CALCULUS OF VARIATIONS
and
TENSOR CALCULUS
(Chapters 2-3)

U.H. GERLACH

Beta Edition
Autumn 2017

CHAPTER 1

FUNDAMENTAL IDEAS

Lecture 1

1.1. Multivariable Calculus as a Prelude to the Calculus of Variations.

1.2. Some Typical Problems in the Calculus of Variations

1.3. Methods for Solving Problems in Calculus of Variations

1.3.1. Method of Finite Differences.

Lecture 2

1.4. The Method of Variations

1.4.1. Variants and Variations.

1.4.2. Variational Derivative.

1.4.3. Euler's Differential Equation.

1.5. Solved Example: *Brachistochrone*

Lecture 3

1.6. Integration of Euler's Differential Equation

CHAPTER 2

2 GENERALIZATIONS

2.1. Functional with Several Unknown Functions

Lecture 4

2.2. Extremum Problem with Side Conditions

2.2.1. Heuristic Solution.

2.2.2. Solution via Constraint Manifold.

Lecture 5

2.2.3. Variational Problems with Finite Constraints.

2.3. Variable End Point Problem

2.3.1. Extremum Principle at a Moment of Time Symmetry.

Lecture 6

2.4. Generic Variable Endpoint Problem

There are variable end point problems in which the endpoint lies on a curve or a surface. For example,

(1) what is the optimal curve from a point to a surface, or for that matter from one surface to another?
(2) How does an extremal meet a terminating surface?
(3) Given that a functional extends over two different media,

$$J = \Delta_a^c F(x, y, y')\, dx + \Delta_c^b F^*(x, y, y')\, dx \ ,$$

how must an extremal in $x < c$ be related to to its continuation in $c < x$, i.e. what are the junction conditions that prevail at the interface. For example, what path does light take as it passes from one medium to another? These questions and others can be answered within the framework

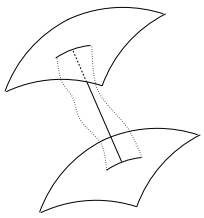


FIGURE 2.4.1. Optimal curve and its variants between two surfaces.

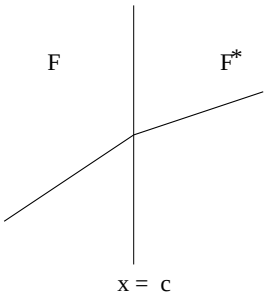


FIGURE 2.4.2. Curve crossing the boundary between two media characterized by F and F^*.

of the following generic variable end point problem:
Find the curve $\{y_1(x), \ldots, y_n(x)\}$ in an $n+1$ dimension space for which the variational integral

$$J[y_1, \ldots, y_n] = \Delta_{x_0}^{x_1} F(x, y, \ldots, y_n, y_n', \ldots, y_n')dx = \text{extremum}$$

subject to the constraint that the starting and end points $\{y_i(x_0)\}$ and $\{y_i(x_1)\}$ lie on two prescribed curves (or surfaces).
This variational principle asks us to consider changes in J not only due to variation in the curves,

$$y_i \to y_i + \delta y_i,$$

but also due to variations in the points

$$\begin{aligned} x_0 &\to x_0 + \delta x_0 \\ x_1 &\to x_1 + \delta x_1, \end{aligned}$$

where the curves start and finish. The determination of these changes in J is a task which one must do before the endpoint constraints are imposed.
To simplify discussion we consider the simplest case

$$J = \Delta_{x_0}^{x_1} F(x, y, y')dx = \text{extremum}$$

and calculate the first variation δJ allowing the endpoints to vary. Consider the curves

		starts	ends
curve:	$y(x)$	$P_0 = (x_0, y_0)$	$P_1 = (x_1, y_1)$
its variant:	$y^*(x)$	$P_0^* = (x_0 + \delta x_0, y_0 + \delta y_0)$	$P_1^* = (x_1 + \delta x_1, y_1 + \delta y_1)$

and the corresponding difference,

$$h(x) = y^*(x) - y(x) = \text{difference at fixed x.}$$

2.4.1. General Variations in the Functional. The total variation in the functional to be extremized is

$$\begin{aligned} \Delta J &= \Delta_{x_0+\delta x_0}^{x_1+\delta x_1} F(x, y+h, y'+h')dx - \Delta_{x_0}^{x_1} F(x, y, y')dx \\ &= \Delta_{x_0}^{x_1}[F(x, y+h, y'+h') - F(x, y, y')]dx \\ &+ \Delta_{x_0+\delta x_0}^{x_0} F(x, y+h, y'+h')dx + \Delta_{x_1}^{x_1+\delta x_1} F(x, y+h, y'+h')dx \end{aligned}$$

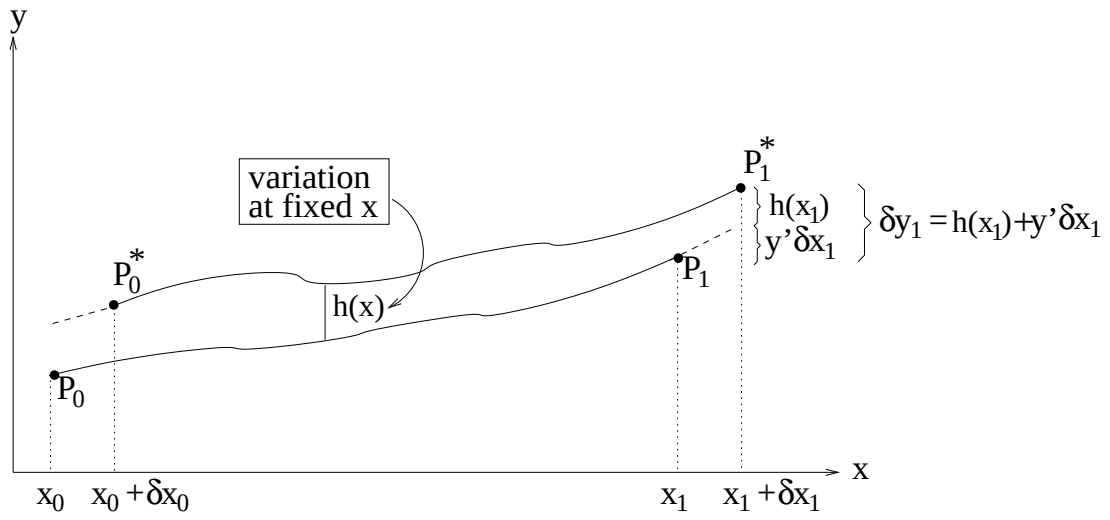


FIGURE 2.4.3. A curve and its generic variant.The total vertical variation δy_1 includes only the principal linear part. With that stipulation one has $h(x_1) = h(x_1 + \delta x_1)$.

whose principal linear part is

$$\Delta J = \Delta_{x_0}^{x_1} \left[F_y - \frac{d}{dx} F_{y'} \right] h(x)\, dx + F_{y'} h(x)|_{x_0}^{x_1} + F(x, y, y')\delta x|_{x_0}^{x_1}.$$

The last term is new. It is proportional to the shift in x :

$$\delta x|^{x_0} = \delta x_0 \quad \text{and} \quad \delta x|^{x_1} = \delta x_1.$$

Referring to Figure 2.4.3, we express any type of end point variation in terms of the independent variations δy_i and δx_i :

$$\begin{aligned} h(x_0) = &\quad \delta y_0 - y'(x_0)\delta x_0 \\ h(x_1) = &\quad \delta y_1 - y'(x_1)\delta x_1 \end{aligned}$$

Thus one obtains for the total first variation

$$\begin{aligned} \Delta J \quad = \quad & \Delta_{x_0}^{x_1} \frac{\delta J}{\delta y} h(x)dx + F_{y'}\delta y \Big|_{x_0}^{x_1} \\ & + \left(F - y' \frac{\partial F}{\partial y'} \right) \delta x \Big|_{x=x_0}^{x=x_1} \end{aligned} \tag{2.4.1}$$

We have two endpoint terms because in general both y and x are allowed to change at the boundary.
This is very useful formula; it has many consequences. It is used, among others,

(a): to establish junction conditions,
(b): to establish boundary conditions,
(c): as a starting point for the "canonical" formulation of mechanics,
(d): as a starting point for the Hamilton-Jacobi formulation of mechanics.

2.4.2. Junction Conditions. Suppose one wished to find an optimal curve between two points in adjoining regions characterized by different variational integrands. Thus we consider an optimal curve which leaves one region characterized by F and enters another characterized by F^* so that

$$J = \Delta_a^c F(x, y, y')dx + \Delta_c^b F^*(x, y, y')dx = \text{extremum}.$$

The boundary between the two adjoining regions is $x = c.$, and it is depicted in Figure 2.4.2 on page 6. The extremum principle establishes a relationship between the properties of the curve in the two regions. To determine this relation we, as usual, consider the total first variation

$$\begin{aligned}\Delta J &= \Delta_a^c \frac{\delta J}{\delta y}\delta y(x)dx + \Delta_c^b \frac{\delta J}{\delta y}\delta y(x)dx \\ &\quad + F_{y'}\delta y|_a^{c^-} + F^*_{y'}\delta y|_{c^+}^b \\ &\quad + \left(F - y'\frac{\partial F}{\partial y'}\right)\delta x\Big|_a^{c^-} + \left(F^* - y'\frac{\partial F}{\partial y'}\right)\delta x\Big|_{c^+}^b .\end{aligned}$$

The fact that the curve is optimal implies

$$\frac{\delta J}{\delta y} = 0.$$

The fact that the boundary between the two regions is $x = c$, implies that $\delta x(c) = 0$, i.e. it is fixed and given.
Consequently, the extremal condition implies

$$0 = \frac{\Delta J}{\delta y(c)} = F_{y'}\Big|_{c^-}^{c^+}$$

Thus

$$F_{y'}|_{c^+} = F_{y'}|_{c^-} \quad \text{“ Junction Condition”}$$

EXAMPLE. *(Snell's Law).*

Consider a light beam propagating from one medium where the refractive index is $n(x, y)$ to another where it is $n^*(x, y)$. The variational principle is

$$J = \Delta_a^c n(x, y)\sqrt{1 + y'^2}dx + \Delta_c^b n^*(x, y)\sqrt{1 + y'^2}dx = \text{extremum!}$$

The junction condition is

$$n(x, y)\frac{y'}{\sqrt{1 + y'^2}}\Big|_{x=c^-} = n^*(x, y)\frac{y'}{\sqrt{1 + y'^2}}\Big|_{x=c^+},$$

which has the geometrical interpretation

$$n \sin\theta = n^* \sin\theta^*,$$

which is Snell's law of refraction.

EXAMPLE. *(Segmented string)*

Consider a string with discontinuous tension. Such a discontinuity would prevail at the junction of two sting segments $0 \le x < c$ and $c < x \le \ell$ which are joined at $x = c$ by a frictionless ring. This ring (having neglegible mass) is constrained to

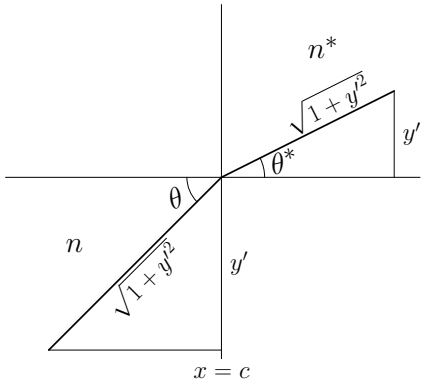


FIGURE 2.4.4. Ray refraction at the interface between two media having constant refractive indeces n and n^*.

slide without friction along a rod transverse to the segmented string. However, at the same time this rod pulls with non-zero force along the longitudinal direction of that string. Consequently, there is a corresponding non-zero jump in the tension of the string segments along their junction. In fact, the difference, $T_2(c^+) - T_1(c^-)$, in the tension

$$T(x) = \begin{cases} T_1(x) & x < c \\ T_2(x) & c < x \end{cases}$$

on the two sides of the junction $x = c$ equals the longitudinal force provided by the frictionless rod.

The equation for the amplitude profile of a normal mode of any string is

$$\frac{d}{dx}T(x)\frac{dy}{dx} + [q(x) - \lambda\rho(x)]\, y = 0 \tag{2.4.2}$$

Q: What are the junction conditions across the frictionless support rod?
A: The variational principle for the amplitude profile is

$$\text{P.E.} = \frac{1}{2}\,\Delta_0^c \left[T_1\, y'^2 + q\, y^2\right] dx + \frac{1}{2}\,\Delta_c^\ell \left[T_2\, y'^2 + q\, y^2\right] dx = \text{extremum} \tag{2.4.3}$$

subject to

$$\Delta_0^\ell\, \rho\, y^2 dx = 1. \tag{2.4.4}$$

The Euler equationfor this "isoperimetric" problem is given by Eq.(2.4.2), where λ is the Lagrange multiplier for the given constraint, Eq.(2.4.4). The junction condition

$$F_{y'}|_{x=c^+} = F_{y'}|_{x=c^-}$$

yields

$$T_2(x)\, \left.\frac{dy}{dx}\right|_{c^+} = T_1(x)\, \left.\frac{dy}{dx}\right|_{c^-}.$$

In other words, the *transverse* force on the string parallel to the boundary $x = c$ is continuous across the boundary.

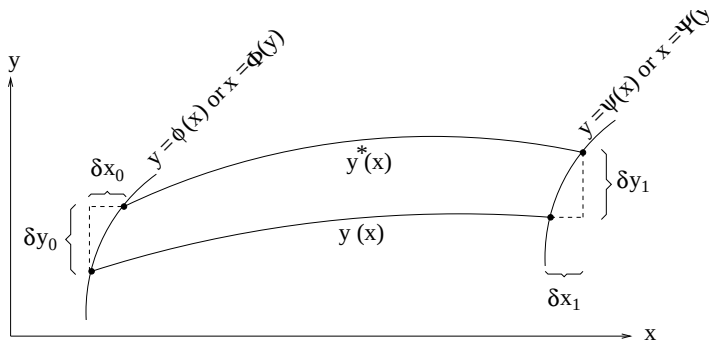


FIGURE 2.4.5. A curve and its variant with end point constraints.

2.4.3. Transversality Conditions. Let us now subject that first variation to the prescribed endpoint constraints. We require that the paths start and terminate on the two respective given curves

$$y = \varphi(x) \tag{2.4.5}$$

and

$$y = \psi(x) \tag{2.4.6}$$

or equivalently,

$$x = \Phi(y) \tag{2.4.7}$$

and

$$x = \Psi(y) \tag{2.4.8}$$

In that case the vertical and horizontal variations in the end point are related by

$$\delta y_0 = \varphi'(x_0)\delta x_0 \quad \text{and} \quad \delta y_1 = \psi'(x_1)\delta x_1 \tag{2.4.9}$$

Consequently, the first variation in J is

$$\begin{aligned} \Delta J &= \Delta_{x_0}^{x_1} \frac{\delta J}{\delta y} h(x)dx + \left[F_{y'}\psi'(x) + (F - y'F_{y'}\right]_{x=x_1} \delta x_1 \\ &\quad - \left[F_{y'}\varphi'(x) + (F - y'F_{y'}\right]_{x=x_0} \delta x_0 \end{aligned}$$

If $y(x)$ is an optimal curve, then $\Delta J = 0$ for arbitrary variations $h(x)$ and arbitrary variations δx_i at the end points. The former implies that $y(x)$ satisfies the Euler equation. The latter implies

$$\begin{aligned} \left[F + (\psi' - y')F_{y'}\right]_{x=x_1} &= 0 \\ \left[F + (\psi' - y')F_{y'}\right]_{x=x_0} &= 0, \end{aligned}$$

which are known as the transversality conditions.

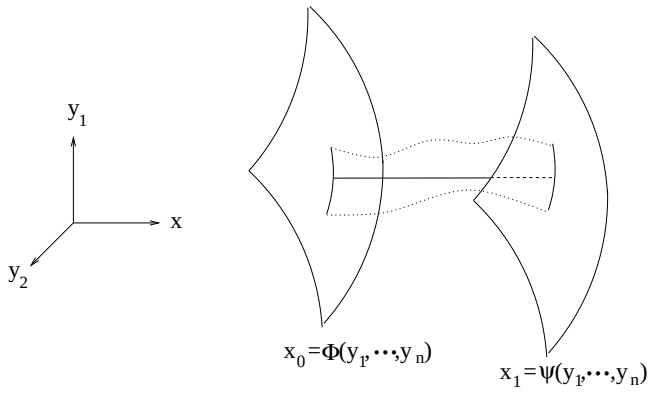


FIGURE 2.5.1. A curve and its variant between surfaces.

2.5. Many Degrees of Freedom

One would like to generalize the variable end point problem from extremal paths in the plane to extremal path in a higher dimensional space. In that case the two curves (2.4.7) and (2.4.8) generalize to the two surfaces

$$x = \Phi(y_1, \ldots, y_n) \tag{2.5.1}$$

and

$$x = \Psi(y_1, \ldots, y_n) \tag{2.5.2}$$

for the loci of points where the extremal path starts and finishes.

The corresponding generalization for the total first variation consists of replacing the equation (2.4.1) with

$$\begin{aligned} \Delta J &= \Delta_{x_0}^{x_1} \sum_{i=1}^{n} \frac{\delta J}{\delta y_i(x)} \delta y_i(x) dx + \sum_{i=1}^{n} F_{y_i'} \delta y_i \Big|_{x_0}^{x_1} \\ &\quad + \left(F - \sum_{j=1}^{n} y_j' \frac{\partial F}{\partial y_j'} \right) \delta x \Big|_{x_0}^{x_1} \end{aligned}$$

Conceptual unit economy demands that it is best not to single out one of the y_i's over all the others. Consequently, we express the variation in δx at in terms of the variations of all the δy_i's in the surfaces (2.5.1) and (2.5.2),

$$\begin{aligned} \delta x_0 &= \sum_{i=1}^{n} \frac{\partial \Phi}{\partial y_i} \delta y_i \\ \delta x_1 &= \sum_{i=1}^{n} \frac{\partial \Psi}{\partial y_i} \delta y_i. \end{aligned}$$

With the help of these expressions the transversality conditions at these starting and terminating surface are given by

$$\left[F_{y_i'} + \frac{\partial \Phi}{\partial y_i} \left(F - \sum_{j=1}^{n} y_j' \frac{\partial F}{\partial y_j'} \right) \right]_{x=x_0} = 0, \qquad i = 1, \ldots, n$$

and analogous condition at $x = x_1$, where the gradient of the surface is $\frac{\partial \Psi}{\partial y_i}$.

EXAMPLE. *(Fermat's Principle of Least Time).*
Consider Fermat's principle of least time for a light beam propagating through a medium whose refractive index is $n(x,y,z)$. Let the light beam start at (x_1,y_1,z_1) and finish at some point x_2,y_2,z_2 on the surface

$$x=\Psi(y,z)$$

such that the path minimizes the time of travel. The quantity to be extemized is

$$\begin{aligned} J[y,z] &= \frac{1}{c}\,\Delta_1^2\, n(x,y,z)\sqrt{dx^2+dy^2+dz^2} \\ &= \frac{1}{c}\,\Delta_{x_1}^{x_2}\, n(x,y,z)\sqrt{1+y'^2+z'^2}dx \end{aligned}$$

One applies the transversality conditions to the variational integrand. We have

$$\begin{aligned} F &= n(x,y,z)\sqrt{1+y'^2+z'^2} \\ \frac{\partial F}{\partial y'} &= \frac{ny'}{\sqrt{1+y'^2+z'^2}}; \qquad \frac{\partial F}{\partial z'}=\frac{nz'}{\sqrt{1+y'^2+z'^2}} \\ F-y'F_{y'}-z'F_{z'} &= n\sqrt{1+y'^2+z'^2}-\frac{y'^2+z'^2}{\sqrt{1+y'^2+z'^2}} \\ &= \frac{n}{\sqrt{1+y'^2+z'^2}} \end{aligned}$$

There are two transversality conditions at $x=x_2$. They are

$$\frac{ny'}{\sqrt{1+y'^2+z'^2}}+\frac{\partial\Psi}{\partial y}\frac{n}{\sqrt{1+y'^2+z'^2}}=0$$

and

$$\frac{nz'}{\sqrt{1+y'^2+z'^2}}+\frac{\partial\Psi}{\partial z}\frac{n}{\sqrt{1+y'^2+z'^2}}=0\ .$$

Thus after cancellation we obtain

$$(y',z')=(-\frac{\partial\Psi}{\partial y},-\frac{\partial\Psi}{\partial z})$$

or

$$\text{"tangent"}\equiv(x',y',z')=(1,-\frac{\partial\Psi}{\partial y},-\frac{\partial\Psi}{\partial z})=\overrightarrow{\nabla}g$$

The left hand side is $\frac{d\vec{x}}{dx}$, the tangent to the light beam. the right hand side is the gradient of the function $g(x,y,z)=x-\Psi(y,z)$. We know that his gradient is $\perp$ to its isograms[1]. Thus the transversality condition says that the optimal light beam is the one which strikes the surface perpendicularly.

Lecture 7

:

[1] *isogram:* locus of points where the function has a constant value.

2.6. Parameterization Invariant Problem

In geometrical problems where one is dealing with the determination of paths which are optimal, it is better if one does not single any one coordinate as an independent variable. Instead, one represents these paths parrametrically by

$$x = x(\lambda)$$

$$y = y(\lambda).$$

This parameterization is arbitrary, i.e. different parametrizations represent the same curve, provided the function x(λ) and y(λ) do not satisfy

$$\frac{dx}{d\lambda} = 0 \qquad \text{and} \qquad \frac{dy}{d\lambda} = 0.$$

simultaneously. This restriction guarantees that λ expresses non-zero progress along the curve.

2.6.1. Parametrization Invariance via Homogeneous Function. Suppose we consider the simplest variational problem whose variational integral has the form

$$J = \Delta_{x_0}^{x_1} F\left(x, y, \frac{dy}{dx}\right) dx = \Delta_{\lambda_0}^{\lambda_1} G\left(x, y, \frac{dx}{d\lambda}, \frac{dy}{d\lambda}\right) d\lambda,$$

where

$$G = \frac{dx}{d\lambda} F\left(x, y, \frac{\frac{dy}{d\lambda}}{\frac{dx}{d\lambda}}\right) \tag{2.6.1}$$

This integrand is a function which is "homogeneous" in the variables $\frac{dx}{d\lambda}$ and $\frac{dy}{d\lambda}$ and its "degree of homogeneity" equals one. It satisfies the homogeneity relation

$$G\left(x, y, k\frac{dx}{d\lambda}, k\frac{dy}{d\lambda}\right) = k^n G\left(x, y, \frac{dx}{d\lambda}, \frac{dy}{d\lambda}\right). \tag{2.6.2}$$

The integer exponent n is the degree of homogeneity. In our case$n = 1$. By differentiating this expression with respect to k, and then setting k=1 one obtains Euler's homogeneous function theorem. In our case it reads

$$x'G_{x'} + y'G_{y'} \quad = \quad G, \tag{2.6.3}$$

because $n = 1$. Here

$$x' = \frac{dx}{d\lambda} \qquad \text{and} \qquad y' = \frac{dy}{d\lambda}.$$

Conversely, suppose that G is some homogeneous function of degree one in x' and y', i.e G satisfies Eq. (2.6.2), then the variational principle

$$\Delta_{\lambda_0}^{\lambda_1} G d\lambda = \text{extremum}$$

determines a parametrization invariant path. This is so because the parameter change $\lambda = \lambda(\overline{\lambda})$ yields

$$\begin{aligned}\Delta_{\lambda_0}^{\lambda_1} G\left(x,\, y,\, \frac{dx}{d\overline{\lambda}},\, \frac{dy}{d\overline{\lambda}}\right) d\overline{\lambda} &= \Delta_{\lambda_0}^{\lambda_1} G\left(x,\, y,\, x'\frac{d\lambda}{d\overline{\lambda}},\, y'\frac{d\lambda}{d\overline{\lambda}}\right) d\overline{\lambda} \\ &= \Delta_{\lambda_0}^{\lambda_1} G\left(x, y, x', y'\right) \frac{d\lambda}{d\overline{\lambda}} d\overline{\lambda} \\ &= \Delta_{\lambda_0}^{\lambda_1} G(x, y, x', y') d\lambda\,.\end{aligned}$$

We conclude that the variational problem is invariant under a change in parametrization, and that the extremal path is independent of one's choice of parametrization used to represent them. The extremal path satisfies the two Euler equations

$$G_x - \frac{d}{d\lambda} G_{x'} = 0 \qquad \text{and} \qquad G_y - \frac{d}{d\lambda} G_{y'} = 0 \tag{2.6.4}$$

These equations must be equivalent to the single equation

$$F_y - \frac{d}{dx} \frac{\partial F}{\partial\left(\frac{dy}{dx}\right)} = 0,$$

where F is related to G by Eq.(2.6.1) on page on the preceding page. This means that some how the two equations (2.6.4) cannot be independent equations. There must exist an identity between them. In fact, one finds with the help of Eq. (2.6.3) that

(2.6.5)

$$\frac{d}{d\lambda}\left[x' G_{x'} + y' G_{y'} - G\right] = 0 \quad \Rightarrow \quad x'\left(G_x - \frac{d}{d\lambda} G_{x'}\right) + y'\left(G_y - \frac{d}{d\lambda} G_{y'}\right) = 0$$

is an identity. It holds for an arbitrary path, extremal or non-extreamal. Let us apply these ideas to the geometrical problem of finding the geodesics in a space coordinatized by curvilinear coordinates.

2.7. Variational Principle for a Geodesic

Consider flat Euclidean space coordinatized by (x,y,z) = $\vec{x}$. Introduce curvilinear coordinates (x^1, x^2, x^3) as an alternative means of labelling the points

$$\left.\begin{array}{l} x(x^1, x^2, x^3) \\ y(x^1, x^2, x^3) \\ z(x^1, x^2, x^3) \end{array}\right\} \vec{r}(x^i).$$

We would like to reexpress distance

$$(ds)^2 = dx^2 + dy^2 + dz^2$$

in terms of these curvilinear coordinates. Thus we consider

$$dx = \sum_{i=1}^{3} \frac{\partial x}{\partial x^i} dx^i$$
$$dy = \sum_{i=1}^{3} \frac{\partial y}{\partial x^i} dx^i$$
$$dz = \sum_{i=1}^{3} \frac{\partial z}{\partial x^i} dx^i$$

The distance formula becomes

$$\begin{aligned}(ds)^2 &= \sum_i \sum_j dx^i \frac{\partial x}{\partial x^i}\frac{\partial x}{\partial x^j} dx^j + dx^i \frac{\partial y}{\partial x^i}\frac{\partial y}{\partial x^j} dx^j + dx^i \frac{\partial z}{\partial x^i}\frac{\partial z}{\partial x^j} dx^j \\ &= \sum_i \sum_j dx^i \frac{\partial \vec{r}}{\partial x^i} \cdot \frac{\partial \vec{r}}{\partial x^j} dx^j \\ &= g_{ij}(x^k) dx^i dx^j\end{aligned}$$

REMARK. The last line intoduces the *Einstein summation convention* in which a *pair of repeated indeces* implies a summation over the relevent coordinates. Also note that the coefficients

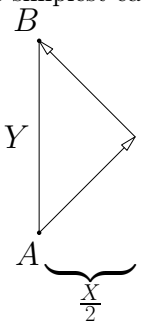
$$g_{ij} = \frac{\partial \vec{r}}{\partial x^i} \cdot \frac{\partial \vec{r}}{\partial x^j} = \vec{e}_i \cdot \vec{e}_j \tag{2.7.1}$$

are the inner products between ith and jth coordinate tangent vectors. They would not be mutually orthonormal in an oblique curvilinear coordinate system.

Q:: How does on determine the optimal path between two given points A and B?

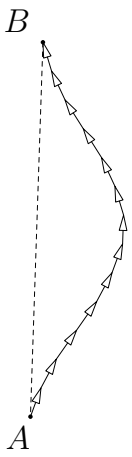
A:: The optimal path extremizes the path length between A and B.

The simplest case is a broken curve consisting of two parts:



$$\begin{aligned}distance &= 2\sqrt{\left(\frac{Y}{2}\right)^2 + \left(\frac{X}{2}\right)^2} = \sqrt{Y^2 + X^2} \\ &= \text{extremum for the direct } (X{=}0) \text{ path}\end{aligned}$$

For broken curves consisting of several pieces, we have



$$\sum_i \sqrt{\Delta x_i^2 + \Delta y_i^2} = extremum.$$

Among arbitrary curves beween A and B, the path length between A and B

$$\Delta_A^B \left[g_{ij} dx^i dx^j \right]^{1/2} = \begin{pmatrix} \text{minimum for a straightline} \\ \text{as compared to} \\ \text{any of its variants} \end{pmatrix} ,$$

i.e. *extremal length* is an indicator of straightness.

Remark. The introduction of curvilinear coordinates in Euclidean space is optional. In such a space one can always find a global coordinate system x,y,z such that

$$(ds)^2 = dx^2 + dy^2 + dz^2.$$

In other spaces, e.g. a curved surface or more generally a Rimannian space, there is no option. The non-Euclidean nature of the space makes the introduction of curvilinear coordiantes mandatory. Relative to such coordinates one necessarily has

$$(ds)^2 = g_{11} \left(dx^1\right)^2 + 2g_{12} dx^1 dx^2 + g_{22} \left(dx^2\right)^2 + \cdots$$

A particularly exotic example is the interior of a star. In terms of spherical coordinates (r,θ,ϕ), the distance formula is not the expected expression $(ds)^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$, but instead is given by

$$(ds)^2 = \frac{dr^2}{1 - \frac{8\pi}{3}\frac{G}{c^2}\rho r^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

where the constants ρ, G, and c are

$$\begin{aligned} \rho &= \frac{\text{mass}}{\text{volume}} \\ G &= \text{Newton's gravitational constant } \left(= \frac{1}{15,\ 000,\ 000} \text{ in cgs units} \right) \\ c &= \text{speed of light } (= 3 \times 10^{10} \text{ in cgs units}) \end{aligned}$$

This expression is a consequence of Einstein's law of graviation.

Introducing the spatial scale parameter

$$a = \left(\frac{8\pi}{3}\frac{G}{c^2}\rho\right)^{-1/2}, \tag{2.7.2}$$

the distance formula has the form

$$(ds)^2 = \frac{dr^2}{1-\frac{r^2}{a^2}} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right).$$

This suggests that one let

$$r = a\sin\chi$$

for the purpose of simplifying the distance formula. Doing so results in

$$(ds)^2 = a^2\left(d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\chi d\phi^2\right)\right).$$

One can show that this is the element of distance on part of a three-sphere S^3

$$x^2 + y^2 + z^2 + w^2 = a^2$$

of radius a, and which is coordinatized by

$$\begin{aligned} x &= a\ \sin\chi\sin\theta\cos\phi \\ y &= a\ \sin\chi\sin\theta\sin\phi \\ z &= a\ \sin\chi\cos\theta \\ w &= a\ \cos\chi \quad . \end{aligned}$$

The remarkable feature of this three-sphere is that its radius a is determined by the stars mass density[2] in Eq.(2.7.2).

2.8. Equation of Geodesic Motion

To determine the curve having extremal distance between A and B, we pick a curve with the suspected property and deform it slightly, but arbitrarily

$$\begin{aligned} \text{original curve: } x^i &= a^i(\lambda) \\ \text{variant = deformed curve: } x^i &= a^i(\lambda) + \delta a^i(\lambda). \end{aligned}$$

(1) Along either curve the distance is

$$s = \Delta_A^B\, ds = \Delta_0^1\left[g_{ij}\frac{dx^i}{d\lambda}\frac{dx^j}{d\lambda}\right]^{1/2} d\lambda. \tag{2.8.1}$$

(2) At *fixed* λ we find that
   (a) the metric coefficient $g_{ij}\left(x^b(\lambda)\right)$ differs from one cuve to the other by

$$\delta g_{ij} \equiv g_{ij}\left(a^k(\lambda) + \delta a^k(\lambda)\right) - g_{ij}\left(a^k(\lambda)\right) = \frac{\partial g_{ij}}{\partial x^k}\delta a^k(\lambda),$$

   (b) the components $\frac{dx^i}{d\lambda}$ of the tangent vector differ by

$$\delta\left(\frac{dx^i}{d\lambda}\right) \equiv \frac{d(a^i + \delta a^i)}{d\lambda} - \frac{da^i}{d\lambda} = \frac{d}{d\lambda}(\delta a^i),$$

[2]For example, with a density of 1 gram/cc, that "radius" would be ∼60 solar radii.

(c) the application of these changes in g_{ij} and $\frac{dx^i}{d\lambda}$ to the integrand of Eq.(2.8.1) results in the following change in the integrated distance, Eq.(2.8.1),

$$\delta s = \Delta_0^1 \frac{\frac{1}{2} g_{ij} \frac{d}{d\lambda}\left(\delta a^i\right) \frac{da^j}{d\lambda} + \frac{1}{2} g_{ij} \frac{da^i}{d\lambda} \frac{d}{d\lambda}\left(\delta a^j\right) + \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \delta a^k \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda}\right]^{1/2}} d\lambda \,,$$

The first two term are equal. Integrate them by parts and omit the boundary term because both paths must pass through A and B:

$$\delta a^i(0) = \delta a^i(1) = 0 \,.$$

Thus obtain

$$\delta s = \Delta_0^1 f_k(\lambda) \delta a^k \left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda}\right]^{1/2} d\lambda \,,$$

where

$$f_k(\lambda) = \frac{-1}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda}\right]^{1/2}} \frac{d}{d\lambda} \frac{g_{kj} \frac{da^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda}\right]^{1/2}} + \frac{\frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}}{\left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda}\right]} \,. \tag{2.8.2}$$

An extremum is achieved, and the first order change δs vanishes for every first order deformation δa^k of the path $a^k(\lambda)$, when the three quantities that multiply the δa^k all vanish. Thus one obtains the n conditions

$$f_k(\lambda) = 0 \qquad k = 1, 2, 3, \cdots, n$$

for the determination of an optimal path.

In 3-dimensional Euclidean or Riemannian space these are three equations. In 4-dimensional Lorentz space-time these are four equations, even though in this latter case the space-time interval is given by

$$(d\tau)^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \,,$$

provided gravitation is absent.

Lecture 8

2.8.1. Parameterization Invariance. The equations $f_k(\lambda) = 0$ constitute overkill, more equations than necessary. Why? Because their number is more than enough to express the extremal nature of

$$s = \Delta_0^1 \left[g_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda}\right]^{1/2} d\lambda \,.$$

This follows from the parametrization independence of this integral. The reparametrization

$$\begin{aligned} \lambda \to \bar{\lambda} &= \lambda + h(\lambda) \\ \bar{\lambda}(0) &= 0 \\ \bar{\lambda}(1) &= 1 \Rightarrow h(0) = h(1) = 0 \end{aligned}$$

does not change the integral. It corresponds to a mere "repositioning of beads along a string" (= reparametrization)

$$s = \Delta_0^1 \left[g_{ij} \frac{dx^i}{d\bar{\lambda}} \frac{dx^j}{d\bar{\lambda}} \right]^{1/2} d\bar{\lambda}.$$

The change in $a^i(\lambda)$ brought about by such a reparametrization is

$$a^i(\lambda) \to a^i(\lambda + h(\lambda)) = a^i(\lambda) + \delta a^i(\lambda)$$

where

$$\delta a^i(\lambda) = \frac{da^i}{d\lambda} h(\lambda)$$

The fact that such variations cannot change the variational integral for arbitrary reparametrization magnitudes $h(\lambda)$ implies that

$$f_k(\lambda) \frac{da^k}{d\lambda} = 0 \tag{2.8.3}$$

The remarkable feature of this equation is that it holds for any path $a^k(\lambda)$, even if it is not optimal.

REMARK. An equation which holds true whether or not the quantities obey any differential equation is called an *identity*.

Thus instead of needing n independent equations, one needs only $n-1$ equations to specify an extremal path for n degrees of freedom. Indeed, we have seen this idea already in the context surrounding Eq.(2.6.1) on page 13.

2.9. Parametrization in Terms of Curve Length

The reparametrization freedom can be exploited to introduce a physically more interesting parameter, the length parameter itself,

$$ds = \left[g_{mn} \frac{da^m}{d\lambda} \frac{da^n}{d\lambda} \right]^{1/2} d\lambda \,.$$

Let $x^i = a^i(\lambda)$ be the optimal path. With s as the new path parameter, call this extremal path $x^i(s)$. Making this replacement in Eq.(2.8.2) for the differential equations

$$f_k(\lambda) = 0 \,,$$

one obtains

$$0 = g_{kj} \frac{d^2x^j}{ds^2} + \frac{\partial g_{kj}}{\partial x^m} \frac{dx^m}{ds} \frac{dx^j}{ds} - \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \frac{dx^i}{ds} \frac{dx^j}{ds} \,. \tag{2.9.1}$$

The middle term can be rewritten as

$$\frac{1}{2} \frac{\partial g_{kj}}{\partial x^m} \frac{dx^m}{ds} \frac{dx^j}{ds} + \frac{1}{2} \frac{\partial g_{kj}}{\partial x^m} \frac{dx^j}{ds} \frac{dx^m}{ds}$$

In the second term let $m = \bar{j}$, $j = \bar{m}$ and then drop the bar[3]. The result is

$$0 = g_{kj} \frac{d^2x^j}{ds^2} + \frac{1}{2} \left(\frac{\partial g_{kj}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{mj}}{\partial x^k} \right) \frac{dx^m}{ds} \frac{dx^j}{ds} \,.$$

[3]Workers in the field call this relabelling of the summation ("dummy") indeces "index gymnastics".

Introduce the inverse g^{lk} of the matrix g_{kj}:

$$g^{lk}g_{kj} = \delta^l_j\,.$$

The equation of the geodesic becomes

$$\frac{d^2x^l}{ds^2} + \Gamma^l_{mj}\frac{dx^m}{ds}\frac{dx^j}{ds} = 0\,, \tag{2.9.2}$$

where

$$\Gamma^l_{mj} = \frac{1}{2}g^{lk}\left(\frac{\partial g_{kj}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^j} - \frac{\partial g_{mj}}{\partial x^k}\right)$$

is the so called "Christoffel symbol of the 2^{nd} kind".

REMARK. The identitiy (2.8.3), or equivalently,

$$f_k\frac{dx^k}{ds} = 0\,, \tag{2.9.3}$$

should not come as a total surprise. Recall that the Pythagorean theorem, $(\Delta s)^2 = g_{ij}\Delta x^i\Delta x^j$, implies

$$g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 1\,.$$

This says that, in light of the arclength parametrization, the tangent has been normalized to unit length all along the curve. It therefore is an integral of motion of the system of differential equations, Eqs.(2.9.1) or (2.9.2). Indeed, differentiating with respect to s yields

$$\frac{d}{ds}\left(g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}\right) = f_k\frac{dx^k}{ds}\,.$$

The result is therefore this:
If $f_k = 0$ is satisfied, then $g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}$ is a constant, an integral of motion. Conversely, given *any* s-parametrized curve along which $g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}$ is constant, then Eq.(2.9.3) must be satisfied, even if $f_k \neq 0$.

Lecture 9

2.10. Physical Significance of the Equation for a Geodesic

Q: What is the physical significance and purpose of the Γ-coefficients in the equation for a geodesic?

A: These coefficients express the acceleration of a free particles relative to a given frame of reference. There are three archetypical reference frames: (i) free float coordinate frames, e.g. an orbiting space capsule not rotating relative to the fixed stars, (ii) rotating coordinate frames, and (iii) linearly accelerated coordinate frames.

2.10.1. Free float frame. Also called "inertial" frames of reference, such frames are defined by Newton's 1st Law of Motion:
This means that relative to a free float frame free particle move uniformly along straight lines, i.e. these particles obey Newton's 1st law of motion. Mathematically we have

$$\text{mass} \times \text{acceleration}{=}0 \;\Rightarrow\; \frac{dx^i}{ds} = (const)^i \Leftrightarrow (\text{straight line motion})$$

where the travelled distance is

$$s = (constant) \times \text{time} \Leftrightarrow (\text{uniform motion})$$

2.10.2. Rotating Frame. Consider a frame of reference which rotates with angular velocity $\vec{\omega}$ relative to the fixed stars. First, consider a vector, say $\vec{G}$, which is rigidly attached to this rotating frame. The vector will rotate relative to the static inertial frame, which is static relative to the stars. In fact, during a time interval dt this rotating vector $\vec{G}$ will have changed by an smount which is give by

$$(d\vec{G})_{static} = dt\vec{\omega} \text{ x } \vec{G}$$

This expresses an infinitesimal rotation around the axis $\vec{\omega}$. Second, consider the circumstance in which the vector $\vec{G}$ is not rigidly attached to the rotating frame. Instead, let it change by the amount $(d\vec{G})_{rot}$ during the time interval dt, then the total change in this vector will be

$$(d\vec{G})_{static} = (d\vec{G})_{rot} + dt\vec{\omega} \text{ x } \vec{G}$$

relative to the inertial frame static with respect to the fixed stars. Thus

$$\frac{d}{dt}\Big)_{static} = \frac{d}{dt}\Big)_{rot} + \vec{\omega}\text{x}$$

Finally, apply this transformation law to the postion vector $\vec{R}$,

$$\frac{d\vec{R}}{dt}\Big)_{static} = \frac{d\vec{R}}{dt}\Big)_{rot} + \vec{\omega}\text{x}\vec{R}$$

and then to the velocity vector $\frac{d\vec{R}}{dt}\Big)_{static} \equiv \vec{v}_s$. Assuming that $\vec{\omega}$ is constant, one obtains that the equation of motion for a free particle is

$$0 = m\frac{d^2\vec{R}}{dt^2}\Big)_{static} = m\left[\vec{a}_{rot} + 2\vec{\omega} \text{ x } \vec{v}_{rot} + \vec{\omega} \text{ x } (\vec{\omega} \text{ x } \vec{R})\right]$$

with

$$\vec{v}_{rot} = \frac{d\vec{R}}{dt}\Big)_{rot}$$

and

$$\vec{a}_{rot} = \frac{d\vec{v}_{rot}}{dt}\Big)_{rot}$$

In terms of components one has

$$m\frac{d^2x^i}{dt^2}\Big)_{rot} = \underbrace{-2m[\vec{\omega} \text{ x } \vec{v}_{rot}]^i}_{coriolis\ force}\underbrace{-m[\vec{\omega} \text{ x } (\vec{\omega} \text{ x } \vec{R})]^i}_{centrifugal\ force} \qquad (8.8) \tag{2.10.1}$$

relative to the *rotating* coordinate frame. The two terms on the right hand side of this equation are the "coriolis force" and the "centrifugal force" in these frames. Suppose we try to compare these Newtonian equations of motion with those of a geodesic,

$$\frac{d^2x^i}{ds^2} = -\Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds}$$

This is a good thing to do because both sets of equations have seond derivatives on the left hand side. The obstacles are the independent variables, time t for Newton's equations, and geodesic length s for the geodesic equations. But this difference is no problem because for each free particle path, length and time are linearly related:

$$\begin{aligned} s &= (constant) \times \text{time} \\ &\equiv (constant)\, t \end{aligned} \tag{2.10.2}$$

Consequently, the geodesic equations are equivlaent to

$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{jk}\frac{dx^i}{dt}\frac{dx^k}{dt} \qquad (8.10) \tag{2.10.3}$$

$$\frac{d^2t}{dt^2} = 0\,. \qquad (8.11) \tag{2.10.4}$$

Suppose we introduce time

$$x^0 = t\,.$$

as another coordinate so that

$$\frac{dx^0}{dt} = 1$$

and

$$\frac{d^2x^0}{dt^2} = 0\,. \tag{2.10.5}$$

We now consolidate the two equations (2.10.3) and (2.10.5) into the single four-component equation

$$\frac{d^2x^\mu}{dt^2} = -\sum_{\alpha=0}^{3}\sum_{\beta=0}^{3}\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} \qquad \mu = 0,1,2,3 \tag{2.10.6}$$

$$= -\Gamma^\mu_{00}\frac{dx^0}{dt}\frac{dx^0}{dt} - \sum_{k=1}^{3}(\Gamma^\mu_{0k}+\Gamma^\mu_{k0})\frac{dx^0}{dt}\frac{dx^k}{dt} - \sum_{j=1}^{3}\sum_{k=1}^{3}\Gamma^\mu_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} \tag{2.10.7}$$

The fact that $\frac{dx^0}{dt} = 1$ results in

$$\frac{d^2x^\mu}{dt^2} = -\Gamma^\mu_{00} - \sum_{k=1}^{3}(\Gamma^\mu_{0k}+\Gamma^\mu_{k0})\frac{dx^k}{dt} - \sum_{j=1}^{3}\sum_{k=1}^{3}\Gamma^\mu_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt}$$

What are these Γ-coefficients? They are determined by doing a component-by-component comparison between this four-component equation with the three-plus-one set of Eqs.(2.10.1) and (2.10.5) and requiring that these two sets are equivalent for all particles. For example, the first component ($i = 1$) of Eq.(2.10.1) is

$$\frac{d^2x^1}{dt^2} = -2(\omega^2\frac{dx^3}{dt} - \omega^3\frac{dx^2}{dt}) - (\vec{\omega}\cdot\vec{\omega}x^1 - \vec{\omega}\cdot\vec{R}\omega^2)$$

For $\mu = 0$ equivalence implies

$$\Gamma^0_{\alpha\beta} = 0 \qquad \alpha,\beta = 0,1,2,3\,,$$

while for $\mu = i\,(= 1,2,3)$ one finds that in the rotating frame

$$\Gamma^i_{0k}\frac{dx^k}{dt} = \left(\vec{\omega}\ \text{x}\ \frac{d\vec{R}}{dt}\right)^i \qquad i = 1,2,3 \quad \text{(coriolis acceleration)}$$

so that[4]

$$\begin{matrix} i=1 \to \\ i=2 \to \\ i=3 \to \end{matrix} \left[\quad \Gamma^i_{0k} \quad \right] \begin{bmatrix} dx^1/dt \\ dx^2/dt \\ dx^3/dt \end{bmatrix} = \begin{bmatrix} 0 & \omega^3 & -\omega^2 \\ -\omega^3 & 0 & \omega^1 \\ \omega^2 & -\omega^1 & 0 \end{bmatrix} \begin{bmatrix} dx^1/dt \\ dx^2/dt \\ dx^3/dt \end{bmatrix} .$$

Furthermore,

$$\Gamma^i_{00} = \left[\vec{\omega} \text{ x } (\vec{\omega} \text{ x } \vec{R})\right]^i \qquad i = 1,2,3 \quad \text{(centrifugal acceleration)} .$$

In other words,

$$\begin{aligned} \Gamma^i_{00} &\neq 0 \Leftrightarrow \text{centrifugal force} \neq 0 \\ \Gamma^i_{0k} &\neq 0 \Leftrightarrow \text{coriolis force} \neq 0 \end{aligned}$$

This says that the "Christoffel symbols" of the equations for a geodesic are an expression of centrifugal and coriolis accelerations, which are associated with a rotating coordinate frame. Let us now extend these same considerations to an accelerated frame of reference.

2.10.3. Uniformly Accelerated Frame. Recall that relative to a free float ("inertial") frame the equation for the free particle moving along the x-direction is

$$\frac{d^2x}{dt^2} = 0 . \tag{2.10.8}$$

Consider the motion of such a particle relative to an accelerated frame.

Q: What is the equation of motion of this particle relative to a frame accelerating into the positive x-direction with acceleration g?

A: To find out, consider a meter rod parallel to the x-direction, but attached to the accelerated frame.

Let ξ be the coaccelerating coordinate displacement along this meter rod. Thus a coaccelerating point on this meter stick is measured to be ξ units from the floor.

However, relative to the nonaccelerated inertial frame the x-coordinate of this point on the meter stick is

$$x = \xi + \frac{1}{2} g\tau^2$$

with

$$t = \tau$$

These two equations comprise the coordinate transformation $(\tau, \xi) \to (t, x)$ between the two frames. Let

$$\xi = \xi(\tau)$$

[4]The coefficients Γ^μ_{0k} and of Γ^μ_{k0} of the quadratic form in the geodesic equation occur only in the symmetric combination $(\Gamma^\mu_{0k} + \Gamma^\mu_{k0})$. Consequently, one may assume without loss of generality that $\Gamma^\mu_{0k} = \Gamma^\mu_{k0}$.

be the particle trajectory as measured in the accelerated frame. Appling the coordinate transformation to Newton's Eq.(2.10.8) for the free particle yields

$$0 = \frac{d^2}{d\tau^2}[\xi(\tau) + \frac{1}{2}g\tau^2] = \frac{d^2\xi}{d\tau^2} + g\,.$$

Furthermore,

$$0 = \frac{d^2t}{d\tau^2}\,.$$

Let $x^0 = t$ and $x^1 = \xi$. The equations assume the form

$$\begin{aligned} \frac{d^2x^0}{d\tau^2} &= 0 \\ \frac{d^2x^1}{d\tau^2} &= -g \end{aligned}$$

Compare them to the equation for a geodesic in two dimensions:

$$\begin{aligned} \frac{d^2x^0}{d\tau^2} &= -\Gamma^0_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \\ \frac{d^2x^1}{d\tau^2} &= -\Gamma^1_{00}\left(\frac{dx^0}{d\tau}\right)^2 - 2(\Gamma^1_{01} + \Gamma^1_{10})\frac{dx^0}{d\tau}\frac{dx^1}{d\tau} - \Gamma^1_{11}\left(\frac{dx^1}{d\tau}\right)^2 \end{aligned}$$

Assume they apply to all possible particle trajectories, we find

(2.10.9) $$\Gamma^1_{00} = g = \text{"inertial acceleration"} \qquad (8.15)$$

All the other Γ's are zero. One concludes therefore that relative to a uniformly accelerated frame

$$\Gamma^1_{00} \neq 0 \Leftrightarrow \text{"non-inertial force"} \neq 0\,.$$

2.11. 6.2.4 Equivlence Principle and "Gravitation"="Geometry"

Einstein directed attention to the fact that there is no way that one can tell the difference between (i) the motion of free particles in a uniformly accelerated frame, and (ii) the motion of particle falling freely in a uniform gravitational field.

This holds true regardless of the particle's composition. In other words, the particle trajectories are identical, no matter whether a particle is made of gold, aluminum, or snakewood. The measurements were first done by Lorand von *Eötvös*. Thus the inertial force is equvalent to, i.e. physically 'indistinguishable from, the gravitational force.

This equivalence is called *the equivalence principle*. It implies that

(2.11.1) $$\text{"inertial force"} = \text{"gravitational force"}\,,$$

or

$$\text{mass}_{inertial} \times \text{"inertial acceleration"} = \text{mass}_{gravitational} \times \text{gravitational field}$$

This equation holds regarless of whether the inertial mass and the gravitational mass refer to particles made of gold, aluminum, etc. Thus

$$\frac{m_{in}}{m_{acc}} = 1\,,$$

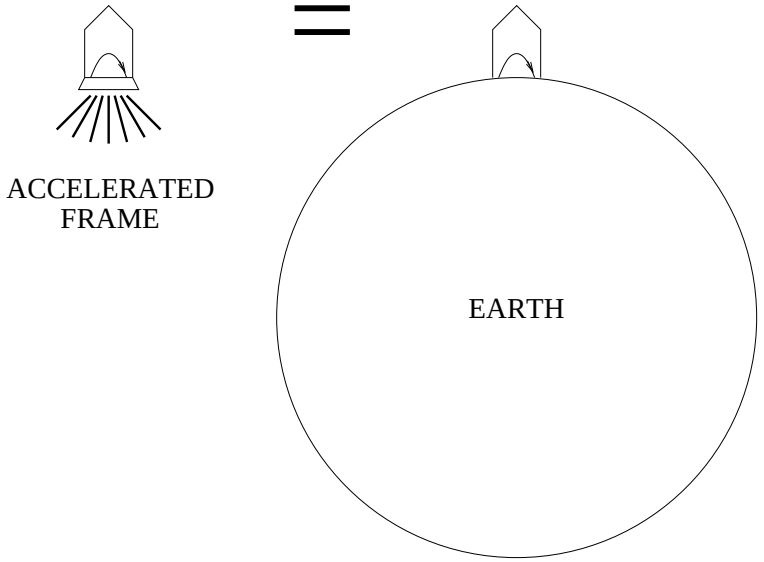


FIGURE 2.11.1. Equivalence Principle: the trajectories of particles in an accelerated frame and in a uniform gravitational field are indistinguishable regardless of their composition.

which expresses the fact that this ratio is independent of the composition of the paricles. Doing the cancellation and appying Eq.(2.10.9) to Eq. (2.11.1), we find that the Christoffel symbol Γ^1_{00} is to be identified with the graviational field, i.e.

$$\Gamma^i_{00} = (\text{"gravitational field"})^i \quad i = 1, 2, 3 \qquad (8.18a) \tag{2.11.2}$$

if we had extended our considerations to motion in three dimenssions.

Apply this result to a static gravitational field. In that case

$$\Gamma^i_{00} = \sum_{\alpha=0}^{3} \frac{1}{2} g^{i\alpha}(g_{\alpha 0,0} + g_{\alpha 0,0} - g_{00,\alpha})$$

The fact that the gravitatioal field is independent of time means that all time-derivatives must vanish. Consequently, the expression reduces to

$$\Gamma^i_{00} = \sum_{j=1}^{3} \frac{1}{2} g^{ij}(-) g_{00,i} \tag{2.11.3}$$

If the gravitational field is weak, then we can say that distance measurements in three dimenssional space are governed by Euclidean geometry as expressed by the theorem of Pythagoras,

$$(ds)^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \,.$$

This implies that the matrix, Eq. (2.7.1), on page 15 has the simple form

$$[g_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Consequently, its inverse $[g^{ij}]$is also the identity matrix,

$$[g^{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Applying this to the simplified Christoffell symbol, Eq.(2.11.3), give the result

$$\Gamma^i_{00} = -\frac{1}{2}\frac{\partial g_{00}}{\partial x^i} \qquad i = 1, 2, 3$$

Introduce this simplified result into Eq. (2.11.2) and obtain

$$-\frac{1}{2}\frac{\partial g_{00}}{\partial x^i} = (\text{gravitational field})^i \qquad i = 1, 2, 3\,.$$

Recall that the gravitational field is the negative gradient of Newton's gravitational potential $\Phi(x^1, x^2, x^3)$,

$$(\text{gravitational field})^i = -\frac{\partial}{\partial x^i}\Phi\,.$$

Consequently,

$$\Phi(x^1, x^2, x^3) = \frac{1}{2}g_{00}(x^1, x^2, x^3)\,.$$

This is a new result. It says that

$$\text{“gavitational potential"} = \text{“metric coefficient"}\,.$$

In other words, by examing the motion of particles we find that gravitation manifests itself by altering the properties of space (and time) as expressed by the coefficients in the “meteric"

$$(ds)^2 = g_{\alpha\beta}dx^\alpha dx^\beta$$

and by the concomitant Christoffell symbols

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\gamma}(g_{\gamma\alpha,\beta} + g_{\gamma\beta,\alpha} - g_{\alpha\beta,\gamma}) \qquad \mu, \alpha, \beta = 0, 1, 2, 3$$

One summarizes this result by saying that Einstein's *equivalence principle* leads to the conclusion that

$$\text{“gravitation"} = \text{“geometry"}\,.$$

Remark. The reasoning process that leads from (a) the equivalence between (i) trajectories of free particles in an accelerated frame and (ii) the trajectories of particles in a gravitaqtional field to (b) the identification of gravitation as a geometrical property is an example of the scientific method, induction based on experiments and/or observation.

CHAPTER 3

Variational Formulation of Mechanics

3.1. Hamilton's Principle

Lecture 10

Recall Newton's second law of motion applied to a force derived from a potential $U(x, y, z, t)$ is

$$\frac{d}{dt}\left(m\dot{\vec{x}}\right) = -\vec{\nabla}U(x, y, z, t)\,.$$

Like all differential equations, this law expresses – in mathematical form – a causal relation in the world. Here the change in momentum is caused by a force. This law governs the change in the particle's momentum along every point of the particle's trajectory. This law is therefore a local one, it is a differential statement about the motion of the particle. By integrating this differential equation one obtains the global trajectory, an integral curve. This is another causal relation, this time between the particle's initial position and velocity on one hand and its final position and velocity on the other. If one knows the initial velocity at point event A, then the integration of Newton's equation yields a unique curve which terminates at, say, point event B.

One now asks: given that there exists a unique curve through A and B, does there exist a principle which singles out this curve over any other – a principle with a mathematical starting point different from Newton's local equations of motion?

The answer is given by the principle of least (more generally, stationary) action, also called *Hamilton's Principle*:

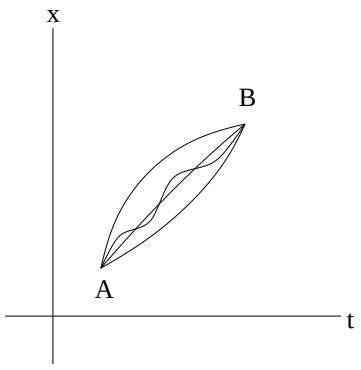


FIGURE 3.1.1. Optimal and non-optimal paths in spacetime

Of all possible paths along which a dynamical system may (be imagined to) move from one point to another within a specified time interval (consistent with any constraints), the actual path followed is that which minimizes the time integral of the difference between the kinetic and potential energies.

Supplementary Lecture: Why $\Delta(K.E. - P.E.)dt = minimum$?

EXAMPLE 1. In its simplest manifestations, Hamilton's principle is a mathematical expression of two fundamental physical aspects of nature: Newton's First Law of Motion and the Equivalence Principle. This can be seen from the following considerations:
A. Launch a particle vertically from x_1 at time t_1, watch it reach its maximum height, and then catch it at time t_2 at the instant it is located at x_2.

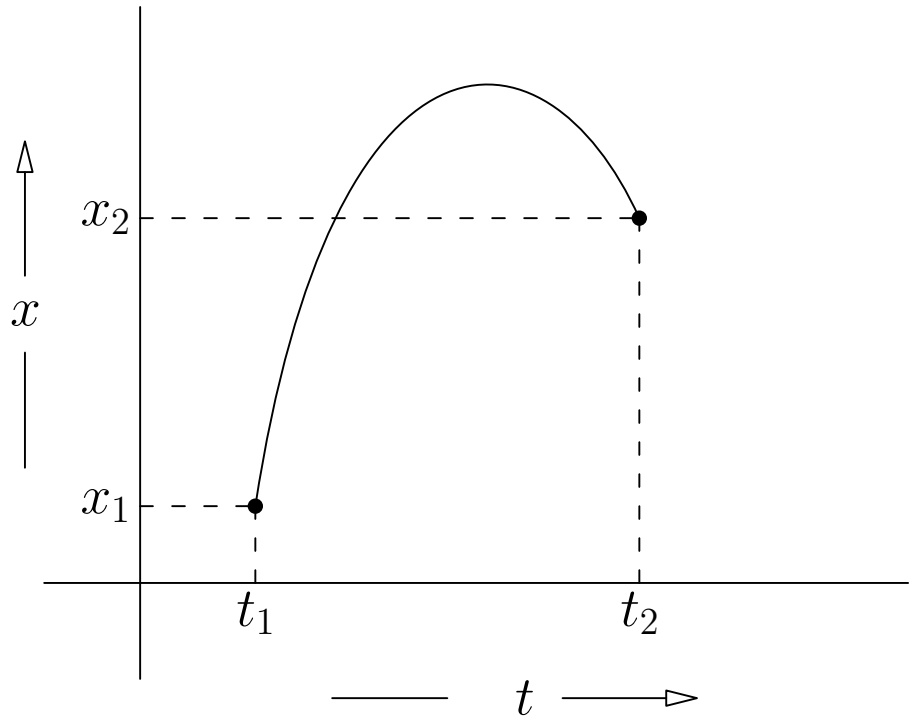


FIGURE 3.1.2. Spacetime trajectory of a particle thrown into the air.

From Galileo we learned that in its travel from (t_1, x_1) to (t_2, x_2) the particle traces a space-time trajectory which is given by a parabola. Why so? Answer:

(1) Newton's 1st Law: Every body continues in its state of rest, or of uniform motion in a straight line, unless it iscompelled to change that state by forces impressed upon it.
(2) The principle of equivalence.

A. Simpler case: Free Particle.
Consider the motion of a particle moving freely in a free float ("inertial") frame. This particle moves with constant velocity, i.e. its space-time trajectory is a straight line.

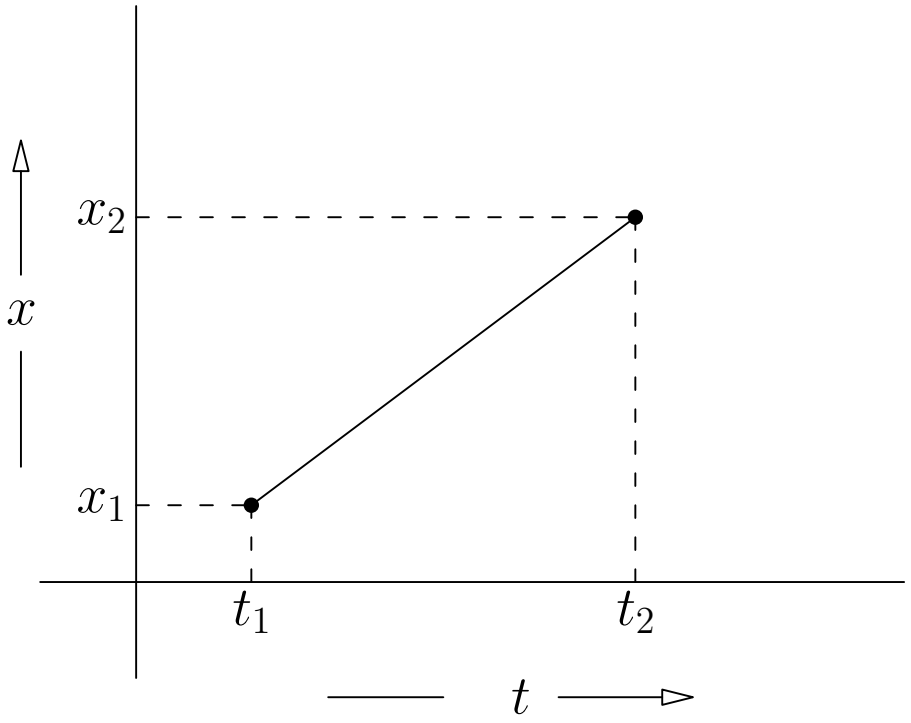


FIGURE 3.1.3. Spacetime trajectory of a free particle is a straight line.

The implication of this fact is that for such a curve the integral

$$\frac{1}{(t_2-t_1)}\,\Delta_{t_i}^{t_2}\left(\frac{dx(t)}{dt}\right)^2 dt \equiv \langle v^2\rangle = min\,!$$

as compared to other curves having the same starting and termination points.
Q: Why?

A: All such curves have the same average velocity,

$$\langle \bar{v}\rangle = \frac{1}{(t_2-t_1)}\,\Delta_{t_1}^{t_2}\frac{d\bar{x}}{dt}dt = \frac{1}{(t_2-t_1)}\,\Delta_{t_1}^{t_2}\frac{dx}{dt}dt = \langle v\rangle = v,$$

which means that the area under the curves $\bar{v}(t)$and $v(t) = v$ are the same.

Applying this fact to the positivity of the averaged squared deviation (away from the average),

$$0 \leqslant \langle(\bar{v}-\langle\bar{v}\rangle)^2\rangle = \langle\bar{v}^2\rangle - (\langle\bar{v}\rangle)^2 = \langle\bar{v}^2\rangle - (\langle v\rangle)^2\,,$$

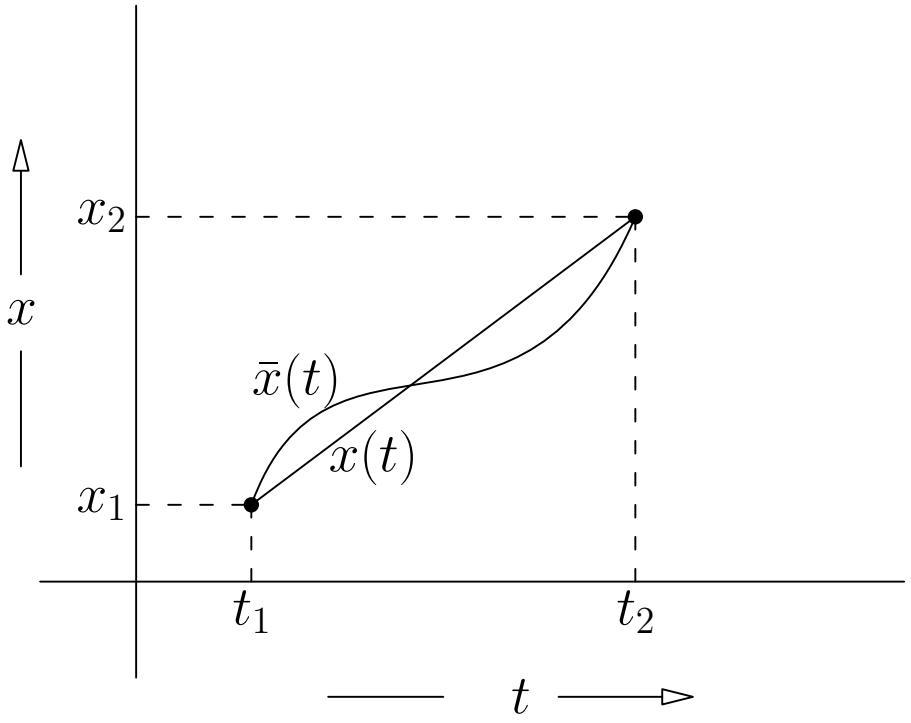


FIGURE 3.1.4. Straight line $x(t)$ and its variant $\bar{x}(t)$ have the same average velocity: $\langle \bar{v} \rangle = v \, (= const.)$.

one has

$$\langle \bar{v}^2 \rangle \geqslant \left(\langle v \rangle \right)^2 = v^2,$$

or

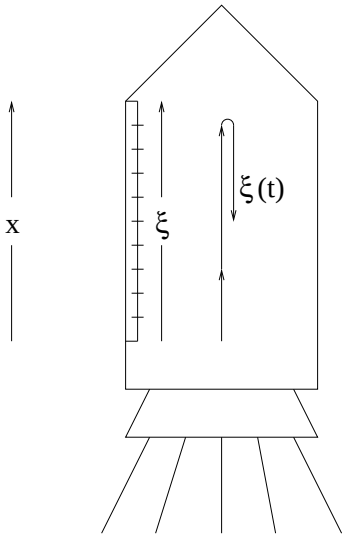
$$\Delta_{t_i}^{t_2} \left(\frac{d\bar{x}(t)}{dt} \right)^2 dt \geqslant \Delta_{t_i}^{t_2} \left(\frac{dx(t)}{dt} \right)^2 dt.$$

This says that a free particle moves so that the integral of its kinetic energy is a minimum:

$$\Delta_{t_i}^{t_2} K.E.\, dt \equiv \Delta_{t_i}^{t_2} \frac{1}{2} m \left(\frac{dx(t)}{dt} \right)^2 dt = min\,!$$

B. Free particle in an accelerated frame.

Consider the motion of the same particle moving freely in a frame accelerated uniformly with acceleration g.



A point ξ fixed in the accelerated frame will move relative to the free float frame according to

$$x = \xi + \frac{1}{2}gt^2.$$

It follows that, relative to the accelerated frame, the spacetime trajectory of the particle, $\xi(t)$, is given by

$$x(t) = \xi(t) + \frac{1}{2}gt^2. \tag{3.1.1}$$

Here $x(t)$ is the linear trajectory in Figure 3.1.3.

The to-be-minimized integral takes the form

$$\begin{aligned} min = \Delta_{t_i}^{t_2}\left(\frac{dx(t)}{dt}\right)^2 dt &= \Delta_{t_i}^{t_2}\left(\frac{d\xi}{dt} + gt\right)^2 dt \\ &= \Delta_{t_i}^{t_2}\left\{\left(\frac{d\xi}{dt}\right)^2 + 2gt\frac{d\xi}{dt} + g^2t^2\right\} dt \\ &= \Delta_{t_i}^{t_2}\left\{\left(\frac{d\xi}{dt}\right)^2 - 2g\xi\right\} dt + 2gt\xi|_{t_1}^{t_2} + \frac{1}{3}gt^2|_{t_1}^{t_2} \end{aligned}$$

The last line is the result of an integration by parts. The last two terms are the same for all trajectories passing through the given points (t_1, x_1) and (t_2, x_2). Consequently,

$$\Delta_{t_i}^{t_2}\frac{1}{2}m\left(\frac{dx(t)}{dt}\right)^2 dt = min \iff \Delta_{t_i}^{t_2}\left\{\frac{m}{2}\left(\frac{d\xi}{dt}\right)^2 - mg\xi\right\} dt = min$$

C. Free particle in an *equivalent* gravitational field.

The *equivalence principle* is an observation of the fact that in an accelerated frame the laws of moving bodies are the same as those in a homogeneous gravitational field.

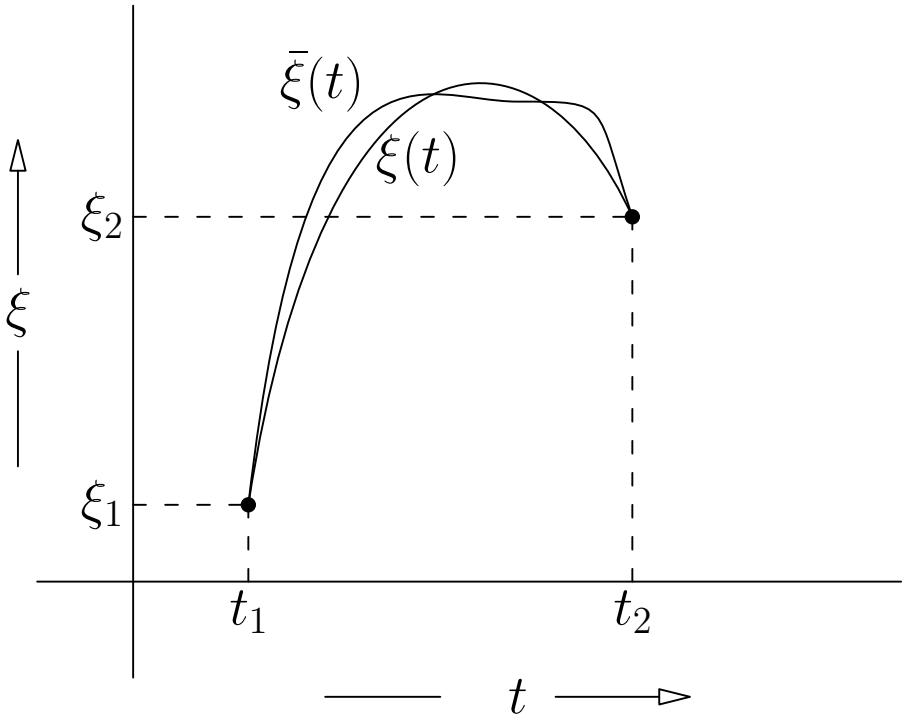


FIGURE 3.1.5. Minimizing trajectory $\xi(t)$ and one of its variants $\bar{\xi}(t)$.

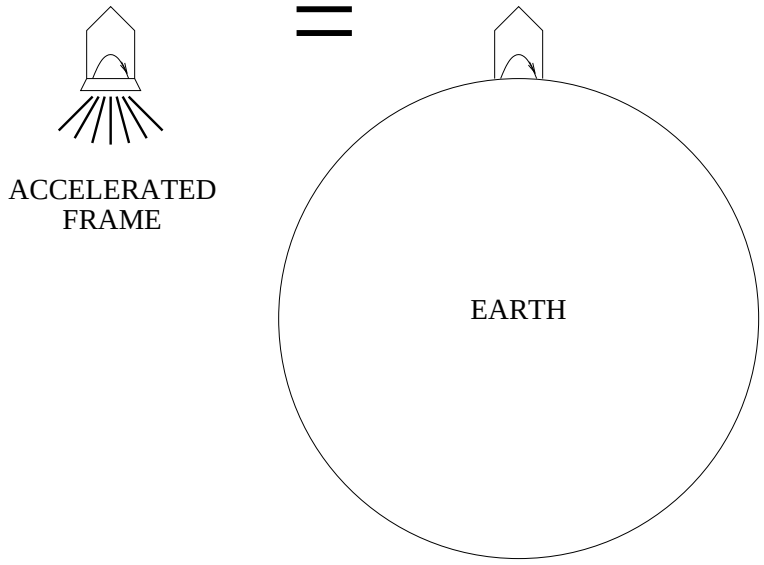


FIGURE 3.1.6. Trajectories in an accelerated frame are indistinguishable from those in a gravitational field. In particular the motion of particles of different composition (gold, aluminum, snakewood, etc.) is independent of their composition.

Recall that in a gravitational field

$$mg\xi = P.E.$$

represents the potential energy of a mass m at a height ξ. Consequently, the trajectory of a particle in a gravitational field is determined by

$$\Delta_{t_i}^{t_2}(\, K.E. \, - P.E.\,)\, dt \equiv \Delta_{t_i}^{t_2} L(\dot{x},\, x,\, t)\, dt = min.$$

In fact, the trajectory of a particle which satisfies this minimum condition satisfies the Euler-Lagrange

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x},$$

which is Newton's second law of motion

$$ma = F$$

for the one-dimensional motion of a particle.
Nota bene:

(1) The same minimum priciple hold even if g, and hence the potential energy $P.E.$, depends explicitly on time.
(2) This principle is a special case of what is known as Hamilton's principle of least action. The difference is that the latter also accomodates motion which are subject to constraints.

D. Extension to multi dimensions and generic potentials.

The Lagrangian formulation opens new vistas on the notion of bodies. It can be fruitfully implemented for more general motions and potentials. These generalizations are alternate but equivalent formulations of Newtonian mechanics. They are simply expressed by the statement that

$$\Delta_{t_i}^{t_2}(\, K.E. \, - P.E.\,)\, dt = min$$

with

$$K.E. = \frac{1}{2}\sum_{i=1}^{n} m_i \dot{\vec{x}}_i \cdot \dot{\vec{x}}_i$$

$$P.E. = U(t,\, \vec{x}_i)$$

on the class of all system trajectories having fixed endpoints.

The advantage of Lagrangian Mechanics becomes evident in the process of setting up Newton's equations of motion. In Newtonian Mechanics one must do this for each force component separately, a task which becomes non-trivial relative to curvilinear coordinate frames (spherical, cylindrical, etc.). By contrast, in the Lagrangian approach one merely identifies the two scalars $K.E$ and $P.E.$ relative to the coordinate frame of one's choice. The remaining task of setting up the differential equations of motion is done automatically by merely writing down the Euler-Lagrange equations.

In terms of the calculus of variations, the principle of stationary action is simply

$$\Delta_B^A(T - U)dt = \text{extremum!}$$

Here $T(\dot{x}^i)$ is the kinetic energy (*K.E.*) and $U(x^i, t)$ is the potential energy (*P.E.*). The integrand $T(\dot{x}^i) - U(x^i, t) \equiv L(\dot{x}^i, x^i, t)$ is the *Lagrangian* of the system. Thus the stationary action principle can be stated as

$$\delta \, \Delta_A^B \, L(x^i, \dot{x}^i, t) dt = 0 \, .$$

End of Supplementary Lecture

Hamilton's principle lends itself to a treatment within the frame work of the calculus of variations if one identifies the action with the variational integral

$$J[y_i] = \Delta_{x_0}^{x_1} F(x, y_i, y_i') dx$$

By making the identificaiton

$$\begin{aligned} x &\rightarrow t \\ y_i(x) &\rightarrow x^i(t) \\ y_i'(x) &\rightarrow \dot{x}^i(t) \\ F(x, y_i, y_i^{'}) &\rightarrow L(x^i, \dot{x}^i, t) \end{aligned}$$

one obtains the Euler-Lagrange equations,

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} = 0 \, .$$

These equations of motion reproduce the Newtonian equations of motion. This fact is illustrated by means of the following example. For n particles, each with mass m_i one has relative to cartesian coordinates (x,y,z)

$$T = \frac{1}{2} \sum_{i=1}^{n} m_i \left(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2 \right)$$

and

$$U = U(t, x_1, y_1, z_1, \cdots, x_n, y_n, z_n) \, .$$

The Euler-Lagrange equations are

$$\begin{aligned} -\frac{\partial U}{\partial x_i} &= m_i \ddot{x}_i \\ -\frac{\partial U}{\partial y_i} &= m_i \ddot{y}_i \\ -\frac{\partial U}{\partial z_i} &= m_i \ddot{z}_i \qquad i = 1, 2, 3 \end{aligned}$$

Thus one has $3n$ differential equations for $3n$ unknown functions.
These equations are Newton's equations of motion for a system of n particles because

$$\left(-\frac{\partial U}{\partial \dot{x}_i}, -\frac{\partial U}{\partial \dot{y}_i}, -\frac{\partial U}{\partial \dot{z}_i} \right) = \overrightarrow{Force} \text{ on the ith particle} \, .$$

The advantage of the variational formulation of mechanics via Hamilton's principle over Newton's formulation is twofold:

(1) From the viewpoint of physics it opens up new vistas. In particular, Hamilton's principle is a *bridge to quantum mechanics.* In fact, it is quantum mechanics which sanctions and justifies Hamilton's principle of least action. We shall return to that point later.

(2) From the viewpoint of mathematics this formulation of mechanics constitutes an enormous streamlining, both in setting up and in solving the equations of motion for complex Newtonian dynamical systems.
   (a) Setting up the equations has been reduced to constructing a single scalar function, the Lagrangian, and its derivatives.
   (b) Solving the resulting system of differential equations is greatly facilitated by the flexibility inherent in a scalar function. The function, and hence the system of equations, can be exhibited relative to any system of coordinates one chooses, including those relative to which the equations are so simple that they can be readily analysed, if not solved.

The conceptual unit economy, and hence the technical advantage and the mathematical power of Hamilton's principle, arises from the fact that the Lagrangian is a single scalar

$$L = K.E. - P.E.\,,$$

and from the fact that it may be expressed in terms of virtually any type of coordinates. These "generalized coordinates"

$$q^1(t), q^2(t), \cdots, q^s(t)$$

and their time derivatives

$$\dot{q}^1(t), \dot{q}^2(t), \cdots, \dot{q}^s(t)$$

characterize the state of motion of the mechanical system. Quite often they are (but need not be) related to, say, the cartesian coordinates, $\vec{x} = (x, y, z)$ by

$$\vec{x}_i = \vec{x}_i(q^1, \cdots, q^s, t) = \vec{x}_i(q^j, t)$$

and their time derivative

$$\dot{\vec{x}}_i = \dot{\vec{x}}_i(\dot{q}^j, q^k, t)\,.$$

In any case, in terms of generalized coordinates Hamilton's principle becomes

$$\delta\, \Delta_{t1}^{t_2}\, L(q^j, \dot{q}^j, t) = 0\,.$$

The associated Euler-Lagrange equations of motion are

$$\frac{\partial L}{\partial q^j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j}\,. \tag{3.1.2}$$

The ease with which the equations of motion can be set up is illustrated by the following simple

Example. Consider a disk rolling down a plane of lenth ℓ inclined at an angle α relative to the horizontal . The task of setting up the equations for the system is archetypical:

(1) The total kinetic energy breaks up into its translational and its rotational part (around the center of mass)

$$\begin{aligned} T &= \frac{1}{2}M\dot{y}^2 + \frac{1}{2}I\dot{\theta}^2 \\ &= \frac{1}{2}M\dot{y}^2 + \frac{1}{4}MR^2\dot{\theta}^2 \end{aligned}$$

Here $I = \frac{1}{2}MR^2$ is the moment of inertia of the disc whose radius is R and whose mass is M.

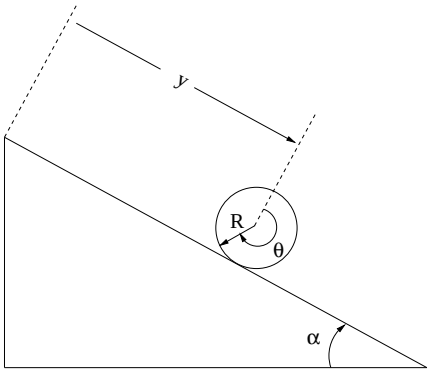


FIGURE 3.1.7. Disk rolling down an inclined plane

(2) The potential energy is

$$U = Mg\,(\ell - y)\,\sin(\alpha)$$

so that U=0 at the bottom ($y = \ell$) of the inclined plane.

(3) The Lagrangian is

$$L = T - U = \frac{1}{2}M\dot{y}^2 + \frac{1}{4}MR^2\dot{\theta}^2 + Mg(y-\ell)\sin(\alpha)\,.$$

(4) The equation of constraint between tranlation and rotation is

$$G(y,\theta) = y - R\theta = 0\,.$$

(5) Finally, the Euler-Lagrange equations are

$$\begin{aligned}\frac{d}{dt}\frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} &= \lambda(t)\frac{\partial G}{\partial y}\\ \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} &= \lambda(t)\frac{\partial G}{\partial \theta}\end{aligned}\qquad .$$

Explicitly, they are

$$M\ddot{y} - Mg\,\sin\alpha = \lambda \tag{3.1.3}$$

$$\frac{1}{2}MR^2\ddot{\theta} = -\lambda R\,. \tag{3.1.4}$$

The constraint is

$$y = R\theta\,. \tag{3.1.5}$$

By differentiating the last equation, using the result to rewrite the antecedent equation, and introducing the resultant expression for λ, these three equations yield

$$\begin{aligned}\ddot{\theta} &= \frac{\ddot{y}}{R}\\ \lambda &= -\frac{1}{2}M\ddot{y}\\ \ddot{y} &= \frac{2g\,\sin\alpha}{3}\,.\end{aligned}$$

Consequently,

$$\begin{aligned} \lambda &= \frac{-Mg\,\sin\alpha}{3} \\ \ddot{\theta} &= \frac{2g\,\sin\alpha}{3R}\,. \end{aligned}$$

REMARK. Note that the introduction of the Lagrange multiplier λ into the variational principle reveals physical aspects of the mechanical system which would have stayed hidden without that multiplier. In this example we see that on the right hand side of Eq.(3.1.3), λ is a *constraint force*, and on the right hand side of Eq.(3.1.4), $-\lambda R$ is a constraint torque. These constraint-induced quantities (i.e. λ and -λR) are non-zero in order to guarantee that the constraint, namely, *no slippage*, stays enforced. In particular, we see that λ is a constraint force along the y-direction. Furthermore, we see that -λR in Eq.(10.8b) is a positive torque. This guarantees that the disk spins up in the positive θ direction.

3.2. Hamilton-Jacobi Theory

Lecture 11

Newton created mechanics and formulated it in terms of his differential laws of motion. Using Hamilton's Principle, Lagrange streamlined Newton's formulations by giving it an innovative mathematical formulation which is based on a single scalar, the Lagrangian. As a consequence, there is an infinite flexibility in describing the mechanical system; in other words, the equations of motion can be exhibitied relative to any chosen set of generalized coordinates.

Hamilton-Jacobi theory, which also springs forth from Hamilton's principle, constitutes an additional step forward in physics as well as in mathematics. The advance in mathematics consists of the fact that H-J theory constitutes a powerful and systematic way of integrating the equations of motion of mechanics. The advance in physics consists of the fact that H-J theory provides a bridge between classical mechanics and quantum mechanics, as well as between ray optics and electromagnetic wave theory (wave optics).

It is difficult to point to a mathematical theory which brings into our grasp a larger and more diverse number of aspects of nature than H-J theory. No wonder many physicists express their admiration and enthusaism for H-J theory in a way which borders on the poetic. For example, Cornelius Lanczos in his book "The Variational Principles of Mechanics" (1949), writes: "We now are in the rarified atmosphere of theories of excessive beauty and we are nearing a high plateau on which geometry, optics, mechanics, and wave mechanics meet on common ground. Only concentrated thinking, and a considerable amount of re-creation, will reveal the full beauty of our subject in which the last words have not yet been spoken. We start with the integration theory of Jacobi and continue with Hamilton's own investigations in the realm of goemetrical optics and mechanics. The combination of these two approaches leads to de Broglie's and Schroedinger's great discoveries, and we come to the end our journey."

3.3. The Dynamical Phase

The equations of motion of a dynamical system are obtained from Hamilton's principle, namely the require ment that the action functional

$$I = \Delta_{x',t'}^{x,t} L\left(x, \frac{dx}{dt}, t\right) dt \tag{3.3.1}$$

be an extremum for the actual path along which the system travels in space and in time. This actual path we call an *extremal* path. Suppose we consider only extremal paths, in particular those that have the same common starting point (x', t') in space and time. The termination point of each extemal path is different, and we designate it by the variable point (x, t). Thus an extremal path is characterized by

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0$$

and the two end point conditions

$$\begin{aligned} x(t') =& x' \\ x(t) =& x\,. \end{aligned}$$

For each one of these extremal paths the action integral has a (unique) extemal value, namely

$$\Delta_{x',t'}^{x,t} L\left(x, \frac{dx}{dt}, t\right) dt = I_{extremum}\,.$$

For each endpoint (x, t) of an extemal path there is number $I_{extremum}(x, t)$. Thus we have a real valued function which maps (x, t) into the reals:

$$S(x, t) \equiv I_{extemum}(x, t)\,.$$

The level surfaces (isograms) of this scalar function are easy to constuct. Suppose we wish to constuct the isogram

$$S(x, t) = 17.1\,.$$

We consider the set of extemal paths emanating from the common point (x', t'). At that point $S(x', t') = 0$. We now move along a particular extemal path and evaluate the action integral, Eq.(3.3.1), along this path. When the value of this integral is $I = 17.1$, we stop. This stopping point is a point on the isogram $S(x, t) = 17.1$.

The function defined by this construction is called the "dynamical phase", or the "action" of the system at (x, t). It also is called the "eikonal" in optics. This function is easy to remember because in wave mechanics it is the phase of the system. In fact, we assert[1] that its relation to the wave function of the system is give by

$$\Psi(x, t) = \mathcal{A}e^{iS(x,t)/\hbar}.$$

Here $\hbar$ is a constant which guarantees that the exponent is dimensionless. For a mechanical system this constant is Planck's constant (6.1×10^{-27}erg sec) divided by 2π. The wave function Ψ satisfies a wave equation. Whether this wave equation is the Schroedinger equation of the system, the scalar wave equation for a light ray, or some other linear wave equation is of great importance, but of no concern to us at present. For us, our present interest lies only in the dynamical phase $S(x, t)$,

(1) in the equation for $S(x, t)$ (Section 3.5) and

[1]The validity of this assertion depends on the conceptualization of experimental and observational evidence.

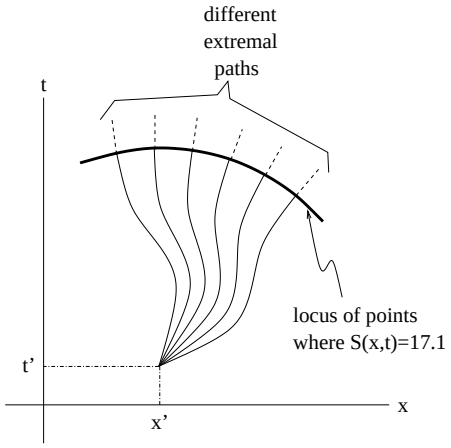


FIGURE 3.3.1. Isogram of the dynamial phase S(x,t)

(2) in the reconstruction of the extremal paths (Sections 3.6 and 3.11) from $S(x,t)$.

The physical perspective on our reconstruction of the extremal paths is very different from that of Newton. Whereas his focus is on the particle aspects of bodies, ours is on their wave aspects. In spite of that difference our wave perspective started with Newton. It grew by a well-defined process of inductive reasoning consisting of Newton →Euler-Lagrange →Hamilton →Jacobi + experimental observations about the dynamical behavior of matter – observations which were performed during the first half of the twentieth century. In spite of the great conceptual distance between the particle and the wave perspectives, the former is a correspondence (asymptotic, shortwave length) limit of the latter. This fact is developed in Sections 3.6-3.11 and summarized in Figure 3.6.3 and 3.8.1 on pages 52 and 55 respectively.

3.4. Momentum and the Hamiltonian

We now proceed with the process of deriving the partial differential equation which the scalar function $S(x,t)$ satisfies. In this process we shall repeatedly encounter two quantities. They are (i) the *generalized momentum* and (ii) the *Hamiltonian* of the system. Specifically we let

$$\frac{\partial L}{\partial \dot{q}^i} \equiv p_i \qquad i = 1, ...s \tag{3.4.1}$$

be the ith generalized momentum. Its importance follows from the equation of motion. If the Lagrangian is independent of the ith generalized coordinate q^i, i.e.

$$\frac{\partial L}{\partial q^i} = 0$$

then the Euler-Lagrange equations of motion demand that

$$\frac{dp_i}{dt} = 0\,.$$

In other words, the ith generalized momentum of the system does not change; this *momentum is "conserved"*.

EXAMPLE. Consider a system of n particles in a potential. The Lagrangian is

$$L = \frac{1}{2}\sum_{j=1}^{n} m_j \left(\dot{x}_j^2 + \dot{y}_j^2 + \dot{z}_j^2\right) - U\left(x_j, y_j, z_j, t\right) .$$

The x-component of the jth particle *momentum* is

$$\frac{\partial L}{\partial \dot{x}_j} = m\dot{x}_j \equiv p_{xj} .$$

The other repeatedly occuring quantity is

$$\sum_{i=1}^{s} \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} - L \equiv H , \tag{3.4.2}$$

the *Hamiltonian* of the system. This quantity is conserved if the Lagrangian of the system has no explicit time dependence. Indeed, differentiating each term of that expression for the Hamiltonian and keeping in mind that L depends on q^i, $\dot{q}^i$, and t, one obtains

$$\frac{dH}{dt} = \sum_{i=1}^{s}\left[\ddot{q}^i \frac{\partial L}{\partial \dot{q}^i} + \dot{q}^i \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i\right] - \frac{\partial L}{\partial t} .$$

Here the two middle terms cancel because of the E-L equations of motion. Consequently, we are left with

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} . \tag{3.4.3}$$

We conclude that if $\frac{\partial L}{\partial t} = 0$, then

$$H = constant$$

along the extremal path of the system.

REMARK. If the system has only a single degree of freedom ($s = 1$) then $H = \dot{q}\frac{\partial L}{\partial \dot{q}} - L$, and Eq.(3.4.3) is simply the "2nd form of the Euler equation", which was considered at the end of Chapter I.

The generalized momenta p_j , $j = 1, \cdots, s$ as well as the Hamiltonian H will play the key roles in establishing the partial diffential equation which the dynamical phase $S(q^1, q^2, \cdots, q^s, t)$ must satisfy. In fact, in order to obtain this equation, we shall find that one must reexpress the Hamiltonian, Eq.(3.4.2), in terms of the generalized momenta defined by Eq.(3.4.1) in such a way that the Hamiltonian has no functional dependence on the generalized velocities $\dot{q}^j$, $j= 1,\cdots,s$. This goal is achieved by solving the defining equations (3.4.1) for the generalized velocities

$$\dot{q}^i = \dot{q}^i(q^k, p_j)$$

in terms of the momenta, and then introducing these expressions into the Hamiltonian:

$$\sum_{i=1}^{s} \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} - L = H(q^k, p_j, t) . \tag{3.4.4}$$

To verify that H does indeed depend only on q^k and p_j, but not on $\dot{q}^i$, we observe that[2]

$$\begin{aligned} dH &= d\left(\dot{q}^i p_i - L\right) \\ &= d\dot{q}^i p_i + \dot{q}^i dp_i - \frac{\partial L}{\partial q^i} dq^i - \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i - \frac{\partial L}{\partial t} dt \,. \end{aligned}$$

The first and the fourth term cancel with the help of Eq.(3.4.1). What remains is

$$dH = \dot{q}^i dp_i - \frac{\partial L}{\partial q^i} dq^i - \frac{\partial L}{\partial t} dt \,.$$

There are two benefits. First, this shows that, with the help of the defining relation (3.4.1), the Hamiltonian is functionally independent of $\dot{q}^j$, and depends only on q^k and p_j instead. Second, by comparing this expression with the differential of Eq.(3.4.4),

$$dH = \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial t} dt \,,$$

by using the E-L equations of motion (3.1.2),

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt} p_i \,,$$

and by taking advantage of the linear independence of the differentials $\{dp_i, dq^i, dt\}$, one obtains the three equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}, \quad \text{and } \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \,.$$

These are *Hamilton's equations of motion.* They govern the dynamics of the mechanical system. Its motion is represented by the trajectory of a moving point $\{q^i(t), p_i(t)\}_1^s$ in the $2s$-dimensional space spanned by $\{q^i, p_i\}$. The rate of change of the Hamiltonian function $H(q^i, p_i, t)$ along the trajectory is

$$\begin{aligned} \frac{dH}{dt} &= \frac{\partial H}{\partial q^i}\dot{q}^i + \frac{\partial H}{\partial p_i}\dot{p}_i + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q^i} + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial t} \,. \end{aligned}$$

Thus, if t is a cyclic coordinate, i.e. the Hamiltonian is manifestly independent of time ($\partial H/\partial t = 0$), then H is evidently a constant of motion.

The Hamiltonian equations of motion are equivalent to those of Euler-Lagrange. The important difference is that the former is a system of *2s first* order equations, while the latter is one of *s second* order equations.

EXAMPLE. (*Particles in the Field of a Potential)*
The Lagrangian is

$$L = \sum_{i=1}^{n} \frac{1}{2} m_i \vec{\dot{x}}_i \vec{\dot{x}}_i - U(\vec{x}_k, t) \,.$$

[2]In the interest of conceptual unit economy we are introducing the Einstein summation convention according to which a summation is implied whenever a pair of repeated (summation!) indices occur in a product expression. For example, $\sum_{i=1}^{s} \dot{q}^i p_i \equiv \dot{q}^i p_i$.

For each of the particles the momentum is

$$\vec{p}_i = (p_{xi},\, p_{yi},\, p_{zi}) = m_i \dot{\vec{x}}_i \,. \tag{3.4.5}$$

The Hamiltonian is

$$\begin{aligned} H &= \sum_j \dot{q}^j \frac{\partial L}{\partial q_j} - L \\ &= \sum_{i=1}^{n} (\dot{x}_i m_i \dot{x}_i + \dot{y}_i m_i \dot{y}_i + \dot{z}_i m_i \dot{z}_i) - \frac{1}{2} \sum_{i=1}^{n} (\dot{x}_i m_i \dot{x}_i + \dot{y}_i m_i \dot{y}_i + \dot{z}_i m_i \dot{z}_i) + U(\vec{x}_k, t) \\ &= \frac{1}{2} \sum_{i=1}^{n} m_i \dot{\vec{x}}_i \cdot \dot{\vec{x}}_i + U(\vec{x}_k, t) \\ &= K.E. + P.E. = \text{Total Energy} \,. \end{aligned}$$

With the momentum defined by Eq. (3.4.5), the Hamilitonian assumes the form

$$H(\vec{p}_i, \vec{x}_i, t) = \sum_{i=1}^{n} \frac{\vec{p}_i \cdot \vec{p}_i}{2m_i} + U(\vec{x}_k, t) \,,$$

and the two sets of Hamiltonian equations of motion are

$$\begin{aligned} \frac{d\vec{x}_i}{dt} &= \frac{\vec{p}_i}{m} \\ \frac{d\vec{p}_i}{dt} &= -\vec{\nabla}_i \, U(\vec{x}_k, t) \,. \end{aligned}$$

If the potential energy is time-independent and hence depends only on the position $\vec{x}_i$ of the particles, then H is cyclic in t:

$$0 = \frac{\partial H}{\partial t} \,.$$

Consequently,

$$H = T.E. = \text{constant} \,.$$

Thus the total energy of a system of particles moving in a time independent potential is a constant of motion, a conclusion one also infer directly from the fact that the Lagrangian does not contain any explicit dependence on time t.

Remark. That the Hamiltonian is the total energy of the system is true in a non-rotating coordinate frame. However, in a rotating coordinate frame the Hamiltonian is not equal to the system's total energy. However, if the Lagrangian has no explicit time dependence ("time invariant system") then the Hamiltonian is conserved nevertheless. To illustrate the issue, consider the above example of particles-in-a-potential problem from the perspective of a rotating frame, relative to which one achieves substantial simplifcation in the analysis of charged particles circulating in a magnetic field or of satellites in an earth-centered reference frame.

Problem. (*Particle motion in a rotating frame)*
Given a frame (x', y', z') which rotates with constant angular frequency ω relative to a given inertial frame (x,y,z),

$$\begin{aligned} x &= x' \cos\omega t - y' \sin\omega t \\ y &= x' \sin\omega t + y' \sin\omega t \\ z &= z' \end{aligned}$$

determine for a single particle in a rotationally symmetric potential $U\left(x^2+y^2,\, z\right)$ (i) its Lagrangian, (ii) its Hamiltonian relative the rotating frame, and then (iii) compare the two Hamiltonians in the two frames.
Solution. The Lagrangian of a particle moving in the potential U is

$$L = \frac{m}{2}\left(\dot{x}^2+\dot{y}^2+\dot{z}^2\right) - U\left(x^2+y^2,\, z\right)\,.$$

Relative to the rotating frame this Lagrangian,

$$L = \frac{m}{2}\left(\dot{x}'^2+\dot{y}'^2+\dot{z}'^2\right) - U\left(x'^2+y'^2,\, z'\right) + \frac{m}{2}\omega^2\left(x'^2+y'^2\right) + \omega m\left(x'\dot{y}'-y'\dot{x}'\right),$$

has acquired an additional centrifugal potential energy, and a velocity dependent potential energy. In the rotating frame the Hamiltonian is

$$\begin{aligned} H' &= \dot{x}'\frac{\partial L}{\partial \dot{x}'} + \dot{y}'\frac{\partial L}{\partial \dot{y}'} + \dot{z}'\frac{\partial L}{\partial \dot{z}'} - L \\ &= \frac{m}{2}\left(\dot{x}'^2+\dot{y}'^2+\dot{z}'^2\right) + U\left(x'^2+y'^2,\, z'\right) - \frac{m}{2}\omega^2\left(x'^2+y'^2\right)\,. \end{aligned}$$

Introducing the defining relations, Eq.(3.4.1)

$$\begin{aligned} p'_x &= m\dot{x}' - m\omega y' \quad &\to \dot{x}' = \frac{p'_x}{m} + \omega y' \\ p'_y &= m\dot{y}' + m\omega x' \quad &\to \dot{y}' = \frac{p'_y}{m} - \omega x' \\ p'_z &= m\dot{z}' \quad &\to \dot{z}' = \frac{p'_z}{m}\,, \end{aligned}$$

one finds that this Hamiltonian is

$$H' = \frac{(p'_x)^2}{2m} + \frac{(p'_y)^2}{2m} + \frac{(p'_z)^2}{2m} + U\left(x'^2+y'^2,\, z'\right) - \omega\left(x'p'_y - y'p'_x\right).$$

The particle's inertial Hamiltonian

$$\begin{aligned} H &= \frac{(p_x)^2}{2m} + \frac{(p_y)^2}{2m} + \frac{(p_z)^2}{2m} + U\left(x^2+y^2,\, z\right) \\ &= \text{Kinetic Energy} + \text{Potential Energy} \end{aligned}$$

expresses its total energy. It is an integral of motion. By contrast the particle's rotational Hamiltonian,

$$H' = (\text{Kinetic Energy})' + (\text{Potential Energy})' - \omega L_z\,,$$

even though it is also an integral of motion, does not express a mere total energy, kinetic plus potential in the rotating frame. There is an additional energy-like quantity $-\omega L_z$ which expresses a kind of interaction between (a) whatever aspect

(proportional to ω) is responsible for the curved meniscus in Newton's bucket experiment[3] and (b) the particle's moment of momentum (i.e. angular momentum),

$$\begin{aligned} L_z &= \left(x' p'_y - y' p'_x\right) \\ &= \left(x\, p_y - y\, p_x\right) \end{aligned}$$

It is evident that this quantity is the same in the rotating and the inertial frame. It therefore refers to a property independent of any rotational aspect of the frame. Furthermore, it is an integral of motion, a fact which is implied by Hamilton's equations of motion. Thus there are three integrals of motion: (i) H', (ii) (K.E.)$'$+(P.E.)$'$, (iii) L_z and any two imply the third.

3.5. The Hamilton-Jacobi Equation

We now return to the task of determining the equation that governs the dynamical phase for a system with a single degree of freedom. The extension to several degrees takes little effort by comparison.

3.5.1. Single Degree of Freedom. Consider a set of extremal paths emanating from event A, i.e. they all start at the same time t' at the same point x' in space. These paths being extremal, they all satisfy the same Euler-Lagrange equation. They also pass through the same starting event, but they have different initial velocities. We evaluate the "action" along each extremal path,

$$I_{extremum}(t, t') = \Delta_{t'}^{t}\, L\left(x(\lambda), \dot{x}(\lambda), \lambda\right) d\lambda\,.$$

Consider the locus of points where, on different extremals, the integral $I_{extreme}$ has the same value, say

$$I_{extremum} = I_{extremum.}(t, t')\,.$$

This locus of points forms the level surfaces of a function *S(x,t)* where

$$S(x,t) = \text{extremum value of } \Delta_{x',t'}^{x,t}\, L(x, \dot{x}, t) dt\,.$$

We shall now see that the partial derivatives of this scalar function are the momenta and the negative Hamiltonian (usually, but not always, the energy), i.e.

$$\delta S = p \delta x - H \delta t$$

We consider two *extemal* paths. They have the same starting point A. But their end points are respectively (x,t) and (x+δx,t+δt). The difference in the dynamical phase (action) of these endpoints is

$$\begin{aligned} \delta S &= \delta I_{extremum} \\ &= \Delta_A^{t+\delta t}\, L(x+h, \dot{x}+\dot{h}, t) dt - \Delta_A^t\, L(x, \dot{x}, t) dt \\ &= L\delta t + \Delta_A^t \left(\frac{\partial L}{\partial x} h + \frac{\partial L}{\partial \dot{x}} \dot{h}\right) dt \\ &= L\delta t + \frac{\partial L}{\partial \dot{x}} h + \Delta_A^t \left(\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right) h\, dt\,. \end{aligned}$$

[3]Newton considered a water-filled bucket rotating (around its axis) relative to a free float (i.e. inertial) frame – a frame relative to which all free bodies obey Newton's first law of motion, but a frame which today we recognize as having the additional feature of not rotating relative to the fixed stars. The rotating frame associated with this bucket is characterized by a single distinguishing feature, its angular velocity ω. In this frame the surface of unmoving water curves upward with increasing radius. This curvature is absent in the inertial frame, which is not rotating relative to the fixed stars.

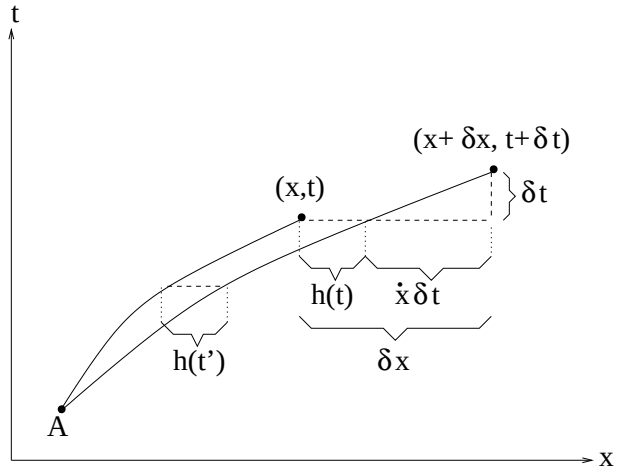


FIGURE 3.5.1. Two nearby extremal paths.

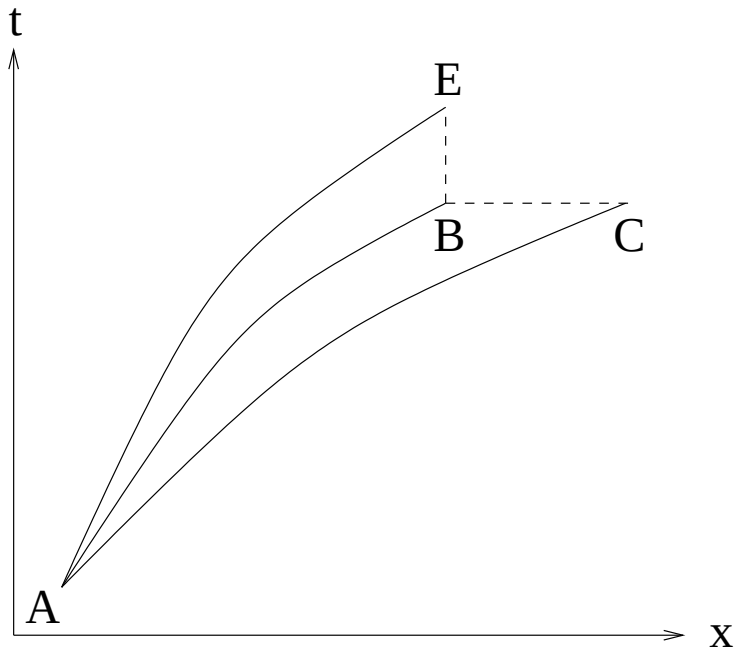


FIGURE 3.5.2. Three extremal curves

We only consider extremal curves. Thus $\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0$, and the integral vanishes. The variations δx and δt refer to the coordinate differences of arbitrarily located endpoint of the two extremals. But the variation $h(t)$ connects pairs of points having the same curve parameter t. Consequently, as one can see in Fig. 3.5.1,

$$h(t) = \delta x - \dot{x}\delta t \, .$$

Thus

$$\begin{aligned}\delta S &= \frac{\partial L}{\partial \dot{x}}\delta x - \left(\frac{\partial L}{\partial \dot{x}}\dot{x} - L\right)\delta t \\ &\equiv p\delta x - H\delta t\end{aligned}$$

An inquiry as to how S(x,t) changes into the x-direction, keeping t fixed, yields

$$\left(\begin{array}{c}\text{rate of change of}\\ \text{dynamical phase}\\ \text{with position}\end{array}\right) = (\text{momentum}) \tag{3.5.1}$$

$$= \lim_{\substack{C\to B\\ t=fixed}} \frac{\delta S}{\delta x} = \frac{\partial S}{\partial x} \tag{3.5.2}$$

$$= \frac{\partial L(x,\dot{x},t)}{\partial \dot{x}} = p\,. \tag{3.5.3}$$

Similarly,

$$-\left(\begin{array}{c}\text{rate of change of}\\ \text{dynamical phase}\\ \text{with time}\end{array}\right) = (\text{energy}) \tag{3.5.4}$$

$$= -\lim_{\substack{E\to B\\ x=fixed}} \frac{\delta S}{\delta t} = -\frac{\partial S}{\partial t} \tag{3.5.5}$$

$$= \frac{\partial L}{\partial \dot{x}}\dot{x} - L = H\,. \tag{3.5.6}$$

These two equations, yield the "dispersion relation" at (x,t). Indeed, in Eq. (3.5.6) replace the $\dot{x}$-dependence with a dependence on p by solving Eq. (3.5.3) for $\dot{x}$. The result is

$$(\text{energy}) = H(x,p,t)\,.$$

Finally introduce into this equation the partial derivatives calculated in Eqs. (3.5.2) and (3.5.5):

$$-\frac{\partial S}{\partial t} = H\left(x, \frac{\partial S}{\partial x}, t\right)\,.$$

This is the Hamilton-Jacobi equation, a single first order differential equation for the dynamical phase $S(x,t)$. This equation is a dispersion relation because it expresses how the time rate of change of the phase depends on its rate of change with position. For a system with several degrees of freedom the H-J equation is

$$H\left(q^i, \frac{\partial S}{\partial q^j}, t\right) + \frac{\partial S}{\partial t} = 0.$$

Lecture 12

PROBLEM. (*Particle in a potential*)

Set up and solve the Hamilton-Jacoby equation for a particle in a one dimensional potential $U(x)$.

Solution. Setting up the H-J equation is a three step process.

(1) Exhibit the Lagrangian:

$$L = \frac{1}{2}m\dot{x}^2 - U(x)\,.$$

(2) Determine the momentum and the Hamiltonian:

$$\begin{aligned} p &= \frac{\partial L}{\partial \dot{x}} \\ &= m\dot{x}\,; \\ H &= \dot{x}\frac{\partial L}{\partial \dot{x}} - L \\ &= \frac{1}{2}m\dot{x}^2 + U(x)\,. \end{aligned}$$

(3) Express the Hamiltonian in terms of the momentum:

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

(4) Write down the H-J equation $-\frac{\partial S}{\partial t} = H\left(x, \frac{\partial S}{\partial x}\right)$:

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + U(x)\,.$$

This a first order non-linear partial differential equation that needs to be solved for the scalar function $S(x,t)$.
This p.d.e. lends itself to being solved by the method of separation of variables according to which one finds solutions of the form

$$S(x,t) = T(t) + X(x)\,. \tag{3.5.7}$$

Introducing this form into the H-J equation, one finds

$$-\frac{dT(t)}{dt} = \frac{1}{2m}\left(\frac{dX(x)}{dx}\right)^2 + U(x)\,.$$

This equation says that the left hand side is independent of x, while the right hand side is independent of t. Being equal, the l.h.s. is also independent of x. Being independent of both t and x, it is a constant. Letting this "separation" constant be equal to E, one obtains two equations

$$\begin{aligned} -\frac{dT(t)}{dt} &= E \\ \frac{1}{2m}\left(\frac{dX(x)}{dx}\right)^2 + U(x) &= E\,. \end{aligned}$$

These are two ordinary equations for T and X. Inserting these equations into Eq. (3.5.7), one obtains the sought after solution to the H-J equation,

$$S(x,t) = -Et + \Delta^x \sqrt{2m\,(E - U(x'))}dx' + \delta(E)\,.$$

Here the "integration constant" $\delta(E)$ is an arbitrary function of E. Furthermore, observe that S depends on E also. This means that one has an E-parametrized family of solutions. Thus, properly speaking, separation of variables yields many solutions to the H-J equation, in fact, a one-parameter family of them

$$S(x,t) = S_E(x,t)\,.$$

3.5.2. Several Degrees of Freedom. We shall see in a subsequent section that whenever the H-J for a system with several degrees of freedom, say $\{q^i\}$, lends itself to being solved by the method of the separation of variables, i.e.

$$S(q^i, t) = T(t) + \sum_{i=1}^{s} Q_i(q^i)\,,$$

the solution has the form

$$S = -\,\Delta^t\, E dt + \sum_{i=1}^{s} \Delta^{q^i}\, p_i(x^i; E, \alpha_1, \cdots, \alpha_{s-1}) dq^i + \delta\,(E, \alpha_1, \cdots, \alpha_{s-1})$$

Here δ is an arbitrary function of E and the other separation constants that arise in the process of solving the H-J equation. We see that for each choice of $(E, \alpha_1, \cdots, \alpha_{s-1})$ we have a different solution S. Thus, properly speaking, we have $S_{E,\alpha_1,\cdots,\alpha_{s-1}}$, a multi-parametrized family of solutions to the H-J equation.

We shall now continue our development and show that Hamilton-Jacobi Theory is

a) A new and rapid way of integrating the E-L equations

b) The bridge to wave (also "quantum") mechanics.

The virtue of Hamilton's principle is that once the kinetic and potential energy of the system are known, the equations of motion can be set up with little effort. These Euler-Lagrange equations are Newton's equations of motion for the system. Although setting up the equations of motion for a system is a routine process, solving them can be a considerable challenge. This task can be facilitated considerably by using an entirely different approach. Instead of setting up and solving the set of coupled Newtonian ordinary differential equations, one sets up and solves a single partial differential equation for a single scalar function. Once one has this scalar function, one knows everything there is to know about the dynamical system. In particular, we shall see that by differentiating this scalar function (the dynamical phase, the Hamilton-Jacobi function, the eikonal) one readily deduces all possible dynamical evolutions of the system.

3.6. Hamilton-Jacobi Description of Motion

Hamilton-Jacobi theory is an example of the *principle of unit economy*[4], according to which one condenses a vast amount of knowledge into a smaller and smaller number of principles. Indeed, H-J theory condenses all of classical mechanics and all of wave mechanics (in the asymptotic high-frequency/short-wavelength (a.k.a. W.K.B.) approximation) into two conceptual units,, (i) the H-J equation and (ii) the principle of constructive interference. These two units are a mathematical expression of the fact that classical mechanics is an asymptotic limit of wave mechanics.

---

[4]The *principle of unit economy*, also known informally as the "crow epistemology", is the principle that stipulates the formation of a new concept

(1) when the description of a set of elements of knowledge becomes too complex,
(2) when the elements comprising the knowledge are used repeatedly, and
(3) when the elements of that set require further study.

Pushing back the frontier of knowledge and successful navigation of the world demands the formation of a new concept under any one of these three circumstances.

Hamilton thinking started with his observations of numerous known analogies between "particle world lines" of mechanics and "light rays" of geometric optics. These observations were the driving force of his theory. With it he developed classical mechanics as an asymptotic limit in the same way that ray optics is the asymptotic limit of wave optics. Ray optics is a mathematically precise asymptotic limit of wave optics. Hamilton applied this mathematical formulation to classical mechanics. He obtained what nowadays is called the Hamilton-Jacobi formulation of mechanics. Even though H-J theory is a mathematical limit of wave mechanics, in Hamilton's time there was no logical justification for attributing any wave properties to material particles. (That justification did not come until experimental evidence to that effect was received in the beginning of the 20th century.) The most he was able to claim was that H-J theory is a mathematical method with more unit economy than any other formulation of mechanics. The justification for associating a wave function with a mechanical system did not come until observational evidence to that effect was received in the beginning of the 20th century.

We shall take advantage of this observation (in particular by Davidson and Germer, 1925) implied association by assigning to a mechanical system a wave function. For our development of the H-J theory it is irrelevant whether it satisfies the Schroedinger, the Klein-Gordon, or some other quantum mechanical wave equation. Furthermore, whatever the form of the wave equation governing this wave function, our focus is only on those circumstances where the wave function has the form

$$\Psi_E(x,t) = \underbrace{\mathcal{A}(x,t)}_{\textit{slowly}\text{ varying function of } x \text{ and } t} \times \underbrace{\exp\left(\frac{i}{\hbar}S_E(x,t)\right)}_{\textit{rapidly}\text{ varying function of } x \text{ and } t} \tag{3.6.1}$$

This circumstance is called the "high frequency" limit or the "semi-classical" approximation. It can be achieved by making the energy E of the system large enough. In that case

$$1 << \frac{S_E(x,t)}{\hbar}$$

with the consequence that the phase factor oscillates as a function of x and t rapidly indeed. The existence of such a wave function raises a non-trivial problem:
If the wave and its dynamical phase, and hence the wave intensity, is defined over all of space-time, how is it possible that a particle traces out a sharp and well defined path in space-time when we are left with three delemas?

(1) The large magnitude ($S \gg \hbar$=1.05×10^{-27}[erg sec]) of the action for a classical particle is certainly of no help.
(2) Neither is the simplicity of the H-J equation

$$\frac{\partial S}{\partial t} + H(x, \frac{\partial S}{\partial x}, t) = 0$$

which governs the dynamical phase in

$$\Psi = \mathcal{A}\exp\left(i\,\frac{S}{\hbar}\right),$$

(3) Nor is the simplicity of the solution S for a particle of energy E,

$$S(x,t) = -Et + \Delta_{x_o}^{x}\sqrt{2m(E-U(x))}dx + \delta(E)$$

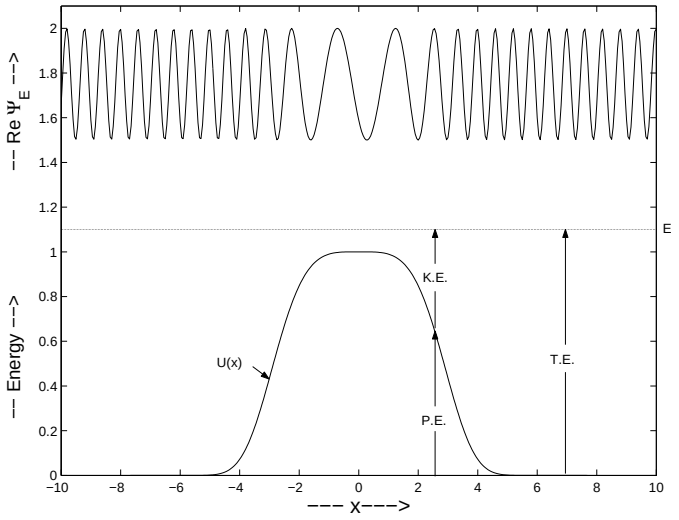


FIGURE 3.6.1. The spatial oscillation rate of the wave function $Re\,\Psi_E$ at $t = const.$ is proportional to its x-momentum, whose square is proportional to the kinetic energy (K.E.=T.E.-P.E.).

of any help in identifying a localized trajectory ("world line") of the particle in space-time coordinatized by x and t.

What *is* of help is the basic implication of associating a wave function with a moving particle, namely, it is a linear superposition of monochromatic waves, Eq. (3.6.1), which gives rise to a travelling wave packet – a localized moving wave packet whose history is the particle's world line. To validate this claim we shall give two heuristic arguements (i-ii), one application (iii), a more precise argument (iv) and an observation (v).

(i): The most elementary superposition monochomatic waves is given by the sum wave trains with different wavelengths

$$\Psi(x,t) = \Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots .$$

(ii): In space-time one has the following system of level surfaces for $S_E(x,t)$ and $S_{E+\Delta E}(x,t)$

Destructive interference between different waves comprising Ψ(x,t) occures everywhere except where the phase of the waves agree:

$$S_E(x,t) = S_{E+\Delta E}(x,t)$$

At the locus of events satisfying this condition, the waves interfere constructively and wave packet has non-zero amplitude. The quantum priciple says that this condition of constructive interference

$$0 = \lim_{\Delta E \to 0} \frac{S_{E+\Delta E}(x,t) - S_E(x,t)}{\Delta E} = \frac{\partial S_E(x,t)}{\partial E}$$

yields a Newtonian worldline, i.e. an extremal paths.

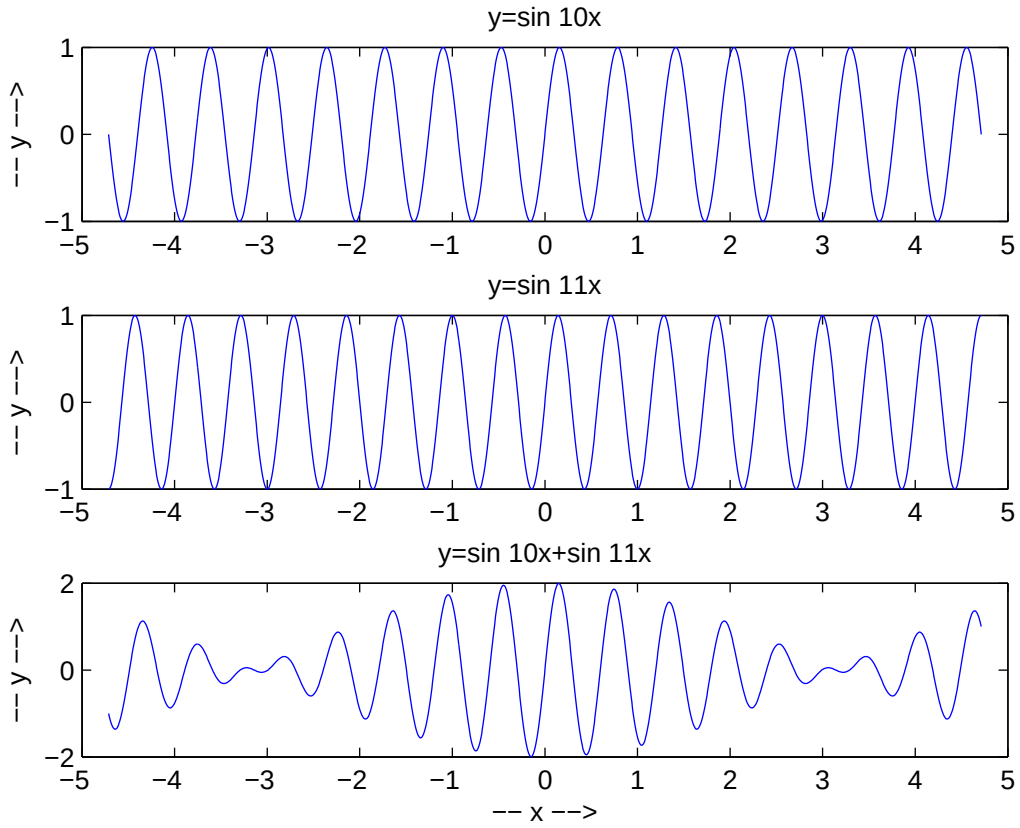


FIGURE 3.6.2. Photographic snapshot in space of two interfering wave trains and their resulting wave packet.

(iii): Apply this condition to the action S(x,t) of a single particle. One obtains the time the particle requires to travel to point x,

$$0 = -t + \Delta_{x_0}^{x} \sqrt{\frac{m}{2}} \left(\frac{1}{E - U(x)} \right)^{\frac{1}{2}} dx + t_0$$

with

$$t_0 \equiv \frac{\partial \delta(E)}{\partial E} \, .$$

This condition yields the Newtonian worldline indeed. The precise argument is Lecture 13. The additional observation is on p13 Lecture 13.

Lecture 13

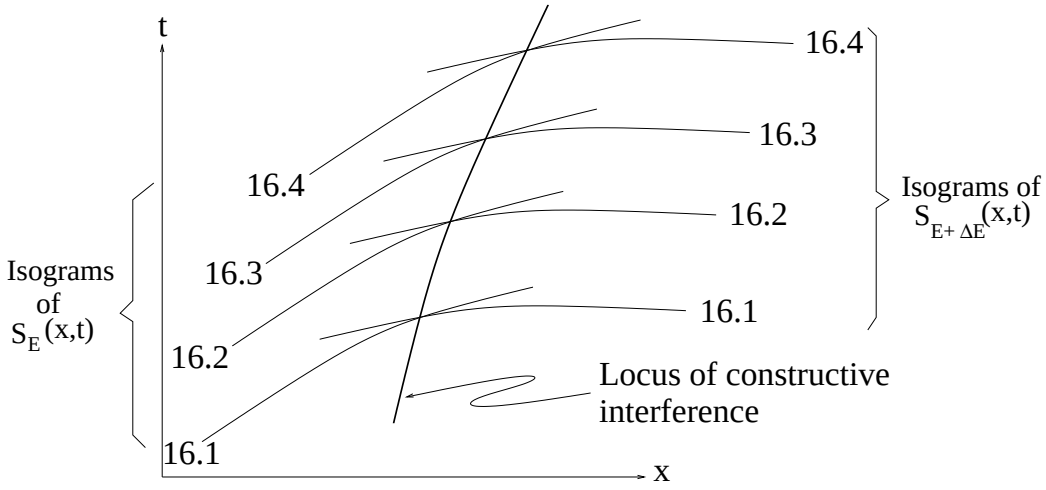


FIGURE 3.6.3. Constructive interference represented in space-time. The intersection of the respective isograms of $S_E(x,t)$ and $S_{E+\Delta E}(x,t)$ locates the events (x,t) which make up the trajectory of the particle in x-t space – the locus of constructive interference.

3.7. Constructive Interference

Our formulation of constructive interference is based on a picture in which at each time t a superposition of wave trains

$$\Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots \equiv \Psi(x,t)$$

yields a wave packet at time t. The principle of constructive interference itself,

$$\frac{\partial S_E(x,t)}{\partial E} = 0$$

is a condition which at each time t locates where the maximum amplitude of the wave packet is.

It is possible to bring into much sharper focus the picture of superposed wave trains and thereby not only identify the location of the resultant wave packet maximum, but also width of that packet.

3.8. Spacetime History of a Wave Packet

The sharpened formulation of this picture consists of replacing a *sum* of superposed wave amplitudes with an *integral* of wave amplitudes

$$\begin{aligned} \Psi(x,t) &= \text{"}\Psi_E(x,t) + \Psi_{E+\Delta E}(x,t) + \cdots\text{"} \\ &= \Delta_{-\infty}^{\infty}\, f(E)\, e^{\frac{i}{\hbar}S_E(x,t)} dE\,. \end{aligned} \tag{3.8.1}$$

A very instructive example is that of a superpostion of monochromatic ("single energy") wavetrains, each one weighted by the amplitude $f(E)$ of a Gaussian window in the Fourier ("energy") domain,

$$f(E) = Ae^{-(E-E_0)^2/\epsilon^2} \tag{3.8.2}$$

The dominant contribution to this integral comes from within the window, which is centered around the location of E_0 of the Gaussian maximum and has width 2ϵ,

which is small for physical reasons. Consequently, it suffices to represent the phase function as a Taylor series around that central point E_0, namely

(3.8.3)
$$S_E(x,t) = S_{E_0}(x,t) + \left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0} (E-E_0) + \frac{1}{2}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} (E-E_o)^2 + \begin{matrix} higher \\ order \\ terms \end{matrix} ,$$

and neglect the higher order terms. Keeping only the first three terms and ignoring the remainder allows an exact evaluation of the Gaussian superposition integral. This evaluation is based on the following formula

$$\Delta_{-\infty}^{\infty} e^{\alpha z^2 + \beta z} dz = \sqrt{\frac{\pi}{-\alpha}} e^{-\frac{\beta^2}{4\alpha}} . \tag{3.8.4}$$

Applying it to the superpostion integral, Eq. (3.8.1) together with Eqs. (3.8.2) and (3.8.3), we make the following identification

$$\begin{aligned} z =& E - E_0 ; \quad dz = dE , \\ \alpha =& -\frac{1}{\epsilon^2} + \frac{i}{\hbar}\frac{1}{2}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} \equiv -\frac{1}{\epsilon^2}(1 - i\sigma) , \\ -\frac{1}{\alpha} =& \frac{\epsilon^2}{1 - i\sigma} = \epsilon^2 \frac{1 + i\sigma}{1 + \sigma^2} , \\ \sigma =& \frac{1}{2}\frac{1}{\hbar}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} \epsilon^2 , \\ \beta =& \frac{i}{\hbar}\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0} . \end{aligned} \tag{3.8.5}$$

Inserting these expressions into the righthand side of the formula (3.8.4), one obtains

$$\begin{aligned} \Psi(x,t) &= A\sqrt{\pi}\epsilon\sqrt{\frac{1+i\sigma}{1+\sigma^2}} \exp\left\{ -\frac{1}{4}\left(\frac{\frac{\partial S(x,t)}{\partial E_o}}{\hbar}\right)^2 \epsilon^2 \left(\frac{1+i\sigma}{1+\sigma^2}\right)\right\} e^{i\frac{S_{E_o}(x,t)}{\hbar}} \\ &\equiv \underbrace{\mathcal{A}(x,t)}_{\substack{slowly \\ varying}} \underbrace{e^{\frac{iS_{E_o}(x,t)}{\hbar}}}_{\substack{rapidly \\ varying}} . \end{aligned}$$

This is a *rapidly oscillating* function

$$e^{i\, S_{E_0}(x,t)/\hbar}$$

modulated by a *slowly varying* amplitude $\mathcal{A}(x,t)$. For each time t this product represents a wave packet. The location of the maximum of this wave packet is given implicitly by

$$\left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0} = 0 . \tag{3.8.6}$$

As t changes, the x-location of the maximum changes. Thus we have curve in x-t space of the locus of those events where the slowly varying amplitude $\mathcal{A}$ has a maximum. In other words, this wave packet maximum condition locates those events (= points in spacetime) where constructive interference takes place.

A wave packet has finite extent in space and in time. This extent is governed by its squared modulus, i.e. the squared magnitude of its slowly varying amplitude,

$$|\Psi(x,t)|^2 = |\mathcal{A}|^2 = A^2\pi\epsilon^2\frac{1}{\sqrt{1+\sigma^2}}\exp\underbrace{\left\{-\frac{\epsilon^2}{2}\frac{1}{\sqrt{1+\sigma^2}}\frac{\left(\frac{\partial S_E(x,t)}{\partial E}\Big|_{E_0}\right)^2}{\hbar^2}\right\}}_{\mathbf{E}(x,t)} \tag{3.8.7}$$

We see that this squared amplitude has nonzero value even if the condition for constructive interference, Eq.(3.8.6), is violated. This violation is responsible for the finite width of the wave packet. More precisely, its shape is controlled by the exponent $\mathbf{E}(x,t)$,

$$\mathbf{E}(x,t) \equiv \left\{-\frac{\epsilon^2}{2}\frac{1}{\sqrt{1+\left(\frac{\epsilon^2}{2\hbar}\,\frac{\partial^2 S_E(x,t)}{\partial E^2}\Big|_{E_0}\right)^2}}\frac{\left(\frac{\partial S_E(x,t)}{\partial E}\Big|_{E_0}\right)^2}{\hbar^2}\right\} \neq 0\,.$$

The spacetime evolution of this shape is exhibited in Figure 3.8.1 on the next page. Thus the worldline of the particle is not a sharp one, but instead has a slight spread in space and in time. How large is this spread?
The magnitude of the wave mechanical ("non-classical") spread in the world line is the width of the Gaussian wave packet. This spread is Δx, the amount by which one has to move away from the maximum in order that the amplitude profile change by the factor $e^{\frac{1}{2}}$ from the maximum value. Let us calculate this spread under the circumstance where the effect due to dispersion is a minimum, i.e. when σ is neglegibly small. In that case the condition that $\mathbf{E}(x+\Delta x,t) = -1$ becomes

$$\left|\frac{\epsilon}{\hbar}\frac{\partial S_E(x+\Delta x,t)}{\partial E}\Big|_{E_0}\right| = 1\,.$$

Expand the left hand side to first order, make use of the fact that (x,t) is a point in spacetime where the wavepacket profile has a maximum, i.e. satisfies Eq.(3.8.6). One obtains

$$\left|\epsilon\frac{\partial^2 S}{\partial E\partial x}\Delta x\right| = \hbar$$

or, in light of $\partial S_E(x,t)/\partial x \equiv p(x,t;E)$,

$$\left|\epsilon\frac{\partial p}{\partial E}\Delta x\right| \quad = \quad \hbar\,,$$

and hence

$$\boxed{\Delta p\Delta x = \hbar}$$

Similarly the temporal extent Δt, the amount by which one has to wait (at fixed x) for the wave amplitude profile to decrease by the factor $e^{-1/2}$ from its

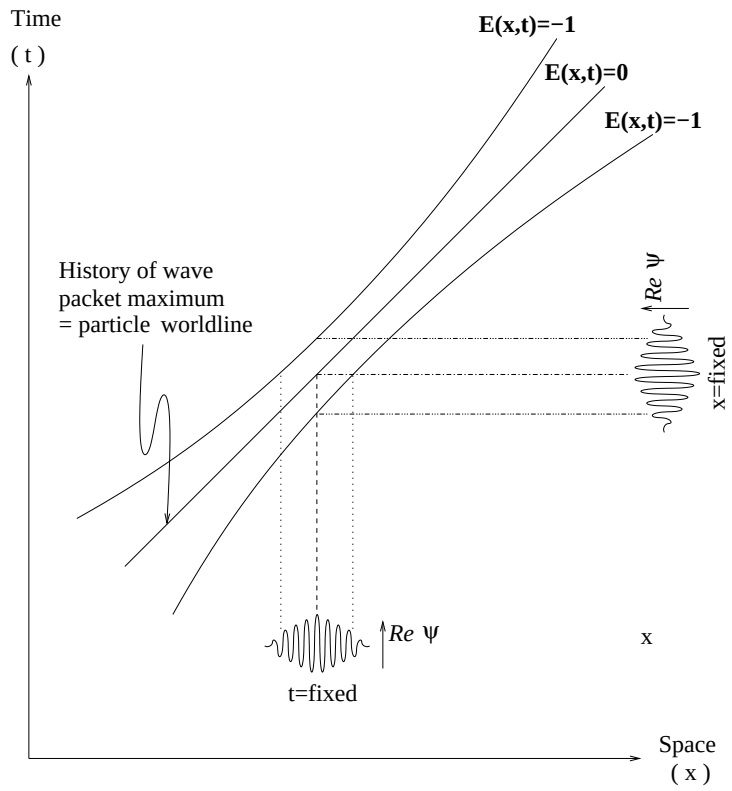


FIGURE 3.8.1. Spacetime particle trajectory ("the $\mathbf{E}(x,t) = 0$ isogram") and the dispersive wave packet amplitude histories surrounding it. The two mutually diverging ones (both characterized by $\mathbf{E}(x,t) = -1$) in this figure refer to the front and the back end of the wave packet at each instant $t = fixed$, or to the beginning and the end of the wave disturbance passing by a fixed location $x = fixed$. The particle and the wave packet maximum are moving with a velocity given by the slope of the $\mathbf{E}(x,t) = 0 = \left.\frac{\partial S_E(x,t)}{\partial E}\right|_{E_0}$ isogram, which is the locus of constructive interference exhibited in Figure 3.6.3

On the other hand, the convergence and subsequent divergence ("dispersion") of the wave packet is controlled (and expressed mathematically) by the behavior of the second derivative, $\left.\frac{\partial^2 S_E(x,t)}{\partial E^2}\right|_{E_0}$ of the dynamical phase $S_E(x,t)$. Whereas the behavior of its first derivative characterizes the difference in the motion of particles launched with difference initial conditions, its second derivative characterizes the intrinsically wave mechanical aspects of each of these particles.

maximum value, satisfies the condition

$$\left| \frac{\epsilon}{\hbar} \frac{\partial S_E(x, t + \Delta t)}{\partial E} \Big|_{E_0} \right| = 1$$

which become

$$\left| \epsilon \frac{\partial^2 S_E}{\partial E \partial t} \Big|_{E_0} \Delta t \right| = \hbar$$

$$\left| \epsilon(-) \frac{\partial E}{\partial E} \Big|_{E_0} \Delta t \right| = \hbar$$

or

$$\boxed{\Delta E \Delta t = \hbar}\,.$$

The two boxed equation are called the Heisenberg indeterminacy relation. Even though we started with the dynamical phase S (see page 38) with $\Psi \sim e^{\frac{iS}{\hbar}}$ to arrive at the extremal path in spacetime, the constant $\hbar$ ("quantum of action") never appeared in the final result for the spacetime trajectory. The reason is that in the limit

$$\frac{S}{\hbar} \to \infty$$

the location of the wave packet reduces to the location of the wave crest. Once one knows the dynamical phase $S(x, t)$ of the system, the condition of constructive interference gives *without approximation* the location of the sharply defined Newtonian world line, the history of this wave crest, an extremal path through spacetime.

3.9. Hamilton's Equations of Motion

To validate the claim that constructive interference leads to the extremal paths determined by the E-L equations, one must first recast them into a form that involves only q^i and p_j instead of $q^i$and $\dot{q}^j$. Holding off that validation until Section 3.11 on page 59, we achieve the transition from $(q^i, \dot{q}^j)$ to (q^i, p_j) as follows: The Lagrangian is a function of $q^i$and $\dot{q}^i$. Consequently,

$$dL = \sum_i \frac{\partial L}{\partial q^i} dq^i + \sum_i \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i$$

which may be rewritten as

$$dL = \sum_i \dot{p}_i dq^i + \sum_i p_i d\dot{q}^i\,,$$

where

$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

and

$$\dot{p}_i = \frac{\partial L}{\partial q^i}$$

by the E-L equations. Using the fact that

$$p_i d\dot{q}^i = d(p_i \dot{q}^i) - \dot{q}^i dp_i\,,$$

one obtains, after a sign reversal, an expression which depends only on q^i and p_i:

$$d\underbrace{\left(\sum_i p_i\dot{q}^i - L\right)}_{H} = -\sum_i \dot{p}_i dq^i + \sum_i \dot{q}^i dp_i \,. \tag{3.9.1}$$

Introduce the Hamiltonian of the system

$$H(p,q,t) \equiv \sum p_i\dot{q}^i - L \,.$$

Compare its differential

$$dH = \frac{\partial H}{\partial q^i}dq^i + \frac{\partial H}{\partial p_i}dp_i + zero \times d\dot{q}^i$$

with the one given above by Eq.(3.9.1). Recall that two differentials are equal if and only if the coefficients of the (arbitrary) coordinate differences (i.e. $dq^1, \cdots, dq^s$, $dp_1, \cdots, dp_s$) are equal. Consequently, one has

$$\boxed{\dot{q}^i = \frac{\partial H}{\partial p_i}\,, \qquad p_i = -\frac{\partial H}{\partial q_i}\,.}$$

These are the *Hamilton's* or the *canonical equations of motion.* They are equivalent to those of E-L. Comment 1: The fact that H does not depend on $\dot{q}^i$ follows directly from

$$\frac{\partial H}{\partial \dot{q}^i} = p_i - \frac{\partial L}{\partial \dot{q}^i} = 0$$

Comment 2: a) It follows from Hamilton's equations of motion that

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial H}{\partial q^i}\dot{q}^i + \frac{\partial H}{\partial p_i}\dot{p}_i + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial t}\end{aligned}$$

In other words, if H is independent of any *explicit* time dependence, i.e. time then H i*s a constant* along the trajectory of the evolving system.
Comment 2: b.) If H is independent of the *generalized coordinate* q^k, then

$$\frac{dp^k}{dt} = 0$$

i.e. p_k *is a constant* for the evolving system.

3.10. The Phase Space of a Hamiltonian System

The 2s-dimentional spaced is spanned by the coordinates

$$[q^1, \cdots, q^s, p_1, \cdots, p_s]$$

is called the phase space of the system. In this phase space, the curve

$$[q^i(t), p_i(t)]$$

is an integral curve of the Hamiltonian or phase path vector field

$$\left(\frac{\partial H(q,p)}{\partial p_i}, -\frac{\partial H(q,p)}{\partial q^i}\right)$$

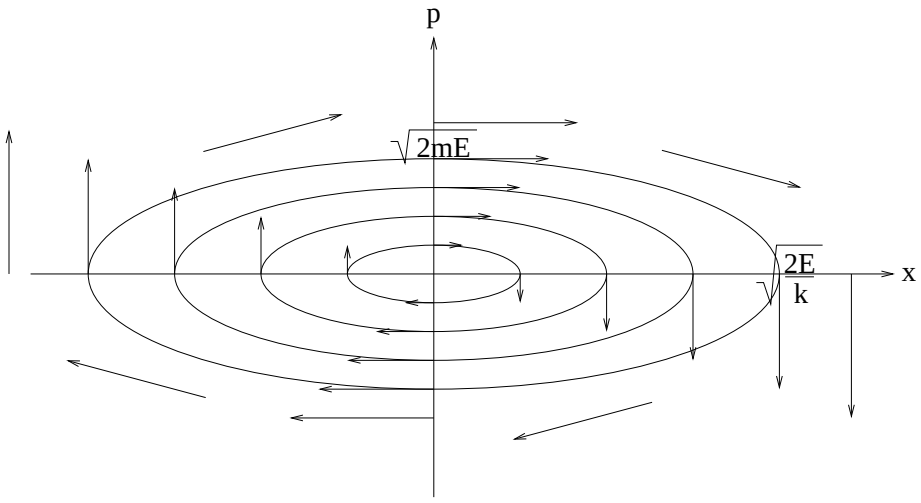


FIGURE 3.10.1. Hamiltonian vector field of a simple harmonic oscillator (s.h.o.) of mass m and spring constant k. The ellipses are integral curves whose tangents at each point are the vectors of that field. The major axis, $\sqrt{\frac{2E}{k}}$, and the minor axis, $\sqrt{2mE}$, of each ellipse are determined by the energy E of the s.h.o. The area of any particular ellipse is $2\pi\, E\, \frac{m}{k}$.

In other words, the tangents to this curve are given by

$$(\dot{q}^i, \dot{p}_i) = \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right)$$

Example: For the simple harmonic oscilator
the Lagarangian is:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

and the Hamiltonian is:

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 = E$$

a) The phase space of this system is spanned by x and p.
The Hamiltonian vector field is

$$\left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right) = \left(\frac{p}{m}, -kx\right)$$

b) The area of one of the phase-path ellipses is

$$\text{area} = \Delta\, pdx \qquad \text{and it has the dimension of "action"}$$

According to quantum mechanics the action of a periodic system must obey the Bohr quantization condition

$$\Delta\, pdx = \left(n + \frac{1}{2}\right) h\,, \qquad n = 1, 2, \cdots \tag{3.10.1}$$

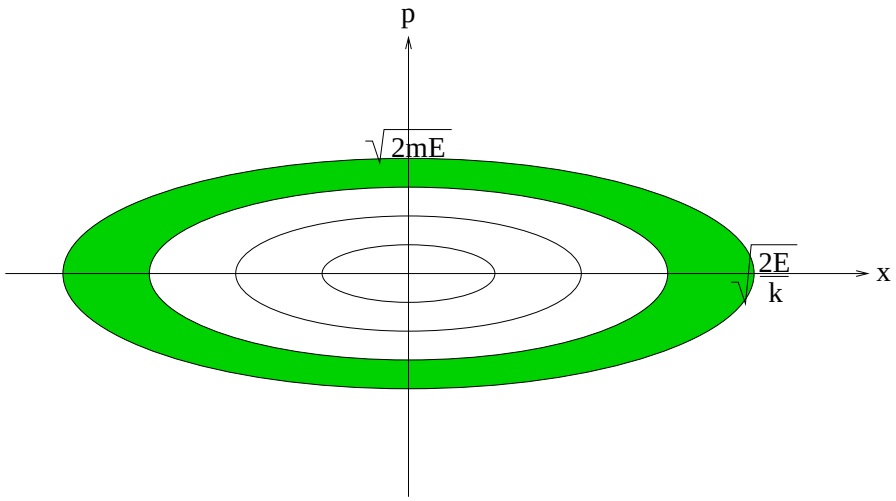


FIGURE 3.10.2. The shaded difference between the areas of adjacent phase space ellipses, Eq.(3.10.1), is precisely $h = 6.27\times10^{-27}$erg sec, which is one quantum of action.

Thus, as depicted in Figure 3.10.2, the quantum mechanically allowed phase space ellipses differ in area from one another by precisely $h = 6.27\times10^{-27}$erg sec, which is one quantum of action.

For the simple harmonic oscilator the area of one of these ellipses is $\Delta\, pdx = \pi\sqrt{2mE}\sqrt{\frac{2E}{k}} = 2\pi E\sqrt{\frac{m}{k}} = 2\pi\frac{E}{\omega}$. Thus the Bohr quantization condition yields

$$2\pi\frac{E}{\omega} = \left(n+\frac{1}{2}\right)h$$

or with $\frac{\omega}{2\pi} = frequency$

$$E = (n+\frac{1}{2})h \times frequency$$

3.11. Consturctive interference ⇒ Hamilton's Equations

The principle of constructive interference provides the bridge between particle and wave mechanics. This fact is validatede by the following theorem.

THEOREM. *Constructive interference conditions imply the Hamilton's equantions of motion and hence determine the existence of an extremal path.*

Proof: Step 1.) Consider a complete integral of the H-J equation

$$S = S(t, q^1, \cdots, q^s, \alpha_1, \cdots, \alpha_s)$$

i.e. a solution which has as many arbitrary constants as there are independent coordinates[5]. The constructive interference conditions are

$$\frac{\partial S}{\partial \alpha_k} = 0 \qquad k = 1, \cdots, s$$

They determine implicitly a trajectory $q^i = q^i(t)$, $i = 1, \cdots, s$.
Step 2.) Take the total derivative and obtain

$$\begin{aligned} 0 &= \frac{d}{dt}\frac{\partial S}{\partial \alpha_k} = \frac{\partial^2 S}{\partial t \partial \alpha_k} + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= -\frac{\partial}{\partial \alpha_k} H(q, \frac{\partial S(t,q,\alpha)}{\partial x^i}, t) + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= -\frac{\partial H}{\partial p_i}\frac{\partial^2 S}{\partial \alpha_k \partial q^i} + \frac{\partial^2 S}{\partial q^i \partial \alpha_k}\frac{dq^i}{dt} \\ &= \frac{\partial^2 S}{\partial \alpha_k \partial q^i}\left(\frac{dq^i}{dt} - \frac{\partial H}{\partial p_i}\right), \end{aligned}$$

which implies the 1st half of Hamilton's equations,

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i},$$

provided $\frac{\partial^2 S}{\partial \alpha_k \partial q^i}$ is non-singular.
Step 3.) Differentiate both sides of the H-J equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial q^i}\left[\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right)\right] \\ &= \frac{\partial}{\partial t}\frac{\partial S}{\partial q^i} + \frac{\partial H}{\partial p_k}\Big)_q \frac{\partial^2 S}{\partial q^k \partial q^i} + \frac{\partial H}{\partial q^i}\Big)_p \\ &= \frac{\partial}{\partial t}\frac{\partial S}{\partial q^i} + \frac{dq^k}{dt}\frac{\partial}{\partial q^k}\left(\frac{\partial S}{\partial q^i}\right) + \frac{\partial H}{\partial q^i}\Big)_p \\ &= \frac{d}{dt}p_i + \frac{\partial H}{\partial q^i} \end{aligned}$$

which is the 2nd half of Hamilton's equation's,

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}.$$

QED. Thus the two Hamilton's equations of motion are implied by the principle of constructive interference indeed.

Lecture 14

3.12. Applications

Two of the most important applications of Hamilton-Jacobi theory are found in

[5]Such independence is expressed mathematically by the fact that

$$det\left|\frac{\partial^2 S}{\partial q^i \partial \alpha_j}\right| \neq 0.$$

This condition would be violated if the dependence on two constants were of the form $S(t, q^i, f(\alpha_1, \alpha_2), \alpha_3, \cdots, \alpha_s)$.

(i) the motion of bodies on the astronomical scale, for example, space craft, comets, or planets moving in a gravitational field, and in.

(ii) the motion of bodies on the atomic scale, for example, a charged particle (electron) moving in the potential of an atomic nucleus or in th electromagnetic field of a pulsed laser.

The mathematical procedure for these and numerous other examples is routine and always the same:
(i) Construct the Hamiltonian for the system
(ii) Write down and solve the H-J equation
(iii) Apply the conditions of constructive interference to obtain the trajectories of the body.
Let us describe how this three step procedure is done in practice.

3.12.1. H-J Equation Relative to Curvilinear Coordinates. In constructing the Hamiltonian one must choose some specific set of coordinates. For a single particle it is is difficult to find an easier way of writing down the H-J equation than the way whose starting point is the element of arclength

$$\begin{aligned} (ds)^2 &= dx^2 + dy^2 + dz^2 \quad \text{(Cartesian coordinates)} \\ &= g_{ij} dx^i dx^j \qquad \text{(curvilinear coordinates)} \end{aligned} \tag{3.12.1}$$

This element of arclength is the best starting point because it is related so closely to the Lagrangian of the system. Indeed, one has

$$\begin{aligned} L &= \frac{1}{2} m \vec{x} \cdot \vec{x} - U \\ &= \frac{1}{2} m \left[\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right] - U \\ &= \frac{1}{2} m \, g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} - U \, . \end{aligned}$$

In other words, the Lagrangian is constructed relative to curvilinear coordinates by inspection. The steps leading to the H-J equation are now quite routine.

The momenta are

$$p_j = \frac{\partial L}{\partial \dot{x}^j} = m \, g_{ij} \dot{x}^i \, .$$

Let g^{kj}be the inverse of g_{ji}: $g^{kj} g_{ji} = \delta_i^k$ so that

$$\dot{x}^i = \frac{1}{m} \, g^{ij} p_j$$

and

$$H = p_j \dot{x}^j - L = \frac{1}{2m} \, g^{ij} p_i p_j + U$$

Thus the Hamilton-Jacobi equation is

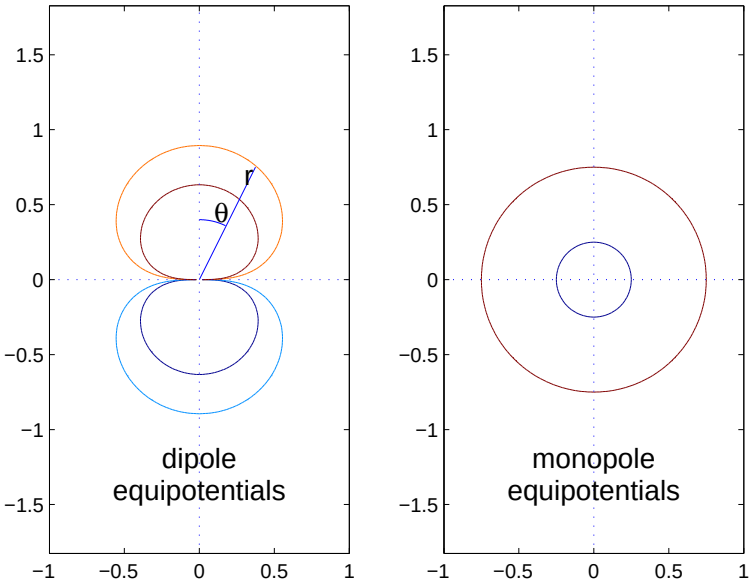


FIGURE 3.12.1. Rotationally symmetric potential as the sum a dipole potential $\left(\mu\frac{\cos\theta}{r^2}\right)$ plus a monopole potential $\left(-\frac{k}{r}\right)$.

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\, g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} + U$$

in terms of the inverse metric.

3.12.2. Separation of Variables. The most important way of solving the H-J equation is by the method of sep[aration of variables. To illustrate this, consider the following

Example (Particle in a dipole potential). Consider the motion of a particle in the potential of a combined dipole and monopole field. Relative to spherical coordinates the metric is

$$(ds)^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2$$

and that potential has the form

$$U(r,\theta) = \mu\frac{\cos\theta}{r^2} - \frac{k}{r}$$

Its equipotential surfaces are rotationally symmetric around the z-axis

The Lagrangian is

$$\begin{aligned} L &= \text{Kinetic Energy } - \text{Potential Energy} \\ &= \frac{1}{2} m\, g_{ij} \dot{q}^i \dot{q}^j - U(q^1, q^2, q^3) \\ &= \frac{1}{2} m\, (\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2\theta \dot{\phi}^2) - \mu \frac{\cos\theta}{r^2} + \frac{k}{r}\,. \end{aligned}$$

The corresponding Hamilto-Jacobi equation is

$$\begin{aligned} 0 &= \frac{\partial S}{\partial t} + H \\ &= \frac{\partial S}{\partial t} + \frac{1}{2m} g^{ij} \frac{\partial S}{\partial q^i} \frac{\partial S}{\partial q^j} + U(q^1, q^2, q^3) \\ &= \frac{\partial S}{\partial t} + \frac{1}{2m} \left[\left(\frac{\partial S}{\partial r} \right)^2 + \frac{1}{r^2} \left(\frac{\partial S}{\partial \theta} \right)^2 + \frac{1}{r^2 \sin^2\theta} \left(\frac{\partial S}{\partial \phi} \right)^2 \right] + \mu \frac{\cos\theta}{r^2} - \frac{k}{r}\,. \end{aligned}$$

This equation can be solved by the method of separation of variables.

A Hamiltonian system whose Hamilton-Jacobi equation has a separable solution is said to be integrable. The method of solving such an equation is by separating the variables.

The Method of Separation of Variables

. This method is condensed into the following three definitions and propositions:

(1) *Definition (Separable variables).* The variables $q^1, q^2, \cdots, q^s$ in the H-J equation

$$0 = \frac{\partial S}{\partial t} + H\left(t,\, q^i,\, \frac{\partial S}{\partial q^j}\right) \equiv \mathcal{H}\left(t,\, \frac{\partial S}{\partial t}, q^i,\, \frac{\partial S}{\partial q^j}\right)$$

are said to be *separable* if it has a "complete" solution of the form

$$S = S_0(t, \alpha_0) + S_1(q^1, \alpha_0, \alpha_1) + S_2(q^2, \alpha_0, \alpha_1, \alpha_2) + S_3(q^3, \alpha_0, \alpha_1, \alpha_2, \alpha_3) \tag{3.12.2}$$

where each S_i depends only on t and q^i respectively.

(2) *Definition (Complete solution).* A solution is said to be *complete* if

$$\det \left| \frac{\partial^2 S}{\partial \alpha_i \partial q^j} \right| \neq 0$$

Remark 1: We saw (in Lecture 13) in the context of reconstructing the classical trajectories of a Hamiltonian system from the principle of constructive interference it was essential that the matrix $\left[\partial^2 S / \partial \alpha_i \partial q^j\right]$ be non-singular.

Remark 2: The solution, Eq.(3.12.2), is complete indeed, because

$$\det \left| \tfrac{\partial^2 S}{\partial \alpha_i \partial q^j} \right| = \det \begin{vmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{vmatrix} \begin{matrix} \\ \neq 0 \\ \\ \end{matrix} \begin{matrix} 1 & 2 & 3 & \leftarrow & j \\ & & & & 0 \\ & & & & 1 \\ & & & & 2 \\ & & & & \uparrow \\ & & & & i \end{matrix}$$

and its diagonal elements are not zero.

(3) *Definition (separability condition).* The Hamilton-Jacobi equation is said to satisfy the separability criterion if its Hamiltonianis is of the form

$$\mathcal{H}\left(t, \frac{\partial S}{\partial t}, q^i, \frac{\partial S}{\partial q^j}\right) = f_3\left(f_2\left(f_1\left(f_0\left(t, \frac{\partial S}{\partial t}\right), q_1, \frac{\partial S}{\partial q^1}\right), q^2, \frac{\partial S}{\partial q^2}\right), q^3, \frac{\partial S}{\partial q^3}\right)$$

(for $s = 3$ degrees of freedom). This functional form is said satisfy the condition of separability because the solution to this first order p.d.e. has the separated form, Eq.(3.12.2). In other words, the claim is

Proposition: Definition 3 implies Definition 1.
Proof: Step 1.) According to Definition 3 the H-J equation is

$$f_3\left(f_2\left(f_1\left(f_0\left(t, \frac{\partial S}{\partial t}\right), q_1, \frac{\partial S}{\partial q^1}\right), q^2, \frac{\partial S}{\partial q^2}\right), q^3, \frac{\partial S}{\partial q^3}\right) = 0\,. \tag{3.12.3}$$

The method of solution via separation of variables starts by solving for f_0. One finds

$$f_0\left(t, \frac{\partial S}{\partial t}\right) = \text{an expression involving } q^1, q^2, q^3, \frac{\partial S}{\partial q^1}, \frac{\partial S}{\partial q^2}, \text{and} \frac{\partial S}{\partial q^3}\,.$$

Assume the solution to have the form

$$S = T(t) + S'(q^1, q^2, q^3)\,. \tag{3.12.4}$$

This assumption is the first step towards success because the resulting common value of

$$\underbrace{f_0\left(t, \frac{dT(t)}{dt}\right)}_{\substack{\text{independent} \\ \text{of } q^1, q^2, q^3}} = \underbrace{\text{an expression that depends only on } q^1, q^2, q^3}_{\text{independent of t}}$$

is independent of all variables. This common independence implies that f_0 is a constant, say, α_0:

$$f_0\left(t, \frac{dT(t)}{dt}\right) = \alpha_0\,.$$

Solving for $T(t)$, one obtains

$$T(t) = S_0(t, \alpha_0)\,. \tag{3.12.5}$$

Step 2.) Using this function in Eq.(3.12.4) and introduce S into Eq.(3.12.3), which now becomes

$$f_3\left(f_2\left(f_1\left(\alpha_0, q^1, \frac{\partial S'}{\partial q^1}\right), q^2, \frac{\partial S'}{\partial q^2}\right), q^3, \frac{\partial S'}{\partial q^3}\right) = 0\,.$$

Solving for f_1 one finds

$$f_1\left(\alpha_0, q^1, \frac{\partial S'}{\partial q^1}\right) = \text{an expression involving } q^2, q^3, \frac{\partial S'}{\partial q^2}, \text{and} \frac{\partial S'}{\partial q^3}\,.$$

Let

$$S' = Q_1(q^1) + S''(q^2, q^3)\,.$$

Consequently,

$$\underbrace{f_1\left(\alpha_0, q^1, \frac{dQ_1(q^1)}{dq^1}\right)}_{\substack{\text{independent}\\ \text{of } q^2, q^3}} = \underbrace{\text{an expression that depends only on } q^2, q^3}_{\text{independent of } q^1}.$$

This common independence implies that f_1 is a constant, say, α_1:

$$f_1\left(\alpha_0, q^1, \frac{dQ_1(q^1)}{dq^1}\right) = \alpha_1\,.$$

Solving for $Q_1(q^1)$, one obtains

$$Q_1(q^1) = S_1(q^1, \alpha_1)\,. \tag{3.12.6}$$

Consequently, the solution, Eq.(3.12.4), to the H-J equation has the form

$$S = T(t) + Q_1(q^1) + S''(q^2, q^3)\,.$$

Step 3.) Repeat Step 2.) two more times to obtain

$$f_2\left(\alpha_0, \alpha_1, q^2, \frac{dQ_2(q^2)}{dq^2}\right) = \alpha_2$$

$$f_3\left(\alpha_0, \alpha_1, \alpha_2, q^3, \frac{dQ_3(q^3)}{dq^3}\right) = \alpha_3\,.$$

Notice, however, that the H-J Eq.(3.12.3) implies that $\alpha_3 = 0$, always. Consequently, there are only *three* independent separation constants, $(\alpha_0, \alpha_1, \alpha_2)$, as many as there are degrees of freedom, (namely,q^1, q^2, q^3), while the number of independent variables, (t, q^1, q^2, q^3) , is four. It follows that

$$S = T(t) + Q_1(q^1) + Q_2(q^2) + Q_3(q^3)$$

and hence wth Eqs.(3.12.5), (3.12.6), etc.

$$= S_0(t, \alpha_0) + S_1(q^1, \alpha_0, \alpha_1) + S_2(q^2, \alpha_0, \alpha_1, \alpha_2) + S_3(q^3, \alpha_0, \alpha_1, \alpha_2, \alpha_3 = 0)\,.$$

Thus the dynamical phase S has indeed the separated form whenever its H-J equation has the form (3.12.3)