IIASA
International Institute for
Applied Systems Analysis
www.iiasa.ac.at

Convexity and Duality in Hamilton-Jacobi Theory

Rockafellar, R.T. and Wolenski, P.R.

IIASA Interim Report
August 1998

International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-057 / August

Convexity and Duality in Hamilton-Jacobi Theory

R. Tyrrell Rockafellar (rtr@math.washington.edu)
Peter R. Wolenski (wolenski@math.lsu.edu)

Approved by
Arkadii Kryazhimskii(kryazhim@iiasa.ac.at)
Principal Investigator, *Dynamic Systems*

Contents

Abstract

Value functions propagated from initial or terminal costs and constraints by way of a differential inclusion, or more broadly through a Lagrangian that may take on ∞, are studied in the case where convexity persists in the state argument. Such value functions, themselves taking on ∞, are shown to satisfy a subgradient form of the Hamilton-Jacobi equation which strongly supports properties of local Lipschitz continuity, semidifferentiability and Clarke regularity. An extended 'method of characteristics' is developed which determines them from Hamiltonian dynamics underlying the given Lagrangian. Close relations with a dual value function are revealed.

Keywords. Convex value functions, dual value functions, subgradient Hamilton-Jacobi equations, extended method of characteristics, nonsmooth Hamiltonian dynamics, viscosity solutions, variational analysis, optimal control, generalized problems of Bolza.

AMS subject classification. Primary: 49L25. Secondary: 93C10, 49N15.

About the Authors

R. Tyrrell Rockafellar
University of Washington
Seattle, WA 98195, USA

Peter R. Wolenski
Louisiana State University
Baton Rouge, LA 70803, USA

Acknowledgement

This work was supported in part by the National Science Foundation under grant DMS–9500957 for the first author and grant DMS–9623406 for the second author.

CONVEXITY AND DUALITY IN HAMILTON-JACOBI THEORY

R. TYRRELL ROCKAFELLAR and PETER R. WOLENSKI

University of Washington and Louisiana State University

1. Introduction

Fundamental to optimal control and the calculus of variations are value functions $V : [0,\infty) \times I\!R^n \to \overline{I\!R} := [-\infty, \infty]$ of the type

$$V(\tau,\xi) := \inf\Big\{ g\big(x(0)\big) + \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(\tau)=\xi \Big\}, \qquad V(0,\xi) = g(\xi), \tag{1.1}$$

which propagate an initial cost function $g : I\!R^n \to \overline{I\!R}$ forward from time 0 in a manner dictated by a Lagrangian function $L : I\!R^n \times I\!R^n \to \overline{I\!R}$. The possible extended-real-valuedness of g and L serves in the modeling of the constraints and dynamics involved in this propagation, such as restrictions on $x(0)$ and on $\dot{x}(t)$ relative to $x(t)$. The minimization takes place over the arc space $\mathcal{A}^1_n[0,\tau]$, in the general notation that $\mathcal{A}^p_n[\tau_0,\tau_1]$ consists of all absolutely continuous $x(\cdot) : [\tau_0,\tau_1] \to I\!R^n$ with derivative $\dot{x}(\cdot) \in \mathcal{L}^p_n[\tau_0,\tau_1]$. Value functions of the "cost-to-go" type, which propagate a terminal cost function backward from a time T, are covered by (1.1) through time reversal.

An important issue in Hamilton-Jacobi theory is the extent to which V can be characterized in terms of the Hamiltonian function $H : I\!R^n \times I\!R^n \to \overline{I\!R}$ associated with L, as defined through the Legendre-Fenchel transform by

$$H(x,y) := \sup_v \Big\{ \langle v,y\rangle - L(x,v) \Big\}. \tag{1.2}$$

Under the properties of this transform, $H(x,y)$ is sure to be convex in y. When $L(x,v)$ is convex, proper and lower semicontinuous in v, as is natural for the existence of optimal arcs in (1.1), the reciprocal formula holds that

$$L(x,v) = \sup_y \Big\{ \langle v,y\rangle - H(x,y) \Big\}, \tag{1.3}$$

so L and H are completely dual to each other.

It is well recognized that a function V given by (1.1) can fail to be smooth despite any degree of smoothness of g and L, or for that matter, H. Much of modern Hamilton-Jacobi theory has revolved around this fact, especially in coming up with generalizations of the Hamilton-Jacobi PDE that might pin down V, which of course was the historical motivation for that equation. Little attention has been paid, though, to ascertaining circumstances in which $V(\tau,\xi)$ is convex in ξ for each $\tau \geq 0$, and to exploring the consequences of such convexity. The convex case merits study for several reasons, however.

Convexity is a crucial marker in classifying optimization problems, and it's often accompanied by interesting phenomena of duality. It can provide powerful support in matters of computation and approximation. Moreover, it has a prospect here of enabling V to be characterized via H in other ways, complementary to the Hamilton-Jacobi PDE, such as versions of the method of characteristics in which convex analysis can be brought to bear. Efforts in the convex case could therefore shed light on topics in nonsmooth Hamilton-Jacobi theory that so far have been overshadowed by PDE extensions.

The convexity of $V(\tau,\xi)$ in ξ entails, for $\tau = 0$, the convexity of the initial function g, but what does it need from the Lagrangian L? The simplest, and in a certain sense the only robust assumption for this is the joint convexity of $L(x,v)$ in x and v, which corresponds under (1.2) and (1.3) to pairing the natural convexity of $H(x,y)$ in y with the concavity of $H(x,y)$ in x. This is what we work with, along with mild conditions of semicontinuity and growth that can readily be dualized.

Our convexity assumptions ensure that the optimization problem appearing in (1.1) fits the theory of generalized problems of Bolza of convex type as developed in Rockafellar [1], [2], [3], [4]. That duality theory, dating from the early 1970's and based entirely on convex analysis [5], hasn't previously been utilized in the Hamilton-Jacobi setting. It had to wait for advances toward handling robustly, by means of subgradients, not only the convexity of $V(\tau,\xi)$ in ξ but also its nonconvexity in (τ,ξ). Such advances have since been through the labor of many researchers, and the time is therefore ripe for investigating the Hamilton-Jacobi aspects of convexity and duality.

Relying on the background of variational analysis in [6], we make progress in several ways. We demonstrate the existence of a dual value function $\tilde{V}$, propagated by a dual Lagrangian $\tilde{L}$, such that the convex functions $V(\tau,\cdot)$ and $\tilde{V}(\tau,\cdot)$ are conjugate to each other under the Legendre-Fenchel transform for every τ. We use this in particular to derive a subgradient Hamilton-Jacobi equation satisfied directly by V, and a dual one for $\tilde{V}$, despite the unboundedness of these functions and their pervasive ∞ values. At the same time we establish a new subgradient form of the "method of characteristics" for determining these functions from the Hamiltonian H.

Central to our approach is a generalized Hamiltonian ODE associated with H, which is actually a differential inclusion in terms of subgradients instead of gradients. By focusing on $V_\tau = V(\tau,\cdot)$ as a convex function on $\mathbb{R}^n$ that varies with τ, we bring to light the remarkable fact that the graph of the subgradient mapping ∂V_τ evolves through nothing more nor less than its "drift" in the (set-valued) flow in $\mathbb{R}^n \times \mathbb{R}^n$ induced by this generalized Hamiltonian dynamical system.

Our treatment of V, although limited to the convex case, contrasts with other work on generalized Hamilton-Jacobi equations which, in coping with discontinuities and ∞ values, has required $H(x,y)$ to be not just convex in y but also positively homogeneous in y; see Frankowska [7], [8], where ∞ is admitted directly, or more recently Bardi and Capuzzo-Dolcetta [9; Chap. V, §5], where the conditions on H are narrower and ∞ is suppressed by nonlinear rescaling. Rescaling isn't compatible with an emphasis on convexity.

While the interior of the set of points where $V < \infty$ could be empty, we prove that if it isn't, then properties of semidifferentiability, Clarke regularity and local Lipschitz continuity hold for V on that open set under our assumptions. Also, we identify through duality the situations in which coercivity or global finiteness is preserved for all $\tau > 0$.

For simplicity and to illuminate clearly the new features stemming from convexity, we keep to the case of a time-independent Lagrangian L, although extensions of the results to accommodate time dependence would be possible.

2. Hypotheses and Main Results

In formulating the conditions that will be invoked throughout this paper, we abbreviate lower semicontinuous by "lsc" and refer to an extended-real-valued function as *proper* when it's not the constant function ∞ yet nowhere takes on $-\infty$. Thus, a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper if and only if its effective domain $\operatorname{dom} f := \{v \mid f(v) < \infty\}$ is nonempty and, on this set, f is finite. Equivalently, f is proper if and only if its epigraph $\operatorname{epi} f := \{(v,s) \mid s \in \mathbb{R},\ f(v) \le s\}$ is nonempty and contains no (entire) vertical lines. The convexity of a function f corresponds to the convexity of the set $\operatorname{epi} f$, while the lower semicontinuity of f corresponds to the closedness of f. The convexity of f implies the convexity of $\operatorname{dom} f$ (and the convexity of the restriction of f to that set), but the lower semicontinuity of f need not entail the closedness of $\operatorname{dom} f$. (This can happen for instance when $f(v)$ approaches ∞ as v approaches the boundary of $\operatorname{dom} f$ from within.)

We denote the Euclidean norm by $|\cdot|$ and call f *coercive* when it is bounded from below and has $f(v)/|v| \to \infty$ as $|v| \to \infty$. Coercivity of a proper nondecreasing function θ on $[0,\infty)$ means that $\theta(s)/s \to \infty$ as $s \to \infty$. For a proper convex function f on $\mathbb{R}^n$, coercivity is equivalent to the finiteness of the conjugate convex function f^* on $\mathbb{R}^n$.

Basic Assumptions.

(A0) *The initial function g is convex, proper and lsc on $\mathbb{R}^n$.*

(A1) *The Lagrangian function L is convex, proper and lsc on $\mathbb{R}^n \times \mathbb{R}^n$.*

(A2) *The set $F(x) := \operatorname{dom} L(x,\cdot)$ is nonempty for all x, and there is a constant ρ such that $\operatorname{dist}(0, F(x)) \le \rho\big(1 + |x|\big)$ for all x.*

(A3) *There are constants α and β and a coercive, proper, nondecreasing function θ on $[0,\infty)$ such that $L(x,v) \ge \theta\big(\max\{0,\ |v| - \alpha|x|\}\big) - \beta|x|$ for all x and v.*

The joint convexity of L with respect to x and v in (A1) contrasts with the more common assumption of convexity merely with respect to v. It is vital to our duality-based methodology. In combination with the convexity in (A0), it ensures that the functional

$$J_\tau\big(x(\cdot)\big) \ := \ g\big(x(0)\big) + \int_0^\tau L\big((x(t), \dot{x}(t)\big)dt \tag{2.1}$$

is convex on $\mathcal{A}_n^1[0,\tau]$. It also, as a side benefit, guarantees that J_τ is well defined. That follows because $L(x(t), \dot{x}(t))$ is measurable in t when L is lsc, whereas L majorizes at least one affine function on $\mathbb{R}^n \times \mathbb{R}^n$ through its convexity and properness. Then there exist $(w,y) \in \mathbb{R}^n \times \mathbb{R}^n$ and $c \in \mathbb{R}$ with $L(x(t), \dot{x}(t)) \ge \langle x(t), w\rangle + \langle \dot{x}(t), y\rangle - c$, the expression on the right being summable in t. The integral thus has an unambiguous value in $(-\infty, \infty]$, and so then does $J_\tau(x(\cdot))$.

In (A2), the mapping F gives the differential inclusion that's implicit in the Lagrangian L. Obviously $J_\tau(x(\cdot)) = \infty$ unless the arc $x(\cdot)$ satisfies the constraints:

$$\dot{x}(t) \in F(x(t)) \text{ a.e. } t, \text{ with } x(0) \in D := \operatorname{dom} g. \tag{2.2}$$

Note that the graph of F, which is the set $\operatorname{dom} L \subset \mathbb{R}^n \times \mathbb{R}^n$, is convex by (A1), although not necessarily closed. Similarly, the initial set D in these implicit constraints is convex by (A0), but need not be closed. Of course, in the special case where L is finite everywhere, the graph of F is all of $\mathbb{R}^n \times \mathbb{R}^n$ and the condition $\dot{x}(t) \in F(x(t))$ trivializes; likewise, if g is finite everywhere the condition $x(0) \in D$ trivializes.

The nonempty-valuedness of F in (A2) means that there are no state constraints implicitly imposed by L. The growth condition in (A2) will be seen to imply that the differential inclusion in (2.2) has no "forced escape time": from any point it provides at least one trajectory over the infinite time interval $[0,\infty)$. The nonemptiness of $F(x)$ didn't really have to be mentioned separately from this growth condition, inasmuch as the distance of any point to the empty set is ∞.

The function $L(x,\cdot)$ on $\mathbb{R}^n$, which for each x is convex by (A1) and proper by (A2), is coercive under the growth condition in (A3). Note that this growth condition is much weaker than the commonly imposed Tonelli-type condition in which $L(x,v) \ge \theta(|v|)$ for a coercive, proper, nondecreasing function θ. For instance, it covers the case of $L(x,v) = L_0(v - Ax) + L_1(x)$ for coercive L_0 and a function L_1 that does not go down to $-\infty$ at more than an linear rate, whereas the Tonelli-type condition would not do that unless $A = 0$ and L_1 is bounded from below.

The following consequence of our assumptions sets the stage for our analysis of the value function V as giving a "continuously moving" convex function on $\mathbb{R}^n$.

Theorem 2.1 (convexity in the value function)**.** *Under assumptions (A0), (A1), (A2) and (A3), the function $V_\tau = V(\tau,\cdot)$ on $\mathbb{R}^n$ is for every $\tau \in [0,\infty)$ proper, lsc and convex. Moreover V_τ depends epi-continuously on τ. In particular, V is proper and lsc as a function on $[0,\infty)\times\mathbb{R}^n$, and V_τ epi-converges to g as $\tau\searrow 0$.*

This theorem will be proved in §5. The epi-continuity in its statement refers to the continuity of the set-valued mapping $\tau \mapsto \operatorname{epi} V_\tau$ with respect to Painléve-Kuratowski set convergence. It amounts to the following assertion (here, as elsewhere in this paper, we consistently use superscript $\nu = 1,2,\ldots \to \infty$ in describing sequences):

$$\text{whenever } \tau^\nu \to \tau \text{ with } \tau^\nu \geq 0, \text{ one has} \\ \begin{cases} \liminf_\nu V(\tau^\nu,\xi^\nu) \geq V(\tau,\xi) & \text{for every sequence } \xi^\nu \to \xi, \\ \limsup_\nu V(\tau^\nu,\xi^\nu) \leq V(\tau,\xi) & \text{for some sequence } \xi^\nu \to \xi, \end{cases} \tag{2.3}$$

where the first limit property is the lower semicontinuity of V on $[0,\infty)\times\mathbb{R}^n$. An exposition of the theory of epi-convergence of functions on $\mathbb{R}^n$ is available in Chapter 7 of [6].

Observe that the epi-convergence in Theorem 2.1 answers the question of how the initial condition $V_0 = g$ should be coordinated with the behavior of V when $\tau > 0$. Pointwise convergence of V_τ to V_0 as $\tau\searrow 0$ isn't a suitable property for a context of semicontinuity and extended-real-valuedness.

Epi-convergence has implications also for the subgradients of the functions V_τ. Recall that for a proper convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point x, a vector $y \in \mathbb{R}^n$ is a *subgradient in the sense of convex analysis* if

$$f(x') \geq f(x) + \langle y, x'-x\rangle \quad \text{for all } x' \in \mathbb{R}^n. \tag{2.4}$$

The set of such subgradients is denoted by $\partial f(\bar{x})$. (This is, in particular, empty when $\bar{x} \notin \operatorname{dom} f$ but nonempty when $\bar{x} \in \operatorname{ri}\operatorname{dom} f$, the relative interior of the convex set $\operatorname{dom} f$; see [5], [6].) The *subgradient mapping* $\partial f : x \mapsto \partial f(x)$ has graph

$$\operatorname{gph}\partial f := \big\{(x,y)\,\big|\, y \in \partial f(x)\big\} \subset \mathbb{R}^n\times\mathbb{R}^n. \tag{2.5}$$

When f is lsc as well as proper and convex, ∂f is a maximal monotone mapping, and $\operatorname{gph}\partial f$ is therefore an globally Lipschitzian manifold of dimension n in $\mathbb{R}^n\times\mathbb{R}^n$; see [6; Chapter 12]. Furthermore, epi-convergence of functions corresponds in this picture to graphical convergence of their subgradient mappings, i.e., Painlevé-Kuratowski set convergence of their graphs; [6; 12.35].

Corollary 2.2 (subgradient manifolds)**.** *Under (A0), (A1), (A2) and (A3), the graph of the subgradient mapping ∂V_τ is, for every $\tau \in [0,\infty)$, a globally Lipschitzian manifold of dimension n in $\mathbb{R}^n\times\mathbb{R}^n$. Moreover this set $\operatorname{gph}\partial V_\tau$ depends continuously on τ.*

The epigraphical continuity in the motion of V_τ in Theorem 2.1 thus corresponds to graphically continuity in the motion of ∂V_τ. Not just "continuous" aspects of this motion, but "differential" aspects need to be understood, however. For that purpose the Hamiltonian function H in (1.2) is an indispensable tool.

A better grasp of the nature of H under our assumptions is essential. Because $L(x,\cdot)$ is lsc, proper and convex under (A1) and (A2), the reciprocal formula in (1.3) does hold, and every property of L must accordingly have some exact counterpart for H. The following fact will be verified in §3. It describes the class of functions H such that, when L is defined from H by (1.3), L will be the unique Lagrangian for which (A1), (A2) and (A3) hold, and for which H is the associated Hamiltonian expressed by (1.2).

Theorem 2.3 (identification of the Hamiltonian class)**.** *A function $H : \mathbb{R}^n\times\mathbb{R}^n \to \overline{\mathbb{R}}$ is the Hamiltonian for a Lagrangian L satisfying (A1), (A2) and (A3) if and only if $H(x,y)$ is everywhere finite, concave in x, convex in y, and the following growth conditions hold, where (a) corresponds to (A3), and (b) corresponds to (A2):*

(a) *There are constants α and β and a finite, convex function φ such that*

$$H(x,y) \leq \varphi(y) + (\alpha|y|+\beta)|x| \quad \textit{for all } x,\ y.$$

(b) *There are constants γ and δ and a finite, concave function ψ such that*

$$H(x,y) \geq \psi(x) - (\gamma|x|+\delta)|y| \quad \textit{for all } x,\ y.$$

The finite concavity-convexity in Theorem 2.3 implies H is locally Lipschitz continuous on $\mathbb{R}^n \times \mathbb{R}^n$; cf. [5; §35].

Concave-convex Hamiltonian functions first surfaced as a significant class in connection with generalized problems of Bolza and Lagrange of convex type; cf. [2]. In the study of such problems, a subgradient form of Hamiltonian dynamics turned out to be crucial in characterizing optimality. Only subgradients of convex analysis are needed in expressing such dynamics. The generalized Hamiltonian *system* is

$$\dot{x}(t) \in \partial_y H(x(t), y(t)), \qquad -\dot{y}(t) \in \tilde{\partial}_x H(x(t), y(t)), \tag{2.6}$$

with $\partial_y H(x,y)$ the usual set of 'lower' subgradients of the convex function $H(x,\cdot)$ at y, but $\tilde{\partial}_x H(x,y)$ the analogously defined set of 'upper' subgradients of the concave function $H(\cdot,y)$ at x. A Hamiltonian *trajectory* over $[\tau_0,\tau_1]$ is an arc $(x(\cdot),y(\cdot)) \in \mathcal{A}^1_{2n}[\tau_0,\tau_1]$ that satisfies (2.6) for almost every t. The associated Hamiltonian *flow* is the one-parameter family of (generally) set-valued mappings S_τ for $\tau \geq 0$ defined by

$$S_\tau(\xi_0,\eta_0) := \big\{(\xi,\eta) \,\big|\, \exists \text{ Hamiltonian trajectory over } [0,\tau] \text{ from } (\xi_0,\eta_0) \text{ to } (\xi,\eta)\big\}. \tag{2.7}$$

Details and alternative expressions of the dynamics in (2.6) will be worked out in §6. Appropriate extensions to nonsmooth Hamiltonians $H(x,y)$ that aren't concave in x, and thus correspond to Lagrangians $L(x,v)$ that aren't jointly convex in x and v, can be found in [10] and [11]. Here, we confine ourselves to stating how, under our assumptions, the graph of the subgradient mapping ∂V_τ, namely

$$\operatorname{gph} \partial V_\tau := \big\{(\xi,\eta) \,\big|\, \eta \in \partial V_\tau(\xi)\big\} \subset \mathbb{R}^n \times \mathbb{R}^n, \tag{2.8}$$

evolves through such dynamics from the graph of the subgradient mapping $\partial V_0 = \partial g$.

Theorem 2.4 (Hamiltonian evolution of subgradients)**.** *Under (A0), (A1), (A2) and (A3), one has* $\eta \in \partial V_\tau(\xi)$ *if and only if, for some* $\eta_0 \in \partial g(\xi_0)$*, there exists a Hamiltonian trajectory* $(x(\cdot),y(\cdot))$ *over* $[0,\tau]$ *with* $(x(0),y(0)) = (\xi_0,\eta_0)$ *and* $(x(\tau),y(\tau)) = (\xi,\eta)$*. Thus, the graph of* ∂V_τ *is the image of the graph of* ∂g *under the flow mapping* S_τ*:*

$$\operatorname{gph} \partial V_\tau = S_\tau(\operatorname{gph} \partial g) \quad \textit{for all } \tau \geq 0. \tag{2.9}$$

Theorem 2.4 is the basis for a generalized *method of characteristics* for determining V uniquely from g and H. It will be proved in §6, where the method will be laid out in full. Especially noteworthy is the global nature of the description in Theorem 2.4, which is a by-product of convexity and underscores why the convex case deserves special attention. Classical forms of the method of characteristics are usually only local.

To go from the characterization in Theorem 2.4 to a description of the motion of V_τ in terms of a generalized Hamilton-Jacobi PDE, we need to bring subgradients beyond those of convex analysis. The notation and terminology of the book [6] will be adopted.

Consider any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and let x be any point at which $f(x)$ is finite. A vector $y \in \mathbb{R}^n$ is a *regular subgradient* of f at x, written $y \in \hat{\partial} f(x)$, if

$$f(x') \;\geq\; f(x) + \langle y,\, x' - x\rangle + o(|x' - x|). \tag{2.10}$$

It is a (*general*) *subgradient* of f at x, written $y \in \partial f(x)$, if there is a sequence of points $x^\nu \to x$ with $f(x^\nu) \to f(x)$ for which regular subgradients $y^\nu \in \hat{\partial} f(x^\nu)$ exist with $y^\nu \to y$.

These definitions refer to 'lower' subgradients, which are usually all that we need. To keep the notation uncluttered, we take 'lower' for granted, and in the few situations where 'upper' subgradient sets (analogously defined) are called for, we express them by

$$\tilde{\partial} f(x) = -\partial[-f](x), \qquad \tilde{\hat{\partial}} f(x) = -\hat{\partial}[-f](x). \tag{2.11}$$

For a convex function f, $\hat{\partial} f(x)$ and $\partial f(x)$ reduce to the subgradient set defined earlier through (2.4). In the case of the value function V, the "partial subgradient" notation

$$\partial_\xi V(\tau,\xi) = \big\{\eta \,\big|\, \eta \in \partial V_\tau(\xi)\big\} \;\text{ for }\; V_\tau = V(\tau,\cdot)$$

can thus, through Theorem 2.1, be interpreted equally in any of the senses above.

Theorem 2.5 (generalized Hamilton-Jacobi equation). *Under (A0), (A1), (A2) and (A3), the subgradients of V on $(0,\tau)\times I\!R^n$ have the property that*

$$
\begin{aligned}
(\sigma,\eta)\in\partial V(\tau,\xi) \quad &\Longleftrightarrow \quad (\sigma,\eta)\in\hat{\partial}V(\tau,\xi)\\
&\Longleftrightarrow \quad \eta\in\partial_\xi V(\tau,\xi),\ \ \sigma=-H(\xi,\eta).
\end{aligned}
\tag{2.12}
$$

In particular, therefore, V satisfies the generalized Hamilton-Jacobi equation:

$$
\sigma+H(\xi,\eta)=0 \quad \textit{for all} \quad (\sigma,\eta)\in\partial V(\tau,\xi) \quad \textit{when} \quad \tau>0. \tag{2.13}
$$

This theorem will be proved in §7. By virtue of the first equivalence in (2.12), the equation in (2.13) could be stated with $\hat{\partial}V(\tau,\xi)$ in place of $\partial V(\tau,\xi)$, but we prefer the ∂V version because of the dominance of general subgradients in so much of the variational analysis and subdifferential calculus in [6]. The $\hat{\partial}V$ version would effectively turn (2.13) into the one-sided 'viscosity' form of Hamilton-Jacobi equation used for lsc functions by Barron and Jensen [12] and Frankowska [8], in distinction to earlier forms for continuous functions that called for pairs of inequalities, cf. Crandall, Evans and Lions [13]. The book of Bardi and Capuzzo-Dolcetta [9] gives a broad picture of viscosity theory in its current state, including the relationships between such different forms.

The extent to which the generalized Hamiltonian equation (2.13) (or its viscosity version), along with the initial condition (interpreted as the epi-convergence of V_τ to g as $\tau\searrow 0$), might suffice to determine V uniquely, isn't yet understood in the framework we have adopted. The strongest result so far available for lsc solutions to (2.13), allowing $V(\tau,\xi)$ to take on ∞ when $\tau>0$, is that of Frankowska [8]. Among problems satisfying our convexity assumptions, however, it only covers ones in which $H(x,y)=\langle Ax,y\rangle+h(y)$ for some matrix A and finite, convex function h that is positively homogeneous, or equivalently, $L(x,v)$ is the indicator $\delta_C(v-Ax)$ corresponding to a differential inclusion $\dot{x}(t)\in Ax(t)+C$ with C a nonempty, compact, convex set.

The arcs $y(\cdot)$ that are paired with the arcs $x(\cdot)$ in the Hamiltonian dynamics are related to the forward propagation of the conjugate initial function g^*, satisfying

$$
g^*(y):=\sup_x\Big\{\langle x,y\rangle-g(x)\Big\},\qquad g(x):=\sup_y\Big\{\langle x,y\rangle-g^*(y)\Big\}, \tag{2.14}
$$

with respect to the *dual* Lagrangian $\tilde{L}$, satisfying

$$
\begin{aligned}
\tilde{L}(y,w)&=L^*(w,y)=\sup_{x,v}\Big\{\langle x,w\rangle+\langle v,y\rangle-L(x,v)\Big\},\\
L(x,v)&=\tilde{L}^*(v,x)=\sup_{y,w}\Big\{\langle x,w\rangle+\langle v,y\rangle-\tilde{L}(y,w)\Big\}.
\end{aligned}
\tag{2.15}
$$

The reciprocal formulas here follow from (A0) and (A1). We'll prove in §5 that the value function $\tilde{V}$ defined as in (1.1), but with g^* and $\tilde{L}$ in place of g and L, has $\tilde{V}_\tau$ conjugate to V_τ for every τ. This duality will be a workhorse in our analysis of other basic properties.

A virtue of our assumptions (A0), (A1), (A2) and (A3) is that they carry over symmetrically to the dual setting. Alternative bundles of assumptions could fail to accomplish this. To put this another way, the class of Hamiltonians that we work with, as described in Theorem 2.3, is no accident, but carefully tuned to obtaining the broadest possible results of duality in Hamilton-Jacobi theory, at least with respect to time-independent Hamiltonians.

3. Elaboration of the Convexity and Growth Conditions

Conditions (A1), (A2) and (A3) can be viewed from several different angles, and a better understanding of them is required before we can proceed. Their Hamiltonian translation in Theorem 2.3 has to be verified, but they be also useful as applied to functions other than L. A broader, not merely Lagrangian, perspective on them must be attained.

We'll draw on some basic concepts of variational analysis, and convex analysis in particular. For any nonempty subset $C \subset I\!R^n$, the *horizon cone* is

$$C^\infty := \limsup_{\lambda \searrow 0} \lambda C = \left\{ w \in I\!R^n \,\middle|\, \exists x^\nu \in C,\ \lambda^\nu \searrow 0, \text{ with } \lambda^\nu x^\nu \to w \right\}.$$

This is always a closed cone. When C convex, it is a convex cone and, for any $\bar{x} \in \operatorname{ri} C$ (the relative interior of C) it consists simply of the vectors w such that $\bar{x} + \lambda w \in C$ for all $\lambda > 0$. When C is convex and closed, C^∞ coincides with the "recession cone" of C. See [5, §6], [6, Chap. 3].

For any function $f : I\!R^n \to \overline{I\!R}$, $f \not\equiv \infty$, the *horizon function* f^∞ is the function having as its epigraph the set $(\operatorname{epi} f)^\infty$, where $\operatorname{epi} f$ is the epigraph of f itself. This function is always lsc and positively homogeneous. When f is convex, f^∞ is convex as well and, for any $\bar{x} \in \operatorname{ri}(\operatorname{dom} f)$, is given by $f^\infty(w) = \lim_{\lambda \to \infty} f(\bar{x} + \lambda w)/\lambda$. When f is convex and lsc, f^∞ is the "recession function" of f in convex analysis. Again, see [5, §6], [6, Chap. 3].

It will be important in the context of conditions (A1), (A2) and (A3) to view L not just as a function on $I\!R^n \times I\!R^n$ but in terms of the associated function-valued mapping $x \mapsto L(x, \cdot)$ that assigns to each $x \in I\!R^n$ the function $L(x, \cdot) : I\!R^n \to \overline{I\!R}$. A function-valued mapping is a 'bifunction' in the terminology of [5].

Definition 3.1 [4] (regular convex bifunctions)**.** *A function-valued mapping from $I\!R^n$ to the space of extended-real-valued functions on $I\!R^n$, as specified in the form $x \mapsto \Lambda(x, \cdot)$ by a function $\Lambda : I\!R^n \times I\!R^n \to \overline{I\!R}$, is called a regular convex bifunction if*

(a1) *Λ is proper, lsc and convex as a function on $I\!R^n \times I\!R^n$;*
(a2) *for each $w \in I\!R^n$ there is a $z \in I\!R^n$ with $(w, z) \in (\operatorname{dom} \Lambda)^\infty$;*
(a3) *there is no $z \neq 0$ with $(0, z) \in \operatorname{cl}(\operatorname{dom} \Lambda^\infty)$.*

Proposition 3.2 [4] (bifunction duality)**.** *For $\Lambda : I\!R^n \times I\!R^n \to \overline{I\!R}$, suppose that the mapping $x \mapsto \Lambda(x, \cdot)$ is a regular convex bifunction. Then for the conjugate function $\Lambda^* : I\!R^n \times I\!R^n \to \overline{I\!R}$, the mapping $y \mapsto \Lambda^*(\cdot, y)$ is a regular convex bifunction.*

Indeed, conditions (a2) and (a3) of Definition 3.1 are dual to each other in the sense that, under (a1), the first mapping satisfies (a2) if and only if the second satisfies (a3), whereas the first satisfies (a3) if and only if the second satisfies (a2).

Proof. This was shown as part of Theorem 4 of [4]; for the duality between (a2) and (a3), see the proof of that theorem. □

Lemma 3.3 [4] (domain selections)**.** *For a function $\Lambda : I\!R^n \times I\!R^n \to \overline{I\!R}$ satisfying condition (a1) of Definition 3.1, condition (a2) is equivalent to the existence of a matrix $A \in I\!R^{n \times n}$ and vectors $a \in I\!R^n$ and $b \in I\!R^n$ such that*

$$\big(x,\ Ax + |x|a + b\big) \in \operatorname{ri}(\operatorname{dom} \Lambda) \ \textit{for all}\ x \in I\!R^n. \tag{3.1}$$

Proof. See the first half of the proof of Theorem 5 of [4] for the necessity. The sufficiency is clear because (3.1) implies $\big(x,\ Ax + |x|a\big) \in (\operatorname{dom} \Lambda)^\infty$ for all $x \in I\!R^n$. □

Proposition 3.4 (Lagrangian growth characterization)**.** *A function $L : I\!R^n \times I\!R^n \to \overline{I\!R}$ satisfies (A1), (A2) and (A3) if and only if the mapping $x \mapsto L(x, \cdot)$ is a regular convex bifunction. Specifically in the context of Definition 3.1 with $\Lambda = L$, (A1) corresponds to (a1), and then one has the equivalence of (A2) with (a2) and that of (A3) with (a3).*

Proof. When $\Lambda = L$, (A1) is identical to (a1). Assuming this property now, we argue the other equivalences.

(A2) $\Rightarrow$ (a2). For any $w \in I\!R^n$ and any integer $\nu > 0$ there exists by (A2) some $v^\nu \in F(\nu w)$ with $|v^\nu| \leq \rho(1 + \nu|w|)$. Let $x^\nu = \nu w$ and $\lambda^\nu = 1/\nu$. We have $(x^\nu, v^\nu) \in \operatorname{dom} L = \operatorname{dom} \Lambda$ and $\lambda^\nu(x^\nu, v^\nu) = (w, (1/\nu)v^\nu)$ with $(1/\nu)|v^\nu| \leq \rho(1 + |w|)$. The sequence of pairs $\lambda^\nu(x^\nu, v^\nu)$ is therefore bounded in $I\!R^n \times I\!R^n$ and has a cluster point, which necessarily is of the form (w,z) for some $z \in \mathbb{R}^n$. Furthermore $(w,z) \in (\operatorname{dom}\Lambda)^\infty$ by definition. Thus, (a2) is fulfilled.

(a2) $\Rightarrow$ (A2). Applying Lemma 3.3, we get the existence of a matrix A and vectors a and b such that $Ax + |x|a + b \in F(x)$ for all x. Then $\operatorname{dist}(0, F(x)) \le |A||x| + |x||a| + |b|$, so we can get the bound in (A2) by taking $\rho \ge \max\{|b|, |A| + |a|\}$.

(A3) $\Rightarrow$ (a3). Let $(\bar{x}, \bar{v}) \in \operatorname{ri}(\operatorname{dom} L) = \operatorname{ri}(\operatorname{dom}\Lambda)$. For any (w,z) we have $\Lambda^\infty(w,z) = \lim_{\lambda\to\infty} \Lambda(\bar{x} + \lambda w, \bar{v} + \lambda z)/\lambda$. On the basis of (A3) this yields

$$\begin{aligned}
\Lambda^\infty(w,z) &\ge \lim_{\lambda\to\infty} \lambda^{-1}\big[\Lambda\big([|\bar{v}+\lambda z| - \alpha|\bar{x}+\lambda z|]_+\big) - \beta|\bar{x}+\lambda w|\big] \\
&= \lim_{\lambda\to\infty} \big[\lambda^{-1}\Lambda\big(\lambda[|\lambda^{-1}\bar{v}+z| - \alpha|\lambda^{-1}\bar{x}+z|]_+]\big)\big] - \beta|\lambda^{-1}x + z|\big] \\
&= \begin{cases} -\beta|w| & \text{if } [|z| - \alpha|w|]_+ = 0, \\ \infty & \text{if } [|z| - \alpha|w|]_+ > 0. \end{cases}
\end{aligned}$$

Hence $\operatorname{dom}\Lambda^\infty \subset \big\{(z,w)\big\{|z| \le \alpha|w|\big\}$. Any $(0,z) \in \operatorname{cl}(\operatorname{dom}\Lambda^\infty)$ then has $|z| \le \alpha|0|$, hence $z = 0$, so (a3) holds.

(a3) $\Rightarrow$ (A3). According to Proposition 3.2, condition (a3) on the mapping $x \mapsto \Lambda(x,\cdot)$ is equivalent to condition (a2) on the mapping $y \mapsto \Lambda^*(\cdot, y)$. By Lemma 3.3, the latter provides the existence of a matrix A and vectors a and b such that

$$(Ay + |y|a + b, y) \in \operatorname{ri}(\operatorname{dom}\Lambda^*) \text{ for all } y \in \mathbb{R}^n.$$

Any convex function is continuous over the relative interior of its effective domain, so the function $y \mapsto \Lambda^*(Ay + |y|a + b, y)$ is (finite and) continuous on $\mathbb{R}^n$ (although not necessarily convex). Define the function ψ on $[0,\infty)$ by $\psi(r) = \max\big\{\Lambda^*(Ay + |y|a + b, y) \,\big|\, |y| \le r\big\}$. Then ψ is finite, continuous and nondecreasing. Because

$$\Lambda(x,v) = \Lambda^{**}(x,v) = \sup_{z,y}\Big\{\langle x, z\rangle + \langle v, y\rangle - \Lambda^*(z,y)\Big\}$$

under (a1), we have

$$\begin{aligned}
\Lambda(x,v) &\ge \sup_y\Big\{\langle x,\ Ay + |y|a + b\rangle + \langle v, y\rangle - \Lambda^*(Ay + |y|a + b, y)\Big\} \\
&\ge \sup_y\Big\{-|x|(|A||y| + |y||a| + |b|) + \langle v, y\rangle - \psi(|y|)\Big\} \\
&= \sup_y\Big\{-|x||y|(|A| + |a|) - |x||b| + |v||y| - \psi(|y|)\Big\} \\
&= -|x||b| + \sup_{r\ge 0}\Big\{r\big[\,|v| - (|A| + |a|)|x|\,\big] - \psi(r)\Big\} \\
&= \psi^*\big([\,|v| - (|A| + |a|)|x|\,]_+\big) - |b||x|.
\end{aligned}$$

Let $\alpha = |A| + |a|$, $\beta = |b|$ and $\theta = \psi^*$ on $[0,\infty)$. Then the inequality in (A3) holds for $L = \Lambda$. The function θ has $\theta(0) = -\psi(0)$ (finite) and is the pointwise supremum of a collection of affine functions of the form $s \mapsto rs - \psi(r)$ with $r \ge 0$ and $\psi(r)$ always finite. Hence θ is convex, proper, nondecreasing and in addition has $\lim_{s\to\infty} \theta(s)/s \ge r$ for all $r \ge 0$, which implies coercivity. ◻

Proposition 3.5 (Lagrangian dualization). *If the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ satisfies (A1), (A2) and (A3), then so too does the dual Lagrangian $\tilde{L} : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ in (2.15). Indeed, (A1) for L yields (A1) for $\tilde{L}$ and the reciprocal formula in (2.15), and then (A2) for L corresponds to (A3) for $\tilde{L}$, whereas (A3) for L corresponds to (A2) for $\tilde{L}$. Furthermore, the dual Hamiltonian*

$$\tilde{H}(y,x) := \sup_w\Big\{\langle x, w\rangle - \tilde{L}(y,w)\Big\} \tag{3.2}$$

associated with $\tilde{L}$ is then related to the Hamiltonian H for L by

$$\tilde{H}(y,x) = -H(x,y). \tag{3.3}$$

Proof. Combine Proposition 3.4 with Proposition 3.2 to get the dualization of (A1), (A2) and (A3) to $\tilde{L}$. Note next that since $L(x,\cdot)$ is by (A1), (A2) and (A3) a proper, lsc, convex and coercive function on $I\!R^n$, its conjugate function, which is $H(x,\cdot)$, is finite on $I\!R^n$. The joint convexity of $L(x,v)$ in x and v corresponds to $H(x,y)$ being not just convex in y, as always, but also concave in x; see [5; 33.3] or [6; 11.48]. For the Hamiltonian relationship in (3.3), observe through (2.15) and the formula (1.2) for H that

$$\tilde{L}(y,w) = \sup_x \Big\{ \langle x,w\rangle + H(x,y) \Big\}. \tag{3.4}$$

Fix any y and let $h = -H(\cdot,y)$, noting that h is a finite convex function on $I\!R^n$. According to (3.4), we have $L(y,\cdot) = h^*$, and from (3.3) we then have $h^{**} = \tilde{H}(y,\cdot)$. The finiteness and convexity of h ensures that $h^{**} = h$, so that $\tilde{H}(y,\cdot) = -H(\cdot,y)$ as claimed in (3.3). $\square$

Proof of Theorem 2.3. Finite convex functions correspond under the Legendre-Fenchel transform to the proper convex functions that are coercive. Having $H(x,\cdot)$ be a finite convex function on $I\!R^n$ for each $x \in I\!R^n$ is equivalent therefore to having H be the Hamiltonian associated by (1.2) with a Lagrangian L such that $L(x,\cdot)$ is, for each $x \in I\!R^n$, a proper, convex function that is coercive; the function L is recovered from H by (1.3). Concavity of $H(x,y)$ in x corresponds then to joint convexity of $L(x,v)$ in x and v, as already pointed out in the proof of Proposition 3.5; see [5; 33.3] or [6; 11.48].

Thus in particular, any finite, concave-convex function H is the Hamiltonian for some Lagrangian L satisfying (A1), while on the other hand, if L satisfies (A3) along with (A1) (and therefore has $L(x,\cdot)$ always coercive) its Hamiltonian H is finite concave-convex.

It will be demonstrated next that in the case of a Lagrangian L satisfying (A1), condition (A3) is equivalent to the growth condition in (a). This will yield through the duality in Proposition 3.5 the equivalence (A2) with the growth condition in (b), and all claims will thereby be justified. Starting with (a), define $\psi(r) = \max\{\varphi(y) \,\big|\, |y| \le r\}$ to get a finite, nondecreasing, convex function ψ on $[0,\infty)$. The inequality in (a) yields $H(x,y) \le \psi(|y|) + (\alpha|y|+\beta)|x|$ and consequently through (1.3) that

$$\begin{aligned} L(x,v) &\ge \sup_y \Big\{ \langle v,y\rangle - \psi(|y|) - (\alpha|y|+\beta)|x| \Big\} \\ &= \sup_{r\ge 0} \sup_{|y|\le r} \Big\{ \langle v,y\rangle - \psi(|y|) - (\alpha|y|+\beta)|x| \Big\} \\ &= \sup_{r\ge 0} \Big\{ |v|r - \psi(r) - (\alpha r+\beta)|x| \Big\} = \psi^*\big([\,|v| - \alpha|x|\,]_+\big) - \beta|x|, \end{aligned}$$

where ψ^* is coercive, proper and nondecreasing on $[0,\infty)$. Taking $\theta = \psi^*$, we get (A3).

Conversely from (A3), where it can be assumed without loss of generality that $\alpha \ge 0$, we can retrace this pattern by estimating through (1.2) that

$$\begin{aligned} H(x,y) &\le \sup_v \Big\{ \langle v,y\rangle - \theta\big(\big[|v| - \alpha|x|\big]_+\big) - \beta|x| \Big\} \\ &= \sup_{s\ge 0} \sup_{|v|\le s} \Big\{ \langle v,y\rangle - \theta\big([\,|v| - \alpha|x|\,]_+\big) - \beta|x| \Big\} \\ &= \sup_{s\ge 0} \Big\{ s|y| - \theta\big([s - \alpha|x|\,]_+\big) - \beta|x| \Big\}, \end{aligned}$$

and on changing to the variable $r = s - \alpha|x|$ obtain

$$\begin{aligned} H(x,y) &\le \sup_{r\ge -\alpha|x|} \Big\{ (r+\alpha|x|)|y| - \theta\big([r]_+\big) - \beta|x| \Big\} \\ &= \sup_{r\ge 0} \Big\{ r|y| - \theta(r) \Big\} + (\alpha|y|+\beta)|x| = \theta^*(|y|) + (\alpha|y|+\beta)|x|, \end{aligned}$$

where θ^* is finite, convex and nondecreasing. The function $\varphi(y) = \theta^*(|y|)$ is then convex on $I\!R^n$ (see [5; 15.3] or [6; 11.21]). Thus, we have the growth condition in (a). $\square$

4. Consequences for Bolza Problem Duality

The properties we have put in place for L and H lead to stronger results about duality for the generalized problems of Bolza of convex type. These improvements, which we lay out next, will be a platform for our investigation of value function duality in §5.

The duality theory in [1] and [3], as expressed over a fixed interval $[0,\tau]$, centers (in the autonomous case) on a problem of the form

$$(\mathcal{P}) \qquad \text{minimize } J\big(x(\cdot)\big) := \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt + l\big(x(0),x(\tau)\big) \text{ over } x(\cdot)\in\mathcal{A}^1_n[0,\tau],$$

where the endpoint function $l : I\!R^n \times I\!R^n \to \overline{I\!R}$ is proper, lsc and convex, and the corresponding dual problem

$$(\tilde{\mathcal{P}}) \qquad \text{minimize } \tilde{J}\big(y(\cdot)\big) := \int_0^\tau \tilde{L}\big(y(t),\dot{y}(t)\big)dt + \tilde{l}\big(y(0),y(\tau)\big) \text{ over } y(\cdot)\in\mathcal{A}^1_n[0,\tau],$$

where the dual endpoint function $\tilde{l} : I\!R^n \times I\!R^n \to \overline{I\!R}$ is generated through conjugacy:

$$\begin{aligned} \tilde{l}(\eta,\eta') &= l^*(\eta,-\eta') = \sup_{\xi',\xi}\Big\{\langle\eta,\xi'\rangle - \langle\eta',\xi\rangle - l(\xi',\xi)\Big\}, \\ l(\xi',\xi) &= \tilde{l}^*(\xi',-\xi) = \sup_{\eta,\eta'}\Big\{\langle\eta,\xi'\rangle - \langle\eta',\xi\rangle - \tilde{L}(\eta',\xi)\Big\}. \end{aligned} \tag{4.1}$$

A major role in characterizing optimality in the generalized Bolza problems $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ is played by the *generalized Euler-Lagrange condition*

$$(\dot{y}(t),y(t)) \in \partial L(x(t),\dot{x}(t)) \text{ for a.e. } t, \tag{4.2}$$

which can also be written in the dual form $(\dot{x}(t),x(t)) \in \partial\tilde{L}(y(t),\dot{y}(t))$ for a.e. t. The Euler-Lagrange conditions are know to be equivalent in turn to the *generalized Hamiltonian condition* (2.6) being satisfied over the time interval $[0,\tau]$; cf. [2]. They act in combination with the *generalized transversality condition*

$$(y(0),-y(\tau)) \in \partial l(x(0),x(\tau)), \tag{4.3}$$

which likewise has an equivalent dual form, $(x(0),-x(\tau)) \in \partial\tilde{l}(y(0),y(\tau))$. The basic facts about optimality are the following.

Theorem 4.1 [1]**,** [2] (optimality conditions)**.** *For any functions L and l that are proper, lsc and convex on $I\!R^n \times I\!R^n$, the optimal values in $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ satisfy $\inf(\mathcal{P}) \le -\inf(\tilde{\mathcal{P}})$. Moreover, for arcs $x(\cdot)$ and $y(\cdot)$ in $\mathcal{A}^1_n[0,\tau]$, the following properties are equivalent:*

(a) *$(x(\cdot),y(\cdot))$ is a Hamiltonian trajectory satisfying the transversality condition;*

(b) *$x(\cdot)$ solves $(\mathcal{P})$, $y(\cdot)$ solves $(\tilde{\mathcal{P}})$, and $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$.*

Proof. Basically this is Theorem 5 of [1], but we've used Theorem 1 of [2] to translate the Euler-Lagrange condition to the Hamiltonian condition. □

Theorem 4.1 gives us the sufficiency of the Hamiltonian condition and transversality condition for optimality of arcs in $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$, but not the necessity. We can get that to the extent we are able to establish that optimal arcs do exist for these problems, and $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$. Criteria for that have been furnished in [3] in terms of certain "constraint qualifications," but this is where we can make improvements now in consequence of our working assumptions.

The issue concerns the *fundamental* function $E : [0,\infty) \times I\!R^n \times I\!R^n \to \overline{I\!R}$ defined for the Lagrangian L by

$$\begin{aligned} E(\tau,\xi',\xi) &:= \inf\Big\{ \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(0)=\xi',\ x(\tau)=\xi \Big\}, \\ E(0,\xi',\xi) &:= \begin{cases} 0 & \text{if } \xi'=\xi, \\ \infty & \text{if } \xi'\ne\xi, \end{cases} \end{aligned} \tag{4.4}$$

where the minimization is over all arcs $x(\cdot) \in \mathcal{A}^1_n[0,\tau]$ satisfying the initial and terminal conditions. At the same time it concerns fundamental function $\tilde{E} : [0,\infty) \times I\!R^n \times I\!R^n \to \overline{I\!R}$ associated with the dual Lagrangian $\tilde{L}$,

$$\begin{aligned} \tilde{E}(\tau,\eta',\eta) &:= \inf\Big\{ \int_0^\tau \tilde{L}\big(y(t),\dot{y}(t)\big)dt \,\Big|\, y(0)=\eta',\ y(\tau)=\eta \Big\}, \\ \tilde{E}(0,\eta',\eta) &:= \begin{cases} 0 & \text{if } \eta'=\eta, \\ \infty & \text{if } \eta' \neq \eta, \end{cases} \end{aligned} \tag{4.5}$$

with the minimization taking place over $y(\cdot) \in \mathcal{A}^1_n[0,\tau]$. The constraint qualifications in [3] are stated in terms of the sets

$$C_\tau := \big\{(\xi',\xi) \,\big|\, E(\tau,\xi',\xi) < \infty\big\}, \qquad \tilde{C}_\tau := \big\{(\eta',\eta) \,\big|\, \tilde{E}(\tau,\eta',\eta) < \infty\big\}. \tag{4.6}$$

They revolve around the overlap between these sets and the sets $\operatorname{dom} l$ and $\operatorname{dom} \tilde{l}$. In this respect the next result provides vital information.

Proposition 4.2 (growth of the fundamental function)**.** *Suppose (A1), (A2) and (A3) hold. Then the following properties of $E(\tau,\cdot,\cdot)$ hold for all $\tau \geq 0$ and guarantee that for all ξ and ξ' the functions $E(\tau,\xi',\cdot)$ and $E(\tau,\cdot,\xi)$ are proper, lsc, convex and coercive:*

(a) *$E(\tau,\cdot,\cdot)$ is proper, lsc and convex on $I\!R^n \times I\!R^n$.*

(b) *There is a constant $\rho(\tau) \in (0,\infty)$ such that*

$$\begin{aligned} \operatorname{dist}\big(0, \operatorname{dom} E(\tau,\xi',\cdot)\big) \;&\leq\; \rho(\tau)(1+|\xi'|) \ \textit{ for all } \xi' \in I\!R^n, \\ \operatorname{dist}\big(0, \operatorname{dom} E(\tau,\cdot,\xi)\big) \;&\leq\; \rho(\tau)(1+|\xi|) \ \textit{ for all } \ \xi \in I\!R^n. \end{aligned}$$

(c) *There are constants $\alpha(\tau)$, $\beta(\tau)$, and a coercive, proper, nondecreasing function $\theta(\tau,\cdot)$ on $[0,\infty)$ such that*

$$\left.\begin{aligned} E(\tau,\xi',\xi) \;&\geq\; \theta\big(\tau, [\,|\xi| - \alpha(\tau)|\xi'|\,]_+\big) - \beta(\tau)|\xi'| \\ E(\tau,\xi',\xi) \;&\geq\; \theta\big(\tau, [\,|\xi'| - \alpha(\tau)|\xi|\,]_+\big) - \beta(\tau)|\xi| \end{aligned}\right\} \ \textit{for all } \xi',\xi \in I\!R^n.$$

Proof. When the mapping $x \mapsto L(x,\cdot)$ is a regular convex bifunction, both of the mappings $\xi' \mapsto E(\tau,\xi',\cdot)$ and $\xi \mapsto E(\tau,\cdot,\xi)$ are regular convex bifunctions as well, for all $\tau \geq 0$. For $\tau > 0$, this was proved as part of Theorem 5 of [4]. For $\tau = 0$, it is obvious from formula (4.5). On this basis we can appeal to Proposition 3.2 for each of the three function-valued mappings. In the conditions in (a) and (b), we get separate constants to work for $E(\tau,\xi',\cdot)$ and $E(\tau,\cdot,\xi)$, but then by taking a max can get constants that work simultaneously for both, so as to simplify the statements. ◻

Corollary 4.3 (growth of the dual fundamental function)**.** *When L satisfies (A1), (A2) and (A3), the function $\tilde{E}$ likewise has the properties in Proposition 4.2.*

Proof. Apply Proposition 4.2 to $\tilde{L}$ instead of L, using the fact from Proposition 3.5 that $\tilde{L}$, like L, satisfies (A1), (A2) and (A3). ◻

Corollary 4.4 (reachable endpoint pairs)**.** *Under (A1), (A2) and (A3), the sets C_τ and $\tilde{C}_\tau$ in (4.6) have the following property for every $\tau > 0$:*

(a) *The image of C_τ under the projection $(\xi',\xi) \mapsto \xi'$ is all of $I\!R^n$. Likewise, the image of C_τ under the projection $(\xi',\xi) \mapsto \xi$ is all of $I\!R^n$.*

(b) *The image of $\tilde{C}_\tau$ under the projection $(\eta',\eta) \mapsto \eta'$ is all of $I\!R^n$. Likewise, the image of $\tilde{C}_\tau$ under the projection $(\eta',\eta) \mapsto \eta$ is all of $I\!R^n$.*

Proof. We get (a) from the property in Proposition 4.2(b). We get (b) then out of the preceding corollary. ◻

Some generalizations of the conditions in Proposition 4.2 to the case of functions E coming from Lagrangians L that are not fully convex are available in [14].

Theorem 4.5 (strengthened duality for Bolza problems). *Consider* $(\mathcal{P})$ *and* $(\tilde{\mathcal{P}})$ *under the assumption that the Lagrangian* L *satisfies (A1), (A2) and (A3), whereas the endpoint function* l *is proper, lsc and convex.*

(a) *If there exists* ξ *such that* $l(\cdot,\xi)$ *is finite, or there exists* ξ' *such that* $l(\xi',\cdot)$ *is finite, then* $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$*. This value is not* ∞*, and if it also is not* $-\infty$ *there is an optimal arc* $y(\cdot)\in\mathcal{A}^1_n[0,\tau]$ *for* $(\tilde{\mathcal{P}})$*. In particular the latter holds if an optimal arc* $x(\cdot)\in\mathcal{A}^1_n[0,\tau]$ *exists for* $(\mathcal{P})$*, and in that case both* $x(\cdot)$ *and* $y(\cdot)$ *must actually belong to* $\mathcal{A}^\infty_n[0,\tau]$*.*

(b) *If there exists* η *such that* $\tilde{l}(\eta,\cdot)$ *is finite, or there exists* η' *such that* $\tilde{l}(\cdot,\eta')$ *is finite, then* $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$*. This value is not* $-\infty$*, and if it also is not* ∞ *there is an optimal arc* $x(\cdot)\in\mathcal{A}^1_n[0,\tau]$ *for* $(\mathcal{P})$*. In particular the latter holds if an optimal arc* $y(\cdot)\in\mathcal{A}^1_n[0,\tau]$ *exists for* $(\tilde{\mathcal{P}})$*, and in that case both* $x(\cdot)$ *and* $y(\cdot)$ *must actually belong to* $\mathcal{A}^\infty_n[0,\tau]$*.*

Proof. Theorem 1 of [3] will be our vehicle. The conditions referred to as (C_0) and (D_0) in the statement of that result are fulfilled in the case of a finite, time-independent Hamiltonian (cf. p. 11 of [3]), which we have here via Theorem 2.3 (already proved in §3).

If l satisfies one of the conditions in (a), it is impossible in the face of Corollary 4.4(a) for there to exist a hyperplane that separates the convex sets $\operatorname{dom} l$ and $\operatorname{dom} E(\tau,\cdot,\cdot)$. By separation theory (cf. [5,§11]), this is equivalent to having $\operatorname{ri} C_\tau \cap \operatorname{ri}\operatorname{dom} l \neq \emptyset$ and $\operatorname{aff} C_\tau \cup \operatorname{dom} l = I\!R^n \times I\!R^n$, where 'ri' is relative interior as earlier and 'aff' denotes affine hull. According to part (b) of Theorem 1 of [3], this pair of conditions guarantees that $\inf(\mathcal{P})$ and $-\inf(\tilde{\mathcal{P}})$ have a common value which is not ∞, and that if this value is also not $-\infty$, then $(\tilde{\mathcal{P}})$ has a solution $y(\cdot)\in\mathcal{A}^1_n[0,\tau]$. We know on the other hand that whenever $\inf(\mathcal{P})<\infty$ and $(\mathcal{P})$ has a solution $x(\cdot)\in\mathcal{A}^1_n[0,\tau]$, we have $J(x(\cdot))$ finite in $(\mathcal{P})$ (because neither l nor the integral functional in (2.1) can take on $-\infty$), so that $\inf(\mathcal{P})$ is finite. It follows then from Theorem 4.1 that $x(\cdot)$ and $y(\cdot)$ satisfy the generalized Hamiltonian condition, i.e., (2.6). Because H is finite everywhere, this implies by Theorem 2 of [2] that these arcs belong to $\mathcal{A}^\infty_n[0,\tau]$. This proves (a). The claims in (b) are justified in parallel by way of Corollary 4.4(b) and part (a) of Theorem 1 of [3]. $\square$

Corollary 4.6 (best-case Bolza duality). *Consider* $(\mathcal{P})$ *and* $(\tilde{\mathcal{P}})$ *under the assumption that* L *satisfies (A1), (A2) and (A3), whereas* l *is proper, lsc and convex. Suppose* l *has one of the finiteness properties in Theorem 4.5(a), while* $\tilde{l}$ *has one of the finiteness properties in Theorem 4.5(b). Then* $-\infty<\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})<\infty$*, and optimal arcs* $x(\cdot)$ *and* $y(\cdot)$ *exist for* $(\mathcal{P})$ *and* $(\tilde{\mathcal{P}})$*. Moreover, any such arcs must belong to* $\mathcal{A}^\infty_n[0,\tau]$*.*

Proof. This simply combines the conclusions in parts (a) and (b) of Theorem 4.5. $\square$

5. Value Function Duality

The topic we treat next is the relationship between V and the dual value function $\tilde{V}$ generated by $\tilde{L}$ and g^*:

$$\tilde{V}(\tau,\eta) := \inf\Big\{ g^*\big(y(0)\big) + \int_0^\tau \tilde{L}\big(y(t),\dot{y}(t)\big)dt \,\Big|\, y(\tau)=\eta\Big\}, \qquad \tilde{V}(0,\eta)=g^*(\eta), \tag{5.1}$$

where the minimum is taken over all arcs $y(\cdot)\in\mathcal{A}^1_n[0,\tau]$. Henceforth we take assume (A0), (A1), (A2) and (A3) without further mention. Because $\tilde{L}$ and g^* inherit these properties from L and g, everything we prove about V automatically holds in parallel form for $\tilde{V}$.

It will be helpful for our endeavor to note that V can be expressed in terms of E. Indeed, from the definitions of V and E in (1.1) and (4.4) it's easy to deduce the rule that

$$V(\tau,\xi) = \inf_{\xi'}\Big\{V(\tau',\xi') + E(\tau-\tau',\xi',\xi)\Big\} \ \text{ for } 0\le\tau'\le\tau. \tag{5.2}$$

By the same token we also have, through (5.1) and (4.5), that

$$\tilde{V}(\tau,\eta) = \inf_{\eta'}\Big\{\tilde{V}(\tau',\eta') + \tilde{E}(\tau-\tau',\eta',\eta)\Big\} \ \text{ for } 0\le\tau'\le\tau. \tag{5.3}$$

Theorem 5.1 (conjugacy). *For each* $\tau\ge 0$*, the functions* $V_\tau := V(\tau,\cdot)$ *and* $\tilde{V}_\tau := \tilde{V}(\tau,\cdot)$ *are proper and conjugate to each other under the Legendre-Fenchel transform:*

$$\tilde{V}_\tau(\eta) = \sup_\xi\Big\{\langle\xi,\eta\rangle - V_\tau(\xi)\Big\}, \qquad V_\tau(\xi) = \sup_\eta\Big\{\langle\xi,\eta\rangle - \tilde{V}_\tau(\eta)\Big\}. \tag{5.4}$$

Hence in particular, the subgradients of these convex functions are related by

$$\eta \in \partial V_\tau(\xi) \iff \xi \in \partial \tilde{V}_\tau \iff V_\tau(\xi) + \tilde{V}_\tau(\eta) = \langle \xi, \eta \rangle. \tag{5.5}$$

Proof. Fix $\tau > 0$ and any vector $\bar{\eta} \in \mathbb{R}^n$. Let $l(\xi', \xi) = g(\xi') - \langle \xi, \bar{\eta} \rangle$. The corresponding dual endpoint function $\tilde{l}$ has $\tilde{l}(\eta', \eta) = g(\eta')$ when $\eta = \bar{\eta}$, but $\tilde{l}(\eta', \eta) = \infty$ when $\eta \neq \bar{\eta}$. In the Bolza problems we then have

$$-\inf(\mathcal{P}) = \sup_\xi \big\{ \langle \xi, \bar{\eta} \rangle - V(\tau, \xi) \big\}, \qquad \inf(\tilde{\mathcal{P}}) = \tilde{V}(\tau, \bar{\eta}). \tag{5.6}$$

Because $\operatorname{dom} l$ has the form $C \times \mathbb{R}^n$ for a nonempty convex set C, namely $C = \operatorname{dom} g$, the constraint qualification of Theorem 4.5(a) is satisfied, and we may conclude that $-\inf(\tilde{\mathcal{P}}) = \inf(\mathcal{P}) > -\infty$. This yields the first equation in (5.4)—in the case of $\eta = \bar{\eta}$—and ensures that $V_\tau \not\equiv \infty$ and $\tilde{V}_\tau > -\infty$ everywhere. By the symmetry between $(\tilde{L}, g^*)$ and (L, f), we get second equation in (5.4) along with $\tilde{V}_\tau \not\equiv \infty$ and $V_\tau > -\infty$ everywhere.

The subgradient relation translates to this context a property that is known for subgradients of conjugate convex functions in general; cf. [5; 11.3]. □

Proof of Theorem 2.1. Through the conjugacy in Theorem 5.1, we see at once that V_τ is convex and lsc, and of course the same for $\tilde{V}_\tau$. The remaining task is to demonstrate the epi-continuity property (2.3) of V. It will be expedient to tackle the corresponding property of $\tilde{V}$ at the same time and appeal to the duality between V and $\tilde{V}$ in simplifying the arguments. By this approach and by passing to subsequences that tend to τ either from above or from below, we can reduce the challenge to proving that

$$\begin{array}{ll} \text{(a)} & \text{whenever } \tau \geq 0 \text{ and } \tau^\nu \searrow \tau, \text{ one has} \\ & \quad \begin{cases} \limsup_\nu V(\tau^\nu, \xi^\nu) \leq V(\tau, \xi) & \text{for some sequence } \xi^\nu \to \xi, \\ \liminf_\nu \tilde{V}(\tau^\nu, \eta^\nu) \geq \tilde{V}(\tau, \eta) & \text{for every sequence } \eta^\nu \to \eta; \end{cases} \\ \text{(b)} & \text{whenever } \tau > 0 \text{ and } \tau^\nu \nearrow \tau, \text{ one has} \\ & \quad \begin{cases} \limsup_\nu V(\tau^\nu, \xi^\nu) \leq V(\tau, \xi) & \text{for some sequence } \xi^\nu \to \xi, \\ \liminf_\nu \tilde{V}(\tau^\nu, \eta^\nu) \geq \tilde{V}(\tau, \eta) & \text{for every sequence } \eta^\nu \to \eta, \end{cases} \end{array} \tag{5.7}$$

since these "subproperties" yield by duality the corresponding ones with V and $\tilde{V}$ reversed.

Argument for (a) of (5.7). Fix any $\bar{\tau} \geq 0$ and $\bar{\xi} \in \operatorname{dom} V_{\bar{\tau}}$. We'll verify that the first limit in (a) holds for $(\bar{\tau}, \bar{\xi})$. Take any $\hat{\tau} > \bar{\tau}$. By Corollary 4.4(a), the image of the set $C_{\hat{\tau}-\bar{\tau}} = \operatorname{dom} E(\hat{\tau} - \bar{\tau}, \cdot, \cdot)$ under the projection $(\xi', \xi) \mapsto \xi'$ contains $\bar{\xi}$. Hence there exists $\hat{\xi}$ such that $E(\hat{\tau} - \bar{\tau}, \bar{\xi}, \hat{\xi}) < \infty$. Equivalently, there is an arc $x(\cdot) \in \mathcal{A}^1_n[\bar{\tau}, \hat{\tau}]$ such that $\int_{\bar{\tau}}^{\hat{\tau}} L(x(t), \dot{x}(t))dt < \infty$ and $x(\bar{\tau}) = \bar{\xi}$. Then too for every $\tau \in (\bar{\tau}, \hat{\tau})$ we have $E(\tau - \bar{\tau}, \bar{\xi}, x(\tau)) \leq \int_{\bar{\tau}}^{\tau} L(x(t), \dot{x}(t))dt < \infty$ and therefore by (5.2) that

$$V(\tau, x(\tau)) \leq V(\bar{\tau}, \bar{\xi}) + \alpha(\tau) \text{ for } \alpha(\tau) := \int_{\bar{\tau}}^{\tau} L(x(t), \dot{x}(t))dt.$$

Consider any sequence $\tau^\nu \searrow \bar{\tau}$ in $(\bar{\tau}, \hat{\tau})$. Let $\xi^\nu = x(\tau^\nu)$. Then $\xi^\nu \to \bar{\xi}$ and we obtain

$$\limsup_\nu V(\tau^\nu, \xi^\nu) \leq \limsup_\nu \big\{ V(\bar{\tau}, \bar{\xi}) + \alpha(\tau^\nu) \big\} = V(\bar{\tau}, \bar{\xi}),$$

as desired. To establish the second limit in (a) as consequence of this, we note now that the conjugacy in Theorem 5.1 gives $\tilde{V}(\tau^\nu, \cdot) \geq \langle \xi^\nu, \cdot \rangle - V(\tau^\nu, \xi^\nu)$. For any $\bar{\eta}$ and sequence $\eta^\nu \to \bar{\eta}$ this yields

$$\liminf_\nu \tilde{V}(\tau^\nu, \eta^\nu) \;\geq\; \liminf_\nu \big\{ \langle \xi^\nu, \eta^\nu \rangle - V(\tau^\nu, \xi^\nu) \big\} \;\geq\; \langle \bar{\xi}, \bar{\eta} \rangle - V(\bar{\tau}, \bar{\xi}). \tag{5.8}$$

But $\bar{\xi}$ was an arbitrary point in $\operatorname{dom} V(\bar{\tau}, \cdot)$, so we get the rest of what is needed in (a):

$$\liminf_\nu \tilde{V}(\tau^\nu, \eta^\nu) \;\geq\; \sup_\xi \big\{ \langle \xi, \bar{\eta} \rangle - V(\bar{\tau}, \xi) \big\} \;=\; \tilde{V}(\bar{\tau}, \bar{\eta}). \tag{5.9}$$

Argument for (b) of (5.7). Fix any $\bar{\tau}0$ and $\bar{\xi} \in \operatorname{dom} V_{\bar{\tau}}$. We'll verify that the first limit in (a) holds for $(\bar{\tau}, \bar{\xi})$. Let $\varepsilon > 0$. Because $V(\bar{\tau}, \bar{\xi}) < \infty$, there exists $x(\cdot) \in \mathcal{A}_n^1[0, \bar{\tau}]$ with $x(\bar{\tau}) = \bar{\xi}$ and $g(x(0)) + \int_0^{\bar{\tau}} L(x(t), \dot{x}(t))dt < V(\bar{\tau}, \bar{\xi}) + \varepsilon$. Then for all $\tau \in (0, \bar{\tau})$ we have

$$\begin{aligned} V(\tau, x(\tau)) &\leq g(x(0)) + \int_0^\tau L(x(t), \dot{x}(t))dt \\ &\leq V(\bar{\tau}, \bar{\xi}) + \varepsilon - \alpha(\tau) \text{ for } \alpha(\tau) = \int_\tau^{\bar{\tau}} L(x(t), \dot{x}(t))dt. \end{aligned}$$

Consider any sequence $\tau^\nu \nearrow \bar{\tau}$ in $(0, \bar{\tau})$. Let $\xi^\nu = x(\tau^\nu)$. Then $\xi^\nu \to \bar{\xi}$ and we have

$$\limsup_\nu V(\tau^\nu, \xi^\nu) \leq \limsup_\nu \big\{ V(\bar{\tau}, \bar{\xi}) + \varepsilon - \alpha(\tau^\nu) \big\} \leq V(\bar{\tau}, \bar{\xi}) + \varepsilon.$$

We've constructed a sequence with $\xi^\nu \to \bar{\xi}$ with this property for arbitrary ε, so by diagonalization we can get a sequence $\xi^\nu \to \bar{\xi}$ with $\limsup_\nu V(\tau^\nu, \xi^\nu) \leq V(\bar{\tau}, \bar{\xi})$, as required. Fixing such a sequence and returning to the inequality $\tilde{V}(\tau^\nu, \cdot) \geq \langle \xi^\nu, \cdot \rangle - V(\tau^\nu, \xi^\nu)$, we obtain now for every sequence $\eta^\nu \to \bar{\eta}$ that (5.8) holds, and hence by the arbitrary choice of $\bar{\xi} \in \operatorname{dom} V_{\bar{\tau}}$ that (5.9) holds as well. $\square$

The duality theory for the Bolza problems in this setting also provides insights into the properties of the optimal arcs associated with V.

Theorem 5.2 (optimal arcs). *In the minimization problem defining $V_\tau(\xi) = V(\tau, \xi)$, an optimal arc $x(\cdot) \in \mathcal{A}_n^1[0, \tau]$ exists for any $\xi \in \operatorname{dom} V_\tau$. Every such arc $x(\cdot)$ must actually belong to $\mathcal{A}_n^\infty[0, \tau]$ when ξ is such that $\partial V_\tau(\xi) \neq \emptyset$, hence in particular if $\xi \in \operatorname{ri} \operatorname{dom} V_\tau$.*

Proof. Although the theorem is stated in terms of V alone, its proof will rest on the duality between V and $\tilde{V}$. We'll focus actually on proving the $\tilde{V}$ version, since that ties in better with the foundation already laid in the proof of Theorem 5.1.

Returning to the problems $(\mathcal{P})$ and $(\tilde{P})$ specified in that proof, we make further use of the duality results in Theorem 4.5. We showed that our choice of the function l implied $\inf(\tilde{\mathcal{P}}) = -\inf(\mathcal{P}) > -\infty$ in (5.6), but we didn't point out then that it also guarantees through Theorem 4.5(a) that an optimal arc $y(\cdot)$ exists for $(\tilde{\mathcal{P}})$ when, in addition, $\inf(\tilde{P}) < \infty$. Thus, an optimal arc exists for the problem defining $\tilde{V}(\tau, \bar{\eta})$ as long as $\tilde{V}(\tau, \bar{\eta}) < \infty$. Likewise then, an optimal arc exists for the problem defining $V(\tau, \bar{\xi})$ for any choice of $\bar{\xi}$ such that $V(\tau, \bar{\xi}) < \infty$.

Next we use the fact that a vector $\bar{\xi}$ belongs to $\partial \tilde{V}_\tau(\bar{\eta})$ if and only if $\bar{\eta} \in \operatorname{dom} \tilde{V}_\tau$ and $\bar{\xi}$ furnishes the maximum in the expression for $-\inf(\mathcal{P})$ in (5.6). (This is true by (5.4) and (5.5) of Theorem 5.1.) For such a vector $\bar{\xi}$, $V(\tau, \bar{\xi})$ has to be finite, so that there exists, by the argument already furnished, an optimal arc $x(\cdot)$ for the minimizing problem that defined $V(\tau, \bar{\xi})$. That arc $x(\cdot)$ must then be optimal for $(\mathcal{P})$. Theorem 4.5(a) tells us in that case that $x(\cdot)$ and the optimal arc $y(\cdot)$ for $(\tilde{\mathcal{P}})$ are in $\mathcal{A}_n^\infty[0, \tau]$.

To finish up, we merely need to recall that a proper convex function φ has subgradients at every point of $\operatorname{ri} \operatorname{dom} \varphi$, in particular. $\square$

6. Hamiltonian Dynamics and Method of Characteristics

The generalized Hamiltonian ODE in (2.6) now enters the discussion. This dynamical system can be written in the form

$$(\dot{x}(t), \dot{y}(t)) \in G(x(t), y(t)) \quad \text{for a.e. } t \tag{6.1}$$

for the set-valued mapping

$$G : (x, y) \;\mapsto\; \partial_y H(x, y) \times -\tilde{\partial}_x H(x, y), \tag{6.2}$$

which derives from the subgradient mapping $(x, y) \mapsto \tilde{\partial}_x H(x, y) \times \partial_y H(x, y)$. The latter has traditionally been associated in convex analysis with H as a concave-convex function on $I\!R^n \times I\!R^n$. It is known to be nonempty-compact-convex-valued and locally bounded with closed graph (since H is also finite; see [5; §35]). Hence the same holds for G.

Through these properties of G, the theory of differential inclusions [15] ensures the local existence of a Hamiltonian trajectory through every point. The local boundedness of G makes any trajectory $(x(\cdot), y(\cdot))$ over a time interval $[\tau_0, \tau_1]$ be Lipschitz continuous, i.e., belong to $\mathcal{A}^\infty_n[\tau_0, \tau_1]$. Another aspect of the Hamiltonian dynamics in (2.6), or (6.1)–(6.2), is that $H(x(t), y(t))$ is constant along any trajectory $(x(\cdot), y(\cdot))$. This was proved in [2].

Nowadays there are other concepts of subgradient, beyond those of convex analysis, that can be applied to H without separating it into its concave and convex arguments. The general definition in §2 directly assigns a subset $\partial H(x, y) \subset I\!R^n \times I\!R^n$ to each point $(x, y) \in I\!R^n \times I\!R^n$. An earlier definition for this purpose, which was used by Clarke in his work on Hamiltonian conditions for optimality in *nonconvex* problems of Bolza (see [16] and its references), relied on H being locally Lipschitz continuous and utilized what we now recognize as the set $\operatorname{con} \partial H(x, y)$ in such circumstances. (Here 'con' designates the convex hull of a set.) A more subtle form of 'partial convexification' of $\partial H(x, y)$, involving only the x argument in a special way, has been featured in more recent work on Hamiltonians in nonconvex problems of Bolza; cf. [10], [11].

As a preliminary to our further analysis of the Hamiltonian dynamics, we provide a clarification of the relationships between these concepts.

Proposition 6.1 (subgradients of the Hamiltonian)**.** *On the basis of $H(x, y)$ being finite, concave in x and convex in y, one has*

$$\operatorname{con} \partial H(x, y) \;=\; \tilde{\partial}_x H(x, y) \times \partial_y H(x, y), \tag{6.3}$$

this set being everywhere nonempty and compact. Moreover, in terms of the subset D of $I\!R^n \times I\!R^n$ consisting of the points where H is differentiable (the complement of which is of measure zero), one has

$$\operatorname{con} \partial H(x, y) = \partial H(x, y) = \{\nabla H(x, y)\} \quad \textit{for all} \quad (x, y) \in D, \tag{6.4}$$

The gradient mapping ∇H is continuous relative to D, so that H is strictly differentiable on D. Elsewhere,

$$\operatorname{con} \partial H(x, y) \;=\; \operatorname{con}\Big\{(w, v) \,\Big|\, \exists\, (x^\nu, y^\nu) \to (x, y) \ \textit{with}\ \nabla H(x^\nu, y^\nu) \to (w, v)\Big\}. \tag{6.5}$$

Proof. Formula (6.5) is well known to hold for the subgradients of any locally Lipschitz continuous function; cf. [6; 9.61]. The special property coming out of the concavity-convexity of H is that the set-valued mapping

$$T_H : (x, y) \;\mapsto\; [-\tilde{\partial}_x H(x, y)] \times \partial_y H(x, y) \;=\; \partial_x[-H](x, y) \times \partial_y H(x, y) \tag{6.6}$$

is maximal monotone; cf. [6; 12.27]. The points (x, y) where T_H is single-valued are the ones where $\tilde{\partial}_x H(x, y)$ and $\partial H_y(x, y)$ both reduce to singletons, a property which corresponds to $H(\cdot, y)$ being differentiable at x while $H(x, \cdot)$ is differentiable at y; then actually H is differentiable (jointly in the two arguments) at (x, y); cf. [5; 35.6]. Thus, the subset of $I\!R^n \times I\!R^n$ on which T_H is single-valued is D, and on this set we have $T_H(x, y) = (-\nabla_x H(x, y), \nabla_y H(x, y)$. Then by maximal monotonicity, T_H is continuous on D with

$$T_H(x, y) \;=\; \operatorname{con}\Big\{(-w, v) \,\Big|\, \exists\, (x^\nu, y^\nu) \to (x, y) \ \text{with}\ \nabla H(x^\nu, y^\nu) \to (w, v)\Big\};$$

see [6; 12.63, 12.67]. We thereby obtain (6.3) from (6.5) and at the same time have (6.4), from which H must be strictly differentiable on D by [6; 9.18]. □

Corollary 6.2 (single-valuedness in the Hamiltonian system). *The mapping G in the differential inclusion (6.1)–(6.2) has the form*

$$G(x,y) = \{(v,-w) \mid (w,v) \in \operatorname{con} \partial H(x,y)\} \tag{6.7}$$

and is single-valued almost everywhere. Indeed, $G(x,y) = \{(\nabla_y H(x,y), -\nabla_x H(x,y))\}$ at all points where the Hamiltonian H is differentiable, whereas in general,

$$G(x,y) = \operatorname{con}\Big\{(v,-w) \,\Big|\, \exists\, (x^\nu,y^\nu) \to (x,y) \text{ with } (\nabla_y H(x^\nu,y^\nu), -\nabla_x H(x^\nu,y^\nu)) \to (v,-w)\Big\}. \tag{6.8}$$

Despite the typical single-valuedness of G, situations exist in which there can be more than one Hamiltonian trajectory from a given starting point. The flow mappings S_τ for this system, as defined in (2.7), can well have values that are not singleton sets, and indeed, can even be nonconvex sets consisting of more than finitely many points. It's rather surprising, then, that they nonetheless capture with precision the behavior of the Lipschitzian manifolds $\operatorname{gph} \partial V_\tau$ in Corollary 2.2. We're prepared now to prove this fact.

Proof of Theorem 2.4. Fix $\tau > 0$ along with any $\bar\xi$ and $\bar\eta$. The relation $\bar\eta \in \partial V_\tau(\bar\xi)$ that we wish to characterize is equivalent by Theorem 5.1 to $\bar\xi \in \partial \tilde V_\tau(\bar\eta)$, or for that matter to having $\bar\xi \in \operatorname{argmax}_\xi \{\langle \xi, \bar\eta\rangle - V_\tau(\xi)\}$. We saw in the proof of Theorem 5.2 that this corresponded further, in terms of the special Bolza problems $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ introduced in the proof of Theorem 5.1, to the existence of optimal arcs $x(\cdot)$ for $(\mathcal{P})$ and $y(\cdot)$ for $(\tilde{\mathcal{P}})$ such that $x(\tau) = \bar\xi$.

On the other hand, because $-\inf(\mathcal{P}) = (\tilde{\mathcal{P}})$ for these problems, we know from Theorem 4.1 that arcs $x(\cdot)$ and $y(\cdot)$ solve these problems, respectively, if and only if $(x(\cdot), y(\cdot)$ is a Hamiltonian trajectory over $[0,\tau]$ satisfying the generalized transversality condition $(y(0), -y(\tau)) \in \partial l(x(0), x(\tau))$. Since $l(\xi', \xi) = g(\xi') - \langle \xi, \bar\eta\rangle$ by definition in this case, the transversality condition comes down to the relations $y(0) \in \partial g(x(0)$ and $y(\tau) = \bar\eta$.

In summary, we have $\bar\eta \in \partial V_\tau(\bar\xi)$ if and only if there is a trajectory $(x(\cdot), y(\cdot))$ over $[0,\tau]$ such that $x(\tau) = \bar\eta$, $y(0) \in \partial g(x(0))$ and $y(\tau) = \bar\eta$. □

Further details about the evolution of the subgradient mappings $\partial V_\tau = \partial_\xi V(\tau, \cdot)$ can now be recorded. The equivalence in the next theorem came out in the preceding proof.

Theorem 6.3 (optimality in subgradient evolution). *A pair of arcs $x(\cdot)$ and $y(\cdot)$ gives a Hamiltonian trajectory over $[0,\tau]$ that starts in $\operatorname{gph} \partial g$ and ends at a point $(\xi,\eta) \in \operatorname{gph} \partial V_\tau$ if and only if*

(a) *$x(\cdot)$ is optimal in the minimization problem in (1.1) that defines $V(\tau,\xi)$, and*

(b) *$y(\cdot)$ is optimal in the minimization problem in (5.1) that defines $\tilde V(\tau,\eta)$.*

Corollary 6.4 (persistence of subgradient relations). *When a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ over $[0,\tau]$ has $y(0) \in \partial g(x(0))$, it has $y(t) \in \partial_\xi V(t, x(t))$ for all $t \in [0,\tau]$.*

We turn now, however, to the task of broadening Theorem 2.4 to cover not only the evolution of subgradients but also that of function values. For this, the graph of ∂V_τ in $\mathbb{R}^n \times \mathbb{R}^n$ has to be replaced by an associated subset of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$.

Proposition 6.5 (characteristic manifolds for convex functions). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, proper and lsc, and let*

$$M = \{(x,y,z) \mid y \in \partial f(x),\ z = f(x)\} \subset \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}. \tag{6.9}$$

Then M is an n-dimensional Lipschitzian manifold in the following terms. There is a one-to-one, locally Lipschitz continuous mapping

$$F : \mathbb{R}^n \to M, \qquad F(u) = (P(u), Q(u), R(u)),$$

whose range is all of M and whose inverse is Lipschitz continuous as well, in fact with

$$F^{-1}(x,y,z) = x + y \quad \text{for } (x,y,z) \in M.$$

The components of F are given by

$$P(u) = \operatorname{argmin}_x \Big\{ f(x) + \tfrac{1}{2}|x-u|^2 \Big\}, \qquad Q = I - P, \qquad R = f \circ P, \tag{6.10}$$

where P and Q, like F^{-1}, are globally Lipschitz continuous with constant 1, and R is Lipschitz continuous with constant r on the ball $\{u \mid |u| \le r\}$ for each $r > 0$.

Proof. The mapping $u \mapsto (P(u), Q(u))$ is well known to parameterize the graph of ∂f in the manner described; cf. [6; 12.15]. With this parameterization, the component $z = R(u)$ must be $f(P(u))$, so the additional issue is just the claimed Lipschitz property of this expression. According to the formulas for P and Q in (6.10) we have

$$R(u) = p(u) - \tfrac{1}{2}|Q(u)|^2 \quad \text{for} \quad p(u) := \min_x \Big\{ f(x) + \tfrac{1}{2}|x-u|^2 \Big\}. \tag{6.11}$$

The function p is smooth with gradient $\nabla p(u) = Q(u)$; see [6; 2.26]. Hence R is locally Lipschitz continuous, but what can be said about its Lipschitz modulus? Because P and Q are Lipschitz continuous with constant 1 and satisfy $P + Q = I$, they are differentiable at almost every point u, their Jacobian matrices satisfying $\nabla P(u) + \nabla Q(u) = I$ and having norms at most 1. At any such point u, R is differentiable as well, with $\nabla R(u) = Q(u) - \nabla Q(u) Q(u) = \nabla P(u) Q(u)$, so that $|\nabla R(u)| \le |\nabla P(u)||Q(u)| \le |Q(u)| \le |u|$. Thus, $|\nabla R(u)| \le r$ on the ball $\{u \mid |u| \le r\}$, and consequently R is Lipschitz continuous with constant r on that ball. □

The set M in (6.9) will be called the (first-order) *characteristic manifold* for f, and the mapping F its *canonical parameterization*.

Proposition 6.6 (recovery of a function from its manifold)**.** *Let M be the characteristic manifold of a convex, proper, lsc function f. Then M uniquely determines f as follows:*

(a) *The image C of M under the projection $(x, y, z) \mapsto x$, namely $C = \operatorname{dom} \partial f$, satisfies $\operatorname{ri} \operatorname{dom} f \subset C \subset \operatorname{cl} \operatorname{dom} f$ and thus has $\operatorname{ri} C = \operatorname{ri} \operatorname{dom} f$ and $\operatorname{cl} C = \operatorname{cl} \operatorname{dom} f$.*

(b) *For every x in C, the vectors $(x, y, z) \in M$ all have the same z, which equals $f(x)$.*

(c) *For every $x \in \operatorname{cl} C \setminus C$ and any $a \in \operatorname{ri} C$, one has $x + \varepsilon(a - x) \in \operatorname{ri} C$ for all $\varepsilon \in (0, 1]$ and $f(x + \varepsilon(a - x)) \to f(x)$ as $\varepsilon \searrow 0$.*

(c) *For every $x \notin \operatorname{cl} C$, $f(x) = \infty$.*

Proof. These facts are evident from the definition of M, the well known existence of subgradients at points of $\operatorname{ri} \operatorname{dom} f$, and the way that f can be recovered fully from its values on $\operatorname{ri} \operatorname{dom} f$; see [5; §7, §23]. □

Proposition 6.7 (convergence of characteristic manifolds)**.** *A sequence of convex, proper, lsc functions f^ν on $\mathbb{R}^n$ epi-converges to another such function f if and only if the associated sequence of characteristic manifolds M^ν in $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ converges (in the Painlevé-Kuratowski sense) to the characteristic manifold M for f.*

Proof. Attouch's theorem on convex functions (cf. [6; 12.35]) says that f^ν epi-converges to f if and only if $\operatorname{gph} \partial f^\nu$ converges to $\operatorname{gph} \partial f$ and, for at least one sequence of points $(x^\nu, y^\nu) \in \operatorname{gph} \partial f^\nu$ converging to a point $(x, y) \in \operatorname{gph} \partial f$, one has $f^\nu(x^\nu) \to f(x)$. On the other hand, epi-convergence of convex functions entails latter holding for every such sequence of points (x^ν, y^ν). The convergence of the characteristic manifolds is thus hardly more than a restatement of these facts of convex analysis. □

Our goal in these terms is to describe how the characteristic manifold for V_τ evolves from that of g. We introduce the following extension of the Hamiltonian system (6.1)–(6.2), which we speak of as the *characteristic system* associated with H:

$$(\dot{x}(t), \dot{y}(t), \dot{z}(t)) \in \bar{G}(x(t), y(t)) \ \text{ for a.e. } t \tag{6.12}$$

for the set-valued mapping $\bar{G}$ defined by

$$\bar{G}(x, y) := \big\{ (v, w, u) \,\big|\, (v, w) \in G(x, y), \ u = \langle v, y \rangle - H(x, y) \big\}. \tag{6.13}$$

The trajectories $(x(\cdot), y(\cdot), z(\cdot))$ of this system will be called *characteristic trajectories*. Like G itself, $\bar{G}$ is nonempty-closed-convex-valued and locally bounded with closed graph, so a characteristic trajectory exists, at least locally, through every point of $I\!R^n \times I\!R^n \times I\!R$. The corresponding flow mapping for each $\tau \in [0, \infty)$ will be denoted by $\bar{S}_\tau$:

$$\begin{aligned} \bar{S}_\tau : (\xi_0, \eta_0, \zeta_0) \mapsto & \\ & \Big\{ (\xi, \eta, \zeta) \,\Big|\, \exists \text{ characteristic trajectory } (x(\cdot), y(\cdot), z(\cdot)) \text{ over } [0, \tau] \text{ with} \\ & \qquad (x(0), y(0), z(0)) = (\xi_0, \eta_0, \zeta_0),\ (x(\tau), y(\tau), z(\tau)) = (\xi, \eta, \zeta) \Big\} \end{aligned} \tag{6.14}$$

Theorem 6.8 (subgradient method of characteristics). *Let M_τ be the characteristic manifold for $V_\tau = V(\tau, \cdot)$, with M_0 the characteristic manifold for $g = V_0$. Then*

$$M_\tau \;=\; \bar{S}_\tau(M_0) \quad \textit{for all} \quad \tau \geq 0. \tag{6.15}$$

Moreover M_τ, as a closed subset of $I\!R^n \times I\!R^n \times I\!R$, depends continuously on τ.

Proof. The continuity of the mapping $\tau \mapsto M_\tau$ is immediate from Proposition 6.7 and the epi-continuity in Theorem 2.1. The evolution of ∂V_τ through the drift of its graph in the underlying system (6.1)–(6.2) has already been verified in Theorem 2.4, so the only issue here is what happens when the z component is added as in (6.12)–(6.13). We have

$$\dot{z}(t) \;=\; \langle \dot{x}(t), y(t) \rangle - H(x(t), y(t)) \;=\; L(x(t), \dot{x}(t)) \tag{6.16}$$

when $(\dot{x}(t), \dot{y}(t)) \in G(x(t), y(t))$, since that relation entails $\dot{x}(t) \in \partial_y H(x(t), y(t))$, which is equivalent to the second equation in (6.16) because the convex functions $H(x(t), \cdot)$ and $L(x(t), \cdot)$ are conjugate to each other. The arc $x(\cdot)$ is optimal for the minimization problem that defines $V(\tau, \xi)$, so that

$$V(\tau, \xi) \;=\; g(x(0)) + \int_0^\tau L(x(t), \dot{x}(t)) dt \;=\; z(0) + \int_0^\tau \dot{z}(t)) dt \;=\; z(\tau).$$

The trajectory $(x(\cdot), y(\cdot), z(\cdot))$ does, therefore, carry the point $(x(0), y(0), z(0)) \in M_0$ to the point $(x(\tau), y(\tau), z(\tau)) \in M_\tau$. Conversely, of course, (6.16) is essential for that. □

Theorem 6.8 provides a remarkably global version of the method of characteristics, made possible by convexity. It relies on the one-to-one correspondence between lsc, proper, convex functions and their characteristic manifolds in Proposition 6.5 and on the preservation of such function properties over time, as in Theorem 2.1. By transforming the evolution of functions into the evolution of the associated manifolds, one is able to reduce the function evolution to the drift of those manifolds in the characteristic dynamical system associated with the given Hamiltonian H, or Lagrangian L.

In contrast, the classical method of characteristics requires differentiability at every turn and, in adopting the implicit (or inverse) function theorem as the main tool, is ordinarily limited to local validity. The characteristic manifold M_0 associated with g has to be a smooth manifold, and g must therefore be $\mathcal{C}^2$. The Hamiltonian H has to be $\mathcal{C}^2$ as well, so that the mappings $\bar{S}_\tau$ are single-valued and smooth. But even these assumptions are not enough to guarantee that the characteristic dynamics will carry M_0 into *smooth* manifolds M_τ. The trouble is that the functions V_τ are defined by minimization, and that operation, in its inherent failure to preserve differentiability, simply does not fit well in the framework of classical analysis.

A generalized "method of characteristics" for value functions has also been developed by Subbotin [17], but in a different framework from ours, namely one focused on bounded control dynamics and not convexity, and not revolving around the Hamiltonian function H and its dynamical system.

7. Hamilton-Jacobi Equation and Regularity

The time has come to move beyond subgradients of convex analysis and establish properties of the subgradient mapping ∂V as a whole.

Proof of Theorem 2.5. Our first goal is to prove the equivalence of the conditions $\eta \in \partial_\xi V(\tau, \xi)$ and $\sigma = -H(\xi, \eta)$ with having $(\sigma, \eta) \in \hat{\partial} V(\tau, \xi)$ when $\tau > 0$. Here $\partial_\xi V(\tau, \xi)$ is the same as $\hat{\partial}_\xi V(\tau, \xi)$, since the function $V(\tau, \cdot) = V_\tau$ is convex.

Let $\bar{\eta} \in \partial_\xi V(\bar{\tau}, \bar{\xi})$ with $\bar{\tau} > 0$. We need to show that $(-H(\bar{\xi}, \bar{\eta}), \bar{\eta}) \in \hat{\partial} V(\bar{\tau}, \bar{\xi})$, or in other words that

$$V(\tau, \xi) - V(\bar{\tau}, \bar{\xi}) + (\tau - \bar{\tau}) H(\bar{\xi}, \bar{\eta}) - \langle \xi - \bar{\xi}, \bar{\eta} \rangle \ \geq \ o\big(|(\tau, \xi) - (\bar{\tau}, \bar{\xi})|\big). \tag{7.1}$$

By Theorem 2.4 there is a Hamiltonian trajectory $(x(\cdot), y(\cdot)$ over $[0, \bar{\tau}]$ that starts in $\operatorname{gph} g$ and goes to $(\bar{\xi}, \bar{\eta})$. Through the local existence property of the Hamiltonian system, this trajectory can be extended to a larger interval $[0, \bar{\tau} + \varepsilon]$, in which case $y(\tau) \in \partial_\xi V(\tau, x(\tau))$ for all $\tau \in [0, \bar{\tau} + \varepsilon]$ by Corollary 6.4, so that

$$V(\tau, \xi) \ \geq \ V(\tau, x(\tau)) + \langle \xi - x(\tau), y(\tau) \rangle \quad \text{for all } \xi \in I\!R^n \text{ when } \tau \in [0, \bar{\tau} + \varepsilon]. \tag{7.2}$$

We have $V(\tau, x(\tau)) \ = \ g(x(0)) + \int_0^\tau \big[\langle \dot{x}(t), y(t) \rangle - H(x(t), y(t))\big] dt$ by Theorem 6.8, where $H(x(t), y(t)) \equiv H(x(\bar{\tau}), y(\bar{\tau}))$ because H is constant along Hamiltonian trajectories. Hence

$$V(\tau, x(\tau)) = V(\bar{\tau}, \bar{\xi}) - (\tau - \bar{\tau}) H(\bar{\xi}, \bar{\eta}) + \int_{\bar{\tau}}^{\tau} \langle \dot{x}(t), y(t) \rangle dt \text{ when } \tau \in [0, \bar{\tau} + \varepsilon]. \tag{7.3}$$

Also $\int_{\bar{\tau}}^{\tau} \langle \dot{x}(t), y(t) \rangle dt = \langle x(\tau), y(\tau) \rangle - \langle x(\bar{\tau}), y(\bar{\tau}) \rangle - \int_{\bar{\tau}}^{\tau} \langle x(t), \dot{y}(t) \rangle dt$, so in combining (7.3) with (7.2), we see that the left side of (7.1) is bounded below by the expression

$$\begin{aligned}
-\langle \xi - \bar{\xi}, \bar{\eta} \rangle + \langle \xi - x(\tau), y(\tau) \rangle + \langle x(\tau), y(\tau) \rangle - \langle x(\bar{\tau}), y(\bar{\tau}) \rangle - \int_{\bar{\tau}}^{\tau} \langle x(t), \dot{y}(t) \rangle dt \\
= \langle \xi - \bar{\xi},\, y(\tau) - \bar{\eta} \rangle + \langle \bar{\xi},\, y(\tau) - \bar{\eta} \rangle - \int_{\bar{\tau}}^{\tau} \langle x(t), \dot{y}(t) \rangle dt \\
= \langle \xi - \bar{\xi},\, y(\tau) - y(\bar{\tau}) \rangle - \int_{\bar{\tau}}^{\tau} \langle x(t) - x(\bar{\tau}), \dot{y}(t) \rangle dt.
\end{aligned}$$

This expression is of type $o\big(|(\tau, \xi) - (\bar{\tau}, \bar{\xi})|\big)$ because $x(\cdot)$ and $y(\cdot)$ are continuous and $\dot{y}(\cdot)$ is essentially bounded on $[0, \bar{\tau} + \varepsilon]$. Thus, $(-H(\bar{\xi}, \bar{\eta}), \bar{\eta}) \in \hat{\partial} V(\bar{\tau}, \bar{\xi})$ as claimed.

To argue the converse implication, we consider now any pair $(\bar{\sigma}, \bar{\eta}) \in \hat{\partial} V(\bar{\tau}, \bar{\xi})$. Such a pair satisfies

$$V(\tau, \xi) \ \geq \ V(\bar{\tau}, \bar{\xi}) - (\tau - \bar{\tau}) \bar{\sigma} + \langle \xi - \bar{\xi}, \bar{\eta} \rangle + o\big(|(\tau, \xi) - (\bar{\tau}, \bar{\xi})|\big). \tag{7.4}$$

In particular $\bar{\eta} \in \hat{\partial}_\xi V(\bar{\tau}, \bar{\xi}) = \partial_\xi V(\bar{\tau}, \bar{\xi})$, and we therefore have, as just explained, the existence of a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ for which (7.3) holds. Specializing (7.4) to $\xi = x(\tau)$ and using the expression in (7.3) for $V(\tau, x(\tau))$, we obtain

$$\begin{aligned}
V(\bar{\tau}, \bar{\xi}) - (\tau - \bar{\tau}) H(\bar{\xi}, \bar{\eta}) + \int_{\bar{\tau}}^{\tau} \langle \dot{x}(t), y(t) \rangle dt \\
\geq \ V(\bar{\tau}, \bar{\xi}) - (\tau - \bar{\tau}) \bar{\sigma} + \langle x(\tau) - x(\bar{\tau}), \bar{\eta} \rangle + o\big(|(\tau, x(\tau)) - (\bar{\tau}, x(\bar{\tau}))|\big),
\end{aligned}$$

where the final term is of type $o(|\tau - \bar{\tau}|)$ because $x(\cdot)$ is locally Lipschitz continuous. Then

$$-(\tau - \bar{\tau})(\bar{\sigma} + H(\bar{\xi}, \bar{\eta})) \ \geq \ \int_{\bar{\tau}}^{\tau} \langle \dot{x}(t),\, y(t) - y(\bar{\tau}) \rangle dt + o(|\tau - \bar{\tau}|),$$

with the integral term likewise being of type $o(|\tau - \bar{\tau}|)$. Necessarily, then, $\bar{\sigma} + H(\bar{\xi}, \bar{\eta}) = 0$.

We turn now to showing that $\partial V(\tau,\xi) = \hat{\partial} V(\tau,\xi)$ for all ξ when $\tau > 0$. Since $\hat{\partial} V(\tau,\xi) \subset \partial V(\tau,\xi)$ in general, only the opposite inclusion has to be checked. Suppose $(\sigma,\eta) \in \partial V(\tau,\xi)$. By definition, there are sequences $(\tau^\nu,\xi^\nu) \to (\tau,\xi)$ and $(\sigma^\nu,\eta^\nu) \to (\sigma,\nu)$ with $V(\tau^\nu,\xi^\nu) \to V(\tau,\xi)$ and $(\sigma^\nu,\eta^\nu) \in \hat{\partial} V(\tau^\nu,\xi^\nu)$. We have seen that the latter means $\sigma^\nu = -H(\xi^\nu,\eta^\nu)$ and $\eta^\nu \in \partial_\xi V(\tau^\nu,\xi^\nu)$. Then $\sigma = -H(\xi,\eta)$ by the continuity of H.

On the other hand, the sets $C^\nu = \operatorname{gph} \partial_\xi V(\tau^\nu,\cdot)$ converge to $C = \operatorname{gph} \partial_\xi V(\tau,\cdot)$ by Corollary 2.2. Hence from having $\eta^\nu \in \partial_\xi V(\tau^\nu,\xi^\nu)$ we get $\eta \in \partial_\xi V(\tau,\xi)$. The pair (σ,η) thus satisfies the conditions we have identified as describing the elements of $\hat{\partial} V(\tau,\xi)$. □

Through the duality in Theorem 5.1, the statements in Theorem 2.5 are valid equally for the dual value function $\tilde{V}$. From this we obtain the following.

Theorem 7.1 (dual Hamilton-Jacobi equation)**.** *The dual value function $\tilde{V}$ satisfies*

$$\sigma - H(\xi,\eta) = 0 \quad \textit{for all} \quad (\sigma,\xi) \in \partial \tilde{V}(\tau,\eta) \quad \textit{when} \quad \tau > 0. \tag{7.5}$$

Indeed, for $\tau > 0$ one has $(\sigma,\xi) \in \partial \tilde{V}(\tau,\eta)$ if and only if $(-\sigma,\eta) \in \partial V(\tau,\xi)$.

Proof. In translating Theorem 2.5 to the context of $\tilde{V}$, as justified by Theorem 5.1, we bring into the scene the dual Hamiltonian $\tilde{H}(y,x) = -H(x,y)$ corresponding (in Proposition 3.5) to the dual Lagrangian $\tilde{L}$. The vectors $(\sigma,\xi) \in \partial \tilde{V}(\tau,\eta)$ are characterized by $\xi \in \partial_\eta \tilde{V}(\tau,\eta)$ and $\sigma = -\tilde{H}(\eta,\xi) = H(\xi,\eta)$. Invoking the conjugacy between $V(\tau,\cdot)$ and $\tilde{V}(\tau,\cdot)$ in Theorem 5.1, specifically the relation (5.5), we get the subgradient equivalence. Then (7.5) is immediate from the Hamilton-Jacobi equation already in Theorem 2.5. □

We take up next the issue of what additional properties of continuity, differentiability, etc., the value function V is guaranteed to have beyond the convexity and epi-continuity in Theorem 2.1. We begin with a characterization of the interior of the set

$$\operatorname{dom} V = \big\{ (\tau,\xi) \in [0,\infty) \times I\!R^n \,\big|\, V(\tau,\xi) < \infty \big\}.$$

Proposition 7.2 (domain interiors)**.** *In terms of $V_\tau = V(\tau,\cdot)$, one has that*

$$(\tau,\xi) \in \operatorname{int} \operatorname{dom} V \quad \Longleftrightarrow \quad \tau > 0,\ \xi \in \operatorname{int} \operatorname{dom} V_\tau.$$

Proof. It's evident that "$\Rightarrow$" holds. We focus therefore on "$\Leftarrow$". Consider $\bar{\tau} > 0$ and $\bar{\xi} \in \operatorname{int} \operatorname{dom} V_{\bar{\tau}}$. The epi-convergence of V_τ to $V_{\bar{\tau}}$ as $\tau \to \bar{\tau}$ in Theorem 2.1 entails through the convexity of these functions that V_τ converges pointwise to $V_{\bar{\tau}}$ uniformly on all compact subsets of $\operatorname{int} \operatorname{dom} V_{\bar{\tau}}$; cf. [6; 7.17]. In particular this convergence holds on some open neighborhood U of $\bar{x}$ in $\operatorname{dom} V_{\bar{\tau}}$, so for some open interval I around $\bar{\tau}$ we have $U \subset \operatorname{dom} V_\tau$ for all $\tau \in I$. Then $I \times O$ is an open subset of $\operatorname{dom} V$ containing $(\bar{\tau},\bar{\xi})$, and we conclude that $(\bar{\tau},\bar{\xi}) \in \operatorname{int} \operatorname{dom} V$. □

The argument just given shows further that V is continuous on the interior of $\operatorname{dom} V$, but we're headed toward showing that V is in fact locally Lipschitz continuous there. The agreement between $\partial V(\tau,\xi)$ and $\hat{\partial} V(\tau,\xi)$ in Theorem 2.5 will have a part in this, and it will yield other strong properties besides.

Recall that a locally Lipschitz continuous function is *subdifferentially regular* (in the sense of Clarke regularity of its epigraph) when all its subgradients are regular subgradients, or equivalently, its subderivatives and regular subderivatives coincide everywhere; for background, see [6; Chapters 8 and 9]. The *subderivative* function for V at a point (τ,ξ) is defined in general by

$$dV(\tau,\xi) : (\tau',\xi') \mapsto dV(\tau,\xi)(\tau',\xi') := \liminf_{\substack{\varepsilon \searrow 0 \\ (\tau'',\xi'') \to (\tau',\xi')}} \frac{V(\tau + \varepsilon\tau'', \xi + \varepsilon\xi'') - V(\tau,\xi)}{\varepsilon}.$$

To say that V is *semidifferentiable* at (τ,ξ) is to say that, for all (τ',ξ'), this lower limit exists actually as the full limit

$$\lim_{\substack{\varepsilon \searrow 0 \\ (\tau'',\xi'') \to (\tau',\xi')}} \frac{V(\tau + \varepsilon\tau'', \xi + \varepsilon\xi'') - V(\tau,\xi)}{\varepsilon}.$$

Then $dV(\tau,\xi)(\tau',\xi')$ must be finite and continuous as a function of (τ',ξ'); cf. [6; 7.21].

Theorem 7.3 (regularity consequences). *On* $\operatorname{int}\operatorname{dom} V$, *the subgradient mapping* ∂V *is nonempty-compact-convex-valued and locally bounded, and* V *itself is locally Lipschitz continuous and subdifferentially regular, moreover semidifferentiable with*

$$dV(\tau,\xi)(\tau',\xi') \;=\; \max\Big\{\langle \xi',\eta\rangle - \tau' H(\xi,\eta)\;\Big|\;\eta\in\partial_\xi V(\tau,\xi)\Big\}. \tag{7.7}$$

Indeed, V *is strictly differentiable wherever it is differentiable, which is at almost every point of* $\operatorname{int}\operatorname{dom} V$, *and relative to such points the gradient mapping* ∇V *is continuous.*

Proof. The points $(\tau,\xi)\in\operatorname{int}\operatorname{dom} V$ have been identified in Corollary 7.2 as the ones with $\tau>0$ and $\xi\in\operatorname{int}\operatorname{dom} V(\tau,\cdot)$. Because $V(\tau,\cdot)$ is convex, the mapping $\partial_\xi V(\tau,\cdot)$ is nonempty-compact-valued and locally bounded on $\operatorname{int}\operatorname{dom} V(\tau,\cdot)$, as already known through convex analysis; cf. [2; §24]. These properties carry over to the behavior of $\partial_\xi V$ on $\operatorname{int}\operatorname{dom} V$ because of the epi-continuous dependence of $V(\tau,\cdot)$ on τ in Theorem 2.1; see [2; §24] again. The local boundedness of $\partial_\xi V$, when joined with the formula $\sigma=-H(\xi,\eta)$ in Theorem 5.1 and the continuity of H, gives us the nonempty-compact-valuedness and local boundedness of ∂V.

The local boundedness of ∂V on $\operatorname{int}\operatorname{dom} V$ implies that V is Lipschitz continuous there locally; cf. [6; 9.13]. Then from having $\hat\partial V(\tau,\xi)=\partial V(\tau,\xi)$ in Theorem 2.5 we get the subdifferential regularity of V on $\operatorname{int}\operatorname{dom} V$ and the convexity of $\partial V(\tau,\xi)$ (because $\hat\partial V(\tau,\xi)$ is always convex). Local Lipschitz continuity and subdifferential regularity yield semidifferentiability by [6; 9.16]. Formula (7.7) specializes the semiderivative formula in that result to V by way of the description of $\partial V(\tau,\xi)$ in Theorem 2.5.

By virtue of being locally Lipschitz continuous, V is differentiable almost everywhere on $\operatorname{int}\operatorname{dom} V$. In the presence of subdifferential regularity, the differentiability is strict and the gradient mapping has the stated continuity property; see [6; 9.20]. $\square$

As a complement to this theorem, we develop further information about $\operatorname{int}\operatorname{dom} V$, utilizing Proposition 7.2 to translate the issue into an investigation of when $\operatorname{int}\operatorname{dom} V_\tau\neq\emptyset$. It will be convenient to work with the calculus of relative interiors and the fact that, for a convex set C in a space $I\!R^d$, one has $\operatorname{int} C\neq\emptyset$ if and only $\operatorname{aff} C = I\!R^d$ (i.e., C isn't included in any hyperplane in $I\!R^d$), in which case $\operatorname{int} C=\operatorname{ri} C$ (cf. [6; Chapter 2]).

Additional motivation for the following result, besides facilitating use of Theorem 7.3, comes from the fact that the set $\operatorname{dom} V_\tau = \big\{\xi\;\big|\;V(\tau,\xi)\big\}$ is the *reachable set* at time τ, giving the points $\xi=x(\tau)$ reached by arcs $x(\cdot)\in\mathcal{A}^1_n[0,\tau]$ that start in $\operatorname{dom} g$ and have finite running cost $\int_0^\tau L(x(t),\dot x(t))dt$.

Proposition 7.4 (relative interiors of reachable sets). *For every* $\tau\in[0,\infty)$ *one has*

$$\emptyset\neq\operatorname{ri}\operatorname{dom} V_\tau = \big\{\xi\;\big|\;\operatorname{ri}\operatorname{dom} g\cap\operatorname{ri}\operatorname{dom} E(\tau,\cdot,\xi)\neq\emptyset\big\}. \tag{7.8}$$

Here $\operatorname{ri}\operatorname{dom} V_\tau$ *reduces to* $\operatorname{int}\operatorname{dom} V_\tau$ *if and only if there exists* $\xi\in\operatorname{dom} V_\tau$ *such that* $\operatorname{dom} g\cup\operatorname{dom} E(\tau,\cdot,\xi)$ *does not lie in a hyperplane, that being true then for all* $\xi\in\operatorname{dom} V_\tau$.

Proof. Let $D_\tau=\operatorname{dom} V_\tau$ so $D_0=\operatorname{dom} g$. Clearly D_τ is the image under $(\xi',\xi)\mapsto\xi$ of $C:=\operatorname{dom} E(\tau,\cdot,\cdot)\cap[D_0\times I\!R^n]$, all these sets being convex and nonempty. Then, under the same projection mapping, $\operatorname{ri} D_\tau$ is the image of $\operatorname{ri} C$; cf. [6; 2.44]. For each ξ the convex set $\operatorname{dom} E(\tau,\cdot,\xi)$ is nonempty by Corollary 4.4; likewise for each ξ' the convex set $\operatorname{dom} E(\tau,\xi',\cdot)$ is nonempty. The rule for relative interiors in product spaces (cf. [6; 2.43]) says then that

$$\operatorname{ri}\operatorname{dom} E(\tau,\cdot,\cdot) = \big\{(\xi',\xi)\;\big|\;\xi'\in\operatorname{ri}\operatorname{dom} E(\tau,\cdot,\xi)\big\} = \big\{(\xi',\xi)\;\big|\;\xi\in\operatorname{ri}\operatorname{dom} E(\tau,\xi',\cdot)\big\}. \tag{7.9}$$

This relative interior meets the set $\operatorname{ri}[D_0\times I\!R^n]=\operatorname{ri} D_0\times I\!R^n$, as seen from the second of the expressions in (7.9) by taking any $\xi'\in\operatorname{ri} D_0$ and then any $\xi\in\operatorname{ri}\operatorname{dom} E(\tau,\xi',\cdot)$. The rule for relative interiors of intersections (cf. [6; 2.42]) then yields

$$\operatorname{ri} C = [\operatorname{ri}\operatorname{dom} E(\tau,\cdot,\cdot)]\cap[\operatorname{ri} D_0\times I\!R^n].$$

Returning to the observation that D_τ is the projection of $\operatorname{ri} C$, and utilizing the first of the expressions in (7.9), we get (7.8).

For the claim about interiors, we have to show that the stated condition on a point $\xi \in D_\tau$ is equivalent to the nonexistence of a hyperplane $M \supset D_\tau$. Fix any $\bar\xi \in D_\tau$ and any $\bar\xi' \in D_0$ with $(\bar\xi', \bar\xi) \in \operatorname{dom} E(\tau, \cdot, \cdot)$. A vector ζ gives a hyperplane $M = \{\xi \,|\, \langle \xi, \zeta\rangle = \alpha\}$ that includes D_τ if and only if $\zeta \neq 0$ and $\pm\zeta \in N_{D_\tau}(\bar\xi)$, this being the normal cone to D_τ at $\bar\xi$. Likewise, a vector ζ' gives a hyperplane $M' = \{\xi' \,|\, \langle \xi', \zeta'\rangle = \alpha'\}$ that includes both D_0 and $\operatorname{dom} E(\tau, \cdot, \bar\xi)$ if and only if $\zeta' \neq 0$ and both $\pm\zeta' \in N_{D_0}(\bar\xi')$ and $\pm\zeta' \in N_{\operatorname{dom} E(\tau,\cdot,\bar\xi)}(\bar\xi')$. (Here we appeal to the fact that $\bar\xi'$ belongs to both D_0 and $\operatorname{dom} E(\tau, \cdot, \bar\xi)$.) From the calculus of normals to convex sets (cf. [2; §23], [6; Chapter 6]), the cone $N_{\operatorname{dom} E(\tau,\cdot,\bar\xi)}(\bar\xi')$ is the projection of the cone $N_{\operatorname{dom} E(\tau,\cdot,\cdot)}(\bar\xi', \bar\xi)$:

$$\pm\zeta' \in N_{\operatorname{dom} E(\tau,\cdot,\bar\xi)}(\bar\xi') \iff \exists\, \zeta \text{ with } \pm(\zeta', \zeta) \in N_{\operatorname{dom} E(\tau,\cdot,\cdot)}(\bar\xi', \bar\xi);$$

this relies on the nonemptiness of $\operatorname{dom} E(\tau, \cdot, \xi)$ for all $\xi \in I\!R^n$ (cf. Corollary 4.4), which in turn ensures that ζ' must be nonzero in this formula when $\zeta \neq 0$. Further calculus, utilizing the set relations that were developed above in determining $\operatorname{ri} D_\tau$, reveals that $\pm\zeta \in N_{D_\tau}(\bar\xi)$ if and only if $(0, \pm\zeta) \in N_C(\bar\xi', \bar\xi)$, and on the other hand that

$$\begin{aligned} N_C(\bar\xi', \bar\xi) &= N_{\operatorname{dom} E(\tau,\cdot,\cdot)}(\bar\xi', \bar\xi) + N_{D_0 \times I\!R^n}(\bar\xi', \bar\xi), \\ &\qquad\qquad \text{where } N_{D_0 \times I\!R^n}(\bar\xi', \bar\xi) = N_{D_0}(\bar\xi') \times \{0\}. \end{aligned}$$

Thus, having a $\zeta \neq 0$ such that $\pm\zeta \in N_{D_\tau}(\bar\xi)$ corresponds to having a $\zeta' \neq 0$ such that $\pm\zeta' \in N_{D_0}(\bar\xi')$ and $\pm(\zeta', \zeta) \in N_{\operatorname{dom} E(\tau,\cdot,\cdot)}(\bar\xi', \bar\xi)$. This yields the claimed equivalence. □

Corollary 7.5 (interiors of reachable sets)**.** *If* $\operatorname{int} \operatorname{dom} g \neq \emptyset$, *then for every* $\tau \in [0, \infty)$,

$$\emptyset \neq \operatorname{int} \operatorname{dom} V_\tau = \{\xi \,|\, \operatorname{int} \operatorname{dom} g \cap \operatorname{dom} E(\tau, \cdot, \xi) \neq \emptyset\}.$$

Proof. For convex sets C_1 and C_2 with $\operatorname{int} C_2 \neq \emptyset$, one has $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 \neq \emptyset$ if and only if $C_1 \cap \operatorname{int} C_2 \neq \emptyset$. Then too, $C_1 \cup C_2$ cannot lie in a hyperplane. □

Corollary 7.6 (propagation of finiteness)**.**

(a) *If g is finite on $I\!R^n$, then V is finite on $[0, \infty) \times I\!R^n$.*

(b) *If L is finite on $I\!R^n \times I\!R^n$, then V is finite on $(0, \infty) \times I\!R^n$.*

Proof. We get (a) immediately from Corollary 7.5 as the case where $\operatorname{int} \operatorname{dom} g = I\!R^n$. We get (b) by observing that, for $\tau > 0$, $\operatorname{dom} E(\tau, \cdot, \cdot)$ is all of $I\!R^n \times I\!R^n$ when L is finite. □

Corollary 7.7 (propagation of coercivity)**.**

(a) *If g is coercive, then V_τ is coercive for every $\tau \in [0, \infty)$.*

(b) *If L is coercive, then V_τ is coercive for every $\tau \in (0, \infty)$.*

Proof. We rely on the fact that a proper convex function is coercive if and only if its conjugate is finite [6; 11.5]. The claims are justified then by the duality between V_τ and $\tilde V$ in Theorem 5.1 and that between L and $\tilde L$ in (2.15). □

References

1. R.T. Rockafellar, "Conjugate convex functions in optimal control and the calculus of variations," J. Math. Analysis Appl. 32 (1970), 174–222.
2. R. T. Rockafellar, "Generalized Hamiltonian equations for convex problems of Lagrange," Pacific J. Math. 33 (1970), 411–428.
3. R. T. Rockafellar, "Existence and duality theorems for convex problems of Bolza," Trans. Amer. Math. Soc. 159 (1971), 1–40.
4. R.T. Rockafellar, "Semigroups of convex bifunctions generated by Lagrange problems in the calculus of variations," Math. Scandinavica 36 (1975), 137–158.
5. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
6. R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, 1997.
7. H. Frankowska, "Optimal trajectories associated with a solution of the contingent Hamilton-Jacobi equation," Appl. Math. Optim. 19 (1989), 291–311.
8. H. Frankowska, "Lower semicontinuous solutions of Hamilton-Jacobi-Bellman equations," SIAM J. Control. Opt. 31 (1993), 257-272.
9. M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, 1997.
10. R. T. Rockafellar, "Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis," SIAM J. Control Opt. 34 (1996), 1300–1315.
11. R. T. Rockafellar and P. D. Loewen, "New necessary conditions for the generalized problem of Bolza," SIAM J. Control Opt. 34 (1996), 1496–1511.
12. E. N. Barron and R. Jensen, "Semicontinuous viscosity solutions for Hamilton-Jacobi equations with convex Hamiltonians," Commun. PDE 15 (1990), 1713–1742.
13. M. G. Crandall, L. C. Evans and P.-L. Lions, "Some properties of viscosity solutions of Hamilton-Jacobi equations," Trans. Amer. Math. Soc. 282 (1984), 478–502.
14. R.T. Rockafellar, "Optimal arcs and the minimum value function in problems of Lagrange," J. Optim. Theory & Appl. 12 (1973), 53–83.
15. J.-P. Aubin and A. Cellina, *Differential Inclusions*, Springer-Verlag, Berlin, 1984.
16. F. H. Clarke, "Hamiltonian analysis of the generalized problem of Bolza," Trans. Amer. Math. Soc. 301 (1987), 385–400.
17. A. I. Subbotin, "Generalization of the main equation of differential game theory," J. Opt. Theory Appl. 43 (1984), 103–133.