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 point is simply that such differences in viewpoint exist, since belonging to this or another group has no influence upon modelling. If, though, we are organizing knowledge, and we believe that all modellers should do this, we need to go further along the terminological track.

First let us remove the conception that only principles of mechanics that are differential lead to differential equations. It is sufficient to note that Hamilton's principle, an integral principle, yields Lagrange's equations, which are obviously differential (for details see section 4.3.1). There are no fundamental differences between integral and differential principles, but there are some differences in detail; differential variational principles establish the distinction between the real motion and the variational motions at a given instant of time, whereas integral variational principles, on the other hand, establish the distinction between the actual motion and the variational motions over a certain time period.

4.2.2 Fundamental concepts

4.2.2.1 Virtual displacement and admissible variations

It seems to be no exaggeration to say that from all the important concepts in mechanics, that of **virtual displacement** causes the most problems to a beginner in mechanics. One encounters unfortunate phrases such as *virtual displacement takes place in no time* or *it occurs infinitely quickly*. Alternatively, if sufficiently stubborn, he would find a refined definition of the virtual displacement, as *an element of the space $T_g(M)$, that is, the tangent space to the manifold M at point q*. This kind of definition, expressed in the language of modern mathematics, can be encountered, though more by students of physics than by students of engineering. The latter are condemned to deal with notions originating at the turn of the eighteenth century, when statics was just undergoing transformation from the science of simple machines into a branch of mechanics. It is in this period that discussions conducted with **Pierre de Varignon** (1654–1722) led **Johann Bernoulli** (1667–1748) to formulate in 1717, the **Principle of Virtual Work**, in which virtual displacement plays the key role. Thus in the problems of statics, the configuration of a mechanical system does not change with time and that is why virtual displacements do not involve time and this motion has a purely geometrical meaning. However, in dynamics, the science of motion, the configuration changes in time. Nevertheless, virtual displacement does not involve time in dynamics, either. It is very important that from the concept of a displacement not involving time one can still obtain useful and fruitful information for dynamics.

Our suggestion consists in taking the middle way, i.e. in showing that virtual displacement is connected with the notion of variation of a function, the latter being a term well founded in the branch of mathematics called variational calculus, whose origins are due to Johann Bernoulli himself.

Consider therefore a function $y(x)$ of variable x. Let $\tilde{y}(x)$ be a certain another function of the some argument differing from $y(x)$ by an arbitrary quantity in every point of the interval (x_1, x_2). The difference $\tilde{y}(x) - y(x)$ brought about by the change of function form (see Fig. 4.7) is called the **variation of the function** and is denoted by δy. Thus, we have

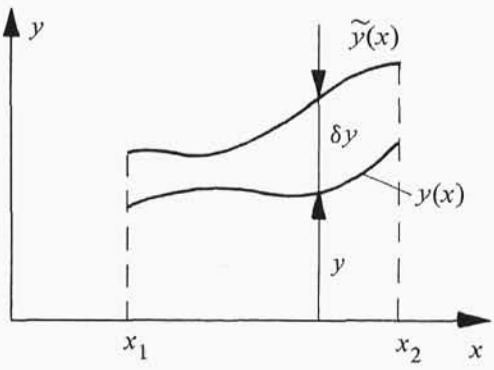

Fig. 4.7.

$$\delta y(x) = \tilde{y}(x) - y(x). \tag{4.37}$$

Let us emphasize an important fact, namely that the variation δy differs from the differential dy in that dy is the increment along the curve y due to an elementary small increment dx, whereas δy is the difference in y between two curves for any given value x. Since argument x does not undergo change in formation of δy, which means that

$$\delta x = 0, \tag{4.38}$$

we refer to variation (4.37) as variation of a function without variation of the argument. This information is important here, but later we will encounter variations where we will have to forgo the condition (4.38).

Note now that if we assign the meaning of time to the independent variable x, it becomes clear how one should understand those unfortunate phrases of *motion in no time*, i.e. simply as denoting the condition of

$$\delta t = 0. \tag{4.39}$$

Variation of a function in which condition (4.39) is preserved will be defined as **synchronous variation.**

Once this is accepted, we can make a subsequently important step in clarifying the concept of virtual displacement. This step consists in introducing the notion of **admissible variations**. Namely, the traditional definition of virtual displacement tells us that it is any imagined small displacement, consistent with any constraints of the system. All, however, becomes understandable when the notion is treated literally, not with verbal expressions, but with mathematical operations. Since this is the most important moment, at least in this section, we shall start with an introductory example.

Imagine a particle constrained to move on the surface determined by the equation

$$f(x, y, z, t) = 0. \tag{4.40}$$

If we assume that a particle is always located on such a surface then equation (4.40) is a constraint equation. In order to express the fact that the particle moves we must introduce its velocity. We shall calculate, for this purpose, the total time derivative of function (4.40):

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial y}\dot{y} + \frac{\partial f}{\partial z}\dot{z} + \frac{\partial f}{\partial t} = 0, \tag{4.41}$$

which, after introduction of the notation

$$\mathbf{v} = (\dot{x}, \dot{y}, \dot{z}), \qquad \operatorname{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right), \tag{4.42}$$

can be expressed in the form

$$\mathbf{v} \cdot \operatorname{grad} f + \frac{\partial f}{\partial t} = 0. \tag{4.43}$$

Note that equation (4.43) describes this essential fact that velocity of a particle is not just arbitrary—the degree of arbitrariness depends to some extent upon the nature of the constraints through quantities $\operatorname{grad} f$ and $\partial f/\partial t$. Dependence upon $\operatorname{grad} f$ is always valid, while the second term is only for rheonomic constraints.

Now let us consider the notion of variation of a function. Since we are dealing with a motion in space of a particle, we have to deal with three functions of one variable, time, i.e. $x(t)$, $y(t)$, $z(t)$. In connection with this one should also take into account three variations, $\delta x(t)$, $\delta y(t)$, $\delta z(t)$. For a vector out of them:

$$\delta\mathbf{r} = (\delta x, \delta y, \delta z). \tag{4.44}$$

Can these variations be arbitrary? They can be partly so, but the particle must follow the constraints constituted by the surface (4.40). The question remains as to the condition which should be fulfilled by vector (4.44) in order for it to describe this fact. The condition can be obtained in the following way: rewrite equation (4.43) in the form

$$\delta\mathbf{r} \cdot \operatorname{grad} f + \frac{\partial f}{\partial t}\,\mathrm{d}t = 0. \tag{4.45}$$

Equation (4.45) defines the differential $\mathrm{d}\mathbf{r}$ of the real displacement $\mathbf{r}$. Now, any vector $\delta\mathbf{r}$ satisfying the equation

$$\delta\mathbf{r} \cdot \operatorname{grad} f = 0 \tag{4.46}$$

will be called a **virtual displacement**.

Note that the fact that the expression $(\partial f/\partial t)\,\mathrm{d}t$ is not preserved between (4.45) and (4.46) means that in the general case (that is in the case of rheonomic constraints) $\delta\mathbf{r} \neq \mathrm{d}\mathbf{r}$. The absence in the definition (4.46) of a term analogous to $(\partial f/\partial t)\delta t$ means that virtual displacement is a synchronous variation which, in verbal descriptions, is expressed through the unfortunate phrases mentioned at the beginning of this section. Variations (4.44) that satisfy the condition (4.46) are called admissible variations, since only such variations are admitted by constraints. Thus, in this perspective, virtual displacements are simply admissible variations.

It remains now only to broaden the relation (4.46) to encompass the general case of holonomic constraints, which are described by the equation

$$f_\alpha(t,\mathbf{r}_\nu)=0, \qquad \alpha=1,\ldots,a, \quad \nu=1,\ldots,n. \tag{4.47}$$

We can presently say that virtual displacement is any vector $\delta\mathbf{r}_\nu=(\delta x_\nu,\delta y_\nu,\delta z_\nu)$ which satisfies the equation

$$\sum_{\nu=1}^{n}\delta\mathbf{r}_\nu\cdot\operatorname{grad}_\nu f_\alpha=0, \tag{4.48}$$

where

$$\operatorname{grad}_\nu f_\alpha=\frac{\partial f_\alpha}{\partial r_\nu}=\left(\frac{\partial f_\alpha}{\partial x_\nu},\frac{\partial f_\alpha}{\partial y_\nu},\frac{\partial f_\alpha}{\partial z_\nu}\right). \tag{4.49}$$

The relations (4.46) or (4.48) can be interpreted in various ways, which are not necessary for modelling and which can be found in most of the handbooks of analytical mechanics. Here two questions remain to be discussed.

(1) What happens to definition (4.48) when generalized coordinates are introduced into the description of motion? The description using position vectors is not used in modelling of technical systems.
(2) What are the limitations on virtual displacements imposed by nonholonomic constraints? These may act alongside holonomic constraints.

Answering the first of these questions is elementary. It is sufficient to recall the discussion of generalized coordinates in section 2.2.3, where any holonomic constraints (4.47) were eliminated. This means that equation (4.48) applies only to the remaining nonholonomic constraints and, consequently, variations of the generalized coordinates can be arbitrary—which is a very significant fact, as we shall see in the modelling of holonomic systems.

The second question becomes simple, too, if we accept the reasoning presented for holonomic constraints. Take, therefore, linear nonholonomic constraints expressed with the help of generalized coordinates in differential form, that is:

$$\sum_{\sigma=1}^{s}B_{\beta\sigma}\,\mathrm{d}q_\sigma+B_\beta\,\mathrm{d}t=0, \qquad \beta=1,\ldots,b \tag{4.50}$$

and thus

$$\sum_{\sigma=1}^{s}B_{\beta\sigma}\delta q_\sigma+B_\beta\delta t=0, \tag{4.51}$$

which, having accepted (4.39), results in

$$\sum_{\sigma=1}^{s}B_{\beta\sigma}\delta q_\sigma=0, \qquad \beta=1,\ldots,b. \tag{4.51}$$

Thus, variations of generalized coordinates in the case of nonholonomic constraints of the type (4.50) are no longer arbitrary and the term 'admissible variations' means those that satisfy condition (4.52).

Before finishing this section, we should present an explanation of the number of degrees of freedom, previously introduced in section 2.2.3. This number should be considered, undoubtedly, to be the number of independent variations of generalized coordinates. In the case of holonomic systems the number of degrees of freedom is equal to the number of independent generalized coordinates. If, therefore, we have s such coordinates, then, on the grounds of constraint conditions (4.52) there are b fewer independent variations. Thus, the number of degrees of freedom of a nonholonomic system is $l = s - b$ (see (2.41)).

4.2.2.2 The mnemonic rule for calculating variations

We shall give now a very useful method for calculating variations of various quantities and this will be used in many transformations. We call this method the mnemonic rule for it reduces to the performance of certain formal operations, whose bases are constituted by the following sequence of manipulations:

(1) consider a certain function

$$F = F(t, x_1, \ldots, x_n), \tag{4.53}$$

written more simply as

$$F = F(t, x_i), \qquad i = 1, \ldots, n; \tag{4.54}$$

(2) form the total differential of this function:

$$\mathrm{d}F = \frac{\partial F}{\partial t}\,\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}\,\mathrm{d}x_i; \tag{4.55}$$

(3) replace the symbol of differentiation 'd' by the symbol of variation δ

$$\delta^{\mathrm{a}} F = \frac{\partial F}{\partial t}\delta t + \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}\delta x_i, \tag{4.56}$$

where the superscript 'a' denotes asynchronous variation;

(4) assume that variation is synchronous (see (4.39)), thus we obtain

$$\delta^{\mathrm{a}} F = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}\delta x_i, \tag{4.57}$$

where the superscript 's' denotes synchronous variation.

The superscripts are omitted in further considerations, for we shall mainly make use of synchronous variation.

To illustrate the validity of this method of calculating variation we shall present several examples.

Example 4.2. The equation of holonomic rheonomic constraints has the form

$$f_\alpha(t, \mathbf{r}_\nu) = 0. \tag{4.58}$$

We are looking for conditions satisfied by variations $\delta\mathbf{r}_\nu$.

By applying, step by step, the rule proposed before, we obtain

$$\mathrm{d}f_\alpha = \frac{\partial f_\alpha}{\partial t}\,\mathrm{d}t + \sum_{\nu=1}^{n} \frac{\partial f_\alpha}{\partial r_\nu}\,\mathrm{d}\mathbf{r}_\nu = 0,$$

$$\frac{\partial f_\alpha}{\partial f}\,\delta t + \sum_{\nu=1}^{n} \frac{\partial f_\alpha}{\partial r_\nu}\,\delta\mathbf{r}_\nu = 0,$$

$$\sum_{\nu=1}^{n} \frac{\partial f_\alpha}{\partial r_\nu}\,\delta\mathbf{r}_\nu = 0. \tag{4.59}$$

This yields the formal definition of virtual displacements (see (4.48)).

Example 4.3. In section 2.2.3 we quoted the condition (2.21) to be satisfied by generalized coordinates. This condition has the form:

$$\mathbf{r}_\nu = \mathbf{r}_\nu(t, q_\sigma), \qquad \nu = 1, \ldots, n, \quad \sigma = 1, \ldots, s.$$

We are now looking for a relation between variations of the radius vector and those of the generalized coordinates.

On the basis of the rule proposed we have, consecutively,

$$\mathrm{d}\mathbf{r}_\nu = \frac{\partial \mathbf{r}_\nu}{\partial t}\,\mathrm{d}t + \sum_{\sigma=1}^{s} \frac{\partial \mathbf{r}_\nu}{\partial q_\sigma}\,\mathrm{d}q_\sigma,$$

$$\delta^{\mathrm{a}}\mathbf{r}_\nu = \frac{\partial \mathbf{r}_\nu}{\partial t}\,\delta t + \sum_{\sigma=1}^{s} \frac{\partial \mathbf{r}_\nu}{\partial q_\sigma}\,\delta q_\sigma, \qquad \delta t = 0,$$

$$\delta\mathbf{r}_\nu = \sum_{\sigma=1}^{s} \frac{\partial \mathbf{r}_\nu}{\partial q_\sigma}\,\delta q_\sigma, \qquad \nu = 1, \ldots, n, \tag{4.60}$$

yielding ultimately a relation which is quite popular in transformations.

Example 4.4. A relation between quasi-velocities and generalized velocities may appear in the form

$$\overset{\circ}{\vartheta}_\sigma = \sum_{\rho=1}^{s} B_{\sigma\rho}\dot{q}_\rho, \qquad \sigma = 1, \ldots, s. \tag{4.61}$$

We are now looking for a relation between variations of quasi-coordinates and generalized coordinates.

Here, before applying the mnemonic rule, we shall first multiply both sides of (4.61) by dt to get

$$\mathrm{d}\vartheta_\sigma = \sum_{\rho=1}^{s} B_{\sigma\rho}\, \mathrm{d}q_\rho. \tag{4.62}$$

Hence, it is sufficient to only carry out step (3):

$$\delta\vartheta_\sigma = \sum_{\rho=1}^{s} B_{\sigma\rho}\delta q_\rho, \qquad \sigma = 1,\ldots, s \tag{4.63}$$

in order to obtain a relation important for deriving equations of motion in quasi-coordinates.

4.2.2.3 Forces of reaction of constraints

The existence of constraints leads to the notion of forces of reaction of constraints. We already know from section 2.2.2 that all constraints limit the freedom of motion of particles of a mechanical system. We shall consider in the present section the manner in which geometric functioning of constraints is expressed in dynamics.

If we require *a priori* that the particle will not leave the surface:

$$f(x, y, z, t) = 0, \tag{4.64}$$

thus th emotion of this particle cannot be determined solely on the basis of the equations of motion for a free particle. It could be that the general solution will not contain the curve which lies on the surface $f = 0$.

In general, we would say that adherence to the condition that a body does not violate constraints (for instance, does not leave a surface) requires a certain action which would oppose the motion of a body directed away from the constraints. Thus, e.g. a railway carriage on a bend pushes on the rails, causing reaction forces in the opposite direction. We may therefore generally assume that the fulfilment of constraint equations is the effect of reaction forces exerted by the devices used for this purpose. Thus, one should distinguish those forces causing motion (which can therefore be called active or given) from those additional forces originated by constraints which will be called **reactions of constraints.**

In order to confirm that the introduction of such forces is necessary, let us note that in the simplest case of a mass particle subject to constraint of the form (4.64), acceleration **a** of this particle has to satisfy the condition

$$\mathbf{a}\cdot\ \mathrm{grad} f = -\frac{\partial^2 f}{\partial t^2} - \mathbf{v}\cdot\ \mathrm{grad}\ \frac{\mathrm{d}f}{\mathrm{d}t}, \tag{4.65}$$

where velocity **v** of the particle is subject to the constraint

$$\mathbf{v}\cdot\ \mathrm{grad} f + \frac{\partial f}{\partial t} = 0. \tag{4.66}$$

(Both these conditions are obtained through differentiation with respect to time of the constraint equation (4.64).) As can be seen from condition (4.65) the component normal to the surface of constraints,

$$\mathbf{a}_\mathrm{n} = -\frac{\operatorname{grad} f}{|\operatorname{grad} f|^2}\left(\frac{\partial^2 f}{\partial t^2} + \mathbf{v}\cdot \operatorname{grad} \frac{\mathrm{d}f}{\mathrm{d}t}\right), \tag{4.67}$$

is determined entirely by the form of the function f appearing in constraint equation (4.64). On the other hand, though, this component has to equal the normal component of acceleration, determined from the equation of motion $m\mathbf{a} = \mathbf{F}$, where $\mathbf{F}$ is the active force (given directly). Hence can be concluded that if the force $\mathbf{F}$ is not given in an implicit way, the equality of the accelerations ($\mathbf{a}_\mathrm{n}$ and $\mathbf{a}$) is out of the question. There is therefore a need for a modification consisting of the introduction of a certain additional force $\mathbf{R}$ known as reaction of the constraints. Thus, the equation of motion for an individual constrained particle, i.e. when the reaction of the constraints is considered, can be presented in the form

$$m\mathbf{a} = \mathbf{F} + \mathbf{R}. \tag{4.68}$$

This equation shows that from the point of view of dynamics, a non-free particle can be treated as a free one, taking place under the influence of given forces and reactions of constraints. The very same statement is the essence of the so-called postulate of freeing from constraints (known also as the postulate of reaction), which is very popular in statics. Our knowledge of constraints then clarifies the meaning of the postulate of reaction. Let us emphasize that this postulate also concerns nonholonomic constraints.

However, in making use of equation (4.68) we encounter a problem, since at this point the reaction of constraints is not known. Only the component of forces of reaction is known through the analytical form of the equation of constraints, and is

$$\mathbf{R}_\mathrm{n} = m\mathbf{a}_\mathrm{n} - \mathbf{F}_\mathrm{n}, \tag{4.69}$$

where $\mathbf{a}_\mathrm{n}$ is given by formula (4.67), and $\mathbf{F}_\mathrm{n}$ is the normal component of force $\mathbf{F}$. We say that constraints are perfect if the total force of reaction of constraints reduces to the form (4.69).

As we have seen, the existence of constraints introduces difficulties in modelling. If we are interested only in the motion of a mechanical system, then reaction forces are only 'hindering' quantities, and so we attempt to eliminate as far as possible.

The most popular method of elimination is based upon the assumption that forces of reaction must be selected in such a way as to make the motion of a system conform with constraints. Since one way of expressing conformation with constraints is as admissible variations, we introduce first the quantity called the *virtual work of forces of reaction of constraints*

$$\delta W = \sum_{\nu=1}^{n} \mathbf{R}_\nu \cdot \delta \mathbf{r}_\nu, \tag{4.70}$$

where n denotes the number of mass particles of the system considered; then we require that

$$\delta W = 0. \tag{4.71}$$

Constraints satisfying condition (4.71) are called **ideal**. Note that this approach is different from the one shown before.

4.2.2.4 Commutability of the variational (δ) *and differential* (d) *operators*

Discussing the variation of a function in section 4.2.2.1 we considered the fact that the operations of calculation of variation and differentiation do not always commute. Let us now have a closer look at this issue. For holonomic systems the property of commutability is valid, since all the trajectories lying in the vicinity of the actual trajectory are kinematically admissible from the point of view of constraint equations. In the case of nonholonomic constraints this is not guaranteed, since not all curves located in the neighbourhood of the actual trajectory are kinematically admissible as comparative trajectories, because they are subject to the nonholonomic constraints. This is due to the fact that variations of generalized coordinates are subject to constraints (4.52). In this situation it is not clear what is the meaning of $\mathrm{d}(\delta d_\sigma)/\mathrm{d}t$, since after leaving the trajectory the derivative along time can no longer be calculated.

In mechanics the operation d denotes differentiation with respect to time. In this connection, it is defined only by points lying on the curve $q_\sigma = q_\sigma(t)$ on which motion takes place. On the other hand the virtual variation operator δ denotes in mechanics any of infinitely many operations which conform with condition (4.52). This is why the operation δ is defined in every point of the space configurations. Thus of the two operations $\delta\mathrm{d}$ and $\mathrm{d}\delta$, only the operation $\mathrm{d}\delta$ is defined at every point lying on an arbitrary (actual or kinematically admissible) trajectory. These considerations can be interpreted geometrically using position vectors (Fig. 4.8, where for the sake of clarity, only one mass particle is considered). Vector $\overrightarrow{BN}$ presented in the figure defines the operations $\delta(\mathbf{r} + \mathrm{d}\mathbf{r})$ which, until now, in accordance with previous considerations, is undefined. This definition means that point N may be attained either via the trajectory ABN or via AMN. Then $\overrightarrow{AN} = \mathrm{d}\mathbf{r} + \delta(\mathbf{r} + \mathrm{d}\mathbf{r}) = \delta\mathbf{r} + \mathrm{d}(\mathbf{r} + \delta\mathbf{r})$ and we obtain

$$\mathrm{d}\delta\mathbf{r} - \delta\mathrm{d}\mathbf{r} = 0. \tag{4.72}$$

If instead of one point we are dealing with a set of n points, equation (4.72) takes the form

$$\mathrm{d}\delta\mathbf{r}_\nu - \delta\mathrm{d}\mathbf{r}_\nu = 0, \qquad \nu = 1,\ldots,n. \tag{4.73}$$

In order to present the equation (4.73) using generalized coordinates, we shall (i) differentiate with respect to time formula (4.60) and (ii) calculate in a formal way the variation of the velocity vector

$$\mathbf{v}_\nu = \frac{\mathrm{d}\mathbf{r}_\nu}{\mathrm{d}t} = \sum_{\sigma=1}^{s} \frac{\partial \mathbf{r}_\nu}{\partial q_\sigma}\dot{q}_\sigma + \frac{\partial \mathbf{r}_\nu}{\partial t}. \tag{4.74}$$

The difference of the two expressions mentioned is

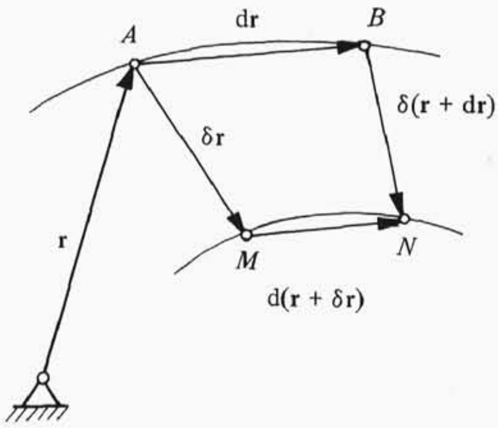

Fig. 4.8.

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{r}_\nu - \delta\frac{\mathrm{d}\mathbf{r}_\nu}{\mathrm{d}t} = \sum_{\sigma=1}^{s}\frac{\partial\mathbf{r}_\nu}{\partial q_\sigma}\left(\frac{\mathrm{d}}{\mathrm{d}t}\delta q_\sigma - \delta\frac{\mathrm{d}q_\sigma}{\mathrm{d}t}\right), \tag{4.75}$$

and in the differential form is

$$\mathrm{d}(\delta\mathbf{r}_\nu) - \delta(\mathrm{d}\mathbf{r}_\nu) = \sum_{\sigma=1}^{s}\frac{\partial\mathbf{r}_\nu}{\partial q_\sigma}\left(\mathrm{d}\delta q_\sigma - \delta\mathrm{d}q_\sigma\right), \qquad \nu = 1,\ldots,n. \tag{4.76}$$

On the basis of (4.75) or (4.76), the difference (4.73) can be presented in the form

$$\mathrm{d}\delta q_\sigma - \delta\mathrm{d}q_\sigma = 0, \tag{4.77}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta q_\sigma - \delta\frac{\mathrm{d}q_\sigma}{\mathrm{d}t} = 0, \qquad \sigma = 1,\ldots,s. \tag{4.78}$$

Equations of the type of (4.77) or (4.78) defining the difference $\mathrm{d}\delta - \delta\mathrm{d}$ are called relations of **commutability** of operations d and δ. Relations obtained are often noted symbolically as '$\mathrm{d}\delta = \delta\mathrm{d}$'.

We wish to emphasize the very important fact that operations d and δ can be arbitrary away from the trajectory $q_\sigma = q_\sigma(t)$ ($\sigma = 1,\ldots, s$), but they have to be precisely. We should say, though, that various solutions of this problem shall lead to different forms of relations of commutability, and consequently to various forms of variational principles and resulting equations (for details see Neimark and Fufaev (1972)).

4.2.2.5 Euler–Lagrange equation

Numerous problems appearing in integrated mechanics can be formulated in terms of extremum (minimum or maximum) of integrated quantities. We have already encountered such notions in sections 4.1.2 and 4.1.3: the optical path expressed via the integral

(4.9), the mass of the rod expressed by the integral (4.21), the time for the sliding down of a particle, integral (4.27). These quantities are called functionals. We shall discuss in the present section, as briefly as possible, the essentials of the variational approach as encountered for such problems. Our main task is to show how to extract the curve for which the given functional takes its extremal values. We will show that this reduces to calculating the variation of a function.

For the sake of simplicity, consider only the one-dimensional case. Let us find a curve $y = y(x)$ which corresponds to the extremum over the segment $x \in (x_1, x_2)$ of the curvilinear integral of the given function $f(x, y, y')$, where $y' = dy/dx$. In other words for the function sought $y(x)$ the integral

$$\int_{x_1}^{x_2} f(x, y, y')\, dx = I[y(x)] \tag{4.79}$$

should take on extremum (minimum or maximum) value. In order to focus our attention later we shall refer uniquely to the minimum. The search for a maximum can be reduced to the search for minimum, since

$$\max I[y(x)] = \min\{-I[y(x)]\}. \tag{4.80}$$

Note that the use of square brackets at (4.79) implies that I depends functionally upon y, and is not just simply a function of points located on this curve. Curves $y(x)$ belong to a certain set Y. Which set of curves is admitted, depends upon the nature of the problem. Thus, for instance, in the problem of the brachistochrone we assumed that the curves have to be smooth. Quite often, though, mainly in the problem of control, it is assumed only that curves are piecewise-smooth. In further considerations, we assume that $y(t)$ is continuous together with its first derivative, and $f(x, y, y')$ is a function having continuous partial derivatives up to the second order inclusively with respect to all the variables.

The type of variational problem is influenced by the form of the integrated function and by the conditions concerning the values of curves at the extremes of the interval, or, as one might say, at their ends. These ends may be free or fixed. If beginnings and ends of all curves y are common (see Fig. 4.9) then we say that we are dealing with the problem with fixed ends. In the present considerations we shall assume that ends are settled, which means that we shall be interested only in such curves $y(x)$ for which

$$y(x_1) = y_1 \qquad \text{and} \qquad y(x_2) = y_2. \tag{4.81}$$

The problem composed of the functional (4.79) and conditions (4.81) will be called the **simplest variational problem.** The problem of the brachistochrone, formulated in section 4.1.3, is of this type.

Thus, the general statement of the simplest type of variational problem is as follows: given function $f(x, y, y')$ we try to find that $y = (x)$ which renders (4.79) minimum, and satisfies the boundary conditions (4.81).

The search for the minimum of the functional $I[y]$ means finding a curve y^* for which

$$I[y^*] \leq I[y], \qquad y \in Y. \tag{4.82}$$

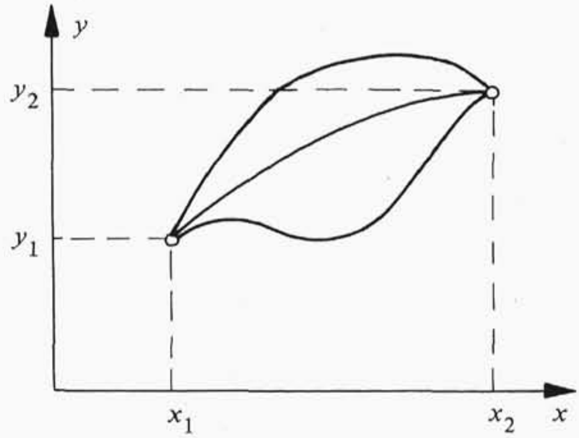

Fig. 4.9.

The question remains: for which curves y does condition (4.82) have to be satisfied? Depending upon the answer to this question we distinguish various types of extrema (for details see Gelfand and Fomin (1963)).

We can now consider another important notion in the framework of the variational approach, namely the notion of variation of a functional, and for this purpose we need the notion of variation of a function, as discussed in section 4.2.2.1. By the variation $\delta y(x)$ of a function $y(x)$ we mean, in general, any difference between a given function $y(x)$ and a function $\tilde{y}(x)$, 'close' to the previous one, thus:

$$\delta y(x) = \tilde{y}(x) - y(x). \tag{4.83}$$

In the case here considered—and recall that we are dealing with fixed ends—variation (4.83) must fulfil yet an additional condition, namely vanishing at the ends of the interval, i.e.

$$\delta y(x_1) = 0, \qquad \delta y(x_2) = 0, \tag{4.84}$$

Frequently, especially in elementary courses, in the place of function $\delta\tilde{y}(x)$ some other function $\eta(x)$ is taken, which is quite arbitrary except that it vanishes at $x = x_1$ and $x = x_2$. Then, if ε is an infinitesimal parameter, the curve defined by $y(x) + \varepsilon\eta(x)$ will be close to $y(x)$ as shown in Fig. 4.10. The function $\varepsilon\eta(x)$, which represents a small change in the overall shape of the original function, is called the variation in y; and therefore according to (4.83) we have

$$\delta y(x) = \varepsilon\eta(x). \tag{4.85}$$

We can also define the difference in the slopes of the y curves at any x to be

$$\delta y' = \tilde{y}' - y'. \tag{4.86}$$

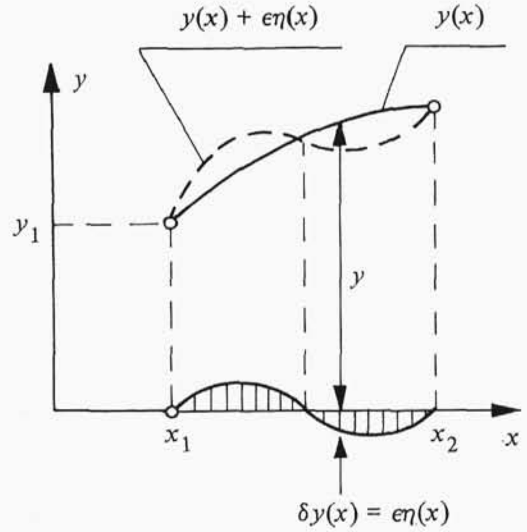

Fig. 4.10.

In order to emphasize the universality of definition (4.86) we have left the second end loose (see Fig. 4.11). However, we could equally well leave both ends loose of fixed.

Now we express the function f in the integral (4.79) along the varied curve $\tilde{y}$ by expanding it about the original curve y. Using the Taylor series we can write

$$f(x, y+\delta y, y'+\delta y') = f(x, y, y') + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial y'}\delta y' + \ \dots \ . \tag{4.87}$$

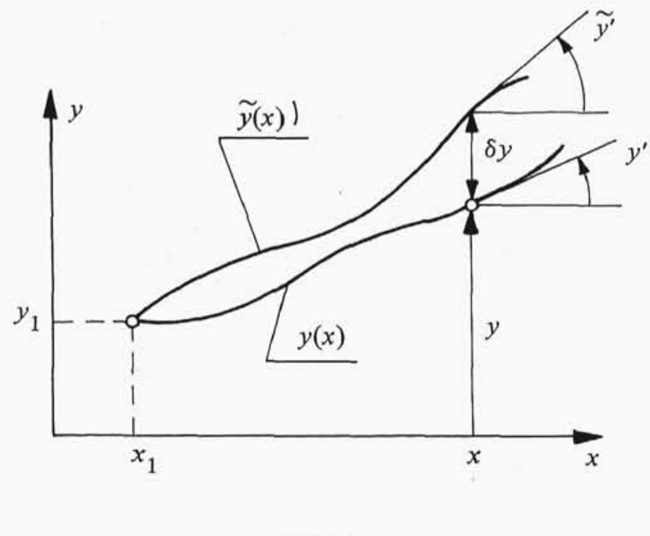

Fig. 4.11.

The quantity

$$\delta f = \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial y'}\delta y' \tag{4.88}$$

is called the **first variation** of function $f(x, y, y')$; this quantity is obtained from (4.87) due to preservation of only its linear components.

Finally if one already gets through the obstacles of variation of a function, the first variation of integral functional I is simply

$$\delta I[y] = \int_{x_1}^{x_2} \delta f \,\mathrm{d}x. \tag{4.89}$$

In order to formalize this crucial notion, for the variational principles for the integral, we shall consider yet another approach using a more natural definition (4.85). For the functional (4.79) we then have

$$I[y + \varepsilon\eta] = \int_{x_1}^{x_2} f(x, y + \varepsilon\eta, y' + \varepsilon\eta')\,\mathrm{d}x, \tag{4.90}$$

so that

$$\Delta I = I[y + \varepsilon\eta] - I[y] = \int_{x_1}^{x_2} [f(x, y + \varepsilon\eta, y' + \varepsilon\eta') - f(x, y, y')]\,\mathrm{d}x. \tag{4.91}$$

If the right-hand side is expanded by the Taylor series we will obtain

$$\Delta I = (\delta I)\varepsilon + \frac{1}{2!}(\delta^2 I)\varepsilon^2 + \ \dots\ . \tag{4.92}$$

Then, δI is called the **first variation in** I, $\delta^2 I$ is called the second variation in I, etc. If we refer to a variation in I without any qualification, we will mean the first variation.

The notion of variation is therefore in a sense analogous to the notion of differential from the conventional differential calculus. The variation of a functional represents the linear part of the increment of the functional when the function $y(x)$ increases by δy (i.e. in a similar manner to the differential of a function $y(x)$ representing the linear part of the increment of this function when independent variable x increases by $\mathrm{d}x$)

$$I[\tilde{y}] = J(\varepsilon) = \int_{x_1}^{x_2} [f(x, y + \varepsilon\eta, y' + \varepsilon\eta')\,\mathrm{d}x, \tag{4.93}$$

where notation $J(\varepsilon)$ has been introduced, since for given $y(x)$ and $\eta(x)$ the integral is now simply a function of ε. Then, the proposition that $I[\tilde{y}]$ be extremal for $\tilde{y}(x) = y(x)$ implies that $J(\varepsilon)$ be extremal at $\varepsilon = 0$. For this latter condition we have

$$J'(\varepsilon)\big|_{\varepsilon=0} = \frac{\mathrm{d}J}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_{x_1}^{x_2} [f(x, y + \varepsilon\eta, y' + \varepsilon\eta')\,\mathrm{d}x\big|_{\varepsilon=0}. \tag{4.94}$$

In (4.92) we briefly introduced the δ notation and we defined δI as the first variation in $I[\tilde{y}]$. In the present case, then, for extremal $I[\tilde{y}]$

$$\delta I = \varepsilon J'(0) = 0, \tag{4.95}$$

so that the extremal character of I and the vanishing of its first variation mean exactly the same thing. It is this manner of proceeding that was made use of in section 4.1.2 in deriving Snell's law from Fermat's principle.

Returning to equations (4.89) and (4.88), we have

$$\delta I = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial y'} \delta y' \right) \mathrm{d}x. \tag{4.96}$$

The task now consists in obtaining the form of (4.99) irrespective of the admissible variation δy. For this purpose the second term in (4.96) can be integrated by parts using the relations

$$\frac{\mathrm{d}}{\mathrm{d}x} \delta y = \delta \frac{\mathrm{d}y}{\mathrm{d}x} \tag{4.97}$$

and

$$\delta \int y \, \mathrm{d}x = \int \delta y \, \mathrm{d}x, \tag{4.98}$$

yielding

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'} \delta y' \, \mathrm{d}x = \frac{\partial f}{\partial y'} \delta y \bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y'} \delta y \, \mathrm{d}x, \tag{4.99}$$

so that the final expression for δI becomes

$$\delta I = \frac{\partial f}{\partial y'} \delta y \bigg|_{x_1}^{x_2} + \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y'} \right) \delta y \, \mathrm{d}x. \tag{4.100}$$

On the basis of equation (4.95) and assumptions (4.84) we obtain

$$\int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y'} \right) \delta y \, \mathrm{d}x = 0. \tag{4.101}$$

Now it suffices to apply the so-called **fundamental lemma** of the variational calculus (sometimes also called the lemma of **Du Bois–Reymond**) in order to complete the work. The essence of the reasoning is as follows: if in equation

$$\int_{x_1}^{x_2} F(x) \eta(x) \, \mathrm{d}x = 0, \tag{4.102}$$

where $\eta(x)$ is an arbitrary function, $F(x)$ does not vanish over the whole interval (meaning that the proof is not direct nor trivial) and is, for instance, positive over a portion of this interval, then, by choosing $\eta(x)$ e.g. as in Fig. 4.12, we would obtain $F(x)\eta(x)\,\mathrm{d}x > 0$, which is contrary to the assumption made. Thus, condition

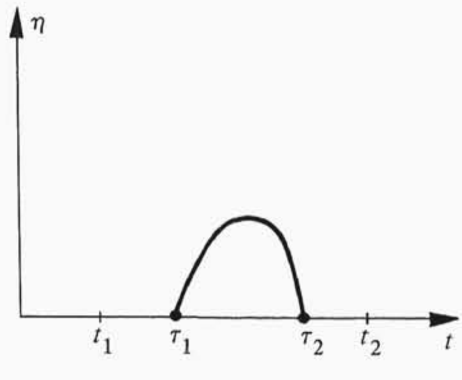

Fig. 4.12.

$$\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} = 0 \tag{4.103}$$

must hold. This is the famous **Euler–Lagrange equation** and its solution is called the **extremal**.

All our results till now may be generalized in a natural way for the multidimensional case—that is, the case of a multidimensional functional space of functions that still depend upon just one variable, x. If we then denote by $\{y\}$ the set of functions $y_1,\ldots,y_n$, we will be analysing the extremum of the functional $I[\{y\}]$. After an adequate generalization of the notions of proximity, variation of a functional etc., we can obtain the Euler–Lagrange equations in the form

$$\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y_i'} = 0, \qquad i = 1,\ldots,n. \tag{4.104}$$

4.2.3 Differential variational principles

4.2.3.1 The common property of differential principles

Before we pass over to consideration of selected differential principles we would like to turn attention to some questions which, in our opinion, are essential. First of all we recall the remark from section 4.2.1 that variations should be understood in a broader sense and must not necessarily mean extremalization. That is why there is no objection to including differential non-extremal principles to variational ones. It is only essential that admissible variations of certain functions appear; in classical mechanics they may be those of the positions of mass particles (the d'Alembert principle), and in thermodynamics they are the variations of the so-called local dissipative potentials (the Onsager principle).

The fact that in section 4.2.2.1 only virtual displacements were considered does not imply that they are the 'construction material' of all the differential principles. True, virtual displacements are the central concept and a difficult one, and that is why they were taken up at the beginning of this section. This should not hinder the proper perception of the fact that other quantities could be equally 'good', for instance virtual velocity