A CURSE-OF-DIMENSIONALITY-FREE NUMERICAL METHOD FOR SOLUTION OF CERTAIN HJB PDES*

WILLIAM M. MCENEANEY †

Abstract. In previous work of the author and others, max-plus methods have been explored for solution of first-order, nonlinear Hamilton-Jacobi-Bellman partial differential equations (HJB PDEs) and corresponding nonlinear control problems. These methods exploit the max-plus linearity of the associated semigroups. In particular, although the problems are nonlinear, the semigroups are linear in the max-plus sense. These methods have been used successfully to compute solutions. Although they provide certain computational-speed advantages, they still generally suffer from the curse-of-dimensionality. Here we consider HJB PDEs where the Hamiltonian takes the form of a (pointwise) maximum of linear/quadratic forms. The approach to solution will be rather general, but in order to ground the work, we consider only constituent Hamiltonians corresponding to long-run average-cost-per-unit-time optimal control problems for the development. We obtain a numerical method not subject to the curse-of-dimensionality. The method is based on construction of the dual-space semigroup corresponding to the HJB PDE. This dual-space semigroup is constructed from the dual-space semigroups corresponding to the constituent linear/quadratic Hamiltonians. The dual-space semigroup is particularly useful due to its form as a max-plus integral operator with kernel obtained from the originating semigroup. One considers repeated application of the dual-space semigroup to obtain the solution.

Key words. partial differential equations, curse-of-dimensionality, dynamic programming, max-plus algebra, Legendre transform, Fenchel transform, semiconvexity, Hamilton-Jacobi-Bellman equations, idempotent analysis.

AMS subject classifications. 49LXX, 93C10, 35B37, 35F20, 65N99, 47D99

1. Introduction.

One approach to nonlinear control is through Dynamic Programming (DP). With DP, solution of the control problem "reduces" to solution of the corresponding partial differential equation (PDE). In the case of deterministic optimal control or deterministic games where one player's feedback is prespecified, the PDE is a Hamilton-Jacobi-Bellman (HJB) PDE. If one can solve the HJB PDE, then this approach is ideal in that one obtains the optimal control for the given criterion as opposed to a control meeting only some weaker goal such as stability. The problem is that one must solve the HJB PDE! We should remark that such HJB PDEs also arise in Robust/H_∞ nonlinear filtering and Robust/H_∞ control under partial information.

Various approaches have been taken to solution of the HJB PDE. First note that it is a fully nonlinear, first-order PDE. Consequently, the solutions are generally nonsmooth (with the exception of the linear/quadratic case of course), and one must use the theory of viscosity solutions [3], [9], [10], [11], [21]. One approach to solution is through generalized characteristics (cf. [37], [38], as well as [16], [22] for classical treatments). This approach can obtain the solution very quickly at a single point *if* the solution is smooth. However, the nonsmoothness introduces tremendous difficulties which appear, to the author, to be difficult to handle in an automated approach. In particular, the projections of the characteristics into the state space can cross and/or may not cover the entire state space (in analogy with shocks and rarefaction waves).

The most common methods by far all fall into the class of grid-based methods (cf. [3], [13], [14], [21], [26] among many others). These require that one generate a grid over some bounded region of the state-space. In this general class of methods, we include finite-difference methods, finite element methods, and those dynamic programming based methods which map the continuum problem onto some discrete space. Although higher-order grid-based methods are being explored (c.f. [5], [15], [17]), there are still hard lower limits to the computational growth as a function of space-dimension. In particular, suppose the region over which one constructs the grid is rectangular, say square for simplicity. Further, suppose one uses 100

* Research partially supported by NSF Grant DMS-0307229 and AFOSR Grant FA9550-06-1-0238.

†Dept. of Mechanical and Aerospace Engineering, University of California San Diego, 9500 Gilman Dr., La Jolla, CA, 92093-0411, USA (wmceneaney@ucsd.edu).

grid points per dimension. (Clearly 50 would be the minimum acceptable, and 100 could be a bit sparse.) If the state dimension is n, then one has 100^n grid points. Thus the computations grow exponentially in state-space dimension n. If the computations per grid point grew with state-space dimension like 2^n, then the computations would grow like $(200C)^n$ for some constant, C. For concreteness, we discuss only the steady-state PDE case here. If the state-space dimension is 3, these problems are feasible to solve on current generation machinery. However, the computations will grow by more than 8×10^6 in going from a dimension 3 problem to a dimension 6 problem. Parallel algorithms can alleviate this problem to some extent (c.f. [4]). However, there can only be rather limited improvement in the dimension of problems which can be handled by such techniques.

In recent years, an entirely new class of numerical methods for HJB PDEs has emerged [20], [35], [1], [24], [33], [32], [34], [31], [29]. These methods exploit the max-plus (or min-plus [8], [33]) linearity of the associated semigroup. They employ a max-plus basis function expansion of the solution, and the numerical methods obtain the coefficients in the basis expansion. We will refer to these methods as *max-plus basis methods.* Much of the work has concentrated on the (harder) steady-state HJB PDE class where (for both max-plus basis and grid-based methods), one propagates forward in "time" to obtain the steady-state limit solution. With the max-plus basis methods, the number of basis functions required still typically grows exponentially with space dimension. For instance, one might use 25 basis functions per space dimension. Consequently, one still has the curse-of-dimensionality. With the max-plus basis methods, the "time-step" tends to be much larger than what can be used in grid-based methods (since it encapsulates the action of the semigroup propagation on each basis function), and so these methods can be quite fast on small problems. Even with a max-plus basis approach, the curse-of-dimensionality growth is so fast that one cannot expect to solve general problems of more than say dimension 5 on current machinery, and again the computing machinery speed increases expected in the foreseeable future cannot do much to raise this.

Many researchers have noticed that the introduction of even a single, simple nonlinearity into an otherwise linear control problem of high dimensionality, say n, has disastrous computational repercussions. Specifically, one goes from solution of an n-dimensional Riccati equation to solution of a grid-based or max-plus basis method over a space of dimension n. While the Riccati equation may be "relatively" easily solved for large n, the max-plus and grid-based methods have no hope on general problems of dimension, say, $n \geq 6$. This has been a frustrating, counter-intuitive situation for decades.

This paper discusses an approach to certain nonlinear HJB PDEs which is not subject to the curse-of-dimensionality. Although this approach also utilizes the max-plus algebra, the method is largely unrelated to the max-plus basis approaches discussed above. In fact, for this new method, the computational growth in state-space dimension is on the order of n^3. There is of course no "free lunch", and there is exponential computational growth in a certain measure of complexity of the Hamiltonian. Under this measure, the minimal complexity Hamiltonian is the linear/quadratic Hamiltonian – corresponding to solution by a Riccati equation. If the Hamiltonian is given as a pointwise maximum or minimum of M linear/quadratic Hamiltonians, then one could say the *complexity* of the Hamiltonian is M. One could also apply this approach to a wider class of HJB PDEs with semiconvex Hamiltonians (by approximation of the Hamiltonian by a finite number of quadratic forms), but that is certainly beyond the scope of this paper.

The approach has been applied on some simple nonlinear problems. A steady-state HJB PDE comprised of 2 linear/quadratic components was solved in dimensions 2-3 in under 5-10 seconds on a standard PC and in 20 seconds over $\mathbf{R}^4$. A few simple examples comprised of 3 linear/quadratic components were solved in 10-20 seconds over $\mathbf{R}^3$ and 10-45 seconds over $\mathbf{R}^4$. For these particular problems, the solution was obtained over the entire space (as opposed to a rectangular region) with the resulting errors in the *gradients* growing linearly in $|x|$. (See Section 7 for more information on specific examples.) These speeds are of course unprecedented in standard general approaches to nonlinear PDEs. This code was not optimized, and there are many computational cost reduction methods that one could employ to further reduce computational growth. Further, the computational growth in going from $n = 4$ up to say $n = 6$

would be on the order of $6^3/4^3 \simeq 4$ as opposed to say more than 10^4 for a grid-based method.

We will be concerned here with HJB PDEs of the form $0 = \widetilde{H}(x, \operatorname{grad} V)$, where the Hamiltonians are given or approximated as

$$\widetilde{H}(x, \operatorname{grad} V) = \max_{m \in \{1,2,\ldots,M\}} \{H^m(x, \operatorname{grad} V)\}.$$

In order to make the problem tractable, we will concentrate on a single class of HJB PDEs – those for long-run average-cost-per-unit-time problems. However, the theory can obviously be expanded to a much larger class.

Since the development of the proposed method in the following sections takes quite a few pages, we briefly outline the main points here. First, recall that the solution of the above PDE is the eigenfunction of the corresponding semigroup, that is

$$0 \otimes V = V = \widetilde{S}_\tau[V]$$

where $\oplus, \otimes$ denote max-plus addition and multiplication, and we note that $\widetilde{S}_\tau$ is max-plus linear (cf. [20], [28], [33]). The Legendre/Fenchel transform maps this to the dual space eigenfunction problem

$$0 \otimes e = \widetilde{\mathcal{B}}_\tau \odot e$$

where we use the $\odot$ notation to indicate $\widetilde{\mathcal{B}}_\tau \odot e \doteq \int_{\mathbb{R}^n}^{\oplus} \widetilde{\mathcal{B}}_\tau(x,y) \otimes e(y)\, dy$ where $\int^{\oplus}$ denotes max-plus integration (maximization). Then one approximates $\widetilde{\mathcal{B}}_\tau \simeq \bigoplus_{m \in \mathcal{M}} \mathcal{B}_\tau^m$ where $\mathcal{M} \doteq \{1, 2, \ldots, M\}$ and the $\mathcal{B}_\tau^m$ correspond to the H^m. The max-plus power method ([12], [25], [33]) suggests that the solution is approximated by the form

$$e \simeq \lim_{N \to \infty} \left[\bigoplus_{m \in \mathcal{M}} \mathcal{B}_\tau^m \right]^N \odot 0 = \lim_{N \to \infty} \left[\bigoplus_{\{m_i\}_{i=1}^N} \mathcal{B}_\tau^{m_1} \otimes \mathcal{B}_\tau^{m_2} \otimes \ldots \mathcal{B}_\tau^{m_N} \right] \odot 0$$

where the N superscript denotes the $\odot$ operation N times, and 0 represents the zero function. Given linear/quadratic forms for each of the H^m, the $\mathcal{B}_\tau^m$ are obtained by Riccati equations. Let $e_N \doteq \left[\bigoplus_{m \in \mathcal{M}} \mathcal{B}_\tau^m \right]^N \odot 0$. Then $e_N \to e$. The convergence rate does not depend on space dimension, but on the dynamics of the problem. There is no curse-of-dimensionality. The exponential growth is in $M = \#\mathcal{M}$. Given the solution of the Riccati equations for the H^m, the computation of each product, $\mathcal{B}_\tau^{m_1} \otimes \mathcal{B}_\tau^{m_2} \otimes \ldots \mathcal{B}_\tau^{m_N}$, is analytical, modulo $n \times n$ matrix inversions (and hence the n^3 computational growth rate).

In Section 2, the class of control problems and HJB PDEs which we will use to demonstrate the theory will be given. We will also review the existing theory relevant to our problem there. In Section 3 the relation between solution of the HJB PDEs and their corresponding semiconvex dual problems will be discussed. In Section 4, a discrete-time approximation of the semigroup for the problem of interest will be introduced, and convergence of the solutions of the approximate problems to the original problem will be obtained. The algorithm itself will be developed in Section 5. The basic algorithm is not subject to the curse-of-dimensionality. However, practical implementation requires some additional work; some initial remarks on this appear in Section 6. The algorithm is applied to some simple examples in Section 7. Finally, Section 8 sketches some future directions.

2. Sample problem class and review of theory. There are certain conditions which must be satisfied for solutions to exist and the method to apply. In order that the assumptions are not completely abstract, we will work with a specific problem class – the infinite time-horizon H_∞ problem with fixed feedback. This class consists of long-term average-cost-per-unit-time problems. Moreover, it is a problem class where there already exists a good deal of results, and so less analysis will be required for application of the new method.

As indicated above, we suppose the individual H^m are linear/quadratic Hamiltonians. Consequently, consider a finite set of linear systems

$$\dot{\xi}^m = A^m \xi^m + \sigma^m w, \quad \xi_0^m = x \in I\!R^n. \tag{2.1}$$

Let $w \in \mathcal{W} \doteq L_2^{loc}([0,\infty); I\!R^m)$ where we recall that $L_2^{loc}([0,\infty); I\!R^m) = \{w : [0,\infty) \to I\!R^m \, : \, \int_0^T |w_t|^2 \, dt < \infty \; \forall T < \infty\}$. Let the cost functionals be

$$J^m(x,T;w) \doteq \int_0^T \tfrac{1}{2}\xi_t^m D^m \xi_t^m - \frac{\gamma^2}{2}|w_t|^2 \, dt, \tag{2.2}$$

and let the value function (also known as the available storage in this context) be

$$V^m(x) = \sup_{w\in\mathcal{W}} \sup_{T<\infty} J^m(x,T;w) = \lim_{T\to\infty} \sup_{w\in\mathcal{W}} J^m(x,T;w). \tag{2.3}$$

We remark that a generalization of the second term in the integrand of the cost functional to $\frac{1}{2}w^T C^m w$ with C^m symmetric, positive definite is not needed since this is equivalent to a change in σ^m in the dynamics (2.1). Obviously J^m and V^m require some assumptions in order to guarantee their existence. The assumptions will hold throughout the paper. Since these assumptions only appear together, we will refer to this entire set of assumptions as Assumption Block $(A.m)$, and this is:

$(A.m)$

Assume that there exists $c_A \in (0,\infty)$ such that

$$x^T A^m x \le -c_A |x|^2 \qquad \forall\, x \in I\!R^n,\; m \in \mathcal{M}.$$

Assume that there exists $c_\sigma < \infty$ such that

$$|\sigma^m| \le c_\sigma \qquad \forall\, m \in \mathcal{M}.$$

Assume that all D^m are positive definite, symmetric, and let c_D be such that

$$x^T D^m x \le c_D |x|^2 \qquad \forall\, x \in I\!R^n,\; m \in \mathcal{M}$$

(which is obviously equivalent to all eigenvalues of the D^m being no greater than c_D). Lastly, assume that $\gamma^2/c_\sigma^2 > c_D/c_A^2$.

Note that these assumptions guarantee the existence of the V^m as locally bounded functions which are zero at the origin (cf. [36]). (These assumptions could be weakened by using the specific linear/quadratic structure, but that would distract from the goal of this paper.) The corresponding HJB PDEs are

$$\begin{aligned} 0 &= -H^m(x, \operatorname{grad} V) = -\left\{ \tfrac{1}{2}x^T D^m x + (A^m x)^T \operatorname{grad} V + \max_{w\in I\!R^m} \left[(\sigma^m w)^T \operatorname{grad} V - \frac{\gamma^2}{2}|w|^2 \right] \right\} \\ &= -\left\{ \tfrac{1}{2}x^T D^m x + (A^m x)^T \operatorname{grad} V + \tfrac{1}{2} \operatorname{grad} V^T \Sigma^m \operatorname{grad} V \right\} \\ V(0) &= 0 \end{aligned} \tag{2.4}$$

where $\Sigma^m \doteq \frac{1}{\gamma^2}\sigma^m(\sigma^m)^T$. Let $I\!R^- \doteq I\!R \cup \{-\infty\}$. Recall that a function, $\phi : I\!R^n \to I\!R^-$ is semiconvex if given any $R \in (0,\infty)$ there exists $k_R \in I\!R$ such that $\phi(x) + \frac{k_R}{2}|x|^2$ is convex over $\overline{B}_R(0) = \{x \in I\!R^n : |x| \le R\}$. For a fixed choice of c_A, c_σ, $\gamma > 0$ satisfying the above assumptions, and for any $\delta \in (0,\gamma)$ we define

$$\mathcal{G}_\delta = \left\{ V : I\!R^n \to [0,\infty) \,\middle|\, V \text{ is semiconvex and } V(x) \le \frac{c_A(\gamma-\delta)^2}{c_\sigma^2}|x|^2,\; \forall x \in I\!R^n \right\}.$$

From [36] (undoubtedly among many others), each value function (2.3) is the unique viscosity solution of its corresponding HJB PDE (2.4) in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$.

From the structure of the running cost and dynamics, it is easy to see (c.f. [42], [36]) that each V^m satisfies

$$V^m(x) = \sup_{T<\infty} \sup_{w\in\mathcal{W}} J^m(x,T;w) = \lim_{T\to\infty} \sup_{w\in\mathcal{W}} J^m(x,T;w) \doteq \lim_{T\to\infty} V^{m,f}(x,T), \tag{2.5}$$

and that each $V^{m,f}$ is the unique continuous viscosity solution of (cf. [3], [21])

$$0 = V_T - H^m(x, \operatorname{grad} V), \quad V(0,x) = 0. \tag{2.6}$$

It is easy to see that these solutions have the form $V^{m,f}(x,t) = \frac{1}{2}x^T P_t^{m,f} x$ where each $P^{m,f}$ satisfies the differential Riccati equation

$$\dot{P}^{m,f} = (A^m)^T P^{m,f} + P^{m,f} A^m + D^m + P^{m,f}\Sigma^m P^{m,f}, \quad P_0^{m,f} = 0. \tag{2.7}$$

By (2.5) and (2.7), the V^m take the form $V(x) = \frac{1}{2}x^T P^m x$ where $P^m = \lim_{t\to\infty} P_t^{m,f}$. With this form, and (2.4) (or (2.7)), we see that the P^m satisfy the algebraic Riccati equations

$$0 = (A^m)^T P^m + P^m A^m + D^m + P^m \Sigma^m P^m. \tag{2.8}$$

Combining this with the above, one has

THEOREM 2.1. *Each value function (2.3) is the unique classical solution of its corresponding HJB PDE (2.4) in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$. Further, $V^m(x) = \frac{1}{2}x^T P^m x$ where P^m is the smallest symmetric, positive definite solution of (2.8)*

The duality between viscosity (and/or classical) solutions of the HJB PDEs and the corresponding value functions is certainly very important. However, the method we will use to obtain these value functions/HJB PDE solutions will be through the associated semigroups. These semigroups are equivalent to dynamic programming principles (DPPs). Consequently, for each m define the semigroup

$$S_T^m[\phi] \doteq \sup_{w\in\mathcal{W}} \left[\int_0^T \tfrac{1}{2}(\xi_t^m)^T D^m \xi_t^m - \frac{\gamma^2}{2}|w_t|^2\, dt + \phi(\xi_T^m) \right] \tag{2.9}$$

where ξ^m satisfies (2.1). By [36], the domain of S_T^m includes $\mathcal{G}_\delta$ for all $\delta > 0$. The following result is similar to that in [33]; the only significant difference is that, in this case, $V^m(x) = \frac{1}{2}x^T P^m x$ is smooth.

THEOREM 2.2. *Fix any $T > 0$. Each value function, V^m, is the unique smooth solution of $V = S_T^m[V]$ in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$. Further, given any $V \in \mathcal{G}_\delta$, $\lim_{T\to\infty} S_T^m[V](x) = V^m(x)$ for all $x \in I\!R^n$ (uniformly on compact sets).*

Recall that the HJB PDE problem of interest is

$$0 = -\widetilde{H}(x, \operatorname{grad} V) \doteq -\max_{m\in\mathcal{M}} H^m(x, \operatorname{grad} V), \quad V(0) = 0. \tag{2.10}$$

The corresponding value function is

$$\widetilde{V}(x) = \sup_{w\in\mathcal{W}} \sup_{\mu\in\mathcal{D}_\infty} \widetilde{J}(x,w,\mu) \doteq \sup_{w\in\mathcal{W}} \sup_{\mu\in\mathcal{D}_\infty} \sup_{T<\infty} \int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\, dt \tag{2.11}$$

where $l^{\mu_t}(x) = \frac{1}{2}x^T D^{\mu_t} x$, $\mathcal{D}_\infty = \{\mu : [0,\infty) \to \mathcal{M} : \text{ measurable}\}$, and ξ satisfies

$$\dot{\xi} = A^{\mu_t}\xi + \sigma^{\mu_t}w_t, \quad \xi_0 = x. \tag{2.12}$$

THEOREM 2.3. *Value function $\widetilde{V}$ is the unique viscosity solution of (2.10) in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$.*

REMARK 2.4. The proof of Theorem 2.3 is nearly identical to the proof of Theorems 2.5 and 2.6 from [36], with only trivial changes, and so is not included. In particular, rather than choosing any $w \in \mathcal{W}$, one chooses both any $w \in \mathcal{W}$ and any $\mu \in \mathcal{D}_\infty$.

Define the semigroup

$$\widetilde{S}_T[\phi] = \sup_{w\in\mathcal{W}} \sup_{\mu\in\mathcal{D}_T} \left[\int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt + \phi(\xi_T) \right] \tag{2.13}$$

where $\mathcal{D}_T = \{\mu : [0,T) \to \mathcal{M} : \text{ measurable}\}$. In analogy with Theorem 2.2, one has:

THEOREM 2.5. *Fix any $T > 0$. Value function $\widetilde{V}$ is the unique continuous solution of $V = \widetilde{S}_T[V]$ in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$. Further, given any $V \in \mathcal{G}_\delta$, $\lim_{T\to\infty} \widetilde{S}_T[V](x) = \widetilde{V}(x)$ for all $x \in I\!R^n$ (uniformly on compact sets).*

The proof is nearly identical to the proof of a similar result in [33], and so is not included. In particular, the only change is the addition of the supremum over $\mathcal{D}_T$ – which makes no substantial change in the proof. Importantly, we also have the following.

THEOREM 2.6. *There exists $c_V > 0$ such that $\widetilde{V}(x) - \frac{1}{2}c_V|x|^2$ is strictly convex.*

Proof. Fix any $x, \nu \in I\!R^n$ with $|\nu| = 1$ and any $\delta > 0$. Let $\varepsilon > 0$. Given x, let $w^\varepsilon \in \mathcal{W}$, $\mu^\varepsilon \in \mathcal{D}_\infty$ be ε–optimal for $\widetilde{V}(x)$. Then

$$\begin{aligned} &\widetilde{V}(x-\delta\nu) - 2\widetilde{V}(x) + \widetilde{V}(x+\delta\nu) \\ &\geq \widetilde{J}(x-\delta\nu, w^\varepsilon, \mu^\varepsilon) - 2\widetilde{J}(x, w^\varepsilon, \mu^\varepsilon) + \widetilde{J}(x+\delta\nu, w^\varepsilon, \mu^\varepsilon) - 2\varepsilon. \end{aligned} \tag{2.14}$$

Let $\xi^\delta, \xi^0, \xi^{-\delta}$ be solutions of dynamics (2.12), but with initial conditions $\xi_0^\delta = x + \delta\nu$, $\xi_0^0 = x$ and $\xi_0^{-\delta} = x - \delta\nu$, respectively, where the inputs are w^ε and μ^ε for all three processes. Then

$$\dot{\xi}^\delta - \dot{\xi}^0 = A^{\mu_t^\varepsilon}[\xi^\delta - \xi^0], \quad\text{and}\quad \dot{\xi}^0 - \dot{\xi}^{-\delta} = A^{\mu_t^\varepsilon}[\xi^0 - \xi^{-\delta}]. \tag{2.15}$$

Letting $\Delta_t^+ \doteq \xi_t^\delta - \xi_t^0$, one also has $\xi_t^0 - \xi_t^{-\delta} = \Delta_t^+$, and by linearity one finds $\dot{\Delta}^+ = A^{\mu_t^\varepsilon}\Delta^+$. Also, using (2.14) and (2.11)

$$\widetilde{V}(x-\delta\nu) - 2\widetilde{V}(x) + \widetilde{V}(x+\delta\nu) \geq \int_0^\infty (\Delta^+)^T D^{\mu_t^\varepsilon}\Delta^+\,dt - 2\varepsilon. \tag{2.16}$$

Also, by the finiteness of $\mathcal{M}$, there exists $K < \infty$ such that

$$\frac{d}{dt}|\Delta^+|^2 = 2(\Delta^+)^T A^{\mu_t^\varepsilon}\Delta^+ \geq -K|\Delta^+|^2$$

which implies

$$|\Delta^+|^2 \geq e^{-Kt}\delta^2 \qquad \forall\, t \geq 0. \tag{2.17}$$

Let $\lambda_D \doteq \min\{\lambda \in I\!R : \lambda \text{ is an eigenvalue of a } D^m\}$. By the positive definiteness of the D^m and finiteness of $\mathcal{M}$, $\lambda_D > 0$. Consequently, by (2.16), and then (2.17),

$$\widetilde{V}(x-\delta\nu) - 2\widetilde{V}(x) + \widetilde{V}(x+\delta\nu) \geq \int_0^\infty \lambda_D|\Delta^+|^2\,dt - 2\varepsilon \geq \frac{\lambda_D}{K}\delta^2 - 2\varepsilon. \tag{2.18}$$

Since $\varepsilon > 0$ and $|\nu| = 1$ were arbitrary, one obtains the result. ∎

3. Max-plus spaces and dual operators. Again, recall that a function, $\phi : I\!R^n \to I\!R^-$ is semiconvex if given any $R \in (0,\infty)$ there exists $\beta_R \in I\!R$ such that $\phi(x) + \frac{\beta_R}{2}|x|^2$ is convex over $\overline{B}_R(0) = \{x \in I\!R^n : |x| \le R\}$. We will modify this definition by allowing the β_R to be $n \times n$, symmetric, positive or negative definite matrices. We will denote the set of such matrices as $\mathcal{D}_n$. We say ϕ is uniformly semiconvex with (symmetric, definite matrix) constant $\beta \in \mathcal{D}_n$ if $\phi(x) + \frac{1}{2}x^T\beta x$ is convex over $I\!R^n$. Let $\mathcal{S}_\beta = \mathcal{S}_\beta(I\!R^n)$ be the set of functions mapping $I\!R^n$ into $I\!R^-$ which are uniformly semiconvex with (symmetric, definite matrix) constant β. (A negative definite semiconvexity constant corresponds to functions which are still convex after subtracting a convex quadratic.) Also note that $\mathcal{S}_\beta$ is a max-plus vector space (also known as a moduloid) [20], [33], [2], [7], [27]. For instance, $\alpha_1 \otimes \phi_1 \oplus \alpha_2 \otimes \phi_2 \in \mathcal{S}_\beta$ for all $\alpha_1, \alpha_2 \in I\!R^-$ and all $\phi_1, \phi_2 \in \mathcal{S}_\beta$. Combining Theorems 2.1 and 2.6, we have the following.

THEOREM 3.1. *There exists $\overline{\beta} \in \mathcal{D}_n$ such that given any β such that $\beta - \overline{\beta} > 0$ (i.e., $\beta - \overline{\beta}$ positive definite), $\widetilde{V} \in \mathcal{S}_\beta$ and $V^m \in \mathcal{S}_\beta$ for all $m \in \mathcal{M}$. Further, one may take β negative definite (i.e., $\widetilde{V}, V^m$ are convex).*

We henceforth assume we have chosen β such that $\beta - \overline{\beta} > 0$.

Throughout the remainder, we will employ certain transform kernel functions, $\psi : I\!R^n \times I\!R^n \to I\!R$ which take the form

$$\psi(x,z) = \tfrac{1}{2}(x-z)^T C (x-z)$$

with nonsingular, symmetric C satisfying $C+\beta < 0$ (i.e., $C+\beta$ negative definite). The following semiconvex duality result [20], [32], [33] requires only a small modification of convex duality and Legendre/Fenchel transform results [39], [40].

THEOREM 3.2. *Let $\phi \in \mathcal{S}_\beta$. Let C and ψ be as above. Then, for all $x \in I\!R^n$,*

$$\phi(x) = \max_{z \in I\!R^n} [\psi(x,z) + a(z)] \tag{3.1}$$

$$= \int_{I\!R^n}^{\oplus} \psi(x,z) \otimes a(z)\, dz = \psi(x,\cdot) \odot a(\cdot) \tag{3.2}$$

where for all $z \in I\!R^n$

$$a(z) = -\max_{x \in I\!R^n} [\psi(x,z) - \phi(x)] \tag{3.3}$$

$$= -\int_{I\!R^n}^{\oplus} \psi(x,z) \otimes [-\phi(x)]\, dx = -\{\psi(\cdot,z) \odot [-\phi(\cdot)]\} \tag{3.4}$$

which using the notation of [7]

$$= \left\{\psi(\cdot,z) \odot [\phi^-(\cdot)]\right\}^- . \tag{3.5}$$

We will refer to a as the *semiconvex dual* of ϕ (with respect to ψ).

REMARK 3.3. We note that $\phi \in \mathcal{S}_\beta$ implies that ϕ is locally Lipschitz (cf. [19]). We also note that if $\phi \in \mathcal{S}_\beta$ and if there is any $x \in I\!R^n$ such that $\phi(x) = -\infty$, then $\phi \equiv -\infty$. Henceforth, we will ignore the special case of $\phi \equiv -\infty$, and assume that all functions are real-valued.

Semiconcavity is the obvious analogue of semiconvexity. In particular, a function, $\phi : I\!R^n \to I\!R \cup \{+\infty\}$, is uniformly semiconcave with constant $\beta \in \mathcal{D}_n$ if $\phi(x) - \frac{1}{2}x^T\beta x$ is concave over $I\!R^n$. Let $\mathcal{S}^-_\beta$ be the set of functions mapping $I\!R^n$ into $I\!R \cup \{+\infty\}$ which are uniformly semiconcave with constant β.

LEMMA 3.4. *Let $\phi \in \mathcal{S}_\beta$ (still with $C+\beta < 0$), and let a be the semiconvex dual of ϕ. Then $a \in \mathcal{S}^-_d$ for some $d \in \mathcal{D}_n$ such that $C + d < 0$.*

Proof. A proof only in the case $\phi \in C^2$ is provided; in the more general case, a mollification argument can be employed.

Noting that $\phi \in \mathcal{S}_\beta$ and $-C-\beta > 0$, there exists a unique minimizer,

$$\overline{x}(z) = \operatorname*{argmin}_{x \in I\!R}[\phi(x) - \psi(x,z)],$$

and one has

$$a(z) = \phi(\overline{x}(z)) - \psi(\overline{x}(z), z). \tag{3.6}$$

Fix any $z, \nu \in I\!R^n$ with $|\nu| = 1$. Define $a^s : I\!R \to I\!R$ and $\overline{x}^s : I\!R \to I\!R^n$ by

$$a^s(\delta) \doteq a(z+\delta\nu) \quad \textit{and} \quad \overline{x}^s(\delta) = \overline{x}(z+\delta\nu). \tag{3.7}$$

We will obtain a lower bound on the second derivative of a^s, and this will prove the result. Differentiating a^s, one has

$$\begin{aligned}\left.\frac{da^s}{d\delta}\right|_{\delta=0} &= \frac{d}{d\delta}\left[\phi(\overline{x}(z+\delta\nu)) - \psi(\overline{x}(z+\delta\nu), z+\delta\nu)\right] \\ &= \operatorname{grad}_x \phi(\overline{x}(z)) \cdot \frac{d\overline{x}^s}{d\delta} - \operatorname{grad}_x \psi(\overline{x}(z), z) \cdot \frac{d\overline{x}^s}{d\delta} - \operatorname{grad}_z \psi(\overline{x}(z), z)\cdot\nu,\end{aligned}$$

which using the fact that $\operatorname{grad}_x \phi(\overline{x}(z)) - \operatorname{grad}_x \psi(\overline{x}(z), z) = 0$,

$$= -\operatorname{grad}_z \psi(\overline{x}(z), z)\cdot\nu.$$

Differentiating again, one finds

$$\left.\frac{d^2a^s}{d\delta^2}\right|_{\delta=0} = -\sum_{i=1}^n \left\{ \sum_{j=1}^n \psi_{z_i x_j}(\overline{x}(z), z)\frac{d\overline{x}_j^s}{d\delta}\nu_i + \sum_{k=1}^n \psi_{z_i z_k}(\overline{x}(z), z)\nu_k\nu_i \right\} \tag{3.8}$$

$$= -\nu^T C\nu + \nu^T C \left.\frac{d\overline{x}^s}{d\delta}\right|_{\delta=0}. \tag{3.9}$$

Now, differentiating both sides of $\operatorname{grad}_x \phi(\overline{x}(z+\delta\nu)) - \operatorname{grad}_x \psi(\overline{x}(z+\delta\nu), z+\delta\nu) = 0$ yields

$$\sum_{j=1}^n \phi_{x_i x_j}\frac{d\overline{x}_j^s}{d\delta} - \sum_{k=1}^n \psi_{x_i x_k}\frac{d\overline{x}_k^s}{d\delta} - \sum_{l=1}^n \psi_{x_i z_l}\nu_l = 0 \qquad \forall\, i,$$

which yields

$$\frac{d\overline{x}^s}{d\delta} = -\left[\phi_{xx}(\overline{x}(z), z) - C\right]^{-1} C\nu. \tag{3.10}$$

Substituting (3.10) into (3.9), one obtains

$$\frac{d^2a^s}{d\delta^2} = \nu^T\left\{-C + C\left[C - \phi_{xx}(\overline{x}(z), z)\right]^{-1} C\right\}\nu. \tag{3.11}$$

Now, $\phi \in \mathcal{S}_\beta$ implies $-\phi_{xx}(\overline{x}(z)) - \beta < 0$ which implies that there exists $k_0 > 0$ such that

$$\nu^T\left[-\beta - \phi_{xx}(\overline{x}(z))\right]\nu < -k_0|\nu|^2 \qquad \forall\,\nu.$$

Combining this with the fact that $C + \beta < 0$ implies that

$$\nu^T[C - \phi_{xx}(\overline{x}(z))]\nu < -k_0|\nu|^2 \qquad \forall\,\nu.$$

Now, $C - \phi_{xx}(\overline{x}(z))$ being symmetric, negative definite implies that $C - \phi_{xx}(\overline{x}(z)) = U\Lambda U^T$ for some diagonal Λ (with all diagonal entries negative of course) and some real, unitary U. Consequently, $[C - \phi_{xx}(\overline{x}(z))]^{-1} = U\Lambda^{-1}U^T < 0$, that is negative definite. Then, since $\zeta^T[C - \phi_{xx}(\overline{x}(z))]^{-1}\zeta < 0$ for all $\zeta \in I\!R^n$, $\zeta \neq 0$, one sees that

$$\nu^T C[C - \phi_{xx}(\overline{x}(z))]^{-1} C\nu < 0$$

for all $\nu \in I\!R^n$, $\nu \neq 0$, and so

$$C[C - \phi_{xx}(\overline{x}(z))]^{-1} C < 0. \tag{3.12}$$

Let $d \doteq -C + \frac{1}{2}C[C - \phi_{xx}(\overline{x}(z))]^{-1}C$. Then, by (3.12), $C + d = \frac{1}{2}C[C - \phi_{xx}(\overline{x}(z))]^{-1}C < 0$. (Note that if d is not definite, then by addition of εI for arbitrarily small ε, one can make d definite without violating the inequalities.) Further, by (3.11) and (3.12),

$$\left.\frac{d^2 a^s}{d\delta^2}\right|_{\delta=0} = \nu^T \left[d + \tfrac{1}{2}C[C - \phi_{xx}(\overline{x}(z))]^{-1}C\right]\nu < \nu^T d\nu$$

which yields the result. □

REMARK 3.5. Fix any $\delta > 0$ such that $\widetilde{V} \in \mathcal{G}_\delta$, and let $K_\delta = 2\frac{c_A(\gamma-\delta)^2}{c_\sigma^2}$ so that $0 \le \widetilde{V}(x) \le \frac{K_\delta}{2}|x|^2$. Then, using Lemma 3.4 and the monotonicity of the dual operations, the semiconvex dual, $\widetilde{a}$, of $\widetilde{V}$ is in $\mathcal{S}_d^- \cap \mathcal{G}_\delta^-$ for some $d \in \mathcal{D}_n$ such that $C + d < 0$ where $\mathcal{G}_\delta^-$ is the space of semiconcave functions satisfying

$$0 \le \widetilde{a}(z) \le \tfrac{1}{2} z^T Q_\delta^- z$$

where $Q_\delta^- \doteq C(C - K_\delta I)^{-1} K_\delta (C - K_\delta I)^{-1} C - K_\delta^2 (C - K_\delta I)^{-1} C (C - K_\delta I)^{-1}$, and where the last term on the right is the dual of $\frac{K_\delta}{2}|x|^2$. Further, by the monotonicity of the dual operations, any $a \in \mathcal{G}_\delta^-$ has dual $V \in \mathcal{G}_\delta$.

LEMMA 3.6. *Let $\phi \in \mathcal{S}_\beta$ with semiconvex dual a. Suppose $b \in \mathcal{S}_d^-$ with $C + d < 0$ is such that $\phi = \psi(x, \cdot) \odot b(\cdot)$. Then $b = a$.*

Proof. Note that $-b \in \mathcal{S}_d$. Therefore, for all $y \in I\!R^n$, $-b(y) = \max_{\zeta \in I\!R^n}[\psi(y,\zeta) + \alpha(\zeta)]$ or equivalently,

$$b(y) = -\max_{\zeta \in I\!R^n}[\psi(y,\zeta) + \alpha(\zeta)] \tag{3.13}$$

where for all $\zeta \in I\!R^n$

$$\alpha(\zeta) = -\max_{y \in I\!R^n}[\psi(y,\zeta) + b(y)]$$

which by assumption

$$= -\phi(\zeta). \tag{3.14}$$

Combining (3.13) and (3.14), and then using (3.3), one obtains

$$b(y) = -\max_{\zeta \in I\!R^n}[\psi(y,\zeta) - \phi(\zeta)] = a(y) \qquad \forall\, y \in I\!R^n.$$

□

We will hereafter refer to the uniqueness of the semiconvex dual in the sense of Lemma 3.6 simply as uniqueness of the semiconvex dual. It will be critical to the method that the functions obtained by application of the semigroups to the $\psi(\cdot, z)$ be semiconvex with less concavity than the $\psi(\cdot, z)$ themselves. In other words, we will want for instance $\widetilde{S}_\tau[\psi(\cdot, z)] \in \mathcal{S}_{-(c+\varepsilon I)}$ for some $\varepsilon > 0$. This is the subject of the next theorem. Also, in order to keep the theorem statement clean, we will first make some definitions. Define

$$\lambda_D \doteq \min\{\lambda \in I\!R : \lambda \text{ is an eigenvalue of } D^m,\ m \in \mathcal{M}\}.$$

Note that the finiteness of $\mathcal{M}$ and positive definiteness of the D^m imply that $\lambda_D > 0$. Let

$$I_C \doteq \left\{ C \in \mathcal{D}_n \,\middle|\, \min_{|\nu|=1} \min_{m\in\mathcal{M}} \nu^T \left[{A^m}^T C + C A^m \right] \nu \geq -\lambda_D/4 \right\}.$$

THEOREM 3.7. *Let $C \in I_C$. Then there exists $\overline{\tau} > 0$ and $\eta > 0$ such that for all $\tau \in [0, \overline{\tau}]$*

$$\widetilde{S}_\tau[\psi(\cdot, z)],\ S^m_\tau[\psi(\cdot, z)] \in \mathcal{S}_{-(C+\eta I \tau)}.$$

REMARK 3.8. If we restrict to $C = cI$ for some $c \in I\!R$, then $C \in I_C$ if one takes $|c| \leq \lambda_D/[8 \max_{m\in\mathcal{M}} |A^m|]$, $c \neq 0$, and so the theorem condition can be satisfied.

Proof. We prove the result only for $\widetilde{S}_\tau$. The proof for S^m_τ is nearly identical and slightly simpler.

The first portion of the proof is similar to the proof of Theorem 2.6. Again, fix any $x, \nu \in I\!R^n$ with $|\nu| = 1$ and any $\delta > 0$. Fix $\tau > 0$ (to be specified below), and let $\varepsilon > 0$. Let $w^\varepsilon, \mu^\varepsilon$ be ε–optimal for $\widetilde{S}_\tau[\psi(\cdot, z)](x)$. Specifically, suppose $\widehat{\mathcal{I}}^\psi(x, \tau, w^\varepsilon, \mu^\varepsilon) \geq \widetilde{S}_\tau[\psi(\cdot, z)](x) - \varepsilon$ where

$$\widehat{\mathcal{I}}^\psi(x, \tau, w, \mu) \doteq \int_0^\tau l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\, dt + \psi(\xi_\tau, z) \tag{3.15}$$

and ξ_t satisfies (2.12). For simplicity of notation, let $\widehat{V}^{\tau,\psi} = \widetilde{S}_\tau[\psi(\cdot, z)]$. Then

$$\begin{aligned} &\widehat{V}^{\tau,\psi}(x - \delta\nu) - 2\widehat{V}^{\tau,\psi}(x) + \widehat{V}^{\tau,\psi}(x + \delta\nu) \\ &\geq \widehat{\mathcal{I}}^\psi(x - \delta\nu, \tau, w^\varepsilon, \mu^\varepsilon) - 2\widehat{\mathcal{I}}^\psi(x, \tau, w^\varepsilon, \mu^\varepsilon) + \widehat{\mathcal{I}}^\psi(x + \delta\nu, \tau, w^\varepsilon, \mu^\varepsilon) - 2\varepsilon. \end{aligned} \tag{3.16}$$

Let $\xi^\delta, \xi^0, \xi^{-\delta}, \Delta^+$ be as given in the proof of Theorem 2.6. Note that

$$\psi(\xi^\delta_\tau, z) - 2\psi(\xi^0_\tau, z) + \psi(\xi^{-\delta}_\tau, z) = (\Delta^+_\tau)^T C \Delta^+_\tau. \tag{3.17}$$

Note also that as in the proof of Theorem 2.6,

$$\tfrac{1}{2}\left[\xi^\delta_t D^{\mu^\varepsilon_t} \xi^\delta_t - 2\xi^0_t D^{\mu^\varepsilon_t} \xi^0_t + \xi^{-\delta}_t D^{\mu^\varepsilon_t} \xi^{-\delta}_t \right] = (\Delta^+_t)^T D^{\mu^\varepsilon_t} \Delta^+_t. \tag{3.18}$$

Combining (3.15), (3.16), (3.17) and (3.18), one obtains

$$\begin{aligned} &\widehat{V}^{\tau,\psi}(x - \delta\nu) - 2\widehat{V}^{\tau,\psi}(x) + \widehat{V}^{\tau,\psi}(x + \delta\nu) \\ &\geq \int_0^\tau (\Delta^+_t)^T D^{\mu^\varepsilon_t} \Delta^+_t\, dt + (\Delta^+_\tau)^T C \Delta^+_\tau - 2\varepsilon. \end{aligned} \tag{3.19}$$

Further, noting as before that $\dot{\Delta}^+ = A^{\mu^\varepsilon_t} \Delta^+$, one has

$$\Delta^+_t = \exp\left\{ \int_0^t A^{\mu^\varepsilon_r}\, dr \right\} \delta\nu \doteq \Lambda^\varepsilon_t \delta\nu. \tag{3.20}$$

Combining (3.19) and (3.20), one has

$$\begin{aligned} &\widehat{V}^{\tau,\psi}(x - \delta\nu) - 2\widehat{V}^{\tau,\psi}(x) + \widehat{V}^{\tau,\psi}(x + \delta\nu) \\ &\geq \delta^2 \left\{ \int_0^\tau \nu^T (\Lambda^\varepsilon_t)^T D^{\mu^\varepsilon_t} \Lambda^\varepsilon_t \nu\, dt + \nu^T (\Lambda^\varepsilon_\tau)^T C \Lambda^\varepsilon_\tau \nu \right\} - 2\varepsilon. \end{aligned}$$

However, since $\lambda_D > 0$ and $\Lambda^\varepsilon_0 = I$, there exists $\bar{\tau} > 0$ such that for all $\tau \in (0, \bar{\tau})$,

$$\geq \delta^2 \left[\frac{\lambda_D}{2}\tau + \nu^T C \nu \right] + \delta^2 \left[\nu^T (\Lambda_\tau^\varepsilon)^T C \Lambda_\tau^\varepsilon \nu - \nu^T C \nu \right] - 2\varepsilon. \tag{3.21}$$

Now define $g_t^\nu \doteq \nu^T (\Lambda_t^\varepsilon)^T C \Lambda_t^\varepsilon \nu - \nu^T C \nu$. Noting that $\frac{d}{dt}[\Lambda_t^\varepsilon] = A^{\mu_t^\varepsilon} \Lambda_t^\varepsilon$, one obviously has

$$\frac{dg^\nu}{dt} = \nu^T \left[(\Lambda_t^\varepsilon)^T (A^{\mu_t^\varepsilon})^T C \Lambda_t^\varepsilon + (\Lambda_t^\varepsilon)^T C A^{\mu_t^\varepsilon} \Lambda_t^\varepsilon \right] \nu,$$

and consequently,

$$g_t^\nu = \int_0^t \nu^T \left[(\Lambda_r^\varepsilon)^T (A^{\mu_r^\varepsilon})^T C \Lambda_r^\varepsilon + (\Lambda_r^\varepsilon)^T C A^{\mu_r^\varepsilon} \Lambda_r^\varepsilon \right] \nu \, dr. \tag{3.22}$$

Also, define

$$\bar{g}_t^\nu \doteq \int_0^t \nu^T \left[(A^{\mu_r^\varepsilon})^T C + C A^{\mu_r^\varepsilon} \right] \nu \, dr.$$

Noting that Λ_t^ε is continuous, and that $\Lambda_0^\varepsilon = I$, one sees that there exist $\hat{\delta} > 0$ and $\hat{\tau} > 0$ such that for all $\tau \in (0, \hat{\tau})$,

$$|g_t^\nu - \bar{g}_t^\nu| \leq \frac{\hat{\delta}}{2} t^2 \qquad \forall\, t \in (0, \hat{\tau}). \tag{3.23}$$

Let $\tilde{\tau} = \min\{\bar{\tau}, \hat{\tau}, \frac{\lambda_D}{4\hat{\delta}}\}$. By (3.21) and the definition of g^ν,

$$\widehat{V}^{\tau,\psi}(x - \delta\nu) - 2\widehat{V}^{\tau,\psi}(x) + \widehat{V}^{\tau,\psi}(x + \delta\nu) \geq \delta^2 \nu^T C \nu + \delta^2 \left[\frac{\lambda_D}{2}\tau + \bar{g}_\tau^\nu - |g_\tau^\nu - \bar{g}_\tau^\nu| \right] - 2\varepsilon$$

which by the definition of $\bar{g}^\nu$ and (3.23)

$$\geq \delta^2 \nu^T C \nu + \delta^2 \left[\frac{\lambda_D}{2}\tau + \int_0^\tau \nu^T \left[(A^{\mu_r^\varepsilon})^T C + C A^{\mu_r^\varepsilon} \right] \nu \, dr - \frac{\hat{\delta}}{2}\tau^2 \right] - 2\varepsilon,$$

which by the definition of $\tilde{\tau}$ and the assumption that $C \in \mathcal{I}_C$

$$\geq \delta^2 \nu^T C \nu + \delta^2 \frac{\lambda_D \tau}{8} - 2\varepsilon \qquad \forall\, \tau \in (0, \tilde{\tau}).$$

Since this is true for all $\varepsilon > 0$, letting $\eta = \lambda_D/8$, one has

$$\widehat{V}^{\tau,\psi}(x - \delta\nu) - 2\widehat{V}^{\tau,\psi}(x) + \widehat{V}^{\tau,\psi}(x + \delta\nu) \geq \delta^2 \nu^T [C + \eta I \tau] \nu \qquad \forall\, \tau \in (0, \tilde{\tau}).$$

□

COROLLARY 3.9. *We may choose $C \in \mathcal{D}_n$ such that $\widetilde{V}, V^m \in \mathcal{S}_{-C}$, and such that with $\psi, \overline{\tau}, \eta$ as in the statement of Theorem 3.7,*

$$\widetilde{S}_\tau[\psi(\cdot, z)], S_\tau^m[\psi(\cdot, z)] \in \mathcal{S}_{-(C+\eta I \tau)} \qquad \forall\, \tau \in [0, \overline{\tau}].$$

Henceforth, we suppose C chosen so that the results of Corollary 3.9 hold. We also suppose τ, η chosen according to the corollary as well.

Now for each $z \in \mathbb{R}^n$, $\widetilde{S}_\tau[\psi(\cdot, z)] \in \mathcal{S}_{-(C+\eta I \tau)}$. Therefore, by Theorem 3.2

$$\widetilde{S}_\tau[\psi(\cdot, z)](x) = \int_{\mathbb{R}^n}^{\oplus} \psi(x, y) \otimes \widetilde{\mathcal{B}}_\tau(y, z) \, dy = \psi(x, \cdot) \odot \widetilde{\mathcal{B}}_\tau(\cdot, z) \tag{3.24}$$

where for all $y \in \mathbb{R}^n$

$$\widetilde{\mathcal{B}}_\tau(y,z) = -\int_{\mathbb{R}^n}^{\oplus} \psi(x,y)\otimes\left\{-\widetilde{S}_\tau[\psi(\cdot,z)](x)\right\}dx = \left\{\psi(\cdot,y)\odot[\widetilde{S}_\tau[\psi(\cdot,z)](\cdot)]^-\right\}^-. \tag{3.25}$$

It is handy to define the max-plus linear operator with "kernel" $\widetilde{\mathcal{B}}_\tau$ (where we do not rigorously define the term kernel as it will not be needed here) as $\widehat{\widetilde{\mathcal{B}}}_\tau[\alpha](z) \doteq \widetilde{\mathcal{B}}_\tau(z,\cdot)\odot\alpha(\cdot)$ for all $\alpha\in\mathcal{S}_{-C}$.

PROPOSITION 3.10. *Let $\phi\in\mathcal{S}_{-C}$ with semiconvex dual denoted by a. Define $\phi^1 = \widetilde{S}_\tau[\phi]$. Then $\phi^1\in\mathcal{S}_{-(C+\eta I\tau)}$, and*

$$\phi^1(x) = \psi(x,\cdot)\odot a^1(\cdot)$$

where

$$a^1(x) = \widetilde{\mathcal{B}}_\tau(x,\cdot)\odot a(\cdot).$$

Proof. The proof that $\phi^1\in\mathcal{S}_{-(C+\eta I\tau)}$ is similar to the proof in Theorem 3.7. Consequently, we prove only the second assertion.

$$\begin{aligned}\phi^1(x) &= \sup_{w\in\mathcal{W}}\sup_{\mu\in\mathcal{D}_\infty}\left[\int_0^\tau l^{\mu_t}(\xi_t)-\frac{\gamma^2}{2}|w_t|^2\,dt+\phi(\xi_\tau)\right]\\ &= \sup_{w\in\mathcal{W}}\sup_{\mu\in\mathcal{D}_\infty}\max_{z\in\mathbb{R}^n}\left[\int_0^\tau l^{\mu_t}(\xi_t)-\frac{\gamma^2}{2}|w_t|^2\,dt+\psi(\xi_\tau,z)+a(z)\right]\\ &= \max_{z\in\mathbb{R}^n}\left\{\widetilde{S}_\tau[\psi(\cdot,z)](x)+a(z)\right\}\end{aligned}$$

which by (3.24)

$$\begin{aligned}&= \max_{z\in\mathbb{R}^n}\max_{y\in\mathbb{R}^n}\left\{\psi(x,y)+\widetilde{\mathcal{B}}_\tau(y,z)+a(z)\right\}\\ &= \int_{y\in\mathbb{R}^n}^{\oplus}\int_{z\in\mathbb{R}^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(y,z)\otimes a(z)\,dz\,\otimes\psi(x,y)\,dy\\ &= \int_{y\in\mathbb{R}^n}^{\oplus} a^1(x)\otimes\psi(x,y)\,dy.\end{aligned}$$

□

THEOREM 3.11. *Let $V\in\mathcal{S}_{-C}$, and let a be its semiconvex dual (with respect to ψ). Then $V=\widetilde{S}_\tau[V]$ if and only if*

$$a(z) = \max_{y\in\mathbb{R}^n}\left[\widetilde{\mathcal{B}}_\tau(z,y)+a(y)\right]$$

which of course

$$= \int_{\mathbb{R}^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(z,y)\otimes a(y)\,dy = \widetilde{\mathcal{B}}_\tau(z,\cdot)\odot a(\cdot) = \widehat{\widetilde{\mathcal{B}}}_\tau[a](z) \qquad \forall\, z\in\mathbb{R}^n.$$

Proof. Since a is the semiconvex dual of V, for all $x\in\mathbb{R}^n$,

$$\begin{aligned}\psi(x,\cdot)\odot a(\cdot) &= V(x) = \widetilde{S}_\tau[V](x)\\ &= \widetilde{S}_\tau[\max_{z\in\mathbb{R}^n}\{\psi(\cdot,z)+a(z)\}](x)\\ &= \sup_{w\in\mathcal{W}}\sup_{\mu\in\mathcal{D}_\infty}\left[\int_0^\tau l^{\mu_t}(\xi_t)-\frac{\gamma^2}{2}|w_t|^2\,dt+\max_{z\in\mathbb{R}^n}\{\psi(\xi_\tau,z)+a(z)\}\right]\end{aligned}$$

$$= \max_{z\in I\!\!R^n}\left[a(z) + \sup_{w\in\mathcal{W}}\sup_{\mu\in\mathcal{D}_\infty}\left\{\int_0^\tau l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt + \psi(\xi_\tau, z)\right\}\right]$$
$$= \max_{z\in I\!\!R^n}\left\{a(z) + \widetilde{S}_\tau[\psi(\cdot,z)](x)\right\}$$
$$= \int_{I\!\!R^n}^{\oplus} a(z)\otimes\widetilde{S}_\tau[\psi(\cdot,z)](x)\,dz$$

which by (3.24)

$$= \int_{I\!\!R^n}^{\oplus} a(z)\otimes\int_{I\!\!R^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(y,z)\otimes\psi(x,y)\,dy\,dz$$
$$= \int_{I\!\!R^n}^{\oplus}\int_{I\!\!R^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(y,z)\otimes a(z)\otimes\psi(x,y)\,dy\,dz$$
$$= \int_{I\!\!R^n}^{\oplus}\left[\int_{I\!\!R^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(y,z)\otimes a(z)\,dz\right]\otimes\psi(x,y)\,dy$$
$$= \left[\int_{I\!\!R^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(\cdot,z)\otimes a(z)\,dz\right]\odot\psi(x,\cdot)$$

where by Proposition 3.10, the first term is in $\mathcal{S}_{-(C+\eta I\tau)}$. Combining this with Lemma 3.6, one has

$$a(y) = \int_{I\!\!R^n}^{\oplus}\widetilde{\mathcal{B}}_\tau(\cdot,z)\otimes a(z)\,dz = \widetilde{\mathcal{B}}_\tau(y,\cdot)\odot a(\cdot) \qquad \forall\, y\in I\!\!R^n.$$

The reverse implication follows by supposing $a(\cdot) = \widetilde{\mathcal{B}}_\tau(y,\cdot)\odot a(\cdot)$, and reordering the above argument. ∎

Corollary 3.12. *Value function $\widetilde{V}$ is given by $\widetilde{V}(x) = \psi(x,\cdot)\odot\widetilde{a}(\cdot)$ where $\widetilde{a}$ is the unique solution of*

$$\widetilde{a}(y) = \widetilde{\mathcal{B}}_\tau(y,\cdot)\odot\widetilde{a}(\cdot) \qquad \forall\, y \in = I\!\!R^n$$

or equivalently, $\widetilde{a} = \widehat{\widetilde{\mathcal{B}}}_\tau[\widetilde{a}]$.

Proof. Combining Theorem 2.5 and Theorem 3.11 yields the assertion that $\widetilde{V}$ has this representation. The uniqueness follows from the uniqueness assertion of Theorem 2.5 and Lemma 3.6. ∎

Similarly, for each $m\in\mathcal{M}$ and $z\in I\!\!R^n$, $S^m_\tau[\psi(\cdot,z)]\in\mathcal{S}_{-(C+\eta I\tau)}$ and

$$S^m_\tau[\psi(\cdot,z)](x) = \psi(x,\cdot)\odot\mathcal{B}^m_\tau(\cdot,z) \qquad \forall\, x\in I\!\!R^n$$

where

$$\mathcal{B}^m_\tau(y,z) = \left\{\psi(\cdot,y)\odot\left[S^m_\tau[\psi(\cdot,z)]\right]^-(\cdot)\right\}^- \qquad \forall\, y\in I\!\!R^n.$$

As before, it will be handy to define the max-plus linear operator with "kernel" $\mathcal{B}^m_\tau$ as $\widehat{\mathcal{B}}^m_\tau[a](z) \doteq \mathcal{B}^m_\tau(z,\cdot)\odot a(\cdot)$ for all $a\in\mathcal{S}_{-C}$. Further, one also obtains analogous results (by similar proofs). In particular, one has the following

Theorem 3.13. *Let $V\in\mathcal{S}_{-C}$, and let a be its semiconvex dual (with respect to ψ). Then $V = S^m_\tau[V]$ if and only if*

$$a(z) = \mathcal{B}^m_\tau(z,\cdot)\odot a(\cdot) \qquad \forall\, z\in I\!\!R^n.$$

Corollary 3.14. *Each value function V^m is given by $V^m(x) = \psi(x,\cdot)\odot a^m(\cdot)$ where each a^m is the unique solution of the problem $a^m(y) = \mathcal{B}^m_\tau(y,\cdot)\odot a^m(\cdot)$ for all $y\in I\!\!R^n$.*

4. Discrete-time approximation. The method developed here will not involve any discretization over space. Of course this is obvious since otherwise one could not avoid the curse-of-dimensionality. The discretization will be over time where approximate μ processes will be constant over the length of each time-step.

We define the operator $\bar{S}_\tau$ on $\mathcal{G}_\delta$ by

$$\begin{aligned}\bar{S}_\tau[\phi](x) &= \sup_{w\in\mathcal{W}} \max_{m\in\mathcal{M}} \left[\int_0^\tau l^m(\xi_t^m) - \frac{\gamma^2}{2}|w_t|^2\,dt + \phi(\xi_\tau^m)\right](x)\\ &= \max_{m\in\mathcal{M}} S_\tau^m[\phi](x)\end{aligned}$$

where ξ^m satisfies (2.1). Let

$$\overline{\mathcal{B}}_\tau(y,z) \doteq \max_{m\in\mathcal{M}} \mathcal{B}_\tau^m(y,z) = \bigoplus_{m\in\mathcal{M}} \mathcal{B}_\tau^m(y,z) \qquad \forall\, y,z\in I\!R^n.$$

The corresponding max-plus linear operator is

$$\widehat{\overline{\mathcal{B}}}_\tau = \bigoplus_{m\in\mathcal{M}} \widehat{\mathcal{B}}_\tau^m.$$

LEMMA 4.1. *For all $z\in I\!R^n$, $\bar{S}_\tau[\psi(\cdot,z)]\in\mathcal{S}_{-(C+\eta I\tau)}$. Further,*

$$\bar{S}_\tau[\psi(\cdot,z)](x) = \psi(x,\cdot)\odot\overline{\mathcal{B}}_\tau(\cdot,z) \qquad \forall\, x\in I\!R^n. \tag{4.1}$$

Proof. We provide the proof of the last statement, and this is as follows.

$$\begin{aligned}\bar{S}_\tau[\psi(\cdot,z)](x) &= \max_{m\in\mathcal{M}} S_\tau^m[\psi(\cdot,z)](x) = \max_{m\in\mathcal{M}} \psi(x,\cdot)\odot\mathcal{B}_\tau^m(\cdot,z)\\ &= \max_{m\in\mathcal{M}}\max_{y\in I\!R^n}[\psi(x,y)+\mathcal{B}_\tau^m(y,z)] = \max_{y\in I\!R^n}\Big[\psi(x,y)+\max_{m\in\mathcal{M}}\mathcal{B}_\tau^m(y,z)\Big]\\ &= \psi(x,\cdot)\odot\Big[\max_{m\in\mathcal{M}}\mathcal{B}_\tau^m(\cdot,z)\Big].\end{aligned}$$

□

We remark that, parameterized by τ, the operators $\bar{S}_\tau$ do not necessarily form a semigroup, although they do form a sub-semigroup (i.e. $\bar{S}_{\tau_1+\tau_2}[\phi](x)\le\bar{S}_{\tau_1}\bar{S}_{\tau_2}[\phi](x)$ for all $x\in I\!R^n$ and all $\phi\in\mathcal{S}_{-C}$). In spite of this, one does have $S_\tau^m\le\bar{S}_\tau\le\widetilde{S}_\tau$ for all $m\in\mathcal{M}$.

With τ acting as a time-discretization step-size, let

$$\begin{aligned}\mathcal{D}_\infty^\tau = \Big\{\mu:[0,\infty)\to\mathcal{M}\mid &\text{ for each } n\in\mathbf{N}\cup\{0\}, \text{ there exists } m_n\in\mathcal{M}\\ &\text{such that } \mu(t)=m_n\ \forall\ t\in[n\tau,(n+1)\tau)\Big\},\end{aligned}$$

and for $T=\bar{n}\tau$ with $\bar{n}\in\mathbf{N}$ define $\mathcal{D}_T^\tau$ similarly but with domain $[0,T)$ rather than $[0,\infty)$. Let $\mathcal{M}^{\bar{n}}$ denote the outer product of $\mathcal{M}$, $\bar{n}$ times. Let $T=\bar{n}\tau$, and define

$$\bar{\bar{S}}_T^\tau[\phi](x) = \max_{\{m_k\}_{k=0}^{\bar{n}-1}\in\mathcal{M}^{\bar{n}}}\left\{\prod_{k=0}^{\bar{n}-1} S_\tau^{m_k}\right\}[\phi](x) = (\bar{S}_\tau)^{\bar{n}}[\phi](0)$$

where the $\prod$ notation indicates operator composition, and the superscript in the last expression indicates repeated application of $\bar{S}_\tau$, $\bar{n}$ times.

We will be approximating $\widetilde{V}$ by solving $V = \bar{S}_\tau[V]$ via its dual problem $a = \widehat{\overline{\mathcal{B}}}_\tau[a]$ for small τ. Consequently, we will need to show that there exists a solution to $V = \bar{S}_\tau[V]$, that the solution is unique, and that it can be found by solving the dual problem. We begin with existence.

THEOREM 4.2. *Let*

$$\overline{V}(x) \doteq \lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[0](x) \tag{4.2}$$

for all $x \in \mathbb{R}^n$ where 0 here represents the zero-function. Then, $\overline{V}$ satisfies

$$V = \bar{S}_\tau[V], \qquad V(0) = 0. \tag{4.3}$$

Further, $0 \le V^m \le \overline{V} \le \widetilde{V}$ for all $m \in \mathcal{M}$, and consequently, $\overline{V} \in \mathcal{G}_\delta$.

Proof. Note that

$$\begin{aligned} V^m(x) &= \lim_{N\to\infty} S^m_{N\tau}[0](x) \le \limsup_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[0] \\ &\le \lim_{N\to\infty} \widetilde{S}_{N\tau}[0](x) = \widetilde{V}(x) \qquad \forall\, x \in \mathbb{R}^n. \end{aligned} \tag{4.4}$$

Also,

$$\begin{aligned} \bar{\bar{S}}^{\tau}_{(N+1)\tau}[0](x) &= \bar{\bar{S}}^{\tau}_{N\tau}[\bar{S}_\tau[0](\cdot)](x) \\ &= \sup_{\hat{w}\in\mathcal{W}} \sup_{\hat{\mu}\in\mathcal{D}_{N\tau}} \int_0^{N\tau} l^{\hat{\mu}_t}(\xi_t) - \frac{\gamma^2}{2}|\hat{w}_t|^2\,dt \\ &\quad + \sup_{w\in\mathcal{W}} \max_{m\in\mathcal{M}} \int_{N\tau}^{(N+1)\tau} l^m(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt \end{aligned} \tag{4.5}$$

which by taking $w \equiv 0$

$$\ge \sup_{\hat{w}\in\mathcal{W}} \sup_{\hat{\mu}\in\mathcal{D}_{N\tau}} \int_0^{N\tau} l^{\hat{\mu}_t}(\xi_t) - \frac{\gamma^2}{2}|\hat{w}_t|^2\,dt = \bar{\bar{S}}^{\tau}_{N\tau}[0](x), \tag{4.6}$$

which implies that $\bar{\bar{S}}^{\tau}_{N\tau}[0](x)$ is a monotonically increasing function of N. Since it is also bounded from above (by (4.4)), one finds

$$V^m(x) \le \lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[0](x) \le \widetilde{V}(x) \qquad \forall\, x \in \mathbb{R}^n \tag{4.7}$$

which also justifies the use of the limit definition of $\overline{V}$ in the statement of the Theorem. In particular, one has $0 \le V^m \le \overline{V} \le \widetilde{V}$, and so $\overline{V} \in \mathcal{G}_\delta$.

Fix any $x \in \mathbb{R}^n$, and suppose there exists $\delta > 0$ such that

$$\overline{V}(x) \le \bar{S}_\tau[\overline{V}](x) - \delta. \tag{4.8}$$

However, by the definition of $\overline{V}$, given any $y \in \mathbb{R}^n$, there exists $N_\delta < \infty$ such that for all $N \ge N_\delta$

$$\overline{V}(y) \le \bar{\bar{S}}^{\tau}_{N_\delta\tau}[0](y) + \delta/4. \tag{4.9}$$

Combining (4.8) and (4.9), one finds after a small bit of work that

$$\overline{V}(x) \le \bar{S}_\tau\big[\bar{\bar{S}}^{\tau}_{N_\delta\tau}[0] + \delta/2\big](x) - \delta$$

which using the max-plus linearity of $\bar{S}_\tau$

$$= \bar{\bar{S}}^{\tau}_{(N_\delta+1)\tau}[0](x) - \delta/2$$

for all $N \geq N_\delta$. Consequently, $\overline{V}(x) \leq \lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[0](x) - \delta/2$ which is a contradiction. Therefore, $\overline{V}(x) \geq \bar{S}_\tau[\overline{V}](x)$ for all $x \in I\!R^n$. The reverse inequality follows in a similar way. Specifically, fix $x \in I\!R^n$ and suppose there exists $\delta > 0$ such that

$$\overline{V}(x) \geq \bar{S}_\tau[\overline{V}](x) + \delta. \tag{4.10}$$

By the monotonicity of $\bar{\bar{S}}^{\tau}_{N\tau}$ with respect to N, for any $N < \infty$,

$$\overline{V}(x) \geq \bar{\bar{S}}^{\tau}_{N\tau}[0](x) \qquad \forall\, x \in I\!R^n.$$

By the monotonicity of $\bar{S}_\tau$ with respect to its argument (i.e. $\phi_1(x) \leq \phi_2(x)$ for all x implying $\bar{S}_\tau[\phi_1](x) \leq \bar{S}_\tau[\phi_2](x)$ for all x), this implies

$$\bar{S}_\tau[\overline{V}] \geq \bar{\bar{S}}^{\tau}_{(N+1)\tau}[0] \qquad \forall\, x \in I\!R^n. \tag{4.11}$$

Combining (4.10) and (4.11) yields

$$\overline{V}(x) \geq \bar{\bar{S}}^{\tau}_{(N+1)\tau}[0](x) + \delta.$$

Letting $N \to \infty$ yields a contradiction, and so $\overline{V} \leq \bar{S}_\tau[\overline{V}]$. □

The following result is immediate.

THEOREM 4.3.

$$\overline{V}(x) = \sup_{\mu\in\mathcal{D}^\tau_\infty} \sup_{w\in\mathcal{W}} \sup_{T\in[0,\infty)} \left[\int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt\right]$$

where ξ_t satisfies (2.12).

THEOREM 4.4. *$\overline{V}(x) - \frac{1}{2}c_V|x|^2$ is strictly convex.*

Proof. The proof is identical to the proof of Theorem 2.6 with the exception that μ^ε is chosen from $\mathcal{D}^\tau_\infty$ instead of $\mathcal{D}_\infty$. □

REMARK 4.5. From the choice of β in Section 3, this immediately implies that $\overline{V} \in \mathcal{S}_\beta$, and of course since $C + \beta < 0$, that $\overline{V} \in \mathcal{S}_{-C}$.

We now address the uniqueness issue. Similar techniques to those used for V^m and $\widetilde{V}$ will prove uniqueness for (4.3) within $\mathcal{G}_\delta$. A slightly weaker type of result under weaker assumptions will be obtained first; this result is similar in form to that of [41].

Suppose $\overline{V}' \neq \overline{V}$, $\overline{V}' \in \mathcal{G}_\delta$ satisfies (4.3). This implies that for all $x \in I\!R^n$ and all $N < \infty$

$$\begin{aligned}\overline{V}'(x) &= \bar{\bar{S}}^{\tau}_{N\tau}[\overline{V}'](x)\\ &= \sup_{w\in\mathcal{W}} \sup_{\mu\in\mathcal{D}^\tau_\infty} \left\{\int_0^{N\tau} l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt + \overline{V}'(\xi_{N\tau})\right\}\end{aligned}$$

which by taking $w^0 \equiv 0$ (with corresponding trajectory denoted by ξ^0)

$$\geq \overline{V}'(\xi^0_{N\tau}). \tag{4.12}$$

However, by (2.12), one has $\dot{\xi}^0 = A^{\mu_t}\xi^0$, and so $|\xi^0_t| \leq e^{-c_A t}|x|$ for all $t \geq 0$ which implies that $|\xi^0_{N\tau}| \to 0$ as $N \to \infty$. Consequently

$$\lim_{N\to\infty} \overline{V}'(\xi^0_{N\tau}) = 0. \tag{4.13}$$

Combining (4.12) and (4.13), one has

$$\overline{V}'(x) \geq 0 \qquad \forall\, x \in I\!R^n. \tag{4.14}$$

Also, by (4.3)

$$\overline{V}'(x) = \lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[\overline{V}'](x) \qquad \forall\, x \in I\!R^n.$$

By (4.14) and the monotonicity of $\bar{\bar{S}}^{\tau}_{N\tau}$ with respect to its argument, this is

$$\geq \lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[0](x) = \overline{V}(x). \tag{4.15}$$

By (4.14), (4.15), one has the uniqueness result analogous to [41], which is as follows.

THEOREM 4.6. *$\overline{V}$ is the unique minimal, nonnegative solution to (4.3).*

The stronger uniqueness statement (making use of the quadratic bound on $l^{\mu_t}(x)$) is as follows. As with $V^m, \widetilde{V}$, the proof is similar to that in [36]. However in this case, there is a small difference in the proof, and this difference requires another lemma. Due to this difference in the case of $\overline{V}$, we include a sketch of the proof (but with the new lemma in full) in Appendix A.

THEOREM 4.7. *$\overline{V}$ is the unique solution of (4.3) within the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$. Further, given any $V \in \mathcal{G}_\delta$, $\lim_{N\to\infty} \bar{\bar{S}}^{\tau}_{N\tau}[V](x) = \overline{V}(x)$ for all $x \in I\!R^n$ (uniformly on compact sets).*

Henceforth, we let $\delta > 0$ be sufficiently small such that $V^m, \widetilde{V}, \overline{V} \in \mathcal{G}_\delta$ for all $m \in \mathcal{M}$.

THEOREM 4.8. *Let $V \in \mathcal{S}_{-C}$, and let a be its semiconvex dual. Then $V = \bar{S}_\tau[V]$ if and only if $a(y) = \overline{\mathcal{B}}_\tau(y,\cdot) \odot a(\cdot)$ for all $y \in I\!R^n$.*

Proof. By the semiconvex duality

$$\begin{aligned} \psi(x,\cdot) \odot a(\cdot) = V(x) &= \bar{S}_\tau[V](x) \\ &= \bar{S}_\tau\Big[\max_{z\in I\!R^n}\{\psi(\cdot,z) + a(z)\}\Big](x) \end{aligned} \tag{4.16}$$

which as in the first part of the proof of Theorem 3.11

$$= \int_{I\!R^n}^{\oplus} a(z) \otimes \bar{S}_\tau[\psi(\cdot,z)](x)\, dz$$

which by Lemma 4.1

$$= \int_{I\!R^n}^{\oplus} a(z) \otimes \int_{I\!R^n}^{\oplus} \psi(x,y) \otimes \overline{\mathcal{B}}_\tau(y,z)\, dy\, dz$$

which as in the latter part of the proof of Theorem 3.11

$$= \left[\int_{I\!R^n}^{\oplus} \overline{\mathcal{B}}_\tau(\cdot,z) \otimes a(z)\, dz\right] \odot \psi(x,\cdot). \tag{4.17}$$

By Lemmas 3.4 and 3.6, this implies

$$a(y) = \overline{\mathcal{B}}_\tau(y,\cdot) \odot a(\cdot) \qquad \forall\, y \in I\!R^n.$$

Alternatively, if $a(y) = \overline{\mathcal{B}}_\tau(y,\cdot) \odot a(\cdot)$ for all y, then

$$V(x) = \psi(x,\cdot) \odot a(\cdot) = \left[\int_{I\!R^n}^{\oplus} \overline{\mathcal{B}}_\tau(\cdot,z) \otimes a(z)\, dz\right] \odot \psi(x,\cdot) \qquad \forall\, x \in I\!R^n,$$

which by (4.16)–(4.17) yields $V = \bar{S}_\tau[V]$. ∎

COROLLARY 4.9. *Value function $\overline{V}$ given by (4.2) is in $\mathcal{S}_\beta \subset \mathcal{S}_{-C}$, and has representation $\overline{V}(x) = \psi(x,\cdot) \odot \overline{a}(\cdot)$ where $\overline{a}$ is the unique solution in $\mathcal{S}_d^- \cap \mathcal{G}_\delta^-$ of*

$$\overline{a}(y) = \overline{\mathcal{B}}_\tau(y,\cdot) \odot \overline{a}(\cdot) \qquad \forall\, y \in I\!R^n \tag{4.18}$$

or equivalently, $\overline{a} = \widehat{\overline{\mathcal{B}}}_\tau[\overline{a}]$.

Proof. The fact that $\overline{V} \in \mathcal{S}_\beta$ follows from Theorem 4.4 and the choice of β. By Theorem 4.7, $\overline{V} \in \mathcal{G}_\delta$ and is the unique solution of (4.3) in $\mathcal{G}_\delta$.

By Theorem 4.8, its semiconvex dual, $\overline{a}$, satisfies (4.18), and by Lemma 3.4, $\overline{a} \in \mathcal{S}_d^-$ for some $d \in \mathcal{D}_n$ such that $C + d < 0$. Suppose there is $\hat{a} \in \mathcal{S}_{\hat{d}}^- \cap \mathcal{G}_\delta^-$ for some $\hat{d} \in \mathcal{D}_n$ such that $C + \hat{d} < 0$, and that $\hat{a}$ satisfies (4.18). Then, by Theorem 4.8 and Remark 3.5, its dual, $\widehat{V}$, is in $\mathcal{G}_\delta$ and $\widehat{V} = \bar{S}_\tau[\widehat{V}]$, $\widehat{V}(0) = 0$. By Theorem 4.7 then, $\widehat{V} = \overline{V}$. By Lemma 3.6, this implies that $\hat{a} = \overline{a}$. ☐

The following result on propagation of the semiconvex dual will also come in handy.

PROPOSITION 4.10. *Let $\phi \in \mathcal{S}_\beta \subset \mathcal{S}_{-C}$ with semiconvex dual denoted by a. Define $\phi^1 = \bar{S}_\tau[\phi]$. Then $\phi^1 \in \mathcal{S}_{-(C+\eta I\tau)}$, and*

$$\phi^1(x) = \psi(x,\cdot) \odot a^1(\cdot)$$

where

$$a^1(y) = \overline{\mathcal{B}}_\tau(y,\cdot) \odot a(\cdot) \qquad \forall\, y \in I\!R^n.$$

Proof. The proof is similar to the proof of Proposition 3.10, and consequently some details are not included. To begin, as in the proof of Proposition 3.10, we note that the proof that $\phi^1 \in \mathcal{S}_{-(C+\eta I\tau)}$ is nearly identical to the proof of Theorem 3.7. In particular, fix any $x, \nu \in I\!R^n$ with $|\nu| = 1$ and any $\delta > 0$. Let $\overline{m} \in \mathcal{M}$ be optimal, and w^ε be ε–optimal, for $\bar{S}_\tau[\phi](x)$. That is, suppose $\overline{\mathcal{I}}^\phi(x,\tau,w^\varepsilon,\overline{m}) \geq \bar{S}_\tau[\phi](x) - \varepsilon$ where

$$\overline{\mathcal{I}}^\phi(x,\tau,w,m) \doteq \int_0^\tau l^{\overline{m}}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\, dt + \phi(\xi_\tau)$$

and ξ satisfies (2.1). Then

$$\begin{aligned} &\bar{S}_\tau[\phi](x-\delta\nu) - 2\bar{S}_\tau[\phi](x) + \bar{S}_\tau[\phi](x+\delta\nu) \\ &\geq \overline{\mathcal{I}}^\phi(x-\delta\nu,\tau,w^\varepsilon,\overline{m}) - 2\overline{\mathcal{I}}^\phi(x,\tau,w^\varepsilon,\overline{m}) + \overline{\mathcal{I}}^\phi(x+\delta\nu,\tau,w^\varepsilon,\overline{m}) - 2\varepsilon. \end{aligned}$$

Let $\xi^\delta, \xi^0, \xi^{-\delta}$ satisfy the dynamics of (2.1) with inputs w^ε and $\overline{m}$, and with initial conditions $\xi_0^\delta = x + \delta\nu$, $\xi_0^0 = x$ and $\xi_0^{-\delta} = x - \delta\nu$, respectively. Letting $\Delta_t^+ \doteq \xi_t^\delta - \xi_t^0$, one finds

$$\phi(\xi_\tau^\delta) - 2\phi(\xi_\tau^0) + \phi(\xi_\tau^{-\delta}) \geq (\Delta_\tau^+)^T C \Delta_\tau^+$$

because $\phi \in \mathcal{S}_{-C}$. One then continues as in the proof of Theorem 3.7, but with $D^{\overline{m}}$ and $A^{\overline{m}}$ replacing $D^{\mu_t^\varepsilon}$ and $A^{\mu_t^\varepsilon}$, respectively. In particular, one has $\Lambda_t^\varepsilon = \exp\{A^{\overline{m}}t\}$. Importantly, one may use the same values of η and τ which were fixed in Section 3.

Now we turn to the second assertion of the proposition. This follows exactly as in the proof of Proposition 3.10 with two minor exceptions: First, the supremum over $\mu \in \mathcal{D}_\infty$ is replaced by a maximum over $m \in \mathcal{M}$. Second, the use of (3.24) is replaced by the invocation of (4.1). ☐

We now show that one may approximate $\widetilde{V}$, the solution of $V = \widetilde{S}_\tau[V]$, to as accurate a level as one desires by solving $V = \bar{S}_\tau[V]$ for sufficiently small τ. Recall that if $V = \bar{S}_\tau[V]$, then it satisfies $V = \bar{\bar{S}}^\tau_{N\tau}[V]$ for all $N > 0$ (while $\widetilde{V}$ satisfies $V = \widetilde{S}_{N\tau}[V]$), and so this is essentially equivalent to introducing a discrete-time $\overline{\mu} \in \mathcal{D}^{\tau}_{N\tau}$ approximation to the μ process in $\widetilde{S}_{N\tau}$. The result will follow easily from the following technical lemma. The lemma uses the particular structure of our example class of problems as given by Assumption Block $(A.m)$. As the proof of the lemma is technical but long, it is delayed to Appendix B.

LEMMA 4.11. *Given $\hat{\varepsilon} \in (0,1]$, $\overline{T} < \infty$, there exist $T \in [\overline{T}/2, \overline{T}]$ and $\tau > 0$ such that*

$$\widetilde{S}_T[V^m](x) - \bar{\bar{S}}^{\tau}_T[V^m](x) \le \hat{\varepsilon}(1+|x|^2) \qquad \forall\, x \in I\!R^n,\ \forall m \in \mathcal{M}.$$

We now obtain the main approximation result.

THEOREM 4.12. *Given $\overline{\varepsilon} > 0$ and $R < \infty$, there exists $\tau > 0$ such that*

$$\widetilde{V}(x) - \overline{\varepsilon} \le \overline{V}(x) \le \widetilde{V}(x) \qquad \forall\, x \in \overline{B}_R(0).$$

Proof. From Theorem 4.2, we have

$$0 \le V^m(x) \le \overline{V}(x) \le \widetilde{V}(x) \le \frac{c_A(\gamma-\delta)^2}{c_\sigma^2}|x|^2 \qquad \forall\, x \in I\!R^n. \tag{4.19}$$

Also, with $T = N\tau$ for any positive integer N,

$$\bar{\bar{S}}^{\tau}_{N\tau}[\phi] \le \widetilde{S}_T[\phi] \qquad \forall\, \phi \in \mathcal{G}_\delta. \tag{4.20}$$

Further, by Theorem 2.5, given $\varepsilon > 0$ and $R < \infty$, there exists $\widehat{T} < \infty$ such that for all $T > \widehat{T}$ and all $m \in \mathcal{M}$

$$\widetilde{S}_T[\widetilde{V}](x) - \varepsilon/2 \le \widetilde{S}_T[V^m](x) \qquad \forall\, x \in \overline{B}_R(0). \tag{4.21}$$

By (4.21) and Lemma 4.11, given $\overline{\varepsilon} > 0$ and $R < \infty$, there exists $T \in [0,\infty)$, $\tau \in [0,T]$ where $T = N\tau$ for some integer N such that for all $|x| \le R$

$$\begin{aligned}\widetilde{V}(x) - \overline{\varepsilon} &= \widetilde{S}_T[\widetilde{V}](x) - \overline{\varepsilon}\\ &\le \widetilde{S}_T[V^m](x) - \overline{\varepsilon}/2\\ &\le \bar{\bar{S}}^{\tau}_T[V^m](x)\end{aligned}$$

where $\hat{\varepsilon}(1+R^2) = \overline{\varepsilon}/2$, and which by (4.19) and the monotonicity of $\bar{\bar{S}}^{\tau}_T[\cdot]$,

$$\le \bar{\bar{S}}^{\tau}_T[\overline{V}](x)$$

which by (4.20)

$$\le \widetilde{S}_T[\overline{V}](x)$$

which by the monotonicity of $\widetilde{S}_T[\cdot]$

$$\le \widetilde{S}_T[\widetilde{V}](x) = \widetilde{V}(x).$$

Noting (from Theorem 4.7) that $\overline{V} = \bar{\bar{S}}^{\tau}_T[\overline{V}]$ completes the proof. ☐

REMARK 4.13. For this class of systems (defined by Assumption Block $(A.m)$), we expect this result could be sharpened to

$$\widetilde{V}(x) \le -\hat{\varepsilon}(1+|x|^2) \le \overline{V}(x) \le \widetilde{V}(x) \qquad \forall\, x \in I\!R^n$$

by sharpening Theorem 2.5. However, this type of result might only be valid for limited classes of systems, and so we have not pursued it here.

5. The algorithm. We now begin discussion of the actual algorithm.

Let $C \in \mathcal{I}_C$ such that $C - c_V I < 0$, and initialize with $\overline{V}^0(x) \doteq \frac{c_V}{2}|x|^2$. From Theorem 4.2, $\overline{V} = \lim_{N\to\infty} \bar{\bar{S}}^\tau_{N\tau}[\overline{V}^0]$. Given $\overline{V}^k$, let

$$\overline{V}^{k+1} \doteq \bar{S}_\tau[\overline{V}^k]$$

so that $\overline{V}^k = \bar{\bar{S}}^\tau_{k\tau}[\overline{V}^0]$ for all $k \ge 1$.

Let $\overline{a}^k$ be the semiconvex dual of $\overline{V}^k$ for all k. Since $\overline{V}^0 = \frac{c_V}{2}|x|^2$, one easily finds the quadratic $\overline{a}^0(\cdot)$. Note also that by Proposition 4.10,

$$\overline{a}^{k+1} = \overline{\mathcal{B}}_\tau(x,\cdot) \odot \overline{a}^k(\cdot) = \widehat{\overline{\mathcal{B}}}_\tau[\overline{a}^k]$$

for all $n \ge 0$.

Recall that

$$\begin{aligned}\overline{\mathcal{B}}_\tau(x,\cdot) \odot \overline{a}^k(\cdot) &= \int^\oplus_{I\!R^n} \overline{\mathcal{B}}_\tau(x,y) \otimes \overline{a}^k(y)\,dy = \int^\oplus_{I\!R^n} \bigoplus_{m\in\mathcal{M}} \mathcal{B}^m_\tau(x,y) \otimes \overline{a}^k(y)\,dy \\ &= \bigoplus_{m\in\mathcal{M}} \int^\oplus_{I\!R^n} \mathcal{B}^m_\tau(x,y) \otimes \overline{a}^k(y)\,dy = \bigoplus_{m\in\mathcal{M}} \left[\mathcal{B}^m_\tau(x,\cdot) \odot \overline{a}^k(\cdot)\right]. \end{aligned} \tag{5.1}$$

By (5.1),

$$\overline{a}^1(x) = \bigoplus_{m\in\mathcal{M}} \widehat{a}^1_m(x)$$

where

$$\widehat{a}^1_m(x) \doteq \mathcal{B}^m_\tau(x,\cdot) \odot \overline{a}^0(\cdot) \qquad \forall\, m. \tag{5.2}$$

By (5.1) and (5.2),

$$\begin{aligned}\overline{a}^2(x) &= \bigoplus_{m_2\in\mathcal{M}} \int^\oplus_{I\!R^n} \mathcal{B}^{m_2}_\tau(x,y) \otimes \left[\bigoplus_{m_1\in\mathcal{M}} \widehat{a}^1_{m_1}(y)\right] dy \\ &= \bigoplus_{\{m_1,m_2\}\in\mathcal{M}\times\mathcal{M}} \int^\oplus_{I\!R^n} \mathcal{B}^{m_2}_\tau(x,y) \otimes \widehat{a}^1_{m_1}(y)\,dy.\end{aligned}$$

Consequently,

$$\overline{a}^2(x) = \bigoplus_{\{m_1,m_2\}\in\mathcal{M}^2} \widehat{a}^2_{\{m_1,m_2\}}(x)$$

where

$$\widehat{a}^2_{\{m_1,m_2\}}(x) \doteq \mathcal{B}^{m_2}_\tau(x,\cdot) \odot \widehat{a}^1_{m_1}(\cdot) \qquad \forall\, m_1, m_2 \tag{5.3}$$

and $\mathcal{M}^2$ represents the outer product $\mathcal{M}\times\mathcal{M}$. Proceeding with this, one finds that in general,

$$\overline{a}^k(x) = \bigoplus_{\{m_i\}_{i=1}^k\in\mathcal{M}^k} \widehat{a}^k_{\{m_i\}_{i=1}^k}(x)$$

where

$$\widehat{a}^k_{\{m_i\}_{i=1}^k}(x) \doteq \mathcal{B}^{m_k}_\tau(x,\cdot) \odot \widehat{a}^{k-1}_{\{m_i\}_{i=1}^{k-1}}(\cdot) \qquad \forall\, \{m_i\}_{i=1}^k \in \mathcal{M}^k. \tag{5.4}$$

Of course one can obtain $\overline{V}^k$ from its dual as

$$\begin{aligned}
\overline{V}^k(x) &= \max_{y\in I\!\!R^n}[\psi(x,y)+\overline{a}^k(y)] \\
&= \max_{y\in I\!\!R^n}\left[\psi(x,y)+\max_{\{m_i\}_{i=1}^k\in\mathcal{M}^k}\widehat{a}^k_{\{m_i\}_{i=1}^k}(y)\right] \\
&= \max_{\{m_i\}_{i=1}^k\in\mathcal{M}^k}\left\{\max_{y\in I\!\!R^n}[\psi(x,y)+\widehat{a}^k_{\{m_i\}_{i=1}^k}(y)]\right\} \\
&\doteq \max_{\{m_i\}_{i=1}^k\in\mathcal{M}^k}\widehat{V}^k_{\{m_i\}_{i=1}^k}(x) \qquad (5.5)
\end{aligned}$$

where

$$\widehat{V}^k_{\{m_i\}_{i=1}^k}=\max_{y\in I\!\!R^n}[\psi(x,y)+\widehat{a}^k_{\{m_i\}_{i=1}^k}(y)]=\int_{I\!\!R^n}^{\oplus}\psi(x,y)\otimes\widehat{a}^k_{\{m_i\}_{i=1}^k}(y)\,dy. \qquad (5.6)$$

The algorithm will consist of the forward propagation of the $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ (according to (5.4)) from $k=0$ to some termination step $k=N$, followed by construction of the value as $\widehat{V}^k_{\{m_i\}_{i=1}^k}$ (according to (5.6)).

It is important to note that the computation of each $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ is analytical. We will indicate the actual analytical computations.

By the linear/quadratic nature of the m-indexed systems, we find that the $S^m_\tau[\psi(\cdot,z)]$ take the form

$$S^m_\tau[\psi(\cdot,z)](x)=\tfrac{1}{2}(x-\Lambda^m_\tau z)^T P^m_\tau(x-\Lambda^m_\tau)+\tfrac{1}{2}z^T R^m_\tau z$$

where the time-dependent $n\times n$ matrices P^m_t, Λ^m_t and R^m_t satisfy $P^m_0=C$, $\Lambda^m_0=I$, $R^m_0=0$,

$$\begin{aligned}
\dot{P}^m &= (A^m)^T P^m+P^m A^m-[D^m+P^m\Sigma^m P^m] \\
\dot{\Lambda}^m &= \left[(P^m)^{-1}D^m-A^m\right]\Lambda^m \\
\dot{R}^m &= (\Lambda^m)^T D^m\Lambda^m.
\end{aligned} \qquad (5.7)$$

We note that each of the $P^m_\tau,\Lambda^m_\tau,R^m_\tau$ need only be computed once.

Next one computes each quadratic function $\mathcal{B}^m_\tau(x,z)$ (one time only) as follows. One has

$$\mathcal{B}^m_\tau=-\max_{y\in I\!\!R^n}\{\psi(y,x)-S^m_\tau[\psi(\cdot,z)](y)\}$$

which by the above,

$$=\min_{y\in I\!\!R^n}\left\{\tfrac{1}{2}(y-x)^T C(y-x)+\tfrac{1}{2}(y-\Lambda^m_\tau z)^T P^m_\tau(y-\Lambda^m_\tau z)+\tfrac{1}{2}z^T R^m_\tau z\right\}. \qquad (5.8)$$

Recall that by Theorem 3.7, this has a finite minimum ($P^m-(C+\eta I\tau)$ positive definite). Taking the minimum in (5.8), one has

$$\mathcal{B}^m_\tau(x,z)=\tfrac{1}{2}\left[x^T M^m_{1,1}x+x^T M^m_{1,2}z+z^T(M^m_{1,2})^T x+z^T M^m_{2,2}z\right]$$

where with shorthand notation $D_\tau\doteq(P^m_\tau-C)^{-1}$,

$$M^m_{1,1}=\left[CD_\tau^{-1}P^m_\tau D_\tau^{-1}C-(D_\tau^{-1}C+I)^T C(D_\tau^{-1}C+I)\right] \qquad (5.9)$$

$$M^m_{1,2}=\left[(D_\tau^{-1}C+I)^T CD_\tau^{-1}P^m_\tau-CD_\tau^{-1}P^m_\tau(D_\tau^{-1}P^m_\tau-I)\right]\Lambda^m_\tau \qquad (5.10)$$

$$M^m_{2,2}=(\Lambda^m_\tau)^T\left[(D_\tau^{-1}P^m_\tau-I)^T P^m_\tau(D_\tau^{-1}P^m_\tau-I)-P^m_\tau D_\tau^{-1}CD_\tau^{-1}P^m_\tau\right]\Lambda^m_\tau+R^m_\tau. \qquad (5.11)$$

Note that given the $P^m_\tau,\Lambda^m_\tau,R^m_\tau$, the $\mathcal{B}^m_\tau$ are quadratic functions with analytical expressions for their coefficients. Also note that all the matrices in the definition of $\mathcal{B}^m_\tau$ may be precomputed.

Now let us write the (quadratic) $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ in the form

$$\widehat{a}^k_{\{m_i\}_{i=1}^k}(x) = \tfrac{1}{2}(x - \widehat{z}^k_{\{m_i\}_{i=1}^k})^T \widehat{Q}^k_{\{m_i\}_{i=1}^k}(x - \widehat{z}^k_{\{m_i\}_{i=1}^k}) + \widehat{r}^k_{\{m_i\}_{i=1}^k}.$$

Then, for each m_{k+1},

$$\begin{aligned}
\widehat{a}^{k+1}_{\{m_i\}_{i=1}^{k+1}} &= \max_{z\in \mathbb{R}^n}\left\{\mathcal{B}^{m_{k+1}}_\tau(x,z) + \widehat{a}^k_{\{m_i\}_{i=1}^k}(z)\right\} \\
&= \max_{z\in \mathbb{R}^n}\Big\{\tfrac{1}{2}\left[x^T M^m_{1,1}x + x^T M^m_{1,2}z + z^T (M^m_{1,2})^T x + z^T M^m_{2,2} z\right] \\
&\quad + \tfrac{1}{2}(x - \widehat{z}^k_{\{m_i\}_{i=1}^k})^T \widehat{Q}^k_{\{m_i\}_{i=1}^k}(x - \widehat{z}^k_{\{m_i\}_{i=1}^k}) + \widehat{r}^k_{\{m_i\}_{i=1}^k}\Big\} \\
&= \tfrac{1}{2}(x - \widehat{z}^{k+1}_{\{m_i\}_{i=1}^{k+1}})^T \widehat{Q}^{k+1}_{\{m_i\}_{i=1}^{k+1}}(x - \widehat{z}^{k+1}_{\{m_i\}_{i=1}^{k+1}}) + \widehat{r}^{k+1}_{\{m_i\}_{i=1}^{k+1}} \qquad (5.12)
\end{aligned}$$

where

$$\begin{aligned}
\widehat{Q}^{k+1}_{\{m_i\}_{i=1}^{k+1}} &= M^{m_{k+1}}_{1,1} - M^{m_{k+1}}_{1,2}\widehat{D}\left(M^{m_{k+1}}_{1,2}\right)^T \qquad (5.13)\\
\widehat{z}^{k+1}_{\{m_i\}_{i=1}^{k+1}} &= -\left(\widehat{Q}^{k+1}_{\{m_i\}_{i=1}^{k+1}}\right)^{-1} M^{m_{k+1}}_{1,2}\widehat{E} \\
\widehat{r}^{k+1}_{\{m_i\}_{i=1}^{k+1}} &= \widehat{r}^k_{\{m_i\}_{i=1}^k} + \tfrac{1}{2}\widehat{E}^T M^m_{2,2}\widehat{z}^k_{\{m_i\}_{i=1}^k} - \tfrac{1}{2}\left(\widehat{z}^{k+1}_{\{m_i\}_{i=1}^{k+1}}\right)^T \widehat{Q}^{k+1}_{\{m_i\}_{i=1}^{k+1}}\widehat{z}^{k+1}_{\{m_i\}_{i=1}^{k+1}} \\
\widehat{D} &= \left(M^{m_{k+1}}_{2,2} + \widehat{Q}^k_{\{m_i\}_{i=1}^k}\right)^{-1} \\
\widehat{E} &= \widehat{D}\widehat{Q}^k_{\{m_i\}_{i=1}^k}\widehat{z}^k_{\{m_i\}_{i=1}^k}.
\end{aligned}$$

Thus we have the analytical expression for the propagation of each (quadratic) $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ function. Specifically, we see that the propagation of each $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ amounts to a set of matrix multiplications (and an inverse). Note that for purely quadratic constituent Hamiltonians considered here (without terms that are linear or constant in the state and gradient variables), one will have $\widehat{z}^k_{\{m_i\}_{i=1}^k} = 0$ and $\widehat{r}^k_{\{m_i\}_{i=1}^k} = 0$, and so computation of these terms is not necessary (unless one adds linear and/or constant terms).

At each step, k, the semiconvex dual of $\overline{V}^k$, $\overline{a}^k$, is represented as the finite set of functions

$$\widehat{\mathcal{A}}_k \doteq \left\{\widehat{a}^k_{\{m_i\}_{i=1}^k} \,|\, m_i \in \mathcal{M}\ \forall i \in \{1,2,\ldots,k\}\right\}.$$

where this is equivalently represented as the set of triples

$$\widehat{\mathcal{Q}}_k \doteq \left\{\left(\widehat{Q}^k_{\{m_i\}_{i=1}^k}, \widehat{z}^k_{\{m_i\}_{i=1}^k}, \widehat{r}^k_{\{m_i\}_{i=1}^k}\right) \,|\, m_i \in \mathcal{M}\ \forall i \in \{1,2,\ldots,k\}\right\}.$$

At any desired stopping time, one can recover a representation of $\overline{V}^k$ as

$$\widehat{\mathcal{V}}_k \doteq \left\{\widehat{V}^k_{\{m_i\}_{i=1}^k} \,|\, m_i \in \mathcal{M}\ \forall i \in \{1,2,\ldots,k\}\right\}$$

where these $\widehat{V}^k_{\{m_i\}_{i=1}^k}$ are also quadratics. In fact, recall

$$\begin{aligned}
\overline{V}^k(x) &= \max_{z\in\mathbb{R}^n}[\overline{a}^k(z) + \psi(x,z)] \\
&= \max_{\{m_i\}_{i=1}^k}\max_{z\in\mathbb{R}^n}\left[\tfrac{1}{2}(z - \widehat{z}^k_{\{m_i\}_{i=1}^k})^T \widehat{Q}^k_{\{m_i\}_{i=1}^k}(z - \widehat{z}^k_{\{m_i\}_{i=1}^k}) + \widehat{r}^k_{\{m_i\}_{i=1}^k} + \tfrac{c}{2}|x-z|^2\right] \\
&\doteq \max_{\{m_i\}_{i=1}^k}\tfrac{1}{2}(x - \widehat{x}^k_{\{m_i\}_{i=1}^k})^T \widehat{P}^k_{\{m_i\}_{i=1}^k}(x - \widehat{x}^k_{\{m_i\}_{i=1}^k}) + \widehat{\rho}^k_{\{m_i\}_{i=1}^k}
\end{aligned}$$

$$\doteq \bigoplus_{\{m_i\}_{i=1}^k} \widehat{V}^k_{\{m_i\}_{i=1}^k}(x)$$

where with $C \doteq cI$

$$\widehat{P}^k_{\{m_i\}_{i=1}^k} = C\widehat{F}\widehat{Q}^k_{\{m_i\}_{i=1}^k}\widehat{F}C + (\widehat{F}C - I)^T C(\widehat{F}C - I) \tag{5.14}$$

$$\widehat{x}^k_{\{m_i\}_{i=1}^k} = -(\widehat{P}^k_{\{m_i\}_{i=1}^k})^{-1}\left[C\widehat{F}\widehat{Q}^k_{\{m_i\}_{i=1}^k}\widehat{G} + (\widehat{F}C - I)^T C\widehat{F}\widehat{Q}^k_{\{m_i\}_{i=1}^k}\right]\widehat{z}^k_{\{m_i\}_{i=1}^k}$$

$$\widehat{\rho}^k_{\{m_i\}_{i=1}^k} = \widehat{r}^k_{\{m_i\}_{i=1}^k} + \tfrac{1}{2}(\widehat{z}^k_{\{m_i\}_{i=1}^k})^T\left[\widehat{G}^T\widehat{Q}^k_{\{m_i\}_{i=1}^k}\widehat{G} + \widehat{Q}^k_{\{m_i\}_{i=1}^k}\widehat{F}C\widehat{F}\widehat{Q}^k_{\{m_i\}_{i=1}^k}\right]\widehat{z}^k_{\{m_i\}_{i=1}^k}$$

$$\widehat{F} \doteq (\widehat{Q}^k_{\{m_i\}_{i=1}^k} + C)^{-1}$$

and

$$\widehat{G} \doteq (\widehat{F}\widehat{Q}^k_{\{m_i\}_{i=1}^k} - I).$$

Thus, $\overline{V}^k$ has the representation as the set of triples

$$\mathcal{P}_k \doteq \left\{\left(\widehat{P}^k_{\{m_i\}_{i=1}^k}, \widehat{x}^k_{\{m_i\}_{i=1}^k}, \widehat{\rho}^k_{\{m_i\}_{i=1}^k}\right) \mid m_i \in \mathcal{M}\ \forall i \in \{1,2,\ldots,k\}\right\}. \tag{5.15}$$

We note that the triples which comprise $\mathcal{P}_k$ are obtained from the triples $(\widehat{Q}^k_{\{m_i\}_{i=1}^k}, \widehat{z}^k_{\{m_i\}_{i=1}^k}, \widehat{r}^k_{\{m_i\}_{i=1}^k})$ by matrix multiplications and an inverse. The transference from triples $(\widehat{Q}^k_{\{m_i\}_{i=1}^k}, \widehat{z}^k_{\{m_i\}_{i=1}^k}, \widehat{r}^k_{\{m_i\}_{i=1}^k})$ to triples $(\widehat{P}^k_{\{m_i\}_{i=1}^k}, \widehat{x}^k_{\{m_i\}_{i=1}^k}, \widehat{\rho}^k_{\{m_i\}_{i=1}^k})$ need only be done once which is at the termination of the algorithm propagation. Again, in the purely quadratic class of problems addressed here, and with the pure quadratic initialization, the $\widehat{x}^k_{\{m_i\}_{i=1}^k}$ and $\widehat{\rho}^k_{\{m_i\}_{i=1}^k}$ terms will be zero. We note that (5.15) is our approximate solution of the original control problem/HJB PDE.

The errors are due to our approximation of $\widetilde{V}$ by $\overline{V}$ (see Theorem 4.12 and Remark 4.13), and to the approximation of $\overline{V}$ by the prelimit $\overline{V}^N$ for stopping time $k = N$. Neither of these errors are related to the space dimension. The errors in $|\widetilde{V} - \overline{V}|$ are dependent on the step size τ. The errors in $|\overline{V}^N - \overline{V}| = |\bar{\bar{S}}^\tau_{N\tau}[0] - \overline{V}|$ are due to premature termination in the limit $\overline{V} = \lim_{N\to\infty} \bar{\bar{S}}^\tau_{N\tau}[0]$. The computation of each triple $(\widehat{P}^k_{\{m_i\}_{i=1}^k}, \widehat{x}^k_{\{m_i\}_{i=1}^k}, \widehat{\rho}^k_{\{m_i\}_{i=1}^k})$ grows like the cube of the space dimension (due to the matrix operations). Thus one avoids the curse-of-dimensionality. Of course if one then chooses to compute $\overline{V}^N(x)$ for all x on some grid over say a rectangular region in $I\!R^n$, then by definition one has exponential growth in this computation as the space dimension increases. The point is that one does not need to compute $\overline{V}^N \simeq \widetilde{V}$ at each such point.

However, the curse-of-dimensionality is replaced by another type of rapid computational cost growth. Here, we refer to this as the curse-of-complexity. If $\#\mathcal{M} = 1$, then all the computations of our algorithm (excepting the solution of the Riccati equation) are unnecessary, and we *informally* refer to this as complexity one. When there are $M = \#\mathcal{M}$ such quadratics in the Hamiltonian, $\widetilde{H}$, we say it has complexity M. Note that

$$\#\left\{\widehat{V}^k_{\{m_i\}_{i=1}^k} \mid m_i \in \mathcal{M}\ \forall i \in \{1,2,\ldots,k\}\right\} \sim M^N.$$

For large N, this is indeed a large number. (We very briefly discuss means for reducing this in the next section.) Nevertheless, for small values of M, we obtain a very rapid solution of such nonlinear HJB PDEs, as will be indicated in the examples to follow. Further, the computational cost growth in space dimension n is limited to cubic growth. We emphasize that the existence of an algorithm avoiding the curse-of-dimensionality is significant regardless of the practical issues.

6. Practical issues. The bulk of this paper develops an algorithm which avoids the curse-of-dimensionality. However, the curse-of-complexity is also a formidable barrier. The purpose of the paper is to bring the existence of this class of algorithms to light. Considering the long development of finite element methods, it is clear that the development of highly efficient methods from this new class could be a further substantial achievement. (Nevertheless, some impressive computational times are indicated in the next section.) In this section, we briefly indicate some practical heuristics that have been helpful, and outline the actual steps in an implementation of the basic algorithm.

6.1. Pruning. The number of quadratics in $\mathcal{Q}_k$ grows exponentially in k. However, in practice (for the cases we have tried) we have found that relatively few of these actually contribute to $\overline{V}^k$. Thus it would be very useful to prune the set.

Note that if

$$\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}(x) \le \bigoplus_{\{m_i\}_{i=1}^k \ne \{\hat{m}_i\}_{i=1}^k} \widehat{a}^k_{\{m_i\}_{i=1}^k}(x) \qquad \forall\, x \in I\!R^n, \tag{6.1}$$

then

$$\int_{I\!R^n}^{\oplus} \overline{\mathcal{B}}_\tau(x,z) \otimes \overline{a}^k(z)\,dz \le \int_{I\!R^n}^{\oplus} \overline{\mathcal{B}}_\tau(x,z) \otimes \left[\bigoplus_{\{m_i\}_{i=1}^k \ne \{\hat{m}_i\}_{i=1}^k} \widehat{a}^k_{\{m_i\}_{i=1}^k}(z) \right] dz.$$

Consequently $\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}$ will play no role whatsoever in the computation of $\overline{V}^k$. Further, it is easy to show that the progeny of $\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}$ (i.e. those $\widehat{a}^{k+j}_{\{m_i\}_{i=1}^{k+j}}$ for which $\{m_i\}_{i=1}^k = \{\hat{m}_i\}_{i=1}^k$) never contribute either. Thus, one may prune such $\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}$ without any loss of accuracy. This shrinks not only the current $\mathcal{Q}_k$, but also the growth of the future $\mathcal{Q}_{k+j}$.

In the examples to follow, we pruned $\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}$ if there existed a single sequence $\{\tilde{m}_i\}_{i=1}^k$ such that $\widehat{a}^k_{\{\hat{m}_i\}_{i=1}^k}(x) \le \widehat{a}^k_{\{\tilde{m}_i\}_{i=1}^k}(x)$ for all x. This significantly reduced the growth in the size of $\mathcal{Q}_k$. However, it clearly failed to prune anywhere near the number of elements that could be pruned according to condition (6.1), and thus much greater computational reduction might be possible. This would require an ability to determine when a quadratic was dominated by the maximum of a set of other quadratic functions.

Also in the examples to follow, an additional heuristic pruning technique was applied employed for a number of iterations to delay hitting the curse-of-complexity growth rate. A function $\widehat{a}^k_{\{m_i\}_{i=1}^k}$ was pruned if it did not dominate at at least one of the corners of the unit cube. Specifically, let $\mathcal{C} = \{x^j\}$ be the corners of the unit cube. The set of functions was pruned down to a subset of $L \le 2^n$ functions, $\{\widehat{a}^k_{\{\hat{m}^l_i\}_{i=1}^k} \,|\, l \le L\}$, such that $\overline{a}^k(x^j) = \max_{l \le L} \widehat{a}^k_{\{\hat{m}^l_i\}_{i=1}^k}(x^j)$ for all $x^j \in \mathcal{C}$. This introduces a component of the calculations which is subject to curse-of-dimensionality growth, but in the examples run so far is reduced the computations over what they were needed without the heuristic. (Also, the curse-of-dimensionality growth due to this heuristic is 2^n rather than on the order of 200^n as in the discussion of other methods in Section 1.)

6.2. Initialization. It is also easy to see that one may initialize with an arbitrary quadratic function less than an $\overline{a}^k(x)$ rather than with $\overline{a}^0 \equiv 0$. Significant savings were obtained by initializing with a set of $M = \#\mathcal{M}$ quadratics, $\{a^m(x)\}$ where the a^m were the convex duals of the V^m (which were each obtained by solution of the corresponding Riccati equation). With $\overline{a}^0(z) \doteq \bigoplus_{m \in \mathcal{M}} a^m(z)$, one starts much closer to the final solution, and so the number of steps where one is encountering the curse-of-complexity is greatly reduced.

6.3. Pseudo-Code for the Algorithm. In this short section, we briefly indicate the actual steps that one would code in an instantiation of the algorithm.

1. Choose a time-step size, τ, and number of steps, K. (We do not address error analysis and stopping-time criteria in this first paper.)
2. For each $m \in \mathcal{M}$, compute P_τ^m from (5.7). Next, for each $m \in \mathcal{M}$, compute $M_{1,1}^m$, $M_{1,2}^m$ and $M_{2,2}^m$ from (5.9)–(5.11). These are used in each iteration update below.
3. Initialize the iteration. One may initialize with $\overline{a}^0(x) \doteq 0$, which is $\widehat{\mathcal{Q}}_0 = \{\widehat{Q}_1^0\}$ with $\widehat{Q}_1^0 = 0$ (the $n \times n$ matrix of zeros). Note that in this pseudo-code, we will index the $\widehat{Q}^k$ by a generic subscript rather than by the sequences $\{m_i\}_{i=1}^k$, as this is more convenient in software. Although this is a simple initialization, the computational time is hugely improved through the use of the initialization described in Section 6.2. In this latter case, we first compute (approximately) the $P_\infty^m \doteq \lim_{t\to\infty} P_t^m$ from (5.7). The initialization is then $\widehat{\mathcal{Q}}_0 = \{\widehat{Q}_j^0\}_{j=1}^M$ where each $\widehat{Q}_j^0$ is obtained from the corresponding P_∞^j, by the dual operation, and in particular is given by

$$\widehat{Q}_j^0 = C(C - P_\infty^j)^{-1} P_\infty^j (C - P_\infty^j)^{-1} C - P_\infty^j (C - P_\infty^j)^{-1} C (C - P_\infty^j)^{-1} P_\infty^j.$$

4. Perform the basic iteration step. That is, given $\widehat{\mathcal{Q}}_k = \{\widehat{Q}_j^k\}_{j=1}^{J_k}$, compute $\widehat{\mathcal{Q}}_{k+1}$ as follows.
   (a) Start with $j = 1$ and $m = 1$. Let $\ell = 1$.
   (b) (Iteration-subloop): Obtain $\widehat{Q}_\ell^{k+1}$ from update equation (5.14), that is

$$\widehat{Q}_\ell^{k+1} = M_{1,1}^m - M_{1,2}^m \left(M_{2,2}^m + \widehat{Q}_j^k \right)^{-1} \left(M_{1,2}^m \right)^T.$$

   (c) Let $\ell = \ell + 1$. If $m < M$, set $m = m + 1$, and go to step 4b. If $m = M$ and $j < J_k$, set $m = 1$, $j = j + 1$, and go to step 4b. If $m = M$ and $j = J_k$, set $J_{K+1} = \ell - 1$; we are done with the iteration step.
5. Repeat step 4, K times.
6. Recover the solution approximation from the dual matrices. That is, given $\widehat{\mathcal{Q}}_K = \{\widehat{Q}_j^K\}_{j=1}^{J_K}$, compute $\mathcal{P}_k = \{\widehat{P}_j^k\}_{j=1}^{J_K}$ from (5.14). The solution approximation is the pointwise maximum $\overline{V}(x) = \max_{j \le J_K} \frac{1}{2} x^T \widehat{P}_j^K x$.

REMARK 6.1. We emphasize that pruning techniques, such as those of Section 6.1 are critical to rapid computational rates, but this is still an open area of research, and we leave instantiation of such to the intrepid researcher.

7. Examples. A number of examples have so far been tested. In these tests, the computational speeds were very great. This is due to the fact that $M = \#\mathcal{M}$ was small. The algorithm as described above was coded in MATLAB. This includes the very simple pruning technique and initialization discussed in the previous section. The quoted computational times were obtained with a standard 2001 PC. The times correspond to the time to compute $\mathcal{V}_N$ or, equivalently, $\mathcal{P}_N$. The plots below require one to compute the value function and/or gradients pointwise on planes in the state space. These plotting computations are not included in the quoted computational times.

We will briefly indicate the results of three similar examples with state space dimensions of 2, 3 and 4. The number of constituent linear/quadratic Hamiltonians for each of them is 3. The structures of the dynamics are similar for each of them so as to focus on the change in dimension.

Example 1: The first case has constituent Hamiltonians with the A^m given by

$$A^1 = \begin{bmatrix} -1.0 & 0.5 \\ 0.1 & -1.0 \end{bmatrix}, \quad A^2 = (A^1)^T, \quad A^3 = \begin{bmatrix} -1.0 & 0.5 \\ 0.5 & -1.9 \end{bmatrix}.$$

The D^m and Σ^m are simply

$$D^1 = D^2 = D^3 = \begin{bmatrix} 1.5 & 0.2 \\ 0.2 & 1.5 \end{bmatrix},$$

and

$$\Sigma^1 = \Sigma^2 = \Sigma^3 = \begin{bmatrix} 0.27 & -0.01 \\ -0.01 & 0.27 \end{bmatrix}.$$

Figure 7.1 depicts the value function and first partial derivative (computed by a simple first-difference on the grid points) over the region $[-1,1] \times [-1,1]$. Note the discontinuity in the first partial along one of the diagonals. Figure 7.2 depicts the second partial and a backsubstitution error over the same region. The second partial also has a discontinuity along the same diagonal as the first. The error plot has been rotated for better viewing due to the high error along the discontinuity in the gradient. The backsubstitution error is computed by taking these approximate partials and substituting them back into the original HJB PDE. Consequently the depicted errors contain components due to the approximate gradient dotted in with the dynamics and the term with the square in the gradient in the Hamiltonian. Perhaps it should be noted that the solutions of such problems *cannot* be obtained by patching together the quadratic functions corresponding to solutions of the corresponding algebraic Riccati equations. The computations required slightly less than 10 seconds.

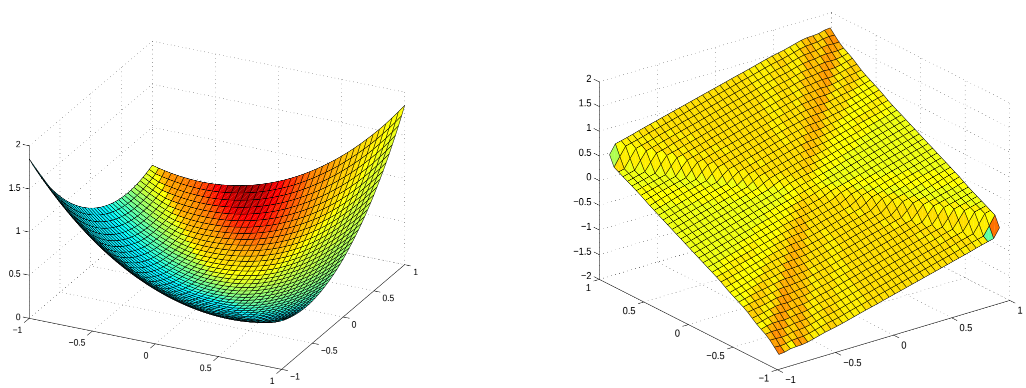

FIG. 7.1. *Value function and first partial (2-D case)*

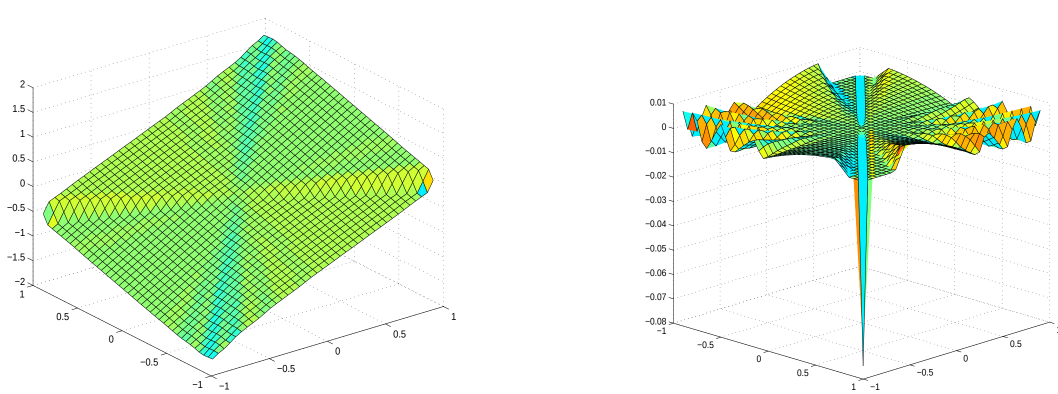

FIG. 7.2. *Second partial and backsubstitution error (2-D case)*

Example 2: We now consider a case where the A^m are given by

$$A^1 = \begin{bmatrix} -1.0 & 0.5 & 0.0 \\ 0.1 & -1.0 & 0.2 \\ 0.2 & 0.0 & -1.5 \end{bmatrix}, \quad A^2 = (A^1)^T, \quad A^3 = \begin{bmatrix} -1.0 & 0.5 & 0.0 \\ 0.1 & -1.0 & 0.2 \\ 0.2 & 0.0 & -1.5 \end{bmatrix},$$

the D^m are

$$D^1 = \begin{bmatrix} 1.5 & 0.2 & 0.1 \\ 0.2 & 1.5 & 0.0 \\ 0.1 & 0.0 & 1.5 \end{bmatrix}, \quad D^2 = \begin{bmatrix} 1.6 & 0.2 & 0.1 \\ 0.2 & 1.6 & 0.0 \\ 0.1 & 0.0 & 1.6 \end{bmatrix}, \quad D^3 = D^1,$$

and the Σ^m are

$$\Sigma^1 = \begin{bmatrix} 0.2 & -0.01 & 0.02 \\ -0.01 & 0.2 & 0.0 \\ 0.02 & 0.0 & 0.25 \end{bmatrix}, \quad \Sigma^2 = \begin{bmatrix} 0.16 & -0.005 & 0.015 \\ -0.005 & 0.16 & 0.0 \\ 0.015 & 0.0 & 0.2 \end{bmatrix}, \quad \Sigma^3 = \Sigma^1.$$

The results of this three-dimensional example appear in Figures 7.3–7.5. In this case, the results have been plotted over the region of the affine plane $x_3 = 3$ given by $x_1 \in [-10, 10]$ and $x_2 \in [-10, 10]$. The backsubstitution error has been scaled by dividing by $|x|^2 + 10^{-5}$. Note that the scaled backsubstitution errors (away from the discontinuity in the gradient) grow only slowly or are possibly bounded with increasing $|x|$. (Recall that the approximate solution is obtained over the whole space.) Since the gradient errors are multiplied by the nominal dynamics in one component of this term (as well as being squared in another), this indicates that the errors in the gradient itself likely grow only linearly (or nearly linearly) with increasing $|x|$. The computations required approximately 13 seconds.

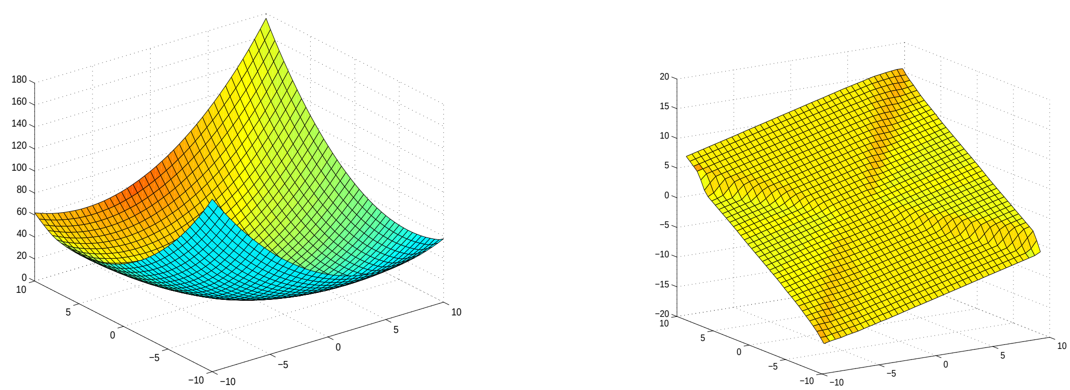

FIG. 7.3. *Value function and first partial (3-D case)*

Example 3: The four-dimensional example has constituent Hamiltonians with the A^m, D^m and Σ^m given by

$$A^1 = \begin{bmatrix} -1.0 & 0.5 & 0.0 & 0.1 \\ 0.1 & -1.0 & 0.2 & 0.0 \\ 0.2 & 0.0 & -1.5 & 0.1 \\ 0.0 & -0.1 & 0.0 & -1.5 \end{bmatrix}, \quad A^2 = (A^1)^T,$$

$$A^3 = \begin{bmatrix} -1.0 & 0.5 & 0.0 & 0.1 \\ 0.1 & -1.0 & 0.2 & 0.0 \\ 0.2 & 0.0 & -1.6 & -0.1 \\ 0.0 & -0.05 & 0.1 & -1.5 \end{bmatrix},$$

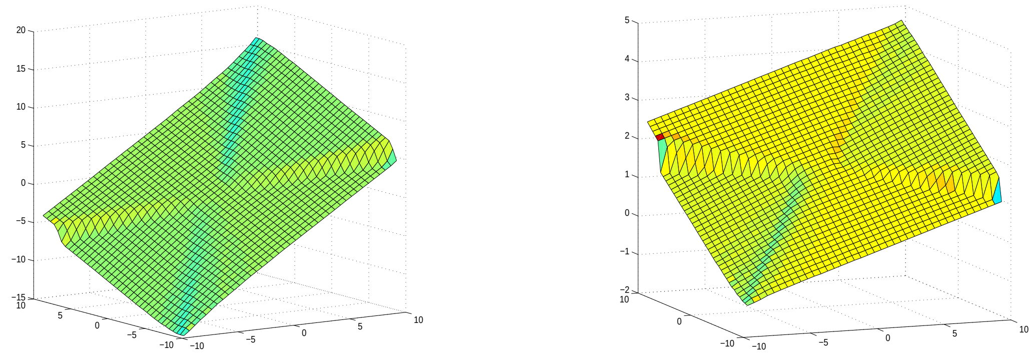

FIG. 7.4. *Second and third partials (3-D case)*

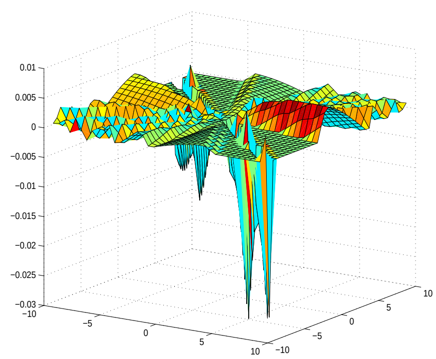

FIG. 7.5. *Scaled backsubstitution error (3-D case)*

$$D^1 = D^2 = D^3 = \begin{bmatrix} 1.5 & 0.2 & 0.1 & 0.0 \\ 0.2 & 1.5 & 0.0 & 0.1 \\ 0.1 & 0.0 & 1.5 & 0.0 \\ 0.0 & 0.1 & 0.0 & 1.5 \end{bmatrix},$$

and

$$\Sigma^1 = \Sigma^2 = \Sigma^3 = \begin{bmatrix} 0.2 & -0.01 & 0.02 & 0.01 \\ -0.01 & 0.2 & 0.0 & 0.0 \\ 0.02 & 0.0 & 0.25 & 0.0 \\ 0.01 & 0.0 & 0.0 & 0.25 \end{bmatrix}.$$

The results for this example appear in Figures 7.6–7.8. In this case, the results have been plotted over the region of the affine plane $x_3 = 3$, $x_4 = -0.5$ given by $x_1 \in [-10, 10]$ and $x_2 \in [-10, 10]$. The backsubstitution error has again been scaled by dividing by $|x|^2 + 10^{-5}$. The computations required approximately 40 seconds. We remark that one cannot change dimension independent of dynamics (except in the trivial case where each component of the system has exactly the same dynamics of the other components with no interdependence), and so one cannot directly compare the computation times of these three examples. However, it is easy to see that the computation time increases are on the order of square to cubic in space dimension, rather than being subject to curse-of-dimensionality type growth.

8. Future directions.

Pruning:. In order to make these methods more practical, algorithms need to be developed for determining when a quadratic function is dominated by the function which is the pointwise maximum

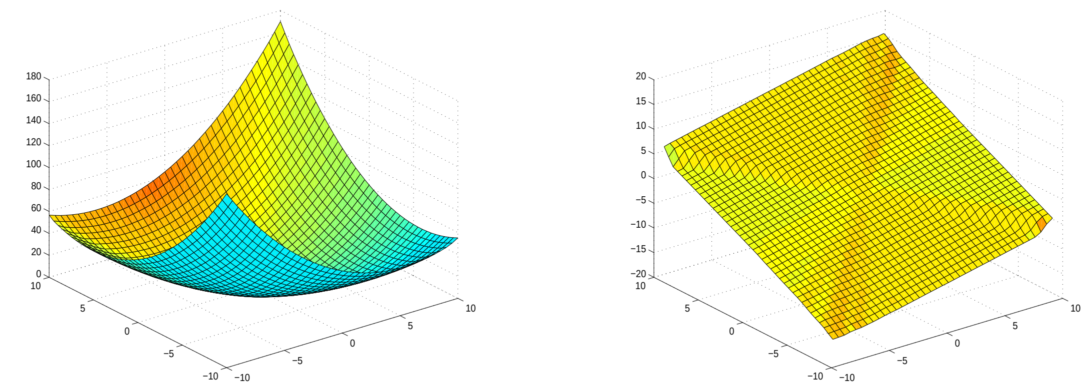

FIG. 7.6. *Value function and first partial (4-D case)*

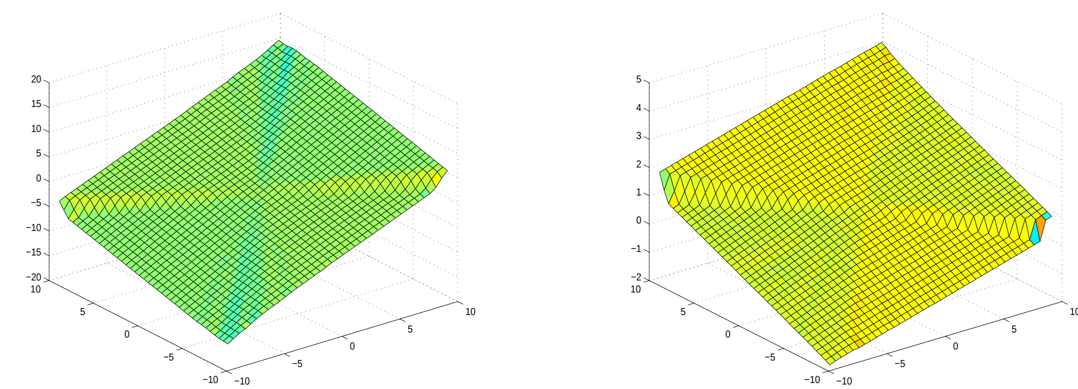

FIG. 7.7. *Second and third partials (4-D case)*

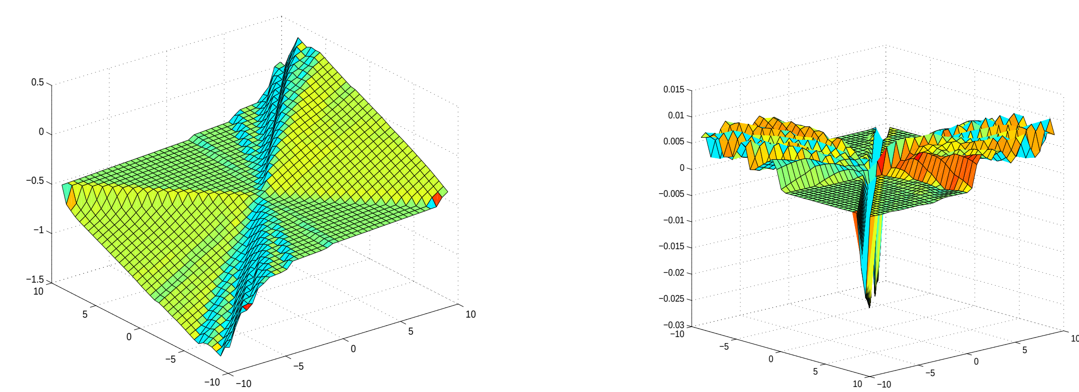

FIG. 7.8. *Fourth partial and scaled backsubstitution error (4-D case)*

of a set of quadratic functions. This has the potential for greatly reducing the effects of the curse-of-complexity, and consequently greatly decreasing computational times.

Constant/Linear Terms:. An instantiation of this class of methods was developed here for a very particular type of Hamiltonian, $\widetilde{H}(x,p) = \max_m\{H^m(x,p)\}$, where the H^m corresponded to a very specific type of linear/quadratic problem. One would like to generalize the H^m to say

$$H^m(x,p) = \tfrac{1}{2}x^T D^m x + \tfrac{1}{2}p^T \Sigma^m p + (A^m x)^T p + (l_1^m)^T x + (l_2^m)^T p + \alpha^m.$$

Clearly certain conditions on $\widetilde{H}(x,p) = \max_m\{H^m(x,p)\}$ would be necessary. It is not obvious that these conditions would need to apply to each of the constituent H^m individually. In the work here, the H^m corresponded to linear/quadratic problems with maximizing controllers/disturbances. It is not clear that the constituent linear/quadratic problems need to be constricted in this way either. For instance, could some or all of the H^m correspond to say game problems?

Convergence/Error Analysis:. Only convergence of the approximation to the solution was obtained here. Estimates of error size and convergence rate need to be determined. For instance, it was hypothesized (and observed in the examples) that one obtains the solution over the whole state space with linear growth rate in the errors in the gradient. Is this true in any generality?

Non-Ergodic Problem:. The algorithm was developed for an infinite time-horizon problem where the dynamics were stable to the origin. One expects the approach would also be applicable to discounted cost problems and exit problems. One would also expect that a similar theory could be developed for finite time-horizon problems such as robust filtering. Max-plus methods have also been discussed for problems corresponding to Variational Inequalities [31]. The analysis and algorithm necessary for a Variational Inequality would be of interest.

Other Nonlinearities:. This work concentrated only on the case of a nonlinearity due to taking the maximum of a set of Hamiltonians for linear/quadratic problems. An obvious question is how well this approach might work for other classes of nonlinearities. What classes of nonlinear HJB PDEs could be best approximated by maxima over reasonably small numbers of linear/quadratic HJB PDEs? Perhaps a single nonlinearity in only one variable (possibly appearing in multiple places) would be the most tractable?

Acknowledgments. The author wishes to thank Prof. J. William Helton for helpful discussions, without which this direction may never have been explored.

Appendix A (sketch of proof of Theorem 4.7):

Fix $\delta > 0$ (used in the definition of $\mathcal{G}_\delta$). Suppose $\overline{V}' \in \mathcal{G}_\delta$ satisfies (4.3). Then,

$$\begin{aligned}\overline{V}'(x) &= \bar{\bar{S}}^{\tau}_{N\tau}[\overline{V}'](x)\\ &= \sup_{w\in\mathcal{W}}\sup_{\mu\in\mathcal{D}^{\tau}_{\infty}}\left\{\int_0^{N\tau} l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2\,dt + \overline{V}'(\xi_{N\tau})\right\} \qquad \forall\, x\in I\!R^n\end{aligned}$$

where ξ satisfies (2.12). Fix $x \in I\!R^n$, and let $\mu^\varepsilon \in \mathcal{D}^\tau_\infty$, $w^\varepsilon \in \mathcal{W}$ be ε–optimal , i.e.

$$\overline{V}'(x) \le \int_0^{N\tau} l^{\mu^\varepsilon_t}(\xi^\varepsilon_t) - \frac{\gamma^2}{2}|w^\varepsilon_t|^2\,dt + \overline{V}'(\xi^\varepsilon_{N\tau}) + \varepsilon$$

where ξ^ε satisfies (2.12) with inputs $\mu^\varepsilon, w^\varepsilon$.

Following the same steps as in [36], one obtains the same lemmas:

Lemma A.1 *For any* $N < \infty$, $\|w^\varepsilon\|^2_{L_2(0,N\tau)} \le \frac{\varepsilon}{\delta} + \frac{1}{\delta}\left[\frac{c_A\gamma^2}{c_\sigma^2}e^{-c_A N\tau} + \frac{c_D}{c_A}\right]|x|^2$.

Lemma A.2 *For any* $N < \infty$,

$$\int_0^{N\tau} |\xi^\varepsilon_t|^2\,dt \le \frac{\varepsilon}{\delta}\frac{c_\sigma^2}{c_A} + \frac{c_\sigma^2}{\delta}\left[\left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2}\right) + \frac{1}{c_A}\right]|x|^2.$$

Lemma A.3 *If $w^\varepsilon, \mu^\varepsilon$ are ε–optimal over $[0, N\tau)$, then they are also ε–optimal over $[0, n\tau)$ for all $n \le N$, i.e.*

$$\int_0^{n\tau} l^{\mu_t^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + \overline{V}'(\xi_{n\tau}^\varepsilon) \ge \overline{V}'(x) - \varepsilon.$$

The independence of the above bounds with respect to N is important. Specifically, since there is a finite bound on the energy (the bound on w^ε) coming in to the trajectories, roughly speaking the ξ^ε "tend" toward the origin.

Now we need a lemma which will replace equation (20) in [36].

Lemma A.4 *For any $N < \infty$,*

$$\sum_{n=1}^N |\xi_{n\tau}^\varepsilon|^2 \le \frac{1}{1 - e^{-c_A\tau}}\left[|x|^2 + (c_\sigma/c_A^2)\|w^\varepsilon\|^2_{L_2(0,N\tau)}\right].$$

Proof. Note that $\frac{d}{dt}|\xi^\varepsilon|^2 \le -c_A|\xi^\varepsilon| + \hat{d}|w^\varepsilon|^2$ with $\hat{d} = c_\sigma^2/c_A$. Solving this on intervals of the form $[n\tau, (n+1)\tau)$, one finds

$$|\xi_\tau^\varepsilon|^2 \le |x|^2 e^{-c_A\tau} + \hat{d}\|w^\varepsilon\|^2_{L_2(0,\tau)},$$
$$|\xi_{2\tau}^\varepsilon|^2 \le |\xi_\tau^\varepsilon|^2 e^{-c_A\tau} + \hat{d}\|w^\varepsilon\|^2_{L_2(\tau,2\tau)},$$

and so on. Continuing this process, and combining the inequalities yields

$$\sum_{n=1}^N |\xi_{n\tau}^\varepsilon|^2 \le \left(\sum_{n=1}^N e^{-nc_A\tau}\right)|x|^2 + \hat{d}\sum_{n=1}^N\left[\left(\sum_{j=0}^{N-n} e^{-jc_A\tau}\right)\|w^\varepsilon\|^2_{L_2((n-1)\tau,n\tau)}\right].$$

Using the standard geometric series limit yields the result. □

Combining Lemmas A.2 and A.4, one obtains a bound on $\sum_{n=1}^N |\xi_{n\tau}^\varepsilon|^2$ which is independent of N. Consequently, at least some of the $|\xi_{n\tau}^\varepsilon|$ can be guaranteed to be arbitrarily small for large N. The remainder of the proof (of Theorem 4.7) then follows as in equations (24) to (28) in [36], but with $N\tau$ replacing T, and $n\tau$ replacing τ. This completes the sketch of the proof.

Appendix B (sketch of proof of Lemma 4.11):

Fix $\delta > 0$ (used in the definition of $\mathcal{G}_\delta$). Fix $m \in \mathcal{M}$. Fix any $T < \infty$ and $x \in \mathbb{R}^n$. Let $\varepsilon = (\hat{\varepsilon}/2)(1 + |x|^2)$. Let $w^\varepsilon \in \mathcal{W}$, $\mu^\varepsilon \in \mathcal{D}_\infty$ be ε–optimal for $\widetilde{S}_T[V^m](x)$, i.e.

$$\widetilde{S}_T[V^m](x) - \left[\int_0^T l^{\mu_t^\varepsilon}(\xi^\varepsilon{}_t) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\xi^\varepsilon{}_T)\right] \le \varepsilon = \frac{\hat{\varepsilon}}{2}(1 + |x|^2) \tag{8.1}$$

where ξ^ε satisfies (2.12) with inputs w^ε, μ^ε.

We will let $\overline{\xi}^\varepsilon$ satisfy (2.12) with inputs w^ε and a $\overline{\mu}^\varepsilon \in \mathcal{D}_\infty^\tau$ (where τ has yet to be chosen). Solving (2.12), one has

$$\xi_t^\varepsilon = \exp\left[\int_0^t A^{\mu_r^\varepsilon}\,dr\right]x + \int_0^t \exp\left[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\right]\sigma^{\mu_r^\varepsilon} w_r^\varepsilon\,dr$$
$$\overline{\xi}_t^\varepsilon = \exp\left[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr\right]x + \int_0^t \exp\left[\int_r^t A^{\overline{\mu}_\rho^\varepsilon}\,d\rho\right]\sigma^{\overline{\mu}_r^\varepsilon} w_r^\varepsilon\,dr.$$

Consequently,

$$
\begin{aligned}
|\xi_t^\varepsilon - \overline{\xi}_t^\varepsilon| \le & \Big|\exp\Big[\int_0^t A^{\mu_r^\varepsilon}\,dr\Big] - \exp\Big[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr\Big]\Big|\,|x| \\
& + \Big\{\int_0^t \Big|\exp\Big[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\Big]\sigma^{\mu_r^\varepsilon} - \exp\Big[\int_r^t A^{\overline{\mu}_\rho^\varepsilon}\,d\rho\Big]\sigma^{\overline{\mu}_r^\varepsilon}\Big|^2\,dr\Big\}^{1/2} \|w^\varepsilon\|_{L_2(0,t)}.
\end{aligned} \tag{8.2}
$$

We now simply show that this can be made arbitrarily small by taking τ small. We will use the boundedness of $\|w^\varepsilon\|$ and $\|\xi^\varepsilon\|$ which is independent of t for this class of systems [36].

Consider the first term on the right in (8.2). Note that

$$
\Big|\exp\Big[\int_0^t A^{\mu_r^\varepsilon}\,dr\Big] - \exp\Big[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr\Big]\Big| = \Big|\exp\Big[\int_0^t A^{\mu_r^\varepsilon}\,dr\Big]\Big|\,\Big|1 - \exp\Big[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr - \int_0^t A^{\mu_r^\varepsilon}\,dr\Big]\Big|. \tag{8.3}
$$

Fix $\tau > 0$. For any subset of $I\!R$, $\mathcal{I}$, let $\mathcal{L}(\mathcal{I})$ be the Lebesgue measure of $\mathcal{I}$. Let N be the largest integer such that $N\tau \le t$. Given $m \in \mathcal{M}$, let

$$
\mathcal{I}^m = \{r \in [0, N\tau) \,|\, A^{\mu_r^\varepsilon} = A^m\} \quad \text{and} \quad \lambda^m = \mathcal{L}(\mathcal{I}^m).
$$

Let $n_0 = 0$. For $1 \le k < M = \#\mathcal{M}$, let n_k be the largest integer such that $n_k\tau \le \lambda^k + n_{k-1}\tau$. For $m < M$ let

$$
\overline{\mu}_r^\varepsilon = m \qquad \forall\, t \in [n_{m-1}\tau, n_m\tau).
$$

Let $\overline{\mu}^\varepsilon{}_r = M$ for all $t \in [n_{M-1}\tau, t) = [n_{M-1}\tau, N\tau) \cup [N\tau, t)$. With this choice of $\overline{\mu}^\varepsilon$, one finds

$$
\Big|1 - \exp\Big[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr - \int_0^t A^{\mu_r^\varepsilon}\,dr\Big]\Big| < \beta_\tau^1 \tag{8.4}
$$

where $\beta_\tau^1 \to 0$ as $\tau \to 0$ independent of t. We skip the details.

Let $y \in I\!R^n$. Define $F_t = \exp\big[\int_0^t A^{\mu_r^\varepsilon}\,dr\big]$. Then, using Assumption Block $(A.m)$,

$$
\begin{aligned}
\frac{d}{dt}\,[y^T F_t^T F_t y] &= y^T\,\Big[F_t^T \dot{F}_t + \dot{F}_t^T F_t\Big]\,y = 2y^T\,\Big[F_t^T A^{\mu_t^\varepsilon} F_t\Big]\,y \\
&= 2(F_t y)^T A^{\mu_t^\varepsilon}(F_t y) \le -2c_A|F_t y|^2 = -2c_A\big[y^T F_t^T F_t y\big].
\end{aligned}
$$

Solving this ordinary differential inequality, one finds $\big[y^T F_t^T F_t y\big] \le |y|^2 e^{-2c_A t}$. Since this is true for all $y \in I\!R^n$, we have

$$
\Big|\exp\Big[\int_0^t A^{\mu_r^\varepsilon}\,dr\Big]\Big| \le e^{-c_A t} \qquad \forall\, t \ge 0. \tag{8.5}
$$

By (8.3), (8.4) and (8.5)

$$
\Big|\exp\Big[\int_0^t A^{\mu_r^\varepsilon}\,dr\Big] - \exp\Big[\int_0^t A^{\overline{\mu}_r^\varepsilon}\,dr\Big]\Big| \le \beta_\tau^1 e^{-c_A t} \qquad \forall\, t \ge 0. \tag{8.6}
$$

We now turn to the second term on the right-hand side of (8.2). Note that

$$
\begin{aligned}
& \Big\{\int_0^t \Big|\exp\Big[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\Big]\sigma^{\mu_r^\varepsilon} - \exp\Big[\int_r^t A^{\overline{\mu}_\rho^\varepsilon}\,d\rho\Big]\sigma^{\overline{\mu}_r^\varepsilon}\Big|^2\,dr\Big\}^{1/2} \\
\le & \Big\{2\int_0^t \Big|\exp\Big[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\Big]\Big|^2\,\Big|\sigma^{\mu_r^\varepsilon} - \sigma^{\overline{\mu}_r^\varepsilon}\Big|^2\,dr \\
& \quad +2\int_0^t \Big|\exp\Big[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\Big] - \exp\Big[\int_r^t A^{\overline{\mu}_\rho^\varepsilon}\,d\rho\Big]\Big|^2\,\Big|\sigma^{\overline{\mu}_r^\varepsilon}\Big|^2\,dr\Big\}^{1/2}
\end{aligned}
$$

and proceeding as above

$$\leq \left\{ 2\int_0^t e^{-2c_A(t-r)} \left| \sigma^{\mu_r^\varepsilon} - \sigma^{\overline{\mu}_r^\varepsilon} \right|^2 dr + 2\beta_\tau^1 \int_0^t e^{-2c_A(t-r)} \left| \sigma^{\overline{\mu}_r^\varepsilon} \right|^2 dr \right\}^{1/2}$$

$$\leq \left\{ 2\left[\int_0^t e^{-4c_A(t-r)}\,dr\right]^{1/2} \left[\int_0^t \left| \sigma^{\mu_r^\varepsilon} - \sigma^{\overline{\mu}_r^\varepsilon} \right|^4 dr\right]^{1/2} + 2\beta_\tau^1 c_\sigma^2 \int_0^t e^{-2c_A(t-r)}\,dr \right\}^{1/2}.$$

Further, there exists β_τ^2 such that $[\int_0^t |\sigma^{\mu_r^\varepsilon} - \sigma^{\overline{\mu}_r^\varepsilon}|^4\,dr]^{1/2} \leq \beta_\tau^2$ where $\beta_\tau^2 \to 0$ as $\tau \to 0$, and we skip the obvious but technical proof. Consequently,

$$\left\{ \int_0^t \left| \exp\Big[\int_r^t A^{\mu_\rho^\varepsilon}\,d\rho\Big]\sigma^{\mu_r^\varepsilon} - \exp\Big[\int_r^t A^{\overline{\mu}_\rho^\varepsilon}\,d\rho\Big]\sigma^{\overline{\mu}_r^\varepsilon} \right|^2 dr \right\}^{1/2}$$

$$\leq \left\{ 2\beta_\tau^2 (4c_A)^{-1/2} + 2\beta_\tau^1 c_\sigma^2 (2c_A)^{-1} \right\}^{1/2} \leq \beta_\tau^3 \tag{8.7}$$

where $\beta_\tau^3 \to 0$ as $\tau \to 0$ (independent of t).

Combining (8.2), (8.6) and (8.7), one has

$$|\xi_t^\varepsilon - \overline{\xi}_t^\varepsilon| \leq \beta_\tau^1 e^{-c_A t}|x| + \beta_\tau^3 \|w^\varepsilon\|_{L_2(0,t)}. \tag{8.8}$$

Now, by the system structure given by Assumption Block $(A.m)$ and by the fact that the V^m are in $\mathcal{G}_\delta$, one obtains the following lemmas exactly as in [36]. These are also analogous to their counterparts in Appendix A.

Lemma B.1 *For any* $t < \infty$, $\|w^\varepsilon\|_{L_2(0,t)}^2 \leq \frac{\varepsilon}{\delta} + \frac{1}{\delta}\left[\frac{c_A\gamma^2}{c_\sigma^2} e^{-c_A N\tau} + \frac{c_D}{c_A}\right]|x|^2$.

Lemma B.2 *For any* $t < \infty$,

$$\int_0^t |\xi_r^\varepsilon|^2\,dt \leq \frac{\varepsilon}{\delta}\frac{c_\sigma^2}{c_A} + \frac{c_\sigma^2}{\delta}\left[\left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2}\right) + \frac{1}{c_A}\right]|x|^2.$$

Let $c_1 \doteq \varepsilon/\delta$ and $c_2 \doteq \frac{1}{\delta}[\frac{c_A\gamma^2}{c_\sigma^2} e^{-c_A N\tau} + \frac{c_D}{c_A}]$. By Lemma B.1 and (8.8), for all $t < \infty$ one has

$$|\xi_t^\varepsilon - \overline{\xi}_t^\varepsilon| \leq \beta_\tau^1 e^{-c_A t}|x| + \beta_\tau^3 (c_1 + c_2|x|^2)^{1/2}$$

and by proper choice of β_τ^4,

$$\leq \beta_\tau^4 (1 + |x|) \tag{8.9}$$

where $\beta_\tau^4 \to 0$ as $\tau \to 0$ (independent of $t > 0$).

Now,

$$\int_0^T l^{\mu_t^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\xi_T^\varepsilon) - \int_0^T l^{\overline{\mu}_t^\varepsilon}(\overline{\xi}_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\overline{\xi}_T^\varepsilon)$$

$$= \int_0^T \xi_t^\varepsilon D^{\mu_t^\varepsilon} \xi_t^\varepsilon - \overline{\xi}_t^\varepsilon D^{\overline{\mu}_t^\varepsilon} \overline{\xi}_t^\varepsilon\,dt + (\xi_T^\varepsilon)^T P^m \xi_T^\varepsilon - (\overline{\xi}_T^\varepsilon)^T P^m \overline{\xi}_T^\varepsilon. \tag{8.10}$$

Note that the integral term on the right-hand side in (8.10) is

$$\int_0^T (\xi_t^\varepsilon)^T D^{\mu_t^\varepsilon}(\xi_t^\varepsilon - \overline{\xi}_t^\varepsilon) + (\xi_t^\varepsilon)^T \left(D^{\mu_t^\varepsilon} - D^{\overline{\mu}_t^\varepsilon}\right)\overline{\xi}_t^\varepsilon + (\xi_t^\varepsilon - \overline{\xi}_t^\varepsilon)^T D^{\overline{\mu}_t^\varepsilon}\overline{\xi}_t^\varepsilon\,dt$$

$$\leq \beta_\tau^4(1+|x|)\int_0^T \left(|D^{\mu_t^\varepsilon}|\,|\xi_t^\varepsilon| + |D^{\overline{\mu}_t^\varepsilon}|\,|\overline{\xi}_t^\varepsilon|\right)dt + \beta_\tau^5 \int_0^T |\xi_t^\varepsilon|\,|\overline{\xi}_t^\varepsilon|\,dt$$

for appropriate $\beta_\tau^5 \to 0$ as $\tau \to 0$, which after some work

$$\leq \beta_\tau^6(1+|x|^2)(1+\sqrt{T}) \tag{8.11}$$

for appropriate choice of $\beta_\tau^6 \to 0$ as $\tau \to 0$ (independent of T).

Similarly, the last two terms on the right-hand side in (8.10) are

$$\begin{aligned} {\xi_T^\varepsilon}^T P^m \xi_T^\varepsilon - {\overline{\xi}_T^\varepsilon}^T P^m \overline{\xi}_T^\varepsilon &= (\xi_T^\varepsilon + \overline{\xi}_T^\varepsilon)^T P^m (\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon) \\ &\leq |P^m| \left[|\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon|^2 + 2|\xi_T^\varepsilon|\, |\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon| \right] \end{aligned}$$

which by (8.9)

$$\leq \beta_\tau^7(1+|x|^2) + \beta_\tau^8|\xi_T^\varepsilon|(1+|x|) \tag{8.12}$$

where $\beta_\tau^7, \beta_\tau^8 \to 0$ as $\tau \to 0$.

We also need the following lemma which is obtained in [36].

Lemma B.3 *Given $\overline{T} < \infty$, there exist $T \in [\overline{T}/2, \overline{T}]$ and ε–optimal $w^\varepsilon \in \mathcal{W}$, $\mu^\varepsilon \in \mathcal{D}_\infty$ for $\widetilde{S}_T[V^m]$ such that*

$$|\xi_T^\varepsilon|^2 \leq \frac{1}{\overline{\overline{T}}} \left\{ \frac{\varepsilon}{\delta} \frac{c_\sigma^2}{c_A} + \frac{c_\sigma^2}{\delta} \left[\left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2} \right) + \frac{1}{c_A} \right] |x|^2 \right\}.$$

Combining (8.12) and Lemma B.3, one finds that there exist $c_3, c_4 < \infty$ such that

$${\xi_T^\varepsilon}^T P^m \xi_T^\varepsilon - {\overline{\xi}_T^\varepsilon}^T P^m \overline{\xi}_T^\varepsilon \leq \beta_\tau^9(1+|x|^2) \tag{8.13}$$

where $\beta_\tau^9 \to 0$ as $\tau \to 0$ (independent of T).

Combining (8.10), (8.11) and (8.13),

$$\begin{aligned} &\int_0^T l^{\mu_t^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\xi_T^\varepsilon) - \int_0^T l^{\overline{\mu}_t^\varepsilon}(\overline{\xi}_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\overline{\xi}_T^\varepsilon) \\ &\leq \beta_\tau^{10}(1+|x|^2)(1+\sqrt{T}) \end{aligned} \tag{8.14}$$

where $\beta_\tau^{10} \to 0$ as $\tau \to 0$ (independent of T).

Combining (8.1) and (8.14), one has

$$\widetilde{S}_T[V^m](x) - \int_0^T l^{\overline{\mu}_t^\varepsilon}(\overline{\xi}_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2\,dt + V^m(\overline{\xi}_T^\varepsilon) \leq \frac{\varepsilon}{2}(1+|x|^2) + \beta_\tau^{10}(1+|x|^2)(1+\sqrt{T})$$

which for τ sufficiently small (depending on T now),

$$\leq \varepsilon(1+|x|^2).$$

This completes the proof of Lemma 4.11.

REFERENCES

[1] M. Akian, S. Gaubert and A. Lakhoua, *A max-plus finite element method for solving finite horizon determinsitic optimal control problems,* Proc. 16^{th} International Symposium on Mathematical Theory of Networks and Systems (MTNS), Leuven, 2004.
[2] F.L. Baccelli, G. Cohen, G.J. Olsder and J.-P. Quadrat, *Synchronization and Linearity*, John Wiley, 1992.
[3] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, Boston, 1997.

[4] F. Camilli, M. falcone, P. Lanucara and A. Seghini, *A domain decomposition method for Bellman equations,* Contemp. Math., **180** (1994), 477-483.

[5] E. Carlini, M. Falcone and R. Ferretti, *An efficient algorithm for Hamilton-Jacobi equations in high dimensions,* Computing and Visualization Sci., **7** (2004), 15–29.

[6] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John-Wiley, 1983.

[7] G. Cohen, S. Gaubert and J.-P. Quadrat, *Duality and separation theorems in idempotent semimodules*, Linear Algebra and Applications, **379** (2004), 395–422.

[8] G. Collins and W.M. McEneaney, *Min-plus eigenvector methods for nonlinear H_∞ problems with active control,* Optimal Control, Stabilization and Nonsmooth Analysis, LNCIS, Vol. 301 Queiroz, Marcio S. de; Malisoff, Michael; Wolenski, Peter (Eds.), Springer (2004), pp. 101–120.

[9] M. G. Crandall, H. Ishii and P.-L. Lions, *Uniqueness of viscosity solutions of Hamilton-Jacobi equations revisited,* J. Math. Society Japan, **39** (1987) pp. 581–595.

[10] M. G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations,* Trans. of the Amer. Math. Soc., **277** (1983) pp. 1–42.

[11] M. G. Crandall and P.-L. Lions, *On existence and uniqueness of solutions of Hamilton–Jacobi equations,* Nonlinear Anal. Theory and Appl., **10** (1986) pp. 353–370.

[12] R.A. Cuninghame-Green, *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems 166, Springer, 1979.

[13] P. Dupuis and M. Boué, *Markov chain approximations for deterministic control problems with affine dynamics and quadratic cost in the control,* SIAM J. on Numerical Analysis, **36** (1999), pp. 667–695.

[14] P. Dupuis and A. Szpiro, *Convergence of the optimal feedback policies in a numerical method for a class of deterministic optimal control problems,* SIAM J. Control and Optim. **40** (2001), pp. 393–420.

[15] P. Dupuis and A. Szpiro, *Second order numerical methods for first order Hamilton-Jacobi equations,* SIAM J. on Num. Analysis, **40** (2002), 113-118.

[16] L.C. Evans, *Partial Differential Equations*, Springer–Verlag, 1998.

[17] M. Falcone and R. Ferretti, *Convergence analysis for a class of high-order semi-lagrangian advection schemes,* SIAM J. Num. Analysis, **35** (1998) ,909-940.

[18] W.H. Fleming, *Max-plus stochastic processes*, Applied Math. and Optim., **. 49** (2004), 159–181.

[19] W.H. Fleming, *Functions of Several Variables*, Springer–Verlag, 1977.

[20] W.H. Fleming and W.M. McEneaney, *A max-plus based algorithm for an HJB equation of nonlinear filtering,* SIAM J. Control and Optim., 38 (2000), pp. 683–710.

[21] W.H. Fleming and H.M. Soner, *Controlled Markov Process and Viscosity Solutions*, Springer-Verlag, New York, 1993.

[22] F. John, *Partial Differential Equations*, Springer-Verlag, New York, 1978.

[23] J.W. Helton and M.R. James, *Extending H_∞ Control to Nonlinear Systems, Control of Nonlinear Systems to Achieve Performance Objectives*, Advances in Design and Control, SIAM, Philadelphia,1999.

[24] M.R. James and W.M. McEneaney, *Max–plus approximation methods in partially observed H_∞ control,* 38th IEEE Conf. on Decision and Control, (1999) pp. 3011–3016.

[25] V.N. Kolokoltsov and V.P. Maslov, *Idempotent Analysis and Its Applications,* Kluwer, 1997.

[26] H.J. Kushner and P. Dupuis, *Numerical methods for Stochastic Control Problems in Continuous Time*, Springer Verlag, New York, 1992.

[27] G.L. Litvinov, V.P. Maslov and G.B. Shpiz, *Idempotent functional analysis: An algebraic approach,* Mathematical Notes, **69**, No. 5 (2001), pp. 696–729.

[28] V.P. Maslov, *On a new principle of superposition for optimization problems,* Russian Math. Surveys, **42** (1987) pp. 43–54.

[29] W.M. McEneaney, *Legendre transforms on max-plus spaces as a tool for nonlinear control problems,* Proc. International Symposium on Mathematical Theory of Networks and Systems (MTNS), 2004.

[30] W.M. McEneaney, *Max-plus summation of fenchel-transformed semigroups for solution of nonlinear Bellman equations,* Systems and Control Letters (to appear).

[31] W.M. McEneaney and P.M. Dower, *A max-plus affine power method for approximation of a class of mixed l_∞/l_2 value functions,* Proc. 42nd IEEE Conf. on Dec. and Control, Maui (2003) pp. 2573 - 2578.

[32] W.M. McEneaney, *Max-plus eigenvector representations for solution of nonlinear H_∞ problems: Error analysis,* SIAM J. Control and Optim., Vol. 43 (2004), 379–412.

[33] W.M. McEneaney, *Max-plus eigenvector representations for solution of nonlinear H_∞ problems: Basic concepts,* IEEE Trans. Auto. Control **48**, (2003) pp. 1150–1163.

[34] W.M. McEneaney, *Error analysis of a max-plus algorithm for a first-order HJB equation,* Stochastic Theory and Control, Proceedings of a Workshop held in Lawrence, Kansas, October 18-20, 2001, Lecture Notes in Control and Information Sciences, Springer-Verlag, B.Pasik-Duncan (Ed.) (2002) pp. 335–352.

[35] W.M. McEneaney and M. Horton, *Max–plus eigenvector representations for nonlinear H_∞ value functions,* 37th IEEE Conf. on Decision and Control (1998), pp. 3506–3511.

[36] W.M. McEneaney, *A uniqueness result for the Isaacs equation corresponding to nonlinear H_∞ control,* Math. Controls, Signals and Systems, **11** (1998), pp. 303–334.

[37] W. M. McEneaney and M. V. Day, *Characteristic characterization of viscosity supersolutions corresponding to nonlinear H_∞ control,* 13^{th} World Congress of International Federation of Automatic Control, (1996) pp. 401–406.

[38] A.A. Melikyan, *Generalized Characteristics of First Order Pdes: Applications in Optimal Control and Differential Games*, Birkhauser, Boston, 1998.

[39] R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional conference Series in Applied Mathematics 16, SIAM (1974).

[40] R.T. Rockafellar and R.J. Wets, *Variational Analysis*, Springer–Verlag, 1997.

[41] P. Soravia, *H_∞ control of nonlinear systems: differential games and viscosity solutions,* SIAM J. Control and Optim., **34** (1996), pp. 1071–1097.

[42] A.J. van der Schaft, *L_2-gain analysis of nonlinear systems and nonlinear state feedback H_∞ control,* IEEE Trans. Auto. Control, **37** (1992) pp. 770–784.