**Control and Cybernetics**

vol. **38** (2009) No. 4A

# Hamiltonian form of the maximum principle*

by

**Revaz V. Gamkrelidze**

Steklov Mathematical Institute
Moscow, Russia
gam@ipsun.ras.ru

**Abstract:** According to the goal of the Conference in Bedlewo, dedicated to the 50-th anniversary of Optimal Control theory, and considering that the 2008 year marked the centennial birthday of Lev Semenovich Pontryagin, I decided to devote my talk to a brief account on the discovery of the maximum principle and to an analysis of its basic feature, the Hamiltonian format.

**Keywords:** optimal control, Pontryagin's maximum principle, Hamiltonian form of the maximum principle.

## 1. Discovery of the maximum principle

### 1.1. Introductory remarks

In mid-1950s Lev Semenovich Pontryagin abandoned topology and completely devoted himself to purely engineering problems of mathematics. He organized at the Steklov Institute of Mathematics a seminar in applied problems of mathematics, often inviting theoretical engineers as speakers, since he considered professional command over the purely engineering part of the problem under investigation to be mandatory for its adequate mathematical handling.

He was led to the formulation of the general time–optimal problem by an attempt to solve a concrete fifth–order system of ordinary differential equations with three control parameters, related to optimal maneuvers of an aircraft, which was proposed to him by two Air Force engineers in the early Spring of 1955. Two of the control parameters entered the equations linearly and were bounded, hence from the beginning it was clear that they could not be found by classical methods, as solutions of the Euler equations. The problem was highly specific, and very soon Lev Semenovich realized that some general guidelines were needed in order to tackle the problem. I remember that he even said half–jokingly, "we must invent a new calculus of variations". As a result, the general time–optimal problem was formulated by him exactly in the form and in notations often used even today.

*Submitted: January 2009; Accepted: November 2009.

### 1.2. Formulation of the problem and the adjoint equation

Initially, it was supposed that the control vector $u$ attains its values from an open set $U \subset \mathbb{R}^r$. The crucial case for the control problems, that of a closed set $U$, was considered later. To denote control parameters, the letter "$u$" was chosen, as the first letter of the Russian word "control" — "upravlenie". Thus the following time–optimal problem was formulated:

$$\frac{dx}{dt} = f(x,u), \quad x \in \mathbb{R}^n, \ u \in U \subset \mathbb{R}^r$$

$$\left.\begin{aligned} &\frac{dx(t)}{dt} = f(x(t),u(t)), \\ &x(t_0) = x_0, x(t_1) = x_1, \\ &u(t) \in U, \ t_1 - t_0 = min. \end{aligned}\right\}$$

It should be noticed that the general optimal problem with an arbitrary integral–type functional is easily reduced to the formulated time–optimal problem, so that by solving the time–optimal problem with fixed boundary conditions we actually overcome all essential difficulties inherent in the general case.

The first and the most important step toward the final solution was made by Lev Semenovich right after the formulation of the problem, during three days, or better to say, during three consecutive sleepless nights: he suffered from severe insomnia and very often used to do maths all night long in bed. As a result, he completely disrupted his sleep and systematically took barbiturates in great quantities.

Thanks to his wonderful geometric insight, he derived from very simple duality considerations about the first order variational equation the initial version of necessary conditions, introducing an auxiliary covector–function $\psi(t)$, subject to the adjoint (linear) differential equation,

$$\frac{d\psi}{dt} = -\psi \frac{\partial f}{\partial x}(x,u).$$

This was the first time the adjoint system, which turned out to be of crucial importance for the whole subject appeared in optimal control theory. Actually, Pontryagin constructed for the first time, for the needs of optimization, what is now called the *Hamiltonian lift* of the initial family of vector fields on the configuration space of the problem into its cotangent bundle, the phase space of the problem.

### 1.3. Formulation of necessary conditions

The initial formulation of necessary conditions reported by Lev Semenovich at the seminar right after they were derived is expressed in formulas

$$\left.\begin{aligned}
&\frac{dx(t)}{dt} = f(x(t), u(t)), && (a)\\
&\frac{d\psi(t)}{dt} = -\psi(t)\frac{\partial f}{\partial x}(x(t), u(t)), && (b)\\
&\psi(t)\frac{\partial f}{\partial u^i}(x(t), u(t)) = 0,\ i = 1, \dots, r; && (c)\\
&x(t_0) = x_0,\ x(t_1) = x_1,\ u(t) \in U,\ \forall t \in [t_0, t_1].
\end{aligned}\right\} \tag{1}$$

This formulation supposes that the set $U$ of admissible values of the control function is open, though, as I already mentioned, from the very beginning it was clear that the ultimate result should have been applicable to closed sets as well. The $r$ "finite" equations $(c)$ were considered as relations which "dynamically" eliminate, "in the generic case", $r$ unknown control parameters as we proceed along the trajectory, thus admitting a unique solution of differential equations $(a)$–$(b)$ with given initial conditions.

The system of equations $(c)$ is solvable with respect to $u^1, \dots, u^r$ in the neighborhood of a given optimal solution $x(t), u(t),\ t_0 \leq t \leq t_1$, if the Hessian matrix of $\psi f$ with respect to the $u$'s along the extremal,

$$\left\|\psi(t)\frac{\partial^2 f}{\partial u^i \partial u^j}(x(t), u(t))\right\| = \psi(t)\frac{\partial^2 f}{\partial u^2}(x(t), u(t)),$$

is nondegenerate — a condition which should be included into the generic case conditions.

I shall describe now Pontryagin's very simple and straightforward geometric arguments leading to the equations (1).

### 1.4. Proof of necessary conditions (1)

Since the set $U$ is open, we can assume that the variation $\delta u$ of a given optimal control $u(t),\ t_0 \leq t \leq t_1$, is an arbitrary piecewise continuous function. Consider the corresponding first order variation

$$\delta x(t),\ t_0 \leq t \leq t_1,\ \delta x(0) = 0,$$

of the optimal trajectory $x(t),\ t_0 \leq t \leq t_0$, i.e. the solution of the *linear (nonhomogeneous) variational equation*,

$$\frac{d}{dt}\delta x = \frac{\partial f}{\partial x}(x(t), u(t))\delta x + \frac{\partial f}{\partial u}(x(t), u(t))\delta u(t), \quad \delta x(t_0) = 0.$$

Introducing fundamental matrices of solutions $\Phi(t), \Psi(t)$ of the corresponding homogeneous and adjoint equations,

$$\left.\begin{array}{c}\dfrac{d}{dt}\Phi = \dfrac{\partial f}{\partial x}(x(t), u(t))\Phi, \quad \dfrac{d}{dt}\Psi = -\Psi\dfrac{\partial f}{\partial x}(x(t), u(t)),\\ \Phi(t) = \Psi^{-1}(t)\ \forall t \in [t_0, t_1],\end{array}\right\}$$

we represent $\delta x(t)$ as

$$\delta x(t) = \Phi(t)\int_{t_0}^{t}\Psi(\tau)\frac{\partial f}{\partial u}(x(\tau), u(\tau))\delta u(\tau)d\tau, \quad t \in [t_0, t_1].$$

The linear mapping $L$ from the vector space of all variations $\delta u(t)$ into $\mathbb{R}^n$, defined by the last expression for $t = t_1$,

$$\{\delta u(t), t_0 \leq t \leq t_1\} \mapsto \delta x(t_1) = L(\delta u(t)),$$

is a subspace $\Gamma \subset \mathbb{R}^n$, and $\Pi = x(t_1) + \Gamma$ is a plain through $x(t_1)$. Since $x(t)$ is optimal, the implicit function theorem immediately implies, provided some general position assumptions are satisfied, that $dim\,\Pi = dim\,\Gamma \leq n-1$. Hence,

$$\exists \chi \neq 0 : \ \chi\delta x(t_1) = 0 \quad \forall \delta x(t_1) \in \Gamma.$$

Finally,

$$\left.\begin{array}{l}\chi\delta x(t_1) = \chi\Phi(t_1)\displaystyle\int_{t_0}^{t_1}\Psi(\tau)\frac{\partial f}{\partial u}(x(\tau), u(\tau))\delta u(\tau)d\tau = \\ \displaystyle\int_{t_0}^{t_1}\chi\Phi(t_1)\Psi(\tau)\frac{\partial f}{\partial u}(x(\tau), u(\tau))\delta u(\tau)d\tau = \\ \displaystyle\int_{t_0}^{t_1}\psi(\tau)\frac{\partial f}{\partial u}(x(\tau), u(\tau))\delta u(\tau)d\tau = 0 \qquad \forall\,\delta u(\tau),\end{array}\right\}$$

where

$$\psi(t) = \chi\Phi(t_1)\Psi(t), \ t_0 \leq t \leq t_1, \quad \psi(t_1) = \chi \neq 0.$$

$$\Downarrow$$

$$\left.\begin{array}{l}\dfrac{d}{dt}\psi(t) = -\psi(t)\dfrac{\partial f}{\partial x}(x(t), u(t)),\\ \psi(t)\dfrac{\partial f}{\partial u}(x(t), u(t)) = 0 \quad \forall t \in [t_0, t_1].\end{array}\right\}$$

Thus we come to the optimality conditions (1). They readily imply the Euler–Lagrange equations for the Lagrange problem of the classical calculus of variations.

### 1.5. The second variation

As soon as equations (1) were obtained, Lev Semenovich recognized, as I already mentioned, the decisive role of the covector–function $\psi(t)$ and the adjoint equation for the whole problem. He considered, in the generic case, $r$ finite equations (1)–($c$) as conditions, which dynamically eliminate the $r$–dimensional control parameter $u$ from system (1) as we proceed along the trajectory, thus making it possible to solve uniquely the $2n$–th order system of differential equations (1)–($a$) – (1)–($b$) with a given initial condition $x(t_0) = x_0$ and an arbitrary (nonzero) initial condition for $\psi$. All such solutions were declared as *extremals of the problem*, among which the optimal solutions were to be looked for.

Further development of this initial picture, based on the equations (1), brought Lev Semenovich to the idea about a universal procedure of elimination of the vector control parameter that reduces the problem of finding extremals to the solution of the system of ordinary differential equations with given boundary conditions. The ultimate realization of this idea found its expression in the Maximum Principle, which was formulated by him several months after his first report at the seminar and was supported by the subsequent advancements achieved meanwhile in the framework of the seminar.

After his talk in the seminar, Pontryagin suggested to V. Boltyanski and me, his former students and close collaborators at that time, to join him in his investigations of the problem. V. Boltyanski held a formal position at the Steklov Institute as Pontryagin's assistant, helping him in everyday computations and manuscript editing; I was a young member of the department of the Steklov Institute, headed by Pontryagin.

I decided to apply Pontryagin's geometric approach to the second order approximation of the optimal solution and thus started to investigate the second variation of the problem.

My second order considerations heavily demanded from the very beginning general position assumptions, which were overcome only in the final version of Boltyanski's proof of the Maximum Principle. The set of admissible values of the control parameters was still assumed to be open.

Take an arbitrary "generic" solution of the optimal problem, $x(t), u(t)$, $t \in [t_0, t_1]$, which means that the plane $\Pi$ introduced above is of *maximal possible dimension* $n-1$ and the trajectory $x(t)$ intersects $L$ at $x(t_1)$ transversally. Hence, $\Pi$ divides $\mathbb{R}^n$ into distinguishable half–spaces, $\mathbb{R}^n_-$ — before $x(t)$ intersects $L$ and $\mathbb{R}^n_+$ — after the intersection. Additionally, we assume that the Hessian of $\psi f(x, u)$ along the extremal under the consideration,

$$\left\| \psi(t) \frac{\partial^2 f}{\partial u^i \partial u^j}(x(t), u(t)) \right\|,$$

is nondegenerate.

It is evident that second order displacements of the endpoint of the optimal trajectory should be considered only for control variations $\delta u(t)$ that annihilate the corresponding first order displacements $\delta_1 x(t_1)$, which we denote from now on with the subscript 1. In other words, we consider the second order displacements of the endpoint of the optimal trajectory on the kernel

$$\mathbf{K} = \left\{ \delta u \middle| \delta_1 x(t_1) = \Phi(t_1) \int_{t_0}^{t_1} \Psi(\tau) \frac{\partial f}{\partial u}(x(\tau), u(\tau)) \delta u(\tau) d\tau = 0 \right\}.$$

The *second variation* $\delta_2 x(t),\ t_0 \leq t \leq t_1$, of $x(t)$ (with the initial condition $\delta_2 x(t_0) = 0$) is defined as the solution of the linear nonhomogeneous equation,

$$\frac{d}{dt}\delta_2 x = \frac{\partial f}{\partial x}(x(t), u(t))\delta_2 x + \delta u(t)^* \left\| \frac{\partial^2 f}{\partial u^i \partial u^j}(x(t), u(t)) \right\| \delta u(t) + \\ \delta_1 x(t)^* \left\| \frac{\partial^2 f}{\partial x^i \partial u^j}(x(t), u(t)) \right\| \delta u(t) + \delta_1 x(t)^* \left\| \frac{\partial^2 f}{\partial x^i \partial x^j} \right\| \delta_1 x(t),$$

which differs from the first order variational equation only by the nonhomogeneous part, quadratic in $\delta u$.

Geometric considerations, similar to those described above, and applied to the same configuration, lead us to the conclusion that the second order displacement $\delta_2 x(t_1)$ of the endpoint is contained in the half–space $\mathbb{R}^n_-$. This yields, for the correctly normalized endpoint $\psi(t_1)$ (directed toward the half–space $\mathbb{R}^n_+$ ), to the inequality,

$$\left. \begin{aligned} \psi(t_1)\delta_2 x(t_1) = \int_{t_0}^{t_1} \Big\{ \delta u(\tau)^* \left\| \psi(\tau) \frac{\partial^2 f}{\partial u^i \partial u^j}(x(\tau), u(\tau)) \right\| \delta u(\tau) + \\ \delta_1 x(\tau)^* \left\| \psi(\tau) \frac{\partial^2 f}{\partial x^i \partial u^j}(x(\tau), u(\tau)) \right\| \delta u(\tau) + \\ \delta_1 x(\tau)^* \left\| \psi(\tau) \frac{\partial^2 f}{\partial x^i \partial x^j}(x(\tau), u(\tau)) \right\| \delta_1 x(\tau) \Big\} d\tau \leq 0 \\ \forall \delta u \in \mathbf{K}. \end{aligned} \right\}$$

After some elaborate investigation of this integral quadratic form, I came to the conclusion that its nonpositivity on $\mathbf{K}$ and general position assumptions imply the nonpositivity on $\mathbf{K}$ of its singular part, hence the pointwise nonpositivity of the nondegenerate quadratic form in $\delta u$,

$$\left. \begin{aligned} \delta u^* \left\| \psi(t) \frac{\partial^2 f}{\partial u^i \partial u^j}(x(t), u(t)) \right\| \delta u \leq 0 \\ \forall \delta u \in \mathbb{R}^r,\ t \in [t_0, t_1]. \end{aligned} \right\} \tag{2}$$

Together with equations (1) we come to the optimality conditions up to the second order,

$$\left.\begin{aligned}
&\frac{dx(t)}{dt} = f(x(t),u(t)), \quad x(t_0)=x_0,\ x(t_1)=x_1,\\
&\frac{d\psi(t)}{dt} = -\psi(t)\frac{\partial f}{\partial x}(x(t),u(t)), \quad \psi(t_0)\neq 0,\\
&\psi(t)\frac{\partial f}{\partial u}(x(t),u(t)) = 0, \quad u(t)\in U,\ \forall t\in[t_0,t_1],\\
&\delta u^* \left\|\psi(t)\frac{\partial^2 f}{\partial u^i \partial u^j}(x(t),u(t))\right\| \delta u \le 0 \quad \forall \delta u \in \mathbb{R}^r.
\end{aligned}\right\}$$

## 1.6. Local form of the maximum principle

Collecting necessary conditions (1), (2) together, I recognized that a certain stable combination of symbols reappears in all of them, namely, the scalar function of three arguments $\psi, x, u$,

$$H(\psi,x,u) = \psi f(x,u). \tag{3}$$

It enables us two rewrite the system (1), $(a)$–$(b)$, as a Hamiltonian system (4) with the Hamiltonian function (3), together with additional conditions (1)–$(c)$, (2), written as (5)–(6):

$$\left.\begin{aligned}
\frac{dx(t)}{dt} &= \frac{\partial H}{\partial \psi}(\psi(t),x(t),u(t)),\\
\frac{d\psi(t)}{dt} &= -\frac{\partial H}{\partial x}(\psi(t),x(t),u(t));
\end{aligned}\right\} \tag{4}$$

$$\left.\begin{aligned}
&\frac{\partial H}{\partial u^i}(\psi(t),x(t),u(t)) = 0,\\
&\forall t\in[t_0,t_1],\ i=1,\dots,r;
\end{aligned}\right\} \tag{5}$$

$$\left.\begin{aligned}
&\delta u^* \left\|\frac{\partial^2 H}{\partial u^i \partial u^j}(\psi(t),x(t),u(t))\right\| \delta u \le 0,\\
&\qquad\qquad \forall\, \delta u \in \mathbb{R}^r.
\end{aligned}\right\} \tag{6}$$

*They assert that generic extremals are solutions of the Hamiltonian system (4), and, according to (5), their points are stationary points of the Hamiltonian (3) with respect to the control parameters $u^i$. Moreover, according to (6), along generic (regular) extremals (for which the form (6) is definite by definition), the function $H$ attains its local maximum with respect to $u$.*

Since the admissible set $U$ is open, we can combine two independent conditions (5)–(6) into one condition and write,

$$H(\psi(t),x(t),u(t)) = \max_{u\in O_t} H(\psi(t),x(t),u), \tag{7}$$

where $O_t$ is a neighborhood of $u(t)$. Furthermore, the equations (4)–(5) imply,

$$\frac{dH}{dt}(\psi(t),x(t),u(t)) = \left(\frac{\partial H}{\partial \psi}\frac{d\psi}{dt} + \frac{\partial H}{\partial x}\frac{dx}{dt}\right) + \frac{\partial H}{\partial u}\frac{du}{dt} \equiv 0.$$

It is also easy to show that $H(\psi(t),x(t),u(t))$, as a function of $t$, is continuous, even if the control function $u(t)$ has jumps. Hence, taking into account the generic character of the solution — the trajectory $x(t)$ is transversal to $L$ at $x(t_1)$ — we obtain,

$$\left.\begin{array}{r} H(\psi(t),x(t),u(t)) \equiv const = \\ \psi(t_1)f(x(t_1),u(t_1)) > 0. \end{array}\right\} \tag{8}$$

After the relations (4), (7) were written, Lev Semenovich realized that the universal elimination method of the control parameters, he was looking for, was found. He replaced the local maximum condition (7) by the global maximum over the whole set $U$, the *"Pontryagin maximum condition"* (9), which made any restrictive assumptions about the admissible set $U$ superfluous,

$$\begin{array}{c} H(\psi(t),x(t),u(t)) = \\ \max_{u\in U} H(\psi(t),x(t),u) \equiv const \geq 0. \end{array} \tag{9}$$

Thus, he came to the final formulation of the Maximum Principle, combining the Hamiltonian system (4) with the maximum condition (9) and dropping off any assumptions about genericity of the solutions or the structure of the admissible set $U$.

### 1.7. Final form of the Maximum Principle

*Suppose a controlled equation is given,*

$$\frac{dx}{dt} = f(x,u),\ x\in\mathbb{R}^n,\ u\in U\subset\mathbb{R}^r,$$

*where the admissible set $U$ is arbitrary. We introduce the Hamiltonian function of the problem (3), which depends on three arguments — the covector $\psi$ and the vectors $x$ and $u$. If $u(t),\ t_0\leq t\leq t_1$, is a time–optimal control, $x(t),\ t_0\leq t\leq t_1$, the corresponding time–optimal trajectory, then there exists a nonzero covector function $\psi(t)$ such that the triple $\psi(t),x(t),u(t),\ t_0\leq t\leq t_1$, is a solution of the Hamiltonian system (4), and the maximum condition (9) holds.*

In this final form, the maximum condition (9) could be viewed not only as a universal elimination method of the undetermined parameter $u$, but also as a generalization of the Legendre transformation from the state–space variables $(x,u)$ to the phase space variables $(\psi,x)$.

It took approximately a year before a full proof of the Maximum Principle in its final formulation was accomplished. This became possible after Boltyanski introduced a new type of variations, the "needle variations" of the control function. These variations are zero everywhere on the time–interval, except on several segments with a small total length, where they can attain arbitrary admissible values, and they have an important property of admitting an operation of convex combination, regardless of the shape of $U$, and still supplying a sufficiently rich first order attainable set. These variations made possible to prove the Maximum Principle in full generality, as formulated above. This was Boltyanski's major contribution to the subject.

Though formulated in 1955, the Maximum Principle was never changed, nor slightly generalized since then. All (first order) advancements were directed toward generalizations of the optimal problem itself, especially toward developing nonsmooth optimization, with corresponding first–order necessary conditions shaped after the Maximum Principle. Still it would be useful to give here an invariant formulation of the maximum principle and a brief analysis of its Hamiltonian format.

## 2. About the Hamiltonian format of the maximum principle

### 2.1. Introductory remarks

The native Hamiltonian format of Pontryagin's maximum principle is its basic feature, inherent in the principle regardless of any regularity conditions imposed on the optimal problem under consideration. It assigns canonically to the problem a family of Hamiltonian systems, indexed with the control parameter, and complements the system with the maximum condition, which makes it possible to solve the initial value problem for the system, "dynamically" eliminating the parameter as we proceed along the trajectory, thus providing the extremals of the problem.

Much was said about the maximum condition since its discovery in 1956, and all achievements in the field were mainly credited to it, whereas the Hamiltonian format of the maximum principle was always taken for granted and never discussed seriously.

Meanwhile, the very possibility of formulating the maximum principle is intimately connected with its native Hamiltonian format and with the parameterization of the problem with the control parameter.

As I described it above, both these starting steps were made by L. S. Pontryagin in 1955 on a completely empty spot, in fact, out of nothing, and they eventually led to the discovery of the maximum principle.

In this section, I shall consider this, now semi-historical, topic and give a short analysis of the Hamiltonian format of the maximum principle.

For the simplicity of considerations, I shall start with an invariant formulation of the Principle for the time optimal autonomous case, to which the general optimal problem (with an integral-type functional) is easily reduced.

### 2.2. Invariant formulation of the maximum principle

A controlled equation is given on the configuration space $M$ of the time-optimal problem,

$$\frac{dx}{dt} = X(x,u) = X,\ x \in M,\ u \in U.$$

Canonically associated with every family of vector fields $X$ (indexed by the control parameter $u$) on $M$ is a scalar-valued function $H_X$ on the cotangent bundle $\pi : T^*M \longrightarrow M$, linear in $\psi$ and indexed by the parameter $u$, the *Hamiltonian* of the problem,

$$H_X(\psi,u) \stackrel{def}{=} < \psi, X(\pi\psi,u) >,\ \psi \in T^*M,\ u \in U,$$

where $< \cdot,\cdot >$ are the duality brackets between the vectors and covectors on $M$, and every fiberwise linear scalar-valued function from $C^\infty(T^*M)$ could be obtained in this way.

Thus, the Hamiltonian of the optimal problem is canonically defined by the problem itself. Denoting by $\overrightarrow{H}_X$ the family of Hamiltonian vector fields generated on $T^*M$ by the family of Hamiltonians $H_X$, we obtain the controlled *Hamiltonian equation of the problem containing the control parameter u,*

$$\frac{d\psi}{dt} = \overrightarrow{H}_X(\psi,u).$$

Finally, the *maximum condition* takes the form,

$$H_X(\psi,u) = \max_{v\in U} H_X(\psi,v).$$

As I described it in Section 1, the canonical construction of the Hamiltonian vector field $\overrightarrow{H}_X$ on the cotangent bundle $T^*M$ was actually introduced by L.S. Pontryagin in 1955, and turned out to be the first step towards the discovery of the maximum principle.

The maximum principle is formulated in the following way.

Pontryagin's maximum principle. *Every extremal over a point $x \in M$ of the time-optimal problem could be obtained as a common solution $\psi(t)$ to the initial value problem $\psi(0) = \psi_0 \neq 0,\ \pi\psi_0 = x$, for the system*

$$\begin{aligned} \frac{d\psi(t)}{dt} &= \overrightarrow{H}_X(\psi(t),u(t)), \\ H_X(\psi(t),u(t)) &= \max_{v\in U} H_X(\psi(t),v). \end{aligned}$$

*Hence, $u(t)$ could be considered to be obtained, in some generalized sense, in the process of "dynamical elimination" of the parameter $u$ by the maximum condition as we proceed along the trajectory $\psi(t)$.*

### 2.3. The regular case

Assume that the maximum condition, written as an equation in $u$,

$$H_X(\psi, u) = \max_{v\in U} H_X(\psi, v), \qquad (*)$$

has a unique smooth solution $u = u(\psi)$ in some region $O \subset T^*M$. Substituting $u = u(\psi)$ into $H_X(\psi, u)$, we obtain the "master Hamiltonian" $\mathcal{H}_X$ (without a parameter) and the corresponding "master Hamiltonian vector field" $\overrightarrow{\mathcal{H}}_X$ of the optimal problem,

$$\mathcal{H}_X(\psi) = H_X(\psi, u(\psi)),\ \overrightarrow{\mathcal{H}}_X = \overrightarrow{H}_X(\psi, u(\psi)),$$
$$\psi \in O \subset T^*M.$$

Then, the extremals over a given point $x \in M$ of the problem are represented as a flow $\psi(t, \psi_0)$ generated by the Hamiltonian vector field $\overrightarrow{\mathcal{H}}_X$ and the given initial conditions,

$$\frac{d}{dt}\psi(t, \psi_0) = \overrightarrow{\mathcal{H}}_X(\psi(t, \psi_0)),$$
$$\psi(0, \psi_0) = \psi_0 \neq 0,\ \pi\psi_0 = x,\ \psi \in O \subset T^*M.$$

Thus, in the regular case, the maximum condition gives a canonical transition from the tangent bundle description of the optimal motion (the "Lagrangian picture") to the cotangent bundle description (the "Hamiltonian picture"), and could be considered to be a generalization of the Legendre transformation.

Still, restriction of our considerations to the regular case only, though it contains all regular problems of the classical calculus of variations, is very limiting and completely unacceptable for the optimal control theory, where the strongly degenerate problems, such as linear control problems, are of paramount importance, which, in fact, stimulated the discovery of the maximum principle and the development of the discipline as a whole.

### 2.4. Identification of the Pontryagin vector field $\overrightarrow{H}_X$

Since the family of vector fields $\overrightarrow{H}_X$ is canonically derived from $X$, it is natural to expect that $\overrightarrow{H}_X$ could be expressed through standard differential-geometric first order invariants of the configuration space $M$. The corresponding expression could be obtained in the following way.

Denote the $C^\infty(M)$-modules of fiberwise constant and fiberwise linear smooth functions on $T^*M$, and their direct sum, respectively by

$$\mathfrak{M}_{(0)} = \pi^* C^\infty(M),\ \mathfrak{M}_{(1)} = Vect\, C^\infty(M),$$
$$\mathfrak{M} = \mathfrak{M}_{(0)} \oplus \mathfrak{M}_{(1)} \subset C^\infty(T^*M).$$

For a given $X$, the derivation $ad_X$ on $Vect\, C^\infty(M)$,

$$ad_X : Y \mapsto ad_X Y = [X, Y],\ ad_X \cdot aY = Xa \cdot Y + a \cdot ad_X Y,$$

where $[\cdot,\cdot]$ are the Lie brackets, could be plainly extended from $\mathfrak{M}_{(1)}$ to the direct sum $\mathfrak{M}$, if we put

$$ad_X \cdot \pi^* a = \pi^* \cdot Xa \quad \forall a \in C^\infty(M).$$

Hence, $ad_X$ preserves both $C^\infty(M)$-submodules, $\mathfrak{M}_{(0)}, \mathfrak{M}_{(1)}$, therefore it could be uniquely extended as a derivation on $C^\infty(T^*M)$, i.e. as a vector field on the cotangent bundle $T^*M$. *We shall assume further under the symbol $ad_X$ exactly this vector field.*

Straightforward computations show that the vector fields $\overrightarrow{H}_X$ and $ad_X$ coincide on fiberwise constant and fiberwise linear scalar valued functions on $T^*M$, hence they are identical on $T^*M$ for every fixed value of the parameter $u$,

$$\overrightarrow{H}_X = ad_X.$$

Thus, for $\forall t$, the vector field $\overrightarrow{H}_X(\psi, u(t))$ is a *Hamiltonian lift* over $X$ to the cotangent bundle $T^*M$,

$$\pi_*(\overrightarrow{H}_X)_\psi = X_{\pi\psi} \quad \forall \psi \in T^*M,$$

and preserves the $C^\infty(M)$-module of vector fields $Vect\, M$ on $M$. Therefore, the flow $G_t$ generated on the cotangent bundle $T^*M$ by the nonstationary Hamiltonian lift $\overrightarrow{H}_X(\psi, u(t))$, where $u(t)$ is an arbitrary control function, is a bundle flow over the flow $g_t$ on $M$, generated by the nonstationary vector field $X(x, u(t))$, i.e. transforms every fiber $T^*_z M$ linearly (and nondegenerate) to the fiber $T^*_{g_t z} M$.

Let $\mathcal{L}_X$ be the Lie derivative over the vector field $X$ (for a fixed value of $u$), i.e. a vector field on the *tangent bundle* $TM$, generating on $TM$ the flow $e^{tX}_*$, the differential of the flow $e^{tX}$ on $M$,

$$e^{t\mathcal{L}_X} = e^{tX}_* : TM \longrightarrow TM.$$

There exists a natural duality between the flows $e^{t\mathcal{L}_X}$ and $e^{t\,ad_X}$ on $TM$ and $T^*M$, which it is convenient to express considering these flows, as well as the flow $e^{tX}$ on $M$, also as flows of the corresponding pullback automorphisms of the algebras $C^\infty(TM), C^\infty(T^*M), C^\infty(M)$, respectively. Then the duality relation could be expressed by the identity

$$e^{tX} < \omega, Y > = < e^{t\mathcal{L}_X}\omega, e^{tad_X}Y >$$
$$\forall Y \in Vect\, M,\ \omega \in \Lambda^{(1)}(M),$$

which is easily derived from the relation

$$Ad\, e^{tX} \cdot Y \overset{def}{=} e^{tX} \circ Y \circ e^{-tX} = e^{tad_X}Y \quad \forall Y \in Vect\, M.$$

The identity asserts that

$$e^{tad_X} = \left(e^{t\mathcal{L}_X}\right)^{*-1} = \left(e^{-t\mathcal{L}_X}\right)^*,$$

where the upper star on the flow denotes conjugation,

$$< \omega, (e^{t\mathcal{L}_X})^*Y > \overset{def}{=} e^{-tX} < e^{t\mathcal{L}_X}\omega, Y >$$
$$\forall X, Y \in Vect\, M,\ \omega \in \Lambda^{(1)}(M).$$

Thus, for every fixed $u$, *the flow* $e^{\overrightarrow{H}_X}$ *generated by the Pontryagin vector field* $\overrightarrow{H}_X$ *is inverse to the conjugate of the differential* $e^{t\mathcal{L}_X}$ *of the flow* $e^{tX}$.

Differentiating the duality relation between the flows given above with respect to $t$ and then putting $t = 0$, we obtain the "infinitesimal" version of the above duality between the corresponding vector fields $\mathcal{L}_X$ and $ad_X = \overrightarrow{H}_X$ (the generalized Leibnitz rule),

$$X < \omega, Y > = < \mathcal{L}_X\omega, Y > + < \omega, ad_XY >$$
$$\forall Y \in Vect\, M,\ \omega \in \Lambda^{(1)}(M).$$

The obtained relations completely identify the Pontryagin vector field $\overrightarrow{H}_X$, hence the Hamiltonian format of the maximum principle.