The Pontryagin Maximum Principle 50 years later

A. A. Agrachev and R.V. Gamkrelidze

1 Introductory remarks

Exactly 50 years ago the Pontryagin Maximum Principle was formulated in [1], which heralded the emergence of the mathematical theory of optimal control. The elapsed half–a–century witnessed a successful development of the theory by many researchers in the field, among whom a prominent position certainly should be credited to the scientist to whom this volume is dedicated. We thought therefore appropriate to present here our vision of the maximum principle and of some of its consequences as we see them today.

Formulation in mid fifties of the past century of an optimal control problem with a closed set of admissible values for the control parameter was a new type of an extremal problem not amenable to the existing methods. From the very beginning it was obvious that an adequate mathematical treatment of the problem would prompt new type of "extremality conditions". It was also clear that the real reason of the difficulties was not the generality of the problem, but rather its specific character related to the fact that, in typical problems, the admissible set of values of the control parameter was closed. Whereas the simplest time–optimal problems for linear second order differential equations with a closed interval on the real line as an admissible set of control seemed to be completely unamenable, the most general control problem with an open set for the admissible values of the control parameter was trivially reduced to the classical Lagrange problem, hence the differential equations of the extremals for the problem were readily at hand.

The Maximum Principle was a mathematical answer to the challenge posed by modern technologies and was fully endorsed by engineers right after its publication as an adequate solution of the problem, though, in the beginning, mathematical communities involved in related fields were quite reluctant, with very few exceptions, to accept the result as something completely new.

Therefore, in mid sixties, there were intensive though unproductive discussions about the possibilities of deriving the maximum principle from classical results of the calculus of variations. In our opinion, there are two independent ways of proving the maximum principle in its natural generality — the original proof of V. G. Boltyansky, [2], employing the needle variations invented especially for this reason, and a second proof, which is based on generalized (relaxed) controls, cf [3]. A crucial feature of both proofs is a special way in which convex combinations of control variations are introduced, completely irrelevant to the structure of the admissible set of the problem, and both providing a sufficiently reach first order attainable sets. At the same time, classical variations heavily depend on the admissible set of the problem, unless the control function is contained in the interior of the admissible set.

The impossibility of deriving the maximum principle from classical constructions certainly could not be considered as a special merit of the principle. We think, the growing scientific demand that the maximum principle enjoyed ever since it has been formulated is due to two of its features that really are remarkable.

First of all, formulation of the maximum principle, though very simple and universally general, without any exceptions, still works at full swing in amazingly many nontrivial cases, often generating whole families of minimizers (the so called optimal synthesis), and thus yielding complete solutions to otherwise unamenable problems. Such an efficiency could serve as a partial explanation of the fact that, despite its simplicity and generality, the maximum principle is quite hard to prove.

As the second remarkable feature of the maximum principle we consider the symplectic invariant form in which it was actually formulated from the very beginning, thus clarifying the real intrinsic meaning of the extremality conditions. The maximum principle canonically assigns to an optimal problem on the configuration manifold a Hamiltonian vector field on the cotangent bundle of the manifold, reducing the solution of the problem to finding trajectories of the Hamiltonian field that satisfy additionally a certain maximum condition. Such procedures that reduce the solution of the initial problem on a manifold to the study of some canonically constructed infinitesimal objects on its cotangent bundle are often encountered in modern analysis under the name of canonical microlocalization. They readily supply us with invariants of the initial problem, as well as of the manifold itself on which the problem is considered. Historically, the maximum principle was one of the first procedures where this transition was explicitly described.

Due to its general character and completely invariant form already in the original formulation, the maximum principle was never subject to any essential generalizations. All advances related to the first order necessary conditions in optimal control were attained by generalizing the problems themselves, mainly in the direction of non–smooth analysis, and deriving for them (first–order) necessary conditions shaped in a larger extent after the maximum principle.

In the next Section 2 of this expository paper we give a brief account of necessary material on Hamiltonian lifts of vector fields on smooth manifolds, and then, in Section 3, an invariant formulation of the maximum principle. The final section Section 4 is devoted to the second–order optimality conditions unified by the notion of the Jacobi curve.

Our exposition is restricted to the time–optimal case, which already contains all the essential difficulties inherent in the general optimal problem (with an arbitrary integral type functional to be minimized), and at the same time could be formulated in a much simpler and geometrically appealing form. Besides, the general optimal problem is easily reduced to the time–optimal problem by simply rescaling the time–parameter. Finally, the time–optimal problem has one additional advantage being intrinsically (canonically) connected with the given dynamical system, hence with the initial manifold on which the system is defined.

2 Hamiltonian lift of vector fields

We start with some notational and terminological comments related to the geometry of tangent and cotangent bundles of a smooth n–dimensional manifold M.

Every (smooth) vector field X on M will be considered, without any further remarks, both as a differentiation of the real algebra $C^\infty(M)$ of smooth functions on M, and as a smooth cross–section of the tangent bundle TM,

$$X : z \mapsto X_z \in T_z M \subset TM, \; z \in M.$$

Thus the family of all (smooth) vector fields on M, $Vect\, M$, will be considered, depending on the context, as a real Lie algebra or as a $C^\infty(M)$–module.

But it is worth mentioning that vector fields have an additional third image important for our purposes. Namely, every vector field X on M is canonically identified, according to its definition, with a scalar–valued function H_X on the cotangent bundle T^*M that is linear on fibers $T^*_z M \subset T^*M, \; z \in M$. Conversely, every such function H is uniquely represented as $H = H_X$. Indeed, the action of X on functions $a \in C^\infty(M)$ is given by the formula

$$(Xa)_z = (< X, da >)_z \stackrel{def}{=} H\left((da)_z\right), \; z \in M,$$

where $(da)_z$ is the value of the differential da at z and $< \cdot, \cdot >$ is the usual duality bracket. Therefore, we shall make no difference between H_X and X, and where appropriate substitute H_X by X.

An arbitrary point of the cotangent bundle T^*M will be denoted ξ; if it belongs to the fiber $T^*_z M$ we also write $\xi = \xi_z$. The canonical projection of T^*M is denoted by $\pi : T^*M \longrightarrow M$, its differential by $d\pi : T(T^*M) \longrightarrow TM$. The $\mathbb{R}$–algebra $C^\infty(T^*M)$ will also be considered, depending on the context, as a $C^\infty(M)$–module, with multiplication by "scalars" defined as

$$aH \overset{def}{=} \pi^* a \cdot H \quad \forall a \in C^\infty(M),\ H \in C^\infty(T^*M),$$

where $\pi^* a = a \circ \pi$ is the pullback of a under the projection π.

Alongside with smooth functions on T^*M that are linear on fibers, i.e. vector fields on M, a special role in the geometry of M is played by smooth functions on T^*M that are constant on fibers, since together they provide us with a sufficient supply of functions to cover the whole space T^*M with "canonical" coordinate neighborhoods.

Indeed, smooth functions constant on fibers are plainly identified with pullbacks $\pi^* a \overset{def}{=} a \circ \pi,\ a \in C^\infty(M)$. Furthermore, for every coordinate neighborhood (O, x) of M, $x = (x^1, \ldots, x^n)\colon O \longrightarrow \mathbb{R}^n$, we canonically construct the *bundle or canonical neighborhood of* T^*M *over* (O, x), the pair

$$\left(\pi^{-1}O, (q, p)\right),\ q = (q^1, \ldots, q^n),\ p = (p_1, \ldots, p_n),$$

with the coordinate mapping

$$(q, p) : \pi^{-1}O \longrightarrow \mathbb{R}^{2n},\ q^i = \pi^* x^i,\ p_j = \frac{\partial}{\partial x^j},\ i, j = 1, \ldots, n.$$

The coordinate functions $q^i,\ i = 1, \ldots, n$ are pullbacks of the corresponding coordinates x^i, hence constant on fibers, their "conjugates" p_i are equal to the partial derivatives with respect to the corresponding x^i, hence linear on fibers. Thus every smooth function H on T^*M is locally represented as $H = F(q, p), F \in C^\infty(R^{2n})$. In particular, the functions H_X introduced above are represented in coordinate neighborhoods $\pi^{-1}O$ as

$$H_X = \sum_{\alpha=1}^{n} \pi^* X^\alpha \cdot p_\alpha = \sum_{\alpha}^{n} X^\alpha p_\alpha,$$

where $X = \sum\limits_{\alpha=1}^{n} X^\alpha \dfrac{\partial}{\partial x^\alpha}$ is a vector field on O.

The $C^\infty(M)$–submodule of $C^\infty(T^*M)$ generated by functions that are constant or linear on fibers will be denoted as

$$\mathfrak{M} = \left\{ a + bH_X \,\middle|\, a, b \in C^\infty(M),\ X \in Vect\, M \right\}.$$

An $\mathbb{R}$–linear operator $D_X : \mathfrak{M} \longrightarrow \mathfrak{M}$ is called a *differentiation of* $\mathfrak{M}$ *over a fixed vector field* $X \in Vect\, M$, iff

$$D_X(a + bH_Y) = Xa + Xb \cdot H_Y + bD_X H_Y \quad \forall a, b \in C^\infty(M),\ Y \in Vect\, M.$$

A vector field $\mathcal{W}$ on the cotangent bundle T^*M will be called *projectable* iff the restriction of the composition $d\pi \circ \mathcal{W}$ on an arbitrary fiber $T_{\xi_z}(T^*M) \subset T(T^*M),\ z \in M$, is constant, hence equal to a tangent vector X_z to M at z,

$$d\pi \cdot \mathcal{W}_{\xi_z} = X_z \quad \forall \xi_z \in T_z^*M,\ z \in M.$$

The uniquely defined vector field X on $M,\ z \mapsto X_z,\ z \in M$, is the *projection of* $\mathcal{W}$ *onto* M and $\mathcal{W}$ is a *lift over* X *in the cotangent bundle* T^*M. If $\mathcal{W}$ is a Hamiltonian vector field on T^*M, we say that it is a *Hamiltonian lift over* X.

For example, partial derivatives $\dfrac{\partial}{\partial q^i}$ are lifts on $\pi^{-1}O$ over corresponding vector fields $\dfrac{\partial}{\partial x^i}$ on O,

$$\left(d\pi \circ \frac{\partial}{\partial q^i} \right)_{\xi_z} x^j = \left(\frac{\partial}{\partial q^i} \right)_{\xi_z} \cdot \pi^* x^j = \left(\frac{\partial}{\partial q^i} \right)_{\xi_z} q^j = \delta_i^j \quad \forall \xi_z \in T_z^*M,\ z \in M.$$

A lift $\mathcal{W}$ over a given vector field $X \in Vect\, M$ is certainly not uniquely defined, since together with $\mathcal{W}$, every vector field $\mathcal{W} + \mathcal{W}'$ is also a lift over X, where $\mathcal{W}'$ is a vertical vector field on T^*M (belongs to the kernel of $d\pi$ — a subbundle of $T(T^*M)$ of rank n). For example, the partial derivatives $\dfrac{\partial}{\partial p_i}$ on $\pi^{-1}O,\ i = 1, \ldots, n$, are vertical. But if we restrict us to Hamiltonian lifts, then the result is much more definitive. Denoting by $\overrightarrow{H}$ the Hamiltonian vector field on T^*M with the Hamiltonian function H, it is easy to see that an arbitrary Hamiltonian lift over a given vector filed $X \in Vect\, M$ has the form $\overrightarrow{H_X + K}$, where the function H_X (linear on fibers) was introduced above and is canonically defined by the vector field X, whereas K is an arbitrary function in $C^\infty(T^*M)$ constant on fibers.

Indeed, representing the vector field $\overrightarrow{H}$ locally, in canonical coordinates, as

$$\overrightarrow{H} = \sum_{\alpha=1}^{n} \left(\frac{\partial H}{\partial p_\alpha} \frac{\partial}{\partial q^\alpha} - \frac{\partial H}{\partial q^\alpha} \frac{\partial}{\partial p_\alpha} \right),$$

we have the relations,

$$d\pi\left(\overrightarrow{H}_{\xi_z}\right) = \sum_{\alpha=1}^{n}\left(\frac{\partial H}{\partial p_\alpha}\right)_{\xi_z}\left(\frac{\partial}{\partial x^\alpha}\right)_z = \sum_{\alpha=1}^{n} X_z^\alpha\left(\frac{\partial}{\partial x^\alpha}\right)_z$$
$$\forall \xi_z \in T_z^*M,\ z \in M.$$

They imply that the functions $\dfrac{\partial H}{\partial p_\alpha}$, $i = 1, \dots, n$, are constant on fibers and equal respectively to $\pi^* X^i$, hence H has the form $H = \sum\limits_{\alpha=1}^{n} p_\alpha X^\alpha + K$, where K is constant on fibers. We can make the procedure unique by taking $K = 0$.

Simultaneously, we have obtained a canonical method of constructing on the cotangent bundle T^*M of a Hamiltonian lift over a vector field X on M : *take the canonically defined scalar function* H_X *on* T^*M *and go to the corresponding Hamiltonian vector field* $\overrightarrow{H_X}$. This is exactly the procedure used in the maximum principle to introduce the Hamiltonian system of the time–optimal problem, though formulated in arbitrary coordinates.

There is a canonical procedure practiced in differential geometry, formally introduced by the universally accepted symbol $ad\, X$, $X \in Vect\, M$, leading to the Hamiltonian lift over X described above. Indeed, $ad\, X$ acts on vector fields $Y \in Vect\, M$ according to the formula,

$$ad\, X \cdot Y = [X, Y] \text{ — Lie brackets of } X, Y,$$

and is a differentiation of the module $\mathfrak{M}$ over the field X,

$$ad\, X(a + bY) \stackrel{def}{=} Xa + [X, bY] = Xa + Xb \cdot Y + b[X, Y] =$$
$$Xa + Xb \cdot H_Y + bH_{[X,Y]}.$$

Evidently, this action could be uniquely extended to a differentiation on the whole algebra $C^\infty(T^*M)$, i.e. $ad\, X$ could be uniquely extended to a vector field on T^*M, which we denote by the same symbol $ad\, X$. A straightforward computation in arbitrary canonical coordinates yields that the vector fields $ad\, X$ and $\overrightarrow{H_X}$ coincide on the submodule $\mathfrak{M}$, hence they coincide as vector fields on T^*M. We thus obtain a definition of the *Hamiltonian lift over* $X \in Vect\, M$ *as the vector field* $ad\, X$ on T^*M.

Remark. A dual construction carried out on the tangent bundle TM leads us in a similar way to a vector field $\mathcal{L}_X$ on the tangent bundle TM over an arbitrary $X \in Vect\, M$, *the Lie derivative over* X, which, restricted to differential 1–forms on M (functions in $C^\infty(TM)$ that are linear on fibers), coincides with the derivation $i_X \circ d + d \circ i_X$, where d is the exterior differential,

i_X is the interior multiplication by X. The duality relation between $ad\,X$ and $\mathcal{L}_X$ takes on the form

$$< ad\,X \cdot Y, \omega > + < Y, \mathcal{L}_X \omega > = X < Y, \omega >,$$

where $Y \in Vect\,M$ and ω is a differential 1–form on M. The corresponding bundle flows in TM and T^*M over the flow e^{tX}, $e^{t\mathcal{L}_X}$ and $e^{tad\,X}$, coincide, respectively, with the differential de^{tX} of e^{tX} and the inverse of the adjoint to the differential de^{tX}.

3 Invariant formulation of the Maximum Principle

We can give now an invariant formulation of the maximum principle based on the geometric images introduced above. We start with the definition of the time–optimal problem.

A *control system on a smooth n–dimensional manifold M*, with an *arbitrary* subset $U \subset \mathbb{R}^r$ for the *admissible set of the values of the control parameter u*, is a family $\{X(u),\ u \in U'\}$ of smooth vector fields on M such that $\overline{U} \subset U'$, where $\overline{U}$ is the closure of U, and the mapping $(z,u) \mapsto X_z(u) \in TM$ is continuous from $M \times U'$ to TM. *Admissible control functions* are, by definition, measurable essentially bounded mappings $t \mapsto u(t) \in U,\ t \geq 0$.

To any admissible control function $u(t)$ we associate a nonautonomous ordinary differential equation on M,

$$\frac{dz}{dt} = X_z(u(t)).$$

Then for any $z_1 \in M,\ t_1 \geq 0$, and sufficiently small closed interval $[t_1, t_2]$ there exists a unique (absolutely continuous) solution $z(t)$ to the above equation defined on the interval $[t_1, t_2]$ such that $z(t_1) = z_1$. This solution is called the trajectory of the control system $X(u),\ u \in U$, associated to the control function $u(t)$ and the initial condition $z(t_1) = z_1$.

We say that the pair $u(t),\ z(t),\ t \in [t_1, t_2]$, is a *solution of the time–optimal problem defined by the control system $X(u),\ u \in U$*, if for any admissible control function $v(t)$ and the corresponding solution $y(t), t_1 \leq t \leq t_2'$, of the initial value problem

$$\frac{dy}{dt} = X_y(v(t)), \quad y(t_1) = z_1,$$

the equality $y(t_2') = z(t_2)$ implies the inequality $t_2' \geq t_2$.

Denote the family of Hamiltonians on T^*M corresponding to the family $X(u)$ by $H_u = H_{X(u)}$. We say that for a given value $\widehat{u}$ of the parameter and a fixed point $\xi \in T^*M$ the *Pontryagin maximum condition* holds iff

$$H_{\widehat{u}}(\xi) = \sup_{u \in U} H_u(\xi).$$

To simplify the formulation of the maximum principle, we assume that the manifold M is naturally embedded into the cotangent bundle as an n–dimensional submanifold, $M \subset T^*M$. Every $z \in M$ is identified with the origin of the fiber $T^*_z M$, hence for an arbitrary control function $u(t),\ t_1 \leq t \leq t_2$, the corresponding functions $H_{X(u(t))} \in C^\infty(T^*M)$ turn to identical zeros on M, and the maximum condition degenerates on M into the identity,

$$H_{\widehat{u}}(\xi) \equiv \sup_{u \in U} H_u(\xi) \equiv 0 \quad \forall \widehat{u},\ \xi \in M.$$

Furthermore, the restriction of the Hamiltonian lift $\overrightarrow{H}_{X(u(t))}$ on M coincides with the nonautonomous vector field $X(u(t))$, and the uniqueness property of the solutions yields that every trajectory $\xi(t), t_1 \leq t \leq t_2$, of the Hamiltonian lift $\overrightarrow{H}_{X(u(t))}$ that has a common point with M is a trajectory of the vector field $X(u(t))$ and is completely contained in M. Therefore we consider this case as an exceptional and call trajectories of the nonautonomous vector field $X(u(t))$ *trivial trajectories of the Hamiltonian lift* $\overrightarrow{H}_{X(u(t))}$. A trajectory $\xi(t),\ t_1 \leq t \leq t_2$, of $\overrightarrow{H}_{X(u(t))}$ is plainly nontrivial iff $\xi(t_1) \neq \pi\xi(t_1)$.

Theorem 3.1 ***The Pontryagin Maximum Principle***. *Assume that a control function $u(t),\ t_1 \leq t \leq t_2$, and the corresponding trajectory $z(t), t_1 \leq t \leq t_2,\ z(t_1) = z_1$, of the nonautonomous vector field $X(u(t))$ constitute a solution of the time–optimal problem defined by the control system $X(u),\ u \in U$. Then there exists a nontrivial trajectory $\xi(t),\ t_1 \leq t \leq t_2$, of the Hamiltonian lift $\overrightarrow{H}_{X(u(t))}$ over the trajectory $z(t)$,*

$$\frac{d\xi(t)}{dt} = \overrightarrow{H}_{X(u(t))}(\xi(t)),\ \pi\xi(t_1) = z_1, \tag{3.1}$$

such that for almost all $t \in [t_1, t_2]$ the Pontryagin maximum condition holds,

$$H_{u(t)}(\xi(t)) = \sup_{u \in U} H_u(\xi(t)) \ a.e.\ on\ [t_1, t_2]. \tag{3.2}$$

If the initial family of vector fields $X(u)$ is autonomous, then the function $\sup_{u \in U} H_u(\xi(t))$ is constant and nonnegative,

$$\sup_{u \in U} H_u(\xi(t)) \equiv const \geq 0,\ t_1 \leq t \leq t_2. \tag{3.3}$$

Every nontrivial solution $\xi(t)$ of the system (3.1)–(3.2) is called an *extremal* of the time–optimal problem defined by the family $X(u)$. The maximum condition could be considered here as a generalized elimination algorithm, "dynamically " excluding the parameter u, as it evolves, at *almost every* instant of time t and thus obtaining the trajectory $\xi(t)$. But it should be emphasized that, in general, sufficiently reach families of extremals could not be incorporated into the flows on the cotangent bundle, as is the case for the regular problem of the classical calculus of variations. For example, the family of all extremals $\xi(t)$ with the initial condition $\pi\xi(t_1) = z_1$, where $z_1 \in M$ is fixed, even for simplest time–optimal problems violate the uniqueness property of the trajectories contained in a flow — two extremals with equal initial conditions do not necessarily coincide.

In our context, the maximum condition could also be considered as a generalization of the Legendre transformation for "setting the solution in canonical coordinates".

4 Jacobi curves

The maximum principle is a first order optimality condition, an elaboration of the classical Lagrange multipliers rule, where $\xi(t)$ plays the role of the Lagrange multiplier. In general, this is not a sufficient condition and an additional second order information is needed to select optimal control functions among those satisfying the maximum principle.

In this section, we demonstrate how to shape the second order optimality conditions after the maximum principle using intrinsic symplectic language and a ramification of the universal elimination principle. We shall assume that $(z, u) \mapsto X_z(u)$ is a smooth map from $M \times U'$ to TM (not only a continuous map as before).

The whole important second order information is collected in so called Jacobi curves. Given an appropriate extremal $\xi(t),\ t_1 \leq t \leq t_2$, Jacobi curve is a family $J_\xi(t),\ t_1 \leq t < t_2$, of Lagrangian subspaces of the symplectic space $T_{\xi(t_1)}(T^*M)$. We give here a general abstract construction of the Jacobi curves; their specifications and effective calculations for regular, singular, and bang-bang extremals can be found in [4, 5, 6]. The construction of a Jacobi curve is based on the "nice variations" of the reference control function.

Let $u(t),\ t_1 \leq t \leq t_2$, be an admissible control function. A *nice variation* of the control function u is the germ at 0 of a twice differentiable at 0 mapping

$$F = (f, \phi) : \mathbb{R}^k \to L^\infty([t_0, t_1]; U \times \mathbb{R})$$

such that

$$f(0)(t) = u(t),\ \phi(0)(t) = 1,\ t_1 \le t \le t_2, \quad \int\limits_{t_1}^{t_2} \phi(x)(t)\, dt = t_2 - t_1,\ x \in \mathbb{R}^k.$$

Given nice variations $F_i : \mathbb{R}^{k_i} \to L^\infty([t_0, t_1]; U \times \mathbb{R}),\ i = 1, 2,\ k_1 \le k_2$, we write $F_1 \prec F_2$ if F_1 is the restriction of F_2 on a k_1-dimensional subspace of $\mathbb{R}^{k_2}$. Obviously, for any pair of nice variations F_1, F_2 there exists a nice variation F such that $F_1 \prec F$ and $F_2 \prec F$. In other words, the partial ordering $\prec$ makes the set of nice variations into a directed set. We denote the directed set of nice variations by $\mathcal{V}$ and use the symbol $\mathcal{V} - \lim$ for the limits of generalized sequences indexed by the elements of the directed set $\mathcal{V}$.

To any nice variation $F = (f, \phi)$ and any $t \in [t_1, t_2)$ we associate the germ of a mapping $\mathcal{F}_t : \mathbb{R}^k \to M$ defined as follows: $\mathcal{F}_t(x)$ equals the value $y(t_1)$ of the solution of the Cauchy problem

$$\dot{y}(\tau) = X_{y(\tau)}(f(x)(\int\limits_{t_1}^{\tau} \phi(\theta)\, d\theta)),\ t_1 \le \tau \le t, \quad y(t) = z(t),$$

where $\dot{z}(t) = X_{z(t)}(u(t)),\ t_1 \le t \le t_2,\ z(t_1) = z_1$. In particular, $\mathcal{F}_t(0) = z_1,\ t_1 \le t < t_2$.

Assume that $u(\cdot)$ satisfies the maximum principle and $\xi(t) \in T^*_{z(t)}M$, $t_1 \le t \le t_2$, is the corresponding extremal. Then $\xi(t_1) D_0 \mathcal{F}_t = 0,\ t_1 \le t < t_2$, where the product $\xi(t_1) D_0 \mathcal{F}_t$ is the composition of linear mappings $D_0\mathcal{F}_t : \mathbb{R}^k \to T_{z_1}M$ and $\xi(t_1) : T_{z_1}M \to \mathbb{R}$. In other words, 0 is a critical point of $\mathcal{F}_t$ and $\xi(t_1)$ is the corresponding Lagrange multiplier. An important numerical invariant of the reference control $u(\cdot)$ is the

$$\operatorname{corank}(u) = \min\{\operatorname{codim} \operatorname{im} D_0\mathcal{F}_t : F \in \mathcal{V},\ t \in [t_0, t_1)\}.$$

Let $\mathcal{L}(\mathcal{F}_t)$ be the $\mathcal{L}$-derivative of $\mathcal{F}_t$ at $(\xi(t_1), 0)$. Recall that the $\mathcal{L}$-derivative $\mathcal{L}(\mathcal{F}_t)$ is a Lagrangian subspace of the symplectic space $T_{z(t_1)}(T^*M)$ obtained by the linearization of the equation $\lambda D_x \mathcal{F}_t = 0$ (with unknown $\lambda \in T^*M$ and parameter $x \in \mathbb{R}^k$) at the point $\lambda = \xi(t_1),\ x = 0$; see [5] for the precise definition. All Lagrangian subspaces of a symplectic space Σ form the *Lagrange Grassmannian* $L(\Sigma)$, a compact $\dim \Sigma(\dim \Sigma + 1)/8$-dimensional manifold. In particular, $\mathcal{L}(\mathcal{F}_t) \in L\left(T_{\xi(t_1)}(T^*M)\right)$.

We say that the extremal $\zeta(\cdot)$ admits the Jacobi curve if for any $t \in [t_1, t_2)$ there exists the limit,

$$J_\xi(t) = \mathcal{V} - \lim_{F \in \mathcal{V}} \mathcal{L}(\mathcal{F}_t).$$

The curve $t \mapsto J_\xi(t)$ in the Lagrange Grassmannian $L\left(T_{\xi(t_1)}\left(T^*M\right)\right)$ is called the *Jacobi curve associated to the extremal* $\xi(t),\ t_1 \le t \le t_2$. The symplectic space $T_{\xi(t_1)}\left(T^*M\right)$ contains also a fixed "vertical" Lagrangian subspace $\Pi_\xi \stackrel{def}{=} T_{\xi(t_1)}\left(T^*_{z_1}M\right)$.

Our next goal is to define the index of a (not necessarily continuous) curve $\Lambda(t),\ t_0 \le t < t_2$, in the Lagrange Grassmannian $L(\Sigma)$ with respect to a fixed Lagrangian subspace $\Pi \in L(\Sigma)$. First we define the index $\mathrm{ind}_\Pi(\Lambda_0, \Lambda_1)$ of the pair $\Lambda_0, \Lambda_1 \in L(\Sigma)$ with respect to Π.

Let $\eta \in (\Lambda_0 + \Lambda_1) \cap \Pi$ so that $\eta = \eta_0 + \eta_1$, where $\eta_i \in \Lambda_i,\ i = 0, 1$. We set $\mathfrak{q}(\eta) = \sigma(\eta_0, \eta_1)$, where σ is the symplectic form of Σ. If $\Lambda_0 \cap \Lambda_1 = 0$, then $\Lambda_0 + \Lambda_1 = \Sigma$, η is any element of Π and η_0, η_1 are uniquely determined by η. This is not true if $\Lambda_0 \cap \Lambda_1 \neq 0$ but $\mathfrak{q}(x)$ is well-defined anyway: $\sigma(\eta_1, \eta_2)$ depends only on $\eta_0 + \eta_1$ since σ vanishes on $\Lambda_i,\ i = 0, 1$.

Thus $\mathfrak{q}$ is a real quadratic form on $(\Lambda_0 + \Lambda_1) \cap \Pi$; we denote by $\mathrm{ind}\,\mathfrak{q}$ its negative inertia index (the number of negative elements in any diagonalization of the quadratic form $\mathfrak{q}$). We set:

$$\mathrm{ind}_\Pi(\Lambda_0, \Lambda_1) = \mathrm{ind}\,\mathfrak{q} + 1/2(\dim(\Pi \cap \Lambda_0) + \dim(\Pi \cap \Lambda_1)) - \dim(\Pi \cap \Lambda_0 \cap \Lambda_1).$$

An important property of this index is the triangle inequality:

$$\mathrm{ind}_\Pi(\Lambda_0, \Lambda_2) \le \mathrm{ind}_\Pi(\Lambda_0, \Lambda_1) + \mathrm{ind}_\Pi(\Lambda_1, \Lambda_2).$$

Now take a curve $\Lambda(\cdot)$ (an arbitrary mapping from $[t_0, t_1]$ into $L(\Sigma)$). For any finite subset $\mathcal{T} = \{\tau_1, \ldots, \tau_k\} \subset [t_0, t_1)$, where $t_0 = \tau_0 < \tau_1 < \cdots < \tau_l < t_1$, we compute the sum $I_\Pi^{\mathcal{T}} = \sum_{i=1}^{l} \mathrm{ind}_\Pi(\Lambda(\tau_{i-1}), \Lambda(\tau_i))$ and set $I_\Pi(\Lambda(\cdot)) = \sup_{\mathcal{T}} I_\Pi^{\mathcal{T}}$, the supremum of these sums for all finite subsets $\mathcal{T}$. The quantity $I_\Pi(\Lambda(\cdot))$ is called the index of the curve $\Lambda(\cdot)$ with respect to Π, this is a nonnegative half-integral number or $+\infty$.

It can be shown that curves with finite I_Π enjoy analytic properties similar to monotone real functions: they have no more than a countable number of discontinuity points, right and left limits at every point, and are differentiable almost everywhere. Index I_Π is closely related to the Maslov index of continuous curves in the Lagrange Grassmannian: see [4, 7] for the explanation of this fact and of the similarity between finite index curves and monotone functions.

Theorem 4.1 *Assume that a pair* $u(t), z(t),\ t_1 \le t \le t_2,\ z(t_1) = z_1$ *is a solution of the time-optimal problem defined by the control system* $X(u),\ u \in U$. *Then there exists an extremal* $\xi(t),\ t_1 \le t \le t_2$, *which satisfies conditions*

(3.1)–(3.3), admits the Jacobi curve $J_\xi(t)$, $t_1 \le t < t_2$, and, moreover, the following inequality holds:

$$I_{\Pi_\xi}(J_\xi(\cdot)) + \dim \bigcap_{t_1 \le t < t_2} J_\xi(t) - \frac{1}{2}\dim(J_\xi(t_2 - 0) \cap \Pi_\xi) < \frac{n}{2} + \operatorname{corank}(u). \quad (4.1)$$

It should be pointed out that already the existence of the finite index Jacobi curve implies all known pointwise optimality conditions for singular extremals while the inequality (4.1) implies conditions involving conjugate points (for all types of extremals).

References

[1] Boltyansky V.G., Gamkrelidze R.V., Pontryagin L.S., *On the Theory of Optimal Processes,* DAN Sov. Ac., **110**, No 1, (1956), 7–10 (Russian)

[2] Boltyansky V.G., *Maximum Principle in the Theory of Optimal Processes,* DAN Sov. Ac., **119**, No 6, (1958), 1070–1073 (Russian)

[3] R. Gamkrelidze, *Principles of Optimal Control Theory*, Plenum Press, 1978

[4] A. A. Agrachev, R. V. Gamkrelidze, *Symplectic geometry and necessary conditions for optimality.* Matem. Sbornik, 1991, v.182 (in Russian). English transl.: Math. USSR Sbornik, 1992, v.72, 29–45

[5] A. A. Agrachev, R. V. Gamkrelidze, *Feedback–invariant optimal control theory and differential geometry, I. Regular extremals.* J. Dynamical and Control Systems, 1997, v.3, 343–389

[6] A. A. Agrachev, *Feedback–invariant optimal control theory and differential geometry, II. Jacobi curves for singular extremals.* J. Dynamical and Control Systems, 1998, v.4, 583–604

[7] A. A. Agrachev, *Geometry of optimal control problems and Hamiltonian systems.* Lect. Notes in Math. Springer Verlag, to appear