1 Introduction

We are familiar with Newton's equation of motion for a particle. A particle is an idealized body having physical attributes, such as, mass, charge etc. but whose dimensions can be neglected in describing its motion, i.e. it is considered to be a geometrical point. Consider a single particle of mass m, whose position at time t is given by $\sim r$ (if we are using the Cartesian coordinates, we specify the position by x,y and z). Since the equation of motion is second order in time, if the position $\sim r$ and the velocity $\sim v = \sim \dot{r}$ of the particle is known at some initial time $t = 0$, we can in principle, by integrating the equation of motion, determine the position and the velocity for all time to come. This can become quite complicated if we consider complex systems, for instance, a system of N particles. Since the position of each particle requires three quantities (such as x,y,z or r,θ,φ), in order to specify the positions of all the N particles simultaneously, we need to specify $3N$ quantities. However, it is likely that the system is under constraints. For instance, if we have two independent particles, we would require 6 coordinates, three for each. However, if the two particles are connected by a rigid rod, the distance between them remains fixed as the two particles move. In such a case we need only five quantities to specify the state of motion of the two particle system. In general, if the system of N particles have k independent constraints, the number of independent quantities would be $3N - k$. This is known as the **degree of freedom** of the system.The position of a system with d degrees of freedom at any time t is completely defined by d *generalized coordinates* $q_1, q_2,...,q_d$. In order to describe the mechanical state of the system, it is not enough to specify the generalized coordinates alone at time $t = 0$ because from a knowledge of these alone one cannot determine the position of the system at a subsequent time. In addition, we need to specify the velocities of each of the particles at this time as well. Since velocity is a vector, we need d number of *generalized velocities* $\dot{q}_1, \dot{q}_2,...,\dot{q}_d$. [Note:

Integrating the equations of motion for such a system gets very complicated . what we need is alternative formulation of the equation of motion. This is provided by Lagrange's equations. The advantage of this "Analytical Mechanics" over the conventional vector mechanics are threefold. Firstly, it provides a single unifying principle to work with. Second, unlike vector mechanics where each particle comprising a system has to be treated separately, the analytical mechanics looks at the system as a whole. Finally, while in vector mechanics the constraints are treated by invoking special forces., here they are taken care of in a natural way.

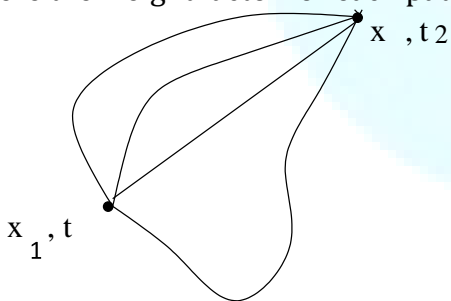## 2   The Variational Principle

Before we discuss the classical variational principle, we would make a short but interesting digression. This has to do with Feynman's formulation of quantum mechanics using his *path integral* method.

We know from our knowledge of elementary quantum mechanics that the state of a particle at a position x at time t is given by the wave function $\psi(x,t)$ which has the interpretation of probability amplitude of the particle to be at that position at time t. (Our discussion can be readily extended to three dimensions). Feynman had provided an interpretation of this amplitude in his Ph.D. thesis entitled "The Principle of Least action in quantum Mechanics", submitted to the Princeton University in 1942. His supervisor Wheeler was so much impressed with this novel interpretation of quantum mechanics that he is said to have remarked to Einstein " Does not this marvellous discovery make you willing to accept quantum mechanics?" (Einstein, who was always skeptical about quantum mechanics had retorted "I still cannot believe that God plays dice, but may be, I have earned the right to make my mistakes").

The idea behind the path integral formulation is as follows: Suppose a particle is at the position x_i at time t_i. The quantum mechanical amplitude for the particle to be found at the position x_f at time t_f is given by summing over *all possible paths* that connect x_i to x_f, i.e.

$$\int Dx(t)e^{iS(x)/\hbar}$$

where the weight factor for each path is determined by the *classical action $S(x,t)$*. The classical equation of motion is determined in a very simple way by taking the $\hbar \to 0$ limit. In this limit the weight factor oscillates very rapidly so that the contribution to the integral comes primarily from those paths for which the action is nearly stationary. We will see later that this is just the way the Euler-Lagrange equations are derived from the classical action.

Fig.1: Possible paths connecting two points

Let us look at a couple of examples before discussing the variational principle. First, let us consider the shortest distance between two points 1 and 2 on a plane. How does one find this, which we intuitively know to be the straight-line joining the two points? We could join the points 1 and 2 by drawing all possible curves, measure the length of each of the paths and find which of them is of minimum length. The length of the curve is given by

$$L = \int_1^2 dl = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$
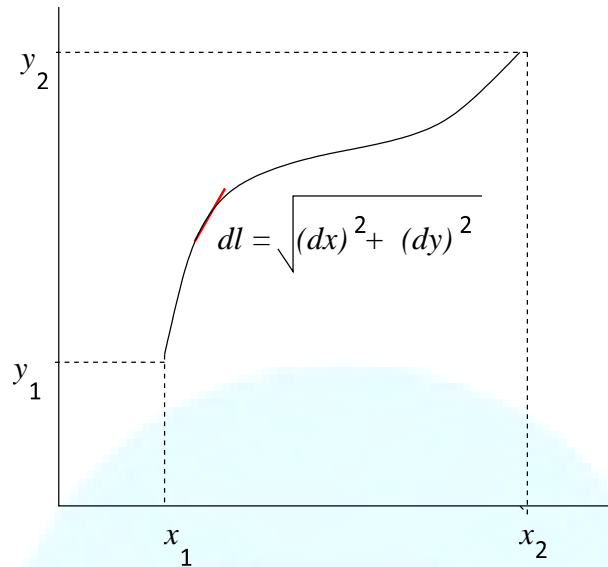
(1)

Figure 2: Geometry of a typical path

We have to determine the function $y(x)$ which minimises the length. $L[y(x)]$ depends on the function $y(x)$ itself. This is known as a **functional**. The way we have written seems to suggest that $L[y(x)]$ is nothing but function of a function. However, there a subtle difference between a functional and function of a function. A function is a one to one mapping from a range of values in its domain to a value in its image plane or the range. A functional takes as its argument not one particular value but a function itself and maps it to a definite value. For different functions as its argument, the functional returns a different value.

Consider a second example. This is Fermat's principle of least time, which states that in travelling between two points, light would take that path for which the travel time is minimum. If the refractive index of the medium $n(x)$ is constant, the shortest time is also the shortest path. However, if $n(x)$ depends on x, the time to travel a distance dl is $dt = dl/v = dl/(c/n) = ndl/c$. Thus the total time taken in travelling from the point 1 to the point 2 is

$$T = \frac{1}{c}\int_1^2 ndl = \frac{1}{c}\int_1^2 n(x)\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

(2)

In both these cases, we need to minimise the integral over different possible paths. There could be situation where we may need to maximise or find a point of inflection. We will determine a general condition for such optimisation.

Recall that for the function of a single variable x, the maximum or the minimum is given by the vanishing of the first derivative $\frac{df}{dx}= 0$. If, however, the second derivative $\frac{d^2 f}{dx^2} = 0$ as well, the function has a point of inflection there.
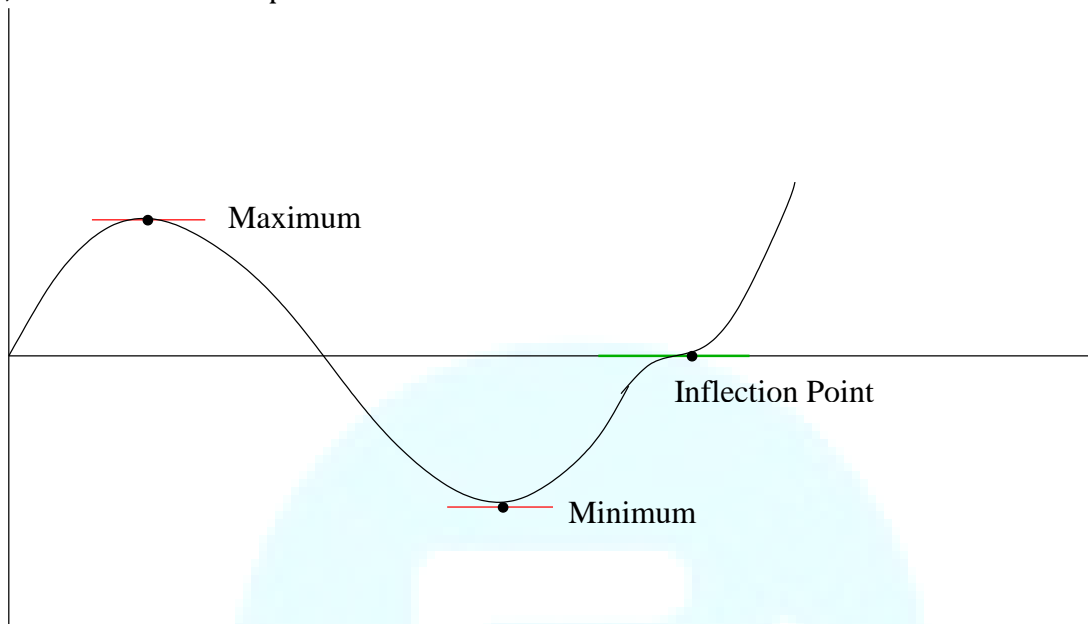

Figure 3: Showing the stationary points

If we consider a function several variables $F(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, ..., x_n)$, using a Taylor expansion, we have,

$$F(\mathbf{x} + \epsilon\xi) - F(\mathbf{x}) = \epsilon \sum_{k=1}^{n} \frac{\partial F}{\partial x_k}\xi_k + O(\epsilon^2)$$

We define a stationary point (maximum, minimum or a point of inflection) as a point at which the first order term in vanishes for all ξ, i.e.

$$\frac{\partial F}{\partial x_k} = 0, \quad \forall k$$

We extend the same idea for a functional. Consider a functional $F[y(x)]$. How does the functional change with the change in the argument, which is a function $y(x)$? Suppose a particular path is given by $y(x)$. The neighboring path can be written as $y(x) + \epsilon h(x)$ where $h(x)$ is some differentiable function. Suppose the end points of all the paths are (x_1, y_1) and (x_2, y_2). Since these two points are common to all the possible paths, we must have, for all the paths $y(x_1) = y_1, y(x_2) = y_2$

which implies

$$h(x_1) = h(x_2) = 0$$

. Consider

$$\delta F = F[y + \epsilon h] - F[y]$$

We can expand δF using a Taylor expansion and see that for small , it is proportional to . For a stationary functional $\delta F = O(\epsilon^2)$.

Let us return to the problem of calculating the shortest distance between two points. In this case, our functional $F[y]$ is given by

$$F[y(x)] = \int_{x_1}^{x_2} \sqrt{1 + (y')^2} dx$$

so that

$$F[y(x) + \epsilon h(x)] = \int_{x_1}^{x_2} \sqrt{1 + (y' + \epsilon h')^2} dx$$

Using a Taylor like expansion, we have,

$$\sqrt{1 + (y' + \epsilon h')^2} = \sqrt{1 + (y')^2} + \epsilon h' \frac{\partial}{\partial y'} \sqrt{1 + (y')^2} \big|_{\epsilon h' = 0} + O(\epsilon^2 h'^2)$$

$$= \sqrt{1 + (y')^2} + \epsilon h' \frac{y'}{\sqrt{1 + (y')^2}} + O(\epsilon^2 h'^2)$$

$$\delta F = F[y + \epsilon h] - F[y] = \epsilon \int_{x_1}^{x_2} dx \frac{h' y'}{\sqrt{1 + (y')^2}} + O(\epsilon^2 h'^2)$$

Thus

As F is stationary, the term proportional to vanishes,

$$\int_{x_1}^{x_2} dx \frac{h' y'}{\sqrt{1 + (y')^2}} = 0$$

It can be shown that a necessary and sufficient condition for the above to be valid is

$$\frac{y'}{\sqrt{1 + (y')^2}} = \text{constant} = \alpha \text{ (say)}$$

The sufficiency is readily proved for then we have

Z $_{x2}$ $\delta F = dx\alpha h^0 = \alpha[h(x_2) - h(x_1)] =$

0

$x1$

Constancy $\frac{y}{\sqrt{1 + (y')^2}}$ ^{0} ofimplies constancy of y^0. Let $y^0 = m$. The solution for y is

then $y = mx + c$. m and c can be determined by substituting the values of y at the end points. We will later show the necessity of the constancy condition.