OPTIMAL TRAJECTORIES AND THE NEOCLASSICAL CALCULUS OF VARIATIONS

R. Tyrrell Rockafellar[1]

Abstract

The calculus of variations, a venerable subject connected for centuries with physics, was largely superseded in the 1960s by the theory of optimal control that arose then from modern engineering. Mathematically, the new subject had in fact been anticipated mostly by the old, but its different applications put much greater emphasis on inequality constraints tied to system dynamics. Inequality constraints had become prominent also, already in the 1950s, in finite-dimensional optimization and the solution methods for that which arrived with the computer revolution.

These developments created a need for convex analysis and eventually its nonconvex extensions, which pushed far beyond the traditions of calculus with its reliance on "smoothness." That, in turn, opened a new outlook on the calculus of variations and optimal control with the realization that the problems there could be returned to neoclassical simplicity by posing them with functions allowed to take on ∞ and appealing to subgradients instead of gradients. Rules for determining subgradients could routinely then handle the many details of structure that problem formulation might be asked to address. Here, that history is reviewed, its accomplishments are put in perspective, and prospects for the future suggested.

Keywords: *variational analysis, calculus of variations, optimal control, existence of optimal trajectories, Euler-Lagrange conditions, Hamiltonian dynamics, dual problems, Hamilton-Jacobi-Bellman equations*

Version of 21 December 2024

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: sites.math.washington.edu/∼rtr/mypage.html

1 Overview and dedication

This paper has been put together as a tribute to Boris S. Mordukhovich for his 75th birthday in appreciation of all that he has done, over decades, to expand the horizons of variational analysis. The topic is fitting, because he has often declared that the neoclassical approach to the calculus of variations and optimal control that began around 1970, relying first on methods of convex analysis and later evolving with broader concepts of subgradients, was a key motivation for his early career. The theory of generalized problems of Bolza, with its duality under full convexity, inspired him in particular with its comprehensive novelty and mathematical beauty. That theory is laid out here with indications of historical origins and subsequent advances in nonconvex trajectory optimization in the same vein. Its echoes for the Hamilton-Jacobi-Bellman equation are explained, as well.

Nowadays, fashion has largely switched to other things, and research in the calculus of variations and optimal control has entered a kind of backwater. But perhaps this paper can help to keep it afloat. Before the engineering applications of control, the classical subject had crucial significance for the physics of mechanical systems, and still does. Its neoclassical reincarnation, so readily able to handle "nonsmoothness," offers prospects of extending those applications into wider practical domains. Moreover, extensions to the PDE-type variational settings that deal with elasticity and stress, for instance, yet may be rife with nonsmoothness, have hardly been touched. So much more is on the mathematical horizon, just waiting. Many more tools in second-order variational analysis are now available, among other things. Hopefully, the review provided here might aid in sparking new research.

The brachistochrone problem of Johann Bernouli in 1696 was perhaps to first to look for an "optimal trajectory." If a ball is to roll by gravity down a curved ramp from a point A to a lower point B, what shape should the curve have for the time taken to be minimal? That was the start of the classical field known somewhat oddly as the *calculus of variations*, because of its central technique of testing optimality by "varying" a solution candidate in one way or another.

A trajectory can be imagined for general purposes as a function from time t in an interval $[t_0, t_1]$ to points $x(t) \in I\!\!R^n$ that represent "states" of a system. In Newtonian notation for rates of change, the first derivative is denoted by $\dot{x}(t)$. An optimization problem can be posed by specifying constraints on the endpoints $x(t_0)$ and $x(t_1)$ along with a "cost" that integrates over time in depending on the states $x(t)$ and velocities $\dot{x}(t)$ as t goes from t_0 to t_1. Questions concern the existence of a trajectory that minimizes the cost subject to the constraints and how to characterize the optimality in the global sense or some local sense.

Such questions turn out to be much harder to answer than might be supposed. A major reason is that the problem is infinite-dimensional and can't even be made rigorous without restricting the trajectories to some good "trajectory space," for which the development of modern functional analysis was essential. It's relatively easy to establish conditions that are necessary for simple versions of local optimality when assuming adequate levels of differentiability, but there's a pitfall. Such conditions can identify a unique trajectory that exhibits such differentiability, and yet that trajectory can fail to be optimal, because lower values of the "cost" can be obtained by a sequence tending, in some respects, toward a trajectory that lacks the specified differentiability. To get around that, bigger spaces of trajectories need to be considered and different notions of convergence have to be sorted out. To handle "costs" expressed by integral functionals, advanced integration theory with its concept of measurability must be brought in. Getting all that in place took many years of effort and a change from looser old-time modes of mathematical analysis to the standards of today. One of the great virtues of the calculus of variations in its history has been the impetus it gave for so much of that.

The basic problem in which the "cost" of a trajectory depends on $x(t)$ and $\dot{x}(t)$ while the interval $[t_0, t_1]$ is fixed can be elaborated, of course, in many ways. The cost could also depend on second derivatives $\ddot{x}(t)$, but tricks can be used to reduce that to just first derivatives, much as in the methodology of ordinary differential equations. The terminal time t_1 can become a variable subject to optimization, too, as in the brachistochrone problem in its search for something quickest. For that, there are again reformulation tricks to reduce back to fixed t_1. More demanding, though, are constraints on $x(t)$ and $\dot{x}(t)$ that operate along the entire trajectory instead of just at the endpoints, or relaxations from fixed endpoints to more complicated conditions the may combine with endpoint costs, perhaps as penalties. Moreover the constraints can be in the form of inequalities, not merely equations. All those complications were nevertheless mastered in mathematical theory, at least to classical satisfaction, as the calculus of variations matured in the middle of the last century and the computer revolution loomed. The important questions seemed to have been answered, and researchers were turning to other endeavors.

Computers brought the birth of modern optimization in the United States in the form of the Dantzig's simplex method for linear programming (1949) and so much that followed from it, with numerical emphasis in finite dimensions in exploring convexity, duality, and the complications caused by inequality constraints. In the Soviet Union, in contrast, optimal control of trajectories was the new topic. It surged with the Pontriagin "maximum principle" being its centerpiece as a necessary condition for local optimality. In fact, most of what became important to understand at that time about inequality constraints or optimal trajectories in control had already been worked out in the "finalized" classical theory. Hestenes, who had been a participant, explained that well in his 1966 book [12]. But a huge shift had arrived in the research paradigm and the vision of applications.

2 Classical framework

In classical analysis, functions are usually assumed, or expected, to be arbitrarily smooth: continuously differentiable as many times as might be helpful. The bedrock problem for trajectories in $\mathbb{R}^n$ is the *problem of Lagrange*. It seeks to

$$\text{minimize} \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt \text{ subject to } x(t_0) = a_0,\ x(t_1) = a_1, \tag{2.1}$$

for choice of points a_0 and a_1 in $\mathbb{R}^n$ and a differentiable function L of $(t, x, v) \in [t_0, t_1] \times \mathbb{R}^n \times \mathbb{R}^n$, which is called the *Lagrangian*. Here a conflict in notation needs to be avoided. The x in $L(t, x, v)$ stands for a vector, so it can't also stand for the trajectory in (2.1). That's denoted therefore instead by $x(\cdot)$. The problem then is to minimize an *integral functional* on a space of trajectories[2] $x(\cdot)$.

But what is that trajectory space? And what might be the distances used in it when designating a *local* minimum? Classically, in taking trajectories $x(\cdot)$ to be continuously differentiable on (t_0, t_1) with the derivative $\dot{x}(\cdot)$ having a continuous extension to $[t_0, t_1]$, one can identify $x(\cdot)$ with the pair $(x(t_0), \dot{x}(\cdot)) \in \mathbb{R}^n \times \mathcal{C}_n[t_0, t_1]$, where $\mathcal{C}_n[t_0, t_1]$ is the Banach space of continuous functions from $[t_0, t_1]$ to $\mathbb{R}^n$ under the uniform norm $||\cdot||_\infty$. A neighborhood of $x(\cdot)$ can be built from a neighborhood of $x(t_0)$ in $\mathbb{R}^n$ and a neighborhood of $\dot{x}(\cdot)$ in $\mathcal{C}_n[t_0, t_1]$. A *weak* local minimum refers traditionally to a neighborhood in that sense, while a *strong* local minimum refers to a neighborhood of $x(\cdot)$ as itself an element of $\mathcal{C}_n[t_0, t_1]$, without invoking closeness of derivatives.

The standard first-order *necessary* condition for a weak local minimum is the *Euler-Lagrange*

[2] Although the term "trajectory" is preferred here, much of the literature speaks of $x(\cdot)$ as an "arc."

equation, which has often been written in the potentially puzzling form

$$\frac{d}{dt}\nabla_v L(t,x(t),\dot{x}(t)) = \nabla_x L(t,x(t),\dot{x}(t)), \tag{2.2}$$

as for instance in the textbook [11]. This involves the partial gradients of the function $L(t,x,v)$ in the x and v arguments and the implicit claim that the composed function $t \mapsto \nabla_v L(t,x(t),\dot{x}(t))$ is differentiable for the candidate trajectory $x(\cdot)$. A better statement of the condition is that there exists a trajectory $p(\cdot)$ such that

$$(\dot{p}(t),p(t)) = \nabla_{x,v} L(t,x(t),\dot{x}(t)). \tag{2.3}$$

This is important because it opens the way to *duality*, with $p(\cdot)$ being called the *adjoint* trajectory associated with $x(\cdot)$ in local optimality. With "full convexity," $p(\cdot)$ will be seen later to solve a problem that's dual to the one solved by $x(\cdot)$. That's noteworthy as an entirely natural development from the perspective of modern optimization, but for which there was no inkling in classical imagination.

Next comes an operation called the *Legendre transform*, which aims to produce from $L(t,x,v)$ a function $H(t,x,p)$ on $[t_0,t_1]\times I\!R^n\times I\!R^n$ called the *Hamiltonian*. The usual description of it is troublesome for the absence of clear assumptions, without which it won't work. The prescription is first to solve the equation $p = \nabla_v L(t,x,v)$ for v as a function of (t,x,p), with $v = V(t,x,p)$ say, and then to set

$$H(t,x,p) = p{\cdot}V(t,x,p) - H(t,x,V(t,x,p)). \tag{2.4}$$

The standard implicit function theorem is evidently the tool for obtaining $V(t,x,p)$, however that's a local thing. Here it seems to be invoked globally — to get H defined everywhere and moreover itself continuously differentiable to whatever degree. Or if H isn't defined everywhere, what might be its domain? The classical literature is silent, but the assumptions needed were finally worked out in [24]. They are much more restrictive and convexity-dependent than textbooks ever indicate. Anyway, the powerful observation is that Euler-Lagrange equation (2.3) can in this way be written equivalently as an ordinary differential equation in the pair $(x(\cdot),p(\cdot))$, which is called the *Hamiltonian* equation:

$$\dot{x}(t) = \nabla_p H(t,x(t),p(t)), \qquad \dot{p}(t) = -\nabla_x H(t,x(t),p(t)). \tag{2.5}$$

In another technical reformulation with valuable consequences, the time t can be recast as another state variable. Details aside, this leads to a companion relation to (2.5) which comes out as

$$\frac{d}{dt}H(t,x(t),p(t)) = \nabla_t H(t,x(t),p(t)). \tag{2.6}$$

A conclusion is that, if $L(t,x,v)$ is just $L(x,v)$, the so-called *autonomous* case, where also $H(t,x,p)$ is just $H(x,p)$, then $H(x(t),p(t))$ must be constant in time. Such properties are famous in physics in application to the dynamics of mechanical systems. There, $p(t)$ is the momentum associated with the velocity $\dot{x}(t)$, and the Hamiltonian represents energy. Its constancy along the trajectory pair $(x(t),p(t))$ is a law of conservation, and there is much more along those lines in connection with the Hamiltonian.

In the case of a strong local minimum, the Euler-Lagrange equation is partnered in necessity with the *Weierstrass* condition, under which the inequality

$$L(t,x(t),v) \geq L(t,x(t),\dot{x}(t)) + p(t){\cdot}[v-\dot{x}(t)] \ \text{ for } \ p(t) = \nabla_v L(t,x(t),\dot{x}(t)) \tag{2.7}$$

must hold for all v. To contemporary eyes, this is clearly a *convexity* type of the property of the function $L(t,x(t),\cdot)$ at the point $\dot{x}(t)$. But convexity was not something perceived in mathematical analysis of the past as a topic worthy of investigation in its own right.

The problem of Lagrange has the trajectory endpoints fixed, but what if they are free, subject to an additional cost tied to their location? The task then is to

$$\text{minimize} \quad \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt + l(x(t_0), x(t_1)) \tag{2.8}$$

for some function l on $I\!R^n \times I\!R^n$. The adjustment needed in characterizing local optimality is to combine the Euler-Lagrange equation with the *transversality* condition

$$(p(t_0), -p(t_1)) = \nabla l(x(t_0), x(t_1)). \tag{2.9}$$

Free endpoints and fixed endpoints are the two extremes, but there can be so much in between. One or the other endpoint could be fixed while the other is free, perhaps with an attached cost. Or parts of each could be fixed while other parts have costs. Besides constraints on the components of $(x(t_0), x(t_1))$, there could be constraints at each time t on the components of $(x(t), \dot{x}(t))$. Altogether, we can contemplate trying to

$$\begin{array}{rl} \text{minimize} & \int_0^T L_0(t, x(t), \dot{x}(t))dt + l_0(x(t_0), x(t_1)) \ \text{ subject to} \\ & L_i(t, x(t), \dot{x}(t)) \begin{cases} \le 0 \ \text{ for } \ i = 1, \ldots, s, \\ = 0 \ \text{ for } \ i = s+1, \ldots, m, \end{cases} \\ & l_j(x(t_0), x(t_1)) \begin{cases} \le 0 \ \text{ for } \ j = 1, \ldots, q, \\ = 0 \ \text{ for } \ j = q+1, \ldots, r. \end{cases} \end{array} \tag{2.10}$$

This broad challenge, covering a myriad of special cases, is a *problem of Bolza*.[3] Of course, a function L_i may depend only on part of the velocity, or part of the state, and a function l_j may involve just a particular aspect of one of the endpoints. In adapting optimality conditions to all these constraints, Lagrange multipliers have to be brought in, and updated methodology was needed for that.

Problems of *optimal control* look different from this, at least on the surface, because they center on a the choice of a control function $u(\cdot)$ from $[t_0, t_1]$ to a set $U(t) \subset I\!R^d$ which determines a trajectory $x(\cdot)$ through an ordinary differential equation (ODE):

$$\begin{array}{rl} \text{minimize} & \int_{t_0}^{t_1} f_0(t, x(t), u(t))dt + l_0(x(t_0), x(t_1)) \ \text{ subject to} \\ & \dot{x}(t) = f(t, x(t), u(t)), \quad u(t) \in U(t), \\ & l_j(x(t_0), x(t_1)) \begin{cases} \le 0 \ \text{ for } \ j = 1, \ldots, q, \\ = 0 \ \text{ for } \ j = q+1, \ldots, r. \end{cases} \end{array} \tag{2.11}$$

The idea behind "control" is that fixing $u(\cdot)$ produces a mapping $F(t, x) = f(t, x, u(t))$, and in that way determines $x(\cdot)$ from the ODE $\dot{x}(t) = F(t, x(t))$ and knowledge of $x(t_0)$ or $x(t_1)$. The endpoint constraints in (2.11) are written very generally, but in particular they could require $x(t_0)$ to be some a_0 or $x(t_1)$ to be some a_1. Conditions on f and $u(\cdot)$ must then be imposed to ensure the properties of F needed for getting $x(\cdot)$ uniquely as a trajectory over the entire interval $[t_0, t_1]$.

But there is an alternative way of looking at (2.11) which provides important insights. Instead of it revolving around a choice of $u(\cdot)$ that acts to generate a trajectory $x(\cdot)$, the problem can be understood in terms of optimizing the choice of both $x(\cdot)$ and $u(\cdot)$ subject to the differential equation

[3] Bolza problems, as recalled in the 1966 textbook of Hestenes [12], can also involve constraints on additional expressions of the kind being minimized in (2.10). However, such problems can be reformulated to the statement here through the incorporation of more state variables.

as a constraint on that joint choice.[4] Moreover, $u(\cdot)$ can harmlessly be posed as the derivative $\dot{y}(\cdot)$ of a trajectory $y(\cdot)$ in $I\!R^d$. That way, (2.11) seeks to

$$\begin{aligned}&\text{minimize} \;\; \textstyle\int_{t_0}^{t_1} f_0(t,x(t),\dot{y}(t))dt + l_0(x(t_0),x(t_1)) \;\; \text{subject to}\\ &\qquad \dot{x}(t) - f(t,x(t),\dot{y}(t)) = 0, \quad \dot{y}(t) \in U(t),\\ &\qquad l_j(x(t_0),x(t_1)) \begin{cases} \leq 0 & \text{for } j=1,\dots,q,\\ =0 & \text{for } j=q+1,\dots,r.\end{cases}\end{aligned} \tag{2.12}$$

When supplemented by a representation of the set $U(t)$ by a system of equations and/or inequalities, (2.12) can be reconstituted as a problem of Bolza with respect to the trajectory $(x(\cdot),y(\cdot))$ in $I\!R^{n+d}$:

$$\begin{aligned}&\text{minimize} \;\; \textstyle\int_{t_0}^{t_1} \widetilde{L}_0(t,x(t),y(t),\dot{x}(t),\dot{y}(t))dt + \widetilde{l}_0(x(t_0),y(t_0),x(t_1),y(t_1)) \;\; \text{subject to}\\ &\qquad \widetilde{L}_i(t,x(t),y(t),\dot{x}(t),\dot{y}(t)) \begin{cases} \leq 0 & \text{for } i=1,\dots,\tilde{s},\\ =0 & \text{for } i=\tilde{s}+1,\dots,\tilde{m},\end{cases}\\ &\qquad \widetilde{l}_j(x(t_0),y(t_0),x(t_1),y(t_1)) \begin{cases} \leq 0 & \text{for } j=1,\dots,\tilde{q},\\ =0 & \text{for } j=\tilde{q}+1,\dots,\tilde{r}.\end{cases}\end{aligned} \tag{2.13}$$

It's from this angle that optimal control theory can be seen as not so different from the classical calculus of variations that preceded it, with the Pontriagin *maximum principle* essentially corresponding to the Weierstrass necessary condition — a point of view emphasized in the book of Hestenes [12]. But optimal control has a different engineering-oriented emphasis that leads to different technical demands.

Simple examples in control indicate that optimality may be out of reach if only continuous functions $u(\cdot)$ are admitted. So-called bang-bang control functions that are piecewise constant, or at least able to make jumps at certain times, may be natural in some applications. Continuous differentiability of trajectories $x(\cdot)$ can't then be taken for granted. But concepts like merely piecewise continuity of derivatives, already contemplated in the calculus of variations and tied there to the *Erdmann* conditions in necessity, aren't adequate either. They can fail to be preserved in taking limits, that makes them unsuitable for establishing the existence of solutions, among other things.

Perhaps, instead of having $\dot{x}(\cdot)$ in the space $\mathcal{C}_n[t_0,t_1]$, it should be allowed to be in a space $\mathcal{L}^p_n[t_0,t_1]$? The most attractive in this direction as trajectory space is

$$\mathcal{A}^1_n[t_0,t_1] = \left\{ x(\cdot) \in \mathcal{C}_n[t_0,t_1] \,\middle|\, \dot{x}(\cdot) \in \mathcal{L}^1_n[t_0,t_1] \right\}, \tag{2.14}$$

which consists of the *absolutely continuous* functions from $[t_0,t_1]$ to $I\!R^n$, characterized by being expressible as $x(t) = x_0 + \int_{t_0}^t v(s)ds$ for some $x_0 \in I\!R^n$ and $v(\cdot) \in \mathcal{L}^1_n[t_0,t_1]$ (in which case $x(t_0) = x_0$ and $\dot{x}(t) = v(t)$ almost everywhere). It's a Banach space with respect to the norm $||x(\cdot)|| = |x(t_0)| + ||\dot{x}(\cdot)||_1$. (For vectors in $I\!R^n$, $|\cdot|$ stands for the canonical norm there.)

3 Neoclassical framework

The development of convex analysis brought many innovations. One of the most striking was allowing a function f on $I\!R^n$ to be extended-real-valued and identifying it then with its epigraph in $I\!R^n \times I\!R$ instead of its graph in $I\!R^n \times [-\infty,\infty]$. That way, constraints could be represented implicitly; minimizing $f(x)$ over $x \in I\!R^n$ is the same as minimizing it over dom f, the set of points x where $f(x) < \infty$. Furthermore, gradients could be replaced by subgradients with their useful calculus.

[4] This could be useful, for example, in an iterative scheme of approximation or computation in which the control function and state trajectory only satisfy the ODE in the limit.

Might the same approach be interesting in the optimization of trajectories? Why not work with a *generalized problem of Bolza*, where the goal is to

$$\text{minimize } \mathcal{J}_{L,l}(x(\cdot)) := \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt + l(x(t_0), x(t_1)) \tag{3.1}$$

with respect to trajectories $x(\cdot) \in \mathcal{A}^1_n[t_0, t_1]$, and both L and l can be *extended-real-valued*? With the understanding (in the technical framework to come) that $\mathcal{J}_{L,l}(x(\cdot)) < \infty$ entails

$$(x(t), \dot{x}(t)) \in \operatorname{dom} L(t, \cdot, \cdot), \qquad (x(t_0), x(t_1)) \in \operatorname{dom} l, \tag{3.2}$$

the classical problem of Bolza in (2.10) would correspond to taking L here to be the sum of L_0 and the indicator of specified constraints on time, state and velocity, and likewise for l and the constraints on endpoints. Maybe subgradient versions of the Euler-Lagrange equation (2.3) and the transversality condition (2.9) could serve then in capturing optimality?

This innocent tactic reduces even the most complicated classical problems to a simple-looking *neo-classical* formulation and at the same time offers fresh perspectives in optimal control. And there's no loss in such notational simplification — from the prospective of subgradient calculus and its Lagrange multiplier component being able in the end to handle the structure behind ∞ values of L and l.

But a serious technical issue has to be confronted right away. Is the functional in (3.1) well defined on $\mathcal{A}^1_n[t_0, t_1]$? The (Lebesgue) measurability of the function $t \mapsto L(t, x(t), \dot{x}(t))$ is required, along with assurance that the integral has an unambiguous value in $(-\infty, \infty]$. The theory of "normal integrands" was developed in [25] (convexity-based) and then more broadly in [35] and [41, Chapter 14] in order to help with this. It relies on the available results about measurable selections from set-valued mappings and the notion that a mapping from $[t_0, t_1]$ to closed subsets $C(t)$ of a space $\mathbb{R}^d$, say, is *measurable* if, for every open set $O \subset \mathbb{R}^d$, the set $\{\, t \,|\, C(t) \cap O \neq \emptyset\}$ is measurable in $[t_0, t_1]$. (This is equivalent to many other useful properties of the mapping.)

A *normal integrand* on $[t_0, t_1] \times \mathbb{R}^d$ is ultimately a function $f(t, u)$ such that the epigraphical mapping $t \mapsto \operatorname{epi} f(t, \cdot) \subset \mathbb{R}^{d+1}$ is closed-valued and measurable, with the closed-valuedness corresponding of course to the lower semicontinuity of $f(t, \cdot)$ on $\mathbb{R}^d$ for each t. A key consequence is indeed the measurability of the function $t \mapsto f(t, u(t))$ for any measurable function $t \mapsto u(t)$. Then, as long as $f(t, u(t))$ is bounded from below by some measurable function $\beta(t)$ that is finitely integrable in the Lebesgue sense, the integral $\int_{t_0}^{t_1} f(t, u(t))dt$ will unambiguously have a value that is either finite or ∞ (not $-\infty$). That will here be termed the *integrability* of $f(\cdot, u(\cdot))$, with finiteness of the value in question called *summability*, in contrast, and entailing that $f(t, u(t)) < \infty$ for a.e. t.

Accordingly, the functional $\mathcal{J}_{L,l}$ in problem (3.1) will be well defined on the space $\mathcal{A}^1_n[t_0, t_1]$ under the assumption henceforth made that

$$\begin{aligned}&l \text{ is lower semcontinous proper on } \mathbb{R}^n \times \mathbb{R}^n, \text{ while } L \text{ is a normal integrand}\\ &\text{satisfying } L(t, x, v) \geq \beta(t) - \gamma(t)(|x| + |v|) \text{ for finitely integrable } \beta \text{ and } \gamma,\end{aligned} \tag{3.3}$$

which moreover will guarantee that (3.2) holds in the almost everywhere sense. Implicit in (3.1) will therefore be the underlying *differential inclusion*

$$\dot{x}(t) \in D(t, x(t)) \ \text{ a.e., where } \ D(t, x) = \operatorname{dom} L(t, x, \cdot), \tag{3.4}$$

and the *state constraint*

$$x(t) \in X(t) \ \text{ a.e., where } \ X(t) = \{\, x \in \mathbb{R}^n \,|\, D(t, x) \neq \emptyset\}. \tag{3.5}$$

These implicit restrictions underscore the great breadth of the problem model, which will come into even brighter light when control ideas return to the discussion. In application to (2.10) for the construction of the corresponding L in (3.1), one can take advantage for instance of the fact that L will be a normal integrand if every L_i is a *Carathéodory* integrand, which means that $L_j(t,x,v)$ is finite, continuous in (x,v) for each t, and measurable in t for each (x,v).

In this repainted picture, what might be the Hamiltonian $H(t,x,p)$ to associate with the Lagrangian $L(t,x,v)$? The usual Legendre transform, invoking the implicit function theorem in solving a gradient equation, is unavailable, but convex analysis offers a substitute that is anyway much better: the *Legendre-Fenchel* transform that's behind conjugate convex functions. Using this, define

$$H(t,x,p) := \sup_v \{\, p{\cdot}v - L(t,x,v)\}, \tag{3.6}$$

noting that then $H(t,x,p)$ is lower semicontinuous convex in p with

$$\begin{aligned} H(t,x,p) > -\infty \ \text{ for all } p \text{ when } x \in X(t), \\ H(t,x,p) = -\infty \ \text{ for all } p \text{ when } x \notin X(t), \end{aligned} \tag{3.7}$$

for the state constraint set $X(t)$ in (3.5), and furthermore

$$L(t,x,v) \text{ convex in } v \implies \sup_p \{\, p{\cdot}v - H(t,x,p)\} = L(t,x,v). \tag{3.8}$$

In presuming to define the Hamiltonian function in the classical theory by way of the Legendre transform, a strong form of convexity of $L(t,x,v)$ in v is taken for granted, perhaps without realizing it; see [24]. For (t,x) such that $L(t,x,v)$ isn't convex in v, but $H(t,x,\cdot)$ isn't the constant function ∞, the right side of the implication (3.8) has to be refined to

$$\begin{aligned} &\sup\nolimits_p \{\, p{\cdot}v - H(t,x,p)\} = \overline{L}(t,x,v), \ \text{ where, for each } (t,x), \\ &\overline{L}(t,x,\cdot) \text{ is the closed convex hull of the function } L(t,x,\cdot). \end{aligned} \tag{3.9}$$

That may seem esoteric, but in fact the classical Weierstrass condition in (2.7), which comes into play for a strong local minimum, effectively demands that $L(t,x(t),\dot{x}(t)) = \overline{L}(t,x(t),\dot{x}(t))$ with $p(t)$ being a subgradient of $\overline{L}(t,x(t),\cdot)$ at $\dot{x}(t)$ in the convex analysis sense. Something really fundamental is thus at stake for trajectories in the convex hull operation in (3.9).

The convexity of $L(t,x,v)$ in v not only obviates any need for taking a convex hull, but also in fact is generally prerequisite to hopes of obtaining the existence of a minimizing trajectory $x(\cdot)$. Without it, the convexification procedure enters as a key form of problem "relaxation."

The Hamiltonian function itself has a major role in existence theory as well by being subjected to the *basic Hamiltonian growth* condition that

$$H(t,x,p) \le \varphi(t,|x|,p), \ \text{ where } \varphi(t,r,p) \text{ is finite and summable in } t. \tag{3.10}$$

Theorem 3.1 (semicontinuity and compactness [33])**.** *Suppose $L(t,x,v)$ is convex in v and $H(t,x,p)$ satisfies the basic growth condition (3.10). Then for every $\alpha \in I\!R$ and $r \in (0,\infty)$ the set*

$$\{\, x(\cdot) \,|\, \mathcal{J}_{L,l}(x(\cdot)) \le \alpha, \ ||x(\cdot)||_\infty \le r\} \tag{3.11}$$

is weakly compact in $\mathcal{A}^1_n[t_0,t_1]$ and strongly compact in $\mathcal{C}_n[t_0,t_1]$.

This assures the attainment of the minimum in problem (3.1) relative to the constraint $||x(\cdot)||_\infty \le r$ for any $r \in I\!R_+$ and therefore in the presence of the state constraint $X(t)$ when the sets $X(t)$ are uniformly bounded in $I\!R^n$. The need for such boundedness can be eliminated by appealing instead to the *strong Hamiltonian growth condition* that

$$H(t,x,p) \leq \mu(t,p) + |x|(\sigma(t) + \rho(t)|p|), \text{ where } \\ \mu(t,p),\ \sigma(t),\ \rho(t), \text{ are nonnegative, } t\text{-summable.} \tag{3.12}$$

Theorem 3.2 (existence of optimal trajectories [33]). *Suppose $L(t,x,v)$ is convex in v and $H(t,x,p)$ satisfies the strong growth condition (3.12). Then, as long as $\mathcal{J}_{L,l} \not\equiv \infty$ (feasibility), there exists a trajectory $x(\cdot)$ satisfying the implicit constraints (3.4)–(3.5) that minimizes $\mathcal{J}_{L,l}$ over $\mathcal{A}^1_n[t_0,t_1]$.*

The growth condition in (3.12) can be compared to the classical one of Tonelli on L, which in translation to a equivalent condition on H requires $H(t,x,p) \leq \gamma|x| + \theta(|p|)$ for some $\gamma \geq 0$ and a finite and nondecreasing convex function θ. That's much tighter and excludes applications to duality such as will be undertaken in Section 4.

Optimal control problems as in (2.11) can be turned into problems of Bolza by following the path of reformulation via (2.12) to (2.13), and from there into problems in the generalized Bolza pattern of (3.1). But optimal control can benefit by starting from an even broader problem statement which was first put forward in [33]:

$$\text{minimize } \mathcal{J}_{K,l}(x(\cdot),u(\cdot)) := \int_{t_0}^{t_1} K(t,x(t),u(t),\dot{x}(t))dt + l(x(t_0),x(t_1)), \tag{3.13}$$

with l as before but K being a normal integrand on $[t_0,t_1] \times I\!R^n \times I\!R^d \times I\!R^n$. The minimization in (3.13) is viewed as taking place jointly with respect to $x(\cdot) \in \mathcal{A}^1_n[t_0,t_1]$ and

$$u(\cdot) \in \mathcal{L}_d[t_0,t_1] = \text{the space of all measurable functions from } [t_0,t_1] \text{ to } I\!R^d. \tag{3.14}$$

The implicit constraints, besides $(x(t_0),x(t_1)) \in \operatorname{dom} l$, are that

$$\begin{aligned} &\dot{x}(t) \in F(t,x(t),u(t)) \text{ a.e., where } F(t,x,u) = \operatorname{dom} K(t,x,u,\cdot), \\ &u(t) \in U(t,x(t)) \text{ a.e., where } U(t,x) = \{\, u \mid F(t,x,u) \neq \emptyset\}, \\ &x(t) \in X(t) \text{ a.e., where } X(t) = \{\, x \mid U(t,x) \neq \emptyset\}. \end{aligned} \tag{3.15}$$

Uniform boundedness of the sets $U(t,x)$ would automatically restrict $u(\cdot)$ to the subspace $\mathcal{L}^\infty_d[t_0,t_1]$ of $L_d[t_0,t_1]$. In particular, $U(t,x)$ might be independent of x and maybe also of t.

This generalized control problem can be reduced to a generalized problem of Bolza (3.1) by following the earlier path of introducing an auxiliary state trajectory $y(\cdot)$ with $\dot{y}(t) = u(t)$. That can be good for some purposes, such as fully propagating the "maximum principle" as a necessary condition for control optimality. But there's a potential disadvantage in treating it that way. The resulting Lagrangian L may lack dynamical convexity, despite convexity of the control space $U(t)$, unless the control equation has $f(t,x,u)$ linear in u. An alternative route to reducing (3.13) to a generalized problem of Bolza is opened by defining, as Lagrangian, the function

$$L(t,x,v) = \inf_u K(t,x,u,v), \tag{3.16}$$

which through (3.16) and (3.6) yields, as Hamiltonian, the function

$$H(t,x,p) = \sup_{u,v}\{\, p{\cdot}v - K(t,x,u,v)\}. \tag{3.17}$$

For this purpose, K should satisfy

$$K(t,x,u,v) \geq \beta(t) - \gamma(t)(|x| + |v|) \text{ for finitely integrable } \beta \text{ and } \gamma \tag{3.18}$$

and the *inf-boundedness* condition that

$$\begin{array}{l}\text{for each } \alpha \in I\!R^r,\ t \in [t_0,t_1] \text{ and bounded subset } B \subset I\!R^n \times I\!R^n,\\ \text{the set } \{\, u \mid \exists\, (x,v) \in B : K(t,x,u,v) \leq \alpha\} \text{ is bounded in } I\!R^d.\end{array} \tag{3.19}$$

Theorem 3.3 (reduction to implicit controls [33])**.** *Under (3.18) and the inf-boundedness condition (3.19) on K, the function L defined by (3.16) is a normal integrand on $[t_0,t_1] \times I\!R^d$ satisfying (3.3) and such that*

$$\mathcal{J}_{L,l}(x(\cdot)) = \inf\Big\{\, \mathcal{J}_{K,l}(x(\cdot),u(\cdot)) \,\Big|\, u(\cdot) \in \mathcal{L}_d[t_0,t_1] \,\Big\}, \tag{3.20}$$

with the infimum in $(\cdot)$ being attained as long as it is not ∞. Thus, as long as $\mathcal{J}_{K,l} \not\equiv \infty$ (feasibility), a pair $(x(\cdot),u(\cdot))$ solves problem (3.13) if and only if the trajectory $x(\cdot) \in \mathcal{A}^1_n[t_0,t_1]$ solves the generalized Bolza problem (3.1) and the control function $u(\cdot) \in \mathcal{L}_d[t_0,t_1]$ selects

$$u(t) \in S(t,x(t),\dot{x}(t)) \ \text{ for a.e. } t, \text{ where } \ S(t,x,v) = \operatorname*{argmin}_u K(t,x,u,v), \tag{3.21}$$

those argmin sets being compact in $I\!R^d$ and such that $t \mapsto S(t,x(t),\dot{x}(t))$ is measurable.

This has liberating content. It reveals that, for purposes of ascertaining the existence of solutions to the very broad problem (3.13) and characterizing optimality of $x(\cdot)$ and $u(\cdot)$ there, one can simply call on the existence theory of existence and optimal conditions for trajectories $x(\cdot)$ in the generalized Bolza problem (3.1) that's extracted from (3.13) by (3.16). Afterward, an appeal to (3.21) recovers optimal controls through an arbitrary measurable selection that's guaranteed to be possible. That depiction may seem odd for the control mindset, but it puts the mathematical focus on the features that are most essential in achieving a fundamental understanding the optimization that's involved.

As an example to illustrate reduction by (3.16) and bring insights to the issue of convexity of $L(t,x,v)$ in v, suppose a given control problem is to

$$\begin{array}{l}\text{minimize } \ g(x(t_1)) \ \text{ over the trajectories } x(\cdot) \text{ starting from } x(t_0) = a_0\\ \text{that are generated from } \dot{x}(t) = f(t,x(t),u(t)) \text{ by controls } u(t) \in U(t).\end{array} \tag{3.22}$$

This corresponds to (3.13) with $l(x_0,x_1)$ being the sum of $g(x_1)$ and the indicator $\delta_{a_0}(x_0)$, while $K(t,x,u,v)$ is the indicator of the graph of f plus the indicator of the graph of $t \mapsto U(t)$. Then, through (3.16), $L(t,x,v)$ is the indicator of the graph of the mapping D defined by

$$D(t,x) = \{\, v \mid \exists\, u \in U(t),\ f(t,x,u) = v\}. \tag{3.23}$$

The reduced problem, with controls suppressed, is therefore to

$$\begin{array}{l}\text{minimize } \ g(x(t_1)) \ \text{ over the trajectories } x(\cdot) \text{ starting from } x(t_0) = a_0\\ \text{that can be generated from the differential inclusion } \dot{x}(t) \in D(t,x(t)).\end{array} \tag{3.24}$$

For any such trajectory, controls generating it can be recovered by measurable selection of $u(t)$ from $U(t) \cap \{\, u \mid f(t,x(t),u) = \dot{x}(t)\}$. Here of course the relations are intended to hold only almost everywhere, and the conditions on $U(t)$ and $f(t,x,u)$ have to be such that $K(t,x,u,v)$ is a normal integrand. But that's easy to identify from the measurability handbook in Chapter 14 of [41].

Whether or not $L(t,x,v)$ is convex in v in this example comes down to whether or not $D(t,x)$ is a convex set. That's a crucial matter in the theory of differential inclusions, as seen for instance in the book of Aubin and Cellina [1]. When $D(t,x)$ isn't convex, the relaxed differential inclusion utilizing the closed convex hull of $D(t,x)$ is an important object of study. The Hamiltonian here is

$$H(t,x,p) = \sup\{\, p{\cdot}v \mid v \in D(t,x,v)\}, \tag{3.25}$$

which as a function of p is the support function of the set $D(t, x)$ and its convex hull.

Formulating control problems with control functions $u(\cdot)$ merely taken to be measurable can seem, for engineering, a step too far for applications. Nobody knows how to "implement" controls of such generality. Wouldn't it be more sensible to limit attention to controls that are piecewise linear, say? The trouble with such wishful niceties is that they interfere with optimality. There's little hope of finding conditions on the Lagrangian that ensure the existence of an optimal trajectory with the desired special properties by insisting on them in advance, instead of them naturally and automatically coming out of the situation itself. Moreover, without that existence, a necessary condition for optimalty that simply assumes the desired property is worthless, even when it leads to a single candidate control function of the wished-for variety (this being a pitfall so often overlooked in the literature on the maximum principle in control). Instead, the correct path is to first understand the characteristics that truly must belong to an optimal trajectory, and second to figure out how such a trajectory might be approximated in minimization a by a nicer one when it isn't already nice.

4 Neoclassical optimality conditions under full convexity

The property of *dynamical convexity* of the Lagrangian in the generalized Bolza problem (3.1), referring to convexity of $L(t, x, v)$ in v, has been observed to have a crucial and almost normalizing role in support of lower semicontinuity of the functional being minimized and the existence of trajectories that achieve the minimum. Its hallmark is the uncompromised conjugate duality between L and the Hamiltonian H in the relations (3.6) and (3.8), which signals that every aspect of L is reflected, somehow or other, by some characteristic of H. Relaxation through the convex hull operation can then be left aside.

In contrast, the property of *full convexity* of the Lagrangian, which refers to $L(t, x, v)$ being convex in (x, v), has a more special role, yet nonetheless a fundamental one, particularly for the deployment of convex analysis. Its Hamiltonian counterpart, on top of the convex conjugacy between $L(t, x, \cdot)$ and $H(t, x, \cdot)$ associated with dynamical convexity, is that

$$\text{full convexity of } L \text{ corresponds to } H(t, x, p) \text{ being concave in } x. \tag{4.1}$$

In the generalized Bolza problem (3.1), with its endpoint cost term added to the Lagrangian integral cost term, full convexity means the combination of the full convexity of L and the convexity of the endpoint function l. It's easy to see that

$$\text{under full convexity in the Bolza problem, } \mathcal{J}_{L,l} \text{ is convex on } \mathcal{A}^1_n[t_0, t_1], \tag{4.2}$$

so this marks the territory in trajectory control and optimization that lies in convex optimization with its powerful extra features. But strangely, that major division has long escaped notice in the history of the calculus of variations and optimal control. Perhaps that's because applications to physics and engineering often don't enjoy full convexity. In modern times, however, other trajectory applications more open to convexity have emerged, such as in economics and operations research. Furthermore, there is much greater appreciation and understanding of convexity and the potential for solving nonconvex problems numerically by iteratively solving convex subproblems.

In the generalized control problem (3.13), convexity of $K(t, x, u, v)$ with respect to (x, u, v) implies through the inf-projection formula (3.16) the convexity of $L(t, x, v)$ in (x, v). That way, full convexity in (3.13), meaning the combination of this with the convexity of l there, implies full convexity in the Bolza problem to which the control problem reduces in Theorem 3.3. But this is just a sufficient condition: it's possible for the reduced problem to be fully convex without the convexity of $K(t,x,u,v)$ in (x,u,v), as seen from the example in (3.22). There, the issue comes down to whether the graphs of the set-valued mappings $x \to D(t,x)$ are convex, but u acts only to parametrize those graphs, and of course a convex set might be parameterized nonconvexly.

The potential for full convexity in a neoclassical format of trajectory optimization was first brought out in the paper [27], from the same era as the publication of the Convex Analysis book [26] and reflecting its break with the traditions. Subgradients could serve in place of gradients, and problems might be dualized, as elsewhere in convex optimization.

Immediately at hand were subgradient reformulations of both the Euler-Lagrange equation (2.3), as the existence of an adjoint trajectory $p(\cdot)$ such that

$$(\dot{p}(t), p(t)) \in \partial_{x.v} L(t, x(t), \dot{x}(t)) \text{ a.e. } t, \tag{4.3}$$

and the associated transversality condition (2.9), as

$$(p(t_0), -p(t_1)) \in \partial l(x(t_0), x(t_1)). \tag{4.4}$$

Although in classical theory the emphasis was on the necessity of such conditions, their sufficiency being a far more difficult matter and requiring something more, here was a complete turnabout: sufficiency elementary, necessity the challenge.

Theorem 4.1 (sufficiency under fully convexity)**.** *In a fully convex problem of Bolza, the existence of $p(\cdot) \in \mathcal{A}^1_n[t_0, t_1]$ such that the Euler-Lagrange condition (4.3) and the transversality condition (4.4) hold is sufficient for the trajectory $x(\cdot)$ to minimize the convex functional $\mathcal{J}_{L,l}$ over $\mathcal{A}^1_n[t_0, t_1]$.*

It's instructive that this follows right from the inequalities in the definition of subgradients in convex analysis. In comparing a trajectory $y(\cdot) \in \mathcal{A}^1_n[t_0, t_1]$ to $x(\cdot)$, one has from (4.3) and (4.4) that

$$\begin{aligned} &L(t, y(t), \dot{y}(t)) \geq L(t, x(t), \dot{x}(t)) + \dot{p}(t)\cdot[y(t) - x(t)] + p(t)\cdot[\dot{y}(t) - \dot{x}(t)] \text{ a.e. } t, \\ &l(y(t_0), y(t_1)) \geq l(x(t_0), x(t_1)) + p(t_0)\cdot[y(t_0) - x(t_0)] - p(t_1)\cdot[y(t_1) - x(t_1)], \end{aligned}$$

and consequently that $\mathcal{J}_{L,l}(y(\cdot)) \geq \mathcal{J}_{L,l}(x(\cdot)) +$

$$\begin{aligned} &\textstyle\int_{t_0}^{t_1} [\dot{p}(t)\cdot[y(t) - x(t)] + p(t)\cdot[\dot{y}(t) - \dot{x}(t)]dt + p(t_0)\cdot[y(t_0) - x(t_0)] - p(t_1)\cdot[y(t_1) - x(t_1)] \\ &\qquad = \textstyle\int_{t_0}^{t_1} \dot{\theta}(t)dt - \theta(t_1) + \theta(t_0) = 0 \text{ for } \theta(t) = p(t)\cdot[y(t) - x(t)]. \end{aligned} \tag{4.5}$$

Another immediate option in the fully convex case, taking advantage of the concavity-convexity in (4.1), is a subgradient formulation of the classical Hamiltonian equations (2.5) in the form of a convex-valued differential inclusion

$$\dot{x}(t) \in \partial_p H(t, x(t), p(t)), \quad -\dot{p}(t) \in \partial_x[-H](t, x(t), p(t)) \quad \text{for a.e. } t. \tag{4.6}$$

In fact, from a basic rule for partial conjugation in convex analysis, (q, p) is a subgradient of the convex function $L(t, \cdot, \cdot)$ at (x, v) if and only if $v \in \partial_p H(t, x, p)$ and $-q \in \partial_x[-H](t, x, p)$, so that

under full convexity, the generalized Hamiltonian condition (4.6) is an equivalent statement of the generalized Euler-Lagrange condition (4.3). (4.7)

Moreover in the autonomous case of $H(x, p)$ instead of $H(t, x, p)$, the function $t \mapsto H(x(t), p(t))$ generated by (4.5) has to be *constant*, as shown in [28]. This is intriguing because such constancy reflects laws of conservation of energy in applications of classical theory to physics. Such laws carry over then to completely different applications, such as perhaps in economics?

The question of when the sufficient conditions in Theorem 4.1 are also necessary can be answered by an appeal to duality and, as is familiar in other areas of convex optimization, an accompanying sort of constraint qualification. Conjugate convex functions have already entered in the relationship between the Lagrangian and the Hamiltonian, but in the framework of full convexity in the Bolza problem, conjugates of the convex functions l and $L(t,\cdot,\cdot)$ can be investigated, too. It turns out that a minor twist in the formulas will be desirable. Instead of passing directly to the conjugate functions l^* and $[L(t,\cdot,\cdot)]^*$, which will be denoted by $L^*(t,\cdot,\cdot)$, it's helpful to introduce

$$m(p_0,p_1) := l^*(p_0,-p_1), \qquad M(t,p,q) = L^*(t,q,p), \tag{4.8}$$

because then, by the rule for subgradients of conjugate functions,

$$\begin{array}{lcl} (\dot{p}(t),p(t)) \in \partial_{x,v}L(t,x(t),\dot{x}(t)) & \iff & (\dot{x}(t),x(t)) \in \partial_{p,q}M(t,p(t),\dot{p}(t)), \\ (p(t_0),-p(t_1)) \in \partial l(x(t_0),x(t_1)) & \iff & (x(t_0),-x(t_1)) \in \partial m(p(t_0),p(t_1)). \end{array} \tag{4.9}$$

That way, the optimality conditions (4.3)-(4.4) for the given problem of Bolza with full convexity translate to the same conditions for the *dual problem of Bolza*

$$\text{minimize } \mathcal{J}_{M,m}(p(\cdot)) := \int_{t_0}^{t_1} M(t,p(t),\dot{p}(t))dt + m(p(t_0),p(t_1)) \tag{4.10}$$

for trajectories $p(\cdot) \in \mathcal{A}^1_n[t_0,t_1]$, namely the existence of $x(\cdot) \in \mathcal{A}^1_n[t_0,t_1]$ such that

$$(\dot{x}(t),x(t)) \in \partial_{p,q}M(t,p(t),\dot{p}(t)) \text{ for a.e. } t, \qquad (x(t_0),-x(t_1)) \in \partial m(p(t_0),p(t_1)). \tag{4.11}$$

The dual problem fits the requirements of this context because one of the virtues motivating the concept of a normal integrand is that this property of $L(t,x,v)$ is preserved in conjugacy.

Theorem 4.2 (dual trajectories in optimality [27])**.** *For fully convex problems of Bolza, the following assertions about a pair of trajectories $x(\cdot)$ and $p(\cdot)$ in $\mathcal{A}^1_n[t_0,t_1]$ are equivalent:*

(a) *the primal Euler-Lagrange and transversality conditions in (4.3)–(4.4) are satisfied,*

(b) *the dual Euler-Lagrange and transversality conditions in (4.11) are satisfied,*

(c) *$x(\cdot)$ minimizes $\mathcal{J}_{L,l}$ in the primal problem, $p(\cdot)$ minimizes $\mathcal{J}_{M,m}$ in the dual problem, and*

$$\mathcal{J}_{L,l}(x(\cdot)) = -\mathcal{J}_{M,m}(p(\cdot)). \tag{4.12}$$

The equation in (4.12) comes directly out of the Fenchel equations that hold for subgradients of conjugate convex functions, here with respect to having, in the case of (a) or (b) of the theorem,

$$\begin{array}{l} L(t,x(t),\dot{x}(t)) + M(t,p(t),\dot{p}(t)) = x(t)\cdot\dot{p}(t) + \dot{x}(t)\cdot p(t), \\ l(x(t_0),x(t_1)) + m(p(t_0),p(t_1)) = x(t_0)\cdot p(t_0) - x(t_1)\cdot p(t_1). \end{array}$$

It corresponds to integrating the first equation over $[t_0,t_1]$, adding that to the second equation, and then pursuing a reduction like the one in (4.5).

All this concerns sufficiency in optimality. What can be said instead about the necessity of the Euler-Lagrange and transversality conditions in Theorems 4.1, which is tied through Theorem 4.2 to the existence of a solution to the dual problem as well as to the primal Bolza problem? It was observed in Section 3 that existence for the primal problem involved having an upper bound of sorts on the

Hamiltonian $H(t,x,p)$. Presumably, then, existence of a solution to the dual problem would involve such an upper bound on the Hamiltonian for that problem, that being the function

$$\widetilde{H}(t,p,x) := \sup_q \{\, x{\cdot}q - M(t,p,q)\}, \;\text{ which calculates to }\; \widetilde{H}(t,p,x) = -H(t,x,p). \tag{4.13}$$

Primally and dually, then, there must presumably be both upper and lower bounds on $H(t,x,p)$ and thus its finiteness. There's no surprise then, that one of the assumptions on the way to necessity in this setting is having

$$H(t,x,p) \;\text{ summable with respect to } t \text{ over } [t_0,t_1]. \tag{4.14}$$

Primal and dual "constraint qualifications" will also enter now as assumptions. They involve the convex sets $\operatorname{dom} l$ and $\operatorname{dom} m$ to which the pairs $(x(t_0),x(t_1))$ and $(p(t_0),p(t_1))$ are implicitly constrained along with the *dynamically attainable* sets

$$\left.\begin{aligned} C_L &= \Big\{ (x_0,x_1) \,\Big|\, \exists\, x(\cdot) \in \mathcal{A}^1_n[t_0,t_1] :\, \textstyle\int_{t_0}^{t_1} L(t,x(t),\dot{x}(t))dt < \infty,\ (x(t_0),x(t_1)) = (x_0,x_1) \Big\}, \\ C_M &= \Big\{ (p_0,p_1) \,\Big|\, \exists\, p(\cdot) \in \mathcal{A}^1_n[t_0,t_1] :\, \textstyle\int_{t_0}^{t_1} M(t,p(t),\dot{p}(t))dt < \infty,\ (p(t_0),p(t_1)) = (p_0,p_1) \Big\}, \end{aligned}\right\} \tag{4.15}$$

which are likewise convex in $I\!\!R^n \times I\!\!R^n$. Obviously, having $\mathcal{J}_{L,l} \not\equiv \infty$ entails $C_L \cap \operatorname{dom} l \neq \emptyset$, while having $\mathcal{J}_{M,m} \not\equiv \infty$ entails $C_M \cap \operatorname{dom} m \neq \emptyset$. The constraint qualifications strengthen that by requiring the relative interiors of these convex sets to have nonempty intersection.

Theorem 4.3 (existence and necessity under full convexity [30]). *Along with the Hamiltonian finiteness condition in (4.14), assume that*

$$\operatorname{ri} C_L \cap \operatorname{ri}[\operatorname{dom} l] \neq \emptyset, \qquad \operatorname{ri} C_M \cap \operatorname{ri}[\operatorname{dom} m] \neq \emptyset. \tag{4.16}$$

Then optimal trajectories exist in both the primal and dual Bolza problems, and the equivalent conditions in Theorem 4.2 hold for them.

The dual constraint qualification can be translated instead into a growth condition on the primal functions L and l. See [30] for that and some wider perspectives on the problem assumptions and relationships.

The need for the Hamiltonian to be finite in (4.14), in securing both primal and dual existence, reveals something important about state constraints. It has already been noted that, because primal existence could be expected to require $H(t,x,p) < \infty$, dual existence could be expected, in view of the formula for the dual Hamiltonian in (4.13), to require $H(t,x,p) > -\infty$. But the latter is equivalent, by the formula for getting H from the Lagrangian L in (3.6) and the formula for the implicit state constraint set $X(t)$ in (3.5) via (3.4), to having $X(t) = I\!\!R^n$. In other words, having $H(t,x,p) > -\infty$ corresponds to having no implicit state constraints in the primal problem — and likewise, having $H(t,x,p) < \infty$ corresponds to having no implicit state constraints in the dual problem!

Here's the fundamental reason that has been discovered to be behind this. A primal state constraint $x(t) \in X(t)$, when active in the sense of $x(t)$ being a boundary point of $X(t)$ can't, in general, be handled by the Euler-Lagrange condition (4.3) with the dual trajectory serving as a kind of Lagrange multiplier element for it, when $p(\cdot)$ is restricted to the space $\mathcal{A}^1_n[t_0,t_1]$ of absolutely continuous trajectories. Trajectories $p(\cdot)$ that aren't absolutely continuous but merely of *bounded variation* have to be brought in, as came to the surface in [31] with its utilization of [29]. The statement of (4.3) must be adapted to that accordingly.

In duality then, with respect to the roles of $x(\cdot)$ and $p(\cdot)$ being reversed in passing to the dual Bolza problem, without the condition $H(t,x,p) < \infty$ the primal problem might have to expanded to admit trajectories $x(\cdot)$ of bounded variation. Generalized problems of Bolza under full convexity with both primal and dual trajectories admitted to be just of bounded variation, and therefore maybe discontinuous with jumps at some times t, have been investigated in [34] and [36]. But much remains to be understood. Special complications arise in handling rigorously the possibility of jumps in states at the initial and final times t_0 and t_1, which seems to undermine the essence of having endpoint constraints.

It was seen in Section 3 that problems of optimal control can be reduced to generalized problems of Bolza in which the control variables are suppressed in the optimization, but can be recovered in the end through measurable selection. Under full convexity, this process can be mirrored in duality. An example will demonstrate it. The dynamical system will be autonomus, for simplicity. Start with the control problem being to

$$\begin{array}{l} \text{minimize } \int_{t_0}^{t_1} k(Cx(t),u(t))dt + l(x(t_0),x(t_1)) \\ \text{subject to } \dot{x}(t) = Ax(t) + Bu(t), \text{ for a.e. } t, \end{array} \tag{4.17}$$

where k is a lower semicontinuous proper convex function of $(y,u) \in I\!R^m \times I\!R^d$ and A, B and C are matrices of appropriate sizes. The role of C is to identify the aspects of the state x that can be "observed." Reduction to a generalized Bolza problem is achieved by taking the Lagrangian to be

$$L(x,v) = \inf_u \{\, k(Cx,u) \,|\, Ax + Bu = v \} \tag{4.18}$$

(with the inf being ∞ when no such u exists). The associated Hamiltonian function is

$$\begin{aligned} H(x,p) &= \sup_v \{\, p{\cdot}v - L(x,v)\} = \sup_u \{\, p{\cdot}[Ax + Bu] - k(Cx,u)\} \\ &= p{\cdot}Ax + h(Cx, B^*p), \end{aligned} \tag{4.19}$$

where B^* is the transpose of B and

$$h(y,z) := \sup_u \{\, z{\cdot}u - k(y,u)\}. \tag{4.20}$$

Suppose that h, which by the convexity of k is itself concave in y and convex in z, is a finite function on $I\!R^m \times I\!R^d$. The Hamiltonian H in (4.19) is then finite on $I\!R^n \times I\!R^n$, which is a special case of the condition (4.14) invoked in Theorem 4.3 and guarantees the conclusions there. The finiteness of h also, separately, guarantees that the infimum (4.18) is attained when not ∞ and similarly attainment of the supremum in (4.19).

To discern the form of the dual Bolza problem, the symmetry in (4.19) can be put to use together with observation through (4.13) that the dual Hamiltonian has to be

$$\widetilde{H}(p,x) = -x{\cdot}A^*p - h(Cx, B^*p). \tag{4.21}$$

Since the function conjugate to k is given by

$$k^*(w,z) = \sup_{y,u} \{\, w{\cdot}y + z{\cdot}u - k(y,u)\} = \sup_y \{\, w{\cdot}y + h(y,z)\},$$

the formula for h in (4.20) is supplemented dually by

$$-h(y,z) = \sup_w \{\, w{\cdot}y - k^*(w,z)\}$$

with the supremum always attained. It can be seen from this and (4.21) that, in analogy to (4.19),

$$\widetilde{H}(p,x) = \sup_{w} \{ x{\cdot}[-A^*p + C^*w] - k^*(w, B^*p)\}.$$

The conclusion is that the dual problem is the Bolza reduction of the *dual control* problem

$$\begin{array}{ll} \text{minimize} & \int_{t_0}^{t_1} k^*(w(t), B^*p(t))dt + m(p(t_0), p(t)1)) \\ \text{subject to} & \dot{p}(t) = -A^*p(t) + C^*w(t), \ \text{ for a.e. } t, \end{array} \tag{4.22}$$

where of course $k^*(w(t), B^*p(t))$ could be written instead as $\widetilde{k}(B^*p(t), w(t))$ to emphasize parallels with the starting problem in (4.17).

The duality between (4.17) and (4.22) as problems of optimal control is remarkable especially in the way the control matrix B and observation matrix C in the primal problem appear in transposed reversal as the control matrix C^* and observation matrix B^* in the dual problem. That adds more shine to a phenomenon long known in the control of linear dynamical systems in which "controllability" and "observability" exhibit duality. This kind of example of duality in neoclassical formulations of optimal control dates back to 1972 in [32] with subsequent elaborations in [38].

5 Neoclassical optimality conditions more generally

Without the functions $L(t,\cdot,\cdot)$ and l being convex, the subgradients in the Euler-Lagrange condition (4.3) and transversality condition (4.4) can't be the ones of convex analysis. For hopes of obtaining necessary conditions in such neoclassical form to apply to generalized problems of Bolza without full convexity, other concepts of subgradient have to be brought in. By now there is a highly developed theory of subdifferentiation in first-order variational analysis that supplies everything that is needed. As explained in [41, Chapter 8], it can be seen as starting with "regular" subgradients, which are defined by an inequality like that for the convex analysis subgradients, which invokes *one-sidedly* the same first-order error term as in the definition of differentiabilty. It produces the subgradients that can serve in general by taking limits of those regular subgradients. Before that confuguration was settled, however, other concepts were tried out.

The pioneer for subgradient conditions for optimal trajectories outside of fully convex problems of Bolza was Clarke [3], [4], [5]. He began by focusing on nonsmooth functions that are finite and Lipschitz continuous, but went on in steps to cover other functions as well. Ultimately his approach amounted to starting from "proximal" subgradients (effectively defined just like regular subgradients but with a *second-order* error term) and producing limiting subgradients from them that turn out to be the same as the ones above. Then he added a convex hull operation designed to promote duality between subgradients and directional derivatives. By the mid 1980s, however, it was apparent that the convexification process ruined hopes for ever developing a robust second-order variational analysis [37]. Furthermore, it was demonstrated around then by Mordukhovich [20] that better rules for calculating subgradients could be put in place without the convexification.

That would seem to settle the shape of the Euler-Lagrange condition (4.3) and transversality condition (4.4) for potentially nonconvex functions $L(t,\cdot,\cdot)$ and l, but not quite. The unsettled aspect has to do with the relationship between the Lagrangian and the Hamiltonian. Should the Hamiltonian condition in the nonconvex setting still be taken in the form of (4.6) with separate and differently oriented subdifferentiation in x and p? Maybe, at least in cases where $H(t,\cdot,\cdot)$ is lower semicontinuous and proper (as under growth conditions that have been noted) it should be taken instead in the form

$$(-\dot{p}(t), \dot{x}(t)) \in \partial_{x,p} H(t, x(t), p(t)) \ \text{ for a.e. } t, \tag{4.23}$$

despite that form not being equivalent to the Euler-Lagrange condition in the presence of full convexity, as was (4.6), yet suggested by some examples in nonlinear control.

Equivalence aside, it was demonstrated in [14] that anyway, for a big class of problems, optimality entailed the *simultaneous* necessity of the Euler-Lagrange condition (4.3), the Hamiltonian condition (4.23), and transversality condition (4.4) for an associated adjoint trajectory $p(\cdot)$. On the other hand, according to [39], (4.23) is equivalent for finite locally Lipschitz continuous $H(t,\cdot,\cdot)$ to posing the Euler-Lagrange condition instead (at least sometimes) as

$$(\dot{p}(t), p(t)) \in \operatorname{con}\Big\{ (-q, v) \,\Big|\, (q, p(t)) \in \partial_{x,v} L(t, x(t), \dot{x}(t)),\ p(t) \in \partial_v L(t, x(t), \dot{x}(t) \Big\}. \tag{4.24}$$

Here the convex hull operation is readmitted, but much more delicately as in Mordukhovich [19], [21].

Later, in explorations of Loewen and the author in [15], the idea came up that a still more subtle adaptation of the Euler-Lagrange condition might be to articulate it as the pair

$$\begin{aligned} &\dot{p}(t) \in \operatorname{con}\{ q \mid (q, p(t)) \in \partial_{x,v} L(t, x(t), \dot{x}(t))\}, \\ &p(t) \in \partial_v L(t, x(t), \dot{x}(t)\}, \end{aligned} \tag{4.25}$$

holding almost everywhere. Then in [40] situations were identified for dynamically convex L in which (4.25) is equivalent to a corresponding version of the Hamiltonian condition, namely

$$\begin{aligned} &\dot{p}(t) \in \operatorname{con}\{ q \mid (q, \dot{x}(t)) \in \partial_{x,p} H(t, x(t), p(t))\}, \\ &\dot{x}(t) \in \partial_p H(t, x(t), p(t)\}. \end{aligned} \tag{4.26}$$

This was utilized in [16]. Results where the starting and ending times t_0 and t_1 can take part in the optimization were obtained in [17]. How the Weierstrass conditon for a strong local minimum might be brought in was studied in [13].

That's more or less as far as this line of research went before other pursuits took over. It would be very interesting to see whether, with so much more now understood in variational analysis, more definitive conclusions could be reached, and the gap between statements of the Euler-Lagrange and Hamiltonian conditions under full convexity and statements without it might be tightened. There's also the question of the extent to which trajectories in an infinite-dimensional Banach space instead of $\mathbb{R}^n$ might be covered in this manner, starting from the strong support in variational analysis provided by Mordukhovich [22, Chapter 6].

Of course, there's no shortage of optimality conditions that have been nicely worked out with the specifics of optimal control in mind. But, as already explained, the neoclassical format offers the possibility of insights that are deeper for trajectory optimization and somehow universal, beyond just the particulars of control.

6 Corresponding innovations in Hamilton-Jacobi-Bellman theory

The Hamilton-Jacobi equation is a partial differential equation in the calculus of variations that aims to describe the evolution of a certain "value function." The Bellman equation, from the 1960s in [2], describes the 'cost-to-go" in an optimization problem formulated as "dynamic programming." The two equations came together in proposed Hamilton-Jacobi-Bellman forms with attractive applications in optimal control — at least in principle. There was serious trouble, however, in the fact that the smoothness postulated in such equations could rarely be expected to prevail in the targeted applications. That was one of the key motivations in optimal control for passing to some form of nonsmooth analysis, as recounted in the textbook of Vinter [45], which is an excellent source of background in this area as well as optimal control in general.

Here only a short foray into the topic is possible, with the emphasis being on how to think about it "neoclassically." For that we stick to autonomous dynamics and a problem in the form:

$$\text{minimize } \; g(x(0)) + \int_0^\tau L(x(t), \dot{x}(t))dt \; \text{ subject to } \; x(\tau) = \xi, \tag{6.1}$$

where $\tau \in (0, \infty)$ and $\xi \in I\!R^n$ are both parameters. The minimum over $x(\cdot) \in \mathcal{A}_n^1[0, \tau]$ is denoted by $V(\tau, \xi)$. The issue is characterizing V, the *value* function associated with (6.1).

In concept, we can look at this in terms of having a function $V_\tau = V(\tau, \cdot)$ on $I\!R^n$ that evolves in time; the dynamics of a "moving function" need to be described. In the setting where differentiability is assumed for g and L and taken for granted for V, and the Hamiltonian H is validly obtained from L somehow by the Legendre transform, the partial differential equation for this purpose would be the classical Hamilton-Jacobi equation

$$\nabla_\tau V(\tau, \xi) + H(\xi, \nabla_\xi V(t, \xi)) = 0. \tag{6.2}$$

However, the differentiability of V ought not to be taken for granted, and optimal control can shove it far out of plausibility. There, g might serve as the indicator of a set or single point, and at the level of generality in Section 3, $L(x, v)$ might be derived from a function $K(x, u, v)$ as in (3.16). Then $H(x, p)$ would given by the maximization formula in (3.17), which would threaten its own differentiability and lead its replacement in (6.2) by an expression defined by maximization over controls u. That, roughly, is where a Hamilton-Jacobi-*Bellman* equation would enter on the scene, substituting for (6.2). But any such development quickly faces conflict with the differentiability of V being far-fetched, and that makes a mystery out of (6.2) and what might be a proper replacement.

The literature of optimal control has anyway not encompassed neoclassical formulations like (6.1), where functions can take on ∞. It has emphasized control dynamics directly, or at least cases corresponding to L being just the indicator of the graph of a differential inclusion. Another difference is that, instead of the "forward" perspective in (6.1) there has usually been a backward (cost-to-go) perspective, where the interval is $[\tau, T]$ and g is applied to $x(T)$, the parameterization being the *initial* time τ and state ξ. Mathematically, these different approaches are equivalent in the end, but the forward version chosen here avoids an unpleasant proliferation of minus signs.

The challenge of finding a broader substitute for (6.2) has led to the concept of *viscosity* solutions [6], in which something like upper and lower subgradients of H are utilized on a track parallel to, but independent of, other developments in variational analysis. Numerous results have been based on that, although in more limited situations than potentially might be captured neoclassically by (6.1).

In contrast, for the fully convex version of (6.1) a completely different approach has been taken in [42], [43]. In this, besides having g and L be convex, proper and lower semicontinuous, two growth conditions are imposed. The first concerns the differential inclusion (3.4) that's implicitly behind L, namely $\dot{x}(t) \in D(x(t))$. There should exist κ such that $\operatorname{dist}(0, D(x)) \leq \kappa(1 + |x|)$. That's altogether normal from the world of ordinary differential equations in making sure that solutions will exist over time intervals of arbitrary length. It also obviously excludes the interference of state constraints (since the distance to the empty set is always ∞). The other growth condition is more in the vein of the kind used in existence theory:

$$L(x, v) \geq \theta(\,\max\{0, |v| - \alpha|x|\}\,) - \beta|x| \tag{6.3}$$

for constants α, β, and a proper nondecreasing function θ on $[0,\infty)$ such that $\theta(s)/s \to \infty$ as $s \to \infty$. These two growth conditions are shown in [42] to translate to the analogous pair of conditions on the dual Lagrangian M and in that way to come out in Hamiltonian terms as the symmetric bounds

$$H(x,p) \leq \varphi(p) + (\alpha|p| + \beta)|x|, \qquad -H(x,p) \leq \tilde{\varphi}(x) + (\tilde{\alpha}|x| + \tilde{\beta})|p|, \tag{6.4}$$

with both φ and $\tilde{\varphi}$ being finite convex functions on $I\!R^n$.

Under these assumptions, it was established in [42] that the value function $V(\tau,\xi)$ is well defined over $[0,\infty) \times I\!R^n$ with the function $V_\tau = V(\tau,\cdot)$ always being convex, starting from $V_0 = g$ and depending epi-continuously on $\tau \in [0,\infty)$, with

$$(\sigma,\eta) \in \partial V(\tau,\xi) \iff \eta \in \partial_\xi V(\tau,\xi),\ \sigma = -H(\xi,\eta) \tag{6.5}$$

when $\tau > 0$. The generalized Hamilton-Jacobi replacement for the classical equation in (6.2) is in this case therefore

$$\sigma + H(\xi,\eta) = 0 \ \text{ for all } \ (\sigma,\eta) \in \partial V(\tau,\xi). \tag{6.6}$$

Moreover the conjugate function V_τ^* is $\widetilde{V}_\tau(\cdot)$ for the value function $\widetilde{V}(\tau,\eta)$ that evolves with respect to the dual Lagrangian M in (4.8) from $\widetilde{V}_0 = g^*$. It has

$$(-\sigma,\xi) \in \partial \widetilde{V}(\tau,\eta) \iff (\sigma,\eta) \in \partial V(\tau,\xi) \tag{6.7}$$

and the generalized Hamilton-Jacobi equation for the dual Hamiltonian in (4.13), with that being

$$\sigma - H(\xi,\eta) = 0 \ \text{ for all } \ (\sigma,\xi) \in \partial \widetilde{V}(\tau,\eta). \tag{6.8}$$

For the example of linear control dynamics at the end of Section 4, with H given by (4.19), the relationships in (6.5) and (6.6) with the maximization in (4.20) signal a Hamilton-Jacobi-Bellman equation, and duality is the centerpiece there as well.

These features of [42] are complemented by the upper and lower envelope representations of the primal and dual value functions furnished in [43] in terms of a *kernel* function and by the parametric extensions in the paper [44]. They deliver, in this setting of full convexity, a Hamilton-Jacobi theory that fulfills every wish and might be a template for neoclassical generalizations without full convexity. How much is known about that?

Long before [42] and [43] with their special utilization of convexity, there were the results on the existence and uniqueness of "viscosity" solutions to Hamilton-Jacobi equations in [6]. Those results weren't tuned to value functions coming from an associated Lagrangian, however, and their requirements of boundedness and uniform continuity on the Hamiltonian and the solution weren't suited for that. Interestingly, the restrictions imposed in viscosity theory make it *inapplicable* to Hamiltonians satisfying (6.4), which besides are concave in x and convex in p in this framework of full convexity and its theme of duality.[5]

But value functions were the center of attention in the 1993 paper of Frankowska [8]. She brought out the importance of Aubin's theory of *viability* for this subject and made many contributions in that direction; see also [7], for instance, and her book [9]. Galbraith [10] picked up on that in a more "neoclassical" setting again. He made strong advances in which H could even take on ∞, and that could have led to very much more. But unfortunately he then switched to a career in finance, so there was no follow-up. His paper [10] would be a good starting point for continued research.

[5]This didn't prevent a reviewer of [42], where the results seriously depended on (6.4), from resisting publication unless it would be rewritten using already-known PDE methodology, so as to make the proofs more "understandable"!

References

[1] Aubin, J.-P., and Cellina, A., *Differential Inclusions*, Springer-Verlag, 1984.

[2] Bellman, R., *Dynamic Programming*, Princeton University Press, 1957.

[3] Clarke, F.H., "Generalized gradients and applications," *Trans. Amer. Math. Soc.* **205** (1975), 247–262.

[4] Clarke, F.H., *Optimization and Nonsmooth Analysis*, Wiley, 1983.

[5] Clarke, F.H., *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Converence Series in Appl. Math. 57, SIAM Publications, 1989.

[6] Crandall, M.G., and Lions, P.-L, "Viscosity solutions to Hamilton-Jacobi equations," *Transactions Amer. Math. Soc.* **277** (1983), 1–42.

[7] Dal Maso, G., and Frankowska, H., "Value functions for Bolza problems with discontinuous Lagrangians and Hamilton-Jacobi inequalities," *ESAIM-COCV* **5** (2000), 369–394.

[8] Frankowska, H., "Lower semicontinuous solutions to Hamilton-Jacobi equations," *SIAM J. Control Optim.* **31** (1993), 257–272.

[9] Frankowska, H., *Value Functions in Optimal Control*, ICTP Lecture Notes, 2002 (Summer School of Mathematical Control Theory, Sep 2–28, 2001).

[10] Galbraith, G., "Extended Hamilton-Jacobi characterization of value functions in optimal control," *SIAM J. Control Opt.* **39** (2001), 281–305.

[11] Gelfand, A. M, and Fomin, S. V., *Calculus of Variations*, 1963 translation in the series *Selected Russian Publications in the Mathematical Sciences*, Prentice-Hall, N.J.

[12] Hestenes, M.R., *Calculus of Variations and Optimal Control Theory*, Wiley, 1966.

[13] Ioffe, A. D, and Rockafellar, R. T., "The Euler and Weierstrass conditions for nonsmooth variational problems," *Calculus of Variations & PDE* **4** (1996), 59–87.

[14] Loewen, P. D, and Rockafellar, R. T., "The adjoint arc in nonsmooth optimization," *Transactions American Math. Soc,* **325** (1991), 39–72.

[15] Loewen, P. D, and Rockafellar, R. T., "Optimal control of unbounded differential inclusions," *SIAM J. Control Opt.* **32** (1994), 442–470.

[16] Loewen, P. D, and Rockafellar, R. T., "New necessary conditions for the generalized problem of Bolza," *SIAM J. Control Opt.* **34** (1996), 1496–1511.

[17] Loewen, P. D, and Rockafellar, R. T., "Bolza problems with general time constraints," *SIAM J. Control Opt.* **35** (1997), 2050–2069.

[18] Mordukhovich, B.S., "The maximum principle in time-optimal control with nonsmooth constraints," *J. Appl. Math. Mechanics* **40** (1976), 960–969.

[19] Mordukhovich, B.S., "Optimization and approximation of differential inclusions," *Cybernetics* **24** (1988), 781–788.

[20] Mordukhovich, B.S., *Approximation Methods in Optimization and Control*, Nauka, Moscow, 1988 (in Russian).

[21] Mordukhovich, B.S., "On variational analysis of differential inclusions," in *Optimization and Nonlinear Analysis* (A.D. Ioffe et al., eds.), Longman, 1992, 199–213.

[22] Mordukhovich, B.S., *Variational Analysis and Generalized Differentiation I, II*, Grundlehren der Mathematischen Wissenschaften **330, 331**, Springer, 2005.

[23] Mordukhovich, B.S., *Second-Order Variational Analysis in Optimization, Variational Stability and Control*, Springer 2024.

[24] Rockafellar, R. T., "Conjugates and Legendre transforms of convex functions," *Canadian J. Math.* **19** (1967), 200–205.

[25] Rockafellar, R. T., "Integrals which are convex functionals," *Pacific J. Math.* **24** (1968), 525-539.

[26] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, 1970.

[27] Rockafellar, R. T., "Conjugate convex functions in optimal control and the calculus of variations," *J. Math. Analysis Appl.* **28** (1970).

[28] Rockafellar, R. T., "Generalized Hamiltonian functions for convex problems of Lagrange," *Pacific J. Math.* **33** (1970), 411–427.

[29] Rockafellar, R. T., "Integrals which are convex functionals, II" *Pacific J. Math.* **39** (1971), 439-469.

[30] Rockafellar, R. T., "Existence and duality theorems for convex problems of Bolza," *Transactions America Math. Soc.* **159** (1971), 1–40.

[31] Rockafellar, R. T., "State constraints in convex problems of Bolza," *SIAM J. Control* **10** (1972), 691–715.

[32] Rockafellar, R. T., "Dual problems of optimal control," in *Techniques of Optimization*, A.V. Balakrishnan (ed.), Academic Press, 1972), 423–431.

[33] Rockafellar, R. T., "Existence theorems for general control problems of Bolza and Lagrange," *Advances in Math.* **15** (1975), 315–333.

[34] Rockafellar, R. T., "Dual problems of Lagrange for arcs of bounded variation," in *Calculus of Variations and Control Theory*, D.L. Russell (ed.), Academic Press, 1976, 155–192.

[35] Rockafellar, R. T., "Integral functionals, normal integrands and measurable selections," in *Nonlinear Operators and the Calculus of Variations*, L. Waelbroeck (ed.), Leture Notes in Math. No. 543, Springer-Verlag, 1976, 155–202.

[36] Rockafellar, R. T., "Optimality conditions for convex control problems with nonnegative states and the possibility of jumps," in *Game theory and Math. Economics*, O. Moeschlin (ed.), Noth-Holland, 1981, 339–349.

[37] Rockafellar, R. T., "Maximal monotone relations and the second derivatives of nonsmooth functions," Ann. INst. H. Poincaré Analyse Non Linéaire **2** (1985), 167–184.

[38] Rockafellar, R. T., "Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs," *SIAM J. Control Opt.* **27** (1989), 1007–1025.

[39] Rockafellar, R. T., "Dualization of subgradient conditions for optimality," *Nonlinear Analysis: Theory, Meth., Appl.* **20** (1993), 627–646.

[40] Rockafellar, R. T., "Equivalent subgradient versions of Hamiltonian and euler-Lagrange equations in variational analysis," *SIAM J. Control Opt.* **34** (1996), 1300–1315.

[41] Rockafellar, R. T., and Wets, R.J-B, *Variational Analysis*, Springer-Verlag, 1997.

[42] Rockafellar, R. T., and Wolenski, P.R, "Convexity in Hamilton-Jacobi theory 1: dynamics and duality," *SIAM J. Control Opt.* **40** (2001), 1323–1350.

[43] Rockafellar, R. T., and Wolenski, P.R, "Convexity in Hamilton-Jacobi theory 2: envelope representations," *SIAM J. Control Opt.* **40** (2001), 1351–1372.

[44] Rockafellar, R. T., "Hamilton-Jacobi theory and parametric analysis in fully convex problems of optimal control," *J. Global Optimization* **28** (2004), 419–431.

[45] Vinter, R., *Optimal Control*, Birkhäuser, 2000.

Data availability: No data was generated or analyzed.