ON THE BRACHISTOCHRONE PROBLEM

OLEG ZUBELEVICH

STEKLOV MATHEMATICAL INSTITUTE OF RUSSIAN ACADEMY OF SCIENCES

DEPT. OF THEORETICAL MECHANICS,
MECHANICS AND MATHEMATICS FACULTY,
M. V. LOMONOSOV MOSCOW STATE UNIVERSITY
RUSSIA, 119899, MOSCOW, MGU
OZUBEL@YANDEX.RU

Abstract. In this article we consider different generalizations of the Brachistochrone Problem in the context of fundamental concepts of classical mechanics. The correct statement for the Brachistochrone problem for nonholonomic systems is proposed. It is shown that the Brachistochrone problem is closely related to vakonomic mechanics.

1. Introduction. The Statement of the Problem

The article is organized as follows.

Section 3 is independent on other text and contains an auxiliary material with precise definitions and proofs. This section can be dropped by a reader versed in the Calculus of Variations.

Other part of the text is less formal and based on Section 3.

The Brachistochrone Problem is one of the classical variational problems that we inherited form the past centuries. This problem was stated by Johann Bernoulli in 1696 and solved almost simultaneously by him and by Christiaan Huygens and Gottfried Wilhelm Leibniz.

Since that time the problem was discussed in different aspects numerous times.

We do not even try to concern this long and celebrated history.

2000 *Mathematics Subject Classification.* 70G75, 70F25,70F20 , 70H30 ,70H03.

Key words and phrases. Brachistochrone, vakonomic mechanics, holonomic systems, nonholonomic systems, Hamilton principle.

The research was funded by a grant from the Russian Science Foundation (Project No. 19-71-30012).

This article is devoted to comprehension of the Brachistochrone Problem in terms of the modern Lagrangian formalism and to the generalizations which such a comprehension involves.

The original version of the problem is as follows. Consider a vertical plane with standard coordinate frame Ox^1x^2. The axis Ox^2 is directed upwards opposite the standard gravity g.

There are two points A, B with coordinates $(0, h)$ and $(l, 0)$ respectively; $h, l > 0$. We can connect these two points by different paths. Let a bead of mass m slide along such a path from A to B without friction. The initial velocity of the bead is zero. What path provides the minimal time of bead's moving from A to B?

So the configuration space of our system is the plane Ox^1x^2. The generalized coordinates are the coordinates of the bead, say (x^1, x^2). This is a system with two degrees of freedom and its Lagrangian is

$$L = \frac{1}{2}m\big((x_t^1)^2 + (x_t^2)^2\big) - mgx^2.$$

Here by the subscript t we denote the derivative in t.

Then we impose an additional ideal constraint that compels the bead to move only along a certain path between A and B. Let this path be given as a parametric equation

$$x = x(s) = (x^1, x^2)^T(s), \quad x(s_1) = A, \quad x(s_2) = B, \quad s \in [s_1, s_2].$$

The law of bead's motion along the path is thus given by the formulas

$$s = s(t), \quad x = x(s(t)).$$

Substituting this equation into the Lagrangian we obtain a system with one degree of freedom and the generalized coordinate s:

$$L'(s, s_t) = \frac{1}{2}m\big((x_s^1(s))^2 + (x_s^2(s))^2\big)s_t^2 - mgx^2(s).$$

This Lagrangian describes the motion of the bead along the chosen path.

That is enough to proceed with the general case.

In the sequel we assume all the functions to be smooth. Recall that a function is smooth in the closed interval $[s_1, s_2]$ if by definition it belongs to $C^\infty(s_1, s_2)$ and all the derivatives are extended to continuous functions in $[s_1, s_2]$.

This assumption is overly strong but we keep it for simplicity of the wording.

1.1. **Holonomic Version of the Problem.** Assume we are given with a Lagrangian system

$$L = T - V, \quad T = \frac{1}{2} x_t^T G(x) x_t, \quad V = V(x). \tag{1.1}$$

Here $x = (x^1, \dots, x^m)^T$ are the local coordinates on a configuration manifold M and $G(x)$ is the matrix of a positively definite quadratic form, $G^T(x) = G(x), \quad \det G(x) \neq 0, \quad \forall x \in M$. All the functions are smooth.

Fix two points $x_1, x_2 \in M$. There are a lot of curves that connect these two points. Let a curve γ be one of them. Assume that this curve is given by the equation

$$x = x(s), \quad x(s_1) = x_1, \quad x(s_2) = x_2. \tag{1.2}$$

So that s is a coordinate in γ.

Impose an additional ideal constraint that makes system (1.1) move along the curve γ only. We obtain a one-degree of freedom system with configuration space γ and the generalized coordinate s. Motion of this system along the curve γ is described by the parameter s such that $s = s(t)$ and

$$x = x(s(t)). \tag{1.3}$$

Substituting this formula in (1.1) we obtain the Lagrangian of this new one-degree of freedom system:

$$L'(s, s_t) = \frac{1}{2}\big(x_s^T(s) G(x(s)) x_s(s)\big) s_t^2 - V(x(s)).$$

We want to choose the curve γ so that system (1.1) spends minimal time passing along γ from x_1 to x_2 at the energy level h:

$$\frac{1}{2}\big(x_s^T(s) G(x(s)) x_s(s)\big) s_t^2 + V(x(s)) = h. \tag{1.4}$$

Let us assume that

$$V(x) < h, \quad \forall x \in M. \tag{1.5}$$

Separating variables in this equation we see that the time of passing is given by the formula

$$\tau(\gamma) = \int_{s_1}^{s_2} \sqrt{\frac{x_s^T(s) G(x(s)) x_s(s)}{2(h - V(x(s))}} ds.$$

Therefore we are looking for a stationary point of the functional τ under the boundary conditions (1.2).

By other words, the Brachistochrone curve is a geodesic of the metric

$$\frac{G(x)}{2(h - V(x))}. \tag{1.6}$$

From proposition 7 (see below) it follows that we can equivalently seek for a stationary point of the functional

$$\tilde{\tau}(x(\cdot))=\int_{s_1}^{s_2}\frac{x_s^T(s)G(x(s))x_s(s)}{2(h-V(x(s)))}ds=\int_{s_1}^{s_2}\frac{T}{h-V}ds$$

with the same boundary conditions.

By the Hamilton principle it follows that the Brachistochrone curve is a trajectory of a dynamical system with Lagrangian

$$\mathcal{L}(x,x_s)=\frac{T}{h-V}.$$

Introduce a function

$$W=-\frac{1}{h-V}.$$

Metric (1.6) presents as follows

$$\frac{G(x)}{2(h-V(x))}=\frac{1}{2}(-W)G.$$

By the principle of least action in the Moupertuis-Euler-Lagrange-Jacobi form [3] we conclude that the Brachistochrone is a trajectory of a system with Lagrangian

$$\tilde{L}=T-W$$

at the zero energy level:

$$T+W=0.$$

1.2. **Holonomic Constraints.** Assume that the Brachistochrone problem is stated for the system with Lagrangian (1.1) and with ideal constraints

$$w(x)=0, \tag{1.7}$$

where $w=(w^1,\dots w^n)^T(x), \quad n<m$ is a vector of smooth functions in M such that

$$\operatorname{rang}\frac{\partial w}{\partial x}(x)=n,\quad \forall x\in M.$$

This statement brings nothing new: equations (1.7) define a smooth submanifold N in M and all constructed above theory works for the corresponding Lagrangian system without constraints on N.

2. The Brachistochrone Problem with Differential Constraints

2.1. **Discussion.** Assume that the Lagrangian system (1.1) is equipped with ideal differential constraints

$$B(x)x_t = 0. \tag{2.1}$$

Here the matrix

$$B(x) = \begin{pmatrix} b_1^1(x) & b_2^1(x) & \cdots & b_m^1(x) \\ b_1^2(x) & b_2^2(x) & \cdots & b_m^2(x) \\ \vdots & \vdots & \ddots & \vdots \\ b_1^n(x) & b_2^n(x) & \cdots & b_m^n(x) \end{pmatrix},$$

here

$$\operatorname{rang} B(x) = n < m, \quad \forall x \in M.$$

Recall that if system (2.1) can equivalently be presented in the form (1.7) then it is referred to as a holonomic system. Otherwise (2.1) is a nonholonomic system. What the word "equivalently" means and when such a presentation is possible is a content of the Frobenius theorem [10].

As above we substitute equation (1.3) into Lagrangian (1.1) and separating the variables in the energy integral (1.4) we get the functional τ. Then we notice condition (1.5).

But now we must pick only motions that satisfy constraints (2.1). Substituting (1.3) in (2.1) we have

$$B(x(s(t))x_s(s(t))s_t = 0.$$

It follows that

$$B(x)x_s = 0. \tag{2.2}$$

We postpone the discussion of the boundary conditions for a while but note that by the same reason (proposition 7) we can replace τ with $\tilde{\tau}$.

Now let us observe that we have obtained the problem of stationary points for the functional $\tilde{\tau}$ defined on a set of curves $x = x(s)$ that obey constraints (2.2).

This problem provoked a lot of confusion and mistakes in classical mechanics. Many researches thought that such a variational problem is equivalent to the Lagrange-d'Alembert equations for the mechanical system with the Lagrangian $\mathcal{L}$ and ideal constraints (2.2) (the variable s plays a role of time). If constrains (2.2) are nonholonomic it is not so.

Here is what in this concern Bloch, Baillieul, Crouch and Marsden write in [5]:

It is interesting to compare the dynamic nonholonomic equations, that is, the Lagrange-dAlembert equations with the corresponding variational nonholonomic equations. The distinction between these two different systems of equations has a long and distinguished history going back to the review article of Korteweg [6] *and is discussed in a more modern context in Arnold, Kozlov, and Neishtadt* [4]. *(For Kozlovs work on vakonomic systems see, e.g.,* [7] *and* [8]*)*

As Korteweg points out, there were many confusions and mistakes in the literature because people were using the incorrect equations, namely the variational equations, when they should have been using the Lagrange-dAlembert equations; some of these misunderstandings persist, remarkably, to the present day. The upshot of the distinction is that the Lagrange-dAlembert equations are the correct mechanical dynamical equations, while the corresponding variational problem is asking a different question, namely one of optimal control.

Perhaps it is surprising, at least at first, that these two procedures give different equations. What, exactly, is the difference in the two procedures? The distinction is one of whether the constraints are imposed before or after taking variations. These two operations do not, in general, commute. We shall see this explicitly with the vertical rolling disk in the next section. With the dynamic Lagrange-dAlembert equations, we impose constraints only on the variations, whereas in the variational problem we impose the constraints on the velocity vectors of the class of allowable curves.

In case of differential constraints (2.2) the situation with boundary conditions is much more complicated than (1.2).

The main question is as follows. Assume that the points $x_1, x_2 \in M$ are connected with a curve $x(s)$. This curve is a stationary point of τ or $\tilde\tau$ on the set of curves that satisfy constraints (2.2) and connect x_1, x_2.

Is the collection of other smooth paths that connect x_1, x_2 and satisfy (2.2) large enough to reduce the variational problem to the differential equations? Or by other words, is this collection large enough to construct the Lagrange multipliers method? In general the answer is "no". See remark 1 below.

The author does not know whether the situation will be fixed if we demand the constraints to be completely nonholonomic. Such questions seem to be closely related to the Rashevsky-Chow theorem [9], [2].

To avoid these hard questions we suggest considering the Brachistochrone Problem with another boundary conditions which guarantee the correct employment of the Lagrange multipliers method in the case of nonholonomic constraints.

2.2. **The Lagrange multipliers method.** Let us restrict our attention to the case when $M \subset \mathbb{R}^m = \{x = (x^1, \dots, x^m)^T\}$ is a domain.

We assume that perhaps after some rearrangement the coordinates can be split in two parts

$$x = \begin{pmatrix} y^1 \\ \vdots \\ y^n \\ z^1 \\ \vdots \\ z^{m-n} \end{pmatrix} = \begin{pmatrix} y \\ z \end{pmatrix}, \tag{2.3}$$

such that

$$B(x)x_s = R(x)y_s + Q(x)z_s, \tag{2.4}$$

where R is an $n \times n$ matrix and $\det R(x) \neq 0, \quad \forall x \in M$.

Representation (2.3), (2.4) is not unique.

Let us impose the boundary conditions:

$$x(s_1) = x_1 = (y_1^T, z_1^T)^T \in M, \quad z(s_2) = z_2. \tag{2.5}$$

Here x_1, z_2 are fixed,

$$\{(y^T, z^T)^T \mid z = z_2\} \cap M \neq \emptyset;$$

and we put no restrictions on $y(s_2)$.

Thus the different choice of representation (2.3), (2.4) leads to physically different statements of the problem.

Theorem 1. *Let $\tilde{x}(s)$ be a stationary point of the functional $\tilde{\tau}$ on the set of functions $x(s)$ that satisfy (2.2), (2.5).*

Then there is a smooth function $\lambda(s) = (\lambda_1, \dots, \lambda_n)(s)$ such that $\tilde{x}$ satisfies the equations

$$\frac{d}{ds}\frac{\partial \mathcal{L}^*}{\partial x_s} - \frac{\partial \mathcal{L}^*}{\partial x} = 0, \quad \mathcal{L}^*(s, x, x_s) = \frac{x_s^T G(x) x_s}{2(h - V(x))} + \lambda(s)B(x)x_s, \tag{2.6}$$

and

$$\frac{p(\tilde{x}(s_2), \tilde{x}_s(s_2))}{h - V(\tilde{x}(s_2))} + \lambda(s_2)R(\tilde{x}(s_2)) = 0, \tag{2.7}$$

where

$$p(x, x_s) = \frac{\partial}{\partial y_s}\Big(x_s^T G(x) x_s\Big).$$

This theorem is a direct consequence from theorem 2.

By proposition 1 the stationary point $\tilde{x}$ preserves the "energy":

$$\frac{\tilde{x}_s^T(s) G(\tilde{x}(s)) \tilde{x}_s(s)}{2(h - V(\tilde{x}(s)))} = \text{const}.$$

Show that system (2.6), (2.2) can be presented in the normal form that is

$$x_{ss} = \Psi_1(x, x_s, \lambda), \quad \lambda_s = \Psi_2(x, x_s, \lambda). \tag{2.8}$$

Indeed, system (2.6) takes the form

$$x_{ss}^T + \lambda_s \tilde{B} = \alpha(x, x_s, \lambda), \quad \tilde{B} = B\Big(\frac{G}{h - V}\Big)^{-1}. \tag{2.9}$$

Differentiate (2.2) to have

$$Bx_{ss} + \gamma(x, x_s) = 0. \tag{2.10}$$

Substituting the second derivatives from (2.9) to (2.10) we obtain

$$\lambda_s \tilde{B} B^T = \psi(x, x_s, \lambda). \tag{2.11}$$

It is clear $\det \tilde{B} B^T \ne 0$ and we can express λ_s from (2.11) and plug it in (2.9).

If we choose $y(s_2)$ and $z_s(s_2)$ then $\lambda(s_2)$ and $y_s(s_2)$ are defined from equations (2.7), (2.2); and at least for $s_2 - s_1$ small, we can solve the Cauchy problem for (2.9), (2.11) backwards. Thus the problem is to choose $y(s_2)$ and $z_s(s_2)$ so that the boundary conditions $x(s_1) = x_1$ are satisfied.

Once representation (2.3), (2.4) fixed the problem meets the Newton principle of determinacy: initial conditions

$$x(s_2), x_s(s_2)$$

determine the trajectory $x(s)$ uniquely. Indeed, from equation (2.7) one finds the value $\lambda(s_2)$. This completes the statement of the Cauchy problem for (2.8).

3. Some Useful Facts From the Calculus of Variations

Here we collect several standard facts from the Calculus of Variations.

Let $\Omega \subset \mathbb{R}^m$ be an open domain with standard coordinates

$$x = (x^1, \dots, x^m)^T.$$

To proceed with formulations we split the vector x in two parts as above (2.3).

Let $F : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be a smooth function.

We are about to state the variational problem for the functional

$$\mathcal{F}\big(x(\cdot)\big) = \int_{s_1}^{s_2} F(x(s), x_s(s))ds \tag{3.1}$$

with boundary conditions

$$z(s_1) = \hat{z}_1, \quad z(s_2) = \hat{z}_2, \quad y(s_1) = \hat{y}_1, \quad s_1 < s_2 \tag{3.2}$$

and constraints

$$a(x, x_s) = 0. \tag{3.3}$$

Here $a = (a^1, \dots, a^n)^T$ is a vector of functions that are smooth in $\Omega \times \mathbb{R}^m$.

There also must be

$$(\hat{y}_1^T, \hat{z}_1^T)^T \in \Omega, \quad \{(y^T, z^T)^T \mid z = \hat{z}_2\} \cap \Omega \neq \emptyset.$$

Assume that

$$\det \frac{\partial a}{\partial y_s}(x, x_s) \neq 0, \quad (x, x_s) \in \Omega \times \mathbb{R}^m \tag{3.4}$$

and equation (3.3) can equivalently be written as

$$y_s = \Phi(y, z, z_s).$$

Definition 1. *Let a smooth function*

$$\tilde{x} : [s_1, s_2] \to \Omega, \quad \tilde{x}(s) = (\tilde{y}^T, \tilde{z}^T)^T(s)$$

be such that

$$a(\tilde{x}(s), \tilde{x}_s(s)) = 0, \quad \tilde{x}(s_1) = \hat{x}_1 = (\hat{y}_1^T, \hat{z}_1^T)^T, \quad \tilde{z}(s_2) = \hat{z}_2.$$

We shall say that $\tilde{x}$ is a stationary point of functional (3.1) with constraints (3.3) and boundary conditions (3.2) if the following holds.

For any smooth function

$$X : [s_1, s_2] \times (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}^m, \quad X(s, \varepsilon) = (Y^T, Z^T)^T(s, \varepsilon), \quad \varepsilon_0 > 0$$

such that

1) $X\big([s_1, s_2] \times (-\varepsilon_0, \varepsilon_0)\big) \subset \Omega$*;*
2) $X(s, 0) = \tilde{x}(s), \quad s \in [s_1, s_2]$*;*
3) $Z(s_1, \varepsilon) = \hat{z}_1, \quad Z(s_2, \varepsilon) = \hat{z}_2, \quad Y(s_1, \varepsilon) = \hat{y}_1, \quad \varepsilon \in (-\varepsilon_0, \varepsilon_0)$ *;*
4) $a(X(s, \varepsilon), X_s(s, \varepsilon)) = 0, \quad (s, \varepsilon) \in [s_1, s_2] \times (-\varepsilon_0, \varepsilon_0)$
we have

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{F}\big(X(\cdot, \varepsilon)\big) = 0.$$

The functions X with properties 1)-4) are referred to as variations.

Theorem 2 ([1])**.** *If the function $\tilde{x}$ is a stationary point of functional (3.1) with constraints (3.3) and boundary conditions (3.2) then there is a smooth function $\lambda(s) = (\lambda_1, \dots, \lambda_n)(s)$ such that $\tilde{x}$ satisfies the equations*

$$\frac{d}{ds}\frac{\partial F^*}{\partial x_s} - \frac{\partial F^*}{\partial x} = 0, \quad F^*(s, x, x_s) = F(x, x_s) + \lambda(s) a(x, x_s),$$

and

$$\frac{\partial F}{\partial y_s}(\tilde{x}(s_2), \tilde{x}_s(s_2)) + \lambda(s_2)\frac{\partial a}{\partial y_s}(\tilde{x}(s_2), \tilde{x}_s(s_2)) = 0. \tag{3.5}$$

This theorem remains valid if the functions a, F depend on s.

For completeness of the exposition sake we prove this theorem in section 3.2.

3.1. **The Homogeneous Case.** In this section it is reasonable to assume the second argument of the functions a, F to be defined on a conic domain $K \subset \mathbb{R}^m$. All the formulated above results and the argument of section 3.2 remain valid under such an assumption.

Recall that by definition the domain K is a conic domain iff

$$x \in K \Longrightarrow \alpha x \in K, \quad \forall \alpha > 0.$$

Proposition 1. *Assume that the function a is homogeneous in the second argument:*

$$a(x, \alpha x_s) = \alpha a(a, x_s), \quad \forall \alpha > 0, \quad \forall (x, x_s) \in \Omega \times K. \tag{3.6}$$

Then the stationary point $\tilde{x}$ preserves the "energy":

$$H(x, x_s) = \frac{\partial F}{\partial x_s} x_s - F$$

that is

$$H(\tilde{x}(s), \tilde{x}_s(s)) = \text{const}.$$

Proof of Proposition 1. Consider a function

$$X(s, \varepsilon) = \tilde{x}(s + \varepsilon \varphi(s))$$

with a smooth function φ such that $\operatorname{supp} \varphi \subset [s_1 + s', s_2 - s']$ and

$$|\varepsilon|, s' > 0$$

are small enough.

The function X satisfies all the conditions of Definition 1. Property (3.6) is needed to check condition 4) of Definition 1.

Furthermore we have

$$\begin{aligned}
X &= \tilde{x}(s) + \varepsilon \varphi(s) \tilde{x}_s(s) + O(\varepsilon^2), \\
X_s &= \tilde{x}_s(s) + \varepsilon \big(\varphi_s(s) \tilde{x}_s(s) + \varphi(s) \tilde{x}_{ss}(s) \big) + O(\varepsilon^2)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{F}(X(\cdot, \varepsilon)) \\
&\quad = \int_{s_1}^{s_2} \Big(\varphi(s) \frac{d}{ds} F(\tilde{x}, \tilde{x}_s) + \varphi_s(s) \frac{\partial F(\tilde{x}, \tilde{x}_s)}{\partial x_s} \tilde{x}_s(s) \Big) ds \\
&\quad = \int_{s_1}^{s_2} H(\tilde{x}(s), \tilde{x}_s(s)) \varphi_s(s) ds = 0.
\end{aligned}$$

Here we use integration by parts.

Since φ is an arbitrary function the proposition is proved.

3.1.1. *The Case of Both Functions a, F Homogeneous in x_s.*

Proposition 2. *Let $\tilde{x}(s)$ be a stationary point of $\mathcal{F}$. Define a function*

$$\bar{x}(\xi) = \tilde{x}(f(\xi)),$$

where $f : [\xi_1, \xi_2] \to [s_1, s_2]$ is a smooth function,

$$f_\xi(\xi) > 0, \quad f(\xi_r) = s_r, \quad r = 1, 2.$$

Then $\bar{x}$ is a stationary point of $\mathcal{F}$ with the integral taken over $[\xi_1, \xi_2]$.

Indeed, such a reparameterization neither changes the shape of constraints (3.3) nor the shape of integral (3.1):

$$\int_{s_1}^{s_2} F(\tilde{x}(s), \tilde{x}_s(s))ds = \int_{\xi_1}^{\xi_2} F(\bar{x}(\xi), \bar{x}_\xi(\xi))d\xi.$$

Proposition 3. *If $F(\tilde{x}(s), \tilde{x}_s(s)) > 0, \quad s \in [s_1, s_2]$ then for any constant $c > 0$ we can choose a parametrization of $\tilde{x}$ such that*

$$F(\bar{x}(\xi), \bar{x}_\xi(\xi)) = c, \quad \xi \in [\xi_1, \xi_2].$$

Indeed, the desired parametrization is obtained from the equation

$$\frac{d\xi}{ds} = \frac{1}{c} F(\tilde{x}(s), \tilde{x}_s(s)).$$

Introduce a functional

$$\mathcal{P}\big(x(\cdot)\big) = \int_{s_1}^{s_2} \big(F(x(s), x_s(s))\big)^2 ds$$

with the same constraints and boundary conditions (3.3), (3.2).

Proposition 4. *Let $\tilde{x}(s)$ be a stationary point of $\mathcal{F}$ and assume that $\tilde{x}(s)$ is parametrized in accordance with Proposition 3:*

$$F(\tilde{x}(s), \tilde{x}_s(s)) = c.$$

Then $\tilde{x}(s)$ is a stationary point of $\mathcal{P}$.

Indeed,

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{P}\big(X(\cdot, \varepsilon)\big) = 2c \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{F}\big(X(\cdot, \varepsilon)\big) = 0. \tag{3.7}$$

Proposition 5. *Let $x^*(s)$ be a stationary point of $\mathcal{P}$. Then F is the energy integral:*

$$F(x^*(s), x_s^*(s)) = \text{const}.$$

Indeed, since F is homogeneous in x_s it follows that

$$\frac{\partial F}{\partial x_s}x_s = F.$$

And the assertion follows from Proposition 1:

$$H = \frac{\partial F^2}{\partial x_s}x_s - F^2 = F^2.$$

Proposition 6. *Let $x^*(s)$ be a stationary point of $\mathcal{P}$ such that*

$$F(x^*(s), x^*_s(s)) = c > 0.$$

Then it is a stationary point of $\mathcal{F}$.

Indeed, it follows from (3.7).

Summing up we obtain the following proposition.

Proposition 7. *If $\tilde{x}(s)$ is a stationary point of $\mathcal{F}$ and $F(\tilde{x}(s), \tilde{x}_s(s)) > 0$ then after some reparametrization it is a stationary point of $\mathcal{P}$.*

If $x^*(s)$ is a stationary point of $\mathcal{P}$ and $F(x^*(s), x^*_s(s)) = c > 0$ then it is a stationary point of $\mathcal{F}$.

3.2. **Proof of Theorem 2.** Introduce a notation

$$[F]_y = -\frac{d}{ds}\frac{\partial F}{\partial y_s} + \frac{\partial F}{\partial y}, \quad [F]_z = -\frac{d}{ds}\frac{\partial F}{\partial z_s} + \frac{\partial F}{\partial z}$$

and correspondingly $[F]_x = ([F]_y, [F]_z)$.

Let us put $Z(s,\varepsilon) = \tilde{z}(s) + \varepsilon\delta z(s)$,

$$\operatorname{supp}\delta z \subset [s_1, s_2]. \tag{3.8}$$

Then the function Y is uniquely determined from the following Cauchy problem

$$Y_s(s,\varepsilon) = \Phi(Y(s,\varepsilon), Z(s,\varepsilon), Z_s(s,\varepsilon)), \quad Y(s_1,\varepsilon) = \hat{y}_1. \tag{3.9}$$

Particularly, it may turn out that $Y_\varepsilon(s_2, 0) \neq 0$.

Remark 1. *That is why we can not impose condition $x(s_2) = \hat{x}_2$ as it is usually done for the holonomic case. The value $Y(s_2,\varepsilon)$ has already been uniquely defined by other boundary conditions and the constraints. In other words if we add the condition $Y(s_2,\varepsilon) = \hat{y}_2$ to the conditions 1)-4) of Definition 1 then the set of variations $\{X(s,\varepsilon)\}$ may turn up to be insufficiently large to prove theorem 2.*

For example, consider a plane $\mathbb{R}^2 = \{x = (y,z)^T\}$. There is a unique smooth path from $x_1 = (0,0)^T$ to $x_2 = (0,1)^T$ that satisfies the equation $y_s = 0$. (Much more complicated example by C. Caratheodory see in [4]*.)*

Cauchy problem (3.9) has the suitable solution at least for $|\varepsilon|$ and $s_2 - s_1$ small. Observe also that

$$Y_\varepsilon(s_1, \varepsilon) = 0. \tag{3.10}$$

Using the standard integration by parts technique and from formulas (3.10), (3.8) we obtain

$$\begin{aligned}\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} &\mathcal{F}\big(X(\cdot, \varepsilon)\big) \\ &= \int_{s_1}^{s_2} \Big([F]_z \delta z + [F]_y Y_\varepsilon\Big) ds + \frac{\partial F}{\partial y_s}(\tilde{x}(s_2), \tilde{x}_s(s_2)) Y_\varepsilon(s_2, 0) = 0.\end{aligned} \tag{3.11}$$

The function $\lambda(s)$ is still undefined but due to condition (3.4) the value $\lambda(s_2)$ is determined uniquely from (3.5).

From condition 4) of definition 1 it follows that

$$A(\varepsilon) = \int_{s_1}^{s_2} \lambda(s) a(X(s, \varepsilon), X_s(s, \varepsilon)) ds = 0.$$

By the same argument as above we have

$$\begin{aligned}\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} &A \\ &= \int_{s_1}^{s_2} \Big([\lambda a]_z \delta z + [\lambda a]_y Y_\varepsilon\Big) ds \\ &+ \lambda(s_2) \frac{\partial a}{\partial y_s}(\tilde{x}(s_2), \tilde{x}_s(s_2)) Y_\varepsilon(s_2, 0) = 0.\end{aligned} \tag{3.12}$$

Summing formulas (3.12) and (3.11) we yield

$$\int_{s_1}^{s_2} \Big([F^*]_z \delta z + [F^*]_y Y_\varepsilon\Big) ds = 0. \tag{3.13}$$

To construct the function λ consider an equation

$$[F^*]_y = 0. \tag{3.14}$$

This is a system of linear ordinary differential equations for λ. Due to assumption (3.4) this system can be presented in the normal form that is

$$\lambda_s = \Lambda(s, \lambda).$$

Since we know $\lambda(s_2)$, by the existence and uniqueness theorem we obtain $\lambda(s)$ as a solution to the IVP for (3.14).

Equation (3.13) takes the form

$$\int_{s_1}^{s_2} [F^*]_z \delta z ds = 0.$$

Since δz is an arbitrary function we get $[F^*]_z = 0$. Together with (3.14) this proves the theorem.

References

[1] N. I. Akhiezer: The Calculus of Variations. Blaisdell, New York, 1962.

[2] W.L. Chow, Uber Systeme von linearen partiellen Differentialgleichungen erster Ordnung, Math. Ann. 117 (1940) 98-115.

[3] V. I. Arnold: Mathematical Methods of Classical Mechanics. Second Edition Translated by K. Vogtmann and A. Weinstein. Springer, 1989.

[4] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt: Dynamical Systems III, Springer- Verlag, Encyclopedia of Math., 3 1988.

[5] A. M. Bloch, J. Baillieul, P. Crouch, J. Marsden: Nonholonomic Mechanics and Control (Interdisciplinary Applied Mathematics). Springer, 2000.

[6] D. Korteweg: Uber eine ziemlich verbreitete unrichtige Behandlungsweise eines Problemes der rollenden Bewegung und insbesondere ber kleine rollende Schwingungen um eine Gleichgewichtslage, Nieuw Archiefvoor Wiskunde. 4, 130155 1899.

[7] V. V. Kozlov: Realization of nonintegrable constraints in classical mechanics, Sov. Phys. Dokl. 28, 735737 1983.

[8] V. V. Kozlov: The problem of realizing constraints in dynamics, J. Appl. Math. Mech. 56 (4), 594600 1992.

[9] P.K. Rashevsky, Any two points of a totally nonholonomic space may be connected by an admissible line, Uch. Zap. Ped. Inst. im. Liebknechta, Ser. Phys. Math. 2 (1938) (in Russian).

[10] S. Sternberg: Lectures on Differential Geometry. 1964 by Presentice-Hall, inc., Englewood Cliffs, New Jersey.

Email address: `ozubel@yandex.ru`