Arch. Mech., 39, 6, pp. 645–662, Warszawa 1987

Algorithms for reactions of nonholonomic constraints and servo-constraints

J. J. SŁAWIANOWSKI (WARSZAWA)

We discuss and compare various procedures of deriving equations of motion of constrained mechanical systems. A geometric interpretation of these procedures is given. Special stress is laid on nonholonomic constraints, both linear and nonlinear. We analyse certain qualitative differences between models of nonholonomic dynamics based on different procedures. Two algorithms are particularly interesting, namely, 1) the d'Alembert principle and its Appell-Tshetajev generalization (used in typical applications, e.g. in rolling problems), and ii) the variational Hamilton principle with subsidiary conditions (more interesting for mathematicians). We argue that the Hamilton principle, although not accepted in traditional technical applications, is more promising, at least as a guiding hint, in generalizations concerning systems with higher differential constraints or more general functional constraints appearing in feedback and control systems.

Przedyskutowano i porównano rozmaite procedury wyprowadzania równań ruchu dla układów mechanicznych z więzami. Przedstawiono geometryczną interpretację tych procedur. Szczególny nacisk położono na więzy nieholonomiczne, zarówno liniowe, jak nieliniowe. Omówiono niektóre jakościowe różnice między modelami dynamiki nieholonomicznej opartymi na różnych procedurach. Szczególne znaczenie mają dwa algorytmy, mianowicie: i) zasada d'Alemberta i jej uogólnienie podane przez Appella i Czetajewa (zasady używane w typowych zastosowaniach, jak zagadnienia toczenia), oraz ii) wariacyjna zasada Hamiltona z dodatkowymi warunkami (ciekawsza z matematycznego punktu widzenia). Utrzymujemy, że zasada Hamiltona, chociaż niestosowna w tradycyjnych zagadnieniach technicznych, jest jednak bardziej obiecująca, przynajmniej jako heurystyczna wskazówka, w uogólnieniach dotyczących układów z wyższymi więzami różniczkowymi, lub jeszcze ogólniejszymi więzami typu funkcjonalnego, pojawiającymi się w zagadnieniach autoregulacji i sterowania.

Обсуждены и сравнены разные процедуры вывода уравнений движения для механических систем со связями. Представлена геометрическая интерпретация этих процедур. Особенное внимание обращено на неголономные связи, так линейные, как и нелинейные. Обсуждены некоторые качественные разницы между моделями неголономной динамики, опирающимися на разные процедуры. Особенное значение имеют два алгоритма, именно: 1) принцип Даламбера и его обобщение, приведенное Аппеллом и Четаевым (принципы, используемые в типичных применениях, как задачи качения), а также 2) вариационный принцип Гамильтона с дополнительными условиями (более интересный с математической точки зрения). Утверждается, что принцип Гамильтона, хотя неприменим в традиционных технических задачах, однако более обещающий, по крайней мере как эвристическое указание, в обобщениях, касающихся систем с высшими дифференциальными связями или еще более общими связями функционального типа, появляющимися в задачах авторегулировки и управления.

1. Introduction

It is really incredible that towards the close of the 20-th century there exists a branch of mechanics with numerous practical applications but without physically justified and unquestionable, commonly accepted foundations. This is the theory of systems with nonholonomic constraints. Its traditional domain of applications includes all problems

of rolling of rigid and deformable bodies, and electromechanical systems with sliding contacts. Nonholonomic constraints appear in rolling problems due to the direct contact mechanism based on friction forces. There is also another, rather new, domain of nonholonomic dynamics. Namely, in navigation and automatic control systems based on feedback phenomena one deals with servo-constraints i.e. holonomic or nonholonomic constraints of non-contact origin. They are realized by systems consisting of sensors, amplifiers, relays (electromagnetic, pneumatic, hydrodynamic) and actuators. Such systems can be also endowed with automatic data processing. In many problems it is possible to neglect the dynamics of the control system itself and represent it by its "final product", i.e. by constraints or a program imposed upon the motion of the controlled mechanical object. There are practically no limitations on the shape of possible constraints. For example, one can produce higher-order nonholonomic constraints (e.g. constraints on accelerations), or even quite general functional constraints if the control system is endowed with diferentiation and integration procedures.

It seems quite natural to use analytical dynamics of nonholonomic systems, a relatively classical topic, as a pattern for the analysis of more complicated constraints appearing, e.g. in automatic control problems. However, when we try to do this, it turns out that traditional procedures do not suit such generalizations; moreover, there are some doubts as to their own correctness and justification.

2. Algorithms for holonomic constraints

To understand the motivation for various algorithms for nonholonomic constraints, we must recapitulate the procedures used in holonomic dynamics.

Let us consider a system with n degrees of freedom, i.e. with the n-dimensional configuration space Q. At this stage Q is a general differential manifold without any additional structure. As usual in analytical mechanics, we use two kinds of state manifolds, namely the Newton state space P_N, i.e. the space of generalized positions and velocities and the Hamilton state space P, i.e. the space of positions and conjugate canonical momenta, called also the phase space. From the mathematical point of view P_N and P are $2n$-dimensional differential manifolds, respectively, the tangent and cotangent bundles TQ, T^*Q over the position space Q.

We shall follow traditional notations; thus the generalized coordinates, i.e. local coordinates on Q, will be denoted by q^i, $i = 1, \ldots, n$, and the corresponding coordinates on P_N and P will be denoted respectively by (q^i, v^i) and (q^i, p_i), $i = 1, \ldots, n$. If calculated along a fixed motion $t \mapsto q(t)$, the velocity $\frac{dq^i}{dt}(t)$ will be denoted shortly by $\dot{q}^i(t)$.

Let us assume that the dynamical structure of our system is described by a pair (L, D), where L is a Lagrangian describing the non-dissipative part of dynamics, and D is a canonical generalized force of dissipative interactions. From the mathematical point of view, L is a scalar function on $P_N = TQ$ and D is a covariant vector on Q, depending on generalized positions and velocities. More rigorously, with any generalized position $q \in Q$ and velocity $v \in T_qQ \subset P_N$, there is assiociated a covector $D(q, v) \in T_q^*Q$ attached at q.

It is important to stress that within an abstract framework based on a general Lagrangian L, the canonical force D cannot be identified with a usual, i.e. contravariant, vector in Q. This can be done only after introducing an additional object in Q, namely, the metric tensor. If the system is free of dissipation, D = 0, then the equations of motion are derivable from the Hamilton principle of stationary action, $\delta \int L dt = 0$, and have the familiar Euler–Lagrange structure,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0, \quad i = 1, \ldots, n. \tag{2.1}$$

If dissipative forces are taken into account, the equations of motion are not derivable from a variational principle and have the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i, \quad i = 1, \ldots, n. \tag{2.2}$$

Let us quote a typical example

$$\begin{aligned} L(q, v) &= \frac{1}{2} g_{ij}(q) v^i v^j + A_i(q) v^i + V(q), \\ D_i(q, v) &= -d_{ij}(q, v) v^j, \end{aligned} \tag{2.3}$$

where g is a positively-definite metric tensor on Q underlying the kinetic energy $T = \frac{1}{2} g_{ij} v^i v^j$, V is a scalar potential (e.g. electrostatic, or oscillatory) and A_i is a vector potential responsible for non-dissipative velocity-dependent forces orthogonal to generalized velocities (magnetic forces, gyroscopic forces). The friction tensor d is positively semi-definite in the g-sense, i.e. the mixed tensor $d^i_j := g^{ik} d_{kj}$ has non-negative eigenvalues. Besides, one assumes that $d(q, 0)$ is finite (no viscous friction at rest). The resulting equations of motion have the form

$$\frac{D^2 q^i}{Dt^2} = g^{ik} F_{kj} \frac{dq^j}{dt} - g^{ik} \frac{\partial V}{\partial q^k} - g^{ik} d_{kj} \frac{dq^j}{dt}, \tag{2.4}$$

where

$$\frac{D^2 q^i}{Dt^2} := \frac{d^2 q^i}{dt^2} + \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} \frac{dq^j}{dt} \frac{dq^k}{dt}$$

is the g-covariant acceleration vector, $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$ being the Christoffel coefficients, and

$$F_{kj} := \frac{\partial A_j}{\partial q^k} - \frac{\partial A_k}{\partial q^j},$$

i.e. the exterior differential of A, is the "magnetic-like" field responsible for non-dissipative velocity-dependent forces. Let us observe that now, after introducing the quantity g, we have identified all forces, both variational and dissipative, with contravariant vectors.

Mechanical energy is given by the familiar formula, $E := v^i \frac{\partial L}{\partial v^i} - L$, for example, in the above example we have $E = T + V$. The energy balance has the form

$$\frac{dE}{dt} = -\frac{\partial L}{\partial t} + D_i \frac{dq^i}{dt}, \tag{2.5}$$

i.e. in the case of time-independent systems the time rate of energy equals the power of non-variational forces D,

$$\frac{dE}{dt} = D_i \frac{dq^i}{dt}. \tag{2.6}$$

Let us now subject the system to holonomic constraints $W \subset Q$, W being an $(n-m)$-dimensional submanifold of Q described by the equations

$$F_a(q^1 \dots q^n) = 0, \quad a = 1, \dots, m; \tag{2.7}$$

the functions F_a are functionally independent in a certain neighbourhood of W. The constraints W restrict the state manifold $P_N = TQ$ to the submanifold $P_N(W) = TW$; similarly, $P = T^*Q$ is replaced by $P(W) = T^*W$. Thus the set of admissible velocities at a configuration $q \in Q$ is restricted to the $(n-m)$-dimensionial linear subspace $T_qW \subset T_qQ$. Obviously velocities compatible with constraints satisfy the following linear equations;

$$\frac{\partial F_a}{\partial q^i} v^i = 0, \quad a = 1, \dots, m. \tag{2.8}$$

Except for some very special relationships between (L, D) and W, the general solution of Eq. (2.2) is incompatible with W; as a rule, trajectories with initial values on TW leave this manifold. Thus to confine the system in W, we must subject it to an additional dynamical influence. This means that at each point $q \in W$ we switch on an additional, in general velocity-dependent, canonical force $R(q, v) \in T_qQ$ with the task to keep constraints without disturbing the along-constraints motion. It is important to stress that R is defined only for constrained mechanical states, thus for $q \in W$ and $v \in T_qW$, nevertheless, *a priori* $R(q, v)$ is a general element of T_qQ, not of T_qW or of a fixed subspace of T_qQ.

The quantities R are traditionally called *reaction forces* or simply *reactions*. In some sense, the search of proper reactions belongs to the theory of program motion, control theory and optimization.

The total system of equations describing constrained motion has the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + R_i, \quad F_a(q) = 0, \quad i = 1, \dots, n, \quad a = 1, \dots, m. \tag{2.9}$$

This system of $n+m$ equations is to be solved with respect to $2n$ quantities q^i, R_i, $i = 1, \dots, n$. The solution is evidently non-unique. To specialize it we must impose certain restrictions on the class of admissible reactions. First of all, the set of admissible trajectories must be derivable from a second-order differential equation on W, as its general solution (dependent on $2(n-m)$ constants). But the fundamental restriction is that the only dynamical effect of R is to maintain constraints without influencing along-constraints modes of motion. Obviously, when formulated in such a way, this demand is rather qualitative. There are a few ways to make it mathematically precise. They are very close to each other, almost equivalent, for holonomic systems, but diverge when constraints are nonholonomic. Overlooking this fact accounts for accumulated historical misunderstandings in nonholonomic mechanics.

2.1. Dual reactions

This is a rather geometric and intuitive way of making precise the demand "without disturbing along-constraints motion". For any $q \in W$, $v \in T_q W$, $R(q, v)$ is an element of the subspace An $T_q W \subset T_q^* Q$, i.e.

$$\langle R(q, v), w\rangle = R(q, v)_i w^i = 0 \tag{2.10}$$

for any $w \in T_q W$. But vectors w tangent to W satisfy

$$\frac{\partial F_a}{\partial q^i} w^i = 0, \quad a = 1, \ldots, m, \tag{2.11}$$

thus

$$R(q, v)_i = \lambda^a(q, v) \frac{\partial F_a}{\partial q^i}, \quad i = 1, \ldots, n, \tag{2.12}$$

λ^a being certain, yet non-definite functions of mechanical state. If Q is endowed with a metric tensor g, for example if we deal with models (2.3), then we can represent any covector f_i by the contravariant vector $\bar{f}^i := g^{ij} f_j$ (if there is no danger of misunderstanding, we shall simply write f^i instead of $\bar{f}^i$). If canonical reactions R_i are dual to W, then their contravariant representatives R^i are g-orthogonal to W:

$$g_{ij} R^i w^j = 0,$$

for any $w \in T_q W$. Such constraints, or rather such forces maintaining them, are traditionally called *ideal*. It is easy to see that ideal reactions are unique. Indeed, substituting Eq. (2.12) to the system (2.9), we obtain equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + \lambda^a \frac{\partial F_a}{\partial q^i}, \quad F_a(q) = 0. \tag{2.13}$$

This system consists of $(n+m)$ equations imposed upon $(n+m)$ variables (q^i, λ^a). Solving them we find a $2(n-m)$-parameter family of motions $q^i(t)$, $i = 1, \ldots, n$ in W, and quantities $\lambda^a(q, v)$, $a = 1, \ldots, m$ for any $q \in W$, $v \in T_q W$.

For historical reasons, let us quote the traditional way of writing Eqs. (2.13) in terms of virtual displacements,

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} - D_i\right) \delta q^i = 0,$$
$$F_a(q) = 0, \quad \frac{\partial F_a}{\partial q^i} \delta q^i = 0 \tag{2.14}$$

(ideal reactions do not work on virtual displacements compatible with constraints).

Let $(x^1 \ldots x^{n-m})$ be coordinates on W; the corresponding parametric equations of W have the form

$$q^i = f^i(x^1 \ldots x^{n-m}), \quad i = 1, \ldots, n.$$

Then the system (2.13) is equivalent to the following system of $(n-m)$ Lagrange equations on W, explicitly free of reactions:

$$\frac{d}{dt} \frac{\partial \bar{L}}{\partial \dot{x}^\mu} - \frac{\partial \bar{L}}{\partial x^\mu} = \bar{D}^\mu, \quad \mu = 1 \ldots (n-m),$$

where $\overline{L}$ and $\overline{D}$ are restrictions of L and D to constraints W,

$$\begin{aligned} \overline{L}(x, \dot{x}) &:= L\left(f^j(x), \frac{\partial f^j}{\partial x^\nu}\dot{x}^\nu\right), \\ D_\mu(x, \dot{x}) &:= D_i\left(f^j(x), \frac{\partial f^j}{\partial x^\nu}\dot{x}^\nu\right)\frac{\partial f^i}{\partial x^\mu}. \end{aligned} \tag{2.16}$$

2.2. Dead reactions, adiabatic reactions

Such reactions do not do any work along curves admitted by constraints (along all of them, not necessarily dynamically admissible). This means that

$$\langle R(q, v), v\rangle = R(q, v)_i v^i = 0 \tag{2.17}$$

for any $v \in T_q W$. Thus we maintain the along-W behaviour of our system without performing mechanical work. Reactions do not influence the energy balance,

$$\frac{dE}{dt} = D_i \frac{dq^i}{dt};$$

energy being understood in the unconstrained sense, $E = v^i \dfrac{\partial L}{\partial v^i} - L$. In particular, energy is conserved if there is no dissipation, $D = 0$. Obviously the condition 2 is weaker than 1, dual reactions do not work along constraints, but not every energy-preserving reaction is dual. Indeed, it is easy to see that Eq. (2.17) is satisfied by all reactions of the form

$$R_i(q, v) = H_{ij}(q, v)v^j + \lambda^a(q, v)\frac{\partial F_a}{\partial q^i}, \tag{2.18}$$

H being a skew-symmetric twice covariant tensor. Such reactions are non-unique. The "magnetic-like" ("gyroscopic-like") term $H_{ij}v^i$ is arbitrary if we do not accept additional criteria. In any case, there is no intrinsic geometric prescription which could assign a particular form of H to given functions F_a.

The force $H_{ij}v^i$ does not contribute to energy balance, nevertheless it influences in some sense the along-constraints motion, because in general it has a non-vanishing projection onto W.

2.3. Constrained Hamilton principle, Lusternik theorem

If there are no dissipation forces, $D = 0$, then we can formulate the restricted variational problem in W,

$$\delta \int \overline{L} dt = 0,$$

L being the restriction of L to TW, $\overline{L} := L|TW$.

The resulting Euler–Lagrange equations coincide with the system (2.15) if $D = 0$. The quantities λ^a (reaction coefficients) then appear as Lagrange multipliers in the Lusternik theorem. Indeed, this theorem tells us that there exist functions of time λ^a, $a = 1, \ldots, m$ such that conditional stationary points of the functional $\int L dt$ (with conditions

$F_a(q(t)) \equiv 0$, $a = 1, \ldots, m$) coincide with the usual (without conditions) stationary points of the functional $\int L[\lambda]\,dt$ with the modified, time-dependent Lagrangian:

$$L[\lambda](q, v, t) := L(q, v) + \lambda^a_{\natural}(t) F_a(q).$$

Calculating Euler–Lagrange equations for $L[\lambda]$ and substituting constraints equations, we obtain the system (2.13) with $D = 0$. Thus, for holonomic non-dissipative systems, the procedures 1 and 3 give identical reactions and identical effective equations of motion.

3. Algorithms for nonholonomic constraints

Let us assume that besides the a priori given interactions described by (L, D), the system is subject to influences which restrict its Newtonian states to certain submanifold $M \subset P_N = TQ$. Such constraints are called *anholonomic*. They are called *nonholonomic* if they are not reducible to holonomic constraints, i.e. if M is not foliated by submanifolds of the form TW, $W \subset Q$. Thus, roughly speaking, nonholonimoc constraints restrict velocities without restricting positions (or, in any case, restrict velocities stronger than positions if they contain an admixture of holonomic constraints).

Let M_q denote the set of admissible velocities attributed to a given position $q \in Q$, thus $M_q := M \cap T_qQ$, and $M = \bigcap_{q \in Q} M_q$. We shall assume that M does not contain any holonomic admixture, i.e. projecting M onto Q we obtain the whole configuration space Q. Thus, for any $q \in Q$ we have: $\dim M = (n-m)$, where m equals the co-dimension of M, $\dim M = (2n-m)$. Analytically the constraints M are described by the equations

$$F_a(q^1 \ldots q^n, v^1 \ldots v^n) = 0, \quad a = 1, \ldots, m. \tag{3.1}$$

For general constraints M the submanifolds M_q are curved, i.e. they are not linear subspaces of T_qQ. Their equations (at least some of them) are nonlinear with respect to velocity arguments. However, in most popular practical applications of analytical mechanics we deal just with nonholonomic constraints M intersecting all fibres T_qQ along linear subspaces. We shall call them *linear nonholonomic constraints*; certain authors use also the term *kinematical constraints*. Such constraints can be pictured as a field of m-dimensional linear subspaces, $Q \ni q \mapsto M_q \subset T_qQ$, i.e. in differential-geometric terms, as an m-dimensional distribution. One can describe them with the help of equations linear in velocities,

$$F_a(q, v) = \omega_{ai}(q) v^i = 0, \quad a = 1, \ldots, m. \tag{3.2}$$

In other words, the set of kinematically admissible trajectories in Q is *a priori* restricted to integral curves of the Pfaff systems

$$\omega_a = 0, \quad a = 1, \ldots, m, \tag{3.3}$$

ω_a being differential forms (covector fields) on Q dual to the distribution $q \mapsto M_q$,

$$\omega_a := \omega_{ai} dq^i. \tag{3.4}$$

If the Pfaff system (3.3) is maximally integrable, then we deal with anholonomic constraints equivalent to a family of holonomic constraints foliating Q. This is the case if and only if

$$d\omega_a \wedge \omega_1 \wedge \ldots \wedge \omega_m = 0. \tag{3.5}$$

Such constraints are sometimes called *semi-holonomic*.

Linear nonholonomic constraints exhaust the most popular and natural class of nonholonomic mechanical problems, namely, the nonsliding rolling of rigid or elastic bodies. Their dynamical properties are still very close to those of holonomic constraints. Nevertheless, in contrast to popular views, there are certain ambiguities and doubts concerning the concept of ideal reactions.

Let us now try to follow three algorithms for reactions, discussed in previous section.

3.1. Dual reactions

There is no submanifold of admissible configurations in Q, but nevertheless, at each configuration $q \in Q$ we have the well-defined linear subspace of admissible velocities $M_q \subset T_qQ$, and, consequently, the space of dual reactions $\mathrm{An}\, M_q \subset T_q^*Q$. R is dual to constraints if

$$\langle R(q, v), w\rangle = R(q, v)_i w^i = 0 \tag{3.6}$$

for any $q \in Q$, $v \in M_q$, $w \in M_q$. Elements of M_q satisfy the equations

$$\langle \omega_{aq}, w\rangle = \omega_{ai}(q) w^i = 0, \tag{3.7}$$

thus,

$$R(q, v)_i = \lambda^a(q, v)\omega_{ai}(q), \tag{3.8}$$

where λ^a are yet unknown functions of admissible states. The resulting equations of motion have the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + \lambda^a \omega_{ai}, \qquad \omega_{ai}\frac{dq^i}{dt} = 0. \tag{3.9}$$

Just as in the holonomic theory, this system of $(n+m)$ equations can be solved with respect to $(n+m)$ quantities (q^i, λ^a).

This is just the commonly used algorithm for linear nonholonomic constraints. By analogy to holonomic systems, the reactions (3.8) are usually called *ideal*. Equations (3.9) reduce to Eq. (2.13) if Eq. (3.2) is an anholonomic description of essentially holonomic constraints, i.e. if

$$\omega_a = dF_a, \qquad \omega_{ai} = \frac{\partial}{\partial q^i} F_a.$$

3.2. Dead reactions, i.e. adiabatic reactions

They do not work along curves compatible with constraints, i.e. along integral curves of ω_a, $a = 1, \ldots, m$, thus,

$$\langle R(q, v), v\rangle = R(q, v)_i v^i = 0 \tag{3.10}$$

for any $q \in Q$, $v \in M_q$. Reactions do not contribute to energy balance, $\frac{dE}{dt} = D_i \dot{q}^i$. By analogy to Eq. (2.18), we have

$$R(q, v)_i = H_{ij}(q, v)v^j + \lambda^a(q, v)\omega_{ai}, \tag{3.11}$$

where $v \in M_q$ and H is skew-symmetric.

Let us observe, however, that nonholonomic systems have certain new features which give rise to intrinsic prescriptions for H. Indeed, we have at disposal m skew-symemtric twice covariant tensors

$$\gamma_a = d\omega_a, \quad \text{i.e.,} \quad \gamma_{aij} = \frac{\partial \omega_{aj}}{\partial q^j} - \frac{\partial \omega_{ai}}{\partial q^j}, \quad a = 1, \ldots, m.$$

Therefore we can postulate

$$H_{ij}(q, v) = \mu^a(q, v)\gamma_{aij}, \tag{3.12}$$

$$R_i(q, v) = \mu^a(q, v)\left(\frac{\partial \omega_{aj}}{\partial q^i} - \frac{\partial \omega_{ai}}{\partial q^j}\right)v^j + \lambda^a(q, v)\omega_{ai}, \tag{3.13}$$

μ^a being yet non-definite functions of admissible states. At each state (q, v), the reaction (3.13) is controlled by $2n$ multipliers μ^a, λ^a $a = 1, \ldots, n$. Therefore the system (2.9) leaves n of them arbitrary.

In general, $R(q, v)$ given by Eq. (3.13) is not dual to M_q, i.e. there are vectors $u \in M_q$ for which $R(q, v)_i u^i \neq 0$. This means, roughly speaking, that the reactions (3.13) modify the "along M_q" — instantaneous behaviour of the system. Nevertheless, there are no reasons to object against them because in the nonholonomic case there is no proper submanifold of constraints $N \subset Q$ and the *instantaneous* "along M_q" effect of $R(q, v)$ cannot be interpreted as a perturbation of the *finite* along-constraints motion. On the contrary, for nonholonomic constraints (3.2) the term $\mu^a(\omega_{ai,j} - \omega_{aj,i})\dot{q}^j$ is exactly as natural as $\lambda^a\omega_{ai}$. The 19-th century philosophy of virtual displacements has brought about some, in our opinion misleading, discussion as to these terms and their mutual relationship. Namely, in analogy to holonomic systems, virtual displacements were defined by the formula

$$\omega_{ai}\delta q^i = 0. \tag{3.14}$$

Time-dependent δq^i were to describe small, compatible with constraints, variations of a given motion $t \mapsto q(t)$. But in the nonholonomic case the modified motion $q^i(t) + \delta q^i(t)$ fails to be compatible with constraints, [1,6], because in a first-order approximation

$$\delta F_a = F_a(q^i + \delta q^i) - F_a(q^i) = \frac{d}{dt}(\omega_{ai}\,\delta q^i) + (\omega_{aj,i} - \omega_{ai,j})\frac{dq^j}{dt}\,\delta q^i. \tag{3.15}$$

The condition (3.14) is not sufficient for vanishing of δF_a. Thus, one argued, virtual displacements connecting nearby curves compatible with constraints had to satisfy *two* algebraic conditions, namely the formula (3.14) and

$$(\omega_{aj,i} - \omega_{ai,j})\delta q^i = 0. \tag{3.16}$$

This view resulted from the unquestionable belief in the magic power of virtual displacements satisfying appropriately chosen *algebraic* conditions. It is obvious, however, that geometrically proper variations δq^i expressing along-constraints deformations of curves satisfy n *differential* conditions

$$\frac{d}{dt}(\omega_{ai}\delta q^i) + (\omega_{aj,i} - \omega_{ai,j})\,\dot{q}^j\delta q^i = 0 \tag{3.17}$$

instead of $2n$ *algebraic* conditions (3.14), (3.16).

The functions $\delta q^i(t)$ compatible with Eq. (3.17) are proper "tangent vectors" at the curve $q(t)$ in the functional manifold of all curves subject to the conditions $\omega_{ai}\dot{q}^i = 0$.

3.3. Hamilton principle, Lusternik theorem

It turns out that this algorithm gives the reactions (3.13) together with a relationship between λ and μ leaving only n of them arbitrary. Indeed, if there is no dissipation, the restricted Hamilton principle

$$\delta \int L(q(t), \dot{q}(t))dt = 0, \quad F_a(q(t), \dot{q}(t)) \equiv 0, \quad a = 1, \ldots, m \tag{3.18}$$

gives the equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = R_i, \quad \omega_{ai}\frac{dq^i}{dt} = 0, \tag{3.19}$$

where

$$R_i = \mu^a\left(\frac{\partial \omega_{aj}}{\partial q^i} - \frac{\partial \omega_{ai}}{\partial q^j}\right)\frac{dq^j}{dt} - \frac{d\mu^a}{dt}\omega_{ai}, \tag{3.20}$$

μ^a being Lagrange multipliers. This is just the special case of Eq. (3.13) with $\lambda^a = -\frac{d}{dt}\mu^a$.

For dissipative systems the obvious generalization of Eqs. (3.19) is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + R_i, \quad \omega_{ai}\frac{dq^i}{dt} = 0 \tag{3.21}$$

with the same reaction (3.20).

The algorithms 1 and 3 give evidently *different* results. For holonomic systems they are equivalent and agree with experimental data concerning the dynamics of naturally occurring constrained systems like rigid bodies and their aggregates. At the same time, the algorithm 3 is a convincing procedure excluding, for holonomic systems, the procedure 2, which, although mathematically possible, is charged with a nonphysical arbitrariness and does not agree with experiments on rigid bodies. Thus we are inclined to believe that the algorithm 3 is also a proper way of describing naturally occurring and technically constructed nonholonomic systems. It is rather surprising that this procedure implies "magnetic-like" ("gyroscopic-like") reaction terms $\mu^a(\omega_{aj,i} - \omega_{ai,j})v^j$ for nonholonomic systems, although it has just excluded them in the holonomic case. The quantity $\mu^a(\omega_{aj,i} - \omega_{ai,j})v^j$ resembles certain expressions occurring in Woronetz equations for linearly nonholonomic systems. However, this similarity is superficial because Woronetz equations are based exactly on dual reactions (3.8), not on reactions (3.13). Equations (3.19) are certainly non-equivalent to Woronetz equations.

Let us observe that in Eqs. (3.19) with the reactions (3.20), the Lagrange multipliers μ^a occur in a non-algebraic way, together with first derivatives. Therefore it does not seem possible to follow the holonomic procedure of solving, i.e., to eliminate μ^a, to solve the remaining equations with respect to q^i, and then to express μ^a as functions of mechanical state. Rather, the parameters μ^a acquire the character of additional, controlling, degrees of freedom.

Let us notice that the Hamilton-Lusternik variational algorithm is based on virtual displacements satisfying the *differential* equations (3.17). What would result if we used *two* independent *algebraic* conditions (3.14), (3.16)? It is easy to see that the standard formula

$$\delta \int L dt = \int \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \delta q^i dt = 0$$

with δq^i subject to Eqs. (3.14), (3.16), gives us exactly Eqs. (3.13) with $2n$ non-determined parameters.

Analysis of nonlinear nonholonomic constraints seems to support our belief in the procedure 3.

Let us now consider general nonholonomic constraints $M \subset TQ$. It is obvious that if the intersections $M_q := M \cap T_q Q$ are *nonlinear*, then the procedure 1 in its literal sense is completely meaningless. Indeed, if M_q is nonlinear, then its linear shell has higher dimension than M_q itself, and in the general case it coincides with the total space $T_q Q$. Therefore, the only covector $R(q, v)$ dual to all elements of M_q is the null element of $T_q Q$. But with vanishing reactions the equations of motion are incompatible with constraints. Many attempts have been made to overcome this difficulty. One of them, from the formal point of view relatively natural, consists in defining the reactions $R(q, v)$ as dual to the tangent space $T_v M_q$, i.e.

$$\langle R(q, v), w \rangle = R(q, v)_i w^i = 0 \tag{3.22}$$

for any $q \in Q$, $v \in M_q$ and $w \in T_v M_q \subset T_q Q$. But elements of $T_v M_q$ satisfy the equations

$$\frac{\partial F_a}{\partial v^i} w^i = 0, \quad a = 1, \ldots, m, \tag{3.23}$$

thus, finally,

$$R(q, v)_i = \lambda^a \frac{\partial F_a}{\partial v^i}, \tag{3.24}$$

λ^a being non-determined coefficients. If (q, v) are kept fixed, the set of all admissible $R(q, v)$'s is m-dimensional. Thus, formally, the system of $(n+m)$ equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + \lambda^a \frac{\partial F_a}{\partial \dot{q}^i}, \quad F_a(q, \dot{q}) = 0, \quad a = 1, \ldots, m, \quad i = 1, \ldots, n, \tag{3.25}$$

can be solved with respect to (q^i, λ^a), uniquely up to initial values of (q, v). Using the virtual displacements philosophy we can write Eqs. (3.25) as

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} - D_i \right) \delta q^i = 0, \quad F_a(q, \dot{q}) = 0, \quad \frac{\partial F_a}{\partial \dot{q}^i} \delta q^i = 0; \tag{3.26}$$

δq^i are to be eliminated.

Constrained systems maintained with forces R of the form (3.24) are called *Appell–Tshetajev systems*. If constraints are linear, the Appell–Tshetajev procedure reduces to the algorithm 1 of dual reactions because then $T_v M_q = M_q$ for any $v \in M_q$. There is some philosophy used in literature for justifying the Appell–Tshetajev procedure. It is based on concepts of acceleration work, acceleration power and acceleration energy. Acceleration

work of a force F along a curve $t \mapsto q(t)$ is defined as $\int F_i \ddot{q}^i \, dt$; the integrand $F_i \ddot{q}^i$ represents the acceleration power. Reactions (3.24) at any point $q \in Q$ have vanishing acceleration power along any instantaneous motion in Q starting at q with vanishing velocity and compatible with constraints. However, this justification is rather misleading because acceleration work is a rather nonphysical concept, which has nothing to do with balance of energy. Moreover, it is not correctly defined unless Q is an affine space, because in a general manifold, second derivatives of $q^i(t)$ do not represent vector or any geometric object at all. Because of this, the expression $F_i \ddot{q}^i$ is dependent on coordinates; thus nongeometric and nonphysical.

The idea of the reactions (3.24) goes back to Gauss who was the first to suggest reactions dual to virtual accelerations. The original Gauss principle was formulated for constrained systems of material points in Euclidean space. All tangent spaces $T_q Q$ are then canonically isomorphic to the standard linear space of translations V, thus $TQ = Q \times V$, $T^*Q = Q \times V^*$. The vectors $(u, w) \in V \times V$ tangent at $(q, v) \in Q \times V$ to constraints M satisfy

$$\frac{\partial F_a}{\partial q^i}(q, v) u^i + \frac{\partial F_a}{\partial v^i}(q, v) w^i = 0. \tag{3.27}$$

Any curve $t \mapsto q^i(t)$, $i = 1 \ldots n$ in Q induces the curve $t \mapsto (q^i(t), \dot{q}^i(t))$ in $TQ = Q \times V$. Vectors tangent to the latter curve have components $(\dot{q}^i(t), \ddot{q}^i(t))$; thus for them $u^i = \dot{q}^i(t)$, $w^i = \ddot{q}^i(t)$. Equation (3.23) can be formally obtained from Eq. (3.27) by restriction to vectors of the form $(0, w^i)$. In this sense w^i are virtual accelerations compatible with constraints [5, 6]. It must be stressed, however, that this interpretation is not to be taken too seriously because vectors $(0, \ddot{q}^i)$ are not tangent to curves in $M \subset Q \times V$ induced from Q.

Quite independently of this, rather incorrect, acceleration philosophy, the Appell–Tshetajev algorithm is formally well-defined for any mechanical system, independently of geometry of Q. However, one can doubt whether it is a natural or technically realizable and optimal program for maintaining realistic constraints M. In any case, the concept of virtual displacement defined by the algebraic conditions (3.26) is a rather artificial and purely formal extrapolation of the corresponding holonomic notion (2.14). It has no convincing mathematical content; for example, it does not represent the tangent vector in the infinite-dimensional functional space of trajectories kinematically compatible with constraints. It has no physical interpretation, either; for example, it has nothing to do with the true mechanical work of reactions and with the energy balance. The vanishing of $R_i \, \delta q^i$, fundamental for Eq. (3.24), has no physical motivation.

The principle of dead reactions (adiabatic reactions) is non-effective because it admits too much arbitrariness for reactions, even in the holonomic case. Besides, if the functions F_a are nonlinear in velocities (even if they are affine, i.e. linear nonhomogeneous functions), it may happen quite easily that equations based on this principle will be inconsistent. In fact, it seems evidently impossible to maintain all kinds of nonholonomic constraints without energetic influence (try to confine the absolute value of velocity or angular velocity without influx or efflux of energy).

Thus it seems that the most promising program for forces maintaining nonholonomic constraints is that based on the Hamilton–Lusternik principle. Let us discuss it briefly.

If there is no dissipation, then, introducing m Lagrange multipliers $\mu^a(t)$, $a = 1, \ldots, m$, and calculating Euler–Lagrange equations for the modified Lagrangian $L[\mu] := L + \mu^a F_a$, we obtain the following system of $(n+m)$ differential equations for $(n+m)$ variables $q^i(t)$, $\mu^a(t)$:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = \mu^a \frac{\partial F_a}{\partial q^i} - \frac{d}{dt}\left(\mu^a \frac{\partial F_a}{\partial \dot{q}^i}\right), \quad i = 1, \ldots, n \tag{3.28}$$

$$F_a(q, \dot{q}) = 0, \quad a = 1, \ldots, m.$$

Reactions contain multipliers μ^a together with their first derivatives;

$$R_i = \mu^a \frac{\partial F_a}{\partial q^i} - \frac{d}{dt}\left(\mu^a \frac{\partial F_a}{\partial \dot{q}^i}\right) \tag{3.29}$$

$$= \mu^a \frac{\partial F_a}{\partial q^i} - \frac{d\mu^a}{dt}\frac{\partial F_a}{\partial \dot{q}^i} - \mu^a \frac{\partial^2 F_a}{\partial \dot{q}^i \partial q^j}\frac{dq^j}{dt} - \mu^a \frac{\partial^2 F_a}{\partial \dot{q}^i \partial \dot{q}^j}\frac{d^2 q^j}{dt^2}.$$

It is rather natural to postulate the same mechanism of reactions in dissipative problems thus the equations of constrained motion have the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = D_i + R_i, \quad F_a(q, \dot{q}) = 0, \tag{3.30}$$

R_i again given by the formula (3.29).

Nonlinearity of nonholonomic constraints with respect to velocities has a remarkable qualitative effect on the dynamical structure of reactions R_i. First of all, R_i contains the term with second derivatives, i.e.

$$-\mu^a \frac{\partial^2 F_a}{\partial \dot{q}^i \partial \dot{q}^j}\frac{d^2 q^j}{dt^2}.$$

Such acceleration-dependent forces modify the inertial properties of the object. Besides, nonlinearity of M may influence the energy balance because, in general, the reactions (3.29) need not annihilate the velocity vectors. Indeed, let us calculate the power of the reactions (3.29) along curves in Q compatible with constraints M. After elementary calculations we obtain

$$R_i \frac{dq^i}{dt} = \mu^a \frac{dF_a}{dt} - \frac{d}{dt}\left(\mu^a \dot{q}^i \frac{\partial F_a}{\partial \dot{q}^i}\right).$$

The first term vanishes in virtue of constraints equations. Thus, finally

$$R_i \dot{q}^i = -\frac{d}{dt}\left(\mu^a \frac{\partial F_a}{\partial \dot{q}^i}\dot{q}^i\right). \tag{3.31}$$

Hence, in general, reactions (3.29) need not be adiabatic. They are adiabatic if constraints are linear. Indeed, linear functions satisfy the equation $v^i \frac{\partial}{\partial v^i} F_a = F_a$, thus Eq. (3.31) takes on the form $-\frac{d}{dt}(\mu^a F_a)$ that evidently vanishes along any curve kinematically compatible with constraints. More generally, Eq. (3.31) vanishes if $v^i \frac{\partial}{\partial v^i} F_a | M = 0$,

i.e. if $v^i \frac{\partial}{\partial v^i} F_a = k_a{}^b F_b$ for certain functions $k_a{}^b$ nonsingular in a neighbourhood of M. But this means that the vector-field $v^i \frac{\partial}{\partial v^i}$ is tangent to constraints M. Let us now notice that $v^i \frac{\partial}{\partial v^i}$ is identical with the generator of the one-parameter dilatation group $v \mapsto e^x v$, $x \in \mathbf{R}$; thus any integral curve of $v^i \frac{\partial}{\partial v^i}$ is either a half-space of a one-dimensional subspace of a certain T_qQ, or just the null element of T_qQ (singular integral curve)([1]). Therefore any manifold M with the above property $\left[v^i \frac{\partial}{\partial v^i} F_a = k_a{}^b F_b\right]$ is plaited of one-dimensional subspaces of fibres T_qQ. Any one-dimensional (half) subspace is either disjoint with M or entirely contained in M. This means that $M_q = M \cap T_qQ$ imposes constraints on directions in T_qQ, i.e. on directions of velocities at $q \in Q$, but not on their magnitudes. Thus we can formulate the following:

THEOREM. *If nonholonomic constraints restrict directions of velocities but leave their magnitudes arbitrary, then their Hamilton–Lusternik reactions are adiabatic, i.e. they do not do any work along curves compatible with constraints.*

This class of constraints is evidently wider than linear nonholonomic constraints.

Equation (3.31) means that the energy balance has the form

$$\frac{d}{dt}\left(E + \mu^a \frac{\partial F_a}{\partial \dot{q}^i} \dot{q}^i\right) = D_i \dot{q}^i. \tag{3.32}$$

The balanced quantity

$$E[L, M] := E + \mu^a \frac{\partial F_a}{\partial \dot{q}^i} \cdot^i \tag{3.33}$$

can be interpreted as the effective energy of the system constrained by the manifold M When M is fixed, $E[L, M]$ does not depend on the particular choice of functions F_a used as left-hand sides of equations of M. Indeed, any change of F_a, $(\ldots F_a \ldots) \mapsto (\ldots \overline{F}_a := := F_b L^b{}_a \ldots)$ is compensated by the corresponding dual change of coefficients μ^a.

The quantity $E[L, M]$ contains two parts: the natural energy $E = v^i \frac{\partial L}{\partial v^i} - L$ of the unconstrained system and the energy of constraints

$$E[M] := \mu^a \frac{\partial F_a}{\partial \dot{q}^i} \dot{q}^i. \tag{3.34}$$

([1]) Indeed, integral curves of $v^i \frac{\partial}{\partial v^i}$ satisfy differential equations

$$\frac{dq^i}{dt} = 0, \qquad \frac{dv^i}{dt} = v^i.$$

Integrating them we obtain

$$q^i = a^i, \qquad v^i = b^i e^t,$$

a^i, b^i being integration constants. But these are just parametric equations of a one-dimensional half-subspace of T_aQ with the directional vector $b = (\ldots b^i \ldots)$. If $b = 0$, this subspace degenerates to the null element of T_aQ.

If there are no dissipative forces, the total energy $E[L, M]$ is a constant of motion. The existence of this constant of motion is just the peculiarity and distinguishing feature of the Hamilton–Lusternik algorithm. Let us observe that $E[L, M]$ can be directly obtained from the modified Lusternik Lagrangian $L[\mu]$. Indeed,

$$E[L[\mu]] = \dot{q}^i \frac{\partial L[\mu]}{\partial \dot{q}^i} - L[\mu] = E[L] + \mu^a \frac{\partial F_a}{\partial \dot{q}^i} \dot{q}^i - \mu^a F_a, \tag{3.35}$$

but the last term vanishes on constraints M.

The mechanical work done by Hamilton–Lusternik reactions has a variational structure; it can be interpreted as the exchange of energy between the system in question and the constraining object. In Gauss–Appell–Tshetajev algorithms based on algebraically-defined virtual displacements there is no nice result like Eq. (3.32).

One can formulate the following question: what is the largest class of nonholonomic constraints with dynamically adiabatic Hamilton–Lusternik reactions? By "dynamically adiabatic" we mean that $R_i \dot{q}^i$ vanishes along all trajectories satisfying the equations of motion (3.28), but not necessarily along all curves compatible with constraints. We have no slightest idea how to answer this question. The point is that Lagrange multipliers μ^a in Eqs. (3.28) are not a priori known; they are to be determined from equations of motion. Thus we cannot exclude the situation that $\mu^a \frac{\partial F_a}{\partial v^i} \dot{q}^i$ will be constant along any solution of Eqs. (3.28) although $v^i \frac{\partial F_a}{\partial v^i}$ will not identically vanish in M. However, this is rather unlikely.

Obviously reactions which do not work along constraints restricting directions of velocities, but have a non-vanishing power when magnitudes of velocities are constrained, seem reasonable and physically intuitive.

In our opinion, the above Hamilton–Lusternik algorithm (variational algorithm) for reactions provides the most natural generalization of the holonomic d'Alembert algorithm. It is also compatible with the ideas of Dirac's generalized Hamiltonian mechanics [4]. Hamilton–Lusternik reactions are not ideal in the traditional sense of vanishing power (except for constraints restricting only directions of velocities) but nevertheless they are "ideal" in some intuitive sense, namely, they have variational structure; together with equations of motion they are derived from the Hamilton principle subject to additional conditions. In contrast to other traditional algorithms (Gauss, Appell, Tshetajev), the Hamilton–Lusternik algorithm is directly applicable to mechanical systems with constraints imposed upon higher derivatives, e.g. accelerations, or to systems with more general functional constraints, e.g. systems with feedback, servomechanisms and automatic data processing.

As an example let us quote a system with dynamical structure (L, D), (L being the usual Lagrangian depending on positions and velocities, and D — non-potential forces) subject to differential constraints

$$F_a(q, \dot{q}, \ddot{q}, \ldots, \overset{(p)}{q}) = 0, \quad a = 1, \ldots, m, \tag{3.36}$$

where $\overset{1)}{q}$ denotes the system of l-th derivatives $\frac{d^l}{dt^l} q^i$, $i = 1, \ldots, n$. The Hamilton–Lusternik

algorithm gives the equations of motion

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i_1} = D_i + R_i, \qquad F_a(q, \dot{q}, \ldots, \overset{(p)}{q}) = 0, \tag{3.37}$$

where

$$R_i = \sum_{l=0}^{p} (-1)^l \frac{d^l}{dt^l}\left(\mu^a(t)\frac{\partial F_a}{\partial q^{i}_{(l)}}\right), \tag{3.38}$$

μ^a denoting Lagrange multipliers.

Let us notice that for such systems there is no other natural algorithm, in particular, there is no natural generalization of traditional procedures like those of Gauss, Appell and Tshetajev. If $p > 1$, i.e. if constraints restrict accelerations or higher derivatives, then, as a rule, reactions R are non-adiabatic, there is no possibility to maintain constraints without performing mechanical work.

According to the Noether theorem, for higher-order differential constraints there exist also the effective energy quantity $E[L, F]$. The novelty is that it depends on higher derivatives, e.g. accelerations. For example, if we consider second-order constraints (restricting accelerations), $F_a(q, \dot{q}, \ddot{q}) = 0$, then, modulo terms vanishing in virtue of constraints equations, we have: $E[L, F] = E[L] + E[F]$, where $E[L]$ is given by the usual formula and

$$\begin{aligned} E[F] = \mu^a \frac{\partial F_a}{\partial \dot{q}^i_1}\dot{q}^i - \frac{d}{dt}\left(\mu^a \frac{\partial F_a}{\partial \ddot{q}^i}\right)\dot{q}^i + \mu^a \frac{\partial F_a}{\partial \ddot{q}^i}\ddot{q}^i \\ = \mu^a \frac{\partial F_a}{\partial \dot{q}^i}\dot{q}^i - \frac{d}{dt}\left(\mu^a \frac{\partial F_a}{\partial \ddot{q}^i}\dot{q}^i\right) + 2\mu^a \frac{\partial F_a}{\partial \ddot{q}^i}\ddot{q}^i. \end{aligned} \tag{3.39}$$

Obviously,

$$\frac{dE[L, F]}{dt} = \frac{dE[L]}{dt} + \frac{dE[F]}{dt} = D_i \dot{q}^i.$$

4. Interpretation problems and open questions

The above algorithm is purely phenomenological, just as all other traditional approaches to the problem of mechanical constraints. One considers constrained motion as a program motion in the control theory sense [5, 7]. It is obvious that from the very formal point of view there is, in general, an infinity of program forces maintaining given constraints and any particular relationship between constraints and reactions is some kind of "constitutive relations" [5, 11]. The whole historical effort concerning constraints was aimed at finding some kind of "philosopher stone", key principle distinguishing some natural shape of maintaining forces among all possible ones. This resulted in phenomenological principles like those known under the names of d'Alembert, Gauss, Appell, Tshetajev and others [1, 5, 6]. This search was based on the belief that geometrically distinguished reactions are best candidates for describing what is actually going on in naturally occurring constrained systems like rigid bodies of our macroscopic environment. Similarly, after

the feedback systems and servomechanisms have been discovered, it seemed quite natural to expect that reactions following from the mentioned geometric principles should provide the simplest or most economic scheme for maintaining required constraints in artificial systems. Obviously this belief concerning both natural and artificial systems is rather correct and physically justified. However, the point is that only for holonomic systems there was a natural consensus as to the key principle, namely d'Alembert's principle. For nonholonomic systems (especially for constraints nonlinear in velocities) all mentioned principles become more doubtful and mutually non-equivalent. They have no generalization to more complicated, e.g. higher-order constraints. The Hamilton–Lusternik principle is in our opinion maximally natural and formally applicable to any functional constraints. It maximally expresses intuitions concerning "ideal" or "optimal" extortion of constraints; ideal in the sense that our intervention into natural behaviour of the system is as small as possible. Nevertheless, even this principle is still phenomenological.

It would be very interesting and instructive to undertake the task of "microscopic" analysis of the problem of constraints. In naturally existing systems, constraints are never exact. Any real system undergoes small, technically negligible deviations from the surface M given by $F_a = 0$, $a = 1, \ldots, m$. The main part of motion, i.e. motion "along M" approximately satisfies autonomous equations independent of the mentioned "transversal" deviations. The peculiarity of systems showing the M-constrained behaviour is that their dynamical structure (L, D) splits in some sense into two parts: (L_M, D_M) responsible for motion along M, and (L_R, D_R) — physical reactions approximately maintaining constraints, i.e. responsible for the "smallness" of deviations from M. Approximate equations satisfied by the "along M" part of motion have the structure

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i_{\mathfrak{z}}} - \frac{\partial L}{\partial q^i} = D_i + R_i, \quad F_a(q, \dot{q}) = 0, \quad i = 1, \ldots, n, \quad a = 1, \ldots, m,$$

where R_i are formal reactions defined only on M and somehow derivable from (L_R, D_R). Let us observe an important difference between (L_R, D_R) and R: Physical reactions (L_R, D_R) are defined in a neighbourhood of M; if the system deviates from M, they force it to return towards M or to perform small oscillations about M. On the manifold M, physical reaction forces vanish. On the contrary, formal reactions R_i are defined only on M itself. They provide a shorthand description of the "macroscopic" effect of (L_R, D_R), i.e. the approximate confinement in M. The problem of deriving the effective algorithm for R from the "micromodel", i.e. from (L_R, D_R) is not yet generally solved. For holonomic systems it has been shown by Arnold that the elastic model of (L_R, D_R) implies in fact d'Alembert's principle. For nonholonomic constraints and general servo-constraints the problem is still open.

The Appell–Tshetajev algorithm and Hamilton–Lusternik algorithm give different results even for nonholonomic constraints linear in velocities. It is an interesting open problem to compare the detailed predictions of both models for the simplest and naturally occurring nonholonomic systems, i.e. for rolling rigid bodies. Simple systems of direct regulation based on contact forces (e.g. frictional speed reducer) and electromechanical systems provide another interesting and non-academic possibility of testing and comparing both algorithms. This is what we intend to do in subsequent papers.

References

1. P. APPELL, *Traité de mécanique rationnelle*, Gauthier-Villars, Paris 1953.
2. V. I. ARNOLD, *Mathematical methods of classical mechanics*, Springer Graduate Texts in Mathematics, 60, Springer-Verlag, New York 1978.
3. M. H. BEGHIN, *Etude théorique des compas gyrostatiques Anschütz et Sperry*, Imprimerie Nationale, Paris 1921.
4. P. A. M. DIRAC, Canadian Journal of Mathematics, 2, 129, 1950.
5. R. GUTOWSKI, *Analytical mechanics*, Polish Scientific Publishers (PWN), Warszawa 1971 [in Polish].
6. G. HAMEL, *Theoretische Mechanik*, Berlin 1949.
7. W. KORJENJEV, *Introduction into mechanics of steerable bodies*, Nauka, Moskva 1964 [in Russian].
8. W. W. KOZLOV, W. P. MAKARYTSHEV, A. W. TIMOFIEJEV, E. J. JURJEVITSH, *Dynamics of steering of robots*, Nauka, Moskva 1984 [in Russian].
9. J. I. NEJMARK, N. A. FUFAJEV, *Dynamics of nonholonomic systems*, Nauka, Moskva 1967 [in Russian].
10. E. T. WHITTAKER, *A treatise on the analytical dynamics of particles and rigid bodies with an introduction to the problem of three bodies*, At the University Press, Cambridge 1952.
11. CZ. WOŹNIAK, Arch. Mech., **26**, 1, 1974.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH.

Received January 9, 1987.