1

Variational approach to a theory of

CLASSICAL PARTICLE TRAJECTORIES ‡

Introduction. The problem central to the classical mechanics of a particle is usually construed to be to discover the function $\boldsymbol{x}(t)$ that describes—relative to an inertial Cartesian reference frame—the positions assumed by the particle at successive times t. This is the problem addressed by Newton, according to whom our analytical task is to discover the solution of the differential equation

$$m\frac{d^2\boldsymbol{x}(t)}{dt^2} = \boldsymbol{F}(\boldsymbol{x}(t))$$

that conforms to prescribed initial data $\boldsymbol{x}(0) = \boldsymbol{x}_0$, $\dot{\boldsymbol{x}}(0) = \boldsymbol{v}_0$. Here I explore an alternative approach to the same physical problem, which we cleave into two parts: we look first for the *trajectory* traced by the particle, and then—as a separate exercise—for its *rate of progress along that trajectory*. The discussion will cast new light on (among other things) an important but frequently misinterpreted variational principle, and upon a curious relationship between the "motion of particles" and the "motion of photons"—the one being, when you think about it, hardly more abstract than the other.

‡ The following material is based upon notes from a Reed College Physics Seminar "Geometrical Mechanics: Remarks commemorative of Heinrich Hertz" that was presented 23 February 1994.

1. “Transit time” in 1-dimensional mechanics. To describe (relative to an inertial frame) the 1-dimensional motion of a mass point m we were taught by Newton to write

$$m\ddot{x} = F(x)$$

If $F(x)$ is “conservative” $\qquad F(x) = -\frac{d}{dx}U(x)$

(which in the 1-dimensional case is automatic) then, by a familiar line of argument,

$$E \equiv \tfrac{1}{2}m\dot{x}^2 + U(x) \quad \text{is conserved:} \quad \dot{E} = 0$$

Therefore the speed of the particle when at x can be described

$$v(x) = \sqrt{\tfrac{2}{m}\big[E - U(x)\big]} \tag{1}$$

and is determined (see the FIGURE 1) by the “local depth $E - U(x)$ of the potential lake.” Several useful conclusions are immediate. The motion of m is

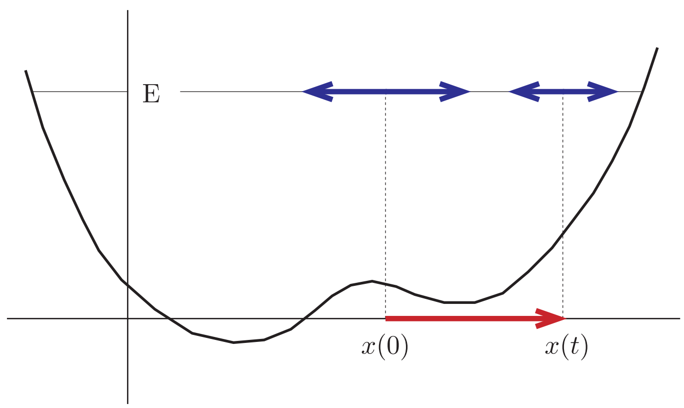

FIGURE 1: *“Potential lake” in 1-dimensional mechanics.*

bounded $a \le x \le b$ by “turning points” a and b where the potential lake has vanishing depth, and excluded from “forbidden regions” where $E - U(x) < 0$; *i.e.*, where the potential has risen above the “lake level” E. And the dynamical time of flight, or “transit time” $x_0 \longrightarrow x$ can be described

$$t(x; x_0, E) = \int_{x_0}^{x} \frac{1}{\sqrt{\frac{2}{m}\big[E - U(y)\big]}}\,dy \tag{2}$$

By functional inversion—when it can be accomplished!—one obtains

$$x = x(t; x_0, E)$$

which provides an explicit description of the E-conserving motion of the particle.

Greater interest attaches, however, to (because they are less familiar) some of the *non*-standard applications/generalizations of (2). For example: functions of the type

$$T(x,p)=\int^x \frac{1}{\sqrt{\frac{2}{m}\big[H(x,p)-U(y)\big]}}\,dy$$

comprise seldom-remarked "natural companions" of the Hamiltonian

$$H(x,p)=\tfrac{1}{2m}p^2+U(x)$$

Indeed, $H(x,p)$ and $T(x,p)$ are "conjugate observables" in the sense that

$$[\,T,H\,]\equiv \tfrac{\partial T}{\partial x}\tfrac{\partial H}{\partial p}-\tfrac{\partial H}{\partial x}\tfrac{\partial T}{\partial p}=1$$

It is by quantization of the Hamiltonian $H(x,p)\longrightarrow \mathsf{H}$ that one prepares to write the Schrödinger equation $\mathsf{H}\psi=i\hbar\partial_t\psi$. The same procedure $T(x,p)\longrightarrow \mathsf{T}$ yields a highly non-standard object: a "time operator," which is conjugate to H in the sense standard to quantum mechanics

$$[\,\mathsf{T},\mathsf{H}\,]\equiv \mathsf{TH}-\mathsf{HT}=i\hbar\,\mathsf{I}$$

Upon this remark hangs a tale which I may tell on some other occasion. It is, however, by *dimensional* generalization of (2) that we are led to the subject matter of present interest.

2. "Transit time" in N-dimensional mechanics. To describe, relative to an inertial frame, the N-dimensional motion[1] of m we write

$$m\ddot{\boldsymbol{x}}=\boldsymbol{F}(\boldsymbol{x})$$

Now

$$\boldsymbol{F}(\boldsymbol{x})=-\boldsymbol{\nabla}U(\boldsymbol{x})$$

is no longer automatic; it is mathematically rare but physically commonplace. It is, in all events, critical to energy conservation

$$\dot{E}=0 \quad\text{with}\quad E\equiv\tfrac{1}{2}mv^2+U(\boldsymbol{x})$$

and will be assumed. Speed v is, as before, determined by the local depth of the potential lake

$$v(\boldsymbol{x})=\sqrt{\tfrac{2}{m}\big[E-U(\boldsymbol{x})\big]} \tag{3}$$

but now the potential lake is (see FIGURE 2) a much more "lake-like" place.

[1] We have immediate interest in the cases $N=2$ and $N=3$, but if we were concerned with the motion of (say) a *pair* of particles we would want to set $N=6$.

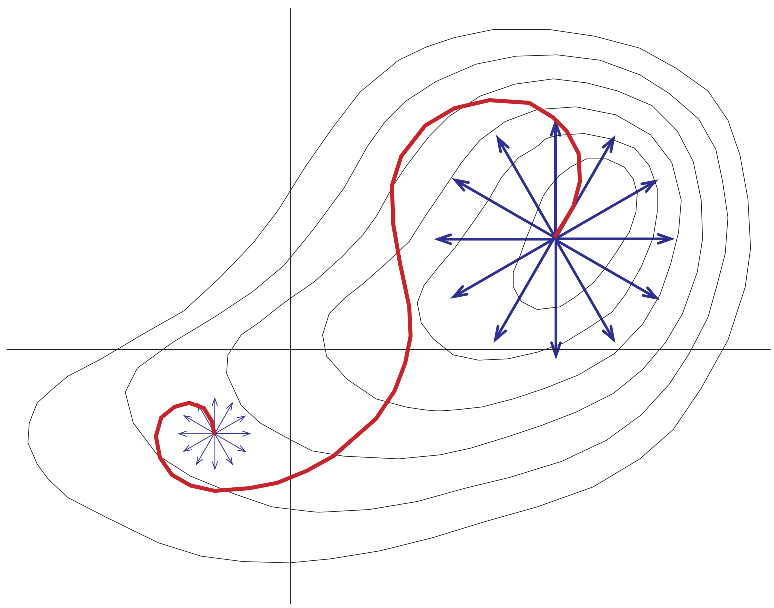

FIGURE 2: *"Topographic map" of a 2-dimensional potential lake. The light curves are equipotentials. The curve which links the points* $\boldsymbol{x}(0) \longrightarrow \boldsymbol{x}(s)$ *has Euclidean arc length* s*.*

It is, in particular, a place where *speed—since it conveys no directional information—is insufficient to determine velocity*.

To every "path" inscribed "on the surface of the E-lake" (*i.e.*, within the region $\mathcal{L}$ bounded by the equipotential $U(\boldsymbol{x}) = E$) we can associate a "transit time" T[path]. To notate this obvious fact it is convenient to adopt arc-length parameterization $ds^2 = dx^2 + dy^2$, writing

$$T[\boldsymbol{x}(s)] = \int_0^s \frac{1}{v(\boldsymbol{x}(s'))}\, ds' \tag{4}$$

Consider now the population $\mathcal{P}$ of paths (of various lengths) $\boldsymbol{a} \longrightarrow \boldsymbol{b}$, as illustrated in FIGURE 3 . We have particular interest in the "dynamical" elements of such populations $\mathcal{P}$, *i.e.*, in the paths which the E-conserving motion of m would trace out in time. More specifically, we have interest in the answer to this question: *Is there a property of* T*[path] that serves to distinguish dynamical paths from paths-in-general?* The question springs naturally to the imagination of anyone passingly familiar with the variational principles of mechanics, particularly because it is so evocative of . . .

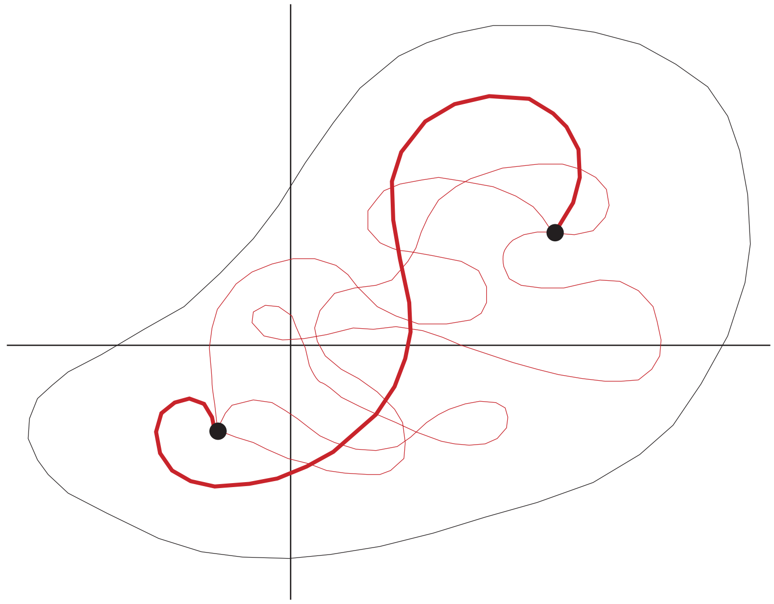

FIGURE 3: *Population $\mathcal{P}$ of curves inscribed on the surface of an E-lake. The distinguished curve was traced by a mass point m; it is "dynamically necessary," the others "dynamically impossible."*

3. Fermat's Principle of Least Time. A satisfactory geometrical optics—a theory of *light rays in isotropically inhomogeneous media* can, whether one proceeds from Maxwellian electrodynamics or in the more phenomenological language of Pierre Fermat, be constructed as follows:

1) To each point $\boldsymbol{x}$ in the medium assign a "speed function"

$$v(\boldsymbol{x}) = \frac{c}{n(\boldsymbol{x})}$$

where $n(\boldsymbol{x})$ is the local "index of refraction";

2) To each hypothetical path $\boldsymbol{x}(s)$ associate a number-valued path functional

$$\begin{aligned} T[\text{path}] &= \text{“transit time”} \\ &= \int_0^s \frac{1}{v(\boldsymbol{x}(s'))}\,ds' = \tfrac{1}{c}\cdot\underbrace{\int_0^s n(\boldsymbol{x}(s'))\,ds'} \\ &\qquad\qquad\qquad\qquad \equiv \text{“optical path length”} \end{aligned}$$

3) Associate optical "rays" with the paths which extremize (or as the informal phrase goes, which "minimize") optical path length.

Turning now to the analytical implementation of Fermat's Principle, it proves convenient (to avoid a certain technical complication, as discussed below) to give up the specialness of s-parameterization in favor of unspecialized/arbitrary λ-parameterization, writing $\boldsymbol{x}(\lambda)$ to describe a path. Then, using $ds = \sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\,d\lambda$ with $\mathring{\boldsymbol{x}} \equiv \frac{d}{d\lambda}\boldsymbol{x}$, we have

$$T[\boldsymbol{x}(\lambda)] = \tfrac{1}{c}\int n(\boldsymbol{x})\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\,d\lambda$$

Then methods standard to the calculus of variations proceed from

$$\text{FERMAT'S PRINCIPLE:}\quad \delta T[\boldsymbol{x}(\lambda)] = 0$$

to the statement

$$\left\{\frac{d}{d\lambda}\frac{\partial}{\partial\mathring{\boldsymbol{x}}} - \frac{\partial}{\partial\boldsymbol{x}}\right\} n(\boldsymbol{x})\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = \boldsymbol{0}$$

Thus are we led to the so-called "ray equations"

$$\frac{d}{d\lambda}\left[n\,\frac{\mathring{\boldsymbol{x}}}{\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}}\right] - \sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\;\boldsymbol{\nabla} n = \boldsymbol{0} \tag{5}$$

Several remarks are now in order:

If at this point we were to revert $s \longleftarrow \lambda$ to arc-length parameterization then $ds = d\lambda$ entails $\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = 1$ and from (5) we obtain

$$\frac{d}{ds}\left[n\,\frac{d\boldsymbol{x}}{ds}\right] - \boldsymbol{\nabla} n = \boldsymbol{0} \tag{6}$$

which in homogeneous media (where $\boldsymbol{\nabla} n = \boldsymbol{0}$) reduces to $d^2\boldsymbol{x}/ds^2 = \boldsymbol{0}$: *rays become straight* (in the Euclidean sense).

Curiously, equation (6) is not itself derivable (except by trickery) from a "Lagrangian." To retain access to the Lagrangian method after adoption of s-parameterization one must treat $\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = 1$ (here $\mathring{\boldsymbol{x}} \equiv \frac{d}{ds}\boldsymbol{x}$) as a "constraint," writing

$$\delta\int\left\{n(\boldsymbol{x}) + \tfrac{1}{2}\lambda\,[\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}} - 1]\right\} = 0$$

where $\lambda(s)$ has now the nature of a "Lagrange multiplier" which has joined $\boldsymbol{x}(s)$ on the list of unknown functions of s which it would be our business to describe.

Multiply $n(\boldsymbol{x})$ into (6) and obtain $\big(n\frac{d}{ds}\big)^2\boldsymbol{x} = \frac{1}{2}\boldsymbol{\nabla} n^2$. Write $n\frac{d}{ds} = \frac{d}{du}$; *i.e.*, give up s-parameterization in favor of u-parameterization, with

$$u(s) = \int^s \frac{1}{n(\boldsymbol{x}(s'))}\,ds'$$

Then (6) reads

$$\left(\frac{d}{du}\right)^2\boldsymbol{x} = \tfrac{1}{2}\boldsymbol{\nabla} n^2$$

which looks very "Newtonian," and can be considered to arise from the following "Lagrangian":

$$L = \tfrac{1}{2}\frac{d\boldsymbol{x}}{du}\cdot\frac{d\boldsymbol{x}}{du} + \tfrac{1}{2}n^2$$

Though Fermat spoke casually of "least time," what he actually gave us is a static *theory of curves*, in which "rays" are distinguished from other curves by their *least optical length*. It is, I think, well to remind ourselves that Fermat wrote in 1657, almost twenty years before Olaf Römer—in 1676, eleven years after Fermat's death—first demonstrated the speed of light to be $c < \infty$. We find it so utterly natural to think of the index of refraction as having to do with the ratio of two speeds

$$\text{index of refraction} = \frac{\text{speed of light in vacuum}}{\text{speed of light in medium}}$$

that we are astonished by the realization that neither Snell, nor Descarte, nor Fermat were in position to entertain the physical imagery that attaches to such a notion. Nothing actually *moved* in optics—I set aside the Newtonian fiction of "corpusles in flight"—until the invention of the dynamical wave theory of light (foreshadowed in 1678 by Huygens), where "rays" arise as "curves normal to surfaces of constant phase," and the "things" which literally move along "rays" are no more "physical" than mere points of intersection!

In mechanics, on the other hand, we confront the "real" motion of (idealized) "real things": mass points. I return now to the mechanical discussion where we left in on page 5, asking . . .

4. Does there exist a mechanical analog of Fermat's Principle? Such a theory, if it existed, would refer presumably to the *geometry of the space curves* $\boldsymbol{x}(s)$ that are traced out (in time) by m. This *separately from all properly "dynamical" matters*, which can be considered to reside in the structure of the function $s(t)$. We adopt, therefore, this non-standard point of view:

$$\begin{array}{ll}
\boldsymbol{X}(s) & \text{describes the "trajectory" of } m\text{, the mechanical analog of a "ray"} \\
\quad\uparrow & \\
\quad s(t) & \text{describes } \textit{motion along} \text{ that trajectory} \\
\downarrow & \\
\boldsymbol{x}(t) \equiv \boldsymbol{X}(s(t)) & \text{is the central object of Newtonian mechanics}
\end{array}$$

To get off the ground we must recall some aspects of the mathematical theory of space curves. Let $\boldsymbol{X}(s)$ serve, relative to a Cartesian frame, to provide the s-parameterized description of such a curve $\mathcal{C}$. Then $\boldsymbol{T}(s) \equiv \mathring{\boldsymbol{X}}(s)$ describes the unit tangent to $\mathcal{C}$ at s. $\mathring{\boldsymbol{T}}(s)$ lies normal to $\boldsymbol{T}(s)$ and in the plane in which $\mathcal{C}$ is locally inscribed. The equation $\mathring{\boldsymbol{T}}(s) = \kappa(s)\boldsymbol{U}(s)$ serves (with $\boldsymbol{U}(s)$ a unit vector) to describe both the direction $\boldsymbol{U}(s)$ and the magnitude $\kappa(s)$ of the local *curvature* of $\mathcal{C}$. Assume $\kappa(s) \neq 0$ and define $\boldsymbol{V}(s) \equiv \boldsymbol{T}(s)\times\boldsymbol{U}(s)$ which serves to

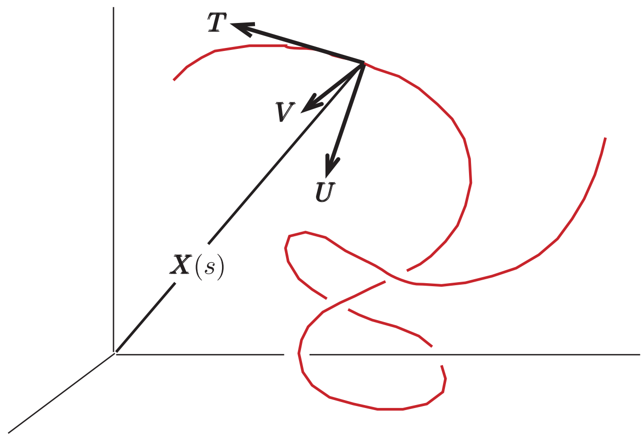

FIGURE 4: *Vectors natural to the description of a space curve* $\mathcal{C}$.

complete the construction of an orthonormal triad at each (non-straight) point s of $\mathcal{C}$. Elementary arguments lead to the conclusions that $\mathring{\boldsymbol{U}} = -\kappa\,\boldsymbol{T} - \tau\boldsymbol{V}$ and $\mathring{\boldsymbol{V}}(s) = \tau\boldsymbol{U}$, where $\tau(s)$ is the *torsion* of $\mathcal{C}$ at s. Briefly

$$\begin{pmatrix} \mathring{\boldsymbol{T}} \\ \mathring{\boldsymbol{U}} \\ \mathring{\boldsymbol{V}} \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{T} \\ \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}$$

which comprise the famous "Frenet-Serret formulae" (1847–1851).[2]

Turning in this language to the subject matter of elementary kinematics, we have

$$\begin{aligned} \dot{\boldsymbol{x}} &= \dot{s}\,\mathring{\boldsymbol{X}} \\ &= \dot{s}\,\boldsymbol{T} \end{aligned}$$

$$\begin{aligned} \ddot{\boldsymbol{x}} &= \ddot{s}\,\boldsymbol{T} + \dot{s}^2\,\mathring{\boldsymbol{T}} \\ &= \ddot{s}\,\boldsymbol{T} + \underbrace{\dot{s}^2\kappa}\,\boldsymbol{U} \end{aligned}$$

$\kappa = 1/R$ with $R =$ "radius of curvature," so $\dot{s}^2\kappa$ is precisely the v^2/R familiar from the elementary theory of uniform circular motion

$$\dddot{\boldsymbol{x}} = (\dddot{s} - \dot{s}^3\kappa^2)\,\boldsymbol{T} + (3\dot{s}\ddot{s}\kappa + \dot{s}^3\mathring{\kappa})\,\boldsymbol{U} - \dot{s}^3\kappa\tau\,\boldsymbol{V}$$

$$\vdots$$

[2] Multi-particle systems can be considered to trace curves in hyperspace, in which connection see my "Frenet-Serret formulæ in higher dimension" (1998).

which we now use to construct a "Newtonian theory of dynamical trajectories." To describe the conservative *motion* of m let us write

$$\ddot{\boldsymbol{x}}(t) = \boldsymbol{G}(\boldsymbol{x}(t))$$

with

$$\boldsymbol{G}(\boldsymbol{x}) \equiv -\tfrac{1}{m}\boldsymbol{\nabla} U(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{\nabla}\Big\{\tfrac{2}{m}\big[E - U(\boldsymbol{x})\big]\Big\}$$

which we are now in position to formulate

$$\dot{s}^2\,\mathring{\boldsymbol{T}} + \ddot{s}\,\boldsymbol{T} = \boldsymbol{G} \tag{7}$$

All reference to the specifically *motional* aspects of the situation resides here in the factors $\dot{s}^2$ and $\ddot{s}$, which we now eliminate to obtain a "theory of trajectories" as a kind of residue. To that end, we note first that by energy conservation

$$\dot{s}^2 = \tfrac{2}{m}\big[E - U(\boldsymbol{x})\big]$$

Also by $\mathring{\boldsymbol{T}} \perp \boldsymbol{T}$ and $\boldsymbol{T}\cdot\boldsymbol{T} = 1$

$$\ddot{s} = \boldsymbol{T}\cdot\boldsymbol{G} = \tfrac{1}{2}\boldsymbol{T}\cdot\boldsymbol{\nabla}\Big\{\tfrac{2}{m}\big[E - U(\boldsymbol{x})\big]\Big\}$$

Partly to reduce notational clutter, but mainly to facilitate comparison with our optical experience, we agree to write

$$\tfrac{2}{m}\big[E - U(\boldsymbol{x})\big] \equiv v^2(\boldsymbol{x}) \equiv \left[\frac{c}{n(\boldsymbol{x}; E)}\right]^2$$

where the c has been introduced from dimensional necessity but immediately drops away, and where it becomes natural to adopt the terminology

$$n(\boldsymbol{x}; E) \equiv \text{ the "}\textit{mechanical}\text{ index of refraction"}$$

The energy-dependence of the mechanical index of refraction is reminiscent of the typical frequency-dependence of its optical analog.

Returning with this information and in this notation to (7) we have

$$\tfrac{1}{n^2}\mathring{\boldsymbol{T}} + \tfrac{1}{2}\big(\boldsymbol{T}\cdot\boldsymbol{\nabla}\tfrac{1}{n^2}\big)\boldsymbol{T} = \underbrace{\tfrac{1}{2}\boldsymbol{\nabla}\tfrac{1}{n^2}}_{=\,\frac{1}{n}\boldsymbol{\nabla}\frac{1}{n}}$$

But

so

$$\tfrac{1}{n}\mathring{\boldsymbol{T}} + \big(\boldsymbol{T}\cdot\boldsymbol{\nabla}\tfrac{1}{n}\big)\boldsymbol{T} = \boldsymbol{\nabla}\tfrac{1}{n}$$

of which

$$\frac{d}{ds}\left[\frac{1}{n}\frac{d\boldsymbol{X}}{ds}\right] - \boldsymbol{\nabla}\frac{1}{n} = \boldsymbol{0} \tag{8}$$

is but a notational variant.

Equation (8) is the "trajectory equation" of Newtonian dynamics. It describes the "design of the tracks" along which m is permitted to move with conserved energy E. To describe the particle's *progress along* such a track $\mathcal{C}_E$ we can appeal to the transit time formalism, writing

$$
\begin{aligned}
t(s) &\equiv \text{transit time } \boldsymbol{X}(0) \longrightarrow \boldsymbol{X}(s) \text{ along } \mathcal{C}_E\\
&= \int_0^s \frac{1}{\sqrt{\frac{2}{m}\big[E - U(\boldsymbol{X}(s'))\big]}}\,ds' = \tfrac{1}{c}\int_0^s n(\boldsymbol{X}(s');E)\,ds'\\
&\downarrow\\
s &= s(t) \quad \text{by functional inversion}
\end{aligned}
$$

We anticipate that there will be occasions when it is the intractability of the functional inversion that prevents our progressing from the trajectory to an explicit description of the motion—occasions, that is to say, when it is relatively easier to solve (8) than it is to solve the associated equations of motion.

The trajectory equation (8) provides the foundation of what might be called "**time-independent Newtonian dynamics**." Interestingly, the phrase is much less familiar than the "time-independent Hamilton-Jacobi equation" and the "time-independent Schrödinger equation" which it calls instantly to mind. Nor are we speaking here of a merely terminological resonance; there exists a sense—which I hope to detail on some other occasion—in which the former subject lies at the theoretical base of the latter two.

5. Variational formulation of time-independent Newtonian mechanics. Equation (8)—the "trajectory equation"—is structurally identical to the "ray equation" (6), from which however it differs in one important respect, which can be symbolized

$$
n(\boldsymbol{x},\nu)_{\text{optical}} \longrightarrow \frac{1}{n(\boldsymbol{x};E)_{\text{mechanical}}} \tag{9}
$$

We have seen that the ray equation can be obtained by specialization $s \longleftarrow \lambda$ of the arbitrarily parameterized Euler-Lagrange equation (5) which issues from the variational principle $\delta\int n(\boldsymbol{x},\nu)\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\,d\lambda = 0$. Similarly,

$$
\delta\int \frac{1}{n(\boldsymbol{x};E)}\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\,d\lambda = 0 \tag{10}
$$

gives

$$
\left\{\frac{d}{d\lambda}\frac{\partial}{\partial\mathring{\boldsymbol{x}}} - \frac{\partial}{\partial\boldsymbol{x}}\right\}\frac{1}{n(\boldsymbol{x};E)}\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = \boldsymbol{0}
$$

whence

$$
\frac{d}{d\lambda}\left[\frac{1}{n}\frac{\mathring{\boldsymbol{x}}}{\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}}\right] - \sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}\,\boldsymbol{\nabla}\frac{1}{n} = \boldsymbol{0} \tag{11}
$$

from which by $s \longleftarrow \lambda$ we recover the trajectory equation (8).

Suppose were were to agree—at high risk of the confusion that I want here to dispel—to notate the variational principle (10) as follows:

$$
\hat{\delta}\int \frac{1}{n}\,ds = 0 \tag{12}
$$

where the ˆ is understood to signify that the variation is to be carried out subject to the constraint $\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = 1$. Using

$$\frac{1}{n} = \sqrt{\frac{2}{mc^2}[E-U]} = \sqrt{\frac{2}{mc^2}}\cdot\sqrt{T} \quad \text{with } T = \text{kinetic energy}$$

we find that Newtonian trajectories of energy E have the property that they extremize "Jacobi's action functional"

$$\begin{aligned} A[\text{path}] &\equiv \sqrt{\tfrac{m}{2}}\int \sqrt{T}\,ds \\ &= \int T\,dt \quad \text{by } ds = \sqrt{\tfrac{2}{m}T}\,dt \end{aligned}$$

It is at this point that the standard literature becomes, by my reading, quite confusing. We have on the one hand

$$\text{HAMILTON'S PRINCIPLE:} \quad \delta S = 0 \quad \text{with} \quad S[\text{path}] = \int (T-U)\,dt$$

and on the other hand

$$\text{JACOBI'S PRINCIPLE:} \quad \hat{\delta} A = 0 \quad \text{with} \quad A[\text{path}] = \int T\,dt \tag{13}$$

—*both* of which are known informally as the "Principle of Least Action," but

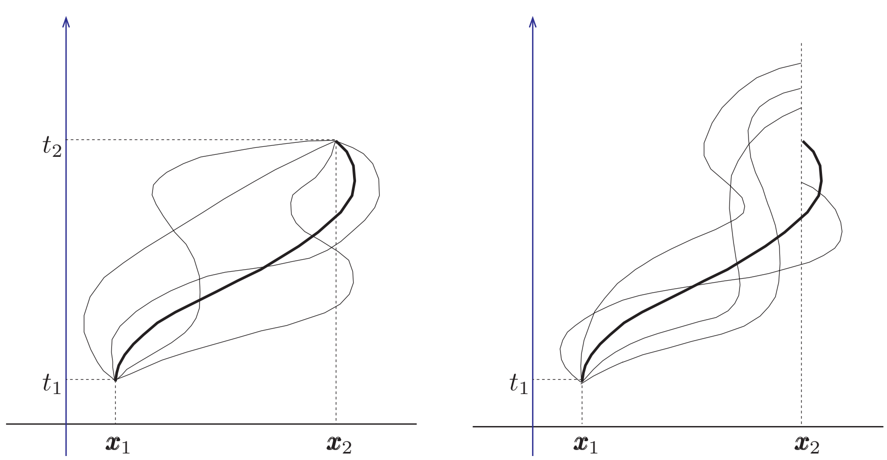

FIGURE 5: *Comparison of the variational processes fundamental to Hamilton's Principle (on the left) and the Jacobi's Principle of Least Action. In both cases the spatial endpoints are specified, but in Hamilton's Principle* transit time *is a constant of the variational process, while Jacobi's Principle requires the conserved* energy E *to be variationally constant.*

the meanings of which are profoundly distinct. Hamilton's principle, as is well-known, has everything to do with the *temporal* aspects of dynamics: it gives us (in Lagrangian form) the *equations of motion*. Jacobi's principle, though deceptively notated to suggest otherwise, has in fact *nothing* to do with temporal dynamics, but everything to do with the *geometrical figure of dynamical trajectories*; this becomes clear when one traces backwards the argument which led from (10) to (13). Sharpened understanding of the point here at issue follows at once from the observation that the integrand

$$\mathcal{A}_E(\boldsymbol{x},\mathring{\boldsymbol{x}}) \equiv \frac{1}{n(\boldsymbol{x};E)}\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}}$$

in (10) is *homogeneous of degree one* in the variables $\mathring{\boldsymbol{x}} \equiv \frac{d}{d\lambda}\boldsymbol{x}$; under arbitrary reparameterization $\lambda = \lambda(\tau) \longleftarrow \tau$ we therefore have

$$\int \mathcal{A}_E(\boldsymbol{x},\tfrac{d\boldsymbol{x}}{d\lambda})\,d\lambda = \int \tfrac{d\tau}{d\lambda}\mathcal{A}_E(\boldsymbol{x},\tfrac{d\boldsymbol{x}}{d\tau})\cdot\tfrac{d\lambda}{d\tau}\,d\lambda = \int \mathcal{A}_E(\boldsymbol{x},\tfrac{d\boldsymbol{x}}{d\tau})\,d\tau$$

according to which not only Jacobi's principle (10) but all of its consequences are *form-invariant with respect to arbitrary reparameterizations*. They therefore are, in particular, form-invariant with respect to arbitrary clock-regraduations $t \longrightarrow \tau = \tau(t)$, and so can have nothing to do with the specifically temporal aspects of mechanics.

We come thus to a conclusion which is, at least in the light of our optical experience, somewhat counterintuitive: the particle elects to pursue not the path $\boldsymbol{a} \longrightarrow \boldsymbol{b}$ which minimizes transit time, but the iso-energetic path which extremizes the Jacobi action A[path]; it pursues the "path of least action," a *geodesic with respect to the action metric*

$$d\sigma = a_E(\boldsymbol{x})ds$$

where $ds = \sqrt{(dx)^2 + (dy)^2 + (dz)^2}$ is Euclidean, and where

$$a_E(\boldsymbol{x}) \equiv \sqrt{\frac{2}{mc^2}\,[E - U(\boldsymbol{x})]} = \frac{1}{n(\boldsymbol{x};E)}$$

With respect to the action metric the dynamical trajectory is "least curved," "straightest possible," generated by "parallel prolongation." The particle pursues its trajectory with Euclidean speed

$$\dot{s} = \sqrt{\tfrac{2}{m}[E - U(\boldsymbol{x})]} = \frac{c}{n(\boldsymbol{x};E)} = c\cdot a_E(\boldsymbol{x})$$

With respect to the action metric its "speed" is $\dot{\sigma} = a_E(\boldsymbol{x})\dot{s}$, and with respect to both metrics the speed is, in general, non-constant. With respect to the "reciprocal metric"

$$d\tilde{\sigma} = \frac{1}{a_E(\boldsymbol{x})}ds$$

speed *is* constant ($\frac{d}{dt}\tilde{\sigma} = c$), but the trajectory is non-geodesic.

6. Theoretical placement of the Principle of Least Action. It is my sense that the profoundly geometrical purport of Jacobi's principle is not widely appreciated, that physicists—even those writing about the subject (see, for example, §8.6 of Goldstein's 2nd edition, or §7.5 of his 1st edition)—typically don't know quite what to make of Jacobi's principle, which they find easy therefore simply to ignore. It is on these grounds I understand the fact that the "Principle of Least Action" terminology is so often misapplied. For example, Richard Feynman gave to the dissertation (Princeton, May 1942) in which he first described what has come to be known as the "Feynman sum-over-paths formalism"; *i.e.*, in which he first had occasion to write

$$K(x,t;x_0,0)=\int e^{\frac{i}{\hbar}S[\text{path}]}\,\mathcal{D}[\text{paths}]$$

. . . the title *The Principle of Least Action in Quantum Mechanics*, though what he clearly had in mind was a quantum generalization of Hamilton's principle.[3] A paper which might more properly have worn Feynman's original title (but to which he in fact gave a different title[4]) was written recently by Richard Crandall, who uses path-integral methods to study the time-independent object

$$G(x,x_0;E)=\frac{1}{i\hbar}\int_0^\infty K(x,t;x_0,0)\,e^{\frac{i}{\hbar}Et}\,dt$$

But the recent work which draws most explicitly upon the geometry of classical trajectories is that having to do with chaos—particularly quantum chaos.

E. T. Whittaker (see §§105–107 of his ANALYTICAL DYNAMICS) has drawn attention to the close kinship which links the Principle of Least Action to (for example) Gauss' "Principle of Least Constraint" and to the even less well known "Appell formalism." This whole corner of physics—of what I have chosen to call "geometrical mechanics"—remains much less studied than (in my view) it deserves to be. Research questions spring easily to mind, and much that is useful (not to say pretty) remains to be worked out.

7. Application to the case of a free particle. This simplest of all dynamical systems arises from setting $U(\boldsymbol{x})=0$. The "mechanical index of refraction" is given then (see again pages 9 and 12) by

$$\frac{1}{n(\boldsymbol{x};E)}=\sqrt{\frac{2E}{mc^2}}\quad:\quad\text{constant}$$

---

[3] When the work was finally published (Rev. Mod. Phys. **20**, 367 (1948)) it wore new title: "Space-time approach to non-relativistic quantum mechanics," where the first adjective refers to the fact that the "paths" in question are inscribed on spacetime, and the second adjective disabuses readers of any presumption that the theory has something therefore to do with relativity.

[4] See "Combinatorial approach to Feynman path integration," J. Phys. A: Math. Gen. **26**, 3627 (1993) and papers there cited.

The trajectory equations (8) therefore assume the simple form

$$\left(\tfrac{d}{ds}\right)^2 x(s) = \left(\tfrac{d}{ds}\right)^2 y(s) = 0$$

and give

$$\begin{aligned} x(s) &= x_0 + x_1 s \\ y(s) &= y_0 + y_1 s \end{aligned}$$

The time of transit $0 \to s$ is given (see again page 10) by

$$t(s) = \int_0^s \frac{1}{\sqrt{\frac{2}{m}E}}\, ds' = \sqrt{m/2E}\, s$$

which by a trivial functional inversion becomes

$$s = \sqrt{2E/m}\, t$$

giving finally the uniform rectilinear *motion*

$$\begin{aligned} x(s(t)) &= x_0 + ut \quad \text{with} \quad u \equiv x_1\sqrt{2E/m} \\ y(s(t)) &= y_0 + vt \quad \text{with} \quad v \equiv y_1\sqrt{2E/m} \end{aligned}$$

enshrined in Newton's 1st Law. So far, so good.

8. Attempted application to the case of a ballistic particle. We turn now to the most familiar instance of the next-simplest case: $U(\boldsymbol{x}) = mgy$. To describe the ballistic motion of a thrown particle m we would, as first-year students of physics, write

$$\begin{aligned} m\ddot{x} &= 0 \\ m\ddot{y} &= -mg \end{aligned}$$

and obtain

$$\begin{aligned} x(t) &= x_0 + ut \\ y(t) &= y_0 + vt - \tfrac{1}{2}gt^2 \end{aligned}$$

To reduce notational clutter we might restrict our attention to the typical case $x_0 = y_0 = 0$, writing

$$\left.\begin{aligned} x(t) &= ut \\ y(t) &= vt - \tfrac{1}{2}gt^2 \end{aligned}\right\} \tag{14.1}$$

to describe the *motion* of the particle. To describe its *trajectory* we have only to eliminate t between those two equations: immediately

$$y = Y(x) \equiv (v/u)x - (g/2u^2)x^2 \tag{14.2}$$

which describes a down-turned parabola that achieves is maximal value at

$$x_{\max} = uv/g, \quad \text{where} \quad y_{\max} \equiv Y(x_{\max}) = \tfrac{1}{2}(v^2/g)$$

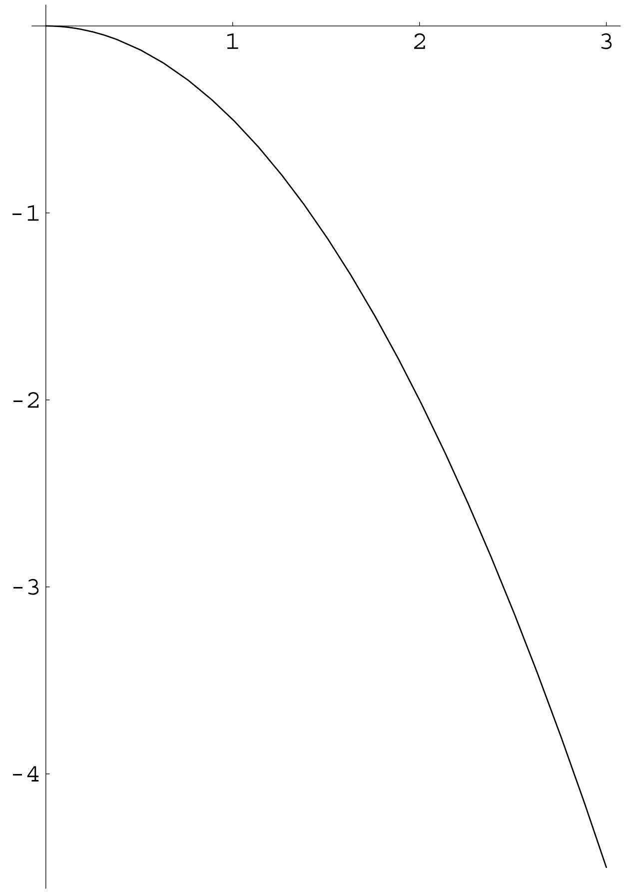

FIGURE 6: *Trajectory* $y = -(g/2u^2)x^2$ *of a ballistic particle in the case* $v = 0$ *and* $g/2u^2 = 1/2$.

We might at this point agree at this point to adopt the additional simplifying assumption that the particle has been launched with no initial $\updownarrow$ velocity: $v = 0$. Then $x_{\rm max} = y_{\rm max} = 0$, which is to say: the apex of the trajectory has been positioned at the origin (see the preceding figure) and equations (14) have become

$$\left.\begin{aligned} x(t) &= ut \\ y(t) &= -\tfrac{1}{2}gt^2 \end{aligned}\right\} \tag{15.1}$$

$$Y(x) = -(g/2u^2)\,x^2 \tag{15.2}$$

The theory developed in §4 is addressed, however, to the construction of an *implicit* description of the trajectory: it has been designed to lead us not to $y(x)$ but to $\{x(s), y(s)\}$, the claim being that those functions are solutions of a pair of differential equations obtained by specialization of (8). In the present instance

$$\frac{1}{n} = \sqrt{\frac{2}{mc^2}\,[E - mgy]}$$

and the trajectory equations (8)—from which, it will be noticed, the artificial c-factors have dropped automatically away (as they always do)—become

$$\begin{aligned}
\frac{d}{ds}\sqrt{E - mgy}\,\frac{dx}{ds} \qquad\qquad\qquad\qquad &= 0\\
\frac{d}{ds}\sqrt{E - mgy}\,\frac{dy}{ds} + \frac{mg}{2\sqrt{E - mgy}} &= 0
\end{aligned}$$

or, after simplifications,

$$\left.\begin{aligned}
(E - mgy)\,\overset{\circ\circ}{x} - \tfrac{1}{2}mg\,\overset{\circ}{x}\overset{\circ}{y} &= 0\\
(E - mgy)\,\overset{\circ\circ}{y} - \tfrac{1}{2}mg\,\overset{\circ}{y}\overset{\circ}{y} &= -\tfrac{1}{2}mg
\end{aligned}\right\} \tag{16}$$

What accounts for the fact that our simple free-fall system has given rise to such an intractable system of coupled non-linear differential equations? The arc-length of a segment of the plane curve $y(x)$ can be described

$$s(a, b) = \int_a^b \sqrt{1 + (dy/dx)^2}\,dx$$

Looking to the parabolic arc shown in FIGURE 6, we have (working from (15.2) with the assistance of *Mathematica*)

$$\begin{aligned}
s(x) \equiv s(0, x) &= \int_0^x \sqrt{1 + k^2y^2}\,dy \quad : \quad k \equiv g/u^2\\
&= x\int_0^1 \sqrt{1 + k^2x^2z^2}\,dz\\
&= \frac{kx\sqrt{1 + k^2x^2} + \text{Arcsinh}(kx)}{2k}
\end{aligned}$$

Functional inversion would supply $x(s)$, which we would insert into $Y(x)$ to obtain $y(s) \equiv Y(x(s))$. But <u>functional inversion is, in this instance, clearly impossible</u>.

Relatedly, we have (see again page 12)

$$\dot{s} = \sqrt{\tfrac{2}{m}[E - mgy(t)]}$$

which presents us with a statement of the form

$$\begin{aligned}
s(t) &= \int_0^t \sqrt{a + b\tau + c\tau^2}\,d\tau\\
&= \frac{2\sqrt{c}\,(b + 2ct)\sqrt{a + bt + ct^2} - (b^2 - 4ac)\log\left[\frac{b+2ct}{\sqrt{c}} + 2\sqrt{a + bt + ct^2}\right]}{8c^{3/2}}\Bigg|_0^t
\end{aligned}$$

Clearly, the functional inversion that would produce $t(s)$ is not feasible, so we would be frustrated if we attempted to proceed

$$\begin{array}{ccc} x(t) & \longrightarrow & x(t(s)) \equiv \mathcal{X}(s) \\ y(t) & & y(t(s)) \equiv \mathcal{Y}(s) \end{array}$$

9. Concluding remarks. On page 10 I speculated that it might sometimes be easier to solve the trajectory equations than to solve the equations of motion. Such occasions, we have reason now to think, must be very rare. For it appears to be the case that the s-parameterized description $\{\mathcal{X}(s), \mathcal{Y}(s), \mathcal{Z}(s)\}$ of a space curve, however abstractly attractive it may be, is generally *too complicated to write down*. To say the same thing another way: the trajectory equations (8) appear in most cases to be *too complicated to solve* (except numerically?).

More perplexing is the puzzle that emerges when one compares the optical "ray equation"

$$\Big\{\frac{d}{d\lambda}\frac{\partial}{\partial \mathring{\boldsymbol{x}}} - \frac{\partial}{\partial \boldsymbol{x}}\Big\} n(\boldsymbol{x}, \nu)\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = \mathbf{0}$$

that results from Fermat's **principle of least time** with the mechanical "trajectory equation"

$$\Big\{\frac{d}{d\lambda}\frac{\partial}{\partial \mathring{\boldsymbol{x}}} - \frac{\partial}{\partial \boldsymbol{x}}\Big\} \frac{1}{n(\boldsymbol{x}; E)}\sqrt{\mathring{\boldsymbol{x}}\cdot\mathring{\boldsymbol{x}}} = \mathbf{0}$$

that results from Jacobi/Hertz's **principle of least action**. Why does the index of refraction appear upstairs in the former, downstairs in the latter? Can this curious circumstance have anything to do with the reciprocal relationship that in the theory of waves

$$\varphi(x,t) = A(k)\, e^{i[kx - \omega(k)t]}$$

is in an important class of cases[5] found to relate

$$\text{phase velocity} = \frac{\omega(k)}{k}$$

to

$$\text{group velocity} = \frac{d\omega(k)}{dk}$$

[5] We note that if

$$\frac{\omega}{k}\cdot\frac{d\omega}{dk} = \text{some constant (velocity)}^2\text{, call it } c^2$$

then integration of $\omega\cdot d\omega = c^2 k\cdot dk$ gives

$$\omega^2 = c^2k^2 + \text{constant of integration, call it } c^2\varkappa^2$$

so we have been led to the relativistic theory of massive particles (Klein-Gordon theory). For discussion of a reciprocity principle that emerges from Hamilton-Jacobi theory, see *Classical Mechanics* (1983), page 364.