SIAM J. CONTROL OPTIM.
Vol. 43, No. 5, pp. 1787–1811

CONVEX OPTIMAL CONTROL PROBLEMS WITH SMOOTH HAMILTONIANS*

RAFAL GOEBEL†

Abstract. Optimal control problems with convex costs, for which Hamiltonians have Lipschitz continuous gradients, are considered. Examples of such problems, including extensions of the linear-quadratic regulator with hard and possibly state-dependent control constraints, and piecewise linear-quadratic penalties are given. Lipschitz continuous differentiability and strong convexity of the terminal cost are shown to be inherited by the value function, leading to Lipschitz continuity of the optimal feedback. With no regularity assumptions on the limiting problem, epi-convergence of costs, which can be equivalently described by pointwise convergence of Hamiltonians, is shown to guarantee epi-convergence of value functions. Resulting schemes of approximating any concave-convex Hamiltonian by continuously differentiable ones are displayed. Auxiliary results about existence and stability of saddle points of quadratic functions over polyhedral sets are also proved. Tools used are based on duality theory of convex and saddle functions.

Key words. optimal control, differentiable Hamiltonian, convex value function, optimal feedback regularity, conjugate duality, epi-convergence, piecewise linear-quadratic function, saddle function

AMS subject classifications. 49N60, 49N10, 49M29, 90C47

DOI. 10.1137/S0363012902411581

1. Introduction.

Given a point $(\tau, \xi) \in (-\infty, T] \times \mathbb{R}^n$, a *terminal cost* $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and a *Lagrangian* $L : \mathbb{R}^{2n} \mapsto \overline{\mathbb{R}}$, consider the generalized problem of Bolza:

$$(1) \qquad \mathcal{P}(\tau, \xi) : \quad \text{minimize} \int_\tau^T L(x(t), \dot{x}(t))\, dt + g(x(T)) \;\; \text{subject to} \;\; x(\tau) = \xi,$$

with the minimization carried out over all absolutely continuous arcs $x : [\tau, T] \mapsto \mathbb{R}^n$. While it is well known that a smooth Lagrangian need not lead to a regular (maximized) *Hamiltonian*, which is defined by

$$(2) \qquad H(x, y) = \sup_{v \in \mathbb{R}^n} \{ y \cdot v - L(x, v) \},$$

it is less appreciated that nonsmooth and infinite-valued L may give rise to a smooth H. We explore this fact here, focusing on problems with convex g and L, and with Hamiltonians for which ∇H is Lipschitz continuous.

Optimal control problems with explicit linear dynamics, hard and possibly state-dependent control constraints, and state and control penalties can be reformulated in Bolza format; see Clarke [10] or Rockafellar [18]. In section 2 we show that a broad range of optimal control problems, including various extensions of the classical linear-quadratic regulator, can lead to a smooth Hamiltonian. This makes the results of section 3 applicable to the control framework.

*Received by the editors July 18, 2002; accepted for publication (in revised form) May 31, 2004; published electronically March 22, 2005. Research carried out at the Department of Mathematics at the University of Washington, the Centre for Experimental and Constructive Mathematics at Simon Fraser University, and the Department of Mathematics at the University of British Columbia.

http://www.siam.org/journals/sicon/43-5/41158.html

†Center for Control Engineering and Computation, ECE, University of California, Santa Barbara, CA 93106-9650 (rafal@ece.ucsb.edu).

Section 3 studies regularity of the *value function* $V : (-\infty, T] \times \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, defined as the optimal value in $\mathcal{P}(\tau, \xi)$ parameterized by the initial condition. Lipschitz continuity of ∇g and ∇H is shown to lead to Lipschitz ∇V; explicit bounds on the constants are given. We stress that no smoothness or even finiteness assumptions are made on L. For comparison, in a nonconvex setting, if the method of characteristics associated with the Hamilton–Jacobi equation has no shocks (in our setting, this automatically holds; see Goebel [14]), the value function inherits continuous differentiability from that of the terminal cost, under further regularity assumptions on L; see Byrnes and Frankowska [7] and also Caroff and Frankowska [8]. We note that while we work with continuously differentiable Hamiltonians, we do not require them to be C^2. This raises an obstacle to Riccati-like descriptions of V as given by Byrnes [6] and Caroff and Frankowska [9] but allows for treatment of problems discussed in section 2 (for those, hard constraints or piecewise linear-quadratic penalties exclude C^2 smoothness of the Hamiltonian).

Our interest in Lipschitz continuity of ∇V comes from the role the gradient plays in constructing optimal feedback. With the regularity of H and V as just mentioned, the adjoint variable to an optimal arc $x(t)$ is just $-\nabla V(t, x(t))$, and the resulting optimal feedback mapping is continuous and Lipschitz in the state variable. Consequently, the classical differential equation tools and existence and uniqueness results apply. This is not the case for the general convex but nonsmooth setting—there, the resulting set-valued feedback may be highly irregular, even for piecewise linear-quadratic costs; see Goebel [14].

In section 4 we show that regular Bolza problems—those with Lipschitz ∇g and ∇H—can approximate any convex problem fitting our mild growth conditions. The approximations are explicit and, together with direct proofs in section 3, they should yield insights to numerical implementation of the method of characteristics. The approximations rely on a more general result concluding the convergence of value functions, defined by any converging to g and L sequences of initial costs and Lagrangians. As the functions in question need not be finite, we rely on the concept of epi-convergence. Its extensions to infinite dimensions, where various topologies have to be considered, have been used to study control problems; see Buttazzo and Dal Maso [5] and Briani [4]. These works, while not requiring full convexity, had stricter growth assumptions and finite cost functions, and dealt, respectively, with convergence of optimal solutions and pointwise convergence of value functions. Moreover, methods used here are significantly different; we employ a dual problem leading to a dual value function, as described by Rockafellar and Wolenski [27]. The symmetry between the primal and dual problem, and the fact that epi-convergence is preserved by convex conjugacy (vaguely speaking, the "lower half" of epi-convergence dualizes to the "upper" and vice-versa), requires us to show just one side (the easier one) of epi-convergence. A similar idea was employed by Joly and Thelin [17] in the study of convex integral functionals; here we keep to a minimum the discussion of such issues, preferring to work with functions on finite-dimensional spaces.

Some of our results are most conveniently handled with the tools related to conjugacy and epi/hypo-convergence of saddle functions; see, respectively, chapters 33–37 in Rockafellar [19], Attouch and Wets [2], and Attouch, Azé, and Wets [1]. We present the necessary background in section 5. In particular, our results on finiteness and differentiability of piecewise linear-quadratic Hamiltonians are closely related to existence and uniqueness of saddle points of an auxiliary quadratic function defined on a product of polyhedral sets. Such a function also appears as a Lagrangian in extended linear-quadratic programming; see Rockafellar [25]. (In convex optimization, Lagrangians are saddle functions used, in particular, to express optimality conditions.)

2. Extended piecewise linear-quadratic optimal control. In this section we illustrate that control problems with constraints and nondifferentiable costs can possess Hamiltonians with desirable smoothness properties. Let us start with the following example.

Example 2.1 (separable smooth Hamiltonian). Suppose that $L(x,v) = k(x) + l(v)$, and l is a convex function. Then the Hamiltonian $H(x,y)$ is differentiable and ∇H is (globally) Lipschitz continuous if and only if k has this property and l is strongly convex (that is, $v \mapsto l(v) - \rho\|v\|^2$ is convex for some $\rho > 0$). Indeed, $H(x,y) = -k(x) + l^*(y)$, where $l^*(y) = \sup_v \{y \cdot v - l(v)\}$ is the convex function conjugate to l. The statement about differentiability of l^*, and the bound $(2\rho)^{-1}$ on its Lipschitz constant, can be found in [26, Proposition 12.60]. Strongly convex functions include functions of the form $l(v) = v \cdot Pv$ for $v \in C$ while $l(v) = +\infty$ for $v \notin C$, where P is a symmetric positive definite matrix and C is any convex set, but the (piecewise) quadratic structure is not necessary. For example, the "barrier function" $l(v) = -\log(1 - |v|)$ for $v \in (-1,1)$, $l(v) = +\infty$ otherwise, is strongly convex (note the nondifferentiability at the origin), we have $l^*(y) = 0$ for $y \in [-1,1]$, $l^*(y) = |y| - \log|y| - 1$ otherwise, and l^* has a Lipschitz continuous gradient.

In the remainder of this section, we discuss control problems with explicit mention of controls, dynamics, constraints, and penalties. Translating such problems to the generalized format of Bolza (1) is possible; see Clarke [10] for a general exposition or Rockafellar [18] for details in the convex case. This enables the translation of results of sections 3 and 4 to the control setting. As finiteness of the Hamiltonian and of the value function implies that an optimal arc $x(\cdot)$ has a bounded derivative—in the control setting below, $u(\cdot)$ is bounded—the potential discrepancy between minimizing over absolutely continuous arcs in $\mathcal{P}(\tau,\xi)$ and over L^2 controls in the linear-quadratic regulator is avoided in most cases under discussion.

Separable Hamiltonians of Example 2.1, and their biaffine perturbations given by $H(x,y) = y \cdot Ax - k(x) + l^*(y)$, appear, for example, in the linear-quadratic regulator with control constraints of the type $u(t) \in U$. However, state-dependent constraints $u(t) \le Cx(t) + d$ or mixed control and state penalties call for the analysis of a more general class of Hamiltonians.

Given vectors p, q; matrices A, B, C, D, P, Q; and sets U, V of appropriate dimensions, consider the following control problem $\mathcal{C}(\tau,\xi)$:

$$
\begin{aligned}
&\min \int_\tau^T \left[p \cdot u(t) + \frac{1}{2} u(t) \cdot Pu(t) + \rho_{V,Q}(q - Cx(t) - Du(t)) \right] dt + g(x(T)) \\
&\text{s.t.} \qquad \dot{x}(t) = Ax(t) + Bu(t), \quad u(t) \in U \text{ a.e.}, \quad x(\tau) = \xi,
\end{aligned} \tag{3}
$$

with the minimization carried out over all integrable controls $u : [\tau, T] \mapsto \mathbb{R}^k$. The convex and possibly infinite-valued penalty function $\rho_{V,Q}(\cdot)$ is given by

$$
\rho_{V,Q}(s) = \sup_{v \in V} \left\{ s \cdot v - \frac{1}{2} v \cdot Qv \right\}. \tag{4}
$$

The key assumptions, guaranteeing not only the convex structure of the problem, but also the piecewise linear-quadratic structure of the resulting Hamiltonian, is stated below. We recall that a set is polyhedral if it is the intersection of finitely many closed half-spaces; consequently, a polyhedral set is always closed and convex (but not necessarily bounded).

$$\text{Matrices } P \text{ and } Q \text{ are symmetric positive semidefinite.} \quad \text{Sets } U \text{ and } V \text{ are nonempty and polyhedral.} \tag{5}$$

Such *extended piecewise linear-quadratic optimal control* format was proposed by Rockafellar [22]. Therein, optimality conditions taking advantage of duality were stated. Their minimax form (related to the structure of the Hamiltonian as outlined in Example 5.1 and the surrounding discussion) facilitates the use of various primal-dual optimization methods to discretized problems; see Rockafellar and Zhu [28], Wright [29], and Zhu [30].

Here, we begin by describing when the control problem (3) fits the convex duality framework of Rockafellar and Wolenski [27], we call upon some of their results in later sections. The Hamiltonian for $\mathcal{C}(\tau,\xi)$ (see Rockafellar [23] or apply (2) to the Lagrangian (11)) is

$$H(x,y) = y\cdot Ax + J^*(B^*y, Cx), \tag{6}$$

where the function J^*, convex in a and concave in b, is given by

$$J^*(a,b) = \sup_{u\in U}\inf_{v\in V}\left\{a\cdot u + b\cdot v - p\cdot u - \frac{1}{2}u\cdot Pu + q\cdot v + \frac{1}{2}v\cdot Qv + v\cdot Du\right\}. \tag{7}$$

Here and in what follows, B^* denotes the transpose of B. The Hamiltonian (and the Lagrangian (11)) are piecewise linear-quadratic: their effective domains are unions of finitely many polyhedral sets, relative to each of which the functions are linear-quadratic (Goebel [12]). Goebel and Rockafellar [15] showed that if a piecewise linear-quadratic Hamiltonian is finite, the control problem fits the framework of [27]. A particular consequence of such a structure of the Hamiltonian, shown in [15], is that the knowledge of $V(\overline{\tau},\cdot)$ at any particular $\overline{\tau}\in(-\infty,T]$ determines V (uniquely) for all times $\tau\in(-\infty,T]$.

In our setting, the finiteness of J^*, which implies that of the Hamiltonian, is described by the following result. For a given set S, the *recession cone* S^∞ consists of all z such that $S+z\subset S$, while for a cone K, the polar cone K^* is $\{w \mid w\cdot z\le 0 \text{ for all } z\in K\}$.

THEOREM 2.2 (finiteness of J^*). *Assume that* (5) *holds. Then, the function J^* is finite if and only if the following is satisfied:*

$$\begin{cases} U^\infty\cap\ker P\cap(-D^*V^\infty)^* = \{0\},\\ V^\infty\cap\ker Q\cap(DU^\infty)^* = \{0\}.\end{cases} \tag{8}$$

Above, $(DU^\infty)^* = \{w \mid D^*w\in U^{\infty *}\}$ and $(-D^*V^\infty) = \{z \mid -Dz\in V^{\infty *}\}$, this comes directly from the definitions. The proof of Theorem 2.2, as well as that of Theorem 2.4, requires some notions of saddle function theory. We present them and the proofs in section 5. Note that if D is the zero matrix (which excludes many modeling options), the function J^* is separable: $J^*(a,b) = \sup_{u\in U}\left\{a\cdot u - \frac{1}{2}u\cdot Pu\right\} - \sup_{v\in V}\left\{-b\cdot v - \frac{1}{2}v\cdot Qv\right\}$, and (8) reduce to known conditions on recession cones and kernels, we mention them in the discussion preceding Example 3.8.

COROLLARY 2.3. *Assume that* (5) *holds.*

(a) *If U is a bounded set, J^* is finite if and only if $V^\infty\cap\ker Q=\{0\}$ (and this holds in particular when V is bounded or Q is positive definite).*

(b) *When sets U and V are cones, J^* is finite if and only if*

$$\begin{cases} U \cap \ker P \cap (-D^* V)^* = \{0\}, \\ V \cap \ker Q \cap (DU)^* = \{0\}. \end{cases}$$

Arguments of Example 2.1 imply that in the separable case, as described before Corollary 2.3, positive definiteness of P and Q is equivalent to the differentiability of J^*. Below, we give a sufficient condition for differentiability, applicable to cases where $D \neq 0$ and not requiring the positive definiteness of P and Q. A somewhat extreme example, showing that this last property is not necessary, is as follows. For a and b one-dimensional, consider J^* with $p = q = P = Q = 0$, $D = 1$, and $U = V = \mathbb{R}$. Direct calculation shows that $J^*(a, b) = ab$.

THEOREM 2.4 (differentiability of J^*). *Assume that the following condition holds:*

$$\begin{cases} \ker P \cap [D^*(V^\infty \cap -V^\infty)]^\perp = \{0\}, \\ \ker Q \cap [D(U^\infty \cap -U^\infty)]^\perp = \{0\}. \end{cases} \tag{9}$$

Then J^* is differentiable and ∇J^* is Lipschitz continuous.

Lipschitz continuity of ∇J^*, while guaranteed by the proof, is automatic in the presence of differentiability of J^*. This is thanks to the piecewise linear-quadratic structure; if J^* is differentiable, then ∇J^* is piecewise affine (and there is finitely many pieces). The piecewise linear-quadratic structure furthermore implies that J^* is not C^2, unless it is in fact quadratic (and this excludes any hard constraints or piecewise linear-quadratic penalties in the underlying problem).

In the remainder of this section, we illustrate the modeling capabilities of the extended piecewise linear-quadratic control, and use Theorem 2.4 to conclude the differentiability of the Hamiltonian for various extensions of the linear-quadratic regulator. Computational methods for such problems in discrete time are of great interest in the engineering literature; see Bemporad et al. [3] and the references therein.

Given symmetric positive semidefinite matrices E and G and a symmetric and positive definite F, this classical problem is as follows:

$$\begin{aligned} &\min \int_\tau^T \frac{1}{2} \left(x(t) \cdot Ex(t) + u(t) \cdot Fu(t)\right) dt + \frac{1}{2} x(T) \cdot Gx(T), \\ &\text{s.t.} \qquad \dot{x}(t) = Ax(t) + Bu(t), \quad x(\tau) = \xi. \end{aligned} \tag{10}$$

Minimization is carried out over all L^2 controls $u(\cdot)$ on $[\tau, T]$ (optimal controls turn out to be bounded, and in fact continuous). The value function for (10) is $V(\tau, \xi) = \frac{1}{2}\xi \cdot S(\tau)\xi$, where the matrix $S(\cdot)$ solves the associated Riccati equation, the Hamiltonian is quadratic, and the optimal feedback is linear in the state. Results of section 3 will show that while constraints and penalties destroy the linear structure, the optimal feedback may still be Lipschitz continuous. Here, we focus on the regularity of the Hamiltonian.

The linear-quadratic regulator can of course be cast in the format (3), by taking

$$P = F,\ Q = I,\ U = \mathbb{R}^k,\ V = \mathbb{R}^n,\ C = \sqrt{E},\ D = 0,\ p = 0,\ q = 0.$$

Indeed, we obtain $\rho_{V,Q}(q - Cu - Dv) = \sup_v\{(-\sqrt{E}u) \cdot v - \frac{1}{2} v \cdot v\} = \frac{1}{2} u \cdot \sqrt{E}^* \sqrt{E}u$. It can be easily verified that conditions (8) and (9) are (obviously) satisfied.

Example 2.5 (fixed control constraints). A linear-quadratic regulator with a constraint $u(t) \in U$, for a nonempty polyhedral set U, certainly fits the format (3).

Thanks to the positive definiteness of $P = F$ and $Q = I$, conditions (8) and (9) hold, and thus the Hamiltonian is finite and differentiable (if U is bounded, the Hamiltonian remains finite but not differentiable if the matrix F is just positive-semidefinite). Direct calculation yields $J^*(a,b) = \rho_{U,F}(a) - \frac{1}{2}\|b\|^2$, and thus the Hamiltonian is $H(x,y) = y \cdot Ax - \frac{1}{2}x \cdot Ex + \rho_{U,F}(B^*y)$. Note that H is not C^2.

Example 2.6 (state-dependent inequality constraints on controls). Consider (10) with the following constraint on the control:

$$u(t) \le C_0 x(t) - q_0,$$

for some matrix C_0. Taking $U = \mathbb{R}^k,\ V = \mathbb{R}^n \times \mathbb{R}^k_+,\ P = F,\ p = 0$, and

$$Q = \begin{bmatrix} I_{n\times n} & 0_{n\times k} \\ 0_{k\times n} & 0_{k\times k} \end{bmatrix},\ q = \begin{bmatrix} 0_{n\times n} \\ q_0 \end{bmatrix},\ C = \begin{bmatrix} \sqrt{E} \\ C_0 \end{bmatrix},\ D = \begin{bmatrix} 0_{n\times k} \\ -I_{k\times k} \end{bmatrix},$$

where 0_n is a zero vector in $\mathbb{R}^n$, $0_{n\times k}$ is the zero matrix of appropriate dimension, etc., casts the problem in the framework of (3). We get, for $s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$ with $s_1 \in \mathbb{R}^n$, $s_2 \in \mathbb{R}^k$,

$$\begin{aligned} \rho_{V,Q}(s) &= \sup_{v\in V}\left\{ s\cdot v - \frac{1}{2}v \cdot Qv \right\} = \sup_{v_1\in\mathbb{R}^n, v_2\in\mathbb{R}^k_+}\left\{ s_1\cdot v_1 + s_2 \cdot v_2 - \frac{1}{2}v_1\cdot v_1 \right\} \\ &= \sup_{v_1\in\mathbb{R}^n}\left\{ s_1\cdot v_1 - \frac{1}{2}v_1 \cdot v_1 \right\} + \sup_{v_2\in\mathbb{R}^k_+}\{s_2\cdot v_2\} = \frac{1}{2}|s_1|^2 + \delta_{\mathbb{R}^k_-}(s_2), \end{aligned}$$

and thus, since

$$q - Cx - Du = \begin{pmatrix} -\sqrt{E}x \\ q_0 - C_0x + u \end{pmatrix},$$

expression $\rho_{V,Q}(q - Cu - Dv)$ equals

$$\frac{1}{2}x\cdot Ex + \delta_{\mathbb{R}^k_-}(q_0 - C_0x + u) = \frac{1}{2}x \cdot Ex + \begin{cases} 0 & \text{if } u \le C_0x - q_0, \\ +\infty & \text{otherwise.} \end{cases}$$

As desired, the penalty function enforces the inequality constraint.

We now check the finiteness and differentiability of the Hamiltonian. First, conditions in both (8) and (9) are satisfied since P is positive definite. We have $V^\infty = V$, $\ker Q = \{0_n\}\times\mathbb{R}^k$, and, since $U^{\infty *} = 0_n$, $(DU^\infty)^* = \{w \mid [0_{n\times n}, I_{k\times k}]w = 0\} = \mathbb{R}^n \times 0_k$; thus the second condition for finiteness is satisfied. Similarly, $[D(U^\infty \cap -U^\infty)]^\perp = \mathbb{R}^n \times 0_k$, and the Hamiltonian is differentiable.

Example 2.7 (state-dependent control constraints through quadratic penalties). Adding to the integrand in (10) the penalty function

$$\sum_{i=1}^{s}\begin{cases} 0 & \text{if } q_i - c_i\cdot x(t) - d_i \cdot u(t) \le 0, \\ \frac{1}{2}\lambda_i\left(q_i - c_i \cdot x(t) - d_i\cdot u(t)\right)^2 & \text{if } q_i - c_i \cdot x(t) - d_i\cdot u(t) > 0, \end{cases}$$

with $\lambda_i > 0$ leads to another problem in the extended piecewise linear-quadratic format. Indeed, set $U = \mathbb{R}^k,\ V = \mathbb{R}^n \times \mathbb{R}^s_+,\ P = F,\ p = 0$, and

$$Q = \begin{bmatrix} I_{n\times n} & 0_{n\times s} \\ 0_{s\times n} & \Lambda^{-1} \end{bmatrix},\ q = \begin{bmatrix} 0_{n\times n} \\ q_1 \\ \vdots \\ q_s \end{bmatrix},\ C = \begin{bmatrix} \sqrt{E} \\ c_1 \\ \vdots \\ c_s \end{bmatrix},\ D = \begin{bmatrix} 0_{n\times k} \\ d_1 \\ \cdots \\ d_s \end{bmatrix},$$

where Λ is a diagonal matrix with diagonal entries λ_i. It can be verified that the corresponding Hamiltonian function is finite and continuously differentiable (but not C^2).

We add that combining penalty functions from Example 2.7, with constraints of either Example 2.5 or 2.6, is possible in the extended piecewise linear-quadratic format. Moreover, these suggested combinations will lead to a differentiable Hamiltonian. In section 3 we will return to the examples above to describe the corresponding optimal feedback mappings.

3. Value function regularity. Techniques used in this and the following sections will rely in part on the Hamilton–Jacobi and duality theories developed for convex control problems in Rockafellar and Wolenski [27]. The required assumptions on the problem $\mathcal{P}(\tau, \xi)$ defined in (1), which we pose throughout this section, are stated below. The growth conditions in (A2), (A3) are quite mild, their detailed discussions can be found in [27] and also Goebel [14].

Assumption 3.1 (basic assumptions).

(A1) The functions $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and $L : \mathbb{R}^{2n} \mapsto \overline{\mathbb{R}}$ are proper, l.s.c., and convex.

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

The symbol $\overline{\mathbb{R}}$ stands for the interval $[-\infty, +\infty]$, a function $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ is said to be proper if it does not take on the value $-\infty$, and its effective domain $\operatorname{dom} f = \{x \mid f(x) < +\infty\}$ is nonempty; a function f is called coercive if $\lim_{|x|\to+\infty} \frac{f(x)}{|x|} = +\infty$.

Example 3.2 (piecewise linear-quadratic Lagrangian). Translating the control problem $\mathcal{C}(\tau, \xi)$ discussed in section 2 to the format of Bolza (1) (see [10] or [18]) leads to the Lagrangian

$$L(x, v) = \inf_u \left\{ p \cdot u + \frac{1}{2} u \cdot Pu + \rho_{V,Q}(q - Cx - Du) \mid v = Ax + Bu, u \in U \right\}. \tag{11}$$

In particular, the value function defined by (1) with the Lagrangian (11) is the same as that defined by (3). If (5) holds and the corresponding Hamiltonian (6) is finite (as is always the case if conditions (8) are in place), then the Lagrangian above satisfies Assumption 3.1; see [15, Corollary 4.5].

A key tool for the analysis of the regularity of the value function V is the global description of the graph of $\partial_\xi V(\tau, \cdot)$ as the image of $\operatorname{gph} \partial g$ under a certain flow mapping. Here, and in what follows, $\partial_\xi V$ denotes the subdifferential in the sense of convex analysis, of the convex function $\xi \mapsto V(\tau, \xi)$; the subdifferentials ∂g and $\partial_y H$ should also be understood in the convex sense; see Rockafellar [19, section 23]. The subdifferential $\tilde{\partial}_x H(x, y)$ of the concave function $H(\cdot, y)$ equals $-\partial_x (-H(x, y))$. If any of the mentioned functions are differentiable, the subdifferential reduces to the gradient. Consider the Hamiltonian inclusion

$$-\dot{y}(t) \in \tilde{\partial}_x H(x(t), y(t)), \quad \dot{x}(t) \in \partial_y H(x(t), y(t)). \tag{12}$$

A pair of absolutely continuous arcs $(x(\cdot), y(\cdot))$ on $[a, b]$ will be called a *Hamiltonian trajectory* if it satisfies (12) for almost all $t \in [a, b]$.

THEOREM 3.3 (flow of the value function). *One has $\eta \in \partial_\xi V(\tau, \xi)$ if and only if, for some $\eta^T \in \partial g(\xi^T)$, there is a Hamiltonian trajectory on $[T - \tau, T]$ from $(\xi, -\eta)$ to $(\xi^T, -\eta^T)$.*

The above result was shown by Rockafellar and Wolenski [27], as Theorem 2.4, in the setting of control problems with an initial cost function, and for which the value function is parameterized by a terminal constraint. A change of variables in the expression for the value function yields the result as described above.

In a less convex setting, descriptions of the (appropriately understood) subdifferential of the value function in the flavor of Theorem 3.3 are possible in some local sense, as long as the image of the subdifferential of the terminal cost under the Hamiltonian flow remains a subdifferential of a function—this is the case in our convex setting for any length of the time interval $[\tau, T]$. Under stronger smoothness assumptions than used here, the Hessian of the value function may then turn out to be a solution of an appropriate matrix Riccati differential equation; see Byrnes [6] and Caroff and Frankowska [9].

To illustrate Theorem 3.3, we show that a piecewise linear-quadratic problem need not yield a piecewise linear-quadratic value function. This is in contrast to discrete time problems.

Example 3.4 (loss of piecewise linear-quadratic structure). Consider a one-dimensional problem of Bolza with the cost functions

$$L(x,v) = \frac{1}{2}v^2 + \begin{cases} 0, & x < 0, \\ \frac{1}{2}x^2, & x \geq 0, \end{cases} \qquad g(x) = \frac{1}{2}(x+3)^2.$$

The corresponding Hamiltonian is piecewise linear-quadratic and differentiable, and its gradient is piecewise linear:

$$H(x,y) = \begin{cases} \frac{1}{2}y^2, & x < 0, \\ -\frac{1}{2}x^2 + \frac{1}{2}y^2, & x \geq 0, \end{cases} \qquad \nabla H(x,y) = \begin{cases} (0,y), & x < 0, \\ (-x,y), & x \geq 0. \end{cases}$$

A Hamiltonian trajectory $(x(\cdot), y(\cdot))$ must satisfy $\dot{x}(t) = y(t) = \mathit{const}$ when $x(t) < 0$, and $x(t) = \alpha e^t + \beta e^{-t}$, $y(t) = \alpha e^t - \beta e^{-t}$ for suitably chosen α, β when $x(t) > 0$.

The segment between $(-2,-1)$ and $(-1,-2)$ is contained in $\mathrm{gph}(-\nabla g)$. Parameterize the segment by $(x_s(T), y_s(T)) = (s-2, -s-1)$ with $s \in [0,1]$. Hamiltonian trajectories terminating at $(x_s(T), y_s(T))$ are given by

$$(x_s(t), y_s(t)) = \begin{cases} \Big((s+1)(T-t) + s - 2, -s-1\Big), & 0 \leq T-t \leq \frac{2-s}{s+1}, \\ (s+1)\Big(\sinh(T-t-\frac{2-s}{s+1}), -\cosh(T-t-\frac{2-s}{s+1})\Big), & T-t \geq \frac{2-s}{s+1}. \end{cases}$$

It is easy to check that for any $t < T-1$, the set $\{(x_s(t), y_s(t)), s \in [0,1]\}$ is not a straight line segment, nor is it a union of segments. But $\{(x_s(t), y_s(t)), s \in [0,1]\} \subset \mathrm{gph}\, -\partial_\xi V(T-t, \cdot)$, and consequently, $V(T-t,\cdot)$ is not piecewise linear-quadratic.

LEMMA 3.5. *Suppose that H is differentiable and ∇H is Lipschitz continuous with constant K. Let $g(x) = \frac{1}{2}a\|x\|^2 + b \cdot x + c$, with $a > 0$. Then, for all $\tau \leq T$,*

(a) *$V(\tau,\cdot)$ is differentiable with $\nabla_\xi V(\tau,\cdot)$ Lipschitz continuous, with constant*

$$a\left[1 + \left(e^{K(T-\tau)} - 1\right)\sqrt{1+a^{-2}}\right]^2;$$

(b) *$V(\tau,\cdot)$ is strongly convex with constant*

$$a\left[1 + \left(e^{K(T-\tau)} - 1\right)\sqrt{1+a^2}\right]^{-2}.$$

Proof. Fix $\tau \leq T$. Pick two points, $\xi_1^T \neq \xi_2^T$ in $\mathbb{R}^n$, and let $\eta_i^T = \nabla g(\xi_i^T) = a\,\xi_i^T + b$, $i = 1, 2$. Let $(x_i(\cdot), y_i(\cdot))$ be the Hamiltonian trajectory on $[\tau, T]$ with $(x_i(T), y_i(T)) = (\xi_i^T, -\eta_i^T)$ for $i = 1, 2$. As ∇H is Lipschitz continuous, the Hamiltonian trajectories and the endpoints just mentioned are well defined. To shorten the notation, let $\alpha(t) = x_1(t) - x_2(t)$, $\beta(t) = y_1(t) - y_2(t)$.

The monotone structure of ∇H implies that $\alpha(t) \cdot \beta(t)$ is a nondecreasing function of t; see Theorem 4 in [20]—this is a distinguishing feature of a convex problem. Consequently,

$$-\|\alpha(\tau)\|\,\|\beta(\tau)\| \leq \alpha(\tau) \cdot \beta(\tau) \leq \alpha(T) \cdot \beta(T) = -a\|\alpha(T)\|^2,$$

and thus $\|\alpha(\tau)\|\,\|\beta(\tau)\| \geq a\|\alpha(T)\|^2$. Lipschitz continuity of ∇H implies that

$$\|\beta(\tau)\| \leq \|\beta(T)\| + \left(e^{K(T-\tau)} - 1\right)\|(a(T), b(T))\|. \tag{13}$$

Maximizing $\|\beta(\tau)\|/\|\alpha(\tau)\|$ subject to the last two inequalities (this is a simple two-dimensional calculus problem) yields

$$\frac{\|\beta(\tau)\|}{\|\alpha(\tau)\|} \leq \frac{\|\beta(T)\| + \left(e^{K(T-\tau)} - 1\right)\|(a(T), b(T))\|}{a\|\alpha(T)\|^2 \Big/ \left[\|\beta(T)\| + \left(e^{K(T-\tau)} - 1\right)\|(a(T), b(T))\|\right]},$$

which simplifies to $\|\beta(\tau)\|/\|\alpha(\tau)\| \leq a\left[1 + \left(e^{K(T-\tau)} - 1\right)\sqrt{1 + a^{-2}}\right]^2$, since $\beta(T) = -a\,\alpha(T)$. Thanks to Theorem 3.3, the last bound is in fact a bound on $\|\eta_1 - \eta_2\|/\|\xi_1 - \xi_2\|$ over all (ξ_i, η_i) such that $\eta_i \in \partial_\xi V(\tau, \xi_i)$, $i = 1, 2$. This shows (a).

A lower bound on $\|\eta_1 - \eta_2\|/\|\xi_1 - \xi_2\|$ and the relationship between strong convexity of a convex function and the Lipschitz continuity of the gradient of its conjugate [26, Proposition 12.60] yield (b); see also Example 4.2. □

THEOREM 3.6 (Lipschitz gradient). *Assume that H is differentiable and ∇H is Lipschitz continuous with constant K.*

(a) *Suppose that g is differentiable and ∇g is Lipschitz with constant γ_0. Then $V(\tau, \cdot)$ is differentiable for all $\tau < T$, and there exists a continuous function $\gamma : (-\infty, T] \mapsto \mathbb{R}$ with $\gamma(T) = \gamma_0$ such that $\nabla_\xi V(\tau, \cdot)$ is Lipschitz with constant $\gamma(\tau)$.*

(b) *Suppose that g is strongly convex with constant δ_0. Then there exists a continuous (and positive) function $\delta : (-\infty, T] \mapsto \mathbb{R}$ with $\delta(T) = \delta_0$ such that for all $\tau < T$, $V(\tau, \cdot)$ is strongly convex with constant $\delta(\tau)$.*

In fact, one can choose $\gamma(\tau) = \frac{c^2-1}{2c}$ with $c = (\gamma_0 + \sqrt{1 + \gamma_0^2})e^{2K(T-\tau)}$, and $\delta(\tau) = \frac{2d}{d^2-1}$ with $d = \left(\delta_0^{-1} + \sqrt{1 + \delta_0^{-2}}\right)e^{2K(T-\tau)}$. In particular, $\gamma(\tau) \leq (\gamma_0 + \frac{1}{2})e^{2K(T-\tau)}$ and $\delta(\tau) \geq \frac{2\delta_0}{2+\delta_0}e^{-2K(T-\tau)}$.

Proof. The gradient of a differentiable convex function f is Lipschitz continuous with constant a if and only if, for all x, x',

$$f(x') \leq f(x) + \nabla f(x) \cdot (x' - x) + \frac{1}{2}a\|x' - x\|^2; \tag{14}$$

see Proposition 12.60 in [26]. If g is as assumed in (a), we have for any x, x', $g(x') \leq \overline{g}^x(x')$, where $\overline{g}^x(x') = g(x) + \nabla g(x) \cdot (x' - x) + \frac{1}{2}\gamma_0\|x' - x\|^2$. Then for any $\tau \leq T$, $V(\tau, \xi) \leq \overline{V}^x(\tau, \xi)$, where $\overline{V}^x(\tau, \cdot)$ is the value function corresponding to the initial cost g_x. The latter value function is differentiable, as shown in Lemma

3.5. Also, $V(\tau,\xi_x) = \overline{V}^x(\tau,\xi_x)$, where ξ_x is the first coordinate of the initial point of the Hamiltonian trajectory on $[\tau,T]$ terminating at $(x,-\nabla g(x))$; this follows from Theorem 3.3 and from the optimality of the first arc constituting the mentioned Hamiltonian trajectory in the definition of both value functions. Consequently, $V(\tau,\cdot)$ is also differentiable at ξ_x, and $\nabla_\xi V(\tau,\xi_x) = \nabla_\xi \overline{V}^x(\tau,\xi_x)$. Now, Lemma 3.5 implies that the gradient of $\overline{V}^x(\tau,\cdot)$ is Lipschitz continuous with constant γ' as described in (a) of the lemma. Combining this, the inequality (14), and the comparison between $V(\tau,\cdot)$ and $\overline{V}^x(\tau,\cdot)$ yields

$$V(\tau,\xi) \le V(\tau,\xi_x) + \nabla_\xi V(\tau,\xi_x) + \frac{1}{2}\gamma'\|\xi-\xi_x\|^2.$$

In light of Theorem 3.3 this bound holds for any ξ, ξ', and thus $\nabla_\xi V(\tau,\cdot)$ is Lipschitz continuous with constant γ'.

The Optimality Principle and time-invariance of the Hamiltonian allow us to derive, through arguments similar to those above, a Lipschitz constant for $\nabla_\xi V(\tau',\cdot)$ given a constant for $\nabla_\xi V(\tau,\cdot)$, with $\tau' < \tau$. Let $\gamma(t)$ denote the (smallest possible) Lipschitz constant for $\nabla_\xi V(t,\cdot)$. Then $\gamma(\tau') \le \gamma(\tau)[1+(e^{K(\tau-\tau')}-1)\sqrt{1+\gamma(\tau)^{-2}}]^2$ whenever $\gamma(\tau) > 0$; a similar bound can be obtained for the case of $\gamma(\tau)=0$ for small values of $\tau-\tau'$ (by estimating $\|a(\tau')\|$, $\|b(\tau')\|$ from the proof of Lemma 3.5 as in (13)). Consequently, we can show that

$$\liminf_{\tau'\to\tau} \frac{\gamma(\tau')-\gamma(\tau)}{\tau'-\tau} \ge -2K\sqrt{1+\gamma^2(\tau)}.$$

Thus $\gamma(\tau) \le \phi(\tau)$, where ϕ is the solution of $\phi'(t) = -2K\sqrt{1+\phi^2(t)}$, $\phi(T) = \gamma_0$. This yields the bound at the end of Theorem 3.6 and proves (a).

A direct proof of (b) is symmetrical to the one just presented for (a), and an alternate approach is explained in Example 4.2. ∎

The factor 2 in the exponent in formulas for c and d at the end of Theorem 3.6 is not surprising. Consider $H(x,y) = x\cdot y$ corresponding to $L(x,v) = \delta_x(v)$. Then for any g, $V(\tau,\xi) = g(e^{T-\tau}\xi)$ and the Lipschitz constant for $\nabla V(\tau,\cdot)$ is $e^{2(T-\tau)}$ times that of ∇g.

Under the assumptions of Theorem 3.6 (a), an arc $x(\cdot)$ is optimal for $\mathcal{P}(\tau,\xi)$ in (1) if and only if

$$x(\tau)=\xi, \quad \dot{x}(t) = \nabla_y H(x(t), -\nabla_\xi V(t,x(t))) \text{ for almost all } t\in[\tau,T]. \tag{15}$$

The properties of the optimal feedback mapping $\Phi : (-\infty,T]\times\mathbb{R}^n$, defined by $\Phi(t,x) = \nabla_y H(x,-\nabla_\xi V(t,x))$, are summarized below. Continuity of ϕ in both variables follows from that of $\nabla_\xi V$, which in turn is a consequence of graphical continuity of $\nabla_\xi V(t,\cdot)$ in t, as stated in [27, Corollary 2.2]; details were worked out in Goebel [14].

COROLLARY 3.7 (Lipschitz optimal feedback). *Suppose that H and g are differentiable and their gradients are Lipschitz continuous. Then the optimal feedback mapping Φ is continuous on $(-\infty,T]\times\mathbb{R}^n$, and there exists a continuous function $\mu : (-\infty,T]\mapsto\mathbb{R}$ such that for all $t\le T$, $\Phi(t,\cdot)$ is Lipschitz continuous with constant $\mu(t)$.*

If the problem of Bolza $\mathcal{P}(\tau,\xi)$ represents a control problem $\mathcal{C}(\tau,\xi)$ in (3) via the transformation (11), an optimal control minimizes the right-hand side in (11). This translates (15) to necessary and sufficient optimality conditions for $\mathcal{C}(\tau,\xi)$ (general case with no smoothness present was discussed in [14]), $x(\tau) = \xi$, $\dot{x}(t) = Ax(t)+Bu(t)$, and

$$u(t) = \nabla_1 J^*(-B^*\nabla_\xi V(t,x(t)), Cx(t)) \quad \text{for almost all } t \in [\tau, T]. \tag{16}$$

Under the assumptions of Theorem 2.4, conclusions similar to those in Corollary 3.7 can be made about $\phi(t,x) = \nabla_1 J^*(-B^*\nabla_\xi V(t,x), Cx)$. In particular, optimal controls turn out to be continuous. To finish this section, we calculate ϕ for some of the examples of section 2.

We will need some properties of $\rho_{V,Q}$ defined in (4) (recall that Q is positive semidefinite and V is polyhedral). The function $\rho_{V,Q}$ is proper, convex, and piecewise linear-quadratic; $\operatorname{dom}\rho_{V,Q} = (V^\infty \cap \ker Q)^*$ and, in particular, $\rho_{V,Q}$ is finite-valued if and only if $V^\infty \cap \ker Q = \{0\}$ (Theorem 2.2 generalizes this fact). If this condition holds, then

$$\partial\rho_{V,Q}(s) = \operatorname*{argmax}_{v\in V}\left\{s\cdot v - \frac{1}{2}v\cdot Qv\right\} = \{v \mid s - Qv \in N_V(v)\} = (Q+N_V)^{-1}(s),$$

where $N_V(v)$ is the normal cone to the set V at v. For details, see Example 11.18 in Rockafellar and Wets [26]. If Q is actually positive definite, and thus invertible, we have, with $\operatorname{proj}_{\sqrt{V}Q}$ being the projection onto $\sqrt{Q}V$,

$$\partial\rho_{Q,V}(s) = \left(\sqrt{Q}\right)^{-1}\operatorname{proj}_{\sqrt{Q}V}\left(\left(\sqrt{Q}\right)^{-1}s\right).$$

Indeed for any convex set C, $(\operatorname{proj}_C)^{-1} = I + N_C$. Then

$$\left[\left(\sqrt{Q}\right)^{-1}\operatorname{proj}_{\sqrt{Q}V}\left(\sqrt{Q}\right)^{-1}\right]^{-1} = \sqrt{Q}\left(\operatorname{proj}_{\sqrt{Q}V}\right)^{-1}\sqrt{Q} = \sqrt{Q}\left(I + N_{\sqrt{Q}V}\right)\sqrt{Q}.$$

The last expression equals $Q+N_V$. It follows from the fact that $\sqrt{Q}N_{\sqrt{Q}V}\sqrt{Q} = N_V$, and this can be deduced from the properties of the normal cone under a change of coordinates.

Example 3.8 (optimal controls in feedback form). The linear-quadratic regulator (10) with a constraint $u(t) \in U$ (Example 2.5) has the following feedback mapping:

$$\phi(t,x) = (F+N_U)^{-1}\left(-B^*\nabla_\xi V(t,x)\right) = \sqrt{F}^{-1}\operatorname{proj}_{\sqrt{F}U}\left(-\sqrt{F}^{-1}B^*\nabla_\xi V(t,x)\right).$$

A similar formula was obtained by Heemels, Van Eijndhoven, and Stoorvogel [16] for a conical U; our regularity results are also stronger than those therein.

Example 2.6 discussed (10) with a constraint $u(t) \leq C_0x(t) - q_0$. With the matrices as defined in the mentioned example, we obtain

$$\begin{aligned}
J^*(a,b) &= \sup_{u\in U}\inf_{v\in V}\left\{a\cdot u + b\cdot v - \frac{1}{2}u\cdot Pu + \frac{1}{2}v\cdot Qv + v\cdot Du\right\}\\
&= \inf_{v\in V}\left\{b\cdot v + \frac{1}{2}v\cdot Qv + \sup_{u\in\mathbb{R}^k}\left\{(a+D^*v)\cdot u - \frac{1}{2}u\cdot Pu\right\}\right\}\\
&= \inf_{v\in V}\left\{b\cdot v + \frac{1}{2}v\cdot Qv + \frac{1}{2}(a+D^*v)\cdot P^{-1}(a+D^*v)\right\}\\
&= \frac{1}{2}a\cdot P^{-1}a - \sup_{v\in V}\left\{(-b-DP^{-1}a)\cdot v - \frac{1}{2}v\cdot(Q+DP^{-1}D^*)v\right\}.
\end{aligned}$$

The matrix $Q + DP^{-1}D^*$ equals $\left[\begin{smallmatrix} I_{n\times n} & 0_{n\times k} \\ 0_{k\times n} & F^{-1} \end{smallmatrix}\right]$, and thus the sup expression above is separable. Also, $DP^{-1} = \left[\begin{smallmatrix} 0_{k\times k} \\ -F^{-1} \end{smallmatrix}\right]$, so $(-b - DP^{-1}a)_1 = -b_1$, $(-b - DP^{-1}a)_2 = -b_2 - F^{-1}a$. Then $J^*(a,b)$ equals

$$\begin{aligned}\frac{1}{2}a\cdot F^{-1}a - \frac{1}{2}\left\|(-b - DF^{-1}a)_1\right\|^2 - \sup_{v_2\in\mathbb{R}^k_+}\left\{(-b - DF^{-1}a)_2\cdot v_2 - v_2\cdot F^{-1}v_2\right\}\\ = \frac{1}{2}a\cdot F^{-1}a - \frac{1}{2}\left\|b_1\right\|^2 - \rho_{\mathbb{R}^k_-,F^{-1}}\left(-b_2 - F^{-1}a\right)\end{aligned}$$

and thus $\nabla_1 J^*(a,b) = F^{-1}\left[a + (N_{\mathbb{R}^k_-} + F^{-1})^{-1}(-b_2 - F^{-1}a)\right]$. Since for $b = Cx$ we have $b_1 = \sqrt{E}x$, $b_2 = C_2 x$, the optimal feedback map is

$$\phi(t,x) = -F^{-1}\left[B^*\nabla_\xi V(t,x) - \left(N_{\mathbb{R}^k_-} + F^{-1}\right)^{-1}\left(F^{-1}B^*\nabla_\xi V(t,x) - C_2 x\right)\right].$$

4. Convergence and approximation of value functions. In this section we study the convergence of value functions defined by sequences of converging costs $\{g_i\}$ and $\{L_i\}$,

$$V_i(\tau,\xi) = \inf\left\{g_i(x(0)) + \int_0^\tau L_i(x(t),\dot{x}(t))\,dt \;\middle|\; x(\tau) = \xi\right\}, \tag{17}$$

to $V(\tau,\xi)$ defined in (1). To treat sequences of possibly infinite-valued functions we use the well appreciated in optimization notion of epi-convergence. A sequence of functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1,2,\ldots$, is said to *epi-converge* to f ($\text{e-lim}_i f_i = f$ for short) if, for every point $x \in \mathbb{R}^n$,

(a) $\liminf_i f_i(x_i) \geq f(x)$ for every sequence $x_i \to x$,

(b) $\limsup_i f_i(x_i) \leq f(x)$ for some sequence $x_i \to x$.

For details, consult Rockafellar and Wets [26, Chapter 7]. We will only need to directly show the "lower" part of epi-convergence of value functions and rely on duality results to complete the argument. Let us briefly introduce the needed concepts.

For a function $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ its convex conjugate is defined by

$$f^*(y) = \sup_{x\in\mathbb{R}^n}\{y\cdot x - f(x)\}.$$

If f is proper, l.s.c., and convex, then so is f^*, and the conjugate of f^* equals f (that is, $f(x) = \sup_{y\in\mathbb{R}^n}\{x\cdot y - f^*(y)\}$). For details, consult Rockafellar [19, section 12]. Relations of certain properties of f to some other properties of f^*, say of coercivity and finiteness, were alluded to in the previous sections; in Example 4.2 we discuss the symmetry between strong convexity of f and Lipschitz continuity of ∇f^*, and revisit Lemma 3.5 and Theorem 3.6. Epi-convergence of a sequence of convex function is equivalent to that of the sequence of conjugates; we will need the following related facts. Below, $\text{e-lim\,inf}_i f_i \geq f$ means that condition (a) in the definition of epi-convergence holds. A sequence $\{f_i\}$ is said to escape epigraphically to the horizon if the epigraphical limit of f_i is equal to $+\infty$ everywhere.

LEMMA 4.1. *Suppose that functions $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and $f_i : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $i = 1,2,\ldots$, are proper, l.s.c., and convex.*

(a) *If $\text{e-lim\,inf}_i f_i \geq f$ and $\text{e-lim\,inf}_i f_i^* \geq f^*$ and neither sequence escapes epigraphically to the horizon, then $\text{e-lim}_i f_i = f$ and $\text{e-lim}_i f_i^* = f^*$.*

(b) *Neither of the sequences f_i, f_i^* escapes epigraphically to the horizon provided there exists a constant $\rho > 0$ such that $f_i(x) \geq -\rho(\|x\|+1)$ and $f_i^*(x) \geq -\rho(\|x\|+1)$ for all x and $i = 1, 2, \ldots$.*

Proof. Statement (a) essentially follows from the statement and proof of Theorem 11.34 in [26]. We show (b). An application of a separation principle (for example, Theorem 11.3 in [19]) implies that for every $i = 1, 2, \ldots$, there exist $\alpha_i \in \mathbb{R}^n$, $\beta_i \in \mathbb{R}$ such that $f_i(x) \geq \alpha_i \cdot x + \beta_i \geq -\rho(\|x\|+1)$ for every $x \in \mathbb{R}^n$. It must be the case that $\|\alpha_i\| \leq \rho$ while $\beta_i \geq -\rho$. We then obtain

$$f_i^*(\alpha_i) = \sup_x \{\alpha_i \cdot x - f(x)\} \leq \sup_x \{\alpha_i \cdot x - \alpha_i \cdot x - \beta_i\} = -\beta_i \leq \rho.$$

Thus $f_i^*(\alpha_i) \leq \rho$ while by assumption, $f_i^*(\alpha_i) \geq -\rho(\|\alpha_i\|+1)$. As $\|\alpha_i\| \leq \rho$, there exists a convergent subsequence of $(\alpha_i, f_i^*(\alpha_i))$, and, consequently, the sequence f_i^* cannot escape to the horizon. A symmetric argument shows the corresponding fact for the sequence f_i. □

For a given initial cost g and Lagrangian L, the *dual value function* $\widetilde{V} : (-\infty, T] \times \mathbb{R}^n \mapsto \mathbb{R}$ is defined in a fashion similar to V,

$$\widetilde{V}(\tau, \eta) = \inf \left\{ g^*(y(0)) + \int_0^\tau \widetilde{L}(y(t), \dot{y}(t))\, dt \;\middle|\; y(\tau) = \eta \right\}, \tag{18}$$

where the dual Lagrangian is

$$\widetilde{L}(y, w) = L^*(w, y) = \sup_{(x,v)\in\mathbb{R}^{2n}} \{w \cdot x + y \cdot v - L(x, v)\}. \tag{19}$$

If L satisfies Assumption 3.1, then so does $\widetilde{L}$ (and consequently $\widetilde{V}(\tau, \cdot)$ is proper, l.s.c., and convex for every $\tau \leq T$), and in fact for any $\tau \leq T$, the functions $V(\tau, \cdot)$ and $\widetilde{V}(\tau, \cdot)$ are conjugate to each other:

$$\widetilde{V}(\tau, \eta) = \sup_{\xi \in \mathbb{R}^n} \big\{\eta \cdot \xi - V(\tau, \xi)\big\}, \qquad V(\tau, \xi) = \sup_{\eta \in \mathbb{R}^n} \big\{\xi \cdot \eta - \widetilde{V}(\tau, \eta)\big\}. \tag{20}$$

These results were shown by Rockafellar and Wolenski [27]. The Hamiltonian $\widetilde{H}$ associated with a dual Lagrangian $\widetilde{L}$ is exactly $\widetilde{H}(y, x) = -H(x, y)$, and thus it has the same smoothness properties as H. Note also that the Lagrangian dual to $\widetilde{L}$ is the original L.

Example 4.2 (strong convexity and Lipschitz differentiability). A convex function f is differentiable and ∇f is Lipschitz continuous with constant σ if and only if f^* is strongly convex with constant $1/\sigma$. This and (20) automatically proves one of the statements (a), (b) of Theorem 3.6 once the other is in place and similarly for Lemma 3.5. For example, we show (b) of 3.6 with the help of (a). Suppose g is strongly convex with constant δ_0, and ∇H is Lipschitz with constant K. Then ∇g^* is Lipschitz with constant $\gamma_0 = 1/\delta_0$, while the dual Hamiltonian $\widetilde{H}$ also has a Lipschitz gradient, with constant K. Application of (a) shows that the dual value function $\widetilde{V}(\tau, \cdot)$ is differentiable, with $\nabla_\eta \widetilde{V}(\tau, \cdot)$ Lipschitz with constant $\gamma(\tau)$ as described at the end of Theorem 3.6. Now (20) implies that $V(\tau, \cdot)$ is strongly convex, with constant $\delta(\tau) = 1/\gamma(\tau)$. This yields the expression for $\delta(\tau)$ as described in the other formula at the end of Theorem 3.6. The lower bound on $\delta(\tau)$ can be obtained in a similar fashion from the upper bound on $\gamma(\tau)$.

We now focus on sequences of Bolza problems. Given a sequence of Lagrangians $\{L_i\}$, for each i we let $\widetilde{L}_i$ be the Lagrangian dual to L_i as described in (19), and H_i be the Hamiltonian corresponding to L_i as suggested by (2). The value function V_i is defined by (17), while $\widetilde{V}_i$ is defined similarly in terms of g_i^* and $\widetilde{L}_i$.

Assumption 4.3 (uniform growth assumption). Each of the functions g_i and L_i, $i = 1, 2, \ldots$, is proper, l.s.c., and convex. There exist functions $\underline{L}$ and $\overline{L}$, each satisfying Assumption 3.1, such that, for every $i = 1, 2, \ldots$,

$$\underline{L} \leq L_i \leq \overline{L}.$$

As L satisfies Assumption 3.1 if and only if $\widetilde{L}$ does, the second condition above is equivalent to the existence of $\underline{M}$ and $\overline{M}$, each satisfying Assumption 3.1, such that $\underline{M} \leq \widetilde{L}_i \leq \overline{M}$; take $\underline{M}$ to be the Lagrangian dual to $\overline{L}$, $\overline{M}$ dual to $\underline{L}$.

LEMMA 4.4 (convergence equivalence). *If Assumption* 4.3 *holds, the following statements are equivalent:*

(a) *Lagrangians L_i epi-converge to L,*

(b) *dual Lagrangians $\widetilde{L}_i$ epi-converge to $\widetilde{L}$,*

(c) *Hamiltonians H_i converge pointwise to H.*

The proof is postponed until section 5. Also there we discuss the convergence of Lagrangians (11) and Hamiltonians (6) corresponding to extended piecewise linear-quadratic functions under perturbations of all defining data; see Theorem 5.6.

Assumption 4.5 (epi-convergence of cost functions). Sequences $\{g_i\}$, $\{L_i\}$ epi-converge, respectively, to g and L.

Equivalently, we could assume that sequences $\{g_i^*\}$ and $\{\widetilde{L}_i\}$ epi-converge, respectively, to g^* and $\widetilde{L}$. We are now ready to state the main result of this section.

THEOREM 4.6 (value function epi-convergence). *Let Assumptions* 4.3 *and* 4.5 *hold. For any $\tau \leq T$ and a sequence $\tau_i \to \tau$ (in particular for $\tau_i = \tau$) we have*

$$\operatorname{e-lim} V_i(\tau_i, \cdot) = V(\tau, \cdot). \tag{21}$$

Equivalently, $\operatorname{e-lim} \widetilde{V}_i(\tau_i, \cdot) = \widetilde{V}(\tau, \cdot)$. *This implies* $\operatorname{e-lim} V_i = V$ *and* $\operatorname{e-lim} \widetilde{V}_i = \widetilde{V}$.

We prove the theorem by taking advantage of the representation

$$V(\tau, \xi) = \inf_{\xi' \in \mathbb{R}^n} \left\{ E(\tau, \xi, \xi') + g(\xi') \right\}, \tag{22}$$

where the *fundamental kernel* $E : (-\infty, T] \times \mathbb{R}^n \times \mathbb{R}^n$ is given by

$$E(\tau, \xi, \xi') = \inf \left\{ \int_\tau^T L(x(t), \dot{x}(t))\, dt \mid x(\tau) = \xi,\ x(T) = \xi' \right\},$$

with the infimum taken over all arcs with prescribed endpoints. A symmetric representation of $\widetilde{V}(\tau, \eta)$ is available, with $\widetilde{E}(\tau, \eta, \eta')$ defined in terms of $\widetilde{L}$. The following conjugacy relationship is a direct consequence of (20); to see this, consider $E(\cdot, \cdot, \xi')$ as the value function associated with the terminal cost $g(x) = \delta_{\xi'}(x)$ and use the fact that $g^*(y) = \xi' \cdot y_0$:

$$\widetilde{E}(\tau, \eta, \eta') = \sup_{\xi, \xi'} \left\{ \eta \cdot \xi - \eta' \cdot \xi' - E(\tau, \xi, \xi') \right\},$$
$$E(\tau, \xi, \xi') = \sup_{\eta, \eta'} \left\{ \xi \cdot \eta - \xi' \cdot \eta' - \widetilde{E}(\tau, \eta, \eta') \right\}.$$

We will need some facts about continuity and convergence of integral functionals. It is known that for a fixed $\tau > 0$, the functional $\Phi(\tau, \cdot)$ defined on the space of absolutely continuous arcs on $[0, \tau]$ by $\Phi(\tau, z(\cdot)) = \int_0^\tau L(z(t), \dot{z}(t))dt$ is weakly sequentially lower semicontinuous. This can be shown as a consequence of the conjugacy between L and H, and by interchanging the integration and maximization,

$$(23)\quad \int_0^\tau \sup_w \{w \cdot \dot{z}(t) - H(z(t), w)\}\, dt = \sup_{w(\cdot)} \int_0^\tau \{w(t) \cdot \dot{z}(t) - H(z(t), w(t))\}\, dt,$$

where the latter supremum is taken over all arcs w in $L^\infty[0, T]$ (see, for example, [26, Theorem 14.60]). Now consider a sequence of functionals $\Phi_i(\tau, z(\cdot)) = \int_0^\tau L_i(z(t), \dot{z}(t))dt$ a sequence of arcs x_i on $[0, \tau]$ weakly convergent to an arc x (meaning that $\dot{x}_i$ converge weakly to $\dot{x}$ in L^1 and $x_i(0)$ converge to $x(0)$). Then

$$(24)\qquad \liminf_i \Phi_i(\tau, x_i(\cdot)) \geq \Phi(\tau, x(\cdot)).$$

We only need to consider the case where $\liminf \Phi_i(\tau, x_i(\cdot)) < +\infty$. As in (23), we have, for any w in $L^\infty[0, T]$, $\Phi_i(\tau, x_i(\cdot)) \geq \int_0^\tau \{w(t) \cdot \dot{x}_i(t) - H_i(x_i(t), w(t))\}\, dt$. Then, as $\dot{x}_i(\cdot)$ converge weakly in L^1 to $\dot{x}(\cdot)$, $x_i(\cdot)$ converge pointwise to $x(\cdot)$, and H_i converge to H pointwise and also uniformly on compact sets (Lemma 5.4), we get $\liminf \Phi_i(\tau, x_i(\cdot)) \geq \int_0^\tau \{w(t) \cdot \dot{x}(t) - H(x(t), w(t))\}\, dt$, and this holds for any w in $L^\infty[0, T]$. By (23), we conclude (24). In the proof of Lemma 4.7 we extend these arguments to varying time intervals.

LEMMA 4.7 (fundamental kernel epi-convergence). *Let E_i and $\widetilde{E}_i$ be the fundamental kernels associated, respectively, with L_i and $\widetilde{L}_i$. Under assumptions of Theorem* 4.6, *for any $\tau < T$ and a sequence $\tau_i \to \tau$ (in particular for $\tau_i = \tau$) we have*

$$(25)\qquad \operatorname{e-lim} E_i(\tau_i, \cdot, \cdot) = E(\tau, \cdot, \cdot).$$

Equivalently, $\operatorname{e-lim} \widetilde{E}_i(\tau_i, \cdot, \cdot) = \widetilde{E}(\tau, \cdot, \cdot)$. *Consequently,* $\operatorname{e-lim} E_i = E$ *and* $\operatorname{e-lim} \widetilde{E}_i = \widetilde{E}$.

Proof. Fix $\tau < T$ and $\tau_i \to \tau$. First, we show that $\operatorname{e-lim\,inf}_i E_i(\tau_i, \cdot, \cdot) \geq E(\tau, \cdot, \cdot)$, that is, for any point $(\xi, \xi') \in \mathbb{R}^{2n}$ and a sequence $(\xi_i, \xi_i') \to (\tau, \xi, \xi')$, we have

$$(26)\qquad \liminf_{i \to \infty} E_i(\tau_i, \xi_i, \xi_i') \geq E(\tau, \xi, \xi').$$

We only need to consider the case where $\liminf_{i \to \infty} E_i(\tau_i, \xi_i, \xi_i') = m < +\infty$, and if necessary we pass to a subsequence so that $E_i(\tau_i, \xi_i, \xi_i') \to m$. There exist arcs x_i on $[\tau_i, T]$ such that $E_i(\tau_i, \xi_i, \xi_i') = \Phi_i(\tau_i, x_i(\cdot)) = \int_0^{\tau_i} L_i(x_i(t), \dot{x}_i(t))dt$. Setting $a_i = (T - \tau_i)/(T - \tau)$ and defining $x_i^0(\tau + s) = x_i(\tau_i + a_i s)$, $L_i^0(x, v) = a_i L_i(x, v/a_i)$ leads to

$$\Phi_i(\tau_i, x_i(\cdot)) = \int_{\tau_i}^T L_i(x_i(t), \dot{x}_i(t))dt = \int_\tau^T L_i^0(x_i^0(t), \dot{x}_i^0(t))dt = \Phi_i^0(\tau, x_i^0(\cdot)),$$

with L_i^0 epiconverging to L [26, Exercise 7.47]. Corresponding Hamiltonians are $H_i^0(x, y) = a_i H(x, y)$, while the dual Lagrangians are $\tilde{L}_i^0(x, v) = a_i \widetilde{L}_i(x, v/a_i)$. As $\{L_i\}$ satisfies Assumption 4.3, so does $\{L_i^0\}$; this is a direct calculation. Consequent uniform growth assumptions imply in particular that some subsequence of rescaled arcs x_i^0 on $[0, \tau]$ weakly converges to an arc x on $[0, \tau]$ with $x(\tau) = \xi$,

$x(T) = \xi'$ (this follows from Theorem 1 in [21]). Moreover, as in (24), we have $\lim_i \Phi_i(\tau_i, x_i(\cdot)) = \lim_i \Phi_i^0(\tau, x_i^0(\cdot)) \geq \Phi(\tau, x(\cdot))$. But the arc x is feasible for the problem defining $E(\tau, \xi, \xi')$, and (26) follows.

The same argument applied to dual problems gives e-lim inf $\widetilde{E}_i(\tau_i, \cdot, \cdot) \geq \widetilde{E}(\tau, \cdot, \cdot)$. Lemma 4.1 (a) will conclude (25) (and the equivalent dual statement) if we show that neither $E_i(\tau_i, \cdot, \cdot)$ nor $\widetilde{E}_i(\tau_i, \cdot, \cdot)$ escapes to the horizon. Uniform growth in Assumption 4.3 and the rescaling arguments above imply that $\{E_i(\tau_i, \cdot, \cdot)\}$ is uniformly bounded below by $\widehat{E}(\tau, \cdot, \cdot)$, a fundamental function corresponding to some Lagrangian satisfying Assumption 3.1. As the latter function is proper and convex, it is bounded below by an affine function. A similar bound is in place for $\widetilde{E}_i(\tau_i, \cdot, \cdot)$, and thus the desired conclusions hold.

Lastly, the very definition of epi-convergence explains that (25) implies e-lim $E_i = E$. □

Proof (Theorem 4.6). As in Lemma 4.7, we begin by showing that for any $(\tau, \xi) \in (0, +\infty) \times \mathbb{R}^n$ and a sequence $(\tau_i, \xi_i) \to (\tau, \xi)$, we have

$$\liminf_i V_i(\tau_i, \xi_i) \geq V(\tau, \xi). \tag{27}$$

It suffices to consider, passing to a subsequence if necessary, the case of $\lim_i V_i(\tau_i, \xi_i) < +\infty$. Recall (22). Functions g_i epi-converge to g by assumption, while Lemma 4.7 and the definition of epi-convergence yield e-lim $\inf_i E_i(\tau_i, \xi_i, \cdot) \geq E(\tau, \xi, \cdot)$. Now by Theorem 7.46 of [26], we obtain

$$\text{e-}\liminf_i \{E_i(\tau_i, \xi_i, \cdot) + g_i(\cdot)\} \geq E(\tau, \xi, \cdot) + g(\cdot). \tag{28}$$

As mentioned in the proof of Lemma 4.7, $\{E_i(\tau_i, \cdot, \cdot)\}$ is uniformly bounded below by $\widehat{E}(\tau, \cdot, \cdot)$, a fundamental kernel corresponding to some Lagrangian satisfying Assumption 3.1. Proposition 4.2 in [27] implies that

$$E_i(\tau_i, \xi_i, \xi') \geq \widehat{E}(\tau, \xi_i, \xi') \geq \theta\left(\max\{0, |\xi'| - \alpha|\xi_i|\}\right) - \beta|\xi_i| \tag{29}$$

for a proper, nondecreasing, and coercive $\theta : [0, +\infty) \mapsto \mathbb{R}$ and constants α, β. As ξ_i converge, there exist a, b such that $\widehat{E}(\tau_i, \xi, \xi') \geq \theta(\max\{0, |\xi'| - a\}) - b$. A similar bound is in place for $E(\tau, \cdot, \cdot)$, and consequently, $E_i(\tau_i, \xi_i, \cdot)$ and $E(\tau, \xi, \cdot)$ are bounded below by a coercive function. Convexity and epi-convergence of g_i to g implies, by 7.34 in [26], that g_i and g are bounded below (uniformly in i) by $-\rho(|\cdot|+1)$, for some constant ρ. As $\inf_{\xi'}\{E_i(\tau_i, \xi_i, \xi') + g_i(\xi')\}$ converge to a finite value, there exists a compact set S such that

$$\inf_{\xi'} \{E_i(\tau_i, \xi', \xi_i) + g_i(\xi')\} = \inf_{\xi' \in S} \{f_i(\xi') + E_i(\tau_i, \xi', \xi_i)\},$$

and a similar condition holds for $E(\tau, \xi, \xi') + g(\xi')$. Consequently, infimum in (22) can be taken over S, similarly for $V_i(\tau_i, \cdot)$. Now (28) and Proposition 7.29 in [26] yield (27).

Growth conditions $g_i(\xi') \geq -\rho(|\xi'| + 1)$, (29), and the fact that since θ in (29) is coercive, there exists $\gamma > 0$ such that $\theta \geq \rho|\cdot| - \gamma$ imply that

$$\begin{aligned} V_i(\tau_i, \xi) &\geq \inf_{\xi'} \{-\rho(|\xi'| + 1) + \rho \max\{0, |\xi'| - \alpha|\xi|\} - \gamma - \beta|\xi|\} \\ &\geq -(\alpha\rho + \beta)|\xi| - (\rho + \gamma). \end{aligned}$$

A similar bound holds for $\widetilde{V}_i(\tau_i,\cdot)$. This and (27) show the desired epi-convergence of $V_i(\tau_i,\cdot)$ as well as $\widetilde{V}_i(\tau_i,\cdot)$, by Lemma 4.1 (b). Epi-convergence of V_i and $\widetilde{V}_i$ follows directly from the definition of epi-convergence. ∎

We now describe how any problems fitting the general Assumption 3.1 can be approximated by problems with value functions possessing regularity as discussed in section 3. We will rely on Moreau–Yosida envelopes of convex and saddle functions. For any proper, l.s.c., and convex f and $\lambda > 0$, $e_\lambda f(x) = \inf_q \left\{ f(q) + \frac{1}{2\lambda}\|x-q\|^2 \right\}$ is finite and differentiable; see [26, Theorem 2.26]. A generalization of this smoothing technique to saddle functions was introduced by Attouch and Wets [2]. Applied to a concave-convex Hamiltonian H (and simplified to single parameter λ vs. the original two), it yields a differentiable concave-convex function

$$e_\lambda H(x,y) = \sup_p \inf_q \left\{ H(p,q) - \frac{1}{2\lambda}\|x-p\|^2 + \frac{1}{2\lambda}\|y-q\|^2 \right\}. \tag{30}$$

(We use the same notation for Moreau–Yosida regularization of convex and saddle functions; it should be clear which one is considered.) The key fact is that $\nabla e_\lambda f$ and $\nabla e_\lambda H$ are globally Lipschitz with constant $1/\lambda$. This is the case since $\nabla e_\lambda f$ is the Yosida regularization of the monotone subdifferential ∂f (see Exercise 12.23 in [26]), while $(x,y) \mapsto (-\nabla_x e_\lambda H(x,y), \nabla_y e_\lambda H(x,y))$ is the Yosida regularization of the monotone mapping $(x,y) \mapsto (-\widetilde{\partial}_x H(x,y), \partial_y H(x,y))$.

Corollary 4.8 (regularization of value functions). *Let L be any Lagrangian satisfying Assumption 3.1; $\widetilde{L}$ be the associated dual Lagrangian; g be any proper, l.s.c., and convex function; and V, $\widetilde{V}$ be the associated value functions. There exists a sequence of finite convex functions g_i and a sequence of Lagrangians L_i satisfying Assumption 4.3 such that the following hold.*

(a) *Conclusions of Theorem 4.6 hold for sequences $\{V_i\}$, $\{\widetilde{V}_i\}$ of value functions corresponding, respectively, to L_i, g_i and their dual costs g_i^*, $\widetilde{L}_i$.*

(b) *For each i, V_i and $\widetilde{V}_i$ are continuously differentiable, and there exist continuous and positive functions $\gamma_i : (-\infty, T] \mapsto \mathbb{R}$, $\delta : (-\infty, T] \mapsto \mathbb{R}$ such that*
   (i) *$\nabla_\xi V_i(\tau,\cdot)$ and $\nabla_\xi \widetilde{V}_i(\tau,\cdot)$ are Lipschitz with constant $\gamma(\tau)$,*
   (ii) *$V(\tau,\cdot)$ and $\widetilde{V}(\tau,\cdot)$ are strongly convex with constant $\delta(\tau)$.*

This can be achieved by considering (with H associated to L)

$$g_i(x) = e_{1/i}\, g(x) + \|x\|^2/i, \qquad H_i(x,y) = e_{1/i}\, H(x,y),$$

and letting L_i and $\widetilde{L}_i$ be the Lagrangians associated with H_i.

Proof. Condition $(A3)$ in Assumption 3.1 (by the proof of Theorem 2.3 in [27]) and the definition of H_i imply, respectively, that

$$H(x,y) \le \theta^*(\|y\|) + (\alpha\|y\| + \beta)\,\|x\|, \quad H_i(x,y) \le \sup_p \left\{ H(p,y) - \frac{1}{2\lambda}\|p-x\|^2 \right\}.$$

Combining the two inequalities yields

$$\begin{aligned} H_i(x,y) &\le \theta^*(\|y\|) + \sup_p \left\{ (\alpha\|y\| + \beta)\,\|p\| - \frac{1}{2\lambda}\|p-x\|^2 \right\} \\ &= \theta^*(\|y\|) + \frac{\lambda}{2}\,(\alpha\|y\| + \beta)^2 + (\alpha\|y\| + \beta)\,\|x\|. \end{aligned}$$

This in turn implies that $(A3)$ holds for L_i, with θ replaced by

$$\theta'(r) = \inf_{s\in[0,r]} \left\{ \theta^*(s) + \frac{\lambda}{2} (\alpha s + \beta)^2 \right\}.$$

Coercivity of both θ^* and the quadratic implies that of θ', which is obviously nondecreasing. A symmetric argument shows that $\widetilde{L}_i$ satisfies $(A3)$ uniformly, and consequently, Assumption 4.3 is satisfied. Moreau–Yosida approximations of H hypo/epi-converge to H, and as all these functions are finite, the convergence is pointwise (Lemma 5.4). Functions g_i epi-converge to g by Theorem 1.25 and Exercise 7.47 in [26]. This shows (a).

To see (b), note that g has a Lipschitz gradient (with constant i) as well as strongly convex (with constant $1/i$). Now, invoke Theorem 3.6 and the symmetry between strong convexity and Lipschitz continuity of the gradient of the dual as outlined in Example 4.2. ◻

Example 4.9 (regularization of control problems). Recall that the Hamiltonian (6) corresponding to an extended piecewise linear-quadratic problem (3) had the special structure $H(x, y) = y \cdot Ax + J^*(B^*y, Cx)$. The regularization, as described in Corollary 4.8, can be applied to such H, but a more explicit smoothing technique is available. One may regularize J^* directly, using the convex-concave counterpart of (30)—the infimum is to be taken over the first variable, supremum over the second. Such regularization, with parameter $1/i$ can be equivalently obtained by defining functions J_i^* in (7) with matrices P and Q replaced, respectively, by positive definite $P + I/i$, $Q + I/i$. (Here I denotes an identity matrix of appropriate size.)

5. Convex analysis tools. We say that a function $K : \mathbb{R}^k \times \mathbb{R}^l \mapsto [-\infty, +\infty]$ is convex-concave if, for any fixed $z \in \mathbb{R}^l$, the function $K(\cdot, z)$ is convex, while for any fixed $w \in \mathbb{R}^k$, $K(w, \cdot)$ is concave. We call a convex-concave function K proper if the effective domain of K, defined as

$$\operatorname{dom} K = \left\{ w \in \mathbb{R}^k \mid K(w, z) < +\infty \ \forall z \in \mathbb{R}^l \right\} \times \left\{ z \in \mathbb{R}^l \mid -\infty < K(w, z) \ \forall w \in \mathbb{R}^k \right\},$$

is nonempty.

Convex function duality gives a one-to-one correspondence between a proper lsc convex function and its conjugate (also proper and l.s.c.). Saddle function duality describes a one-to-one correspondence between *equivalence classes* of proper *closed* saddle functions. Closedness is a notion corresponding, in a sense, to lower semicontinuity of convex functions. For the somewhat technical definition, and the reasons for considering equivalence classes, see Rockafellar [19]. Here, we limit ourselves to the facts crucial to the developments in what follows.

Any equivalence class $[K]$ of closed saddle functions contains the lowest and the highest element, denoted $\underline{K}$ and $\overline{K}$, and consists of all closed saddle functions K such that $\underline{K} \le K \le \overline{K}$. If a saddle function K is finite, then it is closed, $K = \underline{K} = \overline{K}$, and the class $[K]$ of all closed functions equivalent to K is just $\{K\}$. A saddle function k, defined on $W \times Z$, for some nonempty closed convex sets $W \subset \mathbb{R}^K$, $Z \subset \mathbb{R}^L$, gives rise to an equivalence class $[K]$ of saddle functions on $\mathbb{R}^K \times \mathbb{R}^l$, whose lowest and highest elements, $\underline{K}$, $\overline{K}$, are given by

$$\underline{K}(w, z) = \begin{cases} k(w, z) & \text{for } w \in W, z \in Z, \\ -\infty & \text{for } z \notin Z, \\ +\infty & \text{for } w \notin W, z \in Z; \end{cases} \qquad \overline{K}(w, z) = \begin{cases} k(w, z) & \text{for } w \in W, z \in Z, \\ +\infty & \text{for } w \notin W, \\ -\infty & \text{for } w \in W, z \notin Z. \end{cases}$$

Equivalent saddle functions have the same effective domains, on the relative interior of which they are equal to each other (and finite).

For a given saddle function K, the lower conjugate $\underline{K^*}$ and the upper conjugate $\overline{K^*}$ are defined by

$$\underline{K^*}(a,b) = \sup_{u\in\mathbb{R}^k}\inf_{v\in\mathbb{R}^l}\{a\cdot u+b\cdot v-K(u,v)\}, \quad \overline{K^*}(a,b) = \inf_{v\in\mathbb{R}^l}\sup_{u\in\mathbb{R}^k}\{a\cdot u+b\cdot v-K(u,v)\}. \tag{31}$$

The lower and upper conjugate functions are equivalent to each other and are, respectively, the lowest and the highest elements of $[K^*]$, the class of saddle functions conjugate to K. In fact, $\underline{K^*}$, $\overline{K^*}$ do not depend on the choice of $K \in [K]$, so $[K^*]$ should be thought of as conjugate to $[K]$. The lower and upper conjugates of any $K^* \in [K^*]$ are, in turn, the lowest and highest elements of $[K]$.

Example 5.1 (Hamiltonian in terms of a conjugate function). Recall that the Hamiltonian (6) was expressed in terms of a function J^*, which can be viewed as a conjugate of J (a unique conjugate, if we request that J^* be finite), where

$$J(u,v) = p\cdot u + \frac{1}{2}u\cdot Pu + q\cdot v - \frac{1}{2}v\cdot Qv - v\cdot Du \quad \text{for } (u,v)\in U\times V, \tag{32}$$

and has appropriately assigned $\pm\infty$ values outside $U\times V$.

Subdifferentials of K^* are exactly the saddle points in the expressions in (31); see Rockafellar [19, Theorem 37.2]. In particular, as finite saddle functions have nonempty subdifferentials, Theorem 2.2 can be viewed as saying that J has a saddle point on $U\times V$ for any (p,q). In other words, the function J_0 below has a saddle point under any affine perturbation. Similarly, Theorem 2.4 states the Lipschitz continuity of saddle points of J^* under perturbations. From a numerical viewpoint, finding the gradients of the Hamiltonian (6) amounts to solving a quadratic minimax problem.

As the linear terms $p\cdot u$ and $q\cdot v$ in (32) do not influence the finiteness and differentiability of $J^*(\cdot,\cdot)$, in proofs of Theorems 2.2 and 2.4 we work with

$$J_0(u,v) = \frac{1}{2}u\cdot Pu - \frac{1}{2}v\cdot Qv - v\cdot Du.$$

(From (7), we get that $J^*(a,b) = J_0^*(a-p,b+q)$.) We will need the following technical lemma.

LEMMA 5.2. *Assume that sets W and Z in $\mathbb{R}^n$ are polyhedral. Then $W+Z=\mathbb{R}^n$ is equivalent to $W^\infty+Z^\infty=\mathbb{R}^n$. For a linear mapping L we have $(LW)^\infty = LW^\infty$.*

Proof. For a polyhedral set W we can conclude that $W\subset W^\infty+\epsilon_w\mathbb{B}$ for some $\epsilon>0$, this follows for example from Corollary 3.53 in [26]. Thus if $W+Z=\mathbb{R}^n$, then $W^\infty+Z^\infty+(\epsilon_w+\epsilon_z)\mathbb{B}=\mathbb{R}^n$. But since $W^\infty+Z^\infty$ is a cone, we must have $W^\infty+Z^\infty=\mathbb{R}^n$. Now assume the latter. We have $W^\infty\subset W-w$ for any $w\in W$. Similarly for Z. Then $W^\infty+Z^\infty\subset W+Z-(w+z)$, which shows that $W+Z=\mathbb{R}^n$. The fact about linear mappings follows directly from the representation of a polyhedral set in Corollary 3.53 in [26]. ∎

For a proper lsc and convex function f, finiteness of f^* is equivalent to coercivity of f. Generalization of this fact to saddle functions, shown by Goebel [13, Proposition 2.7], states that for a proper closed convex-concave function $K:\mathbb{R}^k\times\mathbb{R}^l\mapsto[-\infty,\infty]$, the following conditions are equivalent:

(a) The class $[K^*]$ of convex-concave functions conjugate to K consists of a unique finite-valued function.

(b) The convex function $\alpha(u) = \sup_v K(u,v)$ and the concave $\beta(v) = \inf_u K(u,v)$ are both proper and coercive (respectively, in the convex and concave sense).

A concave function g is coercive (in the concave sense) if $-g$ is coercive as a convex function. Condition (a) can be translated to the following: for every $(a,b) \in \mathbb{R}^k \times \mathbb{R}^l$, $\underline{K^*}(a,b) = \overline{K^*}(a,b)$, and the common value is finite).

Proof (Theorem 2.2). By the result quoted above, $J_0^*(\cdot,\cdot)$ is finite if and only if the convex function

$$\phi(u) = \sup_{v\in V}\left\{\frac{1}{2}u\cdot Pu - \frac{1}{2}v\cdot Qv - v\cdot Du + \delta_U(u)\right\}$$

and the concave function

$$\psi(v) = \inf_{u\in U}\left\{\frac{1}{2}u\cdot Pu - \frac{1}{2}v\cdot Qv - v\cdot Du - \delta_V(v)\right\}$$

are proper and coercive. By symmetry, it will suffice to analyze $\phi(\cdot)$. We have

$$\phi(u) = \frac{1}{2}u\cdot Pu + \delta_U(u) + \sup_{v\in V}\left\{v\cdot(-Du) - \frac{1}{2}v\cdot Qv\right\}.$$

Let $\phi_1(u) = \frac{1}{2}u\cdot Pu + \delta_U(u)$ and $\phi_2(u) = \sup_{v\in V}\left\{v\cdot(u) - \frac{1}{2}v\cdot Qv\right\}$. Properness of $\phi(\cdot)$ is equivalent to the existence of some $u\in U$ with $\phi_2(-Du)$ finite. As $\operatorname{dom}\phi_2 = (V^\infty \cap \ker Q)^*$ [26, Example 11.18], we get that $\phi(\cdot)$ is proper if and only if $-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset$. Assuming that this holds, we obtain, through Corollary 11.33 in [26], that the conjugate of the function $u \mapsto \phi_2(-Du)$ at a point w is given by

$$\inf_{v\in V}\left\{\frac{1}{2}v\cdot Qv \mid w = -D^*v\right\}$$

and the domain of this function is $-D^*V$. The domain of $\phi_1^*(\cdot)$ is $(U^\infty \cap \ker P)^*$. Then the domain of $\phi^*(\cdot)$ is $(U^\infty \cap \ker P)^* + (-D^*V)$. Now the properness and coercivity of $\phi(\cdot)$ is equivalent to $\operatorname{dom}\phi^*(\cdot) = \mathbb{R}^k$. We get that $\phi(\cdot)$ is proper and coercive if and only if

$$-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset, \qquad -D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k.$$

Analogous statements for $\psi(\cdot)$ follow after analyzing the convex function $-\psi(\cdot)$ in the above way. We obtain

$$D^*V \cap (U^\infty \cap \ker P)^* \neq \emptyset, \qquad DU + (V^\infty \cap \ker Q)^* = \mathbb{R}^l.$$

Now note that $-D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k$ implies $D^*V \cap (U^\infty \cap \ker P)^* \neq \emptyset$. Indeed, since $0 \in \mathbb{R}^k$, there exists a $v \in V$ such that $0 \in -D^*v + (U^\infty \cap \ker P)^*$. But this means that $D^*v \in (U^\infty \cap \ker P)^*$, so $D^*V \cap (U^\infty \cap \ker P)^* \neq \emptyset$. The latter condition is then superfluous, and a similar statement can be made about $-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset$.

Using the properties of polyhedral sets in Lemma 5.2, we can translate the condition $DU + (V^\infty \cap \ker Q)^* = \mathbb{R}^l$ to

$$DU^\infty + (V^\infty \cap \ker Q)^* = \mathbb{R}^l.$$

By polarizing both sides of this equation according to the rules in Corollary 11.25 in [26], we get one of the conditions in (8). The other one is obtained symmetrically from $-D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k$. The expression for $(DU^\infty)^*$ and $(-D^*V^\infty)^*$ also come from Corollary 11.25. ∎

Saddle points in the definition (7) of J^* are exactly the subgradients of that function. This allows us to use a result of Dontchev and Rockafellar [11] on the stability of saddle points; we quote it below in a form specialized for our current setting. By $(a,b) \in \partial^s K(w,z)$ we mean that $a \in \partial_w K(w,z)$, $b \in \tilde{\partial}_z K(w,z)$.

LEMMA 5.3 (see [11, Theorem 3.2]). *Assume that $(\bar{u},\bar{v}) \in \partial^s J_0^*(a,b)$. Then a necessary and sufficient condition for $\partial^s J_0^*$ to be single-valued and Lipschitz continuous on a neighborhood of (a,b) is the following:*

$$\begin{cases} u \in U_0 - U_0, \quad Pu = 0, \quad Du \in [V_0 \cap -V_0]^\perp \quad \Rightarrow \quad u = 0, \\ v \in V_0 - V_0, \quad Qv = 0, \quad D^*v \in [U_0 \cap -U_0]^\perp \quad \Rightarrow \quad v = 0, \end{cases} \tag{33}$$

where $U_0 = T_U(\bar{u}) \cap (a - P\bar{u} + R^*\bar{v})^\perp$ and $V_0 = T_V(\bar{v}) \cap (b + Q\bar{v} + R\bar{u})^\perp$.

The subspace $U_0 - U_0$ is the smallest subspace containing U_0, whereas $U_0 \cap -U_0$ is the largest subspace contained in the cone U_0. Similarly for V_0.

Proof (Theorem 2.4). For a convex set S, the lineality space S_l of S is the set of all those vectors y, such that for all $x \in S$, the line from x in the direction of y is contained in S. If S is a polyhedral set, $S_l = S^\infty \cap -S^\infty$. Using this notation,

$$[D^*(V^\infty \cap -V^\infty)]^\perp = \left\{u \mid Du \in V_l^\perp\right\},$$

and similarly for the other similar expression in condition (9). Thus, this condition can be restated as

$$\begin{cases} Pu = 0, Du \in V_l^\perp \Rightarrow u = 0, \\ Qv = 0, D^*v \in U_l^\perp \Rightarrow v = 0. \end{cases}$$

We first show that for a closed convex set S and any $w \in N_S(s)$, $S_l \subset w^\perp$. The condition for $w \in N_S(s)$ is that for all $x' \in S$, $(x' - x) \cdot w \le 0$, in particular, for every $l \in S_l$, $l \cdot w \le 0$. But S_l is a subspace, so it must be that $l \cdot w = 0$. This shows that $S_l \subset w^\perp$. Also note that $S_l \subset T_S(s)$.

Pick any (a,b) with $J_0^*(a,b)$ finite. As J_0^* is piecewise linear-quadratic, $\partial^s J_0^*(a,b)$ is nonempty. Pick any $(\bar{u},\bar{v}) \in \partial^s J_0^*(a,b)$. This is equivalent to $(a,b) \in \partial^s J_0(\bar{u},\bar{v})$, meaning $a - P\bar{u} + D^*\bar{v} \in N_U(\bar{u})$ and $b + Q\bar{v} + D\bar{u} \in -N_V(\bar{v})$, and consequently $U_l \subset (a - P\bar{u} + D^*\bar{v})^\perp$ and $V_l \in (b + Q\bar{v} + D\bar{u})^\perp$. This implies that $U_l \subset U_0$ and $V_l \subset V_0$, so then $U_l \subset U_0 \cap -U_0$, $V_l \subset V_0 \cap -V_0$ and also $U_l^\perp \supset (U_0 \cap -U_0)^\perp$, $V_l^\perp \supset (V_0 \cap -V_0)^\perp$.

In view of the above inclusions, condition (9) implies that (33) holds everywhere. That is, in the neighborhood of every point where J^* is finite, this function is also differentiable—therefore, in particular, finite. But the domain of J_0^* is a polyhedral, so also closed, set. Then J_0^* is finite and differentiable everywhere. ∎

A corresponding notion of convergence for convex-concave functions is that of epi/hypo-convergence. We will only use it for sequences of convex-concave functions which are modulated (in the sense of Rockafellar [24]), that is, for sequences which satisfy the following: for some $\rho \ge 0$ and some i_0, we have, for all $i > i_0$,

$$\inf_{|w| \le \rho} \overline{K_i}(w,z) \le \rho(1 + |z|) \quad \forall z, \qquad \sup_{|z| \le \rho} \underline{K_i}(w,z) \ge -\rho(1 + |w|) \quad \forall w. \tag{34}$$

Under Assumption 4.3, the sequence of functions $(y,x) \to H_i(x,y)$ is modulated. This can be seen by looking at the equivalent to Assumption 3.1 growth conditions on the Hamiltonian, as described in Rockafellar and Wolenski [27], Theorem 2.3; see also our proof of Corollary 4.8. A sequence of (equivalence classes of) convex-concave functions K_i is said to *epi/hypo-converge* to K if

$$\lim_{\epsilon \searrow 0} \left[\limsup_{z_i \to z, i \to \infty} \left(\inf_{|w_i - w| \le \epsilon} \overline{K_i}(w_i, z_i) \right) \right] \le \overline{K}(w, z), \tag{35}$$

$$\lim_{\epsilon \searrow 0} \left[\liminf_{w_i \to w, i \to \infty} \left(\sup_{|z_i - z| \le \epsilon} \underline{K_i}(w_i, z_i) \right) \right] \ge \underline{K}(w, z). \tag{36}$$

LEMMA 5.4 (convergence of finite saddle functions). *Let K_i, $i = 1, 2, \dots$ and K be finite-valued convex-concave functions on $\mathbb{R}^k \times \mathbb{R}^l$. The following are equivalent:*

(a) *K_i converge epi/hypo-graphically to k,*

(b) *K_i converge pointwise to k,*

(c) *K_i converge uniformly to k on every compact subset of $\mathbb{R}^k \times \mathbb{R}^l$.*

Proof. Assume (a). Subdifferentials of K_i converge graphically to that of K, this follows from an extension of Attouch's theorem for convex functions; see [24, Theorem 4.3]. As subdifferentials of K are convex-valued, Exercise 5.34 in [26] implies the existence of $N > 0$, $\epsilon_0 > 0$ such that, $\|\partial_w K_i(w', z')\| < N$ for $(w', z') \in (w, z) + \epsilon_0 \mathbb{B}$. For $\epsilon < \epsilon_0$ we have $\inf_{|w_i - w| \le \epsilon} K_i(w_i, z_i) \ge K_i(w, z_i) - \epsilon N$. Using this in (35) we get

$$\begin{aligned} K(w, z) &\ge \lim_{\epsilon \searrow 0} \Big[\limsup_{z_i \to z, i \to \infty} (K_i(w, z_i) - \epsilon N) \Big] \\ &\ge \lim_{\epsilon \searrow 0} \Big[\limsup_{i \to \infty} (K_i(w, z) - \epsilon N) \Big] = \limsup_{i \to \infty} K_i(w, z). \end{aligned}$$

Symmetric argument shows that $K(w, z) \le \liminf_{i \to \infty} K_i(w, z)$, and thus K_i converge to K pointwise. Implication (b)⇒(c) was shown in [19, Theorem 35.1], while (c)⇒ (a) is simple—it follows from the uniform continuity of K and the definition of epi/hypo-convergence. ☐

Proof (Lemma 4.4). The equivalence of (a) and (b) follows from the definitions of $\widetilde{L}_i$, $\widetilde{L}$ and the fact that convex conjugacy preserves epi-convergence; see, for example, Theorem 11.34 in [26]. An extension of this fact to partial conjugacy, first shown by Attouch, Aze, and Wets [1] and specialized to modulated sequences in [24, Theorem 4.1], implies that (a) is equivalent to the "hypo/epi-convergence" of H_i to H. As the Hamiltonians are finite, hypo/epi-convergence is equivalent to their pointwise convergence. ☐

We conclude by discussing the convergence of extended piecewise linear-quadratic problems. Let $\mathcal{C}_i(\tau, \xi)$ be defined as in (3) by matrices A_i, B_i, C_i, D_i, P_i, Q_i, vectors p_i, q_i and sets U_i, V_i. To study the convergence of $\{\mathcal{C}_i(\tau, \xi)\}$ to $\mathcal{C}(\tau, \xi)$ one could analyze the sequence of Lagrangians $\{L_i\}$ defined as in (11), with the help of the calculus of epi-convergence, as described for example in [26, Chapter 7]. We propose an alternate way, suggested by Lemma 4.4 and Example 5.1—we focus on Hamiltonians and rely on the lemma below.

LEMMA 5.5 (convergence of constrained saddle functions and their conjugates). *Suppose that*

(a) *$k_i : \mathbb{R}^k \times \mathbb{R}^l \mapsto \mathbb{R}$, $i = 1, 2, \dots,$ are convex-concave functions converging pointwise to a finite-valued convex-concave function k;*

(b) *$W_i \in \mathbb{R}^k$, $Z_i \in \mathbb{R}^l$, $i = 1, 2, \dots,$ are nonempty closed convex sets converging, respectively, to nonempty closed convex sets W, Z.*

Let $[K_i]$ be the equivalence class of convex-concave functions determined by k_i and $W_i \times Z_i$, similarly define $[K]$ by k and $W \times Z$, and assume that $\{[K_i]\}$ is modulated. Then the sequence $\{[K_i]\}$ epi/hypo-converges to K. Consequently, the sequence $\{[M_i]\}$ given by

$$\underline{M_i}(a,b) = \sup_{w\in W_i}\inf_{z\in Z_i}\{a\cdot w+b\cdot z-k_i(w,z)\},\ \overline{M_i}(a,b) = \inf_{z\in Z_i}\sup_{w\in W_i}\{a\cdot w+b\cdot z-k_i(w,z)\}$$

epi/hypo-converges to $[M]$ described by

$$\underline{M}(a,b) = \sup_{w\in W}\inf_{z\in Z}\{a\cdot w+b\cdot z-k(w,z)\},\ \overline{M}(a,b) = \inf_{z\in Z}\sup_{w\in W}\{a\cdot w+b\cdot z-k(w,z)\}.$$

If all four of the functions above are finite-valued, the equivalence classes $[M_i]$ and $[M]$ consist of just one function each, and the convergence is pointwise.

Proof. We show that (35) holds for $\{K_i\}$ and K; the argument for (36) is symmetrical. When $w \notin W$, there is nothing to prove, as $K(w,z) = +\infty$. Suppose that $w \in W$ and fix $\epsilon > 0$. There exists a sequence $\bar{w}_i \to w$ with $\bar{w}_i \in W_n$, and we have

$$\limsup_{z_i\to z, i\to\infty}\left(\inf_{|w_i-w|\le\epsilon}\overline{K_i}(w_i,z_i)\right) \le \limsup_{z_i\to z, i\to\infty}\overline{K_i}(\bar{w}_i,z_i).$$

If $z \notin Z$, any sequence $z_i \to z$ must eventually satisfy $z_i \notin Z_i$, and thus $\overline{K_i}(\bar{w}_i,z_i) = -\infty$. Thus

$$\limsup_{z_i\to z, i\to\infty}\left(\inf_{|w_i-w|\le\epsilon}\overline{K_i}(w_i,z_i)\right) = -\infty = \overline{K}(w,z).$$

Now note that

$$\begin{aligned}\limsup_{z_i\to z, i\to\infty}\left(\inf_{|w_i-w|\le\epsilon}\overline{K_i}(w_i,z_i)\right) &\le \limsup_{z_i\to z, i\to\infty}\overline{K_i}(\bar{w}_i,z_i)\\ &\le \limsup_{z_n\to z, n\to\infty} k_n(\bar{w}_n,z_n) = k(w,z),\end{aligned}$$

where the equality follows from the fact that k_i converge to k uniformly on any compact neighborhood of (w,z) (Lemma 5.4). If $z \in Z$, $k(w,z) = \overline{K}(w,z)$, and this shows the epi/hypo-convergence of $\{K_i\}$ to K.

Epi/hypo-convergence is preserved under saddle function conjugacy [24, Theorem 4.2]. As $\{M_i\}$ are saddle conjugates of $\{K_i\}$ (in (31) the infimum and supremum need to be taken only over the sets where the function values are finite), epi/hypo-convergence of $\{K_i\}$ to K implies that of $\{M_i\}$ to M. The last statement follows from Lemma 5.4. ∎

A related result was shown by Wright [29]. It concluded the convergence of $\{K_i\}$, if each K_i had the form $k_i'(w) - k_i''(z) - w \cdot Dz$ (separable saddle function plus a constant biaffine term); convergence of M_i was not addressed there. Also in [29], epi/hypo-convergence was employed to study discrete approximations of $C(\tau,\xi)$.

THEOREM 5.6 (convergence of piecewise linear-quadratic Hamiltonians). *Assume that matrices A_i, B_i, C_i, D_i, P_i, Q_i, vectors p_i, q_i and sets U_i, V_i defining the problem $\mathcal{C}_i(\tau,\xi)$ converge, respectively, to A, B, C, D, P, Q, p, q, U, V defining $\mathcal{C}(\tau,\xi)$. Suppose also that the data in $\mathcal{C}_i(\tau,\xi)$, $i = 1,2,\ldots,$ and $C(\tau,\xi)$ satisfies the conditions of Theorem 2.2. Then Hamiltonians H_i converge pointwise to H.*

Proof. The sequence of functions J_i corresponding to $\mathcal{C}_i(\tau,\xi)$ as in (32) is modulated (too see this, note that there exist $u_i \in U_i$ converging to some $u \in U$, and for some $\rho > 0$, $\inf_{|u|\leq\rho} \overline{J}_i(u,v)$ is bounded above by $p_i \cdot u_i + \frac{1}{2} u_i \cdot P_i u_i + q_i \cdot v - v \cdot D_i u_i$; this shows the first inequality in (34)). The quadratic expressions defining J_i in (32) converge pointwise (on the whole space) to that of J. Lemma 5.5 guarantees that $\{J_i\}$ as well as $\{J_i^*\}$ epi/hypo-converge to, respectively, J and J^*. As the functions J_i^* and J^* are finite, their convergence is uniform on compact sets by Lemma 5.4. But then, it also implies the pointwise convergence of Hamiltonians H_i. □

REFERENCES

[1] H. Attouch, D. Azé, and R. J.-B. Wets, *Convergence of convex-concave saddle functions: Applications to convex programming and mechanics*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 5 (1988), pp. 537–572.

[2] H. Attouch and R. J.-B. Wets, *A convergence theory for saddle functions*, Trans. Amer. Math. Soc., 280 (1983), pp. 1–41.

[3] A. Bemporad, M. Morari, V. Dua, and E. Pistikopoulos, *The explicit linear quadratic regulator for constrained systems*, Automatica J. IFAC, 38 (2002), pp. 3–20.

[4] A. Briani, *Convergence of Hamilton-Jacobi equations for sequences of optimal control problems*, Commun. Appl. Anal., 4 (2000), pp. 227–244.

[5] G. Buttazzo and G. Dal Maso, *Γ-convergence and optimal control problems*, J. Optim. Theory Appl., 38 (1982), pp. 385–407.

[6] C. Byrnes, *On the Riccati partial differential equation for nonlinear Bolza and Lagrange problems*, J. Math. Systems Estim. Control, 8 (1998), pp. 1–54.

[7] C. Byrnes and H. Frankowska, *Unicité des solutions optimales et absence de chocs pour les équations d'Hamilton-Jacobi-Bellman et de Riccati*, C. R. Acad. Sci. Paris Ser. I Math, 315 (1992), pp. 427–431.

[8] N. Caroff and H. Frankowska, *Optimality and characteristics of Hamilton-Jacobi-Bellman equations*, in Optimization, Optimal Control, and Partial Differential Equations Internat. Ser. Numer. Math., 107, Birkhäuser, Basel, 1992, pp. 169–180.

[9] N. Caroff and H. Frankowska, *Conjugate points and shocks in nonlinear optimal control*, Trans. Amer. Math. Soc., 348 (1996), pp. 3133–3153.

[10] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[11] A. Dontchev and R. Rockafellar, *Primal-dual solution perturbations in convex optimization*, Set-Valued Anal., 9 (2001), pp. 49–65.

[12] R. Goebel, *Convexity, Convergence and Feedback in Optimal Control*, Ph.D. thesis, University of Washington, Seattle, WA, 2000.

[13] R. Goebel, *Convexity in zero-sum differential games*, SIAM J. Control Optim., 40 (2002), pp. 1491–1504.

[14] R. Goebel, *Regularity of the optimal feedback and the value function in convex problems of optimal control*, Set-Valued Anal., 12 (2004), pp. 127–145.

[15] R. Goebel and R. Rockafellar, *Generalized conjugacy in Hamilton-Jacobi theory for fully convex Lagrangians*, J. Convex Anal., 9 (2002), pp. 463–473.

[16] W. Heemels, S. V. Eijndhoven, and A. Stoorvogel, *Linear quadratic regulator with positive controls*, Internat. J. Control, 70 (1998), pp. 551–578.

[17] J. Joly and F. Thelin, *Convergence of convex integrals in l^p spaces*, J. Math. Anal. Appl., 54 (1976), pp. 230–244.

[18] R. Rockafellar, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl., 32 (1970), pp. 174–222.

[19] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[20] R. Rockafellar, *Generalized Hamiltonian equations for convex problems of Lagrange*, Pacific J. Math., 33 (1970), pp. 411–427.

[21] R. Rockafellar, *Existence and duality theorems for convex problems of Bolza*, Trans. Amer. Math. Soc., 159 (1971), pp. 1–40.

[22] R. Rockafellar, *Linear-quadratic programming and optimal control*, SIAM J. Control Optim., 25 (1987), pp. 781–814.

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

[24] R. Rockafellar, *Generalized second derivatives of convex functions and saddle functions*, Trans. Amer. Math. Soc., 322 (1990), pp. 51–77.

[25] R. Rockafellar, *Large-scale extended linear-quadratic programming and multistage optimization*, in Advances in Numerical Partial Differential Equations and Optimization (Merida, 1989), SIAM, Philadelphia, 1991, pp. 247–261.

[26] R. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[27] R. Rockafellar and P. Wolenski, *Convexity in Hamilton–Jacobi theory,* I: *Dynamics and duality*, SIAM J. Control Optim., 39 (2000), pp. 1323–1350.
[28] R. Rockafellar and C. Zhu, *Primal-dual projected gradient algorithms for extended linear-quadratic programming*, SIAM J. Optim., 3 (1993), pp. 751–783.
[29] S. Wright, *Consistency of primal-dual approximations for convex optimal control problems*, SIAM J. Control Optim., 33 (1995), pp. 1489–1509.
[30] C. Zhu, *On a certain parameter of the discretized extended linear-quadratic problem of optimal control*, SIAM J. Control Optim., 34 (1996), pp. 62–73.