CONVEXITY IN HAMILTON-JACOBI THEORY 2: ENVELOPE REPRESENTATIONS

R. TYRRELL ROCKAFELLAR and PETER R. WOLENSKI *

University of Washington and Louisiana State University

Abstract. Upper and lower envelope representations are developed for value functions associated with problems of optimal control and the calculus of variations that are fully convex, in the sense of exhibiting convexity in both the state and the velocity. Such convexity is used in dualizing the upper envelope representations to get the lower ones, which have advantages not previously perceived in such generality and in some situations can be regarded as furnishing, at least for value functions, extended Hopf-Lax formulas that operate beyond the case of state-independent Hamiltonians.

The derivation of the lower envelope representations centers on a new function called the dualizing kernel, which propagates the Legendre-Fenchel envelope formula of convex analysis through the underlying dynamics. This kernel is shown to be characterized by a kind of double Hamilton-Jacobi equation and, despite overall nonsmoothness, to be smooth with respect to time and concave-convex in the primal and dual states. It furnishes a means whereby, in principle, value functions and their subgradients can be determined through optimization without having to deal with a separate, and typically much less favorable, Hamilton-Jacobi equation for each choice of the initial or terminal cost data.

Keywords. Convex value functions, Hamilton-Jacobi equations, dualizing kernels, fundamental kernels, envelope formulas, Hopf-Lax formulas, viscosity solutions, optimal control.

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

Revised: August 2000

---

* This work was supported in part by the National Science Foundation under grant DMS–9803089 for the first author and grant DMS–9623406 for the second author.

1. Introduction

A major goal of Hamilton-Jacobi theory is the characterization of value functions that arise from problems of optimal control and the calculus of variations in which endpoints are treated as parameters. The value function $V : [0,\infty) \times I\!R^n \to \overline{I\!R} := [-\infty,\infty]$ is defined from a Lagrangian $L : I\!R^n \times I\!R^n \to \overline{I\!R}$ and an function $g : I\!R^n \to \overline{I\!R}$ by

$$\begin{aligned} &V(\tau,\xi) := \inf\Big\{ g\big(x(0)\big) + \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(\tau) = \xi \Big\} \;\text{ for } \tau > 0, \\ &V(0,\xi) := g(\xi), \end{aligned} \tag{1.1}$$

with the minimization taking place over all the arcs (i.e., absolutely continuous functions) $x(\cdot) : [0,\tau] \to I\!R^n$ that reach ξ at time τ. Here $V(\tau,\cdot)$ is viewed as an evolving function on $I\!R^n$ which starts as g and describes how g is propagated forward to any time τ in a manner dictated by L.

Similarly, value functions can be considered that describe how g is propagated backward from a future time T, and such a "cost-to-go" formulation is common in optimal control. From a theoretical perspective, of course, backward models are equivalent to forward models through time reversal and do not require separate treatment in basic theory. The expression of control problems in terms of a Lagrangian L in which control parameters do not appear is parallel to the expression of control dynamics in terms of differential inclusions and has generated a substantial literature in nonsmooth optimization, going back to around 1970. More about that can be found in our companion paper [1], which is the springboard for the efforts here.

In the classical context of the calculus of variations, g and L would be smooth (i.e., continuously differentiable). For applications such as in control, however, it is important to allow g and L to be nonsmooth and even to take on ∞, because infinite penalties can systematically be used in incorporating constraints. Under the assumption that g and L are lower semicontinuous (lsc) and proper (i.e., not identically ∞, and nowhere having the value $-\infty$), the integrand $t \mapsto L(x(t),\dot{x}(t))$ is measurable, and the functional $J[x(\cdot)]$ being minimized is well defined. (The usual convention of "inf addition" is followed, in which ∞ dominates in any conflict with $-\infty$.) Then $J[x(\cdot)] = \infty$ unless the arc $x(\cdot)$ satisfies the constraints:

$$\begin{aligned} &x(0) \in D, \;\text{ where }\; D := \big\{ x \,\big|\, g(x) < \infty \big\}, \\ &\dot{x}(t) \in F(x(t)) \;\text{ a.e. } t, \;\text{ where }\; F(x) := \big\{ v \,\big|\, L(x,v) < \infty \big\}. \end{aligned} \tag{1.2}$$

The customary tool for characterizing value functions is the Hamilton-Jacobi PDE in one form or another. It revolves around the Hamiltonian function H associated with L, which is defined through the Legendre-Fenchel transform by

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

Because V typically lacks smoothness, even when g and L are smooth, various generalizations of the classical PDE have been devised, the foremost being "viscosity" versions. The recent book of Bardi and Capuzzo-Dolcetta [2], with its helpful references, provides broad access to that subject. Viscosity theory is able to characterize V in situations far from classical, and sometimes even when V takes on ∞, but uniqueness results are still lacking in many situations of interest for us here, due to the failure of V to satisfy the continuity, boundedness or growth conditions that current results demand.

In this paper, instead of working with a generalized Hamilton-Jacobi PDE for V, we develop basic "envelope representations," which characterize V as the pointwise inf or sup of a family of more elementary functions. In cases where a description of V as a unique Hamilton-Jacobi solution of some sort can indeed be furnished, now or in the future, these formulas become PDE solution formulas. For state-independent Hamiltonians, $H(x,y) \equiv H_0(y)$, they reduce to Hopf-Lax formulas. We aim at contributing to Hamilton-Jacobi theory by opening a way for such classical formulas to be extended to state-dependent Hamiltonians, while exploring representations of value functions in their own right, especially as a potential means of determining value functions and their subgradients through optimization without having to deal with a separate Hamilton-Jacobi equation for each choice of the cost function g.

We look at two kinds of envelope formulas: upper and lower. Both kinds have long been known in the Hopf-Lax setting but haven't systematically been sought outside of that. Upper envelope formulas, involving pointwise minimization, are elementary and easy to obtain very generally. However, in order for them to express V on $(0,\infty)$ as the envelope of a family of *finite* functions, not to speak of smooth or subsmooth functions, significant restrictions are necessary. Lower envelope formulas, involving pointwise maximization, arise by dualization and therefore thrive only in the presence of convexity, as with the original Hopf formula itself. In compensation for assumptions of convexity, though, they offer a number of unusual and attractive features.

Our focus will primarily be on lower envelope formulas, because of their special potential, but we will also investigate properties enjoyed by upper envelope formulas under the convexity assumptions we impose.

Convex analysis [3] will heavily be used, but mostly through the results obtained in our preceding paper [1]. To the extent that broader variational analysis is required, we rely on the book [4].

After introducing the duality scheme and deriving the basic envelope formulas in §2 in terms of the 'fundamental kernel' and the 'dualizing kernel', we concentrate in §3 on the dualizing kernel and its characterization by a double Hamilton-Jacobi equation. The lower envelope formula for V in terms of the dualizing kernel and the properties of that kernel developed in §3 and also in §4, where connections with subgradients of V are brought out, constitute the paper's main results. To complete the picture, relationships with standard Hopf-Lax formulas are discussed in §5,

2. Envelopes and Convexity

Upper envelope formulas rely on the "double" value function $E : [0,\infty) \times I\!R^n \times I\!R^n \to \overline{I\!R}$ that corresponds to two-endpoint, i.e., Lagrangian, minimization problems for L:

$$
\begin{aligned}
&E(\tau,\xi',\xi) := \inf\Big\{ \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(0)=\xi',\ x(\tau)=\xi \Big\} \quad \text{for } \tau>0,\\
&E(0,\xi',\xi) := \begin{cases} 0 & \text{if } \xi=\xi', \\ \infty & \text{otherwise,} \end{cases}
\end{aligned}
\tag{2.1}
$$

where the minimization is over all the arcs $x(\cdot)$ that go from ξ' at time 0 to ξ at time τ.

Theorem 2.1 (upper envelope representation)**.** *The value function V is expressed in terms of E by the formula*

$$V(\tau,\xi) = \inf_{\xi'}\Big\{ g(\xi') + E(\tau,\xi',\xi) \Big\} \quad \textit{for } \tau \ge 0. \tag{2.2}$$

Moreover when $\tau > 0$, an arc $x(\cdot)$ achieves the minimum in the problem defining $V(\tau,\xi)$ in (1.1) if and only if it achieves the minimum in the problem defining $E(\tau,\xi',\xi)$ in (2.1) for some choice of ξ' yielding the minimum in (2.2).

Proof. Elementary and evident. □

We will call E the *fundamental kernel* associated with L. The "kernel" term comes from the far-reaching analogy between minimizing a sum of functions and integrating a product of functions. Formula (2.2) gives a transform whereby g is converted to $V(\tau,\cdot)$ for $\tau > 0$. It is an "upper envelope" formula because it expresses V as the pointwise infimum of a certain family of functions on $[0,\infty) \times I\!R^n$, namely the functions $e_{\xi'} : (\tau,\xi) \mapsto g(\xi') + E(\tau,\xi',\xi)$ indexed by $\xi' \in D$, where D is the effective domain of g as in (1.2). In some situations E may be finite or even smooth on $(0,\infty) \times I\!R^n \times I\!R^n$, and the same then holds for these functions $e_{\xi'}$.

Often E takes on ∞, though, and the upper envelope representation may be difficult to exploit directly. Clearly, $E(\tau,\xi',\xi)$ can't be finite unless there is an arc $x(\cdot)$ that conforms to the differential inclusion in (1.2) and carries ξ' to ξ. Thus, extended-real-valuedness of E is inevitable in applications where the implicit constraints in (1.2) can seriously come into play.

This motivates a search for alternative envelope representations in which troublesome infinite values can be bypassed. Such representations will be generated by way of the function $K : [0,\infty) \times I\!R^n \times I\!R^n \to \overline{I\!R}$ with

$$\begin{aligned} K(\tau,\xi,\eta) &:= \inf\Big\{ \langle x(0),\eta\rangle + \int_0^\tau L\big(x(t),\dot{x}(t)\big)dt \,\Big|\, x(\tau)=\xi \Big\}, \\ K(0,\xi,\eta) &:= \langle \xi,\eta\rangle, \end{aligned} \tag{2.3}$$

which we introduce now as the *dualizing kernel* associated with L. The minimization takes place over all arcs $x(\cdot) : [0,\infty) \to I\!R^n$ that reach ξ at time τ.

For fixed η, $K(\cdot,\cdot,\eta)$ is the value function obtained as in (1.1) but with the linear function $\langle\cdot,\eta\rangle$ in place of g. As a consequence of Theorem 1.1, therefore, we have

$$K(\tau,\xi,\eta) = \inf_{\xi'}\Big\{ \langle \xi',\eta\rangle + E(\tau,\xi',\xi) \Big\}, \tag{2.4}$$

and indeed, this could serve as well as (2.3) in defining K.

Observe that (2.4) dualizes E by employing a variant of the Legendre-Fenchel transform: $-K(\tau,\xi,\eta)$ is calculated by taking the function conjugate to $E(\tau,\cdot,\xi)$ under that transform and evaluating it at $-\eta$. When $E(\tau,\cdot,\xi)$ is lsc, proper and convex, it can be recovered by the reciprocal formula

$$E(\tau,\xi',\xi) = \sup_{\eta}\Big\{ K(\tau,\xi,\eta) - \langle \xi',\eta\rangle \Big\} \tag{2.5}$$

Our strategy is to use such duality between E and K, along with convexity of g, to translate the upper envelope representation in Theorem 2.1 into a lower one involving K and the function g^* conjugate to g. A prerequisite for this, however, is placing assumptions on L that will ensure E has the properties needed for (2.5) to be valid.

Such assumptions have been identified in our paper [1]. In stating them, we call a function $f : I\!R^n \to \overline{I\!R}$ *coercive* if f is bounded from below and has $f(v)/|v| \to \infty$ as $|v| \to \infty$ (where $|\cdot|$ is the Euclidean norm). When applied to a proper, nondecreasing function θ on $[0,\infty)$, coercivity means having $\theta(s)/s \to \infty$ as $s \to \infty$.

Basic Assumptions (A).

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

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

(A2) *The mapping F underlying L in (1.2) is nonempty-valued everywhere, 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 meaning of these assumptions has thoroughly been elucidated in [1], so for present purposes we only need to record some key facts and examples.

An immediate consequence of $L(x,v)$ being, by (A1) and (A2), a convex, proper, lsc function of v for each x is that L can be recovered from H by

$$L(x,v) = \sup_y \Big\{ \langle v, y\rangle - H(x,y) \Big\}. \tag{2.6}$$

The correspondence between Lagrangians and Hamiltonians is thus one-to-one under our conditions. For each H of a certain class, the associated L is uniquely determined by (2.6). The Hamiltonian class is described as follows.

Proposition 2.2 [1] (Hamiltonian conditions)**.** *The Hamiltonians for the Lagrangians L satisfying (A1), (A2) and (A3) are the functions $H : I\!R^n \times I\!R^n \to \overline{I\!R}$ such that*

(a) *$H(x,y)$ is finite, concave in x, and convex in y (hence locally Lipschitz continuous).*

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

$$H(x,y) \;\le\; \varphi(y) + (\alpha|y| + \beta)|x| \quad \text{for all } x,\ y.$$

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

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

Proof. This comes from Theorem 2.3 of [1]. Finite concave-convex functions are locally Lipschitz continuous by [3; 35.1]. □

Example 2.3 (subseparable Lagrangians)**.** *Let the Lagrangian have the form*

$$L(x,v) = G(x) + L_0(v - Ax) \tag{2.7}$$

for $A \in I\!R^{n\times n}$, a finite convex function G on $I\!R^n$, and a proper convex function L_0 on $I\!R^n$ that is lsc and coercive. Then L satisfies (A1), (A2) and (A3), and its Hamiltonian is

$$H(x,y) = \langle Ax, y\rangle - G(x) + H_0(y), \tag{2.8}$$

where H_0 is a finite convex function on $I\!R^n$, namely $H_0 = L_0^*$. Conversely, if H has form (2.8) for finite convex functions G and H_0, then L has the form (2.7) with $L_0 = H_0^*$ and falls in the category described.

Detail. This is evident from Proposition 2.2 and the conjugacy between finite convex functions (always continuous) and proper convex functions that are lsc and coercive. □

Subseparable Lagrangians illustrate also, in a relatively simple case, the way that our framework of Lagrangians and Hamiltonians connects with control theory. An optimal control problem with linear dynamics $\dot{x}(t) = Ax(t) + Bu(t)$ and running cost integral

$$\int_0^T \big\{G(x(t)) + F(u(t))\big\}dt,$$

with F convex, proper, lsc and coercive (but possibly taking on ∞) corresponds the Lagrangian L in (2.7) for

$$L_0(z) = \min\big\{F(u)\,\big|\,Bu = z\big\},$$

and then the Hamiltonian H in (2.8) has $H_0(y) = F^*(B^*y)$, where B^* is the transpose of B and F^* is the convex function conjugate to F, this function being finite because of the coercivity of F. Control constraints are incorporated here through the specification of the set where F is finite. Control formats much more general than this, yet still fully convex and (as may be shown) still fitting with our assumptions, can be found in [5], [6].

Note that if the coercivity condition in (A3) were replaced by a simpler condition like $L(x,v) \geq \theta(|v|)$, Lagrangians of the type in Example 2.3 would have to have $A = 0$, and G would have to be bounded from below.

Of course, there are many more Lagrangians satisfying (A1), (A2) and (A3) than the ones in Example 2.3. An illustration is $L(x,v) = \frac{1}{p}\max\big\{|x|^p, |v|^p\big\}$ with $p \in (1,\infty)$, which for q determined by $(1/p) + (1/q) = 1$ has

$$H(x,y) = \begin{cases} \frac{1}{q}|y|^q & \text{when } |y| \geq |x|^{p-1}, \\ |x||y| - \frac{1}{p}|x|^p & \text{when } |y| \leq |x|^{p-1}. \end{cases} \tag{2.9}$$

Proposition 2.4 [1] (convexity of the fundamental kernel)**.** *Under (A1), (A2) and (A3), $E(\tau,\xi',\xi)$ is a convex, proper, lsc function of (ξ',ξ) for each $\tau \geq 0$. In fact, $E(\tau,\cdot,\xi)$ is proper and coercive for every ξ, and $E(\tau,\xi',\cdot)$ is proper and coercive for every ξ'.*

Proof. This is extracted from Proposition 4.2 and Corollary 4.4 of [1]. □

On the basis of this result we do have the reciprocal formula in (2.5) along with the one in (2.4), and E and K are entirely dual to each other. We are able then to convert the envelope formula in (1.1) into one for functions that are likely to be better behaved. The technique is to apply Fenchel's duality theorem to the minimization problem in (1.1) in order to recast it as a maximization problem.

In the next theorem, and henceforth in this paper, we take assumptions (A) for granted, unless otherwise mentioned.

Theorem 2.5 (lower envelope representation). *The dualizing kernel $K(\tau,\xi,\eta)$ is everywhere finite, convex in ξ and concave in η. The value function V is expressed in terms of K by the formula*

$$V(\tau,\xi) = \sup_{\eta}\Big\{K(\tau,\xi,\eta) - g^*(\eta)\Big\}. \tag{2.10}$$

Proof. For any $(\tau,\xi)\in[0,\infty)\times I\!R^n$, the function $f = E(\tau,\cdot,\xi)$ is lsc, proper, convex and coercive by Proposition 2.4, so its conjugate f^* is finite. We have $-f^*(-\eta) = K(\tau,\xi,\eta)$ by (2.4), hence $K(\tau,\xi,\eta)$ is finite and concave in η. On the other hand, the convexity of $E(\tau,\xi',\xi)$ in (ξ',ξ) in Proposition 2.4 implies the convexity of $K(\tau,\cdot,\eta)$ by the general principle that when the Legendre-Fenchel transform is applied to one argument of a convex function of two arguments, the result is concave in the residual argument; see [3; 33.3] or [4; 11.48]. (The concavity becomes convexity under the changes of sign.)

To obtain the lower envelope representation, we fix ξ along with τ and view the upper envelope representation in (2.2) as expressing $V(\tau,\xi)$ as the optimal value in the problem of minimizing $g(\xi') + f(\xi')$ for $f = E(\tau,\cdot,\xi)$ as above. By Fenchel's duality theorem (cf. [3; 31.1] or [4; 11.41]), one has

$$\inf_{\xi'}\Big\{g(\xi') + f(\xi')\Big\} = \sup_{\eta}\Big\{-f^*(-\eta) - g^*(\eta)\Big\} \tag{2.11}$$

if the relative interiors of the convex sets $\big\{\eta\,\big|\, -f^*(-\eta) > -\infty\big\}$ and $\big\{\eta\,\big|\, g^*(\eta) < +\infty\big\}$ have a point in common. That criterion is met through the finiteness of f^*, which makes the first set be all of $I\!R^n$. Since the inf in (2.11) gives the left side of (2.10) and the sup in (2.11) gives the right side, the equation in (2.10) is confirmed. □

For $\tau = 0$, the lower envelope representation in (2.10) reduces to the Legendre-Fenchel envelope formula

$$g(\xi) = \sup_{\eta}\Big\{\langle\xi,\eta\rangle - g^*(\eta)\Big\}, \tag{2.12}$$

which expresses the proper, lsc, convex function g as the pointwise supremum of all the affine functions majorized by g. For $\tau > 0$, it can be viewed as extending this formula forward in time through a Hamilton-Jacobi propagation of those affine functions into a different family of functions.

In employing the Fenchel duality rule (2.11) as the tool for passing between upper and lower envelope representations, we are in effect invoking a "minimax principle" in a manner reminiscent in Hamilton-Jacobi theory of the duality seen in classical Hopf-Lax formulas (which will be taken up in §5). Indeed, our formulas can be recast as follows.

Theorem 2.6 (envelope formulas in minimax mode). *In terms of the dualizing kernel K, the value function V always has the representation*

$$V(\tau, \xi) = \inf_{\xi'} \sup_{\eta} \Big\{ g(\xi') - \langle \xi', \eta \rangle + K(\tau, \xi, \eta) \Big\}, \tag{2.13}$$

and this even holds for an arbitrary choice of $g : I\!R^n \to \overline{I\!R}$. When g is convex, proper and lsc, however, V also has the representation

$$V(\tau, \xi) = \sup_{\eta} \inf_{\xi'} \Big\{ g(\xi') - \langle \xi', \eta \rangle + K(\tau, \xi, \eta) \Big\}. \tag{2.14}$$

Proof. We get (2.13) by combining the elementary formula (2.2) for V in terms of E with the reciprocal formula (2.5) for E in terms of K, which is valid by Proposition 2.4 under our assumptions. We get (2.14) by combining the representation (2.10) of V in terms of K and g^* with the definition of g^* in terms of g. □

All of duality theory in convex optimization, a very highly developed subject, has the character of a "minimax principle" of course, but there is no single minimax theorem to invoke that would fit all cases. Everything revolves around the precise conditions under which "inf" and "sup" can legitimately be interchanged when the simplest compactness and continuity properties may be absent, as here. Duality of a much deeper kind than in the proof of Theorem 2.5 will be crucial later, for instance, in ascertaining the circumstances in which the supremum in (2.10) is attained and how this can be used in determining the subgradients of V from those of K (cf. Theorem 4.2 and Corollary 4.3 below). Observe that this brings out an important advantage of expressing lower envelope representations as in (2.10) instead of as in (2.14).

The appearance of g^* instead of g in (2.10) shouldn't be regarded as much of a drawback. In many situations g^* can explicitly be determined from g (see [3] and [4; Chap. 11] for the calculus of conjugates), but even if not, there is much that might be made of this formula. Depending on the particular structure of g (in terms of operations like addition and composition), it's common for $g^*(\eta)$ to be expressible as the optimal value in a minimization problem with respect to some other vector, let's call it ζ, in which η is a parameter. When such an expression is substituted into (2.10), one gets a representation of $V(\tau, \xi)$ as the optimal value in a maximization problem involving both η and ζ.

Anyway, as a practical matter, optimization formulas for $V(\tau,\xi)$, whether directly as in (2.10) or with some expansion of the $g^*(\eta)$ term, are generally more favorable for computation than integration formulas, which become intractable numerically in more than a few dimensions. Furthermore, for applications such as to feedback in optimal control the subgradients of V are at least as important as its values. The lower representation in (2.10) affords a much better grip on those than does the upper representation in (2.2), because K is typically far better behaved than E, as will emerge from the results that follow. These better properties suggest that K may be easier to generate than V in a Hamilton-Jacobi context, after which the lower envelope representation in Theorem 2.5 might be used to compute aspects of V as needed, for instance in feedback. Moreover, the same K would be able to serve for every V that relies on the Lagrangian L, no matter what the choice of the initial cost function g.

3. Characterization of the Dualizing Kernel

Turning now to the development of properties of K that underpin the lower envelope representation in Theorem 2.5, we begin with a special kind of Hamilton-Jacobi characterization. Only subgradients in the sense of convex analysis are needed in this characterization, but other subgradients will soon enter the discussion too, so we go straight to a review of the full definitions. For background, see [4].

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

$$f(x') \;\geq\; f(x) + \langle y, x' - x\rangle + o(|x' - x|). \tag{3.1}$$

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 $y^\nu \to y$. (We consistently use superscript ν for sequences; $\nu \to \infty$.) When f is convex, the sets $\hat{\partial} f(x)$ and $\partial f(x)$ are the same and agree with the subgradient set of convex analysis, defined by (3.1) without the "o" term.

These of course are 'lower' subgradients; the corresponding regular and general 'upper' subgradient sets, defined with the opposite inequality in (3.1) and will be denoted here by $\hat{\tilde{\partial}} f(x)$ and $\tilde{\partial} f(x)$; thus

$$\tilde{\partial} f = -\partial[-f]. \tag{3.2}$$

This notation is expedient because most situations can be couched in terms of lower subgradients alone, cf. [4], although just now we'll have something of an exception.

Regular subgradients have been the mainstay in viscosity theory, but general subgradients are the vehicle for many of the strongest results in variational analysis [4].

In the following theorem, $\partial_\xi K(\tau,\xi,\eta)$ refers to subgradients of the convex function $K(\tau,\cdot,\eta)$, whereas $\tilde{\partial}_\eta K(\tau,\xi,\eta)$ refers to subgradients of the concave function $K(\tau,\xi,\cdot)$.

Theorem 3.1 (double Hamilton-Jacobi equation)**.** *The kernel $K(\tau,\xi,\eta)$ is continuously differentiable with respect to τ and satisfies, for $\tau \geq 0$,*

$$\begin{aligned} -\frac{\partial K}{\partial \tau}(\tau,\xi,\eta) &= \begin{cases} H(\xi,\eta') & \text{for all} \quad \eta' \in \partial_\xi K(\tau,\xi,\eta), \\ H(\xi',\eta) & \text{for all} \quad \xi' \in \tilde{\partial}_\eta K(\tau,\xi,\eta), \end{cases} \\ K(0,\xi,\eta) &= \langle \xi,\eta \rangle, \end{aligned} \tag{3.3}$$

where $\partial K/\partial \tau$ is interpreted as the right partial derivative when $\tau = 0$. Moreover, K is the only such function with $K(\tau,\xi,\eta)$ convex in ξ and concave in η.

The proof of Theorem 3.1 will be furnished later in this section, after some additional developments. The continuous differentiability refers to $(\partial K/\partial\tau)(\tau,\xi,\eta)$ depending continuously on $(\tau,\xi,\eta) \in [0,\infty) \times I\!R^n \times I\!R^n$.

The double Hamilton-Jacobi equation in (3.3) has been placed in the elementary picture of subgradients of convex and concave functions and partial derivatives in time, because that seems most conducive to possible uses of the result. What comparison can be made with viscosity versions of Hamilton-Jacobi equations, though? And why two equations instead of one?

The double aspect of the characterization comes from the fact that, through duality, K has an alternative expression in which the roles of ξ and η are interchanged.

Proposition 3.2 (alternative formula for the dualizing kernel)**.** *In minimizing over arcs $y(\cdot) : [0,\tau] \to I\!R^n$, one has*

$$-K(\tau,-\xi,\eta) = \inf\Big\{ \langle \xi, y(0)\rangle + \int_0^\tau L_*\big(y(t),\dot{y}(t)\big)dt \,\Big|\, y(\tau) = \eta \Big\}, \tag{3.4}$$

where $L_*(y,w) = L^*(-w,y)$. Moreover L_*, like L, satisfies (A1), (A2) and (A3), and its Hamiltonian H_* is given by

$$H_*(y,x) = \sup_w \Big\{ \langle w,x\rangle - L_*(y,w) \Big\} = -H(-x,y). \tag{3.5}$$

Proof. The duality theory for convex problems of Bolza [7] will be applied in the form distilled in [1, §4]. The minimization problem that defines $K(\tau,\xi,\eta)$ in (2.3) is

$(\mathcal{P})$ minimize $\displaystyle\int_0^\tau L(x(t),\dot{x}(t))dt + l(x(0),x(\tau))$ over arcs $x(\cdot) : [0,\tau] \to I\!R^n$, where $l(a,b) = \langle a,\eta\rangle$ if $b=\xi$ but $l(a,b)=\infty$ if $b\neq\xi$. The duality theory pairs this with

$$(\tilde{\mathcal{P}})\qquad \text{minimize } \int_0^\tau \tilde{L}(y(t),\dot{y}(t))dt + \tilde{l}(y(0),y(\tau)) \text{ over arcs } y(\cdot):[0,\tau]\to I\!R^n,$$

where $\tilde{L}(y,w) = L^*(w,y)$ and $\tilde{l}(c,d) = l^*(c,-d)$; the latter comes here out as $\tilde{l}(c,d) = -\langle\xi,d\rangle$ if $c=\eta$ but $\tilde{l}(c,d)=\infty$ if $c\neq\eta$. Because $l(\cdot,b)$ is finite on $I\!R^n$ for a certain b, and $\tilde{l}(c,\cdot)$ is finite on $I\!R^n$ for a certain c, the optimal values in the two problems are related by $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$; this holds by [1, Corollary 4.6]. Thus,

$$-K(\tau,\xi,\eta) = \inf\Big\{\int_0^\tau \tilde{L}\big(y(t),\dot{y}(t)\big)dt - \langle\xi,y(\tau)\rangle \,\Big|\, y(0)=\eta\Big\}.$$

By rewriting in terms of $z(t) = y(\tau-t)$, we can convert this to

$$-K(\tau,\xi,\eta) = \inf\Big\{\int_0^\tau \tilde{L}\big(z(t),-\dot{z}(t)\big)dt - \langle\xi,z(0)\rangle \,\Big|\, z(\tau)=\eta\Big\}.$$

It remains only to replace ξ by $-\xi$ and the z notation by y again to obtain (3.4).

The fact that L_* again satisfies (A1), (A2) and (A3) comes from the fact that $\tilde{L}$ inherits these properties from L, as demonstrated in [1; Proposition 3.5]. The expression for the Hamiltonian H_* in terms of H arises similarly from that result, which asserts that the Hamiltonian $\tilde{H}$ for $\tilde{L}$ has $\tilde{H}(y,x) = -H(x,y)$. □

Through results in [1], the value function formulas for K in (2.3) and (3.4) lead to major conclusions about the subgradients of K and in particular to a viscosity version of the double Hamilton-Jacobi equation in Theorem 3.1. This time we use $\partial_{\tau,\xi}K(\tau,\xi,\eta)$ to denote subgradients of the function $K(\cdot,\cdot,\xi)$ on $[0,\infty)\times I\!R^n$, and so forth.

Theorem 3.3 (subgradients of the dualizing kernel)**.** *For $\tau>0$, one has*

$$\begin{aligned}(\sigma,\eta') \in \partial_{\tau,\xi}K(\tau,\xi,\eta) \quad &\Longleftrightarrow \quad (\sigma,\eta') \in \hat{\partial}_{\tau,\xi}K(\tau,\xi,\eta)\\ &\Longleftrightarrow \quad \eta' \in \partial_\xi K(\tau,\xi,\eta),\ \ \sigma = -H(\xi,\eta'),\end{aligned}\tag{3.6}$$

and on the other hand

$$\begin{aligned}(\sigma,\xi') \in \tilde{\partial}_{\tau,\eta}K(\tau,\xi,\eta) \quad &\Longleftrightarrow \quad (\sigma,\xi') \in \tilde{\hat{\partial}}_{\tau,\eta}K(\tau,\xi,\eta)\\ &\Longleftrightarrow \quad \xi' \in \tilde{\partial}_\eta K(\tau,\xi,\eta),\ \ \sigma = -H(\xi',\eta).\end{aligned}\tag{3.7}$$

Proof. We simply apply [1; Theorem 2.5] first to $K(\cdot,\cdot,\eta)$, which is the value function that propagates $\langle\cdot,\eta\rangle$ under L as in (2.3), and second to $-K(\cdot,-\xi,\cdot)$, which by Proposition 3.2 is the value function that propagates $\langle\xi,\cdot\rangle$ under L_*. □

Corollary 3.4 (double viscosity equation)**.** *For* $\tau > 0$, *one has*

$$\begin{cases} \sigma + H(\xi, \eta') = 0 & \textit{for all} \quad (\sigma, \eta') \in \hat{\partial}_{\tau,\xi} K(\tau, \xi, \eta), \\ \sigma + H(\xi', \eta) = 0 & \textit{for all} \quad (\sigma, \xi') \in \hat{\tilde{\partial}}_{\tau,\eta} K(\tau, \xi, \eta). \end{cases} \tag{3.8}$$

It will be established in the next theorem that K is locally Lipschitz continuous. In view of this, the first of the subgradient equations in (3.8) is equivalent, as shown by Frankowska [8], to $K(\cdot, \cdot, \eta)$ being a Hamilton-Jacobi viscosity solution in the sense of satisfying the upper and lower inequalities of Crandall, Evans and Lions [9], with initial $K(0, \cdot, \eta) = \langle \cdot, \eta \rangle$. The second equation has a similar viscosity interpretation relative to a switch in the roles of the ξ and η arguments.

It might be hoped that either of these subgradient equations, by itself, would be enough to determine K uniquely. That could be true, but unfortunately the existing results on uniqueness of viscosity solutions are not fully up to the task. The trouble is that H and K need not satisfy the kinds of growth or boundedness conditions assumed in such results. Because of the initial condition $K(\tau, \xi, \eta)$ is certainly neither globally bounded from above nor globally bounded from below, even for fixed ξ or η. One or the other kind of boundedness would be needed to apply the latest uniqueness theorem of Ishii [10], for example. Anyway, the Hamiltonian can grow at rates like those in (2.9), and this can be problematical as well.

Theorem 3.5 (Lipschitz continuity of the dualizing kernel)**.** *The function* K *is locally Lipschitz continuous on* $[0, \infty) \times I\!R^n \times I\!R^n$.

Proof. By [1; Theorem 2.1] the functions $K(\cdot, \cdot, \eta)$ are lsc on $[0, \infty) \times I\!R^n$ as value functions in the mode of (2.3). Similarly by this result, as applied in the context of Proposition 3.2, the functions $-K(\cdot, -\xi, \cdot)$ are lsc on $[0, \infty) \times I\!R^n$. Hence the functions $K(\cdot, \xi, \cdot)$ are usc on $[0, \infty) \times I\!R^n$, and it follows in particular that $K(\tau, \xi, \eta)$ is continuous in $\tau \in [0, \infty)$ for each (ξ, η). Thus, $K(\tau, \cdot, \cdot)$ converges pointwise to $K(\bar{\tau}, \cdot, \cdot)$ whenever $\tau \to \bar{\tau}$ in $[0, \infty)$. The functions $K(\tau, \cdot, \cdot)$ are finite and convex-concave by Theorem 2.5, and pointwise convergence of such functions on $I\!R^n \times I\!R^n$ implies uniform convergence on bounded sets (see [3; 35.4]). In consequence, K is continuous on $[0, \infty) \times I\!R^n \times I\!R^n$. Furthermore, the convergence implies that the mapping

$$S : (\tau, \xi, \eta) \mapsto \big\{ (\eta', \xi') \,\big|\, \eta' \in \partial_\xi K(\tau, \xi, \eta),\ \xi' \in \tilde{\partial}_\eta K(\tau, \xi, \eta) \,\big\} \tag{3.9}$$

is locally bounded on $[0, \infty) \times I\!R^n \times I\!R^n$ and has closed graph (see [3; 35.7]).

This yields through the continuity of H (in Proposition 2.2) the closed graph property and local boundedness on $[0,\infty)\times I\!R^n$ of the mappings

$$\begin{aligned}(\tau,\xi)&\mapsto\big\{(\sigma,\eta')\,\big|\,\eta'\in\partial_\xi K(\tau,\xi,\eta),\;\;\sigma=-H(\xi,\eta')\big\}\\(\tau,\eta)&\mapsto\big\{(\sigma,\eta')\,\big|\,\eta'\in\tilde\partial_\eta K(\tau,\xi,\eta),\;\;\sigma=-H(\xi',\eta)\big\}.\end{aligned}\tag{3.10}$$

In general, a function f that is finite and lsc on an open set O in a space $I\!R^d$ is Lipschitz continuous with constant κ on any set $X\subset O$ such that $|y|\le\kappa$ for all $y\in\partial f(x)$ when $x\in X$; this holds by [4; 9.2, 9.13]. We invoke this now for $K(\cdot,\cdot,\eta)$ on $(0,\infty)\times I\!R^n$. From the subgradient characterization in (3.6) of Theorem 3.3 and the local boundedness of the first mapping in (3.10), on $[0,\infty)\times I\!R^n$ rather than just $(0,\infty)\times I\!R^n$, we get that $K(\cdot,\cdot,\eta)$ is locally Lipschitz continuous on $(0,\infty)\times I\!R^n$, and more over that the Lipschitz constants don't blow up as $\tau\searrow 0$. Since $K(\cdot,\cdot,\eta)$ is anyway continuous on $[0,\infty)\times I\!R^n$, we conclude it must actually be Lipschitz continuous on $[0,\infty)\times I\!R^n$.

A parallel argument utilizing the dual formula in Proposition 3.2 shows that the functions $K(\cdot,\xi,\cdot)$ are Lipschitz continuous on $[0,\infty)\times I\!R^n$. The two properties of Lipschitz continuity combine to give the Lipschitz continuity of K itself on $[0,\infty)\times I\!R^n\times I\!R^n$. □

The subgradient result in Theorem 3.3 will be complemented now by one about subderivatives. These are defined as follows; see [4] for background. For $f: I\!R^n\to\overline{I\!R}$ and a point x where $f(x)$ is finite, the *subderivative* of f at x for a vector w is

$$df(x)(w):=\liminf_{\substack{\varepsilon\searrow 0\\ w'\to w}}\frac{f(x+\varepsilon w')-f(x)}{\varepsilon}.\tag{3.11}$$

If this 'liminf' coincides with the associated 'limsup' and thus exists as a full limit, f is said to be *semidifferentiable at x for w*, or simply *semidifferentiable at x* if true for all $w\in I\!R^n$. Semidifferentiability at x corresponds to the difference quotient functions $\Delta_\varepsilon f(x): w\mapsto[f(x+\varepsilon w)-f(x)]/\varepsilon$ converging uniformly on bounded subsets of $I\!R^n$, as $\varepsilon\searrow 0$, to a continuous function of w [4; 7.21]. Differentiability is the case where, in addition, the limit function $df(x)$ is linear.

Theorem 3.6 (subderivatives of the dualizing kernel). *On $(0,\infty)\times I\!R^n\times I\!R^n$, K is semidifferentiable everywhere, and its subderivative formula is as follows. For any (τ,ξ,η), the quantities $H(\xi,\eta')$ for $\eta'\in\partial_\xi K(\tau,\xi,\eta)$ and $H(\xi',\eta)$ for $\xi'\in\tilde\partial_\eta K(\tau,\xi,\eta)$ all have the same value, and in denoting it by $k(\tau,\xi,\eta)$ one has*

$$\begin{aligned}dK(\tau,\xi,\eta)(\theta,\omega,\zeta)\;=\;&-\theta k(\tau,\xi,\eta)\\&+\max\big\{\langle\eta',\omega\rangle\,\big|\,\eta'\in\partial_\xi K(\tau,\xi,\eta)\big\}\\&+\min\big\{\langle\xi',\zeta\rangle\,\big|\,\xi'\in\partial_\eta K(\tau,\xi,\eta)\big\}.\end{aligned}\tag{3.12}$$

Proof. The first part of the proof, devoted to the existence of the common value $k(\tau, \xi, \eta)$, will be the basis later for knowing that K is continuously differentiable in τ as claimed in Theorem 3.1 but not yet justified. Let $K_\eta = K(\cdot, \cdot, \eta)$. Since K_η is the value function that propagates a finite convex function under L, namely $\langle \cdot, \eta \rangle$, it is semidifferentiable on $(0, \infty) \times \mathbb{R}^n$ by [1; Theorem 7.3] with formula

$$dK_\eta(\tau, \xi)(\theta, \omega) = \max\big\{ \langle \omega, \eta' \rangle - \theta H(\xi, \eta') \,\big|\, \eta' \in \partial_\xi K_\eta(\tau, \xi) \big\}. \tag{3.13}$$

Likewise for $K^\xi = -K(\cdot, -\xi, \cdot)$ in the context of Proposition 3.2, K^ξ is the value function that propagates $\langle \xi, \cdot \rangle$ under L_* and thus is semidifferentiable on $(0, \infty) \times \mathbb{R}^n$ with formula

$$dK^\xi(\tau, \eta)(\theta, \zeta) = \max\big\{ \langle \xi', \zeta \rangle - \theta H_*(\eta, \xi') \,\big|\, \xi' \in \partial_\eta K^\xi(\tau, \eta) \big\},$$

where $H_*(\eta, \xi') = -H(-\xi', \eta)$. In terms of $\tilde{K}_\xi = K(\cdot, \xi, \cdot)$ the latter can be rewritten as

$$d\tilde{K}_\xi(\tau, \eta)(\theta, \zeta) = \min\big\{ \langle \xi', \zeta \rangle - \theta H(\xi', \eta) \,\big|\, \xi' \in \tilde{\partial}_\eta \tilde{K}_\xi(\tau, \eta) \big\}, \tag{3.14}$$

In particular, K has right and left partial derivatives in τ,

$$\begin{aligned} (\partial^+ K/\partial\tau)(\tau, \xi, \eta) &= dK_\eta(\tau, \xi)(1, 0) = d\tilde{K}_\xi(\tau, \eta)(1, 0), \\ (\partial^- K/\partial\tau)(\tau, \xi, \eta) &= -dK_\eta(\tau, \xi)(-1, 0) = -d\tilde{K}_\xi(\tau, \eta)(-1, 0), \end{aligned}$$

which by (3.13) must satisfy

$$\begin{aligned} (\partial^- K/\partial\tau)(\tau, \xi, \eta) &= \min\big\{ -H(\xi, \eta') \,\big|\, \eta' \in \partial_\xi K(\tau, \xi, \eta) \big\}, \\ (\partial^+ K/\partial\tau)(\tau, \xi, \eta) &= \max\big\{ -H(\xi, \eta') \,\big|\, \eta' \in \partial_\xi K(\tau, \xi, \eta) \big\}, \end{aligned} \tag{3.15}$$

and on the other hand, by (3.14), must satisfy

$$\begin{aligned} (\partial^- K/\partial\tau)(\tau, \xi, \eta) &= \max\big\{ -H(\xi', \eta) \,\big|\, \xi' \in \tilde{\partial}_\eta K(\tau, \xi, \eta) \big\}, \\ (\partial^+ K/\partial\tau)(\tau, \xi, \eta) &= \min\big\{ -H(\xi', \eta) \,\big|\, \xi' \in \tilde{\partial}_\eta K(\tau, \xi, \eta) \big\}. \end{aligned} \tag{3.16}$$

We get $(\partial^- K/\partial\tau)(\tau, \xi, \eta) \leq (\partial^+ K/\partial\tau)(\tau, \xi, \eta)$ from (3.15) but the opposite inequality from (3.16). The partial derivative $(\partial K/\partial\tau)(\tau, \xi, \eta)$ therefore exists and is given by all four expressions on the right in (3.15) and (3.16). In particular, the quantities $H(\xi, \eta')$ and $H(\xi', \eta)$ involved in these expressions must have the same value.

In denoting this common value by $k(\tau, \xi, \eta)$, we have a function that is continuous not only on $(0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$ but has a continuous extension to $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$. That follows from the closed graph property and local boundedness on $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$ of the mappings in (3.10), as demonstrated in the proof of Theorem 3.5. Hence $(\partial K/\partial\tau)(\tau, \xi, \eta)$

exists even at $\tau = 0$, when interpreted there as the right partial derivative, and it depends continuously on $(\tau, \xi, \eta) \in [0, \infty) \times I\!R^n \times I\!R^n$.

Henceforth in proceeding with the proof of Theorem 3.6, we argue solely on the basis of $K(\tau, \xi, \eta)$ being continuously differentiable in τ while convex in ξ and concave in η. This will help with something needed eventually in the proof of Theorem 3.1, although a price must be paid in overlaps with arguments already furnished for Theorem 3.5.

Each of the functions $K(\tau, \cdot, \cdot)$, being finite and convex-concave, is locally Lipschitz continuous on $I\!R^n \times I\!R^n$ by [3; 35.1]. The differentiability of $K(\tau, \xi, \eta)$ with respect to τ entails continuity in τ. Therefore, whenever $\tau \to \bar{\tau}$ in $[0, \infty)$ the functions $K(\tau, \cdot, \cdot)$ converge pointwise on $I\!R^n \times I\!R^n$ to $K(\bar{\tau}, \cdot, \cdot)$. We have already seen in the proof of Theorem 3.6 how this convergence implies that the mapping S in (3.9) is locally bounded on $[0, \infty) \times I\!R^n \times I\!R^n$ with closed graph. This guarantees that the local Lipschitz continuity of the functions $K(\tau, \cdot, \cdot)$ is uniform locally with respect to τ (by virtue of [3; 24.7] as applied in the convex and concave arguments separately). In taking this together with the continuity of $\partial K / \partial \tau$, which ensures the local Lipschitz continuity of $K(\tau, \xi, \eta)$ in τ, we deduce that K is locally Lipschitz continuous as a function of $(\tau, \xi, \eta) \in [0, \infty) \times I\!R^n \times I\!R^n$.

We work next with the difference quotient functions concerned in generating the subderivatives of K:

$$\begin{aligned} \Delta_\varepsilon K(\tau, \xi, \eta)(\theta, \omega, \zeta) = {} & \frac{K(\tau + \varepsilon\theta,\, \xi + \varepsilon\omega,\, \eta + \varepsilon\zeta) - K(\tau,\, \xi + \varepsilon\omega, \eta + \varepsilon\zeta)}{\varepsilon} \\ & + \frac{K(\tau,\, \xi + \varepsilon\omega,\, \eta + \varepsilon\zeta) - K(\tau, \xi, \eta)}{\varepsilon}. \end{aligned} \tag{3.17}$$

When $\varepsilon \searrow 0$, the first expression in the sum in (3.17), as a function of (θ, ω, ζ), converges uniformly over bounded sets to the function

$$(\theta, \omega, \zeta) \mapsto (\partial K / \partial \tau)(\tau, \xi, \eta)\theta$$

because of the continuity of $\partial K / \partial \tau$ (through a classical argument using the mean value theorem). The second expression in the sum in (3.17), as a function of (ω, ζ) that is convex-concave, is known from convex analysis [3; 35.6] to converge pointwise to the function

$$(\omega, \zeta) \;\mapsto\; \max_{\eta' \in \partial_\xi K(\tau, \xi, \eta)} \langle \eta', \omega \rangle + \min_{\xi' \in \partial_\eta K(\tau, \xi, \eta)} \langle \xi', \zeta \rangle.$$

The convergence must then be uniform over bounded subsets of $I\!R^n \times I\!R^n$ (by [3; 35.4]). Thus, as $\varepsilon \searrow 0$, the functions $\Delta_\varepsilon K(\tau, \xi, \eta)$ do converge uniformly on bounded sets to the function described by the right side of (3.12) with $k = (\partial K / \partial \tau)$. Hence K is semidifferentiable with this as its formula. □

Proof of Theorem 3.1. The continuous differentiability of $K(\tau,\xi,\eta)$ has been demonstrated in the first part of the proof of Theorem 3.6 along with the double formula for $(\partial K/\partial\tau)$ in (3.3), the common value on the right side of (3.3) being the expression $k(\tau,\xi,\eta)$ introduced in the statement of Theorem 3.6. The remaining task is to show the uniqueness in this characterization. Let $J(\tau,\xi,\eta)$ on $[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n$ be convex in ξ, concave in η, and continuously differentiable in τ, satisfying (3.3). We have to prove that $J=K$.

As a tool in this endeavor, we can use the fact that J, like K, has the subderivative properties in Theorem 3.6, since those properties depend only on the facts now being assumed; see the remark in the middle of the proof of Theorem 3.6 (in the paragraph starting with "Henceforth"). Thus

$$\begin{aligned} dJ(\tau,\xi,\eta)(\theta,\omega,\zeta) &= \theta\,(\partial J/\partial\tau)(\tau,\xi,\eta) \\ &\quad + \max\big\{\langle\eta',\omega\rangle \,\big|\, \eta'\in\partial_\xi J(\tau,\xi,\eta)\big\} \\ &\quad + \min\big\{\langle\xi',\zeta\rangle \,\big|\, \xi'\in\partial_\eta J(\tau,\xi,\eta)\big\}. \end{aligned} \tag{3.18}$$

In addition we can take J to be locally Lipschitz continuous, because that property was likewise seen there to be a consequence of the current assumptions.

Fix (τ,ξ,η). Certainly $J(\tau,\xi,\eta)=K(\tau,\xi,\eta)$ when $\tau=0$, so suppose $\tau>0$. The infimum in the definition (2.3) of $K(\cdot,\cdot,\eta)$ as the value function propagating $\langle\xi,\cdot\rangle$ is attained by an arc $x(\cdot)$ on $[0,\tau]$ which moreover is Lipschitz continuous; this holds by [1; Theorem 5.2], which under our assumptions (A) applies to value functions at interior points (τ,ξ) of their domains. Then too, for any $\tau'\in(0,\tau)$ and the point $\xi'=x(\tau')$, the restriction of $x(\cdot)$ to $[0,\tau']$ is optimal for the minimization problem defining $K(\tau',\xi',\eta)$ (by the "principle of optimality"). Thus,

$$K(\tau',x(\tau'),\eta)=\langle x(0),\eta\rangle+\int_0^{\tau'} L\big(x(s),\dot{x}(s)\big)dt \quad \text{for } 0\le\tau'\le\tau \tag{3.19}$$

In terms of the functions $\varphi:[0,\tau]\to\mathbb{R}$ and $\psi:[0,\tau]\to\mathbb{R}$ defined by

$$\varphi(t):=K(t,x(t),\eta),\qquad \psi(t):=J(t,x(t),\eta),$$

we have $\varphi(0)=K(0,x(0),\eta)=J(0,x(0),\eta)=\psi(0)$, whereas $\varphi(\tau)=K(\tau,\xi,\eta)$ and $\psi(\tau)=J(\tau,\xi,\eta)$. Furthermore, φ is Lipschitz continuous on $[0,\tau]$, because $x(\cdot)$ has this property and K is locally Lipschitz continuous on $[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n$. Likewise ψ is Lipschitz continuous on $[0,\tau]$. It follows that φ and ψ are the integrals of their derivatives, which exist almost everywhere. Hence

$$K(\tau,\xi,\eta)-J(\tau,\xi,\eta)=\int_0^\tau \varphi'(t)dt-\int_0^\tau \psi'(t)dt. \tag{3.20}$$

On the basis of (3.19), we have

$$\varphi'(t) = L\big(x(t), \dot{x}(t)\big) \text{ for a.e. } t. \tag{3.21}$$

On the other hand, the semidifferentiability of J in (3.18) yields

$$\psi'(t) \;=\; (\partial J/\partial\tau)(t, x(t), \eta) + \max\big\{ \langle \eta', \dot{x}(t)\rangle \,\big|\, \eta' \in \partial_\xi J(t, x(t), \eta)\big\}.$$

For each t let $y(t)$ be a vector ξ' attaining this maximum. Because J satisfies the Hamilton-Jacobi equations in (3.3), we have $(\partial J/\partial\tau)(t, x(t), \eta) = -H(x(t), y(t))$, so that

$$\psi'(t) = -H(x(t), y(t)) + \langle y(t), \dot{x}(t)\rangle. \tag{3.22}$$

Since $L(x(t), \cdot)$ and $H(x(t), \cdot)$ are conjugate convex functions, we know from the reciprocal Legendre-Fenchel formula in (2.6) that $\langle y(t), \dot{x}(t)\rangle - H(x(t), y(t)) \leq L(x(t), \dot{x}(t))$. Therefore $\psi'(t) \leq \varphi'(t)$ by (3.22) and (3.21). When this inequality is combined with (3.20), we arrive at the conclusion that $J(\tau, \xi, \eta) \leq K(\tau, \xi, \eta)$.

So far, we have established that $J \leq K$. To get the opposite inequality, it suffices to show that $-J(\tau, -\xi, \eta) \leq -K(\tau, -\xi, \eta)$ for all (τ, ξ, η). But for this we only need appeal to the alternative value function formula for K in Proposition 3.2 and in such terms reapply the argument just given. □

4. Additional Kernel Properties and Subgradient Formulas

Other facts about the kernels K and E will now be developed, with emphasis on subgradients and regularity. Connections between subgradients of the value function V and those of the dualizing kernel K are featured because of their possible use in applications to feedback in optimal control.

An important role in bringing out such connections is played by the generalized Hamiltonian dynamical system associated with H, which has the form

$$\dot{x}(t) \in \partial_y H(x(t), y(t)), \qquad -\dot{y}(t) \in \tilde{\partial}_x H(x(t), y(t)) \tag{4.1}$$

This dynamical system is the key to characterizing optimality in the theory of generalized problems of Bolza for the Lagrangian L, where it originated in [11]. More on its properties and history can be found in [1] and its references. A Hamiltonian *trajectory* over $[0, \tau]$ is a pair of arcs $x(\cdot)$ and $y(\cdot)$ satisfying (4.1) for almost every t.

Theorem 4.1 (kernel subgradients and Hamiltonian dynamics). *The following properties are equivalent for any* $\tau \geq 0$*:*

(a) $\eta' \in \partial_\xi K(\tau, \xi, \eta)$ *and* $\xi' \in \tilde{\partial}_\eta K(\tau, \xi, \eta)$*;*

(b) $(-\eta, \eta') \in \partial_{\xi', \xi} E(\tau, \xi', \xi)$*;*

(c) *there is a Hamiltonian trajectory* $(x(\cdot), y(\cdot))$ *over* $[0, \tau]$ *from* (ξ', η) *to* (ξ, η')*.*

Proof. The equivalence between (a) and (b) reflects a general principle about how subgradients behave when partial conjugates are taken, as in the passage between E and K in (2.4) and (2.5); cf. [4; 11.48].

The equivalence between (a) and (c) will come out of a result in [1; Theorem 2.4] about the subgradients of value functions V more generally: one has $\eta' \in \partial_\xi V(\tau, \xi)$ if and only if there is a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ over $[0, \tau]$ that starts with $y(0) \in \partial g(x(0))$ and ends at (ξ, η'). Since $K(\cdot, \cdot, \eta)$ is the value function that propagates $\langle \cdot, \eta \rangle$, a function with constant subgradient (gradient) η, we deduce that $\eta' \in \partial_\xi K(\tau, \xi, \eta)$ if and only if there is a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ over $[0, \tau]$ that starts with $y(0) = \eta$ (any $x(0)$) and ends at (ξ, η').

For the remainder, we argue in terms of the dual expression in Proposition 3.2, where $-K(\cdot, -\xi, \cdot)$ is the value function that propagates $\langle \xi, \cdot \rangle$ under L_*, a Lagrangian with Hamiltonian with Hamiltonian H_* given by (3.5). Invoking the same theorem from [1] in this setting, we obtain, after the $\pm$ signs settle down and the trajectories are reversed in time, the fact that $\xi' \in \tilde{\partial}_\eta K(\tau, \xi, \eta)$ if and only if there is a Hamiltonian trajectory over $[0, \tau]$ that starts at (ξ', η) and ends with $x(\tau) = \xi$ (any $y(\tau)$). In putting this together with the earlier statement, we arrive at the description in (c). ☐

Theorem 4.2 (determination of value function subgradients). *For every* $\tau > 0$*, one has*

$$\begin{aligned} \partial V(\tau,\xi) &= \bigcup \left\{ \partial_{\tau,\xi} K(\tau, \xi, \eta) \,\middle|\, \eta \in M(\tau, \xi) \right\}, \\ &\text{where } \; M(\tau, \xi) := \operatorname{argmax}_\eta \left\{ K(\tau, \xi, \eta) - g^*(\eta) \right\}. \end{aligned} \tag{4.2}$$

Therefore, subgradients of V *can be determined from those of* K *by carrying out the maximization in the lower envelope formula, with*

$$(\sigma, \eta') \in \partial V(\tau, \xi) \iff \exists\, \eta \in M(\tau, \xi) \text{ with } \begin{cases} \eta' \in \partial_\xi K(\tau, \xi, \eta), \\ \sigma = -H(\xi, \eta'). \end{cases} \tag{4.3}$$

Proof. Recall from Theorem 3.3 that the subgradients in $\partial_{\tau,\xi} K(\tau, \xi, \eta)$ are of the form $(-H(\xi, \eta'), \eta')$ for $\eta' \in \partial_\xi K(\tau, \xi, \eta)$. A similar result was obtained in [1; Theorem 2.5] for V; its subgradients have the form $(-H(\xi, \eta'), \eta')$ for $\eta' \in \partial_\xi V(\tau, \xi)$. Further, as already noted in the proof of Theorem 4.1, it was demonstrated in [1; Theorem 2.4] that $\eta' \in \partial_\xi V(\tau,\xi)$ if and only if there is a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ that starts with $y(0) \in \partial g(x(0))$ and ends at (ξ, η').

On the other hand, the condition for η to belong to $M(\tau,\xi)$, i.e., to maximize $K(\tau,\xi,\eta) - g^*(\eta)$, can be expressed in subgradient terms as $0 \in \tilde{\partial}_\eta K(\tau,\xi,\eta) - \partial g^*(\eta)$. (This is both necessary and sufficient for optimality because g^* is a convex function while $K(\tau,\xi,\cdot)$ is a finite concave function; see [3; §31].) Equivalently, there exists some $\xi' \in \tilde{\partial}_\eta K(\tau,\xi,\eta) \cap \partial g^*(\eta)$. But for conjugate convex functions we have $\xi' \in \partial g^*(\eta)$ if and only if $\eta \in \partial g(\xi')$. In view of Theorem 4.1, then, we have $\eta \in M(\tau,\xi)$ if and only if there is a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ that starts with $y(0) \in \partial g(x(0))$ and ends at (ξ,η'). This is the same as the condition derived in terms of V, so we conclude that the subgradient formula in the theorem is correct. □

Theorem 4.2 puts the spotlight on the maximizing set $M(\tau,\xi)$ in the lower envelope formula (1.5) and raises questions about the nature of this subproblem of maximization, in particular whether the maximum is actually attained. We address these questions next.

Theorem 4.3 (compactness and attainment in the lower envelope formula)**.** *For any $\tau > 0$ and ξ, the following properties in the lower envelope formula are equivalent:*

(a) *the set $M(\tau,\xi) = \operatorname{argmax}_\eta \{K(\tau,\xi,\eta) - g^*(\eta)\}$ is nonempty and compact;*

(b) *for every $\beta \in I\!R$, the upper level set $\{\eta \mid K(\tau,\xi,\eta) - g^*(\eta) \ge \beta\}$ is compact;*

(c) *$\xi \in \operatorname{int} D(\tau)$ for the set $D(\tau) = \{\xi \mid V(\tau,\xi) < \infty\}$.*

Proof. We return to the proof of Theorem 2.5 and the framework of Fenchel duality in which it was placed, with $f = E(\tau,\cdot,\xi)$ and $-f^*(-\eta) = K(\tau,\xi,\eta)$. It is well known in that theory, in terms of the convex sets $\operatorname{dom} f$ and $\operatorname{dom} g$ (where f and g are finite), that $\operatorname{argmin}_\eta \{f^*(-\eta) + g^*(\eta)\}$ is nonempty and bounded if and only if $0 \in \operatorname{int}(\operatorname{dom} f - \operatorname{dom} g)$ and the infimum is finite (see for instance [4; 11.41, 11.39(b)].) That is in turn equivalent to having $\operatorname{ri}\operatorname{dom} g \cap \operatorname{ri}\operatorname{dom} f \neq \emptyset$ with $\operatorname{dom} g \cup \operatorname{dom} f$ not lying in a hyperplane (cf. [4; 2.45]). In [1; Proposition 7.4] this property has been identified with (c). Thus, (a) is equivalent to (c).

The equivalence between (a) and (b), on the other hand, results from the fact that the function being maximized is concave and upper semicontinuous; cf. [4; 3.27]. □

Corollary 4.4 (finite value functions)**.** *When V is finite on $(0,\infty) \times I\!R^n$, the maximizing set $M(\tau,\xi)$ is nonempty and compact for every $(\tau,\xi) \in (0,\infty) \times I\!R^n$. In particular this is the case when g is finite on $I\!R^n$ or on the other hand when L is finite on $I\!R^n \times I\!R^n$.*

Proof. The first assertion is justified through condition (c) in Theorem 4.3. The rest cites elementary circumstances in which V is know from [1; Cor. 7.6] to be finite. □

We look further now at the fundamental kernel E, first demonstrating a property of epi-continuity. Epi-continuity, which refers to epigraphs depending continuously on a parameter in the sense of Painléve-Kuratowski set convergence, was established in [1; Theorem 2.1] for the dependence of $V(\tau,\cdot)$ on $\tau \in [0,\infty)$. We'll apply that result to the functions $E(\tau,\cdot,\cdot)$ by way of a reformulation trick.

Proposition 4.5 (fundamental epi-continuity). *The function $E(\tau,\cdot,\cdot) : I\!R^n \times I\!R^n \to \overline{I\!R}$ depends epi-continuously on $\tau \in [0,\infty)$: whenever $\tau^\nu \to \tau$ with $\tau^\nu \geq 0$ one has*

$$\begin{cases} \liminf_\nu E(\tau^\nu,\xi'^\nu,\xi^\nu) \geq E(\tau,\xi',\xi) & \textit{for every sequence } (\xi'^\nu,\xi^\nu) \to (\xi',\xi), \\ \limsup_\nu E(\tau^\nu,\xi'^\nu,\xi^\nu) \leq E(\tau,\xi',\xi) & \textit{for some sequence } (\xi'^\nu,\xi^\nu) \to (\xi',\xi). \end{cases}$$

Proof. Although E seems to fit a different pattern than V, in being generated as a value function in terms of a variable pair (ξ',ξ) of initial and terminal points, instead of an initial function g and a terminal point ξ, we can nonetheless obtain results about E from those for V by an adaptation. The trick is to view E as the value function $V_E : [0,\infty) \times I\!R^{2n} \to \overline{I\!R}$ that is generated from the Lagrangian L_E and initial function g_E defined as follows:

$$\begin{aligned} L_E(x',x,v',v) &:= \begin{cases} L(x,v) & \text{if } v' = 0, \\ \infty & \text{if } v' \neq 0, \end{cases} \\ g_E(x',x) &:= \begin{cases} 0 & \text{if } x' = x, \\ \infty & \text{if } x' \neq x. \end{cases} \end{aligned} \tag{4.4}$$

Indeed, under these definitions $V_E(\tau,\xi',\xi)$ is the infimum of $\int_0^\tau L(x(t),\dot{x}(t))dt$ over all arcs $(x'(\cdot),x(\cdot)) \in \mathcal{A}^1_{2n}[0,\tau]$ such that $x'(0) = x(0)$, $\dot{x}'(t) = 0$ a.e., and $x'(\tau) = \xi'$. The latter conditions obviously force $x(0)$ to be ξ'. Note that g_E and L_E satisfy our blanket assumptions (A) (in higher-dimensional interpretation) because L satisfies (A1)–(A3). By this route, we get justification of our claims through epi-continuity result for value functions in [1; Theorem 2.1]. □

In our next result, we record a basic relationship between certain effective domains, which although convex, could in general have empty interior, namely

$$\begin{aligned} \operatorname{dom} E(\tau,\cdot,\cdot) &= \big\{(\xi',\xi) \,\big|\, E(\tau,\xi',\xi) < \infty\big\}, \\ \operatorname{dom} E &= \big\{(\tau,\xi',\xi) \,\big|\, \tau > 0,\ E(\tau,\xi',\xi) < \infty\big\}. \end{aligned} \tag{4.5}$$

Proposition 4.6 (domain interior). *The following properties are equivalent:*

(a) $\tau > 0$ *and* $(\xi', \xi) \in \operatorname{int} \operatorname{dom} E(\tau, \cdot, \cdot)$*;*

(b) $(\tau, \xi', \xi) \in \operatorname{int} \operatorname{dom} E$*.*

Proof. This is [1; Proposition 7.2] as applied to the value function V_E in the reformulation in the proof of Proposition 4.5. □

In the following theorem, subdifferential regularity is a property that a function has when its epigraph is closed and Clarke regular; see [4].

Theorem 4.7 (regularity of the fundamental kernel)**.** *On* $\operatorname{int} \operatorname{dom} E$*, the subgradient mapping* ∂E *is nonempty-compact-convex-valued and locally bounded, and* E *itself is locally Lipschitz continuous and subdifferentially regular, moreover semidifferentiable with*

$$dE(\tau, \xi', \xi)(\tau', \omega', \omega) \;=\; \max\Big\{ \langle \omega, \eta' \rangle - \langle \omega', \eta \rangle - \tau' H(\xi, \eta') \;\Big|\; (-\eta, \eta') \in \partial_{\xi', \xi} E(\tau, \xi', \xi) \Big\},$$

where $H(\xi, \eta')$ *could be replaced by* $H(\xi', \eta)$*. Indeed,* E *is strictly differentiable wherever it is differentiable, which is at almost every point of* $\operatorname{int} \operatorname{dom} E$*, and with respect to such points the gradient mapping* ∇E *is continuous.*

Proof. We apply [1; Theorem 7.3], a result for value functions V in general under our assumptions, to V_E in the pattern of the proof of Proposition 4.6 above. □

Theorem 4.8 (Hamilton-Jacobi equations for the fundamental kernel)**.** *The subgradients of* E *on* $(0, \tau) \times I\!R^n \times I\!R^n$ *have the property that*

$$\begin{aligned} (\sigma, -\eta, \eta') \in \partial E(\tau, \xi', \xi) \quad &\Longleftrightarrow \quad (\sigma, -\eta, \eta') \in \hat{\partial} E(\tau, \xi', \xi) \\ &\Longleftrightarrow \quad (-\eta, \eta') \in \partial_{\xi', \xi} E(\tau, \xi', \xi), \;\; \sigma = -H(\xi, \eta') \\ &\Longleftrightarrow \quad (-\eta, \eta') \in \partial_{\xi', \xi} E(\tau, \xi', \xi), \;\; \sigma = -H(\xi', \eta). \end{aligned} \tag{4.6}$$

In particular, E *is a solution to the generalized double Hamilton-Jacobi equation:*

$$\left. \begin{aligned} \sigma + H(\xi, \eta') = 0 \\ \sigma + H(\xi', \eta) = 0 \end{aligned} \right\} \;\; \textit{for all} \;\; (\sigma, -\eta, \eta') \in \partial E(\tau, \xi', \xi) \;\; \textit{when} \; \tau > 0. \tag{4.7}$$

Proof. We get the equivalence of the first three conditions by applying [1; Theorem 2.5] to V_E, once again following the pattern of reformulation in the proof of Proposition 4.5, but for that purpose it is necessary to know the Hamiltonian H_E for the Lagrangian L_E in (4.4). This calculates out simply to $H_E(x', x, y', y) = H(x, y)$. To add the fourth condition in (4.6), we utilize the subgradient description in Theorem 4.1. Along any Hamiltonian trajectory, H is constant (as proved in [11]), so if the trajectory goes from (ξ', η) to (ξ, η') we must have $H(\xi, \eta') = H(\xi', \eta)$. □

The double Hamilton-Jacobi equation for E isn't surprising in view of the one for K in Theorem 3.1. Indeed, each double equation is essentially equivalent to the other by virtue of the relations in Theorem 4.1. It follows that E is uniquely determined by (4.7) and the initial condition in its definition (2.1). An earlier viscosity version of the double equation for E in simpler cases where E is finite can be seen in the book of Lions [12]. In general cases where E can be discontinuous and take on ∞, however, the Hamilton-Jacobi characterization of K has a major advantage over the one for E in Theorem 4.8, due to the assured finiteness and local Lipschitz continuity of $K(\tau, \xi, \eta)$ (Theorem 3.5), its smoothness in τ (Theorem 3.1), and its semidifferentiability everywhere with respect to all arguments jointly (Theorem 3.6).

5. Application to Hopf-Lax Formulas and Their Generalization

Upper and lower envelope representations of value functions as solutions to Hamilton-Jacobi equations first appeared in works of Hopf [13] and Lax [14] in very particular situations where the Hamiltonian $H(x, y)$ is independent actually of x. We inspect the state-independent case as an example within our framework and then go on to describe how our results cover an extension of the Hopf-Lax formulas beyond that case. The aim is to provide further perspective on how our formulas for value functions tie in with Hamilton-Jacobi theory.

Example 5.1 (formulas of classical Hopf-Lax type)**.** *Suppose that $L(x, v) = L_0(v)$ for a coercive, proper, lsc, convex function $L_0 : I\!R^n \to \overline{I\!R}$, or that $H(x, y) = H_0(y)$ for a finite convex function $H_0 : I\!R^n \to I\!R$, these assumptions being equivalent through the conjugacy relations $H_0 = L_0^*$, $L_0 = H_0^*$. Then the dualizing kernel is given by*

$$K(\tau, \xi, \eta) = \langle \xi, \eta \rangle - \tau H_0(\eta), \tag{5.1}$$

whereas the fundamental kernel is given by

$$E(\tau, \xi', \xi) = \begin{cases} \tau L_0\big(\tau^{-1}[\xi - \xi']\big) & \text{if } \tau > 0, \\ 0 & \text{if } \tau = 0,\ \xi - \xi' = 0, \\ \infty & \text{if } \tau = 0,\ \xi - \xi' \neq 0. \end{cases} \tag{5.2}$$

Thus, for any initial function $g : I\!R^n \to \overline{I\!R}$ one has the upper envelope formula

$$V(\tau, \xi) = \inf_{\xi'} \Big\{ g(\xi') + \tau L_0\big(\tau^{-1}[\xi - \xi']\big) \Big\}, \tag{5.3}$$

while if g is convex, proper, and lsc, one also has the lower envelope formula

$$V(\tau, \xi) = \sup_{\eta} \Big\{ \langle \xi, \eta \rangle - \tau H_0(\eta) - g^*(\eta) \Big\}. \tag{5.4}$$

Proof. Conditions (A1), (A2) and (A3) are fulfilled, since this amounts to Example 2.3 with $A = 0$ and $G \equiv 0$. The formula for K in (5.1) follows at once from the second half of the double Hamilton-Jacobi equation in Theorem 3.1, according to which $(\partial K/\partial \tau)(\tau, \xi, \eta) = -H_0(\eta)$. The formula for E in (5.2) then follows from the general one for E in terms of K in (2.5). Finally, we get the upper envelope representation from Theorem 2.1 and the lower envelope representation from Theorem 2.5. □

The duality between the upper and lower envelope representations in this example can also be seen from the angle that (5.4) can be written as

$$V(\tau, \cdot) = (g^* + \tau H_0)^*, \tag{5.5}$$

whereas the right side of (5.3) gives the well known formula of convex analysis for such a conjugate function in terms of the functions $g^{**} = g$ and $H_0^* = L_0$ (see [4; §11] for instance). In the traditions of Hamilton-Jacobi theory going back to Hopf [13], g^* and H_0^* don't appear and the formulas for these functions are substituted instead. The upper representation comes out then as

$$V(\tau, \xi) = \inf_{\xi'} \sup_{\eta} \Big\{ g(\xi') + \langle \xi - \xi', \eta \rangle - \tau H_0(\eta) \Big\}. \tag{5.6}$$

while the lower representation becomes

$$V(\tau, \xi) = \sup_{\eta} \inf_{\xi'} \Big\{ g(\xi') + \langle \xi - \xi', \eta \rangle - \tau H_0(\eta) \Big\}. \tag{5.7}$$

Nowadays, though, with the Legendre-Fenchel transform so well understood, there's no reason not to simplify these expressions by writing them with conjugate functions. The equality between the "inf sup" in (5.6) and the "sup inf" in (5.7) falls into the pattern of minimax representations of primal and dual optimization problems of convex type for which there is, by now, an enormous literature; see [3] and [4; Chapter 11]. Generally speaking, such an equality is deeply involved with convexity and requires other qualifications besides. Such qualifications are met here because of our assumptions (A).

Although both (5.6) and (5.7) were proposed by Hopf [13] as possible formulas for solutions to a generalized Hamilton-Jacobi PDE in the mode of

$$u_t(t, x) + H_0(u_x(t, x)) = 0, \qquad u(0, x) = g(x),$$

the first of these is often called the Lax formula because of its appearance in a special case in the earlier paper of Lax [14] on hyperbolic conservation laws.

In work since Hopf, the lecture notes of Lions [12] and Evans [15] have provided further treatment of Hopf-Lax formulas. The paper of Bardi and Evans [16] deserves particular mention. Those authors proved that the upper formula in (5.6), or equivalently (5.3), gives the unique viscosity solution to the Hamilton-Jacobi equation in the case of a finite convex function H_0 and a possibly nonconvex function g that is globally Lipschitz continuous; alternatively by Evans [15], g can be merely continuous if H_0 is coercive. (Coercivity of H_0 corresponds in convex analysis to finiteness of L_0.) In Example 5.1, this formula has been seen to give the value function V regardless of such extra conditions on g or H_0.

Bardi and Evans [16] also showed that the lower formula (5.4), or equivalently (5.7), gives the unique viscosity solution as long as g is convex and globally Lipschitz continuous (which is known in convex analysis to correspond to the effective domain of g^* being bounded). Recently Alvarez, Barron and Ishii [17] have removed these restrictions: the assertion holds true for all lsc, proper, convex functions g. In the context of Example 5.1, therefore, it follows that the value function V is the unique viscosity solution—in the sense of Barron and Jensen [18] or Frankowska [8], [19] (who employ a subgradient equation in place of a pair of inequalities involving upper as well as lower subgradients).

In the case of the lower envelope formula, Bardi and Evans [16] don't actually assume that H_0 is convex but just that it is continuous, and they still are able then to identify the unique viscosity solution under their strong assumptions on g. Our framework doesn't cover that feature, because the case is not one of optimization and there is no value function V of type (1.1) as a solution candidate.

We demonstrate now that the formulas in Example 5.1 can be extended to a significantly larger class of situations connected with optimal control (in the manner explained after Example 2.3), where the Lagrangian and Hamiltonian *aren't* state-independent, while maintaining their relatively explicit character. Again we emphasize that in the absence of a uniqueness theorem in Hamilton-Jacobi theory capable of handling all the Hamiltonians and value functions in our framework these formulas, although they uniquely describe value functions, can't yet be claimed to give unique Hamilton-Jacobi solutions.

Example 5.2 (extended Hopf-Lax formulas with linear state dependence)**.** *Suppose that* $L(x,v) = L_0(v - Ax)$ *for a coercive, proper, lsc, convex function* $L_0 : \mathbb{R}^n \to \overline{\mathbb{R}}$, *or that* $H(x,y) = \langle Ax, y\rangle + H_0(y)$ *for a finite convex function* $H_0 : \mathbb{R}^n \to \mathbb{R}$, *these assumptions being equivalent through the conjugacy relations* $H_0 = L_0^*$, $L_0 = H_0^*$. *Here* A *is any* $n \times n$ *matrix. Let* A^* *be the transpose of* A *and define* $\Psi : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$ *by*

$$\Psi(\tau,\eta) := \int_0^\tau H_0\big(e^{-tA^*}\eta\big)\,dt, \tag{5.8}$$

this expression being finite and convex in η. Then the dualizing kernel is given by

$$K(\tau, \xi, \eta) = \langle e^{-\tau A}\xi, \eta\rangle - \Psi(\tau, \eta). \tag{5.9}$$

and the fundamental kernel is given by

$$E(\tau, \xi', \xi) = \Phi(\tau, e^{-\tau A}\xi - \xi'), \tag{5.10}$$

where $\Phi(\tau, \zeta) = \sup_\eta \{\langle \zeta, \eta\rangle - \Psi(\tau, \eta)\}$, or in other words, $\Phi(\tau, \cdot)$ is the convex function conjugate to $\Psi(\tau, \cdot)$. Thus, for any initial function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ one has the upper envelope representation

$$\begin{aligned} V(\tau, \xi) &= \inf_{\xi'} \Big\{ g(\xi') + \Phi(\tau, e^{-\tau A}\xi - \xi') \Big\} \\ &= \inf_{\xi'} \sup_\eta \Big\{ g(\xi') + \langle e^{-\tau A}\xi - \xi',\, \eta\rangle - \Psi(\tau, \eta) \Big\}, \end{aligned} \tag{5.11}$$

while if g is convex, proper, and lsc, one also has the lower envelope representation

$$\begin{aligned} V(\tau, \xi) &= \sup_\eta \Big\{ \langle e^{-\tau A}\xi, \eta\rangle - \Psi(\tau, \eta) - g^*(\eta) \Big\} \\ &= \sup_\eta \inf_{\xi'} \Big\{ g(\xi') + \langle e^{-\tau A}\xi - \xi',\, \eta\rangle - \Psi(\tau, \eta) \Big\}. \end{aligned} \tag{5.12}$$

Proof. Fix (τ, ξ, η) and let $k(t) := K(t, \xi, y(t))$ for $y(t) := e^{(t-\tau)A^*}\eta$. From Theorem 3.5, k is Lipschitz continuous on $[0, \tau]$. We have $k(\tau) = K(\tau, \xi, \eta)$ and

$$k(0) = K(0, \xi, y(0)) = \langle \xi, y(0)\rangle = \langle \xi, e^{-\tau A^*}\eta\rangle = \langle e^{-\tau A}\xi, \eta\rangle. \tag{5.13}$$

Furthermore, from the semidifferentiability of K in Theorem 3.6 and its differentiability with respect to τ we have (almost everywhere)

$$\dot{k}(t) = (\partial K/\partial \tau)(\tau, \xi, y(t)) + \min\big\{ \langle \xi', \dot{y}(t)\rangle \,\big|\, \xi' \in \tilde{\partial}_\eta K(t, \xi, y(t)) \big\}, \tag{5.14}$$

where $\dot{y}(t) = A^* y(t)$. For each $t \in [0, \tau]$ let $x(t)$ denote some vector ξ' for which the minimum in (5.14) is attained. Then

$$\begin{aligned} \dot{k}(t) =& (\partial K/\partial \tau)(\tau, \xi, y(t)) + \langle x(t), \dot{y}(t)\rangle, \\ & \text{where } \langle x(t), \dot{y}(t)\rangle = \langle x(t), A^* y(t)\rangle = \langle Ax(t), y(t)\rangle. \end{aligned} \tag{5.15}$$

The second of the Hamilton-Jacobi equations in Theorem 3.1 gives us

$$(\partial K/\partial \tau)(\tau, \xi, y(t)) = -H(x(t), y(t)) = -\langle Ax(t), y(t)\rangle - H_0(y(t)). \tag{5.15}$$

In combining (5.15) and (5.16) we get $\dot{k}(t) = -H_0(y(t)) = -H_0(e^{(t-\tau)A^*}\eta)$, hence

$$k(\tau) = k(0) - \int_0^\tau H_0\big(e^{(t-\tau)A^*}\eta\big)dt,$$

with the integral equaling $\Psi(\tau,\eta)$ (as seen through time reversal). The desired formula for K in (5.9) comes out now from (5.13) and the fact that $k(\tau) = K(\tau,\xi,\eta)$.

The corresponding formula for E in (5.10) is immediate then from (2.5), and the envelope representations are valid on the basis of Theorems 2.1 and 2.5. □

Example 5.2 may be compared to a recent result of Arisawa and Tourin in [20], extending the upper envelope formula (5.3) to a very special case of state-dependent Hamiltonians of concave-convex type. Those authors take $I\!R^n$ to be $I\!R^m \times I\!R^m$ and treat

$$H(x,y) = H(x_1,x_2;y_1,y_2) = \langle x_2,y_1\rangle + h(y_2)$$

with h a finite *coercive* convex function on $I\!R^m$ having $\min h = h(0) = 0$. For that case they work out a more detailed expression for the fundamental kernel than the one in (5.10).

The formulas in Example 5.2 reduce to the familiar ones in Example 5.1 when $A = 0$, of course. The big difference is that with $A \neq 0$ they can be applied to optimal control problems with dynamics $\dot{x} = Ax + Bu$ through the connection laid out in §2 after Example 2.3. Then $H_0(y) = F^*(B^*y)$ for a finite convex conjugate function F^*, hence

$$\Psi(\tau,\eta) := \int_0^\tau F^*(B^* e^{-tA^*}\eta)\,dt.$$

In many situations it could well be possible to generate the values and even subgradients of Ψ numerically. The lower envelope representation of V in (4.12) could then, in light of Theorems 4.2 and 4.3, furnish an effective way of generating subgradients (or approximate subgradients) of V for potential use in feedback rules, through solving real-time optimization subproblems in $I\!R^n$.

The case in Example 5.2 is still relatively special within our framework. What might be said about value functions that come from state-dependent Hamiltonians more generally under our basic assumptions, as translated through Proposition 2.2? Everything really goes back to Theorem 2.6. The extent to which the basic formulas (2.13) and (2.14) in Theorem 2.6 can be regarded as "explicit" analogs of the classical Hopf-Lax formulas (5.6) and (5.7) hinges on how far one can go in obtaining an "explicit" formula for the dualizing kernel K that we have introduced. This requires an exploration of favorable cases in which the Hamilton-Jacobi characterization of K in Theorem 3.1 can be made to yield an "explicit" expression for K.

Here we have shown that the classical case in Example 5.1, where K is given by (5.1), can be extended with essentially no loss to the case in Example 5.2, where K is given by (5.9). Further research might yield other attractive cases. In the end, though, it must be borne in mind that the notion of what is an explicit expression for a function has evolved considerably in mathematics, and now is more a matter of whether a formula supports insightful analysis tied to modern computational methodology.

References

1. R.T. Rockafellar and P.R. Wolenski, "Convexity in Hamilton-Jacobi theory 1: dynamics and duality," SIAM J. Control Opt.
2. M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, 1997.
3. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
4. R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer-Verlag, 1997.
5. R.T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM J. Control Opt. 25 (1987), 781–814.
6. R.T. Rockafellar, "Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs," SIAM J. Control Opt. 27 (1989), 1007–1025.
7. R.T. Rockafellar, "Existence theorems for general control problems of Bolza and Lagrange," Advances in Math. 15 (1975), 315–333.
8. H. Frankowska, "Hamilton-Jacobi equations: viscosity solutions and generalized gradients," J. Math. Anal. Appl. 141 (1989), 21–26.
9. 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.
10. H. Ishii, "A comparison result for Hamilton-Jacobi equations without growth condition on solutions from above," preprint, 1998.
11. R. T. Rockafellar, "Generalized Hamiltonian equations for convex problems of Lagrange," Pacific J. Math. 33 (1970), 411–428.
12. P.-L. Lions, *Generalized Solutions of Hamilton-Jacobi Equations*, Research Notes in Mathematics, Vol. 62, Pitman, 1993.
13. E. Hopf, "Generalized solutions of non-linear equations of first order," J. Math. Mech. 14 (1965), 201–230.

14. P.D. Lax, "Hyperbolic Systems of Conservation Laws II," Comm. on Pure & Appl. Math. 10 (1957), 537–566.

15. L.C. Evans, *Partial Differential Equations*, Berkeley Mathematics Lecture Notes, Vols. 3A & 3B, Univ. of California at Berkeley, 1993.

16. M. Bardi and L.C. Evans, "On Hopf's formulas for solutions of Hamilton-Jacobi equations," Nonlinear Analysis: Theory, Meth. & Appl. 8 (1984), 1373–1381.

17. O. Alvarez, E.N. Barron and H. Ishii, "Hopf-Lax formulas for semicontinuous data," preprint 1998.

18. E.N. Barron and R. Jensen, "Semicontinuous viscosity solutions for Hamilton-Jacobi equations with convex Hamiltonians," Commun. PDE 15 (1990), 1713–1742.

19. H. Frankowska, "Lower semicontinuous solutions of Hamilton-Jacobi-Bellman equations," SIAM J. Control. Opt. 31 (1993), 257-272.

20. M. Arisawa and A. Tourin, "Regularizing effects for a class of first-order Hamilton-Jacobi equations," Nonlinear Analysis: Theory, Meth. & Appl. 29 (1997) 1405–1419.