# 4

# Modelling using variational principles

## 4.1 FROM NEWTONIAN TO VARIATIONAL MECHANICS

### 4.1.1 Why variational principles?

In Chapter 3 we first studied some basic physical laws indicating how motion or, more generally, state of a mechanical system changes under the influence of forces or, more generally, external stimuli. Then four balance laws, namely those of mass, energy, linear and angular momentum were applied as basic tools of modelling. The examples studied there showed that if a phenomenon is sufficiently complicated, then more than just one balance law must be employed. This means that in general a complex phenomenon cannot be entirely described by means of any single balance law. Additionally, application of balance laws invokes the nontrivial problem of independence of these laws.

Now we would like to focus our attention on some general theorems, which produce the model 'at once'. They are called **variational principles**. The description of real world phenomena by means of variational principles is considered by analysts as most successful in view of its prolific applications. This is related to the observation that physical phenomena have variational character, meaning that certain quantities, closely connected with these phenomena reach stationary or extremal values. Since the branch of mathematics called *variational calculus* is, in the usual brief course of mathematics for engineers, often treated somewhat cursorily, this method of modelling is not familiar to many engineers. We are not going to present here a course on variational calculus. We shall rather try to explain the idea behind the variational approach in a way not to be found in books devoted to variational calculus.

Let us, however, present at the beginning a somewhat broader view. First of all the variational description is very general, for it may be applied both to a single mass particle and complex mechanical systems; it is not only applicable to mechanical systems, but also to thermal, electromagnetic and biological ones. This generality of description based upon variational principles makes it possible, in our opinion, to treat in a unifying way various disciplines of knowledge, thereby inducing a search for analogies, so important because the current trend to rampant specialization leads to a lack of integration.

Secondly, the variational formalism makes it possible to give up the sometimes unnecessary requirements of smoothness (differentiability) of functions describing the system's state, which is unavoidable for the application of the differential (balancing) formalism.

With such an approach there is a possibility of studying via a unified treatment both continuous and discrete processes (e.g. propagation of shock waves, deformation of composite materials, respectively).

Thirdly, variational methods are the basis on which many other specific methods are founded, such as the *finite element method* or *optimal control methods*, the latter referring primarily to the *Pontryagin maximum principle*.

Fourthly, variational methods make possible the wide use of approximate methods, still useful as the basis of numerical procedures.

Finally and perhaps most important from the point of view of the modeller is the possibility of applying variational principles to construction of models of systems characterized by various properties. It is to this that the present chapter will in fact be devoted.

### 4.1.2 Postulating or deducing the principles?

As is known, in order for a thought construct to be a theory it must be founded on axioms. Then, through deduction one formulates theorems and corollaries. If these corollaries are in agreement with facts known from other sources the theory acquires 'citizenship rights'. From the point of view of logic it is quite immaterial which theorems (statements) are considered axioms, and which are conclusions derived therefrom. Nevertheless, for heuristic reasons it is better to accept as axioms those theorems (statements) that are either more obvious or better known. It seems, for instance, that adoption of Newton's second law without a proof provokes less resistance than adoption of Hamilton's principle. This is perhaps because Newton's second law has a life span longer by about 150 years than Hamilton's principle, and it is the starting point used by the vast majority of textbooks on dynamics. In these textbooks Hamilton's principle is derived starting from Newton's second law and passing through the d'Alembert principle. Nevertheless, one could proceed in the reverse direction, i.e. take Hamilton's principle as a postulate and on this basis derive Newton's second law. This way of proceeding is often used in textbooks on analytical mechanics, in which it is assumed that the reader is acquainted with the calculus of variations. Thus, the principles can be either derived—and this manner of introducing them will be called convention $D$ (**derivation**)—or they can be postulated—convention $P$ (**postulation**). The authors of the present book are actually more inclined towards convention $P$, since this way the fundamental importance of principles is emphasized. However, in order to familiarize the reader with this view we shall demonstrate the equivalence of both these types of procedure. The reasons why we are in favour of convention $P$ are as follows. When we are dealing with problems of classical mechanics the advantages of any of the two conventions cannot be seen, but for integrated mechanics an engineer must deal with multiple coupled phenomena. In such situations usually we do not have a choice, for we do not know, *a priori*, any laws governing the phenomenon considered. Hence, we cannot refer to the procedure conforming with convention $D$. To illustrate this, consider the example of light rays. Every college graduate reader recalls, presumably, the laws of geometrical optics concerning reflection and refraction of light. More inquisitive students may also remember that these laws were simply quoted and then verified experimentally in the physical school lab. These laws, however, can be elegantly deduced by postulating a certain principle. That is just what **Pierre de Fermat** (1601–65) did, and this principle, now called **Fermat's principle**, is as follows:

> *a light beam, when travelling between two points, chooses the path which, compared to any other possible paths between these two points, is covered in the minimal time.*

We will show how the refraction law, which determines what happens at the border of two media, results from Fermat's principle (see Fig. 4.1). A beam from $A$ reaches $B$ after refraction at $R$. With the help of Fig. 4.1 we can calculate the time taken by a light beam travelling between points $A$ and $B$. Thus

$$t = \frac{l_1}{v_1} + \frac{l_2}{v_2}, \tag{4.1}$$

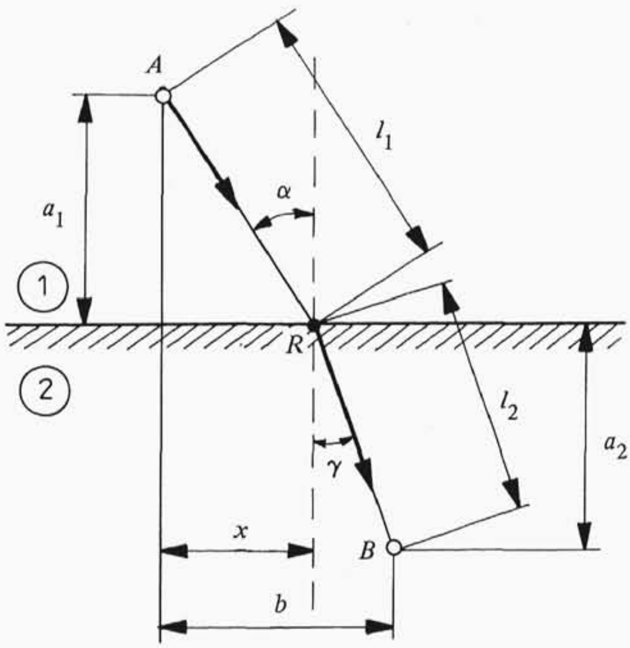


Fig. 4.1.

where $v_1$ and $v_2$ respectively denote the velocity of light in the media 1 and 2. We shall need the notion of optical path; light propagating in a vacuum with velocity $c$ will cover in time $t$ a distance $l = ct$. In the same time in another medium, light will cover a distance of $\lambda = vt$ (where $v$ is the velocity in the given medium). Knowing velocity $v$ and geometrical path $\lambda$ followed by the light beam in a uniform medium we can calculate the distance $l$ that light would cover in the same time in vacuum, namely

$$l = \lambda \frac{c}{v} = \lambda n, \tag{4.2}$$

where $n = c/v$.

Formula (4.1) takes, considering (4.2), the form of

$$t = \frac{n_1 l_1}{c} + \frac{n_2 l_2}{c} = \frac{l}{c},$$

where

$$l = n_1 l_1 + n_2 l_2, \tag{4.3}$$

where $l$ is the *optical path*, $l_1$, $l_2$ *geometrical paths*. It can also be concluded from Fig. 4.1 that

$$l_1 = \sqrt{(a_1^2 + x^2)}, \qquad l_2 = \sqrt{[a_2^2 + (b - x)^2]}. \tag{4.4}$$

Having substituted (4.4) into (4.3) we obtain

$$l = n_1 \sqrt{(a_1^2 + x^2)} + n_2 \sqrt{[a_2^2 + (b - x)^2]}. \tag{4.5}$$

On the basis of Fermat's principle, $l$ must be minimal; note that $l$ is a function of $x$ (changes in $x$ cause changes of incidence and refraction angles). Thus, we can imagine various beam paths leading from $A$ and $B$. Some of them are shown in Fig. 4.2, against the comparative path in Fig. 4.1. Thus, $x$ is selected so as to ensure

$$\frac{\mathrm{d}l}{\mathrm{d}x} = 0. \tag{4.6}$$

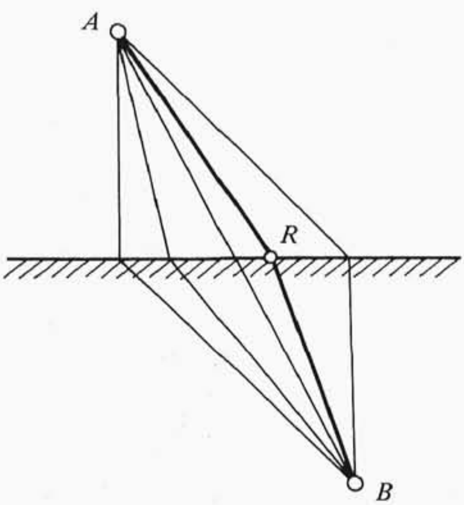


Fig. 4.2.

Having differentiated (4.5) with respect to $x$ we obtain (remembering that locations of points $A$ and $B$ are determined, hence $a_1 = \text{const}$, $a_2 = \text{const}$, $b = \text{const}$):

$$n_1 \frac{x}{\sqrt{(a_1^2 + x^2)}} = n_2 \frac{b - x}{\sqrt{[a_2^2 + (b - x)^2]}}, \tag{4.7}$$

which, on the basis of Fig. 4.1, can be represented in the form of

$$n_1 \sin \alpha = n_2 \sin \gamma,$$

that is

$$\frac{\sin \alpha}{\sin \gamma} = \frac{n_2}{n_1}. \tag{4.8}$$

This is the **refraction law**, called also **Snell's law**. We should therefore remember that only one path from Fig. 4.2 is extremal. The beam will 'select' the path (which is the real path), for which the refraction law is obeyed. Note that the course of the beam is not subject to constraints in both half-spaces and in spite of this there is only one path.

Finally we shall formulate Fermat's principle in a general manner, so as to grasp the core of the variational method. Thus, the optical path of a beam can be presented as

$$S = \int_a^b n \, \mathrm{d}s, \tag{4.9}$$

where the refraction index appears under the integration sign, because in an inhomogeneous medium it is position dependent. The requirement for the optical path to be minimal compared to all the other paths (beginning and terminating at the same points) can be represented in the form

$$\delta S = 0 \qquad \text{with conditions:} \quad \delta q^i(a) = 0, \quad \delta q^i(b) = 0. \tag{4.10}$$

Here in our approach, the symbol $\delta$ appears for the first time. This symbol was introduced by Lagrange to emphasize the virtual character of the variations as opposed to the symbol $d$, which designates differentials.

If a medium is homogeneous (as usually assumed in the fundamental course of physics), then Fermat's principle takes the form

$$\delta \int_a^b \mathrm{d}s = 0 \tag{4.11}$$

for a real trajectory of the light beam, as compared to all the other imaginable paths which begin and end in the same points as the real trajectory.

An important result of this paragraph is finding the real course of the phenomenon with the help of imagined events, and this is actually the essence of variational approach. For Fermat's principle, considered here, the phenomenon in question was the trajectory of a light beam and the alternative events—fictitious beam paths—which could occur but in reality do not do so. It is, however, necessary to define the complete family of contentious paths from which the unique trajectory is selected.

This manner of proceeding is typical in all variational schemes, regardless of the nature of the physical phenomenon. In the subsequent section the approach will be applied to problems of mechanics and in further depth in section 4.2.1.

What can be concluded from considerations presented so far is that although it is true that perception of phenomena from the point of view of variational principles is more accurate, it requires more knowledge. That is why section 4.1.3 has been written in such a way as to introduce an unprepared reader to the universe of notions and methods of variational mechanics without undue formalization.

### 4.1.3 Verbal formulation of Hamilton's principle

Let us consider a particle of mass $m$ moving along the $0x$-axis under the influence of the force

$$F = -\mathrm{grad}\{V(x)\}, \tag{4.12}$$

where the scalar function $V(x)$ depends only on the particle position $x$, and is termed the potential energy of the particle. If, for instance a particle moves near the Earth's surface under the influence of the force of gravitational attraction, its potential energy may be expressed as

$$V(x) = mgx, \tag{4.13}$$

where $x$ is an axis directed vertically upward. Application of Newton's second law to the mass yields:

$$m\ddot{x} = F_x. \tag{4.14}$$

Now taking into account (4.12) and (4.13), we get the equation of motion of the particle:

$$\ddot{x} = -g. \tag{4.15}$$

Assuming $x = x_0$, and $\dot{x} = v_0$ when $t = 0$, the solution (4.15) takes the form

$$x(t) = x_0 + v_0 t - \tfrac{1}{2} g t^2, \tag{4.16}$$

We can introduce here the Lagrangian, $L = T - V$, of the particle under consideration, and

$$L = \tfrac{1}{2} m\dot{x}^2 - mgx = \tfrac{1}{2} m v_0^2 - mgx_0 - 2mv_0 gt + mg^2 t^2. \tag{4.17}$$

Note that the Lagrangian $L$ depends on particle position $x$ and its velocity $\dot{x}$, i.e., generally, $L = L(x, \dot{x})$. Suppose that we are now interested in the motion over a certain time interval $t_1 \le t \le t_2$. The **action in Hamilton's sense**, $S$, of a dynamic system over interval $t_1 \le t \le t_2$ is defined to be

$$S = \int_{t_1}^{t_2} L \, \mathrm{d}t. \tag{4.18}$$

The value of $S$ depends substantially on the functions $x$ and $\dot{x}$ being used for evaluation of $L$. Indeed, if instead of function $x(t)$ describing real motion, any other function $\tilde{x}(t)$ satisfying conditions (see Fig. 4.3):

$$\tilde{x}(t_1) = x(t_1), \qquad \tilde{x}(t_2) = x(t_2) \tag{4.19}$$

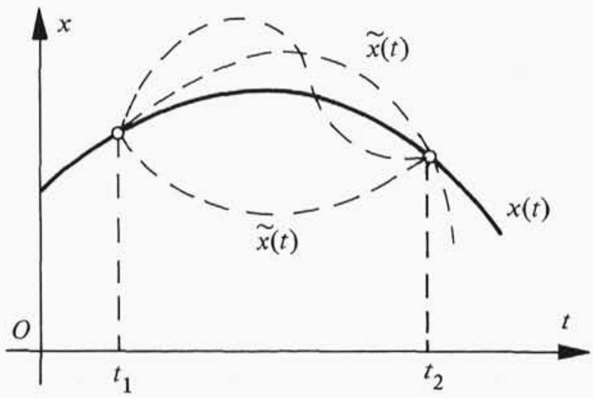


Fig. 4.3.

where substituted into $L$, then, as a rule, we would obtain a different value of the integral $S$. Therefore we say that $S$ is a functional of $x(t)$, which is usually denoted as

$$S = S[x(t)]. \tag{4.20}$$

Another, more practical example of a functional is provided by the mass of a cantilever, whose cross-sectional area $A(x)$ varies along a span $l$ (see Fig. 4.4). Assuming a constant density $\rho$ of the cantilever we get

$$m = \int_0^l \rho A(x)\, \mathrm{d}x. \tag{4.21}$$

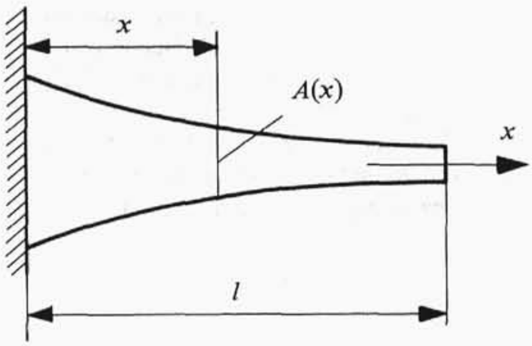


Fig. 4.4.

Using a notation analogous to that of (4.20) we may write

$$m = m[A(x)], \tag{4.22}$$

and note that here mass literally depends on the shape of a cantilever. However, the most fruitful and celebrated functional is that publicized by **Johann Bernoulli** in 1696 within the framework of the **problem of the brachistochrone**:

Two points $O$ and $P$ are given (Fig. 4.5), at different heights but not lying one above the other; it is required to find among all possible curves connecting them, that one along which a particle slides without friction from $O$ to $P$ under the influence of gravity in the shortest possible time.

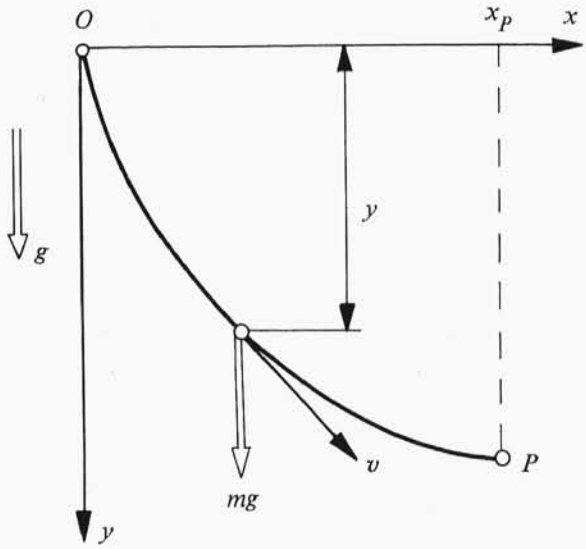


Fig. 4.5.

This problem occupied at the time the leading mathematicians in the whole of Europe: Newton, Leibnitz, Jacob Bernoulli, de L'Hôpital, Hudde and others. From then on, the calculus of variations developed as a special mathematical discipline.

In a suitably chosen coordinate system some such curves $y = f(x)$ joining the points $0$ and $P$ are drawn as possible curves for the fall. Because the motion appears in the homogeneous gravitational field and no resistance forces act on the particle, the principle of conservation of mechanical energy in the form (3.34) may be applied and it gives

$$\tfrac{1}{2} m v^2 = mgy, \tag{4.23}$$

whence

$$v = \sqrt{(2gy)}. \tag{4.24}$$

Remembering that the speed, due to its definition, is $v = \mathrm{d}s/\mathrm{d}t$ and that the element of arc $\mathrm{d}s$ is given by

$$\mathrm{d}s = \sqrt{(1 + y'^2)}\ \mathrm{d}x, \tag{4.25}$$

one obtains (taking account of (4.24))

$$\mathrm{d}t = \frac{\sqrt{(1 + y'^2)}}{\sqrt{(2gy)}}\ \mathrm{d}x. \tag{4.26}$$

Integration of (4.26) gives the time $t$ required for the particle to travel from $O$ to $P$, that is

$$t = \frac{1}{\sqrt{(2g)}} \int_0^{x_P} \sqrt{\left(\frac{1+y'^2}{y}\right)} \, \mathrm{d}x, \tag{4.27}$$

and in the spirit of notation (4.20), this may be written as

$$t = t[y(x)]. \tag{4.28}$$

This means that the sliding time $t$ is a **functional**, since its value depends on the shape of a curve $y(x)$.

The investigation of the brachistochrone led to the problem of finding a function $y(x)$ for which the integral of a second function $f(x, y, y')$ has a smallest, largest, or stationary value, the function being determined by the geometrical, technological or physical domains and called the basic function. In the brachistochrone problem $f(x, y, y') = \sqrt{[(1+y'^2)/y]}$ is the basic functional.

The problem of the calculus of variations is that of finding maxima or minima, and is more difficult than in the differential calculus. The problem is then to find a function for which a certain integral assumes an extremal, or stationary value. The function $y_0(x)$, for which the integral assumes an extreme value, is called an extremal.

Our aim here is not to present the solution of either the particular problem of the brachistochrone or the more general variational problem. However, having traced how specific questions lead in a natural way to the variational problem, we may ask whether other mechanical problems can be formulated in a similar manner. The answer is yes, and for conservative, holonomic dynamical systems, Hamilton's principle provides a framework for variational formulation of many physical problems. It postulates:

*actual path in configuration space between two configurations $q_i(t_1)$ and $q_i(t_2)$ at times $t_1$ and $t_2$, respectively, is that which makes the time integral of the Lagrangian stationary with respect to variations of the path which vanish at end points.*

Thus, this most important integral principle of classical mechanics leads, via the requirement of stationarity, to a variational problem. The requirement of stationarity means that substituting into $L$ in the integral (4.18) different functions fulfilling conditions (4.19) we shall obtain different values, and varying the function about one specific function $x(t)$ shows no change in the value; this function $x(t)$ describes the actual motion of the system.

*Example 4.1.* As an example, consider a case of free fall in a uniform gravitational field. Let the conditions at $t = 0$ be $v_0 = 0$ and $x_0 > 0$. It is known that a particle which is let to fall freely from the height $x_0$ reaches the zero level ($x_0 = 0$) after the time

$$\tau = \sqrt{(2x_0/g)}. \tag{4.29}$$

Consider the class of admissible motions described by

$$\tilde{x} = x_0\left(1 - \left(\frac{t}{\tau}\right)^{\varepsilon}\right), \tag{4.30}$$

where $\varepsilon$ is a positive exponent. Let us check that all motions fulfil the conditions

$$\tilde{x}(0) = x_0, \qquad \tilde{x}(\tau) = 0. \tag{4.31}$$

i.e. all admissible paths pass through given terminal positions.

Fig. 4.6 shows four exemplary admissible curves corresponding to four different values of the parameter $\varepsilon$. Let us find out for what value of the parameter $\varepsilon$ the functional $S[x(t)]$ reaches an extremum, in this case a minimum. In order to do this, we have to calculate $S$ according to the relation (4.18). Thus, step by step we get

$$\dot{x} = -\varepsilon x_0 \frac{t^{\varepsilon-1}}{\tau^{\varepsilon}},$$

$$T = \tfrac{1}{2} m\dot{x}^2 = \tfrac{1}{2} m\varepsilon^2 x_0^2 \frac{t^{2(\varepsilon-1)}}{\tau^{2\varepsilon}},$$

$$V = mgx = mgx_0\left[1-\left(\frac{t}{\tau}\right)^{\varepsilon}\right],$$

$$S = \int_0^{\tau} L\,\mathrm{d}t = \frac{mx_0^2}{2\tau}\frac{\varepsilon^2}{2\varepsilon-1} - mgx_0\tau + mg\tau x_0\frac{1}{\varepsilon+1}. \tag{4.32}$$

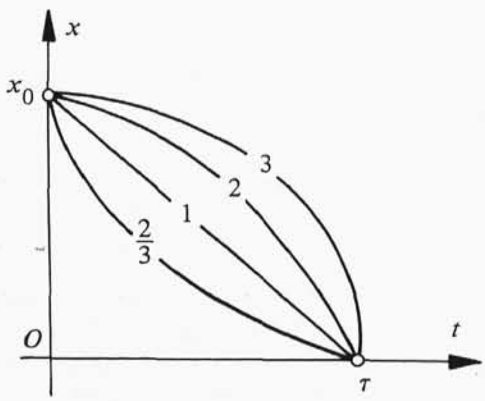


Fig. 4.6.

The expression (4.32) is a function of the parameter $\varepsilon$. Therefore, it is possible to determine its extremum, following the classical procedure,

$$\frac{\mathrm{d}S}{\mathrm{d}\varepsilon} = 0. \tag{4.33}$$

Thus, differentiating (4.32) with respect to $\varepsilon$, then substituting $\tau$ given by (4.29), we obtain the equation

$$\varepsilon^4 + \varepsilon^3 - 9\varepsilon^2 + 7\varepsilon - 2 = 0, \tag{4.34}$$

whose only positive root is $\varepsilon = 2$. This means that the functional $S[x(t)]$ reaches extremum for $\varepsilon = 2$, and the extremal becomes the function

$$x = x_0\left(1 - \left(\frac{t}{\tau}\right)^2\right). \tag{4.35}$$

Having substituted (4.29) into (4.35) we get

$$x = x_0 - \tfrac{1}{2}gt^2, \tag{4.36}$$

i.e. the well-known low of free fall in the uniform gravitational field.

## 4.2 BASIC VARIATIONAL PRINCIPLES

### 4.2.1 Types of principles

Variational principles are conventionally divided into two groups, namely **differential** and **integral** principles. Here they will be classified into **extremal** and **non-extremal** ones, but these divisions are not mutually exclusive. The principle which, it seems, is best known to engineers, the principle of virtual work, is differential and extremal. Another quite popular principle, that of d'Alembert, is differential and non-extremal. A typical example of an integral principle is provided by Hamilton's principle, which has a stationary, that is generally non-extremal, nature. Fermat's principle, known to us from the preceding section, is integral and extremal.

Let us now consider what features dictate the placing of a principle in one group or another, and perhaps even more important; whether and how this influences the manner in which variational principles are applied in modelling?

Let us first note that the distinction into differential and integral principles is not generally accepted by all. There are many authors who consider that only integral principles are variational principles, which is why we deem it proper to present the arguments for acceptance of differential principles as variational. A principle can be considered variational if it contains the requirement of selection from admissible variations. That is, a variational principle considers not just one state (configuration) of the system, but a set of various states (configurations) resulting from carrying out variations that are feasible in terms of constraints (e.g. virtual displacement). Hence the inclusion of d'Alembert's principle and the virtual work principle as differential variational principles is justified. There are of course other principles, but those mentioned here are those which will concern us in detail.

Those, including ourselves, who do in general accept the division into differential and integral principles, quote somewhat different arguments: if a principle relates position, velocity of acceleration of particles of the system in an arbitrary given instant of time, then this principle has a differential nature. If, however, a principle characterizes the motion of the system in a global way, that is, over a certain period of time or space, then this principle has an integral nature. These principles usually require certain functionals, defined on a class of movements given by the boundary conditions to take extremal values. Since the methods of finding the extrema of functionals are provided by the variational calculus, integral principles can also be called variational. In fact, the main