HAMILTON-JACOBI THEORY AND PARAMETRIC ANALYSIS IN FULLY CONVEX PROBLEMS OF OPTIMAL CONTROL

R. T. Rockafellar[1]

Department of Mathematics, Box 354350
University of Washington, Seattle, WA 98195-4350
rtr@math.washington.edu

Abstract

For optimal control problems satisfying convexity conditions in the state as well as the velocity, the optimal value is studied as a function of the time horizon and other parameters. Conditions are identified in which this optimal value function is locally Lipschitz continuous and semidifferentiable, or even differentiable. The Hamilton-Jacobi theory for such control problems provides the framework in which the results are obtained.

Key words: Hamilton-Jacobi theory, value functions, Bolza problems, calculus of variations, optimal control, cost-to-go, variational analysis, convex analysis, nonsmooth analysis

December, 2002

1 Introduction

A very wide variety of problems in optimal control can be posed in the form of a generalized problem of Bolza in the calculus of variations,

$$\text{minimize } \int_0^\tau L(t, x(t), \dot{x}(t))dt + l(x(0), x(\tau)),$$

by allowing the functions L and l in the formulation to take values in $\overline{I\!R} = [-\infty, \infty]$ instead of just $I\!R = (-\infty, \infty)$. For instance a control problem of the type

$$\begin{aligned} &\text{minimize } \int_0^\tau f(t, x(t), u(t))dt + h(x(\tau)) \text{ subject to} \\ &\dot{x}(t) \in F(t, x(t), u(t)),\ u(t) \in U(t),\ x(0) = a,\ x(\tau) \in E, \end{aligned}$$

is covered by letting $l(b, c) = h(c)$ if $b = a$ and $c \in E$, but $l(b, c) = \infty$ otherwise, and letting $L(t, x, v)$ the infimum of $f(t, x, u)$ over all $u \in U(t)$ such that $F(t, x, u) \ni v$. (When there is no such u, the infimum is ∞, by definition.)

*Research partially supported by the NSF Grant DMS–0104055.

In this paper, we concentrate on a class of problems that fit this picture, emphasizing convexity while looking at parameters which influence the solutions. The basic model we adopt is

$$\mathcal{P}(\pi,\tau) \qquad \text{minimize } g(x(0)) + \int_0^\tau L(x(t),\dot{x}(t))dt + h(\pi,\tau,x(\tau)) \text{ over all } x \in \mathcal{A}_n^1[0,\tau],$$

where $\mathcal{A}_n^1[0,\tau]$ is the space of absolutely continuous functions $x(\cdot) : [0,\tau] \to I\!R^n$ (*arcs*), and π is a parameter vector ranging over an open set $O \subset I\!R^d$. Our interest lies in studying the effects of π and the time parameter τ on the optimal value in $\mathcal{P}(\pi,\tau)$. In other words, we aim at understanding properties of the value function p defined by

$$p(\pi,\tau) := \inf \mathcal{P}(\pi,\tau) \quad \text{for} \quad (\pi,\tau) \in O \times (0,\infty). \tag{1}$$

For a function such as p, produced through optimization, continuity cannot usually be expected, let alone differentiability. However, we will be able to identify some situations where p does possess directional derivatives in a strong sense, and even cases where p is smooth, i.e., belongs to $\mathcal{C}^1$. This will be accomplished by relying on convexity assumptions in the state arguments and utilizing tools in convex analysis and general variational analysis [12].

Basic Assumptions (A).

(A0) *The function g is convex, proper and lsc on $I\!R^n$.*

(A1) *The function L is convex, proper and lsc on $I\!R^n \times I\!R^n$.*

(A2) *The set $F(x) := \{v \mid L(x,v) < \infty\}$ is nonempty for all x, and there is a constant ρ such that $\operatorname{dist}(0,F(x)) \leq \rho(1+|x|)$ for all x.*

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

(A4) *The function h is finite on $O \times (0,\infty) \times I\!R^n$, where O is an open subset of $I\!R^d$, and $h(\pi,\tau,\xi)$ is convex with respect to ξ.*

The joint convexity of $L(x,v)$ in x and v in (A1), combined with the convexity in (A0) and (A4), is the hallmark of "full" convexity. Control problems enjoying full convexity were first investigated in depth in the 1970's, cf. [7], [8], [9], [10], [11]. In such problems, locally optimal solutions are globally optimal, and there are numerous other features in the global optimization category as well.

Assumptions (A0)–(A3) come out of the Hamilton-Jacobi theory for fully convex problems of Bolza as presented in [13] and [14] (see also [15] amd [5]), and they go back even earlier to the cited work in the 1970's through [11]. The properness of an extended-real-valued function means that it does not take on $-\infty$, but is not identically ∞; "lsc" abbreviates lower semicontinuous. The growth condition in (A3) serves in place of a Tonelli condition (much stronger), which would be unworkable for control applications. Assumption (A2) imposes a very weak kind of linear growth on the differential inclusion that underlies the problem. Note that it excludes implicit state constraints (which would be signaled by F being empty-valued in some regions of $I\!R^n$).

In terms of the associated Hamiltionian function H, defined through the Legendre-Fenchel transform by

$$H(x,y) := \sup_v \{v\cdot y - L(x,v)\} \tag{2}$$

and yielding L back through the reciprocal formula

$$L(x,v) = \sup_y \{v\cdot y - H(x,y)\}, \tag{3}$$

assumptions (A1)–(A3) correspond to H being finite on $I\!R^n \times I\!R^n$ with $H(x,y)$ convex in x and concave in y, and also satisfying certain mild growth conditions which are symmetric with respect to the x and y arguments; cf. [13, Theorem 2.3].

The connection with Hamilton-Jacobi theory arises through consideration of the auxiliary problem

$$\mathcal{Q}(\tau,\xi) \qquad \text{minimize } g(x(0)) + \int_0^\tau L(x(t),\dot{x}(t))dt \text{ over all } x \in \mathcal{A}_n^1[0,\tau] \text{ having } x(\tau)=\xi$$

and its value function

$$V(\tau,\xi) := \begin{cases} \inf(\mathcal{Q}(\tau,\xi)) & \text{when } \tau > 0, \\ g(\xi) & \text{when } \tau = 0, \end{cases} \tag{4}$$

which represents the forward propagation of g with respect to L. In particular, g could be the indicator function of a given point a: one could have $g(\xi) = 0$ if $\xi = a$, but $g(\xi) = \infty$ if $\xi \neq a$.

Properties of V under assumptions (A0)–(A3) were recently studied in great detail in [13] and [14]. Since the behavior of $V(\tau,\xi)$ with respect to ξ typically has to be distinguished from its behavior with respect to τ, it is helpful to introduce the notation

$$V_\tau := V(\tau,\cdot) : I\!R^n \to \overline{I\!R} \tag{5}$$

and think of V_τ as an extended-real-valued function on $I\!R^n$ which "moves" as τ goes from 0 to ∞. In [13, Theorem 2.1], it was demonstrated that V_τ is convex, proper and lsc, and depends epi-continuously on τ (i.e., its epigraph depends continuously on τ in the sense of set convergence, a topic expounded for instance in [12]).

The "motion" of V_t has been characterized by a *generalized Hamilton-Jacobi equation* in terms of the subgradient mapping ∂V of V as a whole. It was proved in [13, Theorem 2.5] that

$$\sigma + H(\xi,\eta) = 0 \quad \text{for all} \quad (\sigma,\eta) \in \partial V(\tau,\xi) \quad \text{when} \quad \tau > 0, \tag{6}$$

and indeed, the even stronger property holds that

$$(\sigma,\eta) \in \partial V(\tau,\xi) \quad \Longleftrightarrow \quad \eta \in \partial V_\tau(\xi) \text{ and } \sigma = -H(\xi,\eta). \tag{7}$$

The subgradients in (6) follow the definition patterns in [12], which omit the convexification step of Clarke [2], but in the case of V they have actually been shown in [13] to coincide with Clarke's subgradients. In (7), ∂V_τ is the subgradient mapping of convex analysis [6] associated with the convex function V_τ.

In fact, V is the *unique* solution to (6). This was not known in [13], but was established subsequently by Galbraith [3], [4], by way of new uniqueness Hamilton-Jacobi theorems extending beyond the framework of full convexity and also beyond that of viscosity methodology (e.g. as seen in [1]).

An elementary but fundamental relationship between p and the more basic value function V will serve as the key to our analysis here. It concerns the subproblem

$$\hat{\mathcal{P}}(\pi,\tau) \qquad \text{minimize } V(\tau,\xi) + h(\pi,\tau,\xi) \text{ over all } \xi \in I\!R^n,$$

which is aimed at capturing the *finite-dimensional* aspect of the infinite-dimensional optimization problem $\mathcal{P}(\pi,\tau)$. Note that the convexity of $h(\pi,\tau,\cdot)$ in (A4) ensures the convexity of the function of ξ being minimized in $\hat{\mathcal{P}}(\pi,\tau)$.

Proposition 1 (value function reduction). *The optimal value function p for $\mathcal{P}(\tau,\xi)$ is simultaneously the optimal value function for $\hat{\mathcal{P}}(\tau,\xi)$:*

$$p(\pi,\tau) = \inf \hat{\mathcal{P}}(\pi,\tau) = \inf \mathcal{P}(\pi,\tau). \tag{8}$$

Furthermore, optimal solutions to these problems are connected by

$$x(\cdot) \in \operatorname{argmin} \mathcal{P}(\pi,\tau) \iff x(\cdot) \in \operatorname{argmin} \mathcal{Q}(\tau,\xi) \text{ for some } \xi \in \operatorname{argmin} \hat{\mathcal{P}}(\pi,\tau). \tag{9}$$

Proof. These relationships are evident from the definitions. □

This decomposition, along with properties of V and $\mathcal{Q}(\tau,\xi)$ developed in [13] and [14] will furnish the platform for understanding p.

It is known from [13, Theorem 5.2] that $\operatorname{argmin} \mathcal{Q}(\tau,\xi)$, the optimal solution set in $\mathcal{Q}(\tau,\xi)$, is nonempty whenever the pair $(\tau,\xi) \in (0,\infty) \times I\!R^n$ is such that $V(\tau,\xi) < \infty$; moreover, if $\partial V_\tau(\xi) \neq \emptyset$, every $x(\cdot) \in \operatorname{argmin} \mathcal{Q}(\tau,\xi)$ must belong to $\mathcal{A}_n^\infty[0,\tau]$, the space of *Lipschitz continuous* arcs (having $\dot{x}$ in $\mathcal{L}_n^\infty[0,\tau]$ instead of just $\mathcal{L}_n^1[0,\tau]$). Through this result on the existence of solutions $x(\cdot)$ to $\mathcal{Q}(\tau,\xi)$, the question of the existence of solutions to $\mathcal{P}(\pi,\tau)$ is reduced to that of the existence of solutions ξ to $\hat{\mathcal{P}}(\pi,\tau)$.

Optimality conditions for $\mathcal{P}(\pi,\tau)$ likewise can be reduced to those for $\hat{\mathcal{P}}(\pi,\tau)$, which in turn may be derived from convex analysis in terms of subgradients of V and h with respect to their ξ argument. Hamiltonian trajectories give major support in this, because of their tie to the subgradients of V. A *Hamiltonian trajectory* over an interval $I \subset I\!R$ is a trajectory $(x(\cdot),y(\cdot)) \in \mathcal{A}_n^1[I] \times \mathcal{A}_n^1[I]$ of the generalized Hamiltonian dynamical system

$$\dot{x}(t) \in \partial_y H(x(t),y(t)), \qquad -\dot{y}(t) \in \partial_x[-H](x(t),y(t)), \tag{10}$$

where the subgradients are those of convex analysis for the convex functions $H(x,\cdot)$ and $H(\cdot,y)$.

The differential inclusion (10) is very close to a differential equation, because $\partial_y H(x,y)$ and $\partial_x[-H](x,y)$ are singletons for almost every $(x,y) \in I\!R^n \times I\!R^n$; cf. [13, Proposition 6.1]. One has

$$\eta \in \partial V_\tau(\xi) \iff \begin{cases} \exists \text{ Hamiltonian trajectory } (x(\cdot),y(\cdot)) \text{ with} \\ y(0) \in \partial g(x(0)) \text{ and } (x(\tau),y(\tau)) = (\xi,\eta). \end{cases} \tag{11}$$

This prescription, from [13, Theorem 2.4], provides an *extended method of characteristics*, in subgradient form, which operates *globally* for solving the Hamilton-Jacobi equation in (6).

The existence of an arc $y(\cdot)$ satisfying with $x(\cdot)$ the condition in (11) is always sufficient for having $x(\cdot) \in \operatorname{argmin} \mathcal{Q}(\tau,\xi)$, and it is necessary if $\partial V_\tau(\xi) \neq \emptyset$ (which holds in particular if ξ is in the relative interior of the convex set $\operatorname{dom} V_\tau = \{\xi \mid V_\tau(\xi) < \infty\}$); cf. [13, Theorem 6.3].

Another object that will be crucial in our endeavor is the *dualizing kernel* associated with the Lagrangian L, which is the function K on $[0,\infty) \times I\!R^n \times I\!R^n$ defined by

$$K(\tau,\xi,\omega) := \inf\Big\{ x(0)\cdot\omega + \int_0^\tau L(x(t),\dot{x}(t))dt \,\Big|\, x(\tau) = \xi \Big\}. \tag{12}$$

for $\tau > 0$ and extended to $\tau = 0$ by

$$K(0,\xi,\omega) = \xi\cdot\omega. \tag{13}$$

This function, introduced in [14], is known to be finite everywhere, convex with respect to ξ, concave with respect to ω, and continuously differentiable with respect to τ, and it satisfies a generalized Hamilton-Jacobi equation of Cauchy type in the strong form

$$-\frac{\partial K}{\partial \tau}(\tau,\xi,\omega) = H(\xi,\eta) \text{ for all } \eta \in \partial_\xi K(\tau,\xi,\omega), \tag{14}$$

with (13) as initial condition [14, Theorem 3.1]. The results of Galbraith [3], [4], establish that $K(\cdot,\cdot,\omega)$ is the *unique* solution to this Hamilton-Jacobi equation in τ and ξ. Earlier only a weaker version of uniqueness, depending on the convexity-concavity and a dual Hamiltonian-Jacobi equation, had been verified in [14]. The dualizing kernel K yields a *lower envelope representation* of V:

$$V(\tau,\xi) = \sup_\omega\Big\{K(\tau,\xi,\omega) - g^*(\omega)\Big\}, \tag{15}$$

cf. [14, Theorem 2.5], where g^* is the convex function that is conjugate to g under the Legendre-Fenchel transform,

$$g^*(y) := \sup_x\Big\{x\cdot y - g(x)\Big\}, \qquad g(x) := \sup_y\Big\{x\cdot y - g^*(y)\Big\}. \tag{16}$$

In our focus on the parametric analysis of problem $\mathcal{P}(\pi,\tau)$, we will eventually require certain other properties besides the ones already listed in (A).

Additional Assumptions (A′).

(A5) *The function g on $I\!R^n$ is coercive.*

(A6) *The function h on $O \times (0,\infty) \times I\!R^n$ has the property that $h(\pi,\tau,\xi)$ is differentiable with respect to (π,τ) for each ξ, and the gradient in these arguments depends continuously on (π,τ,ξ).*

Coercivity of g in (A5) means that $g(\xi)/|\xi| \to \infty$ as $|\xi| \to \infty$; here $|\cdot|$ denotes the Euclidean norm. This growth condition on g is equivalent to the finiteness of the conjugate function g^*.

The smoothness in (A6) is destined for establishing a property of p called *semidifferentiability.* In general for a function f on an open subset of $I\!R^m$, semidifferentiability means that, at each point z of that subset, the difference quotient functions

$$\Delta_\epsilon f(z)(z') := [f(z+\epsilon z') - f(z)]/\epsilon \quad \text{for } \epsilon > 0$$

(which are defined for z' in a neighborhood of 0 that expands to fill all of $I\!R^m$ as $\epsilon \searrow 0$) converge uniformly on bounded sets to a finite function on $I\!R^m$. This concept is examined from many angles in [12, 7.21]. The limit function, symbolized by $df(z)$ and thus having values denoted by $df(z)(z')$, need not be a linear function, but when it is, semidifferentiability turns into ordinary differentiability. In the presence of local Lipschitz continuity, semidifferentiability is equivalent to the existence of one-sided directional derivatives: one simply has

$$df(z)(z') = \lim_{\epsilon \searrow 0}[f(z+\epsilon z') - f(z)]/\epsilon.$$

In particular, any finite convex function on $I\!R^n$ is locally Lipschitz continuous and semidifferentiable everywhere [12, 9.14 and 7.27]. As another example, the dualizing kernel K was itself shown in [14, Theorem 3.6] to be locally Lipschitz continuous and semidifferentiable with respect to all of its arguments.

2 Main Developments

In obtaining the semidifferentiability of p, along with subgradient properties of p that allow the identification of cases in which p is smooth, several consequences of our assumptions (A4) and (A6) on the terminal cost function h will be needed. These consequences will be gleaned by the methodology of variational analysis in [12].

Proposition 2 (joint properties of the terminal function)**.** *Assumptions (A4) and (A6) on the separate functions*

$$h_{\pi,\tau} = h(\pi,\tau,\cdot), \qquad h_\xi = h(\cdot,\cdot,\xi), \tag{17}$$

guarantee that h has the following properties, involving all of its arguments together.

(a) *h is locally Lipschitz continuous on $O \times (0,\infty) \times I\!R^n$.*

(b) *h is semidifferentiable on $O \times (0,\infty) \times I\!R^n$ with subderivative formula*

$$dh(\pi,\tau,\xi)(\pi',\tau',\xi') = \nabla h_\xi(\pi,\tau)\cdot(\pi',\tau') + dh_{\pi,\tau}(\xi)(\xi'). \tag{18}$$

(c) *h has its subgradients on $O \times (0,\infty) \times I\!R^n$ given by*

$$\partial h(\pi,\tau,\xi) = \{(\rho,\sigma,\eta)\,|\,(\rho,\sigma) = \nabla h_\xi(\pi,\tau),\ \eta \in \partial h_{\pi,\tau}(\xi)\}. \tag{19}$$

(d) *h is subdifferentially regular on $O \times (0,\infty) \times I\!R^n$ (i.e., its epigraph is Clarke regular).*

Proof. Argument for (a). The finite convexity in (A4) implies that $h_{\pi,\tau}$ is locally Lipschitz continuous on $I\!R^n$ for each $(\pi,\tau) \in O \times (0,\infty)$ [12, 9.14]. On the other hand, the smoothness in (A6) implies that h_ξ is locally Lipschitz continuous on $O \times (0,\infty)$ for each $\xi \in I\!R^n$. It is elementary then that $h(\pi,\tau,\xi)$ is locally Lipschitz continuous with respect to (π,τ,ξ).

Argument for (b). By virtue of (A4), $h_{\pi,\tau}$ is semidifferentiable on $I\!R^n$ for each $(\pi,\tau) \in O \times (0,\infty)$ [12, 7.27]. To get the semidifferentiability of h itself, utilizing the differentiability in (A6), we observe that $\Delta_\epsilon h(\pi,\tau,\xi)(\pi',\tau',\xi')$ can be written as

$$\frac{h(\pi+\epsilon\pi',\tau+\epsilon\tau',\xi+\epsilon\xi') - h(\pi,\tau,\xi+\epsilon\xi')}{\epsilon} + \frac{h(\pi,\tau,\xi+\epsilon\xi') - h(\pi,\tau,\xi)}{\epsilon}, \tag{20}$$

where by the mean value theorem the first term in the sum has the representation

$$\frac{h(\pi+\epsilon\pi',\tau+\epsilon\tau',\xi+\epsilon\xi') - h(\pi,\tau,\xi+\epsilon\xi')}{\epsilon} = \nabla_{\pi,\tau} h(\pi+\theta\pi',\tau+\theta\tau',\xi+\epsilon\xi')\cdot(\pi',\tau')$$

for some $\theta \in (0,\epsilon)$ (depending on the various arguments). The continuous dependence of the gradient in (A6) allows us to deduce from this representation that, as a function of (π',τ',ξ') for each ϵ, the first term in the sum in (20) converges uniformly, as $\epsilon \searrow 0$, to the linear function given by the expression $\nabla_{\pi,\tau}(\pi,\tau,\xi)\cdot(\pi',\tau')$. Of course, the second term in the sum in (20), as a function of ξ', converges uniformly as $\epsilon \searrow 0$ because of the semidifferentiability of h in its ξ argument that comes from (A4). Altogether, then, we do have the convergence property that is required by the definition of h being semidifferentiable in all of its arguments. The limit calculations have confirmed also that the semiderivatives are given by (18).

Argument for (c). In the terminology of [12, 8.3], the *regular* subgradient set $\hat{\partial} h(\pi,\tau,\xi)$ consists of all (ρ,σ,η) such that

$$(\rho,\sigma,\eta)\cdot(\pi',\tau',\xi') \leq dh(\pi,\tau,\xi)(\pi',\tau',\xi') \text{ for all } (\pi',\tau',\xi').$$

Through the subderivative formula (18), this comes down to the elements specified on the right side of (19); the right side is thus $\hat{\partial} h(\pi,\tau,\xi)$. By definition, the general subgradient set $\partial h(\pi,\tau,\xi)$ is formed by taking all limits of sequences $\{(\rho^\nu,\sigma^\nu,\eta^\nu)\}_{\nu=1}^\infty$ with $(\rho^\nu,\sigma^\nu,\eta^\nu)\in\hat{\partial} h(\pi^\nu,\tau^\nu,\xi^\nu)$ and $(\pi^\nu,\tau^\nu,\xi^\nu)\to(\pi,\tau,\xi)$ (plus $h(\pi^\nu,\tau^\nu,\xi^\nu)\to h(\pi,\tau,\xi)$, but that is automatic here by (a)). Any such limit (ρ,σ,ξ) must have $(\rho,\sigma)=\nabla h_\xi(\pi,\tau)$ by the gradient continuity in (A6), and it must also have $\eta\in\partial h_{\pi,\tau}(\xi)$; the latter follows because the (finite) convex functions h_{π^ν,τ^ν} converge pointwise to $h_{\pi,\tau}$; see [6, Sec. 24]. Hence $\partial h(\pi,\tau,\xi)=\hat{\partial} h(\pi,\tau,\xi)$.

Argument for (d). Because h is locally Lipschitz continuous (and therefore has no nontrivial "horizon subgradients" [12, 9.13]), the equality between $\partial h(\pi,\tau,\xi)$ and $\hat{\partial} h(\pi,\tau,\xi)$, just verified, guarantees the subdifferential regularity of h [12, 8.11]. □

For the important role it will have in our analysis, we next introduce alongside of $\hat{\mathcal{P}}(\pi,\tau)$ the following *dual problem*:

$$\hat{\mathcal{P}}^*(\pi,\tau) \qquad \text{maximize } j(\pi,\tau,\eta)-V_\tau^*(\eta) \text{ over all } \eta\in\mathbb{R}^n,$$

where V_τ^* is the convex function conjugate to V_τ, and j is the function defined by

$$j(\pi,\tau,\eta)=\inf_\xi\{h(\pi,\tau,\xi)+\eta\cdot\xi\}. \tag{21}$$

Here $j(\pi,\tau,\cdot)$ is the concave conjugate of $-h(\pi,\tau,\cdot)$, so $\hat{\mathcal{P}}(\pi,\tau)$ and $\hat{\mathcal{P}}^*(\pi,\tau)$ are optimization problems dual to each other in the original sense of Fenchel; cf. [6, Sec. 31]. It is interesting to note, although it will not be needed, that V_τ^* can be identified with the value function that is defined like V_τ but for the forward propagation of g^* with respect to a certain Lagrangian dual to L; see [13, Theorem 5.1].

Theorem 1 (parametric optimality)**.** *For every $(\pi,\tau)\in O\times(0,\infty)$, the optimal value in problem $\hat{\mathcal{P}}(\pi,\tau)$, which is $p(\pi,\tau)$, is finite and agrees with the optimal value in the dual problem $\hat{\mathcal{P}}^*(\pi,\tau)$. The optimal solution sets*

$$X(\pi,\tau):=\operatorname{argmin}\hat{\mathcal{P}}(\pi,\tau), \qquad Y(\pi,\tau):=\operatorname{argmax}\hat{\mathcal{P}}^*(\pi,\tau), \tag{22}$$

are nonempty, convex and compact, and they are characterized by

$$(\xi,\eta)\in X(\pi,\tau)\times Y(\pi,\tau) \iff \eta\in\partial V_\tau(\xi),\ -\eta\in\partial h_{\pi,\tau}(\xi). \tag{23}$$

Proof. The coercivity assumed in (A5) makes V_τ be coercive for every $\tau\in(0,\infty)$; this was proved in [13, Corollary 7.7]. In $\hat{\mathcal{P}}(\pi,\tau)$, we are minimizing the sum of this coercive convex function (which is also proper and lsc) and the finite convex function $h(\pi,\tau,\cdot)$. Such a sum is itself a coercive convex function that is proper and lsc, and its minimum is therefore finite and attained on a compact set.

The finiteness of $h_{\pi,\tau}$ entails, on the same grounds, the coercivity of $-j$ and leads us to the conclusion that the maximum in $\hat{\mathcal{P}}^*(\pi,\tau)$ is attained on a compact set. The fact that the maximum agrees with the minimum, and that the optimal solutions are characterized by the subgradient conditions in (23), is a standard feature of Fenchel duality in these circumstances; cf. [6, Sec. 31]. □

To proceed further than in Theorem 1, we need to verify for the function being minimized in $\hat{\mathcal{P}}(\pi,\tau)$ a boundedness condition which is central to the theory of finite-dimensional parametric minimization, as in [12, 1.17].

Proposition 3 (parametric inf-boundedness property). *Let $(\bar\pi,\bar\tau)\in O\times(0,\infty)$, and consider any $\epsilon>0$ small enough that $(\pi,\tau)\in O\times(0,\infty)$ when $|\pi-\bar\pi|\le\epsilon$ and $|\tau-\bar\tau|\le\epsilon$. Then*

$$\forall\,\lambda\in(0,\infty),\ \exists\,\gamma\in(0,\infty)\ \textit{such that}\ |\xi|\le\gamma\ \textit{when}\ \begin{cases} V(\tau,\xi)+h(\pi,\tau,\xi)\le\lambda\ \textit{with}\\ |\pi-\bar\pi|\le\epsilon\ \textit{and}\ |\tau-\bar\tau|\le\epsilon.\end{cases} \tag{24}$$

Proof. We know that V_τ is coercive and depends epi-continuously on τ. This implies that the conjugate convex function V_τ^* is finite and likewise depends epi-continuously on τ (since epi-continuity is preserved under the Legendre-Fenchel transform [12, 11.34]). But finite convex functions epi-converge if and only if they converge pointwise, uniformly on bounded sets [12, 7.18]. It follows that, for any $\epsilon>0$ and $\alpha>0$, there exist $r>0$ and $s>0$ such that

$$V_\tau^*(\eta')\ \le\ V_{\bar\tau}^*(0)+r|\eta'|+s\ \text{ when }\ |\eta'|\le\alpha,\ |\tau-\bar\tau|\le\epsilon.$$

When conjugates are taken on both sides with respect to η', this inequality translates to

$$V_\tau(\xi)\ \ge\ \alpha\max\{0,|\xi|-r\}-V_{\bar\tau}^*(0)-s\ \text{ when }\ |\tau-\bar\tau|\le\epsilon,$$

but all we will really need is the consequence that

$$\forall\,\alpha>0,\ \exists\,\beta\in I\!R\ \text{such that}\ V_\tau(\xi)\ge\alpha|\xi|-\beta\ \text{for all}\ \xi\ \text{when}\ |\tau-\bar\tau|\le\epsilon. \tag{25}$$

Next we observe that, because h is locally Lipschitz continuous (by Proposition 2(a)), there is a Lipschitz constant κ for h on the neighborhood of $(\bar\pi,\bar\tau,0)$ defined by $|\pi-\bar\pi|\le\epsilon$, $|\tau-\bar\tau|\le\epsilon$, $|\xi|\le\epsilon$. In particular, that yields

$$h(\pi,\tau,0)\ge h(0,0,0)-2\kappa\epsilon \tag{26}$$

and $|h(\pi,\tau,\xi')-h(\pi,\tau,\xi)|\le\kappa|\xi'-\xi|$ when $|\xi|\le\epsilon$ and $|\xi'|\le\epsilon$. The latter ensures for the convex function $h_{\pi,\tau}=h(\pi,\tau,\cdot)$ that

$$\eta\in\partial h_{\pi,\tau}(0)\quad\Longrightarrow\quad|\eta|\le\kappa \tag{27}$$

(see [12, 9.14]). The subgradient set in (27) is nonempty (because $h_{\pi,\tau}$ is finite), and its elements η are characterized by the inequality $h_{\pi,\tau}(\xi)\ge h_{\pi,\tau}(0)+\eta\cdot\xi$ holding for all $\xi\in I\!R^n$. The estimates in (26) and (27) yield through this inequality the lower bound:

$$h(\pi,\tau,\xi)\ge-\kappa|\xi|+h(0,0,0)-2\kappa\epsilon\ \text{ for all }\ \xi\ \text{ when }\ |\pi-\bar\pi|\le\epsilon\ \text{ and }\ |\tau-\bar\tau|\le\epsilon.$$

Returning now to (25) and taking $\alpha>\kappa$, we see there will exist a constant μ such that

$$V(\tau,\xi)+h(\pi,\tau,\xi)\ge(\alpha-\kappa)|\xi|-\mu\ \text{ for all }\ \xi\ \text{ when }\ |\pi-\bar\pi|\le\epsilon\ \text{ and }\ |\tau-\bar\tau|\le\epsilon.$$

Then obviously (24) holds, as needed. □

Theorem 2 (Lipschitz continuity and subgradients of the value function)**.** *The function p is locally Lipschitz continuous on $O \times (0,\infty)$, and its subgradients obey the rule that*

$$(\rho,\sigma) \in \partial p(\pi,\tau) \quad\Longrightarrow\quad \begin{cases} (\rho,\sigma + H(\xi,\eta)) = \nabla h_\xi(\pi,\tau) \text{ for} \\ \text{some } (\xi,\eta) \in X(\pi,\tau) \times Y(\pi,\tau). \end{cases} \tag{28}$$

Proof. Let $f(\pi,\tau,\xi) = V(\tau,\xi) + h(\pi,\tau,\xi)$. The property of f in Proposition 3 is known by [12, 1.17] to ensure that the parametric optimal value $\inf_\xi f(\pi,\tau,\xi)$, which again is $p(\pi,\tau)$, is lsc in its dependence on (π,τ). It further yields by [12, 10.13] the estimate

$$\partial p(\pi,\tau) \subset \{(\rho,\sigma) \,|\, (\rho,\sigma,0) \in \partial f(\pi,\tau,\xi) \text{ for some } \xi \in \operatorname{argmin} \hat{\mathcal{P}}(\pi,\tau)\}. \tag{29}$$

Because h is locally Lipschitz continuous by Proposition 2(a), we can apply the subgradient rule in [12, 10.10] to see that $\partial f(\pi,\tau,\xi) \subset (0, \partial V(\tau,\xi)) + \partial h(\pi,\tau,\xi)$. Invoking (7) and the subgradient formula in Proposition 2(c), along with the subgradient condition (23) that characterizes optimality in $\hat{\mathcal{P}}(\pi,\tau)$ as well as $\hat{\mathcal{P}}^*(\pi,\tau)$, we are able then to pass from (29) to (28).

Another consequence of Proposition 3 is that the mapping $(\pi,\tau) \mapsto \operatorname{argmin} \hat{\mathcal{P}}(\pi,\tau) = X(\pi,\tau)$ is locally bounded with respect to any compact subset C of $\{(\pi,\tau) \in O \times (0,\infty) \,|\, p(\pi,\tau) \leq \lambda\}$, for any λ. The mapping $(\pi,\tau) \mapsto \operatorname{argmin} \hat{\mathcal{P}}^*(\pi,\tau) = Y(\pi,\tau)$ is locally bounded then on such a set C as well; this is true because $\eta \in Y(\pi,\tau)$ implies $-\eta \in \partial h_{\pi,\tau}(\xi)$, and the convex functions $h_{\pi,\tau}$ are Lipschitz continuous on a neighborhood of the compact set $X(\pi,\tau)$, locally uniformly with respect to (π,τ) (by Proposition 2(a)).

It follows from the continuity of the Hamiltonian H that the mapping from (π,τ) in such a set C to the set of (ρ,σ) described on the right side of (28) is locally bounded. That guarantees the boundedness of any sequence of subgradients $(\rho^\nu,\sigma^\nu) \in \partial p(\pi^\nu,\tau^\nu)$ with $(\pi^\nu,\tau^\nu) \to (\pi,\tau)$ and $p(\pi^\nu,\tau^\nu) \to p(\pi,\tau)$. Then, however, p has to be locally Lipschitz continuous (because this boundedness eliminates any nontrivial "horizon subgradients") [12, 9.13(a)]. □

The next stage of our analysis requires a minimax representation of the function p.

Proposition 4 (minimax representation)**.** *The function k defined by*

$$k(\pi,\tau,\xi,\omega) := K(\tau,\xi,\omega) - g^*(\omega) + h(\pi,\tau,\xi) \tag{30}$$

is finite on $O \times (0,\infty) \times I\!R^n \times I\!R^n$, convex in ξ, concave in ω, and moreover locally Lipschitz continuous and semidifferentiable with respect to all arguments. It furnishes the representation

$$p(\pi,\tau) \;=\; \min_{\xi \in R^n} \max_{\omega \in R^n} k(\pi,\tau,\xi,\omega) \;=\; \max_{\omega \in R^n} \min_{\xi \in R^n} k(\pi,\tau,\xi,\omega). \tag{31}$$

Furthermore, the associated saddle point set, which is nonempty, convex and compact, has the form $X(\pi,\tau) \times W(\pi,\tau)$ (for the same $X(\pi,\tau)$ as above, but a set $W(\pi,\tau)$ that is new), and is characterized by

$$(\xi,\omega) \in X(\pi,\tau) \times W(\pi,\tau) \quad\Longleftrightarrow\quad \begin{cases} \exists\, \eta \in Y(\pi,\tau) \text{ and } \zeta \in I\!R^n \text{ with } \omega \in \partial g(\zeta), \\ \text{and a Hamiltonian trajectory } (x(\cdot),y(\cdot)) \text{ with} \\ (x(0),y(0)) = (\zeta,\omega) \text{ and } (x(\tau),y(\tau)) = (\xi,\eta). \end{cases} \tag{32}$$

Proof. The initial claims about k follows from the properties already identified for K, g^* and h. For any finite convex-concave function, in this case $k_{\pi,\tau} = k(\pi, \tau, \cdot, \cdot)$, the set of saddle points is always a product of closed, convex sets. We need to demonstrate this product has the form described, and is bounded.

Let $M(\tau, \xi) = \text{argmax}_\omega \{K(\tau, \xi, \omega) - g^*(\omega)\}$. The maximization half of the condition for a saddle point of $k_{\pi,\tau}$ is simply the condition that $\omega \in M(\tau, \xi)$. For any such ω, we have $K(\tau, \xi, \omega) - g^*(\omega) = V(\tau, \xi)$ by (15). Hence

$$k(\pi, \tau, \xi, \omega) = V(\tau, \xi) + h(\pi, \tau, \xi) \text{ when } \omega \in M(\tau, \xi). \tag{33}$$

By subgradient calculus, the elements $\omega \in M(\tau, \xi)$ are characterized by

$$\exists -\zeta \in \partial_\omega[-K](\tau, \xi, \omega) \text{ such that } \zeta \in \partial g^*(\omega). \tag{34}$$

Similarly, let $N(\pi, \tau, \omega) = \text{argmin}_\xi \{K(\tau, \xi, \omega) + h(\pi, \tau, \xi)\}$, so that the minimization half of the condition for a saddle point of $k_{\pi,\tau}$ corresponds to $\xi \in N(\pi, \tau, \omega)$. That is characterized by 0 being a subgradient of the convex function $K(\tau, \cdot, \omega) + h(\pi, \tau, \cdot)$ at ξ, which through subgradient calculus [6] correponds to

$$\exists\, \eta \in \partial_\xi K(\tau, \xi, \omega) \text{ such that } -\eta \in \partial_\xi h(\pi, \tau, \xi). \tag{35}$$

Having (ξ, ω) be a saddle point means having both $\xi \in N(\pi, \tau, \omega)$ and $\omega \in M(\tau, \xi)$. On the other hand, the conditions $\eta \in \partial_\xi K(\tau, \xi, \omega)$ and $-\zeta \in \partial_\omega[-K](\tau, \xi, \omega)$ in (34) and (35) are, by [14, Theorem 4.1], jointly equivalent to the existence of a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ over $[0, \tau]$ that starts at (ζ, ω) and ends at (ξ, η). The condition $\zeta \in \partial g^*(\omega)$ in (34) is itself equivalent, through conjugacy, to $\omega \in \partial g(\zeta)$. Applying (11) and the characterization of $X(\pi, \tau)$ and $Y(\pi, \tau)$ in (23), we obtain the description in (32) of the saddle point set.

This description confirms in particular the nonemptiness of the saddle point set. It yields the boundedness of $W(\pi, \tau)$ through the fact that the Hamiltonian dynamical system in question takes bounded sets into bounded sets, either forward or backward in time. □

Theorem 3 (semidifferentiability of the value function)**.** *The function p is semidifferentiable, with semiderivative formula of minimax type:*

$$\begin{aligned} dp(\pi, \tau)(\pi', \tau') &= \min_{\xi \in X(\pi,\tau)} \max_{\eta \in Y(\pi,\tau)} \left\{ \nabla h_\xi(\tau, \pi) \cdot (\tau', \pi') - \tau' H(\xi, \eta) \right\} \\ &= \max_{\eta \in Y(\pi,\tau)} \min_{\xi \in X(\pi,\tau)} \left\{ \nabla h_\xi(\tau, \pi) \cdot (\tau', \pi') - \tau' H(\xi, \eta) \right\}. \end{aligned} \tag{36}$$

Proof. We apply Gol'shtein's theorem [12, 11.53] to the minimax representation in Proposition 4. The hypothesis of that theorem is satisfied because k is continuous and semidifferentiable, and the saddle point set is bounded. The direct formula obtained by this route is

$$\begin{aligned} dp(\pi, \tau)(\pi', \tau') &= \min_{\xi \in X(\pi,\tau)} \max_{\omega \in W(\pi,\tau)} dk(\pi, \tau, \xi, \omega)(\pi', \tau', 0, 0) \\ &= \max_{\omega \in W(\pi,\tau)} \min_{\xi \in X(\pi,\tau)} dk(\pi, \tau, \xi, \omega)(\pi', \tau', 0, 0). \end{aligned} \tag{37}$$

We calculate that

$$dk(\pi,\tau,\xi,\omega)(\pi',\tau',0,0) = dK(\tau,\xi,\omega)(\tau',0,0) + dh(\pi,\tau,\xi)(\pi',\tau',0), \tag{38}$$

where the final term is merely $\nabla h_\xi(\tau,\pi)\cdot(\tau',\pi')$ by Proposition 2(b). We then recall from the Hamilton-Jacobi theory of K that $dK(\tau,\xi,\omega)(\tau',0,0)$ equals $-\tau' H(\xi,\eta)$ for any $\eta \in \partial_\xi K(\tau,\xi,\omega)$, or for that matter $-\tau' H(\zeta,\omega)$ for any $-\zeta \in \partial[-K](\tau,\xi,\omega)$; cf. [14, Theorem 3.6]. In that way, utilizing the characterization of these two subgradient conditions in terms of Hamiltonian trajectories as in the preceding proof (through [14, Theorem 4.1]), we obtain from (38) the reduction of (37) to (36). □

Theorem 4 (differentiability of the value function)**.** *Suppose that the function $h_{\pi,\tau} = h(\pi,\tau,\cdot)$ is not just convex, but strictly convex and differentiable. Then $X(\pi,\tau)$ and $Y(\pi,\tau)$ reduce to singletons, and p is smooth (continuously differentiable) with*

$$\nabla p(\pi,\tau) = \nabla h_\xi(\pi,\tau) - (0, H(\xi,\eta)) \quad \text{for the unique } (\xi,\eta) \in X(\pi,\tau)\times Y(\pi,\tau). \tag{39}$$

Proof. The strict convexity ensures that $X(\pi,\tau)$ is a singleton, and the differentiability then makes $Y(\pi,\tau)$ be a singleton because having $\eta \in Y(\pi,\tau)$ entails $\eta = -\nabla h_{\pi,\tau}(\xi)$. Then, in the subgradient estimate of Theorem 2, there is only one candidate for membership in $\partial p(\pi,\tau)$. Since p is locally Lipschitz continous, this implies that p is smooth with this candidate element as its gradient [12, 9.18 and 9.19]. □

Corollary (differentiability of Moreau envelopes)**.** *For $\lambda > 0$, the Moreau envelope function*

$$p(\lambda,\zeta,\tau) = e_\lambda V_\tau(\zeta) = \min_{\xi\in\mathbb{R}^n}\left\{V(\tau,\xi) + \frac{1}{2\lambda}|\xi-\zeta|^2\right\}$$

is continuously differentiable with respect to (λ,ζ,τ).

Proof. Here we take $\pi = (\lambda,\zeta) \in (0,\infty)\times\mathbb{R}^n$ and $h(\pi,\tau,\xi) = |\xi-\zeta|^2/2\lambda$. □

References

[1] Bardi, M. and Capuzzo-Dolcetta, I. (1997), *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser.

[2] Clarke, F. H. (1983), *Optimization and Nonsmooth Analysis*, Wiley.

[3] Galbraith, G. (1999), *Applications of Variational Analysis to Optimal Trajectories and Nonsmooth Hamilton-Jacobi Theory*, Ph. D. thesis, University of Washington, Seattle.

[4] Galbraith, G., The Role of Cosmically Lipschitz Mappings in Nonsmooth Hamilton-Jacobi Theory (preprint).

[5] Goebel, R. and Rockafellar, R. T. (2002), Generalized Conjugacy in Hamilton-Jacobi Theory for Fully Convex Lagrangians, *Convex Analysis* 9.

[6] Rockafellar, R. T. (1970), *Convex Analysis*, Princeton University Press, Princeton, N.J.

[7] Rockafellar, R. T. (1970), Conjugate Convex Functions in Optimal Control and the Calculus of Variations, *J. Mathematical Analysis and Applications* 32, 174–222.

[8] Rockafellar, R. T. (1970), Generalized Hamiltonian Equations for Convex Problems of Lagrange, *Pacific J. Mathematics* 33, 411–428.

[9] Rockafellar, R. T. (1971), Existence and Duality Theorems for Convex Problems of Bolza, *Transactions of the American Mathematical Society* 159, 1–40.

[10] Rockafellar, R. T. (1973), Saddle Points of Hamiltonian Systems in Convex Problems of Lagrange, *J. Optimization Theory and Applications* 12, 367–390.

[11] R. T. Rockafellar, R. T. (1975), Semigroups of Convex Bifunctions Generated by Lagrange Problems in the Calculus of Variations, *Mathematica Scandinavica* 36, 137–158.

[12] Rockafellar, R. T. and Wets, R. J-B (1997), *Variational Analysis*, Springer-Verlag, Berlin.

[13] Rockafellar, R. T. and Wolenski, P. R. (2001), Convexity in Hamilton-Jacobi Theory 1: Dynamics and Duality, *SIAM J. Control and Optimization* 40, 1323–1350.

[14] Rockafellar, R. T. and Wolenski, P. R. (2001), Convexity in Hamilton-Jacobi Theory 2: Envelope Representations, *SIAM J. Control and Optimization* 40, 1351–1372.

[15] Rockafellar, R. T. (2001) Convex Analysis in the Calculus of Variations, in Hadjisavvas, N. and Pardalos, P. M. (eds.), *Advances in Convex Analysis and Global Optimization*, Kluwer, 135–152.