_{\rm rms}/\dot{\phi}_{\max}$ is then of order unity; for example, for the latter form of $\phi(t)$ we have $\dot{\phi}_{\rm rms}/\dot{\phi}_{\rm max} = 1/\sqrt{2}$. Thus for times T less than about $(1/\pi)T_0/2$ we have $\delta^2 S > 0$. For the various oscillators studied in Secs. VIII and IX we will see that the half-period is a better estimate of the time limit for which $\delta^2 S > 0$. For example, for the harmonic oscillator we show that for all times up to exactly one half-period, $\delta^2 S > 0$, so that the action is a minimum for times less than a half-period. We show in Sec. VI that for any system the precise time limit for which $\delta^2 S > 0$ for all variations is $t_Q - t_P$, the time to reach the kinetic focus.

Because the location of the initial event P is arbitrary, it follows that the action is a minimum on a short segment anywhere along a true worldline. It is not difficult to show that a necessary and sufficient condition for a curve to be a true worldline is that all short segments have minimum action. This result is valid irrespective of whether the action for the complete worldline is a minimum or a saddle point.

As discussed in Sec. IV, if $U''(x)$ is zero or negative at every x along the worldline, then $\delta^2 S$ in Eq. (19) is always positive, with the result that worldlines of every length have minimum action for particles in these potentials. For example, the gravitational potential energy functions U_1 for vertical motion near Earth's surface and U_2 for radial motion above the Earth (radius r_E) have the standard forms

$$U_1(x) = mgx \quad (0 \leqslant x \ll r_E), \tag{29}$$

$$U_2(x) = -(GMm)/(r_E + x) \quad (0 \leqslant x). \tag{30}$$

In both cases $U''(x)$ is zero or negative everywhere, so that Eq. (19) tells us that worldlines of any length have minimum action. (Further discussion of the nature of the stationary action for trajectories in these gravitational potentials appears in Appendix B; differences arise when we examine two-dimensional trajectories and when we compare the Hamilton action S with the Maupertuis action W.)

There are an infinite number of potential energy functions with the property $U''(x) \leqslant 0$ everywhere [another example is $U(x) = -Cx^2$], leading to minimum action along worldlines of any length. Nevertheless, the class of such functions is small compared with the class of potential energy functions for which $U''(x) > 0$ everywhere (such as the harmonic oscillator potential $U(x) = kx^2/2$), or for which $U''(x)$ is positive for some locations and zero or negative for other locations [such as the Lennard-Jones potential $U(x) = C_{12}/x^{12} - C_6/x^6$]. For this larger class of potentials the particular choice of worldline (length and location) determines whether the action has a minimum or whether it falls into the class for which the action is a saddle point.

VI. MINIMUM ACTION WHEN WORLDLINE TERMINATES BEFORE KINETIC FOCUS

The central result of this section is that the event along a worldline nearest to initial event P at which $\delta^2 S$ goes to zero is the kinetic focus Q. The key idea is that a unique true worldline connects P to Q, a worldline that coalesces with the original worldline and thus satisfies the definition of the kinetic focus in Sec. II. The primary outcome of this proof is that the action is a minimum if a worldline terminates before reaching its kinetic focus.

Our discussion is inspired by the classic work of Culverwell[21] (see Ref. 52 for a textbook discussion). Culverwell and Whittaker focus on the Maupertuis action W. We adapt their work to the Hamilton action S, extend and simplify it, and show that their argument is incomplete.

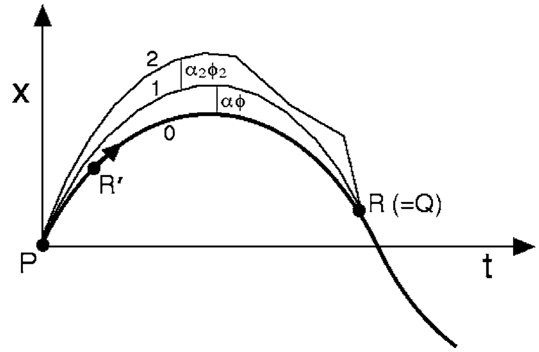

Fig. 10. Let R be the earliest event along the true worldline $x_0(t)$, labeled 0, such that $\delta^2 S = 0$ for worldline PR along $x_0(t)$. The unique variational function achieving $\delta^2 S = 0$ for worldline 0 corresponds to a varied curve labeled 1. We show that for this location of R, curve 1 is a true worldline and R is the kinetic focus Q. The arbitrary curve 2 is used to verify that curve 1 is a true worldline.

Consider a true worldline $x_0(t)$, labeled 0 in Fig. 10. As we have seen in Sec. V, the action S_0 along a sufficiently short segment PR' of worldline 0 is a minimum, which leads to $\delta^2 S > 0$ for all variations. We imagine terminal event R' located at later and later positions along the worldline until it reaches R, the event at which, by hypothesis, $\delta^2 S \to 0$ for the first time for some variation; that is, the integral in Eq. (19) defining $\delta^2 S$ vanishes for some choice of ϕ. We shall find that the earliest event at which $\delta^2 S$ vanishes is connected to the initial event P by a unique type of variation, namely a true worldline that coalesces with the original worldline in the limit for which their initial velocities at P coincide. Hence the earliest event at which $\delta^2 S$ vanishes satisfies the definition of the kinetic focus Q.

As we assumed, $\delta^2 S > 0$ for all R' up to $R' = R$, the first event for which $\delta^2 S = 0$ for a particular variation $\alpha\phi$. To prove that R is the kinetic focus Q, we need to consider small variations because Q involves a coalescing second worldline. In the typical case the integral in Eq. (20) defining $\delta^3 S$ does not vanish, but letting $\alpha \to 0$ will ensure $\delta^3 S$ (proportional to α^3) does not exceed $\delta^2 S$ (proportional to α^2) for R' approaching R.[56] The fact that $\delta^3 S$ does not exceed $\delta^2 S$ keeps $S - S_0 > 0$ (not just $\delta^2 S > 0$) for R' up to R. [The untypical case, in which the integral in Eq. (20) vanishes, is discussed in the following.] In the limit $R' \to R$, we let $\alpha \to 0$ so that the varied curve $x_1(t) = x_0(t) + \alpha\phi(t)$ coalesces with the true worldline $x_0(t)$, and $S - S_0 = 0$ at $R' = R$.

To satisfy the definition of the kinetic focus, we need to show that $x_1(t)$ is a true worldline just short of the limit $\alpha \to 0$ (just short of the limit $R' \to R$), not just at the limit $\alpha = 0$. We will prove this result by contradiction: assume curve $x_1(t)$ (curve 1 in Fig. 10) is not a true worldline. Consider the arbitrary comparison curve 2 in Fig. 10, which differs from 1 by the arbitrary variation $\alpha_2\phi_2$, with α_2 small. Assume (wrongly) that curve 1 is not a true worldline, so that the first-order variation δS in S between curves 1 and 2 is nonzero for arbitrary $\alpha_2\phi_2$, and the sign of α_2 can be chosen to make $S_2 < S_1$. But because $S_1 = S_0$ to second order, we must have $S_2 < S_0$, which is a contradiction; R is the earliest zero of $S - S_0$ for any small variation, so that small variations giving $S-S_0<0$ are impossible. To avoid the contradiction, curve 1 must be a true worldline. Thus we have proven that the unique variation $\alpha\phi$ that connects P and $R=Q$ when $\delta^2 S$ goes to zero for the first time corresponds to a true worldline.

This argument covers the typical case, where the integral in Eq. (20) defining $\delta^3 S$ does not vanish. The only common untypical case[49,57] is the harmonic oscillator, which we will discuss in Sec. VIII. The harmonic oscillator potential $U=kx^2/2$, for which $\partial^3 U/\partial x^3$ in Eq. (20) vanishes, so that $\delta^3 S=0$ identically. Similarly, the variation $\delta^4 S$ and higher variations all vanish because $\partial^4 U/\partial x^4$ and higher potential derivatives vanish. Thus for the harmonic oscillator $\delta^2 S=S-S_0$ and $S-S_0$ remains positive up to $R'=R$ for arbitrary α (not just small α). The preceding argument with $\alpha\to 0$ is valid also for the harmonic oscillator, so that the coalescing true worldline at R again shows that R is the kinetic focus Q. However, it is not necessary here to take the limit $\alpha\to 0$. Figure 3 for the harmonic oscillator shows that all true worldlines beginning at P intersect again where $\delta^2 S$ first vanishes, which is the kinetic focus Q. By varying the amplitude of the alternative true worldlines for the harmonic oscillator, we can find one that coalesces with the original worldline and thus satisfies the definition of the kinetic focus. The argument for other untypical cases[49] is similar to that for the typical case.

In summary we have shown that as terminal event R takes up positions along the worldline farther away from the initial point P, the special varied curve that leads to the earliest vanishing of $\delta^2 S$ is typically a unique[49,58] true worldline that can coalesce with the original worldline. This R satisfies the definition of the kinetic focus Q. Because the varied worldline for which $\delta^2 S=0$ for the first time is unique, it follows that all other curves PQ adjacent to the original worldline have $\delta^2 S>0$.

For bound motion in a time-independent potential, worldlines that can coalesce will typically cross more than once. In the literature all of these sequential limiting crossings are called kinetic foci. The above argument is valid only for the first such crossing, which we refer to as the kinetic focus.

As we have shown, a sufficient condition for the kinetic focus Q is that it is the earliest terminal event R for which $\delta^2 S=0$. In the following section we show the converse necessary condition: Given the definition of the kinetic focus Q as the first event at which a second true worldline can coalesce with PQ, the necessary consequence is that $\delta^2 S=0$ for worldline PQ for the variation leading to coalescence. Taken together, the arguments in these two sections prove the following theorem, which is the fundamental analytical result of our paper:

> A necessary and sufficient condition for Q to be a kinetic focus of worldline PQ is that Q is the earliest event on the worldline for which $\delta^2 S=0$.

This earliest vanishing of $\delta^2 S$ occurs for one special type of variation $\alpha\phi$ (which turns out to correspond to a true worldline), with ϕ unique (up to a factor) and typically $\alpha\to 0$; for all other variations $\delta^2 S$ remains positive at the kinetic focus. The Culverwell-Whittaker argument (more complicated than that just given) is incomplete in that it addresses only the sufficiency part of the theorem (the necessary part given in Sec. VII is new), and it overlooks the usual case where the limit $\alpha\to 0$ is necessary to locate the kinetic focus.

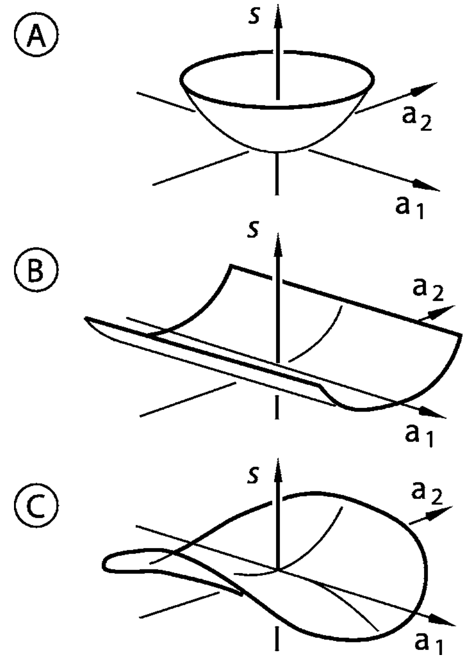

Fig. 11. Schematic illustration of the topological evolution of the minimum A→trough B→saddle C of the action S for two "directions" in function space as the final time t_R increases from $t_R<t_Q$ to $t_R=t_Q$ to $t_R>t_Q$, respectively, where t_Q is the kinetic focus time. (Figure adapted from Ref. 59.)

A simple topological picture of the action landscape in function space is emerging (see Fig. 11). For a short worldline PR, the action S is a minimum: the action increases in all directions away from the stationary point in function space [panel (a) in Fig. 11]. For longer PR, we may reach a kinetic focus $R=Q$ for which S is trough-shaped, that is, flat in one special direction and increasing in all other directions away from the stationary point [panel (b) in Fig. 11]. (The trough is completely flat for the harmonic oscillator, and flat to at least second order[49] for all other systems.) As we shall see in Sec. VII, as R moves beyond Q, the trough bends downward, placing the action at a saddle point; that is, S decreases in one direction in function space and increases in all other directions away from the stationary point [panel (c) in Fig. 11]. Although not discussed in this paper, the pattern may continue as R moves still further beyond the kinetic focus Q. If R reaches a second event Q_2 at which $\delta^2 S=0$ (called the second kinetic focus in the literature), a trough again develops for one special variational function ϕ (different from the first special ϕ); at Q_2 the action is flat in one direction in function space, decreases in one direction, and increases in all other directions. Beyond Q_2 the trough becomes a maximum, and we have a saddle point that is a maximum in two directions and a minimum in all others. Similar topological changes occur if we reach still later kinetic foci events Q_3, Q_4, and so forth, a result in agreement with Morse's theorem,[34] which states that the number of directions n in function space for which the action is a maximum at a saddle point for worldline PR is equal to the number n of kinetic foci between the end events P and R of the worldline.

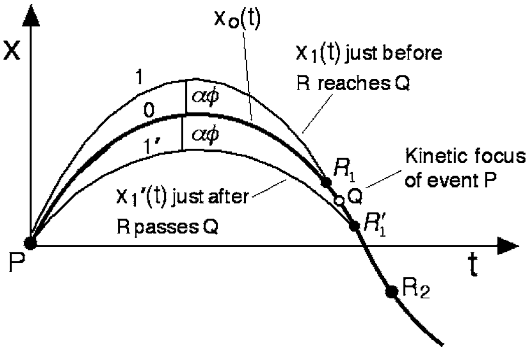

Fig. 12. By definition the kinetic focus Q of the initial event P is the first event at which two adjacent true worldlines $x_0(t)$ and $x_1(t)$ coalesce. We show that $\delta^2 S=0$ at the kinetic focus Q for this variational function ϕ in the limit $R_1 \to Q$, and that the action is a saddle point when the terminal event R_2 lies anywhere on the worldline beyond the kinetic focus Q. (The lower $\alpha\phi$ has the opposite sign from the upper $\alpha\phi$, and the upper and lower functions ϕ are slightly different due to the end-events R_1 and R_1' being slightly different.)

VII. SADDLE POINT IN ACTION WHEN WORLDLINE TERMINATES BEYOND KINETIC FOCUS

In Sec. VI we showed that a sufficient condition for the earliest event to be the kinetic focus Q is that the earliest event at which $\delta^2 S=0$ is connected to the initial event P by a unique true coalescing worldline. All alternative curves PQ lead to $\delta^2 S>0$. In this section we demonstrate the corresponding necessary condition, namely, given an alternative true worldline between P and R that coalesces with the original worldline as $R \to Q$ and therefore defines Q as the kinetic focus, we have $\delta^2 S=0$ for this worldline. By using an extension of this analysis, we also show that when R lies beyond the kinetic focus Q the action of worldline PQR is a saddle point.

The essence of the proof of the first statement is outlined in the following heuristic argument by Routh.[29] Consider two intersecting true worldlines $P \to R$ connecting P to R. Assume R is close to the kinetic focus Q so that the two worldlines differ infinitesimally, as required in the definition of Q. Let the action along the two worldlines be S and $S+\delta S$, respectively. Because both are true worldlines, the first-order variation of each is equal to zero, $\delta S=0$ and $\delta(S+\delta S)=0$. The difference of these two relations gives $\delta^2 S=0$ for R near Q and hence $\delta^2 S=0$ for $R=Q$. In the following we show that this argument is correct in the sense that $\delta^2 S$ vanishes not only at $R=Q$, but also vanishes to $\mathcal{O}(\alpha^2)$ for R near Q, differing from zero by $\mathcal{O}(\alpha^3)$ for R near Q.

To make Routh's argument rigorous, consider the two alternative true worldlines $x_0(t)$ and $x_1(t)$ in Fig. 12 that connect the initial event P to the terminal event R, where R is close to the kinetic focus Q of $x_0(t)$. When R reaches Q, the two worldlines coalesce according to the definition of the kinetic focus Q. For definiteness, take $x_1(t)$ to be the top worldline in Fig. 12, which is closely adjacent to the true worldline $x_0(t)$. Hence, it is a member of the set of adjacent curves used for the variation in Sec. IV, and therefore we can employ the formalism of that section. Conversely, we can regard $x_0(t)$ as a varied curve of $x_1(t)$, because it is closely adjacent to the other true worldline $x_1(t)$. Equations similar to Eqs. (5) and (6) are

$$x_1(t)=x_0(t)+\alpha\phi(t), \quad x_0(t)=x_1(t)-\alpha\phi(t), \tag{31}$$

and

$$\dot{x}_1(t)=\dot{x}_0(t)+\alpha\dot{\phi}(t), \quad \dot{x}_0(t)=\dot{x}_1(t)-\alpha\dot{\phi}(t). \tag{32}$$

We have

$$S_1=S_0+\delta S_0+\delta^2 S_0+\delta^3 S_0+\cdots. \tag{33}$$

In this case both $x_0(t)$ and $x_1(t)$ are true worldlines, so that we can also write the inverse expression

$$S_0=S_1+\delta S_1+\delta^2 S_1+\delta^3 S_1+\cdots. \tag{34}$$

We subtract Eq. (34) from Eq. (33) and use $\delta S_0=0$ and $\delta S_1=0$, because both $x_0(t)$ and $x_1(t)$ are true worldlines. The result is

$$2(S_1-S_0)=(\delta^2 S_0-\delta^2 S_1)+(\delta^3 S_0-\delta^3 S_1)+\cdots. \tag{35}$$

We then find expressions for $\delta^2 S$ from Eq. (19):

$$\delta^2 S_0=\frac{\alpha^2}{2}\int_P^R(-\phi^2 U''(x_0)+m\dot{\phi}^2)dt, \tag{36a}$$

$$\delta^2 S_1=\frac{\alpha^2}{2}\int_P^R(-\phi^2 U''(x_1)+m\dot{\phi}^2)dt. \tag{36b}$$

The first parenthesis on the right side of Eq. (35) has the form

$$\delta^2 S_0-\delta^2 S_1=\frac{\alpha^2}{2}\int_P^R dt[U''(x_1)-U''(x_0)]\phi^2, \tag{37}$$

where terms in $\dot{\phi}$ have cancelled. We expand $U''(x_1)$ to first order in α:

$$U''(x_1)\approx U''(x_0)+U'''(x_0)(x_1-x_0)=U''(x_0)+U'''(x_0)\alpha\phi. \tag{38}$$

If Eq. (38) is substituted into Eq. (37), the resulting integral contributes a further factor of α, yielding a result of $\mathcal{O}(\alpha^3)$:

$$\delta^2 S_0-\delta^2 S_1\approx\frac{\alpha^3}{2}\int_P^R dt U'''(x_0)\phi^3. \tag{39}$$

The fact that the right-hand side of Eq. (39) is proportional to α^3 means that in Eq. (35) we cannot neglect terms in $\delta^3 S$ that are also proportional to α^3. (Later terms are proportional to α^4 or higher.) From Eq. (20) and the signs of α in Eq. (31), we have

$$\delta^3 S_0=-\frac{\alpha^3}{6}\int_P^R\phi^3 U'''(x_0)dt, \tag{40a}$$

$$\delta^3 S_1=+\frac{\alpha^3}{6}\int_P^R\phi^3 U'''(x_1)dt, \tag{40b}$$

and hence the second term on the right-hand side of Eq. (35) becomes

$$\delta^3 S_0 - \delta^3 S_1 = -\frac{\alpha^3}{6}\int_P^R [U'''(x_1) + U'''(x_0)]\phi^3\, dt$$
$$\approx -\frac{\alpha^3}{3}\int_P^R U'''(x_0)\phi^3\, dt, \tag{41}$$

where we have set $U'''(x_1) \approx U'''(x_0)$, which is correct to $\mathcal{O}(\alpha^3)$ in Eq. (41).

The substitution of Eqs. (39) and (41) in Eq. (35) gives

$$S_1 - S_0 \approx \frac{\alpha^3}{12}\int_P^R U'''(x_0)\phi^3\, dt \quad \text{for } R \text{ near } Q \text{ to } \mathcal{O}(\alpha^3). \tag{42}$$

Equation (42) makes precise the earlier heuristic argument due to Routh. We see that $S_1 - S_0$ is $\mathcal{O}(\alpha^3)$ for R near Q and therefore vanishes for $R \to Q$. If we compare Eqs. (42) and (33) (with $\delta S_0 = 0$) and note that the coefficient $\alpha^3/12$ in Eq. (42) differs from the coefficient $-\alpha^3/6$ in Eq. (40a) for $\delta^3 S_0$, we see that not only is $(\delta^2 S_0 - \delta^2 S_1) \propto \alpha^3$, but $\delta^2 S_0$ is also $\mathcal{O}(\alpha^3)$. In Ref. 60 we show that $\delta^2 S_0$ is $\mathcal{O}(\alpha^3)$ more directly but less elegantly (using equations of motion rather than purely variational arguments). The result is that $S_1 - S_0$ is $\mathcal{O}(\alpha^3)$ and hence vanishes for $R \to Q$, which yields the desired necessary condition: Given the definition of Q, which involves two true worldlines coalescing as $R \to Q$, we have $\delta^2 S = 0$ for worldline PQ for the special variation leading to the coalescence that defines the kinetic focus Q.

Next we extend our results to show that when R lies immediately beyond the kinetic focus Q, the action of worldline PQR is a saddle point. For a saddle point to occur the sign of $S_1 - S_0$ in Eq. (42) must change from positive to negative as R passes through the kinetic focus Q. To interpret the sign in Eq. (42) consider the worldline $x_1(t)$, the uppermost worldline in Fig. 12, which crosses the original worldline $x_0(t)$ at R_1 slightly earlier than the kinetic focus Q of worldline $x_0(t)$. We have seen that $\delta^2 S_0 > 0$ for all variations for short PR along $x_0(t)$ and that $\delta^2 S_0$ does not vanish until R reaches Q. Thus $\delta^2 S_0$ and $S_1 - S_0$ are positive for R_1 slightly earlier than Q. Figure 12 shows and Eq. (42) makes quantitative that as R takes positions from R_1 to Q and then positions at Q and beyond Q to R_1', the variational function $\alpha\phi$ vanishes, and then changes sign to become negative. From Eq. (42) we see that when the variation of $x_0(t)$ is the adjacent true worldline $x_{1'}(t)$, we have $S_{1'} - S_0 < 0$ for R_1' slightly later than Q, so that $S(P1'R_1') < S(P0R_1')$. However, there are other variations in $x_0(t)$ that generate $S_{1'} - S_0 > 0$, such as displacing or adding wiggles to a short segment (recall the discussion at the end of Sec. IV). Thus when R_1' is just beyond Q we can increase or decrease the action compared to S_0, depending on which variation we choose, which is the definition of a saddle point: for worldline $x_0(t)$ or $P0R_1'$, the action S_0 is a saddle point for R_1' just beyond Q.

We now demonstrate that worldline $x_0(t)$ has a saddle point in the action not only for R_1' just beyond Q, but also for all terminal events on the worldline beyond Q, no matter how far beyond Q they lie. Imagine a point R_2 further along $x_0(t)$ from R_1' by an arbitrary amount, so that the true worldline in Fig. 12 is now $P0R_1'R_2$. Use the bottom worldline $P1'R_1'$ in Fig. 12 to construct the comparison curve $P1'R_1'R_2$, which is not a true worldline due to the kink at R_1'. Because $P0R_1'R_2$ and $P1'R_1'R_2$ have the segment $R_1'R_2$ in common, and because $S(P1'R_1') < S(P0R_1')$ as shown previously, we have $S(P1'R_1'R_2) < S(P0R_1'R_2)$. Hence we have found a variation $P1'R_1'R_2$ with a smaller action than the original worldline $P0R_1'R_2$. But we know it is easy to find other variations giving a larger action (just add wiggles somewhere). Thus $P0R_1'R_2$ has a saddle point[61] in the action for all R_2 later than the kinetic focus Q.

These demonstrations are valid for the typical case ($\delta^3 S \neq 0$ for the special variation ϕ). The most common atypical case is the harmonic oscillator ($\delta^k S = 0$ for all k for the special variation ϕ); proofs of the previous points are given explicitly in Sec. VIII for this system. Other atypical cases,[49] for example, when $\delta^3 S = 0$ and $\delta^4 S \neq 0$ for the special variation ϕ, so that $S_1 - S_0 \sim \mathcal{O}(\alpha^4)$ for R near Q in place of Eq. (42), require an extension of the previous argument.

We have established, for the typical case and the most common atypical case, the two central results of our paper: (1) The worldline PR has the minimum action if the terminal event R is earlier than the kinetic focus event Q of the initial event P. (2) The worldline PQR has a saddle point in the action when the terminal event R lies beyond the kinetic focus event Q of initial event P.[62] These results are correct wherever on the worldline we freely choose to place the initial event P with respect to which the later kinetic focus event Q of P is established. Note that a true maximum of the action for worldline PR is never found, in agreement with the result of the intuitive argument at the end of Sec. IV. In the literature[26] these results are often expressed as follows: For Lagrangians of type (11) having $\partial^2 L/\partial \dot{x}^2 > 0$, Jacobi's necessary and sufficient condition for a weak[47] local minimum of the stationary action is that a kinetic focus does not occur between the end events P and R.

VIII. HARMONIC OSCILLATOR

For the harmonic oscillator it is particularly easy to use Fourier series to compare the action along a worldline with the action along every (at least every piecewise-smooth) curve alternative to the worldline.[65] The harmonic oscillator Lagrangian has the form

$$L = K - U = \tfrac{1}{2}m\dot{x}^2 - \tfrac{1}{2}kx^2, \tag{43}$$

so that Eq. (19) becomes

$$\delta^2 S = \frac{\alpha^2}{2}\int_P^R (-k\phi^2 + m\dot{\phi}^2)dt = \frac{\alpha^2 m}{2}\int_P^R (\dot{\phi}^2 - \omega_0^2\phi^2)dt, \tag{44}$$

where

$$\omega_0 = \left(\frac{k}{m}\right)^{1/2} = \frac{2\pi}{T_0}, \tag{45}$$

and T_0 is the natural period. All third and higher partial derivatives of L with respect to x and $\dot{x}$ are zero, so that there are only second-order variations in S due to ϕ [see Eq. (20)]. Therefore we have

$$S - S_0 = \delta^2 S \quad \text{for the harmonic oscillator}. \tag{46}$$

We set $P = (x_P, 0)$ and $R = (x_R, t_R)$ and express the variational function $\phi(t)$ using a Fourier series:

$$\phi(t) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{n\omega t}{2}\right), \tag{47}$$

where $\omega = 2\pi/t_R$. The function $\phi(t)$ automatically goes to zero at the initial and final events P and R, respectively. The constants a_n can be chosen arbitrarily, corresponding to our free choice of $\phi(t)$. Because of the completeness of the Fourier series, the a_n can represent every possible piecewise-smooth trial curve alternative to any true worldline. We substitute Eq. (47) into Eq. (44). The squares of $\dot{\phi}$ and ϕ lead to double summations:

$$\delta^2 S = \frac{\alpha^2 m}{2} \sum_{n=1}^{\infty} \sum_{n'=1}^{\infty} a_n a_{n'} \int_0^{t_R} \left[\frac{nn'\omega^2}{4} \cos\left(\frac{n\omega t}{2}\right) \cos\left(\frac{n'\omega t}{2}\right) - \omega_0^2 \sin\left(\frac{n\omega t}{2}\right) \sin\left(\frac{n'\omega t}{2}\right) \right] dt. \tag{48}$$

As t goes from zero to t_R, the arguments of the harmonic functions go from zero to $n\pi$ or $n'\pi$, both of which represent an integer number of half-cycles. Because the different harmonics are orthogonal, terms with $n' \neq n$ integrate to zero for any number of complete half-cycles. Hence Eq. (48) simplifies to

$$\delta^2 S = \frac{\alpha^2 m}{2} \sum_{n=1}^{\infty} a_n^2 \int_0^{t_R} \left[\left(\frac{n\omega}{2}\right)^2 \cos^2\left(\frac{n\omega t}{2}\right) - \omega_0^2 \sin^2\left(\frac{n\omega t}{2}\right) \right] dt. \tag{49}$$

The integrals from $t=0$ to $t=t_R$ in Eq. (49) are over n half-cycles. For any integer number of half-cycles the average cosine squared and the average sine squared are both equal to $1/2$. Therefore we have

$$\delta^2 S = \frac{\alpha^2 m t_R}{4} \sum_{n=1}^{\infty} a_n^2 \left[\left(\frac{n\omega}{2}\right)^2 - \omega_0^2 \right], \tag{50}$$

where ω and ω_0 are defined in Eq. (45) and below Eq. (47). The harmonic oscillator is atypical in that the period does not depend on amplitude, so that all worldlines that start at the same initial event recross at the same kinetic focus, as shown in Fig. 3.

We now use Eq. (50) to give examples and verify, for the atypical case of the harmonic oscillator, the results derived for the typical case in earlier sections of this paper.

Case I, $t_R < T_0/2$, the final time is less than one half-period ($\omega/2 = \pi/t_R > \omega_0 = 2\pi/T_0$). In this case $\delta^2 S > 0$ for all choices of a_n and hence for every adjacent curve. For final times less than $T_0/2$ (that is, for worldlines that terminate before they reach the kinetic focus of the initial event) the action along the worldline is a minimum with respect to every adjacent curve.

Case II, $t_R = T_0/2$, the final time is equal to one half-period ($\omega/2 = \pi/t_R = \omega_0 = 2\pi/T_0$). In this case the final event is the kinetic focus of the initial event, where [see Eq. (46)] $S - S_0 = \delta^2 S$ goes to zero for the first time for a special type of variation. Choose $a_1 = A$ and $a_n = 0$ for $n \neq 1$, giving $\delta^2 S = 0$ in Eq. (50); examples are trial curves 1 and 2 in Fig. 3, which here are also true worldlines. If we take the limit $\alpha \to 0$ (or the zero displacement limit $A \to 0$), the true worldline 1 coalesces with true worldline 0. (All half-period harmonic oscillator worldlines starting from P pass through the same kinetic focus, Fig. 3). For all other choices of the a_n (for example, $a_2 = A$, all other $a_n = 0$), we have $S - S_0 = \delta^2 S > 0$ in Eq. (50), a result established for the typical case in Sec. VI.

Think of each choice of all the coefficients $a_1, a_2, a_3, \ldots$ as a point in function space. Then we see that at the kinetic focus the action is a minimum for all "directions" in function space except for the direction with $a_1 \neq 0$, all other $a_n = 0$, for which $S - S_0 = \delta^2 S = 0$. If we think of each a_n as plotted along a different direction in function space, we can picture this exceptional stationary S case at the kinetic focus as an action trough in function space; that is, flat in one special direction, increasing in all others. (In the typical case[49] the trough is flat only to second order.) In contrast, for case I where $t_R < T_0/2$, $\delta^2 S > 0$ along every direction in function space.

Case III, $t_R > T_0/2$, the final time is greater than one half-period ($\omega/2 = \pi/t_R < \omega_0 = 2\pi/T_0$). In this case we can choose a_n to find adjacent curves with action either greater or less than the action along the original worldline. For $a_1 = A$ and all other $a_n = 0$, $\delta^2 S < 0$, so the action for the worldline is greater than that for the adjacent curve. In contrast, if we choose $a_n = A$ for any term $n = N$ for which $(N\omega/2)^2 > \omega_0^2$ and all other $a_n = 0$, then $\delta^2 S > 0$ and the action for the worldline is smaller than for the curve. In brief, for a final time greater than half a period of the harmonic oscillator, the action for the worldline is neither a true maximum nor a true minimum; it is a saddle point. The corresponding result was shown for the typical case in Sec. VII. Figure 11 shows schematically the evolution of S from case I to case II to case III as t_R increases.

All three cases apply to the second variation $\delta^2 S$ for all harmonic oscillator true worldlines $x_0(t)$, for example, those that do not start from $(x_P, t_P) = (0,0)$, such as $x_0(t) = A_0 \sin(\omega_0 t + \theta_0)$, and include the no-excursion or equilibrium worldline $x_0(t) = 0$. In all cases, for actual worldlines S is a minimum (or a trough as in case II) or a saddle point, never a true maximum, in agreement with the general theory.

For the harmonic oscillator all worldlines starting at initial event $P = (0,0)$ as in Fig. 3, for example, converge next at event $Q = (0, T_0/2)$, which is therefore the kinetic focus of P. We verify this result analytically using the general method of Sec. II. For the harmonic oscillator with $P = (0,0)$, we express the amplitude of displacement in terms of the initial velocity v_0:

$$x = \frac{v_0}{\omega_0} \sin \omega_0 t, \tag{51}$$

where ω_0 is independent of v_0 for the harmonic oscillator. According to Eq. (4), the time t_Q of the kinetic focus is found by taking the partial derivative of Eq. (51) with respect to v_0 and setting the result equal to zero:

$$\frac{\partial x}{\partial v_0} = \frac{1}{\omega_0} \sin \omega_0 t_Q = 0. \tag{52}$$

Therefore the kinetic focus of the initial event $P = (0,0)$ occurs at the time when $\omega_0 t_Q = \pi$ or $t_Q = T_0/2$. (What the literature[27,32] calls the "later kinetic foci" occur for $\omega_0 t = 2\pi, 3\pi, \ldots$, but we limit the term kinetic focus to the first of these.) A similar calculation with $P \neq (0,0)$ gives the same result, that is, $t_Q - t_P = T_0/2$. The fact that $t_Q - t_P$ is independent of the initial event P holds only for the harmonic oscillator and does not hold for nonlinear oscillators. In Appendix B we discuss the location of the kinetic focus for the trajectories of two-dimensional harmonic oscillators.

In summary, four characteristics of the harmonic oscillator worldlines are exceptional; these characteristics are not true for worldlines in most potential energy functions. For an arbitrary initial event P, (1) all worldlines from P pass through the same point (the kinetic focus); (2) the time of the kinetic focus Q of P is half a period $T_0/2$ later; (3) the time interval is $T_0/2$ between all successive kinetic foci; and (4) when the final event R is not a kinetic focus, only one true worldline connects it to P. Underlying these four characteristics is the exceptional property of the harmonic oscillator: the frequency is independent of amplitude, which reflects the linearity of the system.

IX. NONLINEAR OSCILLATORS

We could analyze the action S for the worldlines of an arbitrary oscillator with potential $U(x)$ by methods similar to those used for the harmonic oscillator in Sec. VIII. The second order variation $\delta^2 S$ from the value S_0 for a worldline $x_0(t)$ is given by Eq. (19):

$$\delta^2 S = \frac{\alpha^2}{2}\int_P^R [m\dot{\phi}(t)^2 - U''(x_0(t))\phi(t)^2]dt, \tag{53}$$

where $\alpha\phi(t)$ is an arbitrary variation from $x_0(t)$ that vanishes at the end-events P and R. The analysis is complicated[66] for arbitrary $U(x)$; it is more instructive to consider instead the piecewise-linear oscillator with the potential $U(x)=C|x|$ and the quartic oscillator with a U-shaped potential $U(x)=Cx^4$. Figure 5 illustrates that the piecewise linear oscillator is representative of the class whose period increases with increasing amplitude. Figure 4 illustrates that the period of the quartic oscillator decreases with increasing amplitude.

A. Piecewise-linear oscillator

As an example of a piecewise-linear oscillator, consider a star that oscillates back and forth through the plane of the galaxy and perpendicular to it.[67] We approximate the galaxy as a (freely penetrable) sheet of zero thickness and uniform mass density and express the gravitational potential energy of this configuration as $U(x)=mg|x|$, with the value of $g=C/m$ calculated from the mass density per unit area of the galaxy surface. On the Earth a piecewise-linear potential of the form[68,69] $U(x)=C|x|$ with $C\propto g$ models the horizontal component of the oscillations of a particle sliding without friction between two equal-angle inclined planes that meet at the origin. The same form of the potential roughly models the interaction between two quarks with x their separation. The classical, semiclassical, and quantum motion of three quarks on a line interacting with mutual piecewise-linear potentials have been studied by variational methods (see Ref. 73 and references therein).

From these possible examples of piecewise-linear oscillators, we choose to analyze the star oscillating back and forth perpendicular to the galaxy. We know that the solution to the star's oscillation on either side of the galaxy from elementary analysis of the vertical motion near the surface of the Earth. With the initial event chosen as $P(x_P,t_P)=(0,0)$ and $v_0>0$ as in Fig. 5, the first half-cycle follows the parabolic worldline

$$x = v_0 t - \tfrac{1}{2}gt^2 \quad (t \leqslant T_0/2), \tag{54}$$

where the numerical value of g for galactic oscillation derives from the surface mass density of the galaxy sheet. The time $T_0/2$ of the first half cycle is the time to return to $x=0$:

$$\frac{T_0}{2} = \frac{2v_0}{g}. \tag{55}$$

After crossing into negative values of x, the worldline equation has a form similar to Eq. (54):

$$x = -v_0\left(t - \frac{2v_0}{g}\right) + \frac{1}{2}g\left(t - \frac{2v_0}{g}\right)^2 \quad (T_0/2 \leqslant t \leqslant T_0). \tag{56}$$

If we were considering motion in the region of positive x (or negative x) alone, there would be no kinetic focus because U'' is zero in either region, leading to a positive second order variation in the action derived from Eq. (53) as discussed in Sec. V. It is the infinite second derivative U'' at the origin of the potential $U(x)=C|x|$ that creates the kinetic focus for the piecewise-linear oscillator. The second derivative $U''(x)$ of this potential is

$$U''(x) = 2C\delta(x), \tag{57}$$

where $\delta(x)$ is the Dirac delta function.

Now consider the second order variation $\delta^2 S$ for a worldline with $(x_P,t_P)=(0,0)$ and the time t_R of the terminal point R in the range $T_0/2 \leqslant t_R \leqslant T_0$. As our variational function we choose

$$\phi(t) = a_0 \sin\left(\frac{n\omega t}{2}\right), \tag{58}$$

where $\omega=2\pi/t_R$ and a_0 is arbitrary. The variational function $\phi(t)$ vanishes at the end-points P and R, as it should, and is a slowly oscillating variation for $n=1$ and a rapidly oscillating variation for n large. We substitute Eqs. (58) and (57) in Eq. (53). The integration over $\delta(x)$ is most easily done by changing the integration variable from t to x using $dt=dx/\dot{x}$. The other integration follows the same pattern as in Sec. VIII. We find

$$\delta^2 S = \frac{1}{4}\alpha^2 m a_0^2 t_R\left[\left(\frac{n\omega}{2}\right)^2 - \frac{4}{\pi^2}\frac{T_0}{t_R}\omega_0^2 \sin^2\left(\frac{n\omega T_0}{4}\right)\right], \tag{59}$$

where $\omega_0=2\pi/T_0$ and T_0 is given by Eq. (55). For sufficiently short t_R (that is, sufficiently large $\omega=2\pi/t_R$), the positive $(n\omega/2)^2$ term in Eq. (59) will dominate for any n, so that $\delta^2 S>0$. The action S is therefore a minimum for a worldline with sufficiently short t_R. For large t_R (ω small), the $(n\omega/2)^2$ term will again dominate for a variation with sufficiently large n. In this case we again have $\delta^2 S>0$. But for the $n=1$ variation the negative term in Eq. (59) dominates for sufficiently small ω (t_R sufficiently large). In this case we have $\delta^2 S<0$. Thus for sufficiently large t_R the action is a saddle point. These results are consistent with the general theorems derived earlier.

The dividing line between small and large t_R in the preceding paragraph is the time of the kinetic focus. To find the time t_Q at the kinetic focus Q of the initial event P (see Fig. 5), we use the method developed in Sec. II. Because the oscillator frequency decreases with increasing amplitude, we know that the kinetic focus time t_Q will exceed $T_0/2$. We apply condition Eq. (4) to Eq. (56), giving

$$\frac{\partial x}{\partial v_0} = 0 = -\left(t_Q - \frac{2v_0}{g}\right) + \frac{2v_0}{g} + g\left(t_Q - \frac{2v_0}{g}\right)\left(-\frac{2}{g}\right), \tag{60}$$

which yields

$$t_Q = \frac{8v_0}{3g} = \frac{2T_0}{3} = \frac{4}{3}\left(\frac{T_0}{2}\right). \tag{61}$$

Thus the kinetic focus is later than the half-period by a factor of $\frac{4}{3}$, as shown in Fig. 5.

The spatial location x_Q of the kinetic focus Q of a particular worldline is found from t_Q using Eq. (56):

$$x_Q = -v_0\left(t_Q - \frac{2v_0}{g}\right) + \frac{1}{2}g\left(t_Q - \frac{2v_0}{g}\right)^2. \tag{62}$$

The locus of the various kinetic foci Q of the family of worldlines in Fig. 5 is the *caustic* or *envelope* and can be found by relating v_0 to t_Q. From Eq. (61) we have

$$\frac{2v_0}{g} = \frac{3}{4}t_Q. \tag{63}$$

If we substitute Eq. (63) into Eq. (62), we obtain the relation for the caustic of the family of piecewise-linear oscillator worldlines with $P=(0,0)$ and $v_0>0$:

$$x_Q = -\tfrac{1}{16}gt_Q^2. \tag{64}$$

This caustic is a parabola, shown as the heavy gray line in Fig. 5. It divides space-time. Each final event (x_R, t_R) above the caustic can be reached by one or more worldlines of this family of worldlines; each final event on the caustic can be reached by just one worldline of the family; and each final event below the caustic can be reached by no worldline of the family. For the harmonic oscillator all kinetic foci for a given initial event P fall at the same point (a focal point[37]), the limiting case of a caustic. Caustics for other systems are discussed in Secs. IX B and X and Appendix B.

Unlike the harmonic oscillator, the time $t_Q - t_P$ for the piecewise-linear oscillator to reach a kinetic focus depends on the coordinates (x_P, t_P) of the initial event P. For example, we have shown that $t_Q - t_P = \left(\frac{4}{3}\right)T_0/2$ for $x_P=0$. If $x_P \neq 0$ but still small, we find that $t_Q - t_P < \left(\frac{4}{3}\right)T_0/2$ by approximately $(8x_P/gT_0^2)T_0/2$.

B. Quartic oscillator

Pure quartic potentials are rare in nature,[74] but a mechanical model is easily constructed.[79] A particle is linked by harmonic springs on both sides along the y axis. The equilibrium position is $y=0$ and both springs are assumed to be relaxed in this position. Oscillations along the y axis are harmonic, but for small transverse oscillations in the x direction the potential has the form $U(x) = Cx^4 + \mathcal{O}(x^6)$.

Figure 4 shows a family of worldlines for the quartic oscillator. The second order variation of the action for a worldline $x_0(t)$ is given from Eq. (53) as

$$\delta^2 S = \frac{\alpha^2}{2}\int_0^{t_R} [m\dot{\phi}(t)^2 - 12Cx_0(t)^2\phi(t)^2]dt, \tag{65}$$

where we have taken $t_P=0$ and $\phi(t)$ is an arbitrary variational function. An exact analysis for a general worldline $x_0(t)$ is complicated. We therefore analyze an approximate worldline that brings out the salient points.

Consider a periodic worldline $x_0(t)$ that starts from $P=(0,0)$ with $v_0>0$, as in Fig. 4. For a given energy or amplitude of motion the worldline can be approximated by[12,73]

$$x_0(t) \approx A_0 \sin(\omega_0 t). \tag{66}$$

Unlike the harmonic oscillator, the frequency ω_0 depends on the amplitude A_0. Action principles can be used in the direct (Rayleigh-Ritz) mode[12,73] to estimate ω_0, giving

$$\omega_0 = \frac{2\pi}{T_0} \approx \left(\frac{3C}{4m}\right)^{1/2} A_0. \tag{67}$$

As discussed in Ref. 12, the variational result (67) is accurate to better than 1%; Eqs. (66) and (67) can both be improved systematically with the direct variational method if required. [A direct variational method finds true trajectories from a variational principle (here an action principle) without use of the equations of motion.]

We can analyze $\delta^2 S$ for the quartic oscillator in the same manner as for the piecewise-linear oscillator. We substitute Eqs. (58) and (66) into Eq. (65) and perform the integrations. The results are similar and we omit the details.

As we have discussed, the cut-off time for minimum action trajectories is the kinetic focus time t_Q; beyond this time the action is a saddle point. We recall [see Eq. (4) and the argument there] that for a family of worldlines $x(t, v_0)$ all starting at event P with differing initial velocity v_0, the kinetic focus of the worldline with initial velocity v_0 occurs when $\partial x(t, v_0)/\partial v_0 = 0$. To apply this condition to the worldline in Eq. (66), we first express A_0 and ω_0 in Eq. (66) in terms of v_0. For brevity we write Eq. (67) as $\omega_0 = \beta A_0$, where $\beta = (3C/4m)^{1/2}$. We also have $v_0 = \omega_0 A_0$ from differentiation with respect to time of Eq. (66). From these two relations we obtain $A_0 = \beta^{-1/2}v_0^{1/2}$ and $\omega_0 = \beta^{1/2}v_0^{1/2}$. The condition for the kinetic focus is then

$$\frac{\partial}{\partial v_0}[\beta^{-1/2}v_0^{1/2}\sin(\beta^{1/2}v_0^{1/2}t)] = 0 \tag{68}$$

or

$$\tfrac{1}{2}\beta^{-1/2}v_0^{-1/2}\sin(\beta^{1/2}v_0^{1/2}t) + \beta^{-1/2}v_0^{1/2}\cos(\beta^{1/2}v_0^{1/2}t) \times \left(\tfrac{1}{2}\right)\beta^{1/2}v_0^{-1/2}t = 0. \tag{69}$$

We let $\beta^{1/2}v_0^{1/2} = \omega_0$ and obtain

$$\tan(\omega_0 t) = -\omega_0 t. \tag{70}$$

Equation (70) is satisfied for $t=t_Q$ (and for times of later kinetic foci). The smallest positive root θ_Q of $\tan\theta = -\theta$ is $\theta_Q \approx 0.646\pi$. The time of the kinetic focus is then given by $\omega_0 t_Q \approx 0.646\pi$ or $t_Q \approx 0.646(T_0/2)$ for worldline 0 in Fig. 4; the same fraction of the half-period for the other worldlines is shown there. Because the worldline Eq. (66) is approximate, the location of the kinetic focus is also approximate. Note that t_Q is earlier than the half-period $T_0/2$ for the quartic oscillator.

The location x_Q of the kinetic focus $Q(x_Q,t_Q)$ of a particular worldline is found from t_Q using Eq. (66):

$$x_Q = A_0 \sin(\omega_0 t_Q) = A_0 \sin \theta_Q, \tag{71}$$

where θ_Q=0.646π and A_0 is the amplitude of the particular worldline. The locus of the various kinetic foci $Q(x_Q,t_Q)$ of the family of worldlines in Fig. 4 or caustic can be found by relating A_0 to t_Q. From Eq. (67) we have $A_0 = \omega_0/\beta$, where $\beta=(3C/4m)^{1/2}$ and $\omega_0 t_Q=\theta_Q$. Thus we find

$$x_Q = \frac{B}{t_Q}, \tag{72}$$

where $B=\theta_Q \sin \theta_Q/\beta=1.82(4m/3C)^{1/2}$. This caustic in Fig. 4 is a simple hyperbola and it too divides space-time (see the discussion of the piecewise-linear oscillator caustic).

We have seen examples for which the kinetic focus time is earlier than, equal to, and later than the half-period, corresponding, respectively, to oscillators whose frequency increases with amplitude, is independent of amplitude, and decreases with amplitude.

X. REPULSIVE INVERSE SQUARE POTENTIAL

The previous examples were systems with exclusively bound motions. We now show the corresponding results for a system whose unbound worldlines describe scattering from the potential

$$U(x) = \frac{C}{x^2}, \tag{73}$$

with $C>0$. It might seem surprising that a worldline in a scattering potential, where motion is unbound, can have a kinetic focus, because there is no kinetic focus for free particle worldlines or for worldlines in the scattering potentials $U(x)=Cx$ and $U(x)=-Cx^2$. The difference is due to the curvatures of the potentials: the inverse square potential (73) has $U''(x)>0$, whereas the other potentials have $U''(x)\leq 0$. As discussed qualitatively in Sec. III, potentials with $U''>0$ are stabilizing/focusing, which can lead to a kinetic focus.

For a given initial position x_P and final position x_R in the potential (73), a worldline may be direct (direct motion from x_P to x_R) or indirect (backward motion from x_P to a turning point x_T followed by forward motion from x_T to x_R). For indirect worldlines the turning point (x_T,t_T) occurs where the kinetic energy is equal to zero, so that the total energy E is equal to the potential energy (73), yielding

$$x_T^2 = \frac{C}{E}. \tag{74}$$

The worldlines $x(t)$ for the potential (73) are calculated by integrating the energy conservation relation. We assume $(x_P,t_P)=(x_P,0)$ and find

$$x^2 = x_T^2 + \frac{2E}{m}(t \pm t_T)^2, \tag{75}$$

where $\pm$ apply to the direct/indirect worldline, respectively. For an indirect worldline with the initial event $P(x_P,0)$, Eq. (75) can be solved for the turn-around time t_T:

$$t_T = \left(\frac{m}{2E}\right)^{1/2} (x_P^2 - x_T^2)^{1/2}. \tag{76}$$

Some typical worldlines with $P(x_P,0)$ are shown in Fig. 6. A direct worldline of arbitrary length has minimum action (no kinetic focus). For indirect worldlines the kinetic foci Q are not the minimum x turning points, but rather the tangent points to the straight-line caustic given by

$$x_Q = \left(\frac{2C}{m}\right)^{1/2} \frac{t_Q}{x_P}. \tag{77}$$

The derivation and discussion of $\delta^2 S$ for comparison curves are similar to those given for the piecewise-linear oscillator in Sec. IX.

The derivation of the kinetic foci (x_Q,t_Q) and the caustic equation (77) by our standard method is cumbersome for this potential, so we use an alternative argument (cf. Ref. 54). We can eliminate t_T and x_T from Eq. (75) using Eqs. (74) and (76) and obtain a relation involving x, t, and the (conserved) energy E. By doing some routine algebra, we can solve for E:

$$(2t^2/m)E = x^2 + x_P^2 \pm 2xx_P\left[1 - \frac{2C}{m}\frac{t^2}{x^2x_P^2}\right]^{1/2}. \tag{78}$$

The $\pm$ signs refer either to an indirect/direct pair of worldlines or to two indirect worldlines; both situations are possible as seen in Fig. 6. A kinetic focus arises here when two indirect worldlines coalesce into one; the locus of kinetic foci forms the caustic or envelope in Fig. 6. When the trajectories coalesce, their energies coincide. From Eq. (78) we see that the condition for coinciding energies is the vanishing of the term in square brackets. The caustic relation is thus found to be Eq. (77).

Equation (77) for the caustic is seen to be plausible by the following argument. From Fig. 6, worldline 2, which starts from rest (v_0=0 or zero initial slope), has the caustic as its asymptote. The equation for this asymptote is easily calculated from Eq. (75) by setting t_T=0 and $E=C/x_P^2$ and taking t large.

As we have seen, the indirect worldlines each have a kinetic focus. In contrast to the oscillator systems studied earlier, subsequent kinetic foci do not exist for this system.

XI. GENERALIZATIONS

Extensions of the results of this paper to two- and three-dimensional motion and multiparticle systems are straightforward, primarily because the action and energy are scalars; adding dimensions or particles sums the corresponding scalar quantities. We let x_i denote the coordinates, for example, $(x_1,x_2)=(x,y)$ for motion of a single particle in two dimensions; $(x_1,x_2,x_3)=(x,y,z)$ for motion of a single particle in three dimensions; and $(x_1,\ldots,x_6)$ for two particles in three dimensions, where $(x_1,x_2,x_3)=(x,y,z)$ for particle one and $(x_4,x_5,x_6)=(x,y,z)$ for particle two. Equation (5) generalizes to

$$x_i = x_i^{(0)} + \alpha\phi_i, \tag{79}$$

where $x_i^{(0)}$ and x_i are the coordinates of a point on the actual worldline and varied curve, respectively. This formalism leads to obvious generalizations of the equations of Sec. IV. In particular, for one or more particles of mass m, the Lagrangian is

$$L = \sum_i \frac{1}{2} m\dot{x}_i^2 - U, \qquad (80)$$

where U can be a function of all the x_i and time. Equation (19) generalizes to

$$\delta^2 S = \frac{\alpha^2}{2}\int_P^R \left[-\sum_{ij} \phi_i\phi_j \frac{\partial^2 U}{\partial x_i \partial x_j} + \sum_i m\dot{\phi}_i^2 \right] dt. \qquad (81)$$

The kinetic energy in Eq. (80) is quadratic in the velocities and thus positive, which leads to S having a minimum or a saddle point (never a maximum) for true worldlines (see the argument at the end of Sec. IV). It would be interesting to investigate the possible extension of this result and the corresponding result for the Maupertuis action W (see Appendix A) to more general Lagrangians, including relativistic Lagrangians and Lagrangians containing terms linear in the velocities (for example, magnetic field terms, gyroscopic terms).

Appendix B describes the 2D motion of a particle in two types of gravitational potentials and in harmonic oscillator potentials. The criteria for the minimum action and location of kinetic foci[80] are similar, but, unlike one dimension, two trajectories connecting P to R can both have minimum action for the attractive $1/r$ potential. The analysis generalizes easily to other dimensions and multi-particle systems, but the calculations are more complicated for complex worldlines, for example, when the motion is chaotic. For many-particle systems it is unlikely that we will want to specify in advance the complete final as well as initial configuration, because a major goal of mechanics is to find the final configuration.[81] In such cases these powerful deterministic tools may be less useful than modern statistical mechanical methods, although Helmholtz, Boltzmann, Planck, and others have attempted to base the second law of thermodynamics on action principles for the molecular motions.[73]

We have been careful to use partial derivatives of the potential function with respect to position, because the potential energy $U(x,t)$ can be an explicit function of time. The results of this paper can be applied in principle to motion in time-dependent potentials, in which the energy of the particle may not be a constant of the motion. New qualitative features may arise if U is explicitly time dependent; for example, we expect that $\delta^2 S$ can remain positive for some long worldlines in potentials with $U''>0$.[84] As an example, consider the quartic oscillator with the time-dependent external forcing $F(t)$. The potential becomes

$$U(x,t) = Cx^4 - xF(t) \qquad (82)$$

and has $U''>0$ for all x except $x=0$. A common choice is $F(t)=F_0 \cos \omega t$, but other choices are also of interest, for example, quasiperiodic forcing $F(t)=F_1 \cos \omega_1 t + F_2 \cos \omega_2 t$, with ω_2/ω_1 irrational. Alternatively, we can introduce parametric forcing by modulating C. The unforced oscillator has only equilibrium and periodic worldlines, which are stable. Depending on the initial conditions and potential parameters, the forced oscillator can also have unstable periodic worldlines,[30] quasiperiodic worldlines,[87,88] and chaotic (aperiodic, bounded, exponentially unstable) worldlines.[90] Similarly the potential $U(x)=C|x|$ can be made time dependent.[91] Second variations and kinetic foci have been studied for some worldlines of various oscillators with time dependent potentials,[92–94] but as far as we know not for chaotic worldlines in particular.

Chaotic behavior can arise in higher dimensions even without explicitly time-dependent potentials. As an example, the worldlines for the Henon-Heiles oscillator

$$U(x,y) = \tfrac{1}{2}k(x^2+y^2) + \alpha x^2 y - \tfrac{1}{3}\beta y^3 \qquad (83)$$

are chaotic for certain values of the initial conditions and parameters.[95] Some studies of $\delta^2 S$ and kinetic foci have been done on periodic worldlines for this system,[93] but we are not aware of any studies for chaotic worldlines.

It would be interesting and challenging to study $\delta^2 S$ for chaotic worldlines.[96] We hypothesize that kinetic foci will not exist if the worldline is sufficiently chaotic. In such cases worldlines with incremental difference in velocity at initial event P may recross pseudorandomly in time, but the severe instability may prevent the two worldlines from smoothly coalescing, as required for the existence of a kinetic focus Q. Worldlines PR lacking kinetic foci have $\delta^2 S>0$ for arbitrary final events R, so that the action is expected to remain a minimum in such cases, even for long worldlines in potentials having $U''>0$, such as Eq. (82).

XII. SUMMARY

We have investigated the nature of the stationary value of the Hamilton action S for the worldlines of a single particle in one dimension with potential energy function $U(x)$. We showed that when no kinetic focus exists, the action is a minimum for worldlines of arbitrary length. When a kinetic focus exists, and when a worldline terminates before reaching its kinetic focus, the action is still a minimum. In contrast, when a worldline terminates beyond its kinetic focus, its action is a saddle point. The value of the action S is never a true maximum for a true worldline. These results were illustrated for the harmonic oscillator, two anharmonic oscillators, and a scattering system. Extensions to time-dependent 1D potentials and to multidimensional potentials were discussed briefly. The appendices supply parallel results for spatial orbits described by Maupertuis' action W and give examples for 2D motion for both S and W. Corresponding results for some newer action principles have not yet been derived, and open questions about these newer action principles are sketched in Appendix C.

ACKNOWLEDGMENTS

CGG thanks NSERC for financial support. EFT acknowledges James Reid-Cunningham for providing opportunities for extended thought, and Alar Toomre for tutoring in mathematics. Both authors thank Bernie Nickel for many suggestions, including pointing out the insufficiency of the Culverwell-Whittaker argument as shown in Sec. VI and providing the original analysis elaborated in Sec. VII. We also thank Adam Pound, Slavomir Tuleja, Don Lemons, Donald Sprung, and John Taylor for a careful reading of the manuscript and many useful suggestions. Brenda Law, Igor Tolokh, and Will M. Farr are thanked for their assistance in the production of this paper.

APPENDIX A: THE MAUPERTUIS ACTION PRINCIPLE

There are two major versions of the action and two corresponding action principles. The Hamilton or time-dependent action S and the corresponding Hamilton action principle were introduced in Sec. I. The Maupertuis or time-independent action W is defined along an arbitrary trial trajectory connecting $P(x_P,t_P)$ to $R(x_R,t_R)$ by

$$W=\int_{x_P}^{x_R} p\,dx=\int_{t_P}^{t_R} m\dot{x}\frac{dx}{dt}\,dt=\int_{t_P}^{t_R} 2K\,dt, \qquad \text{(A1)}$$

where the first (time-independent) form is the general definition with $p=\partial L/\partial\dot{x}$ the canonical momentum, and the last (time-dependent) form is valid generally for normal systems[73] in which the kinetic energy K is quadratic in the velocity components. For normal systems W is positive for all trajectories in all potentials (unlike S). The Maupertuis action principle states that in conservative systems W is stationary ($\delta W=0$) for an actual trajectory when comparing trial trajectories all with the same fixed energy E and the same fixed start and end positions x_P and x_R. Note that in Maupertuis' principle the energy E is fixed and the duration $T=(t_R-t_P)$ is not, the opposite conditions of those in Hamilton's principle. The constraint of fixed end positions x_P and x_R is common to both principles. Hamilton's principle is valid for both conservative systems and nonconservative systems with $U=U(x,t)$. The conventional Maupertuis principle is valid only for conservative systems; the extension to nonconservative systems is discussed in Ref. 73. Maupertuis' principle can be used in its time-independent form to find spatial orbits [for example, $(x_P,y_P)\rightarrow(x_R,y_R)$] and in its time-dependent form to find space-time trajectories or worldlines, for example, $(x_P,y_P,t_P)\rightarrow(x_R,y_R,t_R)$.

The Hamilton and Maupertuis action principles can be stated[73] succinctly in terms of constrained variations as $(\delta S)_T=0$ and $(\delta W)_E=0$, respectively, where the constraints of fixed T and fixed E are denoted explicitly as subscripts, and the constraint of fixed end positions x_P and x_R is implicit. Along an arbitrary trial trajectory $P\rightarrow R$, S and W are related[73] by a Legendre transformation, that is,

$$S=W-\bar{E}T, \qquad \text{(A2)}$$

where $\bar{E}=\int_{t_P}^{t_R}H\,dt/T$ is the mean energy along the arbitrary trajectory, $T=(t_R-t_P)$ is the duration, and H is the Hamiltonian. Equation (A2) follows by integrating over time from t_P to t_R along the arbitrary trajectory the corresponding Legendre transform $L=p\dot{x}-H$ and deserves to be better known. Along an actual trajectory of a conservative system, Eq. (A2) reduces to the well-known relation[97] $S=W-ET$, where E is the constant energy of the actual trajectory. Using Eq. (A2) enables us to relate the Hamilton and Maupertuis principles.[73]

Parallel yet distinct discussions have been developed for the second variations of S and W because the kinetic foci, which play such an important role in determining the second variation, can differ for the two actions.[30,98,99]

An intuitive argument why W can never be a true maximum for actual paths was given in Ref. 100 for normal systems. Consider an actual path $x_P\rightarrow A\rightarrow B\rightarrow x_R$ that makes the first form of W in Eq. (A1) stationary. Here A and B are two arbitrary intermediate positions between x_P and x_R. Consider a second trial path $x_P\rightarrow A\rightarrow B\rightarrow A\rightarrow B\rightarrow x_R$ that has an extra "loop" inserted, with the momentum p reversed at every point along $B\rightarrow A$ compared to $A\rightarrow B$. This comparison path satisfies the constraint of having the same energy as the actual path, but has a larger action because pdx is always positive. Thus W for the actual path cannot be a true maximum.

APPENDIX B: TWO-DIMENSIONAL TRAJECTORIES

1. Gravitational fields

We have shown that the Hamilton action S is a minimum for all radial/vertical trajectories in the $1/r$ and linear gravitational potentials discussed in Sec. V. This minimum action property may not hold for 2D trajectories, as we shall discuss. A possible nonminimum in the action is more evident for the Maupertuis action W, Eq. (A1), for which the true trajectories are defined by giving the two end positions (x_P,y_P) and (x_R,y_R) and the energy E; we therefore discuss 2D orbits for W first. We choose the x and y axes in the plane of the orbit.

For $U(x,y)=mgy$, with y the vertical direction and x the horizontal, it is well known that two actual spatial orbits (parabolas) with the same energy E can connect two given positions, the origin $(x_P,y_P)=(0,0)$ say, and the final position (x_R,y_R), provided that (x_R,y_R) lies within the "parabola of safety"—the envelope[43,101] of the parabolic orbits of energy E originating at (0,0) (see Fig. 7). If (x_R,y_R) lies on the parabola of safety, which is the locus of the spatial kinetic foci (x_Q,y_Q) or caustic, there is one actual orbit between the fixed end positions. If (x_R,y_R) lies outside the parabola of safety, no orbit of energy E can connect it to the origin. These conclusions are evident in Fig. 7. As we have seen in Sec. II, the intersection of two paths implies the existence of a (different) kinetic focus for each of the paths. W is a minimum for the path for which (x_R,y_R) precedes its kinetic focus, and is a saddle point for the path for which (x_R,y_R) lies beyond its kinetic focus.

Similarly, as first shown by Jacobi,[22] typically[102] two given positions (x_P,y_P) and (x_R,y_R) in the gravitational potential $(1/r)$ can be connected by two actual orbits (ellipses) (and therefore four paths of less than one revolution) of the same energy E. Again, the fact that more than one true path can connect the two end positions leads to a nonminimum in the action W for actual paths connecting (x_P,y_P) to (x_R,y_R) when (x_R,y_R) lies beyond the kinetic focus. An example is shown in Fig. 13. The intersection points of the orbits show two ellipses connecting point P to other points. The outer curve is the envelope/caustic, which is also elliptical with foci at P and the force center. The spatial kinetic foci for action W lie on this outer ellipse. The second spatial kinetic focus occurs at P itself, following one revolution. There is no envelope for the hyperbolic scattering orbits for the attractive $1/r$ potential.[104] Methods of determining spatial caustics are similar to those for determining space-time caustics and are discussed in Ref. 105.

To discuss 2D space-time worldlines for the Hamilton action S, which depend on the two end positions and the time interval that now specify a worldline, note that for the $1/r$ potential, typically two actual worldlines can connect two

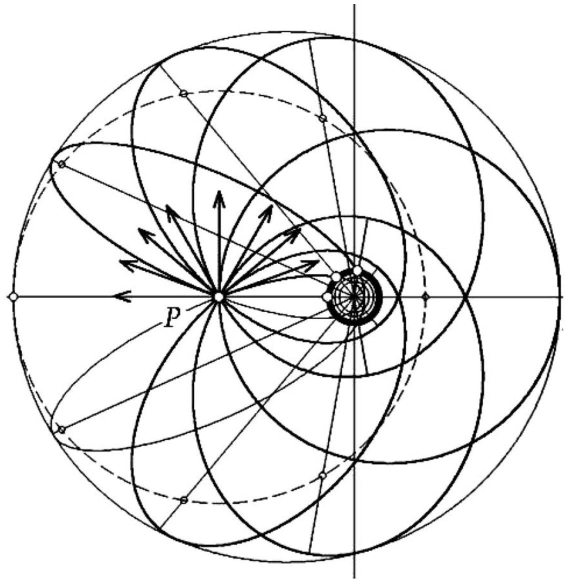

Fig. 13. A family of elliptical trajectories starting at P with the same speed $|\mathbf{v}_0|$ (the directions θ_0 of $\mathbf{v}_0$ differ), and hence the same energy E, the same major axis $2a$, and the same period T_0, in a $1/r$ gravitational potential. The value of v_0 exceeds that necessary to generate a circular orbit. The center of force is the Earth (heavy circle). The dashed circle gives the locus of the second focus of the ellipses (a circle centered at P). The outer ellipse, with foci at P and the Earth, is the envelope of the family of ellipses and the locus of the spatial kinetic foci relevant for action W. The space-time kinetic foci of the family, relevant for action S, all occur at time $t_Q = T_0$. (Figure adapted from Butikov, Ref. 101.)

given positions (x_P, y_P) and (x_R, y_R) in the given time interval $(t_R - t_P)$.[107] The fact that two actual worldlines exist here is illustrated in Fig. 14, which shows two elliptical trajectories that connect the initial and final points in the same time. Choose $t_P = 0$ for simplicity. The kinetic focus time t_Q for S is the period T_0 as is clear intuitively from Fig. 13, which shows a family of trajectories leaving point P and converging back on P in the same time T_0. A rigorous proof that $t_Q = T_0$ can also be given.[110] The space-time kinetic focus here is of the focal point type, as in Fig. 1 for the sphere geodesics and in Fig. 3 for the harmonic oscillator worldlines. As seen in Fig. 14, if two trajectories connect P to R in time $(t_R - t_P) < T_0$, both have the minimum action; the fact that both trajectories have minimum action is in contrast to one dimension, where we have seen in general that when two space-time paths exist, one of them has a saddle point in the

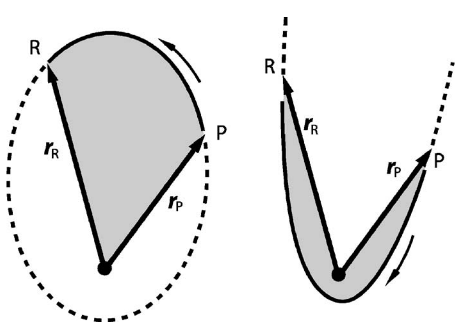

Fig. 14. Two different elliptical trajectories typically can connect $P = (\mathbf{r}_P, t_P)$ to $R = (\mathbf{r}_R, t_R)$ in the same time $(t_R - t_P)$ for the attractive $1/r$ potential. (Adapted from Bate *et al.*, Ref. 108.)

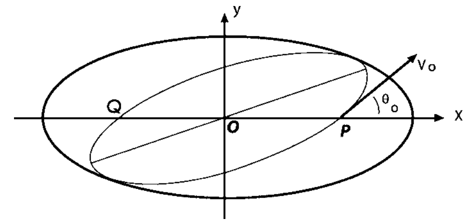

Fig. 15. An elliptical orbit (tilted ellipse) in a 2D isotropic harmonic oscillator potential $U(r) = (\frac{1}{2})kr^2$ with force center at O. A family of trajectories is launched from P with equal initial speeds $|\mathbf{v}_0|$ and various directions θ_0. One member of the family is shown. The envelope of the family is the outer ellipse (heavy line), with foci at P and Q (coordinates $x_Q = -x_P$, $y_Q = y_P = 0$). Points on the envelope are the kinetic foci for the spatial orbits. Point Q, occurring at time $t_Q = T_0/2$, where T_0 is the period, is the kinetic focus for the space-time trajectories (worldlines). (Figure adapted from French, Ref. 101.)

action. For this particular potential, only one space-time path exists for one dimension, which therefore has minimum action S (see Sec. V).

For the potential $U(x,y) = mgy$, only one 2D (parabolic-shaped) actual worldline can connect two given positions (x_P, y_P) and (x_R, y_R) in the given time $(t_R - t_P)$. Kinetic foci for the space-time trajectories therefore cannot arise, and hence S is always a minimum for actual worldlines for this potential.[112] The fact that for this potential only one true worldline can connect two points in a given time is in contrast with the Maupertuis action W for which we have seen that some pairs of positions (x_P, y_P) and (x_R, y_R) can be connected by more than one path of a given energy E so that kinetic foci can exist for the spatial orbits. This contrast between S and W also holds for the vertical 1D paths in the potential $U = mgy$. For S only one actual 1D worldline can connect a given y_P to a given y_R in the given time $(t_R - t_P)$, which leads to the conclusion that all 1D actual worldlines minimize S (see Sec. V). For W typically two actual 1D paths of given energy E can connect y_P to y_R, which leads to the conclusion that not all 1D actual paths minimize W (some are saddle points). For the potential $U(x,y) = mgy$ there is always one actual worldline that can connect two given spatial points in a given time, for both 2D and 1D worldlines (S is always a minimum for these worldlines). In contrast, there may be no actual path that connects two given spatial points for a given energy, for both one and two dimensions. Figure 7 shows examples (final points outside the caustic) of this nonexistence of actual paths.

2. Harmonic oscillators

The potential for a 2D isotropic harmonic oscillator is $U(x,y) = \frac{1}{2}k(x^2 + y^2) \equiv \frac{1}{2}kr^2$. The spatial orbits are ellipses with the force center $(r=0)$ at the center of the ellipse. A family of ellipses launched from P in Fig. 15, with equal values of $|\mathbf{v}_0|$ (and hence equal energies E) and various directions of $\mathbf{v}_0$, has an envelope/caustic that is also elliptical.[101] The envelope is the outer ellipse (heavy line) in Fig. 15. The spatial kinetic foci for the action W lie on the caustic. Reference 105 gives methods for deriving the spatial caustic.

To locate the space-time kinetic focus, which is relevant for the action S, we apply the relation given in Ref. 80 for 2D trajectories $\mathbf{x}(t, \mathbf{v}_0)$. The matrix $\partial x_i / \partial v_{0j}$ is diagonal, so that

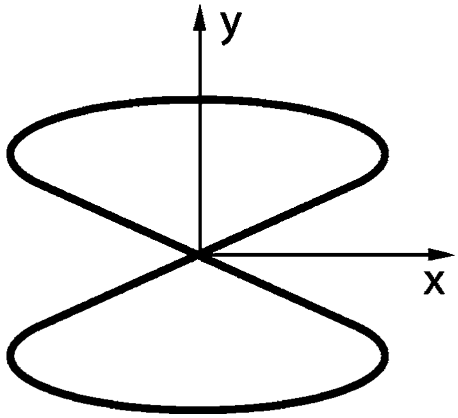

Fig. 16. A periodic orbit of a 2D anisotropic harmonic oscillator with commensurate frequencies (here $\omega_1/\omega_2=2$) (Ref. 113). The space-time kinetic focus occurs at time $t_Q=T_1/2$, where T_1 is the period for x-motion, for any initial event $(x_P,y_P,t_P=0)$ (see text), and the spatial kinetic focus for initial position $(x_P,y_P)=(0,0)$ is located on a parabolic spatial caustic (see Ref. 105).

the determinant condition reduces to $(\partial x/\partial v_{0x})(\partial y/\partial v_{0y})=0$, and we obtain separate 1D conditions for the x and y motions, that is, $\partial x/\partial v_{0x}=0$ or $\partial y/\partial v_{0y}=0$. Choose $t_P=0$ for simplicity. We showed in Sec. VIII that the kinetic focus time t_Q for the 1D harmonic oscillator is $T_0/2$, where $T_0=2\pi/\omega_0$ and $\omega_0=(k/m)^{1/2}$. Thus we have $t_Q=T_0/2$ for the isotropic 2D harmonic oscillator.

The space-time kinetic focus for the trajectories of the 2D anisotropic harmonic oscillator with the potential $U(x,y)=\frac{1}{2}k_1x^2+\frac{1}{2}k_2y^2$ and $k_1\neq k_2$ can be derived similarly. The orbits are Lissajous figures, closed (periodic) for rational ratios ω_1/ω_2 as in Fig. 16, and open (quasiperiodic) for ω_1/ω_2 irrational as in Fig. 17, where $\omega_i=(k_i/m)^{1/2}$. The problem is again separable into x and y motions, and the method of Ref. 80 yields for the kinetic focus time t_Q the value $T_0/2$, where T_0 is the smaller of T_1 and T_2 and $T_i=2\pi/\omega_i$. The determination of the spatial kinetic foci, relevant for W, is more complicated.[105]

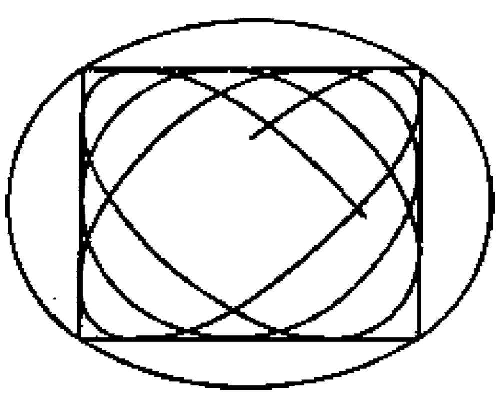

Fig. 17. A quasiperiodic orbit of a 2D anisotropic harmonic oscillator with incommensurate frequencies. The outer ellipse is the equipotential contour $U(x,y)=E$. The rectangle delimits the region of x-y space actually reached by the particular orbit (Ref. 114). The space-time kinetic focus occurs at time $t_Q=T_2/2$, where T_2 is the period for y-motion, for any initial event $(x_P,y_P,t_P=0)$.

APPENDIX C: OPEN QUESTIONS FOR SOME NEWER ACTION PRINCIPLES

Culverwell and Whittaker (see Sec. VI) framed their analysis in terms of the Maupertuis action W. For sufficiently short trajectories (the final position occurs before the kinetic focus) W is always a minimum, and for longer trajectories W is a saddle point. W is never a true maximum. In this paper we amended the Culverwell-Whittaker analysis and adapted it to the Hamilton action S. For times less than the kinetic focus time t_Q, the action S is always a minimum. For longer times S is a saddle point. S is never a true maximum. We refer to these results for W and S as “no-max” theorems.

It may be possible to extend the theorems to several newer action principles.[73,115] To state the newer principles, we first recall the notation for the Hamilton principle and the Maupertuis principle given in Appendix A:

$$(\delta S)_T=0 \quad \text{(Hamilton principle)}, \tag{C1}$$

$$(\delta W)_E=0 \quad \text{(Maupertuis principle)}, \tag{C2}$$

where $T=t_R-t_P$ (the duration) and E (the energy) denote the constraints. The additional constraints of fixed end positions x_P and x_R are implicit in Eqs. (C1) and (C2) and are understood to hold here and in all the action principles discussed in the following.

In recent years the Maupertuis principle (C2) has been extended to a generalized Maupertuis principle,[73,115]

$$(\delta W)_{\bar{E}}=0 \quad \text{(generalized Maupertuis principle)}, \tag{C3}$$

where $\bar{E}=\int_0^T H\,dt/T$ is the mean energy along the arbitrary trial trajectory, with H the Hamiltonian, and where for simplicity we choose $t_P=0$ and $t_R=T$. The constraint of fixed E in Eq. (C2) has been weakened to one of fixed mean energy $\bar{E}$ in Eq. (C3). Conservation of energy for actual trajectories is now a consequence of the principle (C3), rather than an assumption as in the original principle (C2).

Both the generalized Maupertuis principle (C3) and Hamilton principle (C1) have associated reciprocal principles:[73,115]

$$(\delta \bar{E})_W=0 \quad \text{(reciprocal Maupertuis principle)}, \tag{C4}$$

$$(\delta T)_S=0 \quad \text{(reciprocal Hamilton principle)}. \tag{C5}$$

The newer principles in Eqs. (C3)–(C5) have several advantageous features, computational and conceptual, as discussed in Refs. 12, 73, and 115. Additionally, the reciprocal Maupertuis principle (C4) is the direct classical analogue[115,116] (the classical $\hbar\to 0$ limit) of the well known Schrödinger quantum variational principle involving the mean energy.[117]

It would be of interest to prove the existence or nonexistence of a no-max or no-min theorem for these newer action principles. A Routh-type argument (see Appendix A) suggests that the generalized Maupertuis principle (C3) obeys a no-max theorem, but the examples worked out to date[12,73,115] provide no compelling evidence one way or the other for the other principles.

Other newer action principles are discussed in Refs. 73 and 115. We can completely relax the constraints of fixed T in Eq. (C1) and fixed $\bar{E}$ in Eq. (C3) with the help of Lagrange multipliers and obtain an unconstrained Hamilton principle, $\delta S=-E\,\delta T$, and an unconstrained Maupertuis principle, δW

$=T\delta\bar{E}$, where (for conservative systems) the Lagrange multipliers E and T are the energy and duration of the actual trajectory, respectively. The unconstrained Hamilton principle and unconstrained Maupertuis principle can also be written in the more suggestive forms $\delta(S+\lambda T)=0$ and $\delta(W+\lambda\bar{E})=0$, respectively, where $\lambda=E$ or $\lambda=-T$ is the corresponding constant Lagrange multiplier. These unconstrained principles still have the constraint of fixed end positions x_P and x_R; these constraints can be relaxed by introducing additional Lagrange multipliers.[73] It would also be of interest to prove the existence or nonexistence of no-max or no-min theorems for these various principles.

[a]Electronic mail: cgg@physics.uoguelph.ca
[b]Electronic mail: eftaylor@mit.edu

[1]R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1963), Vol. II, Chap. 19.

[2]A. Bork and A. Zellweger, "Least action via computer," Am. J. Phys. **37**, 386–390 (1969).

[3]J. P. Provost, "Least action principle on an air table," Am. J. Phys. **43**, 774–781 (1975).

[4]E. F. Taylor, "A call to action," Am. J. Phys. **71**, 423–425 (2003).

[5]J. Hanc, S. Tuleja, and M. Hancova, "Simple derivation of Newtonian mechanics from the principle of least action," Am. J. Phys. **71**, 386–391 (2003).

[6]J. Hanc, S. Tuleja, and M. Hancova, "Symmetries and conservation laws: Consequences of Noether's theorem," Am. J. Phys. **72**, 428–435 (2004).

[7]J. Hanc, E. F. Taylor, and S. Tuleja, "Deriving Lagrange's equations using elementary calculus," Am. J. Phys. **72**, 510–513 (2004).

[8]J. Hanc and E. F. Taylor, "From conservation of energy to the principle of least action: A story line," Am. J. Phys. **72**, 514–521 (2004).

[9]T. A. Moore, "Getting the most action out of the least action: A proposal," Am. J. Phys. **72**, 522–527 (2004).

[10]J. Ogborn and E. F. Taylor, "Quantum physics explains Newton's laws of motion," Phys. Educ. **40**, 26–34 (2005).

[11]E. F. Taylor, J. Hanc, and S. Tuleja, "Variational mechanics in one and two dimensions," Am. J. Phys. **73**, 603–610 (2005).

[12]C. G. Gray, G. Karl, and V. A. Novikov, "Direct use of variational principles as an approximation technique in classical mechanics," Am. J. Phys. **64**, 1177–1184 (1996) and references therein.

[13]J. L. Lagrange, *Analytical Mechanics (Mécanique Analytique)* (Gauthier-Villars, Paris, 1888–89), 2nd ed. (1811) (Kluwer, Dordrecht, 1997), pp. 183 and 219. The same erroneous statement occurs in work published in 1760–61, *ibid*, p. xxxiii.

[14]We easily found about two dozen texts using the erroneous term "maximum." See, for example, E. Mach, *The Science of Mechanics*, 9th ed. (Open Court, La Salle, IL, 1960), p. 463; A. Sommerfeld, *Mechanics* (Academic, New York, 1952), p. 208; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw Hill, New York, 1953), Part 1, p. 281; D. Park, *Classical Dynamics and its Quantum Analogues*, 2nd ed. (Springer, Berlin, 1990), p. 57; L. H. Hand and J. D. Finch, *Analytical Mechanics* (Cambridge U. P., Cambridge, 1998), p. 53; G. R. Fowles and G. L. Cassiday, *Analytical Mechanics*, 6th ed. (Saunders, Forth Worth, 1999), p. 393; R. K. Cooper and C. Pellegrini, *Modern Analytical Mechanics* (Kluwer, New York, 1999), p. 34. A similar error occurs in A. J. Hanson, *Visualizing Quaternions* (Elsevier, Amsterdam, 2006), p. 368, which states that geodesics on a sphere can have maximum length.

[15]E. F. Taylor and J. A. Wheeler, *Exploring Black Holes: Introduction to General Relativity* (Addison-Wesley Longman, San Francisco, 2000), pp. 1–7 and 3–4.

[16]The number of authors of books and papers using "extremum" and "extremal" is endless. Some examples include R. Baierlein, *Newtonian Dynamics* (McGraw-Hill, New York, 1983), p. 125; L. N. Hand and J. D. Finch, Ref. 14, p. 51; J. D. Logan, *Invariant Variational Principles* (Academic, New York, 1977) mentions both "extremal" and "stationary," p. 8; L. D. Landau and E. M. Lifschitz, *Mechanics*, 3rd ed. (Butterworth Heineman, Oxford, 2003), pp. 2 and 3; see also their *Classical Theory of Fields*, 4th ed. (Butterworth Heineman, Oxford, 1999), p. 25; S. T. Thornton and J. B. Marion, *Classical Dynamics of Particles and Systems*, 5th ed. (Thomson, Brooks/Cole, Belmont, CA, 2004), p. 231; R. K. Cooper and C. Pellegrini, Ref. 14.

[17]F. L. Pedrotti, L. S. Pedrotti, and L. M. Pedrotti, *Introduction to Optics* (Pearson Prentice Hall, Upper Saddle River, NJ, 2007), p. 22; J. R. Taylor, *Classical Mechanics* (University Science Books, Sausalito, 2005), Problem 6.5; P. J. Nahin, *When Least is Best* (Princeton U. P., Princeton, NJ, 2004), p. 133; D. S. Lemons, *Perfect Form* (Princeton U. P., Princeton, NJ, 1997), p. 8; D. Park, Ref. 14, p. 13; F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed. (McGraw Hill, New York, 1976), p. 15; R. Guenther, *Modern Optics* (Wiley, New York, 1990), p. 135; R. W. Ditchburn, *Light*, 3rd ed. (Academic, London, 1976), p. 209; R. S. Longhurst, *Geometrical and Physical Optics*, 2nd ed. (Longmans, London, 1967), p. 7; J. L. Synge, *Geometrical Optics* (Cambridge U. P., London, 1937), p. 3; G. P. Sastry, "Problem on Fermat's principle," Am. J. Phys. **49**, 345 (1981); M. V. Berry, review of *The Optics of Rays, Wavefronts and Caustics* by O. N. Stavroudis (Academic, New York, 1972), Sci. Prog. **61**, 595–597 (1974); V. Lakshminarayanan, A. K. Ghatak, and K. Thyagarajan, *Lagrangian Optics* (Kluwer, Boston, 2002), p. 16; V. Perlick, *Ray Optics, Fermat's Principle, and Applications to General Relativity* (Springer, Berlin, 2000), pp. 149 and 152.

[18]The seminal work on second variations of general functionals by Legendre, Jacobi, and Weierstrass and many others is described in the historical accounts of Refs. 19 and 20. Mayer's work (Ref. 19) was devoted specifically to the second variation of the Hamilton action. In our paper, we adapt Culverwell's work (Ref. 21) for the Maupertuis action W to the Hamilton action. Culverwell's work was preceded by that of Jacobi (Ref. 22) and Kelvin and Tait (Ref. 23).

[19]H. H. Goldstine, *A History of the Calculus of Variations From the 17th Through the 19th Century* (Springer, New York, 1980).

[20]I. Todhunter, *A History of the Progress of the Calculus of Variations During the Nineteenth Century* (Cambridge U. P., Cambridge, 1861) and (Dover, New York, 2005).

[21]E. P. Culverwell, "The discrimination of maxima and minima values of single integrals with any number of dependent variables and any continuous restrictions of the variables, the limiting values of the variables being supposed given," Proc. London Math. Soc. **23**, 241–265 (1892).

[22]C. G. J. Jacobi, "Zür Theorie der Variationensrechnung und der Differential Gleichungen." J. f.Math. **XVII**, 68–82 (1837). An English translation is given in Ref. 20, p. 243, and a commentary is given in Ref. 19, p. 156.

[23]W. Thomson (Lord Kelvin) and P. G. Tait, *Treatise on Natural Philosophy* (Cambridge U. P., Cambridge, 1879, 1912), Part I; reprinted as *Principles of Mechanics and Dynamics* (Dover, New York, 1962), Part I, p. 422.

[24]Gelfand and Fomin (Ref. 25) and other more recent books on calculus of variations are rigorous but rather sophisticated. A previous study (Ref. 26) of the nature of the stationarity of worldline action was based on the Jacobi-Morse eigenfunction method (Ref. 27), rather than on the more geometrical Jacobi-Culverwell-Whittaker approach.

[25]I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, translated by R. A. Silverman (Prentice Hall, Englewood Cliffs, NJ, 1963), Russian edition 1961, reprinted (Dover, New York, 2000).

[26]M. S. Hussein, J. G. Pereira, V. Stojanoff, and H. Takai, "The sufficient condition for an extremum in the classical action integral as an eigenvalue problem," Am. J. Phys. **48**, 767–770 (1980). Hussein *et al.* make the common error of assuming S can be a true maximum. See also J. G. Papastavridis, "An eigenvalue criterion for the study of the Hamiltonian action's extremality," Mech. Res. Commun. **10**, 171–179 (1983).

[27]M. Morse, *The Calculus of Variations in the Large* (American Mathematical Society, Providence, RI, 1934).

[28]As we shall see, the nature of the stationary value of Hamilton's action S (and also Maupertuis' action W) depends on the sign of second variations $\delta^2 S$ and $\delta^2 W$ (defined formally in Sec. IV), which in turn depends on the existence or absence of kinetic foci (see Secs. II and VII). The same quantities (signs of the second variations and kinetic foci) are also important in classical mechanics for the question of dynamical stability of trajectories (Refs. 23 and 29–31), and in semiclassical mechanics where they determine the phase loss term in the total phase of the semiclassical propagator due to a particular classical path (Refs. 32 and 33). The phase loss depends on the Morse (or Morse-Maslov) index, which equals the number of kinetic foci between the end-points of the trajectory (see Ref. 34). Further, in devising computational algorithms to find the stationary points of the action (either S or W), it is useful to know whether we are seeking a minimum or a saddle point, because different algorithms (Ref. 35) are often used for the two cases. As we discuss in this paper, it is the

sign of $\delta^2 S$ (or $\delta^2 W$) that determines which case we are considering. Practical applications of the mechanical focal points are mentioned at the end of Ref. 37.

[29]E. J. Routh, *A Treatise on the Stability of a Given State of Motion* (Macmillan, London, 1877), p. 103; reissued as *Stability of Motion*, edited by A. T. Fuller (Taylor and Francis, London, 1975).

[30]J. G. Papastavridis, "Toward an extremum characterization of kinetic stability," J. Sound Vib. **87**, 573–587 (1983).

[31]J. G. Papastavridis, "The principle of least action as a Lagrange variational problem: Stationarity and extremality conditions," Int. J. Eng. Sci. **24**, 1437–1443 (1986); "On a Lagrangean action based kinetic instability theorem of Kelvin and Tait," Int. J. Eng. Sci. **24**, 1–17 (1986).

[32]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981), p. 143.

[33]M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990), p. 184.

[34]In general, saddle points can be classified (or given an index (Ref. 27)) according to the number of independent directions leading to maximum-type behavior. Thus the point of zero-gradient on an ordinary horse saddle has a Morse index of unity. The Morse index for an action saddle point is equal to the number of kinetic foci between the end-points of the trajectory (Ref. 32, p. 90). Readable introductions to Morse theory are given by R. Forman, "How many equilibria are there? An introduction to Morse theory," in *Six Themes on Variation*, edited by R. Hardt (American Mathematical Society, Providence, RI, 2004), pp. 13–36, and B. Van Brunt, *The Calculus of Variations* (Springer, New York, 2004), p. 254.

[35]F. Jensen, *Introduction to Computational Chemistry* (Wiley, Chichester, 1999), Chap. 14.

[36]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, 4th ed. (originally published in 1904) (Cambridge U. P., Cambridge, 1999), p. 253. The same example was treated earlier by C. G. J. Jacobi, *Vorlesungen Über Dynamik* (Braunschweig, Vieweg, 1884), reprinted (Chelsea, New York, 1969), p. 46.

[37]A closer mechanics-optics analogy is between a kinetic focus (mechanics) and a caustic point (optics) (Ref. 38). The locus of limiting intersection points of pairs of mechanical spatial orbits is termed an envelope or caustic (see Fig. 7 for an example), just as the locus of limiting intersection points of pairs of optical rays is termed a caustic. In optics, the intersection point of a bundle of many rays is termed a focal point; a mechanical analogue occurs naturally in a few systems, for example, the sphere geodesics of Fig. 1 and the harmonic oscillator trajectories of Fig. 3, where a bundle of trajectories recrosses at a mechanical focal point. In electron microscopes (Refs. 39 and 40), mass spectrometers (Ref. 41), and particle accelerators (Ref. 42), electric and magnetic field configurations are designed to create mechanical focal points.

[38]M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1970), pp. 130 and 734; J. A. Luck and J. H. Andrews, "Optical caustics in natural phenomena," Am. J. Phys. **60**, 397–407 (1992).

[39]M. Born and E. Wolf, Ref. 38, p. 738; L. A. Artsimovich and S. Yu. Lukyanov, *Motion of Charged Particles in Electric and Magnetic Fields* (MIR, Moscow, 1980).

[40]P. Grivet, *Electron Optics*, 2nd ed. (Pergamon, Oxford, 1972); A. L. Hughes, "The magnetic electron lens," Am. J. Phys. **9**, 204–207 (1941); J. H. Moore, C. C. Davis, and M. A. Coplan, *Building Scientific Apparatus*, 3rd ed. (Perseus, Cambridge, MA, 2003), Chap. 5.

[41]P. Grivet, Ref. 40, p. 822; J. H. Moore *et al.*, Ref. 40.

[42]M. S. Livingston, *The Development of High-Energy Accelerators* (Dover, New York, 1966); M. L. Bullock, "Electrostatic strong-focusing lens," Am. J. Phys. **23**, 264–268 (1955); L. W. Alvarez, R. Smits, and G. Senecal, "Mechanical analogue of the synchrotron, illustrating phase stability and two-dimensional focusing," Am. J. Phys. **43**, 292–296 (1975); A. Chao *et al.*, "Experimental investigation of nonlinear dynamics in the Fermilab Tevatron," Phys. Rev. Lett. **61**, 2752–2755 (1988). For recent texts, see A. W. Chao, "Resource Letter PBA-1: Particle beams and accelerators," Am. J. Phys. **74**, 855–862 (2006).

[43]V. G. Boltyanskii, *Envelopes* (MacMillan, New York, 1964).

[44]P. T. Saunders, *An Introduction to Catastrophe Theory* (Cambridge U. P., Cambridge, 1980), p. 62.

[45]Systems with subsequent kinetic foci are discussed in Secs. VIII and IX. For examples with only a single kinetic focus, see Figs. 6 and 7.

[46]M. C. Gutzwiller, "The origins of the trace formula," in *Classical, Semiclassical and Quantum Dynamics in Atoms*, edited by H. Friedrich and B. Eckhardt (Springer, New York, 1997), pp. 8–28.

[47]This type of variation, $\delta x = \alpha\phi$, $\delta\dot{x} = \alpha\dot{\phi}$, where δx and $\delta\dot{x}$ vanish together for $\alpha \to 0$, is termed a weak variation. See, for example, C. Fox, *An Introduction to the Calculus of Variations* (Oxford U. P., Oxford, 1950), reprinted (Dover, New York, 1987), p. 3.

[48]H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics*, 3rd ed. (Addison-Wesley, San Francisco, 2002), p. 44.

[49]As discussed in Sec. VI, for a true worldline $x_0(t)$ or PQ, where Q is the (first) kinetic focus, we have $\delta^2 S = 0$ for one special variation (and $\delta^2 S > 0$ for all other variations) as well as $\delta S = 0$ for all variations. Further, we show that $\delta^3 S$, etc., all vanish for the special infinitesimal variation for which $\delta^2 S$ vanishes, and that $S - S_0 = 0$ to second-order for larger such variations. In the latter case, typically $\delta^3 S$ is nonvanishing due to a single coalescing alternative worldline. In atypical (for one dimension (Ref. 58)) cases, more than one coalescing alternative worldline occurs, and the first nonvanishing term is $\delta^4 S$ or higher order—see Ref. 32, pp. 122–127. The harmonic oscillator is a limiting case where $\delta^k S = 0$ for all k for the special variation around worldline PQ, which reflects the infinite number of true worldlines which connect P to Q, and which can all coalesce by varying the amplitude (see Fig. 3). In Morse theory (Refs. 27 and 34) the worldline PQ is referred to as a degenerate critical (stationary) point.

[50]D. Morin, *Introductory Classical Mechanics*, ⟨www.courses.fas.harvard.edu/~phys16/Textbook/⟩ Chap. 5, p. V-8.

[51]This statement and the corresponding one in Ref. 52 must be qualified. It is in general not simply a matter of the time interval $(t_R - t_P)$ being short. The spatial path of the worldline must be sufficiently short. When, as usually happens, more than one actual worldline can connect a given position x_P to a given position x_R in the given time interval $(t_R - t_P)$, for short time intervals only the spatially shortest worldline will have the minimum action. For example, the repulsive power-law potentials $U(x) = C/x^n$ (including the limiting case of a hard-wall potential at the origin for $C \to 0$) and the repulsive exponential potential $U(x) = U_0 \exp(-x/a)$ have been studied (Ref. 53). No matter how short the time interval $(t_R - t_P)$, two different worldlines can connect given position x_P to given position x_R. The fact that two different true worldlines can connect the two points in the given time interval leads to a kinetic focus time t_Q occurring later than t_R for the shorter of the worldlines and a (different) kinetic focus time t'_Q occurring earlier than t_R for the other worldline. For the first worldline S is a minimum and for the other worldline S is a saddle point. Another example is the quartic oscillator discussed in Sec. IX, where an infinite number of actual worldlines can connect given terminal events (x_P, t_P) and (x_R, t_R), no matter how short the time interval $(t_R - t_P)$. Only for the shortest of these worldlines is S a minimum. The situation is different in 2D (see Appendix B).

[52]E. T. Whittaker, Ref. 36, pp. 250–253. Whittaker deals with the Maupertuis action W discussed in Appendix A, whereas we adapt his analysis to the Hamilton action S. In more detail, our Eq. (1) corresponds to the last equation on p. 251 of Ref. 36, with $q_1 \to t$, $q_2 \to x$ and $q'_2 \to \dot{x}$.

[53]L. I. Lolle, C. G. Gray, J. D. Poll, and A. G. Basile, "Improved short-time propagator for repulsive inverse power-law potentials," Chem. Phys. Lett. **177**, 64–72 (1991). In Sec. X and Ref. 54 further analytical and numerical results are given for the inverse-square potential, $U(x) = C/x^2$. For given end positions x_P and x_R, there are two actual worldlines $(x_P, t_P) \to (x_R, t_R)$ for given short times $(t_R - t_P)$. There is one actual worldline for $t_R = t_Q$ when the two worldlines have coalesced into one, and there is no actual worldline for longer times (remember x_P and x_R are fixed).

[54]A. G. Basile and C. G. Gray, "A relaxation algorithm for classical paths as a function of end points: Application to the semiclassical propagator for far-from-caustic and near-caustic conditions," J. Comput. Phys. **101**, 80–93 (1992).

[55]Better estimates can be found using the Sturm and Sturm-Liouville theories; see Papastavridis, Refs. 26 and 66.

[56]This argument can be refined. In Sec. 8 we show that $\delta^2 S$ becomes $\mathcal{O}(\alpha^3)$ for $R \to Q$ but is still larger in magnitude than the $\delta^3 S$ term, which is also $\mathcal{O}(\alpha^3)$. See Ref. 60 for $\delta^2 S$ and Eq. (40a) for $\delta^3 S$.

[57]There are other systems for which $\partial^3 U/\partial x^3$, etc., vanish, for example, $U(x) = C$, $U(x) = Cx$, $U(x) = -Cx^2$. For these systems the worldlines cannot have $\delta^2 S = 0$ [see Eq. (19)], so that kinetic foci do not exist.

[58]In higher dimensions more than one independent variation, occurring in different directions in function space, can occur due to symmetry. In Morse theory the number of these independent variations satisfying $\delta^2 S = 0$ is called the multiplicity of the kinetic focus (Van Brunt, Ref. 34, p. 254; Ref. 32, p. 90). If the mulitplicity is different from unity, Morse's theorem is modified from the statement in Ref. 34 to read as follows: The Morse index of the saddle point in action of worldline PR is equal to the

number of kinetic foci between P and R, with each kinetic focus counted with its multiplicity.

[59] J. M. T. Thompson and G. W. Hunt, *Elastic Instability Phenomena* (Wiley, Chichester, 1984), p. 20.

[60] The fact that $\delta^2 S_0 \to 0$ for $\alpha \to 0$ is not surprising because $\delta^2 S_0$ is proportional to α^2. The surprising fact is that here $\delta^2 S_0$ vanishes as α^3 as $\alpha \to 0$ because the integral involved in the definition, Eq. (36a), of $\delta^2 S_0$ is itself $\mathcal{O}(\alpha)$. We can see directly that $\delta^2 S_0$ becomes $\mathcal{O}(\alpha^3)$ for R near Q for this special variation $\alpha\phi$ by integrating the $\dot{\phi}$ term by parts in Eq. (36a) and using $\phi=0$ at the end-points. The result is $\delta^2 S_0 = -(\alpha^2/2)\int_P^R \phi[m\ddot{\phi} + U''(x_0)\phi]\,dt$. Because x_0 and $x_1 = x_0 + \alpha\phi$ are both true worldlines, we can apply the equation of motion $m\ddot{x} + U'(x) = 0$ to both. We then subtract these two equations of motion and expand $U'(x_0+\alpha\phi)$ as $U'(x_0) + U''(x_0)\alpha\phi + (\frac{1}{2})U'''(x_0)(\alpha\phi)^2 + \mathcal{O}(\alpha^3)$, giving $m\ddot{\phi} + U''(x_0)\phi = -(\frac{1}{2})U'''(x_0)\alpha\phi^2 + \mathcal{O}(\alpha^2)$. (If the ϕ^2, etc., nonlinear terms on the right-hand side are neglected in the last equation, it becomes the Jacobi-Poincaré linear variation equation used in stability studies.) If we use this result in the previous expression for $\delta^2 S_0$, we find to lowest nonvanishing order $\delta^2 S_0 = (\alpha^3/4)\int_P^R dt U'''(x_0)\phi^3$, which is $\mathcal{O}(\alpha^3)$. This result and Eq. (40a) for $\delta^3 S_0$ give the desired result (42) for $S_1 - S_0$.

[61] If we use arguments similar to those of this section and Sec. VI, we can show that $\delta^2 S$ vanishes again at the second kinetic focus Q_2, and that for R beyond Q_2 the wordline PR has a second, independent variation leading to $\delta^2 S < 0$, in agreement with Morse's general theory (Ref. 34).

[62] If we use $L(x,\dot{x}) = p\dot{x} - H(x,p)$, we can rewrite the Hamilton action as a phase-space integral, that is, $S = \int_P^R [p\dot{x} - H(x,p)]\,dt$. We set $\delta S = 0$ and vary $x(t)$ and $p(t)$ independently and find (Ref. 63) the Hamilton equations of motion $\dot{x} = \partial H/\partial p$, $\dot{p} = -\partial H/\partial x$. We can then show (Ref. 64) that in phase space, the trajectories $x(t)$, $p(t)$ that satisfy the Hamilton equations are always saddle points of S, that is, never a true maximum or a true minimum. In the proof it is assumed that H has the normal form $H(x,p) = p^2/2m + U(x)$.

[63] Reference 48, p. 353.

[64] M. R. Hestenes, "Elements of the calculus of variations," in *Modern Mathematics for the Engineer*, edited by E. F. Beckenbach (McGraw Hill, New York, 1956), pp. 59–91.

[65] O. Bottema, "Beisipiele zum Hamiltonschen Prinzip," Monatsh. Math. **66**, 97–104 (1962).

[66] J. G. Papastavridis, "On the extremal properties of Hamilton's action integral," J. Appl. Mech. **47**, 955–956 (1980).

[67] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 318. These authors use the Rayleigh-Ritz direct variational method (see Ref. 12 for a detailed discussion of this method) with a two-term trial trajectory $x(t) = a_1 \sin(\omega t/2) + a_2 \sin(\omega t)$, where $\omega = 2\pi/t_R$ and a_1 and a_2 are variational parameters, to study the half-cycle ($t_R = T_0/2$) and one-cycle ($t_R = T_0$) trajectories. Because the kinetic focus time $t_Q > T_0/2$ for this oscillator, they find, in agreement with our results, that S is a minimum for the half-cycle trajectory (with $a_1 \neq 0$, $a_2 = 0$) and a saddle point for the one-cycle trajectory (with $a_1 = 0$, $a_2 \neq 0$). However, in the figure accompanying their calculation, which shows the stationary points in (a_1, a_2) space, they label the origin $(a_1, a_2) = (0,0)$ a maximum. The point $(a_1, a_2) = (0,0)$ represents the equilibrium trajectory $x(t) = 0$. As we have seen, a true maximum in S cannot occur, so that other "directions" in function space not considered by the authors must give minimum-type behavior of S, leading to an overall saddle point.

[68] The two-incline oscillator potential has the form $U(x) = C|x|$, with $C = mg \sin\alpha \cos\alpha$ and α the angle of inclination. Here x is a horizontal direction. A detailed discussion of this oscillator is given by B. A. Sherwood, *Notes on Classical Mechanics* (Stipes, Champaign, Il, 1982), p. 157.

[69] Other constant force or linear potential systems include the 1D Coulomb model (Ref. 70) $U(x) = q_1 q_2 |x|$, the bouncing ball (Ref. 71), and for $x \geqslant 0$ the constant force spring (Ref. 72).

[70] I. R. Lapidus, "One- and two-dimensional Hydrogen atoms," Am. J. Phys. **49**, 807 (1981). K. Andrew and J. Supplee, "A Hydrogen atom in d-dimensions," Am. J. Phys. **58**, 1177–1183 (1990).

[71] I. R. Gatland, "Theory of a nonharmonic oscillator," Am. J. Phys. **59**, 155–158 (1991) and references therein; W. M. Hartmann, "The dynamically shifted oscillator," Am. J. Phys. **54**, 28–32 (1986).

[72] A. Capecelatro and L. Salzarulo, *Quantitative Physics for Scientists and Engineers: Mechanics* (Aurie Associates, Newark, NJ, 1977), p. 162; C.-Y. Wang and L. T. Watson, "Theory of the constant force spring," Trans. ASME, J. Appl. Mech. **47**, 956–958 (1980); H. Helm, "Comment on 'A constant force generator for the demonstration of Newton's second law'," Am. J. Phys. **52**, 268 (1984).

[73] C. G. Gray, G. Karl, and V. A. Novikov, "Progress in classical and quantum variational principles," Rep. Prog. Phys. **67**, 159–208 (2004).

[74] Nearly pure quartic potentials have been found in molecular physics for ring-puckering vibrational modes (Ref. 75) and for the caged motion of the potassium ion K^+ in the endohedral fullerene complex $K^+@C_{60}$ (Ref. 76), where the quadratic terms in the potential are small. Ferroelectric soft modes in solids are also sometimes approximately represented by quartic potentials (Refs. 77 and 78).

[75] R. P. Bell, "The occurrence and properties of molecular vibrations with $V(x) = ax^4$," Proc. R. Soc. London, Ser. A **183**, 328–337 (1945); J. Laane, "Origin of the ring-puckering potential energy function for four-membered rings and spiro compounds. A possibility of pseudorotation," J. Phys. Chem. **95**, 9246–9249 (1991).

[76] C. G. Joslin, J. Yang, C. G. Gray, S. Goldman, and J. D. Poll, "Infrared rotation and vibration-rotation bands of endohedral Fullerene complexes. $K^+@C_{60}$." Chem. Phys. Lett. **211**, 587–594 (1993).

[77] A. S. Barker, "Far infrared dispersion and the Raman spectra of ferroelectric crystals," in *Far-Infrared Properties of Solids*, edited by S. S. Mitra and S. Nudelman (Plenum, New York, 1970), pp. 247–296.

[78] J. Thomchick and J. P. McKelvey, "Anharmonic vibrations of an 'ideal' Hooke's law oscillator," Am. J. Phys. **46**, 40–45 (1978).

[79] R. Baierlein, Ref. 16, p. 73.

[80] In two dimensions with $\mathbf{x} = (x_1, x_2)$ and $\mathbf{v}_0 = (v_{01}, v_{02})$, our analytic condition (4) for the kinetic focus of worldline $\mathbf{x}(t, \mathbf{v}_0)$ becomes $\det(\partial x_i/\partial v_{0j}) = 0$, where $\det(A_{ij})$ denotes the determinant of matrix A_{ij}. The generalization to other dimensions is obvious. This condition (in slightly different form) is due to Mayer, Ref. 19, p. 269. For a clear discussion, see J. G. Papastavridis, *Analytical Mechanics* (Oxford U. P., Oxford, 2002), p. 1061. For multidimensions a caustic becomes in general a surface in space-time. The analogous theory for multidimensional spatial caustics, relevant for the action W, is discussed in Ref. 105. A simple example of a 2D surface spatial caustic is obtained by revolving the pattern of Fig. 7 about the vertical axis, thereby generating a paraboloid of revolution surface caustic/envelope. Due to axial symmetry, the caustic has a second (linear) branch, that is, the symmetry axis from $y=0$ to $y=Y$. An analogous optical example is discussed by M. V. Berry, "Singularities in waves and rays," in *Physics of Defects*, Les Houches Lectures XXXIV, edited by R. D. Balian, M. Kleman, and J.-P. Poirier (North Holland, Amsterdam, 1981), pp. 453–543.

[81] A dynamics problem can be formulated as an initial value problem. For example, find $x(t)$ from Newton's equation of motion with initial conditions $(x_P, \dot{x}_P)$. It can also be formulated as a boundary value problem; for example, find $x(t)$ from Hamilton's principle with boundary conditions (x_P, t_P) and (x_R, t_R). Solving a boundary value problem with initial value problem methods (for example, the shooting method) is standard (Ref. 82). Solving an initial value problem with boundary value problem methods is much less common (Ref. 83). For an example of a boundary value problem with mixed conditions (prescribed initial velocities and final positions) for about 10^7 particles, see A. Nusser and E. Branchini, "On the least action principle in cosmology," Mon. Not. R. Astron. Soc. **313**, 587–595 (2000).

[82] See, for example, W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge U. P., Cambridge, 1992), p. 749.

[83] H. R. Lewis and P. J. Kostelec, "The use of Hamilton's principle to derive time-advance algorithms for ordinary differential equations," Computer Phys. Commun. **96**, 129–151 (1996); D. Greenspan, "Approximate solution of initial value problems for ordinary differential equations by boundary value techniques," J. Math. Phys. Sci. **15**, 261–274 (1967).

[84] The converse effect cannot occur: a time-dependent potential $U(x,t)$ with $U'' < 0$ at all times always has $\delta^2 S > 0$ as seen from Eq. (19). If $U(x,t)$ is such that U'' alternates in sign with time, kinetic foci (and hence trajectory stability) may occur. An example is a pendulum with a rapidly vertically oscillating support point. In effect the gravitational field is oscillating. The pendulum can oscillate stably about the (normally unstable) upward vertical direction (Ref. 85). Two- and three-dimensional examples of this type are Paul traps (Ref. 86) and quadrupole mass filters (Ref. 85), which use oscillating quadrupole electric fields to trap ions. The equilibrium trajectory $\mathbf{x}(t) = 0$ at the center of the trap is unstable for purely electrostatic fields but is stabilized by using time-dependent electric fields. Focusing by alternating-gradients (also known as strong focus-

ing) in particle accelerators and storage rings is based on the same idea (Ref. 42).
[85]M. H. Friedman, J. E. Campana, L. Kelner, E. H. Seeliger, and A. L. Yergey, "The inverted pendulum: A mechanical analog of the quadrupole mass filter," Am. J. Phys. **50**, 924–931 (1982).
[86]P. K. Gosh, *Ion Traps* (Oxford U. P., Oxford, 1995), p. 7.
[87]J. J. Stoker, *Nonlinear Vibrations in Mechanical and Electrical Systems* (Wiley, New York, 1950), p. 112. Stoker's statements on series convergence need amendment in light of the Kolmogorov-Arnold-Moser (KAM) theory (Ref. 89). See J. Moser, "Combination tones for Duffing's equation," Commun. Pure Appl. Math. **18**, 167–181 (1965); T. Kapitaniak, J. Awrejcewicz and W.-H. Steeb, "Chaotic behaviour in an anharmonic oscillator with almost periodic excitation," J. Phys. A **20**, L355–L358 (1987); A. H. Nayfeh, *Introduction to Perturbation Techniques* (Wiley, New York, 1981), p. 216; A. H. Nayfeh and B. Balachandran, *Applied Nonlinear Dynamics* (Wiley, New York, 1995), p. 234; S. Wiggins, "Chaos in the quasiperiodically forced Duffing oscillator," Phys. Lett. A **124**, 138–142 (1987).
[88]G. Seifert, "On almost periodic solutions for undamped systems with almost periodic forcing," Proc. Am. Math. Soc. **31**, 104–108 (1972); J. Moser, "Perturbation theory of quasiperiodic solutions and differential equations," in *Bifurcation Theory and Nonlinear Eigenvalue Problems*, edited by J. B. Keller and S. Antman (Benjamin, New York, 1969), pp. 283–308; J. Moser, "Perturbation theory for almost periodic solutions for undamped nonlinear differential equations," in *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics*, edited by J. P. Lasalle and S. Lefschetz (Academic, New York, 1963), pp. 71–79; M. S. Berger, "Two new approaches to large amplitude quasi-periodic motions of certain nonlinear Hamiltonian systems," Contemp. Math. **108**, 11–18 (1990).
[89]G. M. Zaslavsky, R. Z. Sagdeev, D. A. Usikov, and A. A. Chernikov, *Weak Chaos and QuasiRegular Patterns* (Cambridge U. P., Cambridge, 1991), p. 30.
[90]See, for example, M. Tabor, *Chaos and Integrability in Nonlinear Dynamics* (Wiley, New York, 1989), p. 35; J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos*, 2nd ed. (Wiley, Chichester, 2002), pp. 310.
[91]For example, the equilibrium position can be modulated. A somewhat similar system is a ball bouncing on a vertically oscillating table. The motion can be chaotic. See, for example, J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1983), p. 102; N. B. Tufillaro, T. Abbott, and J. Reilly, *An Experimental Approach to Nonlinear Dynamics and Chaos* (Addison-Wesley, Redwood City, CA, 1992), p. 23; A. B. Pippard, *The Physics of Vibration* (Cambridge U. P., Cambridge, 1978), Vol. 1, pp. 253, 271.
[92]The forced Duffing oscillator with $U(x,t)=(\frac{1}{2})kx^2+Cx^4-xF_0\cos\omega t$ is studied in Ref. 30. For $k=0$ the Duffing oscillator reduces to the quartic oscillator.
[93]R. H. G. Helleman, "Variational solutions of non-integrable systems," in *Topics in Nonlinear Dynamics*, edited by S. Jorna (AIP, New York, 1978), pp. 264–285. This author studies the forced Duffing oscillator with $U(x,t)=(\frac{1}{2})kx^2-Cx^4-xF_0\cos\omega t$ (note the sign change in C compared to Ref. 92), and the Henon-Heiles oscillator with the potential in Eq. (83).
[94]In Ref. 54 the harmonic potential $U(x,t)=(\frac{1}{2})k[x-x_c(t)]^2$ with an oscillating equilibrium position $x_c(t)$ is studied. The worldlines for this system are all nonchaotic.
[95]M. Henon and C. Heiles, "The applicability of the third integral of the motion: Some numerical experiments," Astron. J. **69**, 73–79 (1964).
[96]There have been a few formal studies of action for chaotic systems, but few concrete examples seem to be available. See, for example, S. Bolotin, "Variational criteria for nonintegrability and chaos in Hamiltonian systems," in *Hamiltonian Mechanics*, edited by J. Seimenis (Plenum, New York, 1994), pp. 173–179.
[97]Reference 48, p. 434.
[98]H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Gauthier-Villars, Paris, 1899), Vol. 3; *New Methods of Celestial Mechanics* (AIP, New York, 1993), Part 3, p. 958.
[99]The situation is complicated because, as Eq. (A1) shows, there are two forms for W, that is, the time-independent (first) form and the time-dependent (last) form. Spatial kinetic foci (discussed in Appendix B) occur for the time-independent form of W. Space-time kinetic foci occur for the time-dependent form of W, as for S. Typically the kinetic foci for the two forms for W differ from each other (Refs. 30 and 98) and from those for S.
[100]E. J. Routh, *A Treatise on Dynamics of a Particle* (Cambridge U. P., Cambridge, 1898), reprinted (Dover, New York, 1960), p. 400.
[101]A. P. French, "The envelopes of some families of fixed-energy trajectories," Am. J. Phys. **61**, 805–811 (1993); E. I. Butikov, "Families of Keplerian orbits," Eur. J. Phys. **24**, 175–183 (2003).
[102]We assume that we are dealing with bound orbits. Similar comments apply to scattering orbits (hyperbolas). Just as for the orbits in the linear gravitational potential discussed in the preceding paragraph, here too there are restrictions and special cases (Ref. 19, p. 164; Ref. 103, p. 122). If the second point (x_R,y_R) lies within the "ellipse of safety" (the envelope (French, Ref. 101)) of the elliptical trajectories of energy E originating at (x_P,y_P), then two ellipses with energy E can connect (x_P,y_P) to (x_R,y_R). If (x_R,y_R) lies on the ellipse of safety, then one ellipse of energy E can connect (x_P,y_P) to (x_R,y_R), and if (x_R,y_R) lies outside the ellipse of safety, then no ellipse of energy E can connect the two points. Usually the initial and final points (x_P,y_P) and (x_R,y_R) together with the center of force at (0,0) (one focus of the elliptical path) define the plane of the orbit. If (x_P,y_P), (x_R,y_R), and (0,0) lie on a straight line, the plane of the orbit is not uniquely defined, and there is almost always an infinite number of paths of energy E in three dimensions that can connect $(x_P,y_P,z_P=0)$ to $(x_R,y_R,z_R=0)$. A particular case of the latter is a periodic orbit where $(x_R,y_R)=(x_P,y_P)$. Because the orbit can now be brought into coalescence with an alternative true orbit by a rotation around the line joining (x_P,y_P) to (0,0), a third kinetic focus arises for elliptical periodic orbits in three dimensions (see Ref. 33, p. 29).
[103]N. G. Chetaev, *Theoretical Mechanics* (Springer, Berlin, 1989).
[104]For the repulsive $1/r$ potential, the hyperbolic spatial orbits have a (parabolic shaped) caustic/envelope (French, Ref. 101).
[105]If the orbit equation has the explicit form $y=y(x,\theta_0)$, or the implicit form $f(x,y,\theta_0)=0$, the spatial kinetic focus is found from $\partial y/\partial\theta_0=0$ or $\partial f/\partial\theta_0=0$, respectively. Here θ_0 is the launch angle at (x_P,y_P) (see Fig. 15 for an example). The derivation of these spatial kinetic focus conditions is similar to the derivation of the space-time kinetic focus condition of Eq. (4) (see Ref. 106, p. 59). In contrast, if the orbit equation is defined parametrically by the trajectory equations $x=x(t,\theta_0)$ and $y=y(t,\theta_0)$, the spatial kinetic focus condition is $\partial(x,y)/\partial(t,\theta_0)=0$. This Jacobian determinant condition is similar to that of Ref. 80 for the space-time kinetic focus (see Ref. 106, p. 73 for a derivation). As an example, consider a family of figure-eight-like harmonic oscillator orbits of Fig. 16, launched from the origin $(x_P,y_P)=(0,0)$ at time $t_P=0$, all with speed v_0 (and therefore the same energy E), at various angles θ_0. The trajectory equations are $x=(v_0/\omega_1)\cos\theta_0\sin\omega_1 t$ and $y=(v_0/\omega_2)\sin\theta_0\sin\omega_2 t$, where $\omega_1=2\omega_2$. The determinant condition for the spatial kinetic focus reduces to $\cos\theta_0\tan\omega_2 t=1$, which locates the kinetic focus (in time) for the orbit with launch angle θ_0. Elimination of θ_0 and t from these three equations leads to the locus of the (first) spatial kinetic foci, the spatial caustic/envelope equation $y^2=(v_0/\omega_2)^2-2(v_0/\omega_2)|x|$, which is a parabolic shaped curve with two cusps on the y-axis.
[106]R. H. Fowler, *The Elementary Differential Geometry of Plane Curves* (Cambridge U. P., Cambridge, 1920).
[107]The finding of the two elliptical (or hyperbolic or parabolic) shaped trajectories from observations giving the two end-positions and the time interval is a famous problem of astronomy and celestial mechanics, solved by Lambert (1761), Gauss (1801–1809), and others (Ref. 108).
[108]R. R. Bate, D. D. Mueller, and J. E. White, *Fundamentals of Astrodynamics* (Dover, New York, 1971), p. 227; H. Pollard, *Celestial Mechanics* (Mathematical Association of America, Washington, 1976), p. 28; P. R. Escobal, *Methods of Orbit Determination* (Wiley, New York, 1965), p. 187. For the elliptical orbits, more than two trajectories typically become possible at sufficiently large time intervals; these additional trajectories correspond to more than one complete revolution along the orbit (Ref. 109).
[109]R. H. Gooding, "A procedure for the solution of Lambert's orbital boundary-value problem," Celest. Mech. Dyn. Astron. **48**, 145–165 (1990).
[110]It is clear from Fig. 13 that a kinetic focus occurs after time T_0. To show rigorously that this focus is the first kinetic focus (unlike for W where it is the second), we can use a result of Gordon (Ref. 111) that the action S is a minimum for time $t=T_0$. If one revolution corresponds to the second kinetic focus, the trajectory $P\rightarrow P$ would correspond to a saddle point. The result $t_Q=T_0$ can also be obtained algebraically by applying the general relation (4) to the relation $r=r(t,L)$ for the radial distance, where we

use angular momentum L as the parameter labeling the various members of the family in Fig. 13. We obtain t_Q from $(\partial r/\partial L)_t=0$. The latter equation implies that $(\partial t/\partial L)_r=0$, because $(\partial r/\partial L)_t=-(\partial r/\partial t)_L(\partial t/\partial L)_r$. At fixed energy E (or fixed major axis $2a$), the period T_0 is independent of L for the attractive $1/r$ potential, so that the solution of $(\partial t/\partial L)_r=0$ occurs for $t=T_0$, which is therefore the kinetic focus time t_Q.

[111] W. B. Gordon, "A minimizing property of Keplerian orbits," Am. J. Math. **99**, 961–971 (1977).

[112] Note that for the actual 2D trajectories in the potential $U(x,y)=mgy$, kinetic foci exist for the spatial paths of the Maupertuis action W, but do not exist for the space-time trajectories of the Hamilton action S. This result illustrates the general result stated in Appendix A that the kinetic foci for W and S differ in general.

[113] A. P. French, *Vibrations and Waves* (Norton, New York, 1966), p. 36.

[114] J. C. Slater and N. H. Frank, *Introduction to Theoretical Physics* (McGraw Hill, New York, 1933), p. 85.

[115] C. G. Gray, G. Karl, and V. A. Novikov, "The four variational principles of mechanics," Ann. Phys. (N.Y.) **251**, 1–25 (1996).

[116] C. G. Gray, G. Karl, and V. A. Novikov, "From Maupertuis to Schrödinger. Quantization of classical variational principles," Am. J. Phys. **67**, 959–961 (1999).

[117] E. Schrödinger, "Quantisierung als eigenwert problem I," Ann. Phys. **79**, 361–376 (1926), translated in E. Schrödinger, *Collected Papers on Wave Mechanics* (Blackie, London, 1928), Chelsea reprint 1982. For modern discussions and applications, see, for example, E. Merzbacher, *Quantum Mechanics*, 3rd ed. (Wiley, New York, 1998), p. 135; S. T. Epstein, *The Variation Method in Quantum Chemistry* (Academic, New York, 1974).