A Curse-of-Dimensionality-Free Numerical Method for Solution of Certain HJB PDEs

William M. McEneaney *
Dept. of Mechanical and Aerospace Eng. and Dept. of Math.
Univ. of California, San Diego
La Jolla, CA 92093-0112, USA

June 22, 2004

Abstract

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 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 $H_\infty$ control with fixed-feedback control. 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. 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.

*Dept. of Mechanical and Aerospace Eng. and Dept. of Math., University of California San Diego, wmceneaney@ucsd.edu Research partially supported by NSF grant DMS-0307229.
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 (such as $H_\infty$ control) 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/$H_\infty$ nonlinear filtering and Robust/$H_\infty$ 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], [7], [8], [9], [17]. One approach to solution is through generalized characteristics (cf. [34], [35], as well as [13], [18] for classical general 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 finite element methods (cf. [3], [11], [12], [17], [22] among many others). These require that one generate a grid over some bounded region of the state-space. Suppose the region over which one constructs the grid is rectangular, say square for simplicity. Suppose one uses 100 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 $8 \times 10^6$ in going from a dimension 3 problem to a dimension 6 problem. Consequently, there is no hope that such techniques could be used for problems in dimension greater than 4 or 5 in the foreseeable future.

In recent years, an entirely new class of numerical methods for HJB PDEs has emerged [16], [32], [20], [30], [31], [29], [28], [27], [1], [25], [26]. These methods exploit the max-plus (or min-plus [6], [30]) 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. Much of the work has concentrated on the (harder) steady-state HJB PDE class where (for both max-plus and finite element methods), one propagates forward in “time” to obtain the steady-state limit solution. With the max-plus
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 methods, the “time-step” tends to be much larger than what can be used in finite element 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 the max-plus approach, the curse-of-dimensionality growth is so fast that one cannot expect to solve problems of more than say dimension 4 or 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 finite element or max-plus method over a space of dimension $n$. While the Riccati equation may be “relatively” easily solved for large $n$, the max-plus and finite element methods have no hope for $n=6$. This has been a frustrating, counter-intuitive situation.

This paper discusses an approach to certain nonlinear HJB PDEs which is not subject to the curse-of-dimensionality. In fact, 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 (or approximated by) a maximum or minimum of $M$ linear/quadratic Hamiltonians, then one could say the complexity of the Hamiltonian (or the approximation of the Hamiltonian) is $M$.

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 $\mathbb{R}^4$. A few simple examples comprised of 3 linear/quadratic components were solved in 10-20 seconds over $\mathbb{R}^3$ and 10-40 seconds over $\mathbb{R}^4$. 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 testing and 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 finite element method.

We will be concerned here with HJB PDEs given or approximated as

$$\bar{H}(x, \nabla V) = \max_{m \in \{1,2,\ldots,M\}} \{H^m(x, \nabla V)\}$$

(1)

or
\[
\vec{H}(x, \nabla V) = \min_{m \in \{1,2,\ldots,M\}} \{ H^m(x, \nabla V) \}.
\]

In order to make the problem tractable, we will concentrate on a single class of HJB PDEs of form (1). However, the theory can obviously be expanded to a much larger class.

To give an idea of the proposed method, recall that the solution of (1) is the eigenfunction of the corresponding semigroup, that is

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

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

\[
0 \otimes e = \bar{B}_\tau \circ e
\]

where we use the \( \circ \) notation to indicate \( \bar{B}_\tau \circ e \doteq f_{\bar{B}_\tau} \) \( (x, y) \otimes e(y) \ dy \) where \( f_{\bar{B}_\tau} \) denotes max-plus integration (maximization). Then one approximates \( \bar{B}_\tau \doteq \bigoplus_{m \in \mathcal{M}} B^m_\tau \) where \( \mathcal{M} = \{1,2,\ldots,M\} \) and the \( B^m_\tau \) correspond to the \( H^m \). The power method ([30], [10], [21]) implies that the solution is approximated by the form

\[
e \doteq \lim_{N \to \infty} \left[ \bigoplus_{m \in \mathcal{M}} B^m_\tau \right]^N \oplus 0
\]

where the \( N \) superscript denotes the \( \oplus \) operation \( N \) times, and \( 0 \) represents the zero function. Given linear/quadratic forms for each of the \( H^m \), the \( B^m_\tau \) are obtained by Riccati equations. Let \( e_N \doteq \left[ \bigoplus_{m \in \mathcal{M}} B^m_\tau \right]^N \otimes 0 \). Note that

\[
e_1 = \bigoplus_{m \in \mathcal{M}} B^m_\tau \oplus 0
\]

\[
e_2 = \bigoplus_{(m_1, m_2) \in \mathcal{M} \times \mathcal{M}} B^{m_1, m_2}_\tau \oplus 0 \doteq \left[ \bigoplus_{m_2 \in \mathcal{M}} B^{m_2}_\tau \right] \oplus \left[ \bigoplus_{m_1 \in \mathcal{M}} B^{m_1}_\tau \right] \oplus 0
\]

\[
e_3 = \bigoplus_{(m_1, m_2, m_3) \in \mathcal{M} \times \mathcal{M} \times \mathcal{M}} B^{m_1, m_2, m_3}_\tau \oplus 0 \doteq \left[ \bigoplus_{m_3 \in \mathcal{M}} B^{m_3}_\tau \right] \oplus \left[ \bigoplus_{(m_1, m_2) \in \mathcal{M} \times \mathcal{M}} B^{m_1, m_2}_\tau \right] \oplus 0
\]

and so on. 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 \( M = \#\mathcal{M} \). (However, we remark that \( B^{m_1, m_2, \ldots, m_N}_\tau \oplus 0 \) which are dominated by others can be deleted from the list of such objects without consequence, which can alleviate some of the growth.) The computation of each \( B^{(m_1)}_\tau \) is analytical given the solution of the Riccati equations for the \( H^m \).

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, the last section indicates some future directions.

2 Review of Theory for Sample Problem Class

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 (ITH) $H_\infty$ problem with fixed feedback. This 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

\[
\dot{\xi}^m = A^m \xi^m + \sigma^m w, \\
\xi^m_0 = x \in \mathbb{R}^m.
\]  
(2)

Let $w \in \mathcal{W} \doteq L^\text{loc}_2([0, \infty); \mathbb{R}^m)$ where we recall that $L^\text{loc}_2([0, \infty); \mathbb{R}^m) = \{ w : [0, \infty) \rightarrow \mathbb{R}^m : \int_0^T |w_t|^2 dt < \infty \forall T < \infty \}$. Let the cost functionals be

\[
J^m(x, T; w) = \int_0^T \frac{1}{2} \xi^m_t D^m \xi^m_t - \frac{\gamma^2}{2} |w_t|^2 dt,
\]  
(3)

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

\[
V^m(x) = \sup_{w \in \mathcal{W}} \sup_{T < \infty} J^m(x, T; w) = \lim_{T \to \infty} \sup_{w \in \mathcal{W}} J^m(x, T; w).
\]  
(4)

We remark that a generalization of the second term in the integrand of the cost functional to $\frac{1}{2} w^T \Gamma T \Gamma w$ with $\Gamma T \Gamma$ positive definite is not needed since this is equivalent to a change in $\sigma^m$ in the dynamics (2). 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:
Assume that there exists $c_A \in (0, \infty)$ such that
\[
x^T A^m x \leq -c_A |x|^2 \quad \forall x \in \mathbb{R}^n, m \in \mathcal{M}.
\]
Assume that there exists $c_\sigma < \infty$ such that
\[
|\sigma^m| \leq c_\sigma \quad \forall m \in \mathcal{M}.
\]
(A.m)

Assume that all $D^m$ are positive definite, symmetric, and let $c_D$ be such that
\[
x^T D^m x \leq c_D |x|^2 \quad \forall x \in \mathbb{R}^n, m \in \mathcal{M}
\]

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

Note that these assumptions guarantee the existence of the $V^m$ as locally bounded functions which are zero at the origin (cf. [33]). (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
\[
0 = H^m(x, \nabla V) = \frac{1}{2} x^T D^m x + (A^m x)^T \nabla V + \max_{w \in \mathbb{R}^m} [(\sigma^m w)^T \nabla V - \frac{\gamma^2}{2} |w|^2]
\]
\[= \frac{1}{2} x^T D^m x + (A^m x)^T \nabla V + \frac{1}{2} \nabla V^T \Sigma^m \nabla V \quad \forall x \in \mathbb{R}^n, m \in \mathcal{M}
\]
\[V(0) = 0
\]
where $\Sigma^m = \frac{1}{\gamma^2} \sigma^m (\sigma^m)^T$. Let $\mathcal{G}_\delta$ be the subset of $C(\mathbb{R}^n)$ such that $0 \leq V(x) \leq \frac{c_A(\gamma - \delta)^2}{c_\sigma^2} |x|^2$ for all $x$. From [33] (undoubtedly among many others),

**Theorem 2.1** Each value function (4) is the unique viscosity solution of its corresponding HJB PDE (5) in the class $\mathcal{G}_\delta$ for sufficiently small $\delta > 0$.

Note from (4) that by considering $w$, which are zero for large $t$, each $V^m$ satisfies
\[
V^m(x) = \lim_{T \to \infty} V^{m,f}(x, T) = \lim_{T \to \infty} \sup_{w \in \mathcal{W}} J^m(x, T; w),
\]
and so (c.f. [33]) for each $m$,
\[
V^m(x) = \lim_{T \to \infty} V^{m,f}(x, T) \quad (6)
\]
where $V^{m,f}$ is the unique continuous viscosity solution of (c.f. [3], [17])
\[ 0 = V_T - H^m(x, \nabla V) \]
\[ V(0, x) = 0. \tag{7} \]

It is easy to see that these solutions have the form \( V^{m,f}(x,t) = \frac{1}{2} x^T P^{m,f}_t x \) where each \( P^{m,f} \) satisfies the differential Riccati equation
\[ \dot{P}^{m,f} = (A^m)^T P^{m,f} + P^{m,f} A^m + D^m + P^{m,f} \Sigma^m P^{m,f} \tag{8} \]
\[ P^{m,f}_0 = 0. \]

By (6) and (8), the \( V^m \) take the form \( V^m(x) = \frac{1}{2} x^T P^m x \) where \( P^m = \lim_{t \to \infty} P^{m,f}_t \). With this form, and (5) (or (8)), we see that the \( P^m \) satisfy the algebraic Riccati equations
\[ 0 = (A^m)^T P^m + P^m A^m + D^m + P^m \Sigma^m P^m. \tag{9} \]

Combining this with Theorem 2.1, one has

**Theorem 2.2** Each value function (4) is the unique classical solution of its corresponding HJB PDE (5) in the class \( G_\delta \) 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 (9).

**Corollary 2.3** Each \( V^m \) is convex. Further, there exists symmetric, positive definite \( C \) and \( \varepsilon > 0 \) such that \( V^m(x) - \frac{1}{2} x^T C x \) is convex for all \( m \in M \).

The duality between viscosity (and/or classical) solutions of the HJB PDEs 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
\[ S^m_T[\phi] = \sup_{w \in \mathcal{W}} \left[ \int_0^T \frac{1}{2} (\xi^m_t)^T D^m \xi^m_t - \frac{\gamma^2}{2} |w_t|^2 dt + \phi(\xi^m_T) \right] \tag{10} \]
where \( \xi^m \) satisfies (2). By [33], the domain of \( S^m_T \) includes \( G_\delta \) for all \( \delta > 0 \). The following result is similar to that in [30].

**Theorem 2.4** Fix any \( T > 0 \). Each value function, \( V^m \), is the unique smooth solution of
\[ V = S^m_T[V] \]
in the class \( G_\delta \) for sufficiently small \( \delta > 0 \). Further, given any \( V \in G_\delta \), \( \lim_{T \to \infty} S^m_T[V](x) = V^m(x) \) for all \( x \in \mathbb{R}^n \) (uniformly on compact sets).
**Proof.** (Sketch of proof.) First we note that for any $0 < \tau < T < \infty$

\[ V^{m,f}(x,T) = \sup_{w \in W} \left[ \int_0^T \frac{1}{2}(\xi_t^m)^T D^m \xi_t^m - \frac{\gamma^2}{2} |w_t|^2 dt + V^{m,f}(\xi_T^m, T - \tau) \right] \]

(11)

where $\xi^m$ satisfies (2). Since the proof of (11) is standard, we do not prove it. (Similar proofs can be found in the standard references. The proof in this particular case appears in [33].)

Recall from (6) that $V^{m,f}(x) = \lim_{T \to \infty} V^{m,f}(x,T)$. Let $T_2 > T_1$. Let $w^\varepsilon$ be $\varepsilon$–optimal for $V^{m,f}(x,T_1)$. Let $w$ be given by

\[ w(t) = \begin{cases} w^\varepsilon(t) & \text{if } t \in [0,T_1] \\ 0 & \text{if } t \in (T_1,T_2) \end{cases} \]

Then,

\[ V^{m,f}(x,T_1) \leq J^m(x,T_1,w^\varepsilon) + \varepsilon \leq J^m(x,T_2,w) + \varepsilon \leq V^{m,f}(x,T_2) + \varepsilon. \]

Since $\varepsilon > 0$ was arbitrary, $V^{m,f}(x, \cdot)$ is monotonically increasing. Using this monotonicity and (6), one has

\[ V^m(x) = \sup_{T<\infty} V^{m,f}(x,T) \]

which by (11)

\[ = \sup_{T<\infty} \sup_{w \in W} \left\{ \int_0^T \frac{1}{2}(\xi_t^m)^T D^m \xi_t^m - \frac{\gamma^2}{2} |w_t|^2 dt + V^{m,f}(\xi_T^m, T - \tau) \right\} \]

where $\xi^m$ satisfies (2)

\[ = \sup_{w \in W} \sup_{T<\infty} \left\{ \int_0^T \frac{1}{2}(\xi_t^m)^T D^m \xi_t^m - \frac{\gamma^2}{2} |w_t|^2 dt + V^{m,f}(\xi_T^m, T - \tau) \right\} \]

\[ = \sup_{w \in W} \left\{ \int_0^T \frac{1}{2}(\xi_t^m)^T D^m \xi_t^m - \frac{\gamma^2}{2} |w_t|^2 dt + V^m(\xi^m_T) \right\} \]

\[ = S^m_T[V^m](x). \]

The smoothness follows from the quadratic form. The proof of uniqueness is nearly identical to the uniqueness proof of Theorem 2.5 in [33], and so we do not include it, but refer the reader to [33]. □

Recall that the HJB PDE of interest is

\[ 0 = \tilde{H}(x, \nabla V) \doteq \max_{m \in M} H^m(x, \nabla V) \]

\[ V(x) = 0. \]

(12)
The corresponding value function is

\[ \bar{V}(x) = \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \bar{J}(x, w, \mu) \]

\[ \triangleq \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 dt \tag{13} \]

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 \( \xi \) satisfies

\[ \dot{\xi} = A^{\mu_t} \xi + \sigma^{\mu_t} w_t \]

\[ \xi_0 = x. \tag{14} \]

**Theorem 2.5** Value function \( \bar{V} \) is the unique viscosity solution to (12) in the class \( \mathcal{G}_\delta \) for sufficiently small \( \delta > 0 \).

**Remark 2.6** The proof of Theorem 2.5 is identical to the proof of Theorems 2.5 and 2.6 from [33] 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 \( \mu \in \mathcal{D}_\infty \). Also, the finite time-horizon PDEs now include maximization over \( m \in \mathcal{M} \). In particular, (31) in [33] now becomes

\[ 0 = \bar{V}^f_T - \max_{m \in \mathcal{M}} \left[ \frac{1}{2} x^T D^m x + (A^m x)^T \nabla \bar{V}^f + \frac{1}{2} (\nabla \bar{V}^f)^T \Sigma^m \nabla \bar{V}^f \right] \]

\[ \bar{V}^f(x, 0) = 0 \]

where previously there was no maximization over \( m \).

Define the semigroup

\[ \bar{S}_T[\phi] = \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_T} \left[ \int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 dt + \phi(\xi_T) \right] \tag{15} \]

where

\[ \mathcal{D}_T = \{ \mu : [0, T) \to \mathcal{M} : \text{measurable} \} . \tag{16} \]

In analogy with Theorem 2.4, one has the following.

**Theorem 2.7** Fix any \( T > 0 \). Value function \( \bar{V} \) is the unique continuous solution of

\[ V = \bar{S}_T[V] \]

in the class \( \mathcal{G}_\delta \) for sufficiently small \( \delta > 0 \). Further, given any \( V \in \mathcal{G}_\delta \), \( \lim_{T \to \infty} \bar{S}_T[V](x) = \bar{V}(x) \) for all \( x \in \mathbb{R}^n \) (uniformly on compact sets).
The proof is nearly identical to the proof of Theorem 2.4, and so is not included. In particular, the only change is the addition of the supremum over $D_T$ – which makes no substantive change in the proof.

Importantly, we also have the following.

**Theorem 2.8** Value function $\bar{V}$ is convex. Further, there exists $c_V > 0$ and $\varepsilon > 0$ such that $\bar{V}(x) - \frac{1}{2}c_V|x|^2$ is convex.

**Proof.** Fix any $x, u \in \mathbb{R}^n$ with $|u| = 1$ and any $\delta > 0$. Let $\varepsilon > 0$. Given $x$, let $w^\varepsilon \in \mathcal{W}$, $\mu^\varepsilon \in \mathcal{D}_\infty$ be $\varepsilon$-optimal for $\bar{V}(x)$ (i.e. so that $\bar{J}(x, w^\varepsilon, \mu^\varepsilon) \geq \bar{V}(x) - \varepsilon$). Then

$$\bar{V}(x - \delta u) - 2\bar{V}(x) + \bar{V}(x + \delta u) \\
\geq \bar{J}(x - \delta u, w^\varepsilon, \mu^\varepsilon) - 2\bar{J}(x, w^\varepsilon, \mu^\varepsilon) + \bar{J}(x + \delta u, w^\varepsilon, \mu^\varepsilon) - 2\varepsilon. \tag{17}$$

Let $\xi^\delta, \xi^0, \xi^{-\delta}$ be solutions of dynamics (14) with initial conditions $\xi^\delta_0 = x + \delta u$, $\xi^0_0 = x$ and $\xi^{-\delta}_0 = x - \delta u$, respectively, where the inputs are $w^\varepsilon$ and $\mu^\varepsilon$ for all three processes. Then

$$\dot{\xi} - \dot{\xi}^0 = A^\mu \xi - \xi^0, \quad \text{and} \quad \dot{\xi}^0 - \dot{\xi}^{-\delta} = A^\mu \xi^0 - \xi^{-\delta}. \tag{18}$$

Letting $\Delta^+_t = \xi^\delta_t - \xi^0_t$, one also has $\dot{\xi}^+_t = \Delta^+_t$, and by linearity one finds $\dot{\Delta}^+ = A^\mu \Delta^+$.

Also, using (17) and (13)

$$\bar{V}(x - \delta u) - 2\bar{V}(x) + \bar{V}(x + \delta u) \\
\geq \frac{1}{2} \int_0^\infty \left[ \xi^\delta_t D^\mu \xi^\delta_t - 2\xi^0_t D^\mu \xi^0_t + \xi^{-\delta}_t D^\mu \xi^{-\delta}_t \right] dt - 2\varepsilon \\
= \int_0^\infty (\Delta^+_t)^T D^\mu \Delta^+_t dt - 2\varepsilon. \tag{19}$$

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

$$\frac{\partial}{\partial t} |\Delta^+_t|^2 = 2(\Delta^+_t)^T A^\mu \Delta^+_t \geq -K |\Delta^+_t|^2$$

which implies

$$|\Delta^+_t|^2 \geq e^{-Kt} \delta^2 \quad \forall t \geq 0. \tag{20}$$

Let $\lambda_D = \min\{\lambda \in \mathbb{R} : \lambda \text{ is an eigenvalue of a } D^m\}$. By the positive definitness of the $D^m$ and finiteness of $\mathcal{M}$, $\lambda_D > 0$. Then, by (19)

$$\bar{V}(x - \delta u) - 2\bar{V}(x) + \bar{V}(x + \delta u) \geq \int_0^\infty \lambda_D |\Delta^+_t|^2 dt - 2\varepsilon$$

which by (20)

$$\geq \frac{\lambda_D}{K} \delta^2 - 2\varepsilon.$$

Since $\varepsilon > 0$ and $|u| = 1$ were arbitrary, one obtains the result. \qed
3 Max-Plus Spaces and Dual Operators

Let $\mathbb{R}^- = \mathbb{R} \cup \{-\infty\}$. Recall that a function, $\phi : \mathbb{R}^n \to \mathbb{R}^-$ is semiconvex if given any $R \in (0, \infty)$ there exists $\beta_R \in \mathbb{R}$ such that $\phi(x) + \frac{\beta_R}{2} |x|^2$ is convex over $B_R(0) = \{ x \in \mathbb{R}^n : |x| \leq R \}$. We say $\phi$ is uniformly semiconvex with constant $\beta$ if $\phi(x) + \frac{\beta}{2} |x|^2$ is convex over $\mathbb{R}^n$. Let $S_\beta = S_\beta(\mathbb{R}^n)$ be the set of functions mapping $\mathbb{R}^n$ into $\mathbb{R}^-$ which are uniformly semiconvex with constant $\beta$. Note that $S_\beta$ is a max-plus vector space (also known as a moduloid) [16], [30], [2], [5], [23]. For instance, $\alpha_1 \otimes \phi_1 + \alpha_2 \otimes \phi_2 \in S_\beta$ for all $\alpha_1, \alpha_2 \in \mathbb{R}^-$ and all $\phi_1, \phi_2 \in S_\beta$. Combining Corollary 2.3 and Theorem 2.8, we have the following.

**Theorem 3.1** There exists $\beta \in \mathbb{R}$ such that given any $\beta > \beta$, $\tilde{V} \in S_\beta$ and $V_m \in S_\beta$ for all $m \in \mathcal{M}$. Further, one may take $\beta < 0$ (i.e. $\tilde{V}, V^m$ are convex).

The following semiconvex duality result [16], [30], [29] requires only a small modification of convex duality and Legendre/Fenchel transform results [36], [37].

**Theorem 3.2** Let $\phi \in S_\beta$. Let $C$ be a symmetric matrix such that $C + \beta I < 0$ (i.e. such that $C + \beta I$ is negative definite). Define $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $\psi(x, z) = -\frac{1}{2} (x - z)^T C (x - z)$. Then, for all $x \in \mathbb{R}^n$,

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

$$= \int_{\mathbb{R}^n} \psi(x, z) \otimes a(z) \, dz = \psi(x, \cdot) \otimes a(\cdot)$$

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

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

$$= -\int_{\mathbb{R}^n} \psi(x, z) \otimes [-\phi(x)] \, dx = -\{ \psi(\cdot, z) \otimes [-\phi(\cdot)] \}$$

which using the notation of [5]

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

We will refer to $a$ as the semiconvex dual of $\phi$ (with respect to $\psi$).

**Remark 3.3** We note that $\phi \in S_\beta$ implies that $\phi$ is locally Lipschitz (c.f. [15]). We also note that if $\phi \in S_\beta$ and if there is any $x \in \mathbb{R}^n$ 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, $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, is uniformly semiconcave with constant $\beta$ if $\phi(x) - \frac{\beta}{2} |x|^2$ is concave over $\mathbb{R}^n$. Let $S^-_\beta$ be the set of functions mapping $\mathbb{R}^n$ into $\mathbb{R} \cup \{+\infty\}$ which are uniformly semiconvex with constant $\beta$. The next lemma is an obvious result of Theorem 3.2.
Lemma 3.4 Let $\phi \in \mathcal{S}_\beta$, and let $a$ be the semiconvex dual of $\phi$. Then $a \in \mathcal{S}^-$.

Lemma 3.5 Let $\phi \in \mathcal{S}_\beta$ with semiconvex dual $a$. Suppose $b \in \mathcal{S}_\beta^-$ is such that $\phi = \psi(x, \cdot) \odot b(\cdot)$. Then $b = a$.

Proof. Note that $-b \in \mathcal{S}_\beta$. Therefore, for all $y \in \mathbb{R}^n$

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

or equivalently,

$$b(y) = -\max_{\zeta \in \mathbb{R}^n} [\psi(y, \zeta) + \alpha(\zeta)] \quad (26)$$

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

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

which by assumption

$$= -\phi(\zeta). \quad (27)$$

Combining (26) and (27), and then using (23), one obtains

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

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 $\mathcal{S}_\tau[\psi(\cdot, z)] \in \mathcal{S}^-_{(c+\varepsilon)}$. This is the subject of the next theorem. For simplicity, we will henceforth specialize to the case where

$$\psi(x, z) \doteq (c/2)|x - z|^2.$$ 

Also, in order to keep the theorem statement clean, we will first make some definitions. Define

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

Note that the finiteness of $\mathcal{M}$ implies that $\lambda_D > 0$. Let

$$\overline{K} \doteq \max_{m \in \mathcal{M}, x \neq 0} \frac{x^T A^m x}{|x|^2}.$$ 

Using the simplifying notational conventions that $-\lambda_D/0 = -\infty$ and $\lambda_D/0 = +\infty$, we define the interval

$$I_{\epsilon_A, \overline{K}} \doteq \begin{cases} 
-\lambda_d/\overline{K}, \infty & \text{if } 0 \leq -2c_A \leq \overline{K}, \\
-\lambda_d/\overline{K}, \lambda_D/(2c_A) & \text{if } -2c_A < 0 < \overline{K}, \\
-\infty, \lambda_D/(2c_A) & \text{if } -2c_A \leq \overline{K} \leq 0.
\end{cases}$$
Theorem 3.6 Let $c \in I_{c_A, K}$. Then there exists $\tau > 0$ and $\eta > 0$ such that for all $\tau \in [0, \tau]$

$$\tilde{S}_\tau[\psi(\cdot, z)], S^m_\tau[\psi(\cdot, z)] \in S_{-(c+\eta\tau)}.$$ 

Remark 3.7 From the proof to follow, one can obtain feasible values for $\tau, \eta$. For instance, if $c > 0, c \in I_{c_A, K}$, then one may take $\eta = \frac{1}{2}(\lambda_D - c_A c)$ and $\tau$ such that $e^{-c_A \tau} = \frac{1}{2}$. However, in practice, such a $\tau$ tends to be highly conservative. Since these estimates are also quite technical, we do not give explicit values.

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

The first portion of the proof is similar to the proof of Theorem 2.8. Again, fix any $x, u \in \mathbb{R}^m$ with $|u| = 1$ and any $\delta > 0$. Fix $\tau > 0$, and let $\varepsilon > 0$. Given $x$, let $w, \mu \in \mathbb{R}^m$ be $\varepsilon$–optimal for $\tilde{S}_\tau[\psi(\cdot, z)](x)$. Specifically, suppose $\tilde{J}^\psi(x, \tau, w, \mu) \geq \tilde{S}_\tau[\psi(\cdot, z)](x) - \varepsilon$ where

$$\tilde{J}^\psi(x, \tau, w, \mu) = \int_0^\tau l^{\mu}(\xi_t) - \frac{\tau^2}{2} |w_t|^2 dt + \psi(\xi_T, z) \quad (28)$$

and $\xi_t$ satisfies (14). For simplicity of notation, let $\tilde{V}_\tau^\psi = \tilde{S}_\tau[\psi(\cdot, z)]$. Then

$$\tilde{V}_\tau^\psi(x - \delta u) - 2\tilde{V}_\tau^\psi(x) + \tilde{V}_\tau^\psi(x + \delta u) \geq \tilde{J}^\psi(x - \delta u, \tau, w, \mu^\varepsilon) - 2\tilde{J}^\psi(x + \delta u, \tau, w, \mu^\varepsilon) + \tilde{J}^\psi(x + \delta u, \tau, w, \mu^\varepsilon) - 2\varepsilon. \quad (29)$$

Let $\xi^\delta, \xi^0, \xi^{-\delta}, \Delta^+$ be as given in the proof of Theorem 2.8. Note that

$$\psi(\xi^\delta, z) - 2\psi(\xi^0, z) + \psi(\xi^{-\delta}, z) = c|\Delta^+|^2. \quad (30)$$

Note also that as in the proof of Theorem 2.8,

$$\frac{1}{2} \left[ \xi^\delta D^{\mu^\varepsilon} \xi^\delta - 2\xi^0 D^{\mu^0} \xi^0 + \xi^{-\delta} D^{\mu^\varepsilon} \xi^{-\delta} \right] = (\Delta^+)^T D^{\mu^\varepsilon} \Delta^+. \quad (31)$$

Combining (28), (29), (30) and (31), one obtains

$$\tilde{V}_\tau^\psi(x - \delta u) - 2\tilde{V}_\tau^\psi(x) + \tilde{V}_\tau^\psi(x + \delta u) \geq \int_0^\tau (\Delta^+)^T D^{\mu^\varepsilon} \Delta^+ dt + c|\Delta^+|^2 - 2\varepsilon$$

$$\geq \int_0^\tau \lambda_D |\Delta^+|^2 dt + c|\Delta^+|^2 - 2\varepsilon. \quad (32)$$

Further, noting again that $\Delta^+ = A^{\mu^\varepsilon} \Delta^+$, one has

$$\frac{\partial}{\partial t} |\Delta^+|^2 = -2(\Delta^+)^T A^{\mu^\varepsilon} \Delta^+.$$ 

Consequently, using Assumption Block (A.m) and the definition of $K$,

$$-2c_A |\Delta^+|^2 \leq \frac{\partial}{\partial t} |\Delta^+|^2 \leq 2K |\Delta^+|^2,$$
and so
\[ \delta^2 e^{-2c_A t} \leq |\Delta^t|^2 \leq \delta^2 e^{2K t}. \]  
(33)

Suppose \( c \geq 0 \). Then by (32) and (33),
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \lambda_D \delta^2 \int_0^T e^{-2c_A t} dt + c\delta^2 e^{-2c_A T} - 2\varepsilon \]
\[ = \delta^2 \hat{f}(\tau) - 2\varepsilon \]
where
\[ \hat{f}(\tau) = \lambda_D \frac{1 - e^{-2c_A \tau}}{2c_A} + ce^{-2c_A \tau}. \]
Note that \( f(0) = 0 \) and \( f'(\tau) = (\lambda_D - 2c_A c)e^{-c_A \tau} \). Since \( f'(0) = \lambda_D - 2c_A c > 0 \), then letting \( \eta \doteq \frac{1}{2}(\lambda_D - 2c_A c) \), one sees that there exists \( \overline{\tau} > 0 \) such that
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \delta^2 [f(0) + \eta \tau] - 2\varepsilon \quad \forall \tau \in [0, \overline{\tau}]. \]
Since this is true for all \( \varepsilon > 0 \),
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \delta^2 [c + \eta \tau] \quad \forall \tau \in [0, \overline{\tau}]. \]  
(34)

Now suppose \( c < 0 \). Then by (32) and (33),
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \delta^2 \hat{f}(\tau) - 2\varepsilon \]
(35)
where
\[ \hat{f}(\tau) = \lambda_D \frac{1 - e^{-2c_A \tau}}{2c_A} + ce^{2c_A \tau}. \]
Note that \( \hat{f}(0) = e \) and \( \hat{f}'(\tau) = \lambda_D e^{-2c_A \tau} + \overline{K}e^{2c_A \tau} \). Since \( \hat{f}'(0) = \lambda_D + \overline{K}c > 0 \), then letting \( \eta \doteq \frac{1}{2}(\lambda_D + \overline{K}c) \), one sees that there exists \( \overline{\tau} > 0 \) such that
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \delta^2 [\hat{f}(0) + \eta \tau] - 2\varepsilon \quad \forall \tau \in [0, \overline{\tau}]. \]
Since this is true for all \( \varepsilon > 0 \),
\[ \hat{V}^{\tau,\psi}(x - \delta u) - 2\hat{V}^{\tau,\psi}(x) + \hat{V}^{\tau,\psi}(x + \delta u) \geq \delta^2 [c + \eta \tau] \quad \forall \tau \in [0, \overline{\tau}]. \]  
(36)
Combining (34) and (36) yields the result if the two conditions \( \lambda_D - 2c_A c > 0 \) when \( c \geq 0 \) and \( \lambda_D + \overline{K}c > 0 \) when \( c < 0 \) are met. The reader can check that these conditions are met if \( c \in I_{c_A, \overline{K}} \). \( \square \)

**Corollary 3.8** We may choose \( c \in \mathbb{R} \) such that \( \overline{V}, V^m \in S_{-c}, \) and such that with \( \psi, \tau, \eta \) as in the statement of Theorem 3.6, 
\[ \overline{S}_\tau[\psi(\cdot, z)], S^m_\tau[\psi(\cdot, z)] \in S_{-(c+\eta \tau)} \quad \forall \tau \in [0, \overline{\tau}]. \]
Henceforth, we suppose $c$ chosen so that the results of Corollary 3.8 hold, and take $\psi(x, z) = \frac{3}{2}|x - z|^2$. We also suppose $\tau, \eta$ chosen according to the corollary as well.

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

$$\tilde{S}_\tau[\psi(\cdot, z)](x) = \int_{\mathbb{R}^n} \psi(x, y) \otimes \tilde{B}_\tau(y, z) \, dy = \psi(x, \cdot) \otimes \tilde{B}_\tau(\cdot, z)$$

(37)

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

$$\tilde{B}_\tau(y, z) = - \int_{\mathbb{R}^n} \psi(x, y) \otimes \{ -\tilde{S}_\tau[\psi(\cdot, z)](x) \} \, dx = \{ \psi(\cdot, y) \otimes [\tilde{S}_\tau[\psi(\cdot, z)](\cdot)]^- \}^-$$

(38)

It is handy to define the max-plus linear operator with “kernel” $\tilde{B}_\tau$ (where we do not rigorously define the term kernel as it will not be needed here) as $\hat{\tilde{B}}_\tau[a](z) = \tilde{B}_\tau(z, \cdot) \circ a(\cdot)$ for all $a \in \mathcal{S}_c$.

**Proposition 3.9** Let $\phi \in \mathcal{S}_c$ with semiconvex dual denoted by $a$. Define $\phi^1 = \tilde{S}_\tau[\phi]$. Then $\phi^1 \in \mathcal{S}_{-(c+\eta\tau)}$, and

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

where

$$a^1(x) = \hat{\tilde{B}}_\tau(x, \cdot) \circ a(\cdot).$$

**Proof.** The proof that $\phi^1 \in \mathcal{S}_{-(c+\eta\tau)}$ is similar to the proof in Theorem 3.6. Consequently, we prove only the second assertion.

$$\phi^1(x) = \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \left[ \int_0^\tau l^\mu(\xi_t) - \frac{\gamma}{2} |w_t|^2 \, dt + \phi(\xi_\tau) \right]$$

$$= \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \max_{z \in \mathbb{R}^n} \left[ \int_0^\tau l^\mu(\xi_t) - \frac{\gamma}{2} |w_t|^2 \, dt + \psi(\xi_\tau, z) + a(z) \right]$$

$$= \max_{z \in \mathbb{R}^n} \left\{ \tilde{S}_\tau[\psi(\cdot, z)](x) + a(z) \right\}$$

which by (37)

$$= \max_{z \in \mathbb{R}^n} \max_{y \in \mathbb{R}^n} \left\{ \psi(x, y) + \tilde{B}_\tau(y, z) + a(z) \right\}$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \tilde{B}_\tau(y, z) \otimes a(z) \, dz \otimes \psi(x, y) \, dy$$

$$= \int_{\mathbb{R}^n} a^1(x) \otimes \psi(x, y) \, dy. \square$$

**Theorem 3.10** Let $V \in \mathcal{S}_c$, and let $a$ be its semiconvex dual (with respect to $\psi$). Then

$$V = \tilde{S}_\tau[V]$$

if and only if

15
\[ a(z) = \max_{y \in \mathbb{R}^n} [\overline{B}_\tau(z, y) + a(y)] \]

which of course

\[ = \int_{\mathbb{R}^n} \overline{B}_\tau(z, y) \otimes a(y) \, dy = \overline{B}_\tau(z, \cdot) \circ a(\cdot) = \widehat{\overline{B}}[a](z) \quad \forall z \in \mathbb{R}^n. \]

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

\[
\psi(x, \cdot) \circ a(\cdot) = V(x) = \overline{S}_\tau[V](x)
\]

\[
= \overline{S}_\tau\left[\max_{z \in \mathbb{R}^n}\{\psi(\cdot, z) + a(z)\}\right](x)
\]

\[
= \sup_{w \in W} \sup_{\mu \in D_{\infty}} \left\{ \int_0^\tau l^{\mu}(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 \, dt + \max_{z \in \mathbb{R}^n}\left\{ \psi(\xi_t, z) + a(z) \right\} \right\}
\]

\[
= \max_{z \in \mathbb{R}^n}\left\{ a(z) + \sup_{w \in W} \sup_{\mu \in D_{\infty}} \left\{ \int_0^\tau l^{\mu}(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 \, dt + \psi(\xi_t, z) \right\} \right\}
\]

\[
= \max_{z \in \mathbb{R}^n}\left\{ a(z) + \overline{S}_\tau[\psi(\cdot, z)](x) \right\}
\]

\[
= \int_{\mathbb{R}^n} a(z) \otimes \overline{S}_\tau[\psi(\cdot, z)](x) \, dz
\]

which by (37)

\[
= \int_{\mathbb{R}^n} a(z) \otimes \int_{\mathbb{R}^n} \overline{B}_\tau(y, z) \otimes \psi(x, y) \, dy \, dz
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \overline{B}_\tau(y, z) \otimes a(z) \otimes \psi(x, y) \, dy \, dz
\]

\[
= \int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^n} \overline{B}_\tau(y, z) \otimes a(z) \, dz \right] \otimes \psi(x, y) \, dy
\]

\[
= \left[ \int_{\mathbb{R}^n} \overline{B}_\tau(\cdot, z) \otimes a(z) \, dz \right] \circ \psi(x, \cdot).
\]

Combining this with Lemma 3.5, one has

\[ a(y) = \int_{\mathbb{R}^n} \overline{B}_\tau(\cdot, z) \otimes a(z) \, dz = \overline{B}_\tau(y, \cdot) \circ a(\cdot) \quad \forall y \in \mathbb{R}^n. \]

**Corollary 3.11** Value function \( \overline{V} \) is given by \( \overline{V}(x) = \psi(x, \cdot) \circ \widehat{a}(\cdot) \) where \( \widehat{a} \) is the unique solution of

\[ \widehat{a}(y) = \overline{B}_\tau(y, \cdot) \circ \widehat{a}(\cdot) \quad \forall y \in \mathbb{R}^n \]

or equivalently, \( \widehat{a} = \widehat{\overline{B}}[\widehat{a}] \).

**Proof.** Combining Theorem 2.7 and Theorem 3.10 yields the assertion that \( \overline{V} \) has this representation. The uniqueness follows from the uniqueness assertion of Theorem 2.7 and Lemma 3.5.
Similarly, for each $m \in M$ and $z \in \mathbb{R}^n$, $S_m^m[\psi(\cdot, z)] \in S_{(c+\eta\tau)}$ and

$$S_m^m[\psi(\cdot, z)](x) = \psi(x, \cdot) \odot B_m^m(\cdot, z) \quad \forall x \in \mathbb{R}^n$$

where

$$B_m^m(y, z) = \left\{ \psi(\cdot, y) \odot [S_m^m[\psi(\cdot, z)] - (\cdot)] \right\}^- \quad \forall y \in \mathbb{R}^n.$$

As before, it will be handy to define the max-plus linear operator with “kernel” $B_m^m$ as $\hat{B}_m^m[a](z) = B_m^m(z, \cdot) \odot a(\cdot)$ for all $a \in S_{-c}$. Further, one also obtains analogous results (by similar proofs). In particular, one has the following

**Theorem 3.12** Let $V \in S_{-c}$, and let $a$ be its semiconvex dual (with respect to $\psi$). Then

$$V = S^m[V]$$

if and only if

$$a(z) = B_m^m(z, \cdot) \odot a(\cdot) \quad \forall z \in \mathbb{R}^n.$$

**Corollary 3.13** 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

$$a_m(y) = B_m^m(y, \cdot) \odot a_m(\cdot) \quad \forall y \in = \mathbb{R}^n.$$

### 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 $\mu$ processes will be constant over the length of each time-step.

We define the operator $\tilde{S}_\tau$ on $G$ by

$$\tilde{S}_\tau[\phi](x) = \sup_{w \in W} \max_{m \in M} \left[ \int_{\tau} l^m(\xi^m_t) - \frac{\tau^2}{2} |w_t|^2 dt + \phi(\xi^m_\tau) \right](x)$$

$$= \max_{m \in M} S_m^m[\phi](x)$$

where $\xi^m$ satisfies (2). Let

$$\bar{B}_\tau(y, z) = \max_{m \in M} B_m^m(y, z) = \bigoplus_{m \in M} B_m^m(y, z) \quad \forall y, z \in \mathbb{R}^n.$$

The corresponding max-plus linear operator is

$$\bar{B}_\tau = \bigoplus_{m \in M} \hat{B}_m^m.$$
Lemma 4.1  For all \( z \in \mathbb{R}^n \), \( \tilde{S}_r[\psi(\cdot, z)] \in \mathcal{S}_-(c+\eta r) \). Further,
\[
\tilde{S}_r[\psi(\cdot, z)](x) = \psi(x, \cdot) \odot \tilde{B}_r(\cdot, z) \quad \forall x \in \mathbb{R}^n.
\]

**Proof.** We provide the proof of the last statement, and this is as follows.
\[
\tilde{S}_r[\psi(\cdot, z)](x) = \max_{m \in M} S^m_r[\psi(\cdot, z)](x) = \max_{m \in M} \psi(x, \cdot) \odot B^m_r(\cdot, z)
\]
\[
= \max_{m \in M} \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + B^m_r(y, z) \right] = \max_{m \in M} \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + \max_{m \in M} B^m_r(y, z) \right]
\]
\[
= \psi(x, \cdot) \odot \left[ \max_{m \in M} \max_{y \in \mathbb{R}^n} \psi(x, y) + B^m_r(y, z) \right].
\]

We remark that, parameterized by \( \tau \), the operators \( \tilde{S}_\tau \) do not necessarily form a semigroup, although they do form a sub-semigroup (i.e. \( \tilde{S}_{\tau_1+\tau_2}[\phi](x) \leq \tilde{S}_{\tau_1} \tilde{S}_{\tau_2}[\phi](x) \) for all \( x \in \mathbb{R}^n \) and all \( \phi \in \mathcal{S}_- \)). In spite of this, one does have \( S^m_r \leq \tilde{S}_r \leq \tilde{\tilde{S}}_r \) for all \( m \in M \).

With \( \tau \) acting as a time-discretization step-size, let
\[
D^\infty_{\tau} = \left\{ \mu : [0, \infty) \to M \mid \text{for each } n \in \mathbb{N} \cup \{0\}, \text{ there exists } m_n \in M \text{ such that } \mu(t) = m_n \forall t \in [n\tau, (n+1)\tau) \right\},
\]
and for \( T = \bar{n}\tau \) with \( \bar{n} \in \mathbb{N} \) define \( D^T_{\tau} \) similarly but with domain \([0, T)\) rather than \([0, \infty)\). Let \( M^\bar{n} \) denote the outer product of \( M \), \( \bar{n} \) times. Let \( T = \bar{n}\tau \), and define
\[
\tilde{\tilde{S}}^T_{\tau}[\phi](x) = \max_{\{m_k\}_{k=0}^{\bar{n}-1} \in M^\bar{n}} \left\{ \prod_{k=0}^{\bar{n}-1} S^{m_k}_r \right\} [\phi](x) = (\tilde{S}_r)^\bar{n}[\phi](0)
\]
where the \( \prod \) notation indicates operator composition, and the superscript in the last expression indicates repeated application of \( \tilde{S}_\tau \), \( \bar{n} \) times.

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

**Theorem 4.2** Let
\[
\bar{V}(x) = \lim_{N \to \infty} \tilde{S}^N_{\tau}[0](x)
\]
for all \( x \in \mathbb{R}^n \) where 0 here represents the zero-function. Then, \( \bar{V} \) satisfies
\[
V = S_\tau[V] \\
V(0) = 0.
\]

Further, \( 0 \leq V^m \leq \bar{V} \leq \tilde{\tilde{V}} \) for all \( m \in M \), and consequently, \( \bar{V} \in \mathcal{G}_\delta \).
Proof. Note that
\[ V^m(x) = \lim_{N \to \infty} S^m_{N \tau}[0](x) \leq \limsup_{N \to \infty} \bar{S}^\tau_{N \tau}[0] \leq \lim_{N \to \infty} \bar{S}_{N \tau}[0](x) = \bar{V}(x) \quad \forall \, x \in \mathbb{R}^n. \quad (41) \]

Also,
\[ \bar{S}^\tau_{(N+1) \tau}[0](x) = \bar{S}^\tau_{N \tau}[\bar{S}_{\tau}[0](\cdot)](x) \]
\[ = \sup_{\bar{w} \in \mathcal{W}} \sup_{\bar{\mu} \in \mathcal{D}_{N \tau}} \int_{0}^{N \tau} \bar{l}_{\bar{\mu}}(\xi_t) - \frac{\gamma^2}{2} |\hat{\bar{w}}_t|^2 \, dt \]
\[ + \sup_{\bar{w} \in \mathcal{W}} \max_{m \in \mathcal{M}_{N \tau}} \int_{N \tau}^{(N+1) \tau} l^m(\xi_t) - \frac{\gamma^2}{2} |\bar{w}_t|^2 \, dt \]
which by taking \( \bar{w} \equiv 0 \)
\[ \geq \sup_{\bar{w} \in \mathcal{W}} \sup_{\bar{\mu} \in \mathcal{D}_{N \tau}} \int_{0}^{N \tau} l_{\bar{\mu}}(\xi_t) - \frac{\gamma^2}{2} |\hat{\bar{w}}_t|^2 \, dt = \bar{S}^\tau_{N \tau}[0](x), \quad (42) \]
which implies that \( \bar{S}^\tau_{N \tau}[0](x) \) is a monotonically increasing function of \( N \). Since it is also bounded from above (by (41)), one finds
\[ V^m(x) \leq \lim_{N \to \infty} \bar{S}^\tau_{N \tau}[0](x) \leq \bar{V}(x) \quad \forall \, x \in \mathbb{R}^n \quad (44) \]
which also justifies the use of the limit definition of \( \nabla \) in the statement of the Theorem.

In particular, one has \( 0 \leq V^m \leq \nabla \leq \bar{V} \), and so \( \nabla \in G_\delta \).

Fix any \( x \in \mathbb{R}^n \), and suppose there exists \( \delta > 0 \) such that
\[ \nabla(x) \leq \bar{S}_\tau[\nabla](x) - \delta. \quad (45) \]

However, by the definition of \( \nabla \), given any \( y \in \mathbb{R}^n \), there exists \( N_\delta < \infty \) such that for all \( N \geq N_\delta \)
\[ \nabla(y) \leq \bar{S}^\tau_{N_\delta \tau}[0](y) + \delta/4. \quad (46) \]
Combining (45) and (46), one finds after a small bit of work that
\[ \nabla(x) \leq \bar{S}_\tau[\bar{S}^\tau_{N_\delta \tau}[0] + \delta/2](x) - \delta \]
which using the max-plus linearity of \( \hat{S}_\tau \)
\[ = \bar{S}^\tau_{(N_\delta+1) \tau}[0](x) - \delta/2 \]
for all \( N \geq N_\delta \). Consequently, \( \nabla(x) \leq \lim_{N \to \infty} \bar{S}^\tau_{N \tau}[0](x) - \delta/2 \) which is a contradiction. Therefore, \( \nabla(x) \geq \bar{S}_\tau[\nabla](x) \) for all \( x \in \mathbb{R}^n \). The reverse inequality follows in a similar way. Specifically, fix \( x \in \mathbb{R}^n \) and suppose there exists \( \delta > 0 \) such that
\[ \nabla(x) \geq \bar{S}_\tau[\nabla](x) + \delta. \quad (47) \]
By the monotonicity of $\bar{S}_{N\tau}$ with respect to $N$, for any $N < \infty$,

$$\bar{V}(x) \geq \bar{S}_{N\tau}[0](x) \quad \forall x \in \mathbb{R}^n.$$ 

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

$$\bar{S}_\tau[\bar{V}] \geq \bar{S}_{(N+1)\tau}[0] \quad \forall x \in \mathbb{R}^n.$$  \hspace{1cm} (48)

Combining (47) and (48) yields

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

Letting $N \to \infty$ yields a contradiction, and so $\bar{V} \leq \bar{S}_\tau[\bar{V}]$. \hspace{1cm} $\square$

The following result is immediate.

**Theorem 4.3**

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

where $\xi_t$ satisfies (14).

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

**Proof.** The proof is identical to the proof of Theorem 2.8 with the exception that $\mu^c$ is chosen from $\mathcal{D}_\infty$ instead of $\mathcal{D}_\infty$. \hspace{1cm} $\square$

We now address the uniqueness issue. Similar techniques to those used for $V^m$ and $\tilde{V}$ will prove uniqueness for (40) 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 [38].

Suppose $\bar{V}' \neq \bar{V}$, $\bar{V}' \in \mathcal{G}_\delta$ satisfies (40). This implies that for all $x \in \mathbb{R}^n$ and all $N < \infty$

$$\bar{V}'(x) = \bar{S}_{N\tau}'[\bar{V}](x)$$

$$= \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \left\{ \int_0^{N\tau} l^\mu_t(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 dt + \bar{V}'(\xi_{N\tau}) \right\}$$

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

$$\geq \bar{V}'(\xi^0_{N\tau}).$$ \hspace{1cm} (49)

However, by (14), one has $\dot{\xi}^0 = A^{\mu^c}e^0$, and so $|\dot{\xi}_t^0| \leq e^{-cA^t|x|}$ for all $t \geq 0$ which implies that $|\xi_{N\tau}^0| \to 0$ as $N \to \infty$. Consequently

$$\lim_{N \to \infty} \bar{V}'(\xi_{N\tau}^0) = 0.$$ \hspace{1cm} (50)
Combining (49) and (50), one has
\[ V'(x) \geq 0 \quad \forall x \in \mathbb{R}^n. \] (51)

Also, by (40)
\[ V'(x) = \lim_{N \to \infty} \tilde{S}^\tau_{N\tau}(V')(x) \quad \forall x \in \mathbb{R}^n. \]

By (51) and the monotonicity of \( \tilde{S}^\tau_{N\tau} \) with respect to its argument, this is
\[ \geq \lim_{N \to \infty} \tilde{S}^\tau_{N\tau}[0](x) = V(x). \] (52)

By (51), (52), one has the uniqueness result analogous to [38], which is as follows.

**Theorem 4.5** \( \bar{V} \) is the unique minimal, nonegative solution to (40).

The stronger uniqueness statement (making use of the quadratic bound on \( l^\mu(x) \)) is as follows. As with \( V^m, \bar{V} \), the proof is similar to that in [33]. 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 \( \bar{V} \), we include a sketch of the proof (but with the new lemma in full) in Appendix A.

**Theorem 4.6** \( \bar{V} \) is the unique solution of (40) within the class \( \mathcal{G}_\delta \) for sufficiently small \( \delta > 0 \). Further, given any \( V \in \mathcal{G}_\delta \), \( \lim_{N \to \infty} \tilde{S}^\tau_{N\tau}[V](x) = \bar{V}(x) \) for all \( x \in \mathbb{R}^n \) (uniformly on compact sets).

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

**Theorem 4.7** Let \( V \in \mathcal{S}_-c \), and let \( a \) be its semiconvex dual. Then
\[ V = \bar{S}_r[V] \]
if and only if
\[ a(y) = B_r(y, \cdot) \odot a(\cdot) \quad \forall y \in \mathbb{R}^n. \]

**Proof.** By the semiconvex duality
\[ \psi(x, \cdot) \odot a(\cdot) = V(x) = \bar{S}_r[V](x) \]
\[ = \bar{S}_r[\max_{z \in \mathbb{R}^n}\{\psi(\cdot, z) + a(z)\}](x) \]
which as in the first part of the proof of Theorem 3.10
\[ = \int_{\mathbb{R}^n} a(z) \otimes \bar{S}_r[\psi(\cdot, z)](x) \, dz \]
which by Lemma 4.1
\[
\int_{\mathbb{R}^n} a(z) \otimes \int_{\mathbb{R}^n} \psi(x, y) \otimes \mathcal{B}_r(y, z) \, dy \, dz
\]
which as in the latter part of the proof of Theorem 3.10
\[
= \left[ \int_{\mathbb{R}^n} \mathcal{B}_r(\cdot, z) \otimes a(z) \, dz \right] \otimes \psi(x, \cdot). \tag{54}
\]
By Lemma 3.5, this implies
\[
a(y) = \mathcal{B}_r(y, \cdot) \otimes a(\cdot) \quad \forall y \in \mathbb{R}^n.
\]
Alternatively, if \(a(y) = \mathcal{B}_r(y, \cdot) \otimes a(\cdot)\) for all \(y\), then
\[
V(x) = \psi(x, \cdot) \otimes a(\cdot) = \left[ \int_{\mathbb{R}^n} \mathcal{B}_r(\cdot, z) \otimes a(z) \, dz \right] \otimes \psi(x, \cdot) \quad \forall x \in \mathbb{R}^n,
\]
which by (53)–(54) yields \(V = \bar{S}_r[V]\). \(\square\)

**Corollary 4.8** Value function \(\bar{V}\) given by (39) is in \(\mathcal{S}_{-c}\), and has representation \(\bar{V}(x) = \psi(x, \cdot) \otimes \bar{a}(\cdot)\) where \(\bar{a}\) is the unique solution of
\[
\bar{a}(y) = \mathcal{B}_r(y, \cdot) \otimes \bar{a}(\cdot) \quad \forall y \in \mathbb{R}^n \tag{55}
\]
or equivalently, \(\bar{a} = \hat{\mathcal{B}}_r[\bar{a}]\).

**Proof.** The representation follows from Theorems 4.2 and 4.7. The uniqueness follows from Theorem 4.6 and Lemma 3.5. \(\square\)

The following result on propagation of the semiconvex dual will also come in handy. The proof is similar to the proof of Proposition 3.9, and so is not included.

**Proposition 4.9** Let \(\phi \in \mathcal{S}_{-c}\) with semiconvex dual denoted by \(a\). Define \(\phi^1 = \bar{S}_r[\phi]\). Then \(\phi^1 \in \mathcal{S}_{-(c+\eta r)}\), and
\[
\phi^1(x) = \psi(x, \cdot) \otimes a^1(\cdot)
\]
where
\[
a^1(y) = \mathcal{B}_r(y, \cdot) \otimes a(\cdot) \quad \forall y \in \mathbb{R}^n.
\]

We now show that one may approximate \(\bar{V}\), the solution of \(V = \bar{S}_r[V]\), to as accurate a level as one desires by solving \(V = \bar{S}_r[V]\) for sufficiently small \(\tau\). Recall that if \(V = \bar{S}_r[V]\), then it satisfies \(V = \bar{S}_{N\tau}[V]\) for all \(N > 0\) (while \(\bar{V}\) satisfies \(V = \bar{S}_{N\tau}[V]\)), and so this is essentially equivalent to introducing a discrete-time \(\bar{\mu} \in \mathcal{D}_{N\tau}\) approximation to the \(\mu\) process in \(\bar{S}_{N\tau}\). 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. We also note that a similar result under different assumptions appears in [26].

22
Lemma 4.10 Given $\hat{\varepsilon} \in (0, 1]$, $T < \infty$, there exist $T \in [T/2, T]$ and $\tau > 0$ such that
\[ \tilde{S}_T[V^m](x) - \tilde{S}_T^T[V^m](x) \leq \hat{\varepsilon}(1 + |x|^2) \quad \forall x \in \mathbb{R}^n, \forall m \in \mathcal{M}. \]

We now obtain the main approximation result.

Theorem 4.11 Given $\varepsilon > 0$ and $R < \infty$, there exists $\tau > 0$ such that
\[ \tilde{V}(x) - \varepsilon \leq \tilde{V}(x) \leq \tilde{V}(x) \quad \forall x \in \overline{B}_R(0). \]

Proof. From Theorem 4.2, we have
\[ 0 \leq V^m(x) \leq \overline{V}(x) \leq \tilde{S}_T[V^m](x) \leq \frac{c_A(\gamma - \delta)^2}{c_2^2} |x|^2 \quad \forall x \in \mathbb{R}^n. \tag{56} \]

Also, with $T = N\tau$ for any positive integer $N$,
\[ \tilde{S}_{N\tau}[\phi] \leq \tilde{S}_T[\phi] \quad \forall \phi \in \mathcal{G}_\delta. \tag{57} \]

Further, by Theorem 2.7, given $\varepsilon > 0$ and $R < \infty$, there exists $\hat{T} < \infty$ such that for all $T > \hat{T}$ and all $m \in \mathcal{M}$
\[ \tilde{S}_T[\overline{V}](x) - \varepsilon/2 \leq \tilde{S}_T[V^m](x) \quad \forall x \in \overline{B}_R(0). \tag{58} \]

By (58) and Lemma 4.10, given $\varepsilon > 0$ 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 $|x| \leq R$
\[ \tilde{V}(x) - \varepsilon = \tilde{S}_T[\overline{V}](x) - \varepsilon \]
\[ \leq \tilde{S}_T[V^m](x) - \varepsilon/2 \]
which with $\hat{\varepsilon}(1 + R^2) = \varepsilon/2$
\[ \leq \tilde{S}_T^T[V^m](x) \]
which by (56) and the monotonicity of $\tilde{S}_T^T[\cdot]$
\[ \leq \tilde{S}_T[\overline{V}](x) \]
which by (57)
\[ \leq \tilde{S}_T[\overline{V}](x) \]
which by the monotonicity of $\tilde{S}_T[\cdot]$
\[ \leq \tilde{S}_T[\overline{V}](x) = \tilde{V}(x). \]

Noting (from Theorem 4.6) that $\overline{V} = \tilde{S}_T^T[\overline{V}]$ completes the proof. \(\square\)

Remark 4.12 For this class of systems (defined by Assumption Block (A.m)), we expect this result could be sharpened to
\[ \tilde{V}(x) \leq -\hat{\varepsilon}(1 + |x|^2) \leq \overline{V}(x) \leq \tilde{V}(x) \quad \forall x \in \mathbb{R}^n \]
by sharpening Theorem 2.7. 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. From Theorem 4.2, \( \mathcal{V} = \lim_{N \rightarrow \infty} \tilde{S}_N^r[0] \). Let \( \mathcal{V}^0 \equiv 0 \). Then \( \mathcal{V}^0 \in \mathcal{S}_{-c} \) of course. Given \( \mathcal{V}^k \), let

\[
\mathcal{V}^{k+1} = \tilde{S}_r[\mathcal{V}^k]
\]

so that \( \mathcal{V}^k = \tilde{S}_r^k[0] \) for all \( k \geq 1 \).

Let \( \mathcal{V}^k \) be the semiconvex dual of \( \mathcal{V}^k \) for all \( k \). Since \( \mathcal{V}^0 \equiv 0 \), one easily finds that \( \mathcal{V}^0(y) = 0 \) for all \( y \in \mathbb{R}^n \). Note also that by Proposition 4.9,

\[
\mathcal{V}^{k+1} = \mathcal{B}_r(x, \cdot) \circ \mathcal{V}^k = \tilde{\mathcal{B}}_r[\mathcal{V}^k]
\]

for all \( n \geq 0 \).

Recall that

\[
\mathcal{B}_r(x, \cdot) \circ \mathcal{V}^k(\cdot) = \int_{\mathbb{R}^n} \mathcal{B}_r(x, y) \otimes \mathcal{V}^k(y) \, dy = \int_{\mathbb{R}^n} \bigoplus_{m \in \mathcal{M}} \mathcal{B}_r^m(x, y) \otimes \mathcal{V}^k(y) \, dy
\]

\[
= \bigoplus_{m \in \mathcal{M}} \int_{\mathbb{R}^n} \mathcal{B}_r^m(x, y) \otimes \mathcal{V}^k(y) \, dy = \bigoplus_{m \in \mathcal{M}} \left[ \mathcal{B}_r^m(x, \cdot) \otimes \mathcal{V}^k(\cdot) \right]. \quad (59)
\]

By (59),

\[
\mathcal{V}^1(x) = \bigoplus_{m \in \mathcal{M}} \hat{\mathcal{V}}^1_m(x)
\]

where

\[
\hat{\mathcal{V}}^1_m(x) = \mathcal{B}_r^m(x, \cdot) \otimes \mathcal{V}^0(\cdot) \quad \forall m. \quad (60)
\]

By (59) and (60),

\[
\mathcal{V}^2(x) = \bigoplus_{m_2 \in \mathcal{M}} \int_{\mathbb{R}^n} \mathcal{B}_r^{m_2}(x, y) \otimes \left[ \bigoplus_{m_1 \in \mathcal{M}} \hat{\mathcal{V}}^1_{m_1}(y) \right] \, dy
\]

\[
= \bigoplus_{\{m_1, m_2\} \in \mathcal{M} \times \mathcal{M}} \int_{\mathbb{R}^n} \mathcal{B}_r^{m_2}(x, y) \otimes \hat{\mathcal{V}}^1_{m_1}(y) \, dy.
\]

Consequently,

\[
\mathcal{V}^2(x) = \bigoplus_{\{m_1, m_2\} \in \mathcal{M}^2} \hat{\mathcal{V}}^2_{\{m_1, m_2\}}(x)
\]

where

\[
\hat{\mathcal{V}}^2_{\{m_1, m_2\}}(x) = \mathcal{B}_r^{m_2}(x, \cdot) \circ \hat{\mathcal{V}}^1_{m_1}(\cdot) \quad \forall m_1, m_2. \quad (61)
\]
and $\mathcal{M}^2$ represents the outer product $\mathcal{M} \times \mathcal{M}$. Proceeding with this, one finds that in general,

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

where

$$
\bar{a}_{\{m_i\}_{i=1}^k}^k(x) = \mathcal{B}^m_\tau(x, \cdot) \odot \bar{a}_{\{m_i\}_{i=1}^{k-1}}^{k-1}(\cdot) \quad \forall \{m_i\}_{i=1}^k \in \mathcal{M}^k.
$$

Of course one can obtain $V^n$ from its dual as

$$
V^n(x) = \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + a^k(y) \right] = \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + \max_{\{m_i\}_{i=1}^k \in \mathcal{M}^k} \left\{ \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + \hat{a}_{\{m_i\}_{i=1}^k}^k(y) \right] \right\} \right] = \max_{\{m_i\}_{i=1}^k \in \mathcal{M}^k} \hat{V}_{\{m_i\}_{i=1}^k}^k(x)
$$

where

$$
\hat{V}_{\{m_i\}_{i=1}^k}^k(x) = \max_{y \in \mathbb{R}^n} \left[ \psi(x, y) + \hat{a}_{\{m_i\}_{i=1}^k}^k(y) \right] = \int_{\mathbb{R}^n} \psi(x, y) \otimes \hat{a}_{\{m_i\}_{i=1}^k}^k(y) \, dy.
$$

The algorithm will consist of the forward propagation of the $\hat{a}_{\{m_i\}_{i=1}^k}^k$ (according to (62)) from $k = 0$ to some termination step $k = N$, followed by construction of the value as $\hat{V}_{\{m_i\}_{i=1}^k}^k$ (according to (64)).

It is important to note that the computation of each $\hat{a}_{\{m_i\}_{i=1}^k}^k$ 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

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

where the time-dependent $n \times n$ matrices $P^m_\tau$, $\Lambda^m_\tau$ and $R^m_\tau$ satisfy $P^m_0 = cI$, $\Lambda^m_0 = I$, $R^m_0 = 0$,

$$
\dot{P}^m = (A^m)^T P^m + P^m A^m - [D^m + P^m \Sigma^m P^m]
\dot{\Lambda}^m = [(P^m)^{-1} D^m - A^m] \Lambda^m
\dot{R}^m = (\Lambda^m)^T [D^m - (A^m)^T P^m - P^m A^m] \Lambda^m.
$$

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 $\mathcal{B}_r^m (x, z)$ (one time only) as follows. One has

$$
\mathcal{B}_r^m = - \max_{y \in \mathbb{R}^n} \{ \psi (y, x) - S_r^m [\psi (\cdot, z)] (y) \}
$$

which by the above

$$
= \min_{y \in \mathbb{R}^n} \left\{ (c/2)(y - x)^T (y - x) + \frac{1}{2} (y - \Lambda_r^m z)^T P_r^m (y - \Lambda_r^m z) + \frac{1}{2} z^T R_r^m z \right\}.
$$

(65)

Recall that by Theorem 3.6, this has a finite minimum $(P_r^m - (c + \eta r) I)$ positive definite. Taking the minimum in (65), one has

$$
\mathcal{B}_r^m (x, z) = \frac{1}{2} \left[ x^T M_{1,1}^m x + x^T M_{1,2}^m z + z^T (M_{1,2}^m)^T x + z^T M_{2,2}^m z \right]
$$

where with shorthand notation $C = c I$ and $D_r = (P_r^m - c I)^{-1}$,

$$
M_{1,1}^m = \left[ CD_r^{-1} P_r^m D_r^{-1} C - (D_r^{-1} C + I)^T C (D_r^{-1} C + I) \right]
$$

$$
M_{1,2}^m = \left[ (D_r^{-1} C + I)^T C D_r^{-1} P_r^m - C D_r^{-1} P_r^m (D_r^{-1} P_r^m - I) \right] \Lambda_r^m
$$

$$
M_{2,2}^m = (\Lambda_r^m)^T \left[ (D_r^{-1} P_r^m - I)^T P_r^m (D_r^{-1} P_r^m - I) - P_r^m D_r^{-1} C D_r^{-1} P_r^m \right] \Lambda_r^m + R_r^m.
$$

Note that given the $P_r^m, \Lambda_r^m, R_r^m$, the $\mathcal{B}_r^m$ are quadratic functions with analytical expressions for their coefficients. Also note that all the matrices in the definition of $\mathcal{B}_r^m$ may be precomputed.

Now let us write the (quadratic) $\hat{\alpha}_{k \{m \}_{i=1}}^k$ in the form

$$
\hat{\alpha}_{k \{m \}_{i=1}}^k (x) = \frac{1}{2} (x - \hat{z}_{k \{m \}_{i=1}}^k)^T \hat{Q}_{k \{m \}_{i=1}}^k (x - \hat{z}_{k \{m \}_{i=1}}^k) + \hat{r}_{k \{m \}_{i=1}}^k
$$

Then, for each $m_{k+1}$,

$$
\hat{\alpha}_{k+1 \{m \}_{i=1}}^k = \max_{z \in \mathbb{R}^n} \left\{ \mathcal{B}_r^{m_{k+1}} (x, z) + \hat{\alpha}_{k \{m \}_{i=1}}^k (z) \right\}
$$

$$
= \max_{z \in \mathbb{R}^n} \left\{ \frac{1}{2} \left[ x^T M_{1,1}^{m_{k+1}} x + x^T M_{1,2}^{m_{k+1}} z + z^T (M_{1,2}^{m_{k+1}})^T x + z^T M_{2,2}^{m_{k+1}} z \right]
$$

$$
+ \frac{1}{2} (x - \hat{z}_{k \{m \}_{i=1}}^k)^T \hat{Q}_{k \{m \}_{i=1}}^k (x - \hat{z}_{k \{m \}_{i=1}}^k) + \hat{r}_{k \{m \}_{i=1}}^k \right\}
$$

$$
= \frac{1}{2} (x - \hat{z}_{k+1 \{m \}_{i=1}}^{k+1})^T \hat{Q}_{k+1 \{m \}_{i=1}}^{k+1} (x - \hat{z}_{k+1 \{m \}_{i=1}}^{k+1}) + \hat{r}_{k+1 \{m \}_{i=1}}^{k+1}
$$

(66)

where

$$
\hat{Q}_{k+1 \{m \}_{i=1}}^{k+1} = M_{1,1}^{m_{k+1}} - M_{1,2}^{m_{k+1}} \hat{D} (M_{1,2}^{m_{k+1}})^T
$$

$$
\hat{z}_{k+1 \{m \}_{i=1}}^{k+1} = - \left( \hat{Q}_{k+1 \{m \}_{i=1}}^{k+1} \right)^{-1} M_{1,2}^{m_{k+1}} \hat{E}
$$

$$
\hat{r}_{k+1 \{m \}_{i=1}}^{k+1} = \hat{r}_{k \{m \}_{i=1}}^k + \frac{1}{2} \hat{E}^T M_{2,2}^{m_{k+1}} \hat{z}_{k \{m \}_{i=1}}^k - \frac{1}{2} \left( \hat{z}_{k+1 \{m \}_{i=1}}^{k+1} \right)^T \hat{Q}_{k+1 \{m \}_{i=1}}^{k+1} \hat{z}_{k+1 \{m \}_{i=1}}^{k+1} - \frac{1}{2} \left( \hat{z}_{k \{m \}_{i=1}}^k \right)^T \hat{Q}_{k \{m \}_{i=1}}^k \hat{z}_{k \{m \}_{i=1}}^k
$$

$$
\hat{D} = \left( M_{2,2}^{m_{k+1}} + \hat{Q}_{k+1 \{m \}_{i=1}}^{k+1} \right)^{-1}
$$

$$
\hat{E} = \hat{D} \hat{Q}_{k \{m \}_{i=1}}^k \hat{z}_{k \{m \}_{i=1}}^k.
$$
Thus we have the analytical expression for the propagation of each (quadratic) \( \hat{a}_{(m_i)}^{k} \) function. Specifically, we see that the propagation of each \( \hat{a}_{(m_i)}^{k} \) amounts to a set of matrix multiplications (and an inverse).

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

\[
\hat{A}_k \doteq \left\{ \hat{a}_{(m_i)}^{k} \mid m_i \in M \forall i \in \{1, 2, \ldots, k\} \right\}.
\]

where this is equivalently represented as the set of triples

\[
\hat{Q}_k \doteq \left\{ \left( \hat{Q}_{(m_i)}^{k}, \hat{z}_{(m_i)}^{k}, \hat{\rho}_{(m_i)}^{k} \right) \mid m_i \in M \forall i \in \{1, 2, \ldots, k\} \right\}.
\]

At any desired stopping time, one can recover a representation of \( V^k \) as

\[
\hat{V}_k \doteq \left\{ \hat{V}_{(m_i)}^{k} \mid m_i \in M \forall i \in \{1, 2, \ldots, k\} \right\}
\]

where these \( \hat{V}_{(m_i)}^{k} \) are also quadratics. In fact, recall

\[
V^k(x) = \max_{z \in \mathbb{R}^n} [\hat{V}^k(z) + \psi(x, z)]
\]

\[
= \max_{\{m_i\}^k_{i=1}} \max_{z \in \mathbb{R}^n} \left[ \frac{1}{2} (z - \hat{z}_{(m_i)}^{k})^T \hat{Q}_{(m_i)}^{k} (z - \hat{z}_{(m_i)}^{k}) + \hat{\rho}_{(m_i)}^{k} + \frac{\zeta}{2} |x - z|^2 \right]
\]

\[
= \max_{\{m_i\}^k_{i=1}} \frac{1}{2} (x - \hat{x}_{(m_i)}^{k})^T \hat{P}_{(m_i)}^{k} (x - \hat{x}_{(m_i)}^{k}) + \hat{\rho}_{(m_i)}^{k}
\]

\[
= \bigoplus_{\{m_i\}^k_{i=1}} \hat{V}_{(m_i)}^{k}(x)
\]

where with \( C \doteq cI \)

\[
\hat{P}_{(m_i)}^{k} = C\hat{F}Q_{(m_i)}^{k} + (FC + I)^T C(FC + I)
\]

\[
\hat{x}_{(m_i)}^{k} = -\left( \hat{P}_{(m_i)}^{k} \right)^{-1} \left[ C\hat{F}Q_{(m_i)}^{k} \hat{G} + (FC + I)^T C\hat{F}Q_{(m_i)}^{k} \right] \hat{z}_{(m_i)}^{k}
\]

\[
\hat{\rho}_{(m_i)}^{k} = \hat{\rho}_{(m_i)}^{k} + \frac{1}{2} \left( \hat{z}_{(m_i)}^{k} \right)^T \left[ \hat{G}^T \hat{Q}_{(m_i)}^{k} \hat{G} + \hat{Q}_{(m_i)}^{k} \hat{F}C\hat{F}Q_{(m_i)}^{k} \right] \hat{z}_{(m_i)}^{k}
\]

\[
\hat{F} = (\hat{Q}_{(m_i)}^{k})^{-1} + C
\]

and

\[
\hat{G} = (\hat{F}Q_{(m_i)}^{k} + I).
\]

Thus, \( V^k \) has the representation as the set of triples

\[
P_k \doteq \left\{ \left( \hat{P}_{(m_i)}^{k}, \hat{x}_{(m_i)}^{k}, \hat{\rho}_{(m_i)}^{k} \right) \mid m_i \in M \forall i \in \{1, 2, \ldots, k\} \right\}.
\]

We note that the triples which comprise \( P_k \) are analytically obtained from the triples \( (\hat{Q}_{(m_i)}^{k}, \hat{z}_{(m_i)}^{k}, \hat{\rho}_{(m_i)}^{k}) \) by matrix multiplications and an inverse. The transference from \( (\hat{Q}_{(m_i)}^{k}, \hat{z}_{(m_i)}^{k}, \hat{\rho}_{(m_i)}^{k}) \) to \( (\hat{P}_{(m_i)}^{k}, \hat{x}_{(m_i)}^{k}, \hat{\rho}_{(m_i)}^{k}) \) need only be done once
which is at the termination of the algorithm propagation. We note that (67) is our approximate solution of the original control problem/HJB PDE.

The errors are due to our approximation of \( \bar{V} \) by \( V \) (see Theorem 4.11 and Remark 4.12), and to the approximation of \( V \) by the prelimit \( V^N \) for stopping time \( k = N \). Neither of these errors are related to the space dimension. The errors in \( |\bar{V} - V| \) depend on the step size \( \tau \). The errors in \( |\bar{V}^N - V| = |S^T_{N\tau}[0] - V| \) are due to premature termination in the limit \( V = \lim_{N \to \infty} S^T_{N\tau}[0] \). The computation of each triple \( (\hat{P}^k_{\{m_i\}_{i=1}}, \hat{x}^k_{\{m_i\}_{i=1}}, \hat{\rho}^k_{\{m_i\}_{i=1}}) \) 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 \( V^N(x) \) for all \( x \) on some grid over say a rectangular region in \( \mathbb{R}^n \), 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 \( V^N \simeq \bar{V} \) 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 \#\( M \) = 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 \( M = \#\{H\} \) such quadratics in the Hamiltonian, \( \tilde{H} \), we say it has complexity \( \tilde{M} \). Note that

\[
\# \{ \hat{V}^k_{\{m_i\}_{i=1}} | m_i \in M \forall i \in \{1, 2, \ldots, k\} \} \sim M^N.
\]

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.

6.1 Pruning

The number of quadratics in \( Q_k \) grows exponentially in \( k \). However, in practice (for the cases we have tried) we have found that relatively few of these actually contribute to \( V^k \).
Thus it would be very useful to prune the set.

Note that if

\[ \hat{a}^k_{\{\hat{m}_i\}_{i=1}^k} (x) \leq \bigoplus_{\{m_i\}_{i=1}^k \neq \{\hat{m}_i\}_{i=1}^k} \hat{a}^k_{\{m_i\}_{i=1}^k} (x) \quad \forall x \in \mathbb{R}^n, \quad (68) \]

then

\[ \int_{\mathbb{R}^n} \mathcal{B}_r(x,z) \otimes \pi^k(z) \, dz \leq \int_{\mathbb{R}^n} \mathcal{B}_r(x,z) \otimes \left[ \bigoplus_{\{m_i\}_{i=1}^k \neq \{\hat{m}_i\}_{i=1}^k} \hat{a}^k_{\{m_i\}_{i=1}^k} (z) \right] \, dz. \]

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

In the examples to follow, we pruned \( \hat{a}^k_{\{\hat{m}_i\}_{i=1}^k} \) if there existed a single sequence \( \{\hat{m}_i\}_{i=1}^k \) such that \( \hat{a}^k_{\{\hat{m}_i\}_{i=1}^k} (x) \leq \hat{a}^k_{\{\hat{m}_i\}_{i=1}^k} (x) \) for all \( x \). This significantly reduced the growth in the size of \( Q_k \). However, it clearly failed to prune anywhere near the number of elements that could be pruned according to condition (68), 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.

### 6.2 Initialization

It is also easy to see that one may initialize with an arbitrary quadratic function less than an \( \bar{\pi}^k(x) \) rather than with \( \bar{\pi}^0 \equiv 0 \). Significant savings were obtained by initializing with a set of \( M = \#M \) 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 \( \bar{\pi}^0(z) = \bigoplus_{m \in M} a^m(z) \), 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.

### 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 \( M = \#M \) 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 will be for a standard 2001 PC. The times correspond to the time to compute \( V_N \) or, equivalently, \( P_N \). 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.

The first example 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 $\Sigma^m$ were 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 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 2 depicts the second partial and a backsubstitution error over the same region. 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.
The second example has constituent Hamiltonians with the $A^m$ 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 $D^m$ were
\[
D^1 = \begin{bmatrix}
1.5 & 0.2 & 0.1 \\
0.2 & 1.5 & 0.0 \\
0.1 & 0.0 & 1.5
\end{bmatrix}, \quad D^2 = \begin{bmatrix}
1.6 & 0.2 & 0.1 \\
0.2 & 1.6 & 0.0 \\
0.1 & 0.0 & 1.6
\end{bmatrix}, \quad D^3 = D^1.
\]

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

The results of the three-dimensional example appear in Figures 3–5. 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 $|x|$. Since the gradient errors are multiplied by the gradient 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 third example has constituent Hamiltonians with the $A^m$ given by
\[
A^1 = \begin{bmatrix}
-1.0 & 0.5 & 0.0 & 0.1 \\
0.1 & -1.0 & 0.2 & 0.0 \\
0.2 & 0.0 & -1.5 & 0.1 \\
0.0 & -0.1 & 0.0 & -1.5
\end{bmatrix}, \quad A^2 = (A^1)^T,
\]
Figure 3: Value function and first partial (3-D case)

Figure 4: Second and third partials (3-D case)

Figure 5: Scaled backsubstitution error (3-D case)
\[ A^3 = \begin{bmatrix} -1.0 & 0.5 & 0.0 & 0.1 \\ 0.1 & -1.0 & 0.2 & 0.0 \\ 0.2 & 0.0 & -1.6 & -0.1 \\ 0.0 & -0.05 & 0.1 & -1.5 \end{bmatrix} . \]

The \( D^m \) and \( \Sigma^m \) were simply

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

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 of the four-dimensional example appear in Figures 6–8. 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 \( |x|^2 + 10^{-5} \).

![Figure 6: Value function and first partial (4-D case)](image)

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 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, \( \tilde{H}(x, p) = \max_m \{H^m(x, p)\} \), where the \( H^m \) corresponded to a very specific type of linear/quadratic problem. One would like to generalize the \( H^m \) to say
\[
H^m(x, p) = \frac{1}{2} x^T D^m x + \frac{1}{2} p^T \Sigma^m p + (A^m x)^T p + (l^m_1)^T x + (l^m_2)^T p + \alpha^m.
\]
Clearly certain conditions on \( \tilde{H}(x, p) = \max_m \{H^m(x, p)\} \) would be necessary. It is not obvious that these conditions would need to apply for 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 so constricted. 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 [27]. 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 would be the most tractable?

Appendix A (sketch of proof of Theorem 4.6):

Fix $\delta > 0$ (used in the definition of $G_\delta$). Suppose $\nabla' \in G_\delta$ satisfies (40). Then,

$$\nabla'(x) = \bar{S}_{N\tau}^\gamma \nabla'(x)$$

$$= \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \left\{ \int_0^{N\tau} l^{\mu}(\xi_\tau) - \frac{\gamma^2}{2} |w_\tau|^2 dt + \nabla'(\xi_{N\tau}) \right\} \quad \forall x \in \mathbb{R}^n$$

where $\xi$ satisfies (14). Fix $x \in \mathbb{R}^n$, and let $\mu^\varepsilon \in \mathcal{D}_\infty$, $w^\varepsilon \in \mathcal{W}$ be $\varepsilon$-optimal, i.e.

$$\nabla'(x) \leq \int_0^{N\tau} l^{\mu^\varepsilon}(\xi^\varepsilon_\tau) - \frac{\gamma^2}{2} |w^\varepsilon_\tau|^2 dt + \nabla'(\xi^\varepsilon_{N\tau}) + \varepsilon$$

where $\xi^\varepsilon$ satisfies (14) with inputs $\mu^\varepsilon, w^\varepsilon$.

Following the same steps as in [33], one obtains the same lemmas:

Lemma A.1 For any $N < \infty$,

$$\|w^\varepsilon\|_{L_2(0,N\tau)} \leq \frac{\varepsilon}{\delta} + \frac{1}{\delta} \left[ \frac{c_\gamma^2}{c_\sigma^2} e^{-c_\gamma N\tau} + \frac{c_D}{c_A} \right] |x|^2.$$
Lemma A.2 For any $N < \infty$,
\[ \int_0^{N\tau} |\xi_t^\varepsilon|^2 dt \leq \frac{\varepsilon}{\delta c_A} \frac{c_\sigma^2}{\delta c_A} + \frac{c_\sigma^2}{\delta c_A} \left[ \left( \frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2} \right) + \frac{1}{c_A} \right] |x|^2. \]

Lemma A.3 If $w^\varepsilon, \mu^\varepsilon$ are $\varepsilon$–optimal over $[0, N\tau)$, then they are also $\varepsilon$–optimal over $[0, n\tau)$ for all $n \leq N$, i.e.
\[ \int_0^{n\tau} \mu_t^\varepsilon (\xi_t^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + \nabla' (\xi_{n\tau}^\varepsilon) \geq \nabla'(x) - \varepsilon. \]

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 $w^\varepsilon$) coming in to the trajectories, roughly speaking the $\xi^\varepsilon$ “tend” toward the origin.

Now we need a lemma which will replace equation (20) in [33].

Lemma A.4 For any $N < \infty$,
\[ \sum_{n=1}^{N} |\xi_{n\tau}^\varepsilon|^2 \leq \frac{1}{1 - e^{-c_A \tau}} \left[ |x|^2 + (c_\sigma/c_A) \|w^\varepsilon\|_{L_2(0,N\tau)}^2 \right]. \]

Proof. Note that
\[ \frac{d}{dt} |\xi_t^\varepsilon|^2 \leq -2c_A |\xi_t^\varepsilon|^2 + 2c_\sigma |\xi_t^\varepsilon| |w^\varepsilon| \leq -c_A |\xi_t^\varepsilon| + \hat{d} |w^\varepsilon|^2 \]
with $\hat{d} = c_\sigma^2/c_A$. Solving this on interval $[n\tau, (n+1)\tau)$ implies that
\[ |\xi_{t}^\varepsilon|^2 \leq |\xi_{n\tau}^\varepsilon|^2 e^{-c_A \tau} + \hat{d} \|w^\varepsilon\|_{L_2(n\tau,(n+1)\tau)}^2 \quad \forall t \in [n\tau, (n+1)\tau]. \]
In particular, one has
\[ |\xi_{\tau}^\varepsilon|^2 \leq |x|^2 e^{-c_A \tau} + \hat{d} \|w^\varepsilon\|_{L_2(0,\tau)}^2, \]
\[ |\xi_{2\tau}^\varepsilon|^2 \leq |\xi_{\tau}^\varepsilon|^2 e^{-c_A \tau} + \hat{d} \|w^\varepsilon\|_{L_2(\tau,2\tau)}^2, \]
and these two inequalities imply
\[ |\xi_{\tau}^\varepsilon|^2 + |\xi_{2\tau}^\varepsilon|^2 \leq (e^{-c_A \tau} + e^{-2c_A \tau}) |x|^2 + \hat{d} (1 + e^{-c_A \tau}) \|w^\varepsilon\|_{L_2(0,\tau)}^2 + \hat{d} \|w^\varepsilon\|_{L_2(\tau,2\tau)}^2. \]
Continuing this process, one finds
\[ \sum_{n=1}^{N} |\xi_{n\tau}^\varepsilon|^2 \leq \left( \sum_{n=1}^{N} e^{-nc_A \tau} \right) |x|^2 + \hat{d} \sum_{n=1}^{N} \left[ \left( \sum_{j=0}^{N-n} e^{-jc_A \tau} \right) \|w^\varepsilon\|_{L_2((n-1)\tau,n\tau)}^2 \right]. \]
Using the standard geometric series limit yields the result. □

Combining Lemmas A.2 and A.4, one obtains a bound on \( \sum_{n=1}^{N} |\xi_{n\tau}|^2 \) which is independent of \( N \). Consequently, at least some of the \( |\xi_{n\tau}| \) can be guaranteed to be arbitrarily small for large \( N \). The remainder of the proof (of Theorem 4.6) then follows as in equations (24) to (28) in [33], but with \( N\tau \) replacing \( T \), and \( n\tau \) replacing \( \tau \). This completes the sketch of the proof.

Appendix B (sketch of proof of Lemma 4.10):

Fix \( \delta > 0 \) (used in the definition of \( G_\delta \)). 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 \( \varepsilon \)-optimal for \( \tilde{S}_T[V^m](x) \), i.e.

\[
\tilde{S}_T[V^m](x) - \left[ \int_0^T l^\varepsilon(\xi_t) - \frac{\gamma^2}{2} |w_t|^2 dt + V^m(\xi_T) \right] \leq \varepsilon = \frac{\hat{\varepsilon}}{2}(1 + |x|^2) \tag{69}
\]

where \( \xi^\varepsilon \) satisfies (14) with inputs \( w^\varepsilon \), \( \mu^\varepsilon \).

We will let \( \xi^\varepsilon \) satisfy (14) with inputs \( w^\varepsilon \) and a \( p^\varepsilon \in \mathcal{D}_\infty \) (where \( \tau \) has yet to be chosen). Solving (14), one has

\[
\xi^\varepsilon_t = \exp \int_0^t A^\mu^\varepsilon r \, dr \, x + \int_0^t \exp \int_r^t A^\mu^\varepsilon p \, dp \, \sigma^\mu^\varepsilon w^\varepsilon r \, dr \\
\bar{\xi}^\varepsilon_t = \exp \int_0^t A^\mu^\varepsilon r \, dr \, x + \int_0^t \exp \int_r^t A^\mu^\varepsilon p \, dp \, \sigma^\mu^\varepsilon w^\varepsilon r \, dr.
\]

Consequently,

\[
|\xi^\varepsilon_t - \bar{\xi}^\varepsilon_t| \leq \left| \exp \int_0^t A^\mu^\varepsilon r \, dr - \exp \int_0^t A^\mu^\varepsilon r \, dr \right| |x| \\
+ \left\{ \int_0^t \left| \exp \int_r^t A^\mu^\varepsilon p \, dp \right| \sigma^\mu^\varepsilon - \exp \int_r^t A^\mu^\varepsilon p \, dp \sigma^\mu^\varepsilon r \right|^2 dr \right\}^{1/2} \|w^\varepsilon\|_{L_2(0,t)}. \tag{70}
\]

We now simply show that this can be made arbitrarily small by taking \( \tau \) small. We will use the boundedness of \( \|w^\varepsilon\| \) and \( \|\xi^\varepsilon\| \) which is independent of \( t \) for this class of systems [33].

Consider the first term on the right in (70). Note that

\[
\left| \exp \int_0^t A^\mu^\varepsilon r \, dr - \exp \int_0^t A^\mu^\varepsilon r \, dr \right| = \left| \exp \int_0^t A^\mu^\varepsilon r \, dr \right| \left| 1 - \exp \int_0^t A^\mu^\varepsilon r - A^\mu^\varepsilon r \, dr \right|.
\tag{71}
\]

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

\[
\mathcal{I}_m = \{ r \in [0, N\tau) \mid A^\mu^\varepsilon r = A^m \}
\]

\[
\lambda^m = \mathcal{L}(\mathcal{I}_m).
\]
Let $n_0 = 0$. For $1 \leq k < M = \# M$, let $n_k$ be the largest integer such that $n_k \tau \leq \lambda^k + n_{k-1} \tau$. For $m < M$ let

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

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

$$1 - \exp \left[ \int_0^t A^\overline{\mu}_r \, dr - \int_0^t A^{\mu_\tau} \, dr \right] < \beta_1^1 \tau$$

(72)

where $\beta_1^1 \to 0$ as $\tau \to 0$ independent of $t$. We skip the details.

Let $y \in \mathbb{R}^n$. Define $F_t = \exp[\int_0^t A^{\mu_\tau} \, dr]$. Then,

$$\frac{d}{dt} \left[ y^T F_t^T F_t y \right] = y^T \left[ F_t^T \dot{F_t} + \dot{F_t}^T F_t \right] y = 2y^T \left[ F_t^T A^{\mu_\tau} F_t \right] y$$

which be Assumption Block (A.m)

$$\leq -2c_A |F_t y|^2 = -2c_A [y^T F_t^T F_t y].$$

Solving this ordinary differential inequality, one finds

$$[y^T F_t^T F_t y] \leq |y|^2 e^{-2c_A t}.$$

Since this is true for all $y \in \mathbb{R}^n$, we have

$$\left| \exp \left[ \int_0^t A^{\mu_\tau} \, dr \right] \right| \leq e^{-c_A t} \quad \forall t \geq 0. \quad (73)$$

By (71), (72) and (73)

$$\left| \exp \left[ \int_0^t A^{\mu_\tau} \, dr \right] - \exp \left[ \int_0^t A^{\overline{\mu}_r} \, dr \right] \right| \leq \beta_1^1 e^{-c_A t} \quad \forall t \geq 0. \quad (74)$$

We now turn to the second term on the right-hand side of (70). Note that

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

$$\leq \left\{ 2 \int_0^t \left| \exp \left[ \int_r^t A^{\mu_\tau} \, d\rho \right] \right|^2 \left| \sigma^{\mu_\tau} - \sigma^{\overline{\mu}_r} \right|^2 \, dr \right.$$

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

and proceeding as above

$$\leq \left\{ 2 \int_0^t e^{-2c_A(t-r)} \left| \sigma^{\mu_\tau} - \sigma^{\overline{\mu}_r} \right|^2 \, dr + 2\beta_1^1 \int_0^t e^{-2c_A(t-r)} \left| \sigma^{\overline{\mu}_r} \right|^2 \, dr \right\}^{1/2}$$

$$\leq \left\{ 2 \int_0^t e^{-4c_A(t-r)} \, dr \right\}^{1/2} \left[ \int_0^t \left| \sigma^{\mu_\tau} - \sigma^{\overline{\mu}_r} \right|^4 \, dr \right]^{1/2} + 2\beta_1^2 e^2 \int_0^t e^{-2c_A(t-r)} \, dr \right\}^{1/2}.$$
Further, there exists $\beta^2_\tau$ such that $[\int_0^t |\sigma^{\mu\nu} - \sigma^{\tau\nu}|^4 \, dr]^{1/2} \leq \beta^2_\tau$ where $\beta^2_\tau \to 0$ as $\tau \to 0$, and we skip the obvious but technical proof. Consequently,

$$\left\{ \int_0^t \left| \exp\left[ \int_r^t A^{\mu\nu}_r \, d\rho \right] \sigma^{\mu\nu} - \exp\left[ \int_r^t A^{\tau\nu}_r \, d\rho \right] \sigma^{\tau\nu} \right|^2 \, dr \right\}^{1/2} \leq \{2\beta^2_\tau (4c_A)^{-1/2} + 2\beta^1_\tau \sigma_\sigma (2c_A)^{-1}\}^{1/2} \leq \beta^3_\tau$$

(75)

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

Combining (70), (74) and (75), one has

$$|\xi^e_t - \xi^\tau_T| \leq \beta^1_\tau e^{-c_A t} |x| + \beta^3_\tau \|w^e\|_{L^2(0,t)}.$$  

(76)

Now, by the system structure given by Assumption Block (A.m) and by the fact that the $V^m$ are in $G_\delta$, one obtains the following lemmas exactly as in [33]. These are also analogous to their counterparts in Appendix A.

**Lemma B.1** For any $t < \infty$,

$$\|w^e\|_{L^2(0,t)}^2 \leq \frac{\varepsilon}{\delta} + \left[ \frac{c_A \gamma^2}{c_A} e^{-c_A N \tau} + \frac{c_D}{c_A} \right] |x|^2.$$

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

$$\int_0^t |\xi^e_r|^2 \, dr \leq \frac{\varepsilon}{\delta} \frac{c_A^2}{c_A} + \frac{c_A^2}{\delta} \left[ \left( \frac{c_D}{c_A} + \frac{\gamma^2}{c_A^2} \right) + \frac{1}{c_A} \right] |x|^2.$$

Let $c_1 = \frac{\varepsilon}{\delta}$ and $c_2 = \frac{1}{\delta} \left[ \frac{c_A \gamma^2}{c_A} e^{-c_A N \tau} + \frac{c_D}{c_A} \right]$. By Lemma B.1 and (76), for all $t < \infty$ one has

$$|\xi^e_t - \xi^\tau_T| \leq \beta^1_\tau e^{-c_A t} |x| + \beta^3_\tau (c_1 + c_2 |x|^2)^{1/2}$$

and by proper choice of $\beta^4_\tau$,

$$\leq \beta^4_\tau (1 + |x|)$$

(77)

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

Now,

$$\int_0^T l^{\mu\nu}(\xi^e_t) - \frac{\gamma^2}{2} |w^e_t|^2 \, dt + V^m(\xi^e_T)$$

$$- \int_0^T \tilde{l}^{\mu\nu}(\xi^e_t) - \frac{\gamma^2}{2} |\tilde{w}^e_t|^2 \, dt + V^m(\xi^\tau_T)$$

$$= \int_0^T \xi_t D^{\mu\nu} \xi_t - \xi_t D^{\mu\nu} \xi_t \, dt + (\xi^e_T)^T P^m \xi^e_T - (\xi^\tau_T)^T P^m \xi^\tau_T.$$

(78)
Note that the integral term on the right-hand side in (78) is
\[
\int_0^T (\xi_T^\varepsilon)^T D\mu^\varepsilon ((\xi_T^\varepsilon - \overline{\xi}_T) + (\xi_T^\varepsilon)^T (D\mu^\varepsilon - D\overline{\mu}_T) \overline{\xi}_T + (\xi_T^\varepsilon - \overline{\xi}_T)^T D\overline{\mu}_T \overline{\xi}_T \, dt
\]
\[
\leq \beta_\tau^1 (1 + |x|) \int_0^T \left( |D\mu^\varepsilon| |\xi_T^\varepsilon| + |D\overline{\mu}_T| |\overline{\xi}_T| \right) \, dt + \beta_\tau^2 \int_0^T |\xi_T^\varepsilon| |\overline{\xi}_T| \, dt
\]
for appropriate $\beta_\tau^5 \to 0$ as $\tau \to 0$, which after some work
\[
\leq \beta_\tau^6 (1 + |x|^2)(1 + \sqrt{T})
\]
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 (78) are
\[
\xi_T^\varepsilon^T Pm^\varepsilon \xi_T^\varepsilon - \overline{\xi}_T^\varepsilon^T Pm^\varepsilon \overline{\xi}_T^\varepsilon = (\xi_T^\varepsilon + \overline{\xi}_T^\varepsilon)^T Pm^\varepsilon (\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon)
\]
\[
\leq |Pm^\varepsilon| \left( |\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon|^2 + 2|\xi_T^\varepsilon| |\xi_T^\varepsilon - \overline{\xi}_T^\varepsilon| \right)
\]
which by (77)
\[
\leq \beta_\tau^7 (1 + |x|^2) + \beta_\tau^8 |\xi_T^\varepsilon| (1 + |x|)
\]
where $\beta_\tau^7, \beta_\tau^8 \to 0$ as $\tau \to 0$.

We also need the following lemma which is obtained in [33].

**Lemma B.3** Given $T < \infty$, there exist $T \in [T/2, T]$ and $\varepsilon$-optimal $w^\varepsilon \in \mathcal{W}, \mu^\varepsilon \in \mathcal{D}_\infty$ for $\tilde{S}_T[V^m]$ such that
\[
|\xi_T^\varepsilon|^2 \leq \frac{1}{T} \left\{ \frac{\varepsilon}{\delta c_A} + \frac{c_\sigma^2}{\delta} \left[ \left( \frac{c_D}{\varpi A} \right)^2 + \frac{\gamma^2}{\varpi A} \right] + \frac{1}{c_A} \right\} |x|^2 \}
\]

Combining (80) and Lemma B.3, one finds that there exist $c_3, c_4 < \infty$ such that
\[
\xi_T^\varepsilon^T Pm^\varepsilon \xi_T^\varepsilon - \overline{\xi}_T^\varepsilon^T Pm^\varepsilon \overline{\xi}_T^\varepsilon \leq \beta_\tau^9 (1 + |x|^2) \]
\[
where \beta_\tau^9 \to 0 \text{ as } \tau \to 0 \text{ (independent of } T\text{).}
\]

Combining (78), (79) and (81),
\[
\int_0^T \nu^\varepsilon (\xi_T^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 \, dt + V^m(\xi_T^\varepsilon) - \int_0^T \nu^\varepsilon (\overline{\xi}_T^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 \, dt + V^m(\overline{\xi}_T^\varepsilon)
\]
\[
\leq \beta_\tau^{10} (1 + |x|^2)(1 + \sqrt{T})
\]
where $\beta_\tau^{10} \to 0$ as $\tau \to 0$ (independent of $T$).

Combining (69) and (82), one has
\[
\tilde{S}_T[V^m](x) - \int_0^T \nu^\varepsilon (\xi_T^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 \, dt + V^m(\overline{\xi}_T^\varepsilon) \leq \varepsilon (1 + |x|^2) + \beta_\tau^{10} (1 + |x|^2)(1 + \sqrt{T})
\]
which for $\tau$ sufficiently small (depending on $T$ now),
\[
\leq \varepsilon (1 + |x|^2).
\]

This completes the proof of Lemma 4.10.
References


