

Convergence Rate for a Curse-of-Dimensionality-Free Method for a Class of HJB PDEs

William M. McEneaney ^{*} L. Jonathan Kluberg [†]

March 20, 2009

Abstract

In previous work of the first author and others, max-plus methods have been explored for solution of first-order, nonlinear Hamilton-Jacobi-Bellman partial differential equations (HJB PDEs) and corresponding nonlinear control problems. Although max-plus basis expansion and max-plus finite-element methods can provide substantial computational-speed advantages, they still generally suffer from the curse-of-dimensionality. Here we consider HJB PDEs where the Hamiltonian takes the form of a (pointwise) maximum of linear/quadratic forms. The approach to solution will be rather general, but in order to ground the work, we consider only constituent Hamiltonians corresponding to long-run average-cost-per-unit-time optimal control problems for the development. We consider a previously obtained numerical method not subject to the curse-of-dimensionality. The method is based on construction of the dual-space semigroup corresponding to the HJB PDE. This dual-space semigroup is constructed from the dual-space semigroups corresponding to the constituent linear/quadratic Hamiltonians. The dual-space semigroup is particularly useful due to its form as a max-plus integral operator with kernel obtained from the originating semigroup. One considers repeated application of the dual-space semigroup to obtain the solution. Although previous work indicated that the method was not subject to the curse-of-dimensionality, it did not indicate any error bounds or convergence rate. Here, we obtain specific error bounds.

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

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

^{*}Dept. of Mechanical and Aerospace Eng., University of California San Diego, wmceneaney@ucsd.edu
Research partially supported by NSF grant DMS-0307229 and AFOSR grant FA9550-06-1-0238.

[†]Operations Research Center, Massachusetts Institute of Technology

1 Introduction

A robust approach to the solution of nonlinear control problems is through the general method of dynamic programming. For the typical class of problems in continuous time and continuous space, with the dynamics governed by finite-dimensional, ordinary differential equations, this leads to a representation of the problem as a first-order, nonlinear partial differential equation, the Hamilton-Jacobi-Bellman equation – or the HJB PDE. If one has an infinite time-horizon problem, then the HJB PDE is a steady-state equation, and this PDE is over a space (or some subset thereof) whose dimension is the dimension of the state variable of the control problem. Due to the nonlinearity, the solutions are generally nonsmooth, and one must use the theory of viscosity solutions [4], [9], [10], [11], [21].

The most intuitive class of approaches to solution of the HJB PDE consists of grid-based methods (cf. [4], [6], [13], [14], [15], [17], [21], [25] among many others). These require that one generate a grid over some bounded region of the state space. In particular, suppose the region over which one constructs the grid is rectangular, say square for simplicity. Further, suppose one uses N grid points per dimension. If the state dimension is n , then one has N^n grid points. Thus the computations grow exponentially in state-space dimension n , and this is referred to as the curse-of-dimensionality.

In [29], [30], [31], [32], a new class of methods for first-order HJB PDEs was introduced, and these methods are not subject to the curse-of-dimensionality. A different class of methods which also utilize the max-plus algebra are those which expand the solution over a max-plus basis, and solve for the coefficients in the expansion via max-plus linear algebra (c.f., [1], [2], [29], [35], [36]). Although this new approach bares a superficial resemblance to these other methods in that it utilizes the max-plus algebra, it is largely unrelated. Most notably, with this new approach, the computational growth in state-space dimension is on the order of n^3 . There is of course no “free lunch”, and there is exponential computational growth in a certain measure of complexity of the Hamiltonian. Under this measure, the minimal complexity Hamiltonian is the linear/quadratic Hamiltonian – corresponding to solution by a Riccati equation. If the Hamiltonian is given as a pointwise maximum or minimum of M linear/quadratic Hamiltonians, then one could say the *complexity* of the Hamiltonian is M . One could also apply this approach to a wider class of HJB PDEs with semiconvex Hamiltonians (by approximation of the Hamiltonian by a finite number of quadratic forms), but that is certainly beyond the scope of this paper.

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

$$\widetilde{H}(x, \nabla V) = \max_{m \in \mathcal{M}} \{H^m(x, \nabla V)\} \quad (1)$$

where $\mathcal{M} = \{1, 2, \dots, M\}$ and the H^m have computationally simpler forms. In order to make the problem tractable, we will concentrate on a single class of HJB PDEs – those for long-run average-cost-per-unit-time problems. However, the theory can clearly be expanded to a much larger class.

In [29], [30], a curse-of-dimensionality-free algorithm was developed in the case where each of the constituent H^m was a quadratic function of its arguments. In particular, we had

$$H^m(x, p) = (A^m x)'p + \frac{1}{2}x'D^m x + \frac{1}{2}p'\Sigma^m p, \quad (2)$$

where the A^m , D^m and Σ^m were $n \times n$ matrices meeting certain conditions which guaranteed existence and uniqueness of a solution within a certain class of functions. First, some existence and uniqueness results were reviewed, and an approximately equivalent form as a fixed point of a semigroup, $\tilde{S}_\tau = \bigoplus_{m \in \mathcal{M}} S_\tau^m$ (where \oplus and \otimes indicate max-plus addition/summation and multiplication) was discussed. The analysis leading to the curse-of-dimensionality-free algorithm was developed. We briefly indicate the main points here, and more details appear in Sections 2 and 3.

In one sense, the curse-of-dimensionality-free method computes the solution of $0 = -\tilde{H}(x, \nabla V)$ with Hamiltonian (1) (and boundary condition $V(0) = 0$) through repeated application of \tilde{S}_τ to some initial function, V^0 , say N times, yielding approximation $\tilde{S}_T V^0 = [\tilde{S}_\tau]^N V^0$ where $T = N\tau$, and the superscript N indicates repeated composition. However, the operations are all carried out in the semiconvex-dual space (where semiconvex duality is defined for the reader in Section 3.2). Suppose $V^1 = \tilde{S}_\tau V^0$, and that V^0 has semiconvex dual a^0 . Then, one may propagate instead in the dual space with $a^1 = \hat{\tilde{\mathcal{B}}}_\tau a^0$, and recover V^1 as the inverse semiconvex dual of a^1 . It is natural to propagate in the dual space because one automatically obtains $\hat{\tilde{\mathcal{B}}}_\tau$ as a max-plus integral operator, that is,

$$\hat{\tilde{\mathcal{B}}}_\tau[a^0](z) = \int_{\mathbb{R}^n}^{\oplus} \tilde{\mathcal{B}}_\tau(z, y) \otimes a^0(y) dy$$

(where the definitions are given below). Importantly, one has $\tilde{\mathcal{B}}_\tau(z, y) \simeq \bigoplus_{m \in \mathcal{M}} \mathcal{B}_\tau^m(z, y)$ where the \mathcal{B}_τ^m are dual to the S_τ^m . The key to the algorithm is that when the H^m are quadratic as in (2), then the \mathcal{B}_τ^m are quadratic functions. If one takes a^0 to be a quadratic function, then $a^1(z) = \hat{\tilde{\mathcal{B}}}_\tau[a^0](z) = \bigoplus_{m \in \mathcal{M}} \int_{\mathbb{R}^n}^{\oplus} \mathcal{B}_\tau^m(z, y) \otimes a^0(y) dy = \bigoplus_{m \in \mathcal{M}} \hat{a}_m^1(z)$ where, as the max-plus integral is a supremum operation, the \hat{a}_m^1 are obtained analytically (modulo a matrix inverse) as maxima of sums of quadratics, $\mathcal{B}_\tau^m(z, y) + a^0(y)$. At the second step, one obtains $a^2(z) = \bigoplus_{m_1, m_2 \in \mathcal{M}^2} \hat{a}_{m_1, m_2}^2(z)$ where the $\hat{a}_{m_1, m_2}^2(z)$ are again obtained analytically. Thus, the computational growth in space dimension is only cubic (due to the matrix inverse). The rapid growth in cardinality of the set of $\hat{a}_{\{m_j\}}^k$ is what we refer to as the curse-of-complexity, and we briefly discuss pruning as a means for complexity attenuation as well (although this is not the focus of this paper). The method allows us to solve HJB PDEs that would otherwise be intractable. A simple example over \mathbb{R}^6 appears in [28].

In [29], [30], the algorithm was explicated. Although it was clear that the method converged to the solution, and that the computational-cost growth was only at a rate proportional to n^3 , no convergence rate or error analysis was performed. In this paper, we obtain error bounds as a function of the number of iterations. In particular, two parameters define the convergence, with the first one, T , being the time-horizon, and

going to infinity. The second, τ , is the time-step size, and it goes to zero. Of course, the number of iterations is T/τ . In Sections 5 and 8, we indicate an error bound as a function of these two parameters.

2 Problem Class

There are certain conditions which must be satisfied for solutions to exist, to be unique within an appropriate class, and for the method to converge to the solution. In order that the assumptions are not completely abstract, we work with a specific problem class – the long-run average-cost-per-unit-time optimal control problem. This is a problem class where there already exist a great many results, and so less analysis is required. More specifically, we are interested in solving HJB PDEs of the form (1), and of course equivalently, the corresponding control problems. We refer to the H^m in (1) as the constituent Hamiltonians. As indicated above, we suppose the individual constituent H^m are quadratic forms. These constituent Hamiltonians have corresponding HJB PDE problems which take the form

$$0 = -H^m(x, \nabla V), \quad V(0) = 0. \quad (3)$$

As the constituent Hamiltonians are given by (2), they are associated, at least formally, with the following (purely quadratic) control problems. In particular, the dynamics take the form

$$\dot{\xi}^m = A^m \xi^m + \sigma^m w, \quad \xi_0^m = x \in \mathbb{R}^n. \quad (4)$$

where the nature of σ^m is specified just below. Let $w \in \mathcal{W} \doteq L_2^{loc}([0, \infty); \mathbb{R}^k)$, and we recall that $L_2^{loc}([0, \infty); \mathbb{R}^k) = \{w : [0, \infty) \rightarrow \mathbb{R}^k : \int_0^T |w_t|^2 dt < \infty \forall T < \infty\}$. The cost functionals in the problems are

$$J^m(x, T; w) \doteq \int_0^T \frac{1}{2} (\xi_t^m)' D^m \xi_t^m - \frac{\gamma^2}{2} |w_t|^2 dt, \quad (5)$$

where we use $|\cdot|$ to indicate vector and induced matrix norms. The value functions are $V^m(x) = \lim_{T \rightarrow \infty} \sup_{w \in \mathcal{W}} J^m(x, T; w)$. Lastly, σ^m and γ are such that $\Sigma^m = \frac{1}{\gamma^2} \sigma^m (\sigma^m)'$.

We remark that a generalization of the second term in the integrand of the cost functional to $\frac{1}{2} w' C^m w$ with C^m symmetric, positive definite is not needed since this is equivalent to a change in σ^m in the dynamics (4).

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

$$x' A^m x \leq -c_A |x|^2 \quad \forall x \in \mathbb{R}^n, m \in \mathcal{M}.$$

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

$$|\sigma^m| \leq c_\sigma \quad \forall m \in \mathcal{M}. \tag{A.m}$$

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

$$x' D^m x \leq c_D |x|^2 \quad \forall x \in \mathbb{R}^n, m \in \mathcal{M}$$

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

We note here that we will take $\sigma^m \doteq \sigma$ (independent of m) in the error estimates.

3 Review of the basic concepts

The theory in support of the algorithm can be found in [29], [30] (without error bounds). We summarize it here.

3.1 Solutions and semigroups

First, we indicate the associated semigroups and some existence and uniqueness results. Assumption Block (A.m) guarantees the existence of the V^m as locally bounded functions which are zero at the origin (cf. [37]). The corresponding HJB PDEs are

$$\begin{aligned} 0 &= -H^m(x, \nabla V) = - \left\{ \frac{1}{2} x' D^m x + (A^m x)' \nabla V + \max_{w \in \mathbb{R}^m} \left[(\sigma^m w)' \nabla V - \frac{\gamma^2}{2} |w|^2 \right] \right\} \\ &= - \left\{ \frac{1}{2} x' D^m x + (A^m x)' \nabla V + \frac{1}{2} \nabla V' \Sigma^m \nabla V \right\} \\ V(0) &= 0. \end{aligned} \tag{6}$$

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

$$\mathcal{G}_\delta = \left\{ V : \mathbb{R}^n \rightarrow [0, \infty) \mid V \text{ is semiconvex and } 0 \leq V(x) \leq \frac{c_A(\gamma - \delta)^2}{2c_\sigma^2} |x|^2, \forall x \in \mathbb{R}^n \right\}.$$

From the structure of the running cost and dynamics, it is easy to see (c.f. [41], [37]) that each V^m satisfies

$$V^m(x) = \sup_{T < \infty} \sup_{w \in \mathcal{W}} J^m(x, T; w) = \lim_{T \rightarrow \infty} \sup_{w \in \mathcal{W}} J^m(x, T; w) \doteq \lim_{T \rightarrow \infty} V^{m,f}(x, T), \quad (7)$$

and that each $V^{m,f}$ is the unique continuous viscosity solution of (cf. [4], [21]) $0 = V_T - H^m(x, \nabla V)$, $V(x, 0) = 0$. It is easy to see that these solutions have the form $V^{m,f}(x, t) = \frac{1}{2}x'P_t^{m,f}x$ where each $P^{m,f}$ satisfies the differential Riccati equation

$$\dot{P}^{m,f} = (A^m)'P^{m,f} + P^{m,f}A^m + D^m + P^{m,f}\Sigma^m P^{m,f}, \quad P_0^{m,f} = 0. \quad (8)$$

By (7) and (8), the V^m take the form $V^m(x) = \frac{1}{2}x'P^m x$ where $P^m = \lim_{t \rightarrow \infty} P_t^{m,f}$. One can show that the P^m are the smallest symmetric, positive definite solutions of their corresponding algebraic Riccati equations. The method we will use to obtain value functions/HJB PDE solutions of the $0 = -\widetilde{H}(x, \nabla V)$ problems is through the associated semigroups. For each m define the semigroup

$$S_T^m[\phi] \doteq \sup_{w \in \mathcal{W}} \left[\int_0^T \frac{1}{2}(\xi_t^m)'D^m\xi_t^m - \frac{\gamma^2}{2}|w_t|^2 dt + \phi(\xi_T^m) \right] \quad (9)$$

where ξ^m satisfies (4). By [37], the domain of S_T^m includes \mathcal{G}_δ for all $\delta > 0$. It has also been shown that V^m is the unique solution in \mathcal{G}_δ of $V = S_T^m[V]$ for all $T > 0$ if $\delta > 0$ is sufficiently small, and that $V^{m,f}(x, t + T) = S_T^m[V^{m,f}(\cdot, t)](x)$.

Recall that the HJB PDE problem of interest is

$$0 = -\widetilde{H}(x, \nabla V) \doteq -\max_{m \in \mathcal{M}} H^m(x, \nabla V), \quad V(0) = 0. \quad (10)$$

Below, we show that the corresponding value function is

$$\widetilde{V}(x) = \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \widetilde{J}(x, w, \mu) \doteq \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_\infty} \sup_{T < \infty} \int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2 dt \quad (11)$$

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

$$\dot{\xi} = A^{\mu_t}\xi + \sigma^{\mu_t}w_t, \quad \xi_0 = x. \quad (12)$$

The computational complexity which will be discussed fully in Section 4, arises from the switching control μ in (11).

Define the semigroup

$$\widetilde{S}_T[\phi] = \sup_{w \in \mathcal{W}} \sup_{\mu \in \mathcal{D}_T} \left[\int_0^T l^{\mu_t}(\xi_t) - \frac{\gamma^2}{2}|w_t|^2 dt + \phi(\xi_T) \right] \quad (13)$$

where $\mathcal{D}_T = \{\mu : [0, T) \rightarrow \mathcal{M} : \text{measurable}\}$. One has the following.

Theorem 3.1 *Value function \tilde{V} is the unique viscosity solution of (10) in the class \mathcal{G}_δ for sufficiently small $\delta > 0$. Fix any $T > 0$. Value function \tilde{V} is also the unique continuous solution of $V = \tilde{S}_T[V]$ in the class \mathcal{G}_δ for sufficiently small $\delta > 0$. Further, given any $V \in \mathcal{G}_\delta$, $\lim_{T \rightarrow \infty} \tilde{S}_T[V](x) = \tilde{V}(x)$ for all $x \in \mathbb{R}^n$ (uniformly on compact sets).*

We remind the reader that the proofs of the results in this section may be found in [29], [30].

Let \mathcal{D}^n be the set of $n \times n$ symmetric, positive or negative definite matrices. We say ϕ is uniformly semiconvex with (symmetric, definite matrix) constant $\beta \in \mathcal{D}^n$ if $\phi(x) + \frac{1}{2}x'\beta x$ is convex over \mathbb{R}^n . Let $\mathcal{S}_\beta = \mathcal{S}_\beta(\mathbb{R}^n)$ be the set of functions mapping \mathbb{R}^n into \mathbb{R}^- which are uniformly semiconvex with (symmetric, definite matrix) constant β . Also note that \mathcal{S}_β is a max-plus vector space [20], [29]. (Note that a max-plus vector space is an example from the set of abstract idempotent spaces, which are also known as idempotent semimodules or as moduloids, c.f., [7], [26].) We have the following.

Theorem 3.2 *There exists $\bar{\beta} \in \mathcal{D}^n$ such that given any β such that $\beta - \bar{\beta} > 0$ (i.e., $\beta - \bar{\beta}$ positive definite), $\tilde{V} \in \mathcal{S}_\beta$ and $V^m \in \mathcal{S}_\beta$ for all $m \in \mathcal{M}$. Further, one may take β negative definite (i.e., \tilde{V}, V^m convex).*

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

3.2 Semiconvex transforms

Recall that the max-plus algebra is the commutative semifield over \mathbb{R}^- given by $a \oplus b \doteq \max\{a, b\}$ and $a \otimes b \doteq a + b$; see [3], [24], [29] for more details. Throughout the work, we will employ certain transform kernel functions, $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ which take the form

$$\psi(x, z) = \frac{1}{2}(x - z)'C(x - z)$$

with nonsingular, symmetric C satisfying $C + \beta < 0$ (i.e., $C + \beta$ negative definite). The following semiconvex duality result [20], [29], [35] requires only a small modification of convex duality and Legendre/Fenchel transform results; see Section 3 of [38], and more generally, [39].

Theorem 3.3 *Let $\phi \in \mathcal{S}_\beta$. Let C and ψ be as above. Then, for all $x \in \mathbb{R}^n$,*

$$\phi(x) = \max_{z \in \mathbb{R}^n} [\psi(x, z) + a(z)] \tag{14}$$

$$\doteq \int_{\mathbb{R}^n}^{\oplus} \psi(x, z) \otimes a(z) dz \doteq \psi(x, \cdot) \odot a(\cdot) \tag{15}$$

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

$$a(z) = - \max_{x \in \mathbb{R}^n} [\psi(x, z) - \phi(x)] \quad (16)$$

$$= - \int_{\mathbb{R}^n}^{\oplus} \psi(x, z) \otimes [-\phi(x)] dx = - \{ \psi(\cdot, z) \odot [-\phi(\cdot)] \}, \quad (17)$$

which using the notation of [7]

$$= \{ \psi(\cdot, z) \odot [\phi^-(\cdot)] \}^-. \quad (18)$$

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

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

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. This is the subject of the next theorem.

Theorem 3.4 *We may choose $C \in \mathcal{D}^n$ such that $\tilde{V}, V^m \in \mathcal{S}_{-C}$. Further, there exists $\bar{\tau} > 0$ and $\eta > 0$ such that*

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

Henceforth, we suppose $C, \bar{\tau}, \eta$ chosen so that the results of Theorem 3.4 hold.

By Theorem 3.3

$$\tilde{S}_\tau[\psi(\cdot, z)](x) = \int_{\mathbb{R}^n}^{\oplus} \psi(x, y) \otimes \tilde{\mathcal{B}}_\tau(y, z) dy = \psi(x, \cdot) \odot \tilde{\mathcal{B}}_\tau(\cdot, z) \quad (19)$$

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

$$\tilde{\mathcal{B}}_\tau(y, z) = - \int_{\mathbb{R}^n}^{\oplus} \psi(x, y) \otimes \{ -\tilde{S}_\tau[\psi(\cdot, z)](x) \} dx = \{ \psi(\cdot, y) \odot [\tilde{S}_\tau[\psi(\cdot, z)](\cdot)]^- \}^-. \quad (20)$$

It is handy to define the max-plus linear operator with “kernel” $\tilde{\mathcal{B}}_\tau$ as $\tilde{\mathcal{B}}_\tau[\alpha](z) \doteq \tilde{\mathcal{B}}_\tau(z, \cdot) \odot \alpha(\cdot)$ for all $\alpha \in \mathcal{S}_{-C}^-$.

Proposition 3.5 *Let $\phi \in \mathcal{S}_\beta$ with semiconvex dual denoted by a . Define $\phi^1 = \tilde{S}_\tau[\phi]$. Then $\phi^1 \in \mathcal{S}_{\beta-\eta I\tau}$, and*

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

where

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

Theorem 3.6 *Let $V \in \mathcal{S}_\beta$, let a be its semiconvex dual (with respect to ψ), and suppose $\tilde{\mathcal{B}}_\tau(z, \cdot) \odot a(\cdot) = \widehat{\tilde{\mathcal{B}}}_\tau[a](z) \in \mathcal{S}_d^-$ for some d such that $C + d < 0$. Then*

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

if and only if

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

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

$$S_\tau^m[\psi(\cdot, z)](x) = \psi(x, \cdot) \odot \mathcal{B}_\tau^m(\cdot, z) \quad \forall x \in \mathbb{R}^n$$

where

$$\mathcal{B}_\tau^m(y, z) = \left\{ \psi(\cdot, y) \odot [S_\tau^m[\psi(\cdot, z)]]^-(\cdot) \right\}^- \quad \forall y \in \mathbb{R}^n. \quad (21)$$

It will be handy to define the max-plus linear operator with “kernel” \mathcal{B}_τ^m as $\widehat{\mathcal{B}}_\tau^m[a](z) \doteq \mathcal{B}_\tau^m(z, \cdot) \odot a(\cdot)$ for all $a \in \mathcal{S}_{-C}^-$. Further, one also obtains analogous results as those for \tilde{V} above.

3.3 Discrete-time approximation

We now discretize over time, and employ approximate μ processes which will be constant over the length of each time-step. We define the operator \bar{S}_τ on \mathcal{G}_δ by

$$\bar{S}_\tau[\phi](x) = \sup_{w \in \mathcal{W}} \max_{m \in \mathcal{M}} \left[\int_0^\tau l^m(\xi_t^m) - \frac{\gamma^2}{2} |w_t|^2 dt + \phi(\xi_\tau^m) \right] (x) = \max_{m \in \mathcal{M}} S_\tau^m[\phi](x) \quad (22)$$

where ξ^m satisfies (4). Let

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

The corresponding max-plus linear operator is $\widehat{\bar{\mathcal{B}}}_\tau = \bigoplus_{m \in \mathcal{M}} \widehat{\mathcal{B}}_\tau^m$.

Lemma 3.7 *For all $z \in \mathbb{R}^n$, $\bar{S}_\tau[\psi(\cdot, z)] \in \mathcal{S}_{-(C+\eta I\tau)}$. Further,*

$$\bar{S}_\tau[\psi(\cdot, z)](x) = \psi(x, \cdot) \odot \bar{\mathcal{B}}_\tau(\cdot, z) \quad \forall x \in \mathbb{R}^n. \quad (23)$$

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

$$\mathcal{D}_\infty^\tau = \left\{ \mu : [0, \infty) \rightarrow \mathcal{M} \mid \text{for each } n \in \mathbb{N} \cup \{0\}, \text{ there exists } m_n \in \mathcal{M} \text{ such that } \mu(t) = m_n \forall t \in [n\tau, (n+1)\tau) \right\}, \quad (24)$$

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

$$\bar{S}_T^\tau[\phi](x) = \max_{\{m_k\}_{k=0}^{\bar{n}-1} \in \mathcal{M}^{\bar{n}}} \left\{ \prod_{k=0}^{\bar{n}-1} S_\tau^{m_k} \right\} [\phi](x) = (\bar{S}_\tau)_{\bar{n}}[\phi](0) \quad (25)$$

where the \prod notation indicates operator composition with the ordering being $\prod_{k=0}^{\bar{n}-1} S_\tau^{m_k} = S_\tau^{m_{\bar{n}-1}} S_\tau^{m_{\bar{n}-2}} \dots S_\tau^{m_1} S_\tau^{m_0}$, and the superscript in the last expression indicates repeated application of \bar{S}_τ , \bar{n} times.

We will be approximating \tilde{V} by solving $V = \bar{S}_\tau[V]$ via its dual problem $a = \bar{\mathcal{B}}_\tau \odot a$ for small τ . In [29], [30], it is shown that there exists a solution to $V = \bar{S}_\tau[V]$, and that the solution is unique. In particular, for the existence part, we have:

Theorem 3.8 *Let*

$$\bar{V}(x) \doteq \lim_{N \rightarrow \infty} \bar{S}_{N\tau}^\tau[0](x) \quad (26)$$

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

$$V = \bar{S}_\tau[V], \quad V(0) = 0. \quad (27)$$

Further,

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

where ξ_t satisfies (12), and $0 \leq V^m \leq \bar{V} \leq \tilde{V}$ for all $m \in \mathcal{M}$ (which implies $\bar{V} \in \mathcal{G}_\delta$). Lastly, with the choice of β above (i.e., such that $C + \beta < 0$), one has $\bar{V} \in \mathcal{S}_\beta \subset \mathcal{S}_{-C}$.

Similar techniques to those used for V^m and \tilde{V} prove uniqueness for (27) within \mathcal{G}_δ . In particular, we have the following results [29], [30].

Theorem 3.9 \bar{V} is the unique solution of (27) within the class \mathcal{G}_δ for sufficiently small $\delta > 0$. Further, given any $V \in \mathcal{G}_\delta$, $\lim_{N \rightarrow \infty} \bar{S}_{N\tau}^\tau[V](x) = \bar{V}(x)$ for all $x \in \mathbb{R}^n$ (uniformly on compact sets).

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

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

where

$$a^1(y) = \bar{\mathcal{B}}_\tau(y, \cdot) \odot a(\cdot) \quad \forall y \in \mathbb{R}^n.$$

We will solve the problem $V = \bar{S}_\tau[V]$, or equivalently $0 = -\tilde{H}(x, \nabla V)$, with boundary condition $V(0) = 0$, by computing $\bar{a}^k = [\bar{\mathcal{B}}_\tau]^k \odot \bar{a}^0$ with appropriate initial \bar{a}^0 (such as $\bar{a}^0 = 0$), where the k superscript of the right-hand side represents operator composition k times. From this, one obtains approximation $\bar{V}^k(x) \doteq \psi(x, \cdot) \odot \bar{a}^k$. We will show that for k sufficiently large and τ sufficiently small, \bar{V}^k will approximate \tilde{V} within an error bound of $\varepsilon(1 + |x|^2)$ for as small an ε as desired.

4 The algorithm

We summarize the mechanics of the algorithm. The full development may be found in [29], [30]. We start by computing $P_\tau^m, \Lambda_\tau^m, R_\tau^m$ for each $m \in \mathcal{M}$ where these are defined by

$$S_\tau^m[\psi(\cdot, z)](x) = \frac{1}{2}(x - \Lambda_\tau^m z)' P_\tau^m (x - \Lambda_\tau^m z) + \frac{1}{2} z' R_\tau^m z,$$

and where the time-dependent $n \times n$ matrices P_t^m, Λ_t^m and R_t^m satisfy $P_0^m = C, \Lambda_0^m = I, R_0^m = 0$,

$$\begin{aligned} \dot{P}^m &= (A^m)' P^m + P^m A^m - [D^m + P^m \Sigma^m P^m] \\ \dot{\Lambda}^m &= [(P^m)^{-1} D^m - A^m] \Lambda^m \\ \dot{R}^m &= (\Lambda^m)' D^m \Lambda^m. \end{aligned} \tag{28}$$

The $P_\tau^m, \Lambda_\tau^m, R_\tau^m$ may be computed from these ordinary differential equations via a Runge-Kutta algorithm (or other technique) with initial time $t = 0$ and terminal time $t = \tau$. We remark that each of the $P_\tau^m, \Lambda_\tau^m, R_\tau^m$ *need only be computed once*.

Next, noting that each \mathcal{B}_τ^m is given by (21), one has

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

where with shorthand notation $D_\tau \doteq (P_\tau^m - C)$,

$$M_{1,1}^m = -C D_\tau^{-1} P_\tau^m \tag{29}$$

$$M_{1,2}^m = -C D_\tau^{-1} P_\tau^m \Lambda_\tau^m \tag{30}$$

$$M_{2,2}^m = R_\tau^m - (\Lambda_\tau^m)' C D_\tau^{-1} P_\tau^m \Lambda_\tau^m. \tag{31}$$

Note that each of the $M_{1,1}^m, M_{1,2}^m, M_{2,2}^m$ *need only be computed once*.

In order to reduce notational complexity, for the moment we suppose that we initialize the following iteration with an \bar{a}^0 (the dual of \bar{V}^0) which consists of a single quadratic, that is $\bar{a}^0(x) = \hat{a}^0(x)$ where \hat{a}^0 takes the form $\hat{a}^0(x) = \frac{1}{2}(x - \hat{z}^0)' \hat{Q}^0 (x - \hat{z}^0) + \hat{r}^0$. Next we note that, recalling $\bar{a}^k = \bar{\mathcal{B}}_\tau \odot \{[\bar{\mathcal{B}}_\tau]^{k-1} \odot \bar{a}^0\}$ with $\bar{\mathcal{B}}_\tau = \bigoplus_{m \in \mathcal{M}} \mathcal{B}_\tau^m$,

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

where each

$$\begin{aligned} \hat{a}_{\{m_i\}_{i=1}^k}^k(x) &\doteq \mathcal{B}_\tau^{m_k}(x, \cdot) \odot \hat{a}_{\{m_i\}_{i=1}^{k-1}}^{k-1}(\cdot), \\ &= \frac{1}{2}(x - \hat{z}_{\{m_i\}_{i=1}^k}^k)' \hat{Q}_{\{m_i\}_{i=1}^k}^k (x - \hat{z}_{\{m_i\}_{i=1}^k}^k) + \hat{r}_{\{m_i\}_{i=1}^k}^k, \end{aligned}$$

where for each m_{k+1} and each $\{m_i\}_{i=1}^k$,

$$\hat{Q}_{\{m_i\}_{i=1}^{k+1}}^{k+1} = M_{1,1}^{m_{k+1}} - M_{1,2}^{m_{k+1}} \widehat{D} (M_{1,2}^{m_{k+1}})' \tag{32}$$

$$\begin{aligned}
\widehat{z}_{\{m_i\}_{i=1}^{k+1}}^{k+1} &= - \left(\widehat{Q}_{\{m_i\}_{i=1}^{k+1}}^{k+1} \right)^{-1} M_{1,2}^{m_{k+1}} \widehat{E} \\
\widehat{r}_{\{m_i\}_{i=1}^{k+1}}^{k+1} &= \widehat{r}_{\{m_i\}_{i=1}^k}^k + \frac{1}{2} \widehat{E}' M_{2,2}^{m_{k+1}} \widehat{z}_{\{m_i\}_{i=1}^k}^k - \frac{1}{2} \left(\widehat{z}_{\{m_i\}_{i=1}^{k+1}}^{k+1} \right)' \widehat{Q}_{\{m_i\}_{i=1}^{k+1}}^{k+1} \widehat{z}_{\{m_i\}_{i=1}^{k+1}}^{k+1} \\
\widehat{D} &= \left(M_{2,2}^{m_{k+1}} + \widehat{Q}_{\{m_i\}_{i=1}^k}^k \right)^{-1} \\
\widehat{E} &= \widehat{D} \widehat{Q}_{\{m_i\}_{i=1}^k}^k \widehat{z}_{\{m_i\}_{i=1}^k}^k.
\end{aligned}$$

We see that the propagation of each $\widehat{a}_{\{m_i\}_{i=1}^k}^k$ amounts to a set of matrix multiplications and an inverse. (Note in the case of purely quadratic constituent Hamiltonians here, one will have all $\widehat{z}_{\{m_i\}_{i=1}^k}^k = 0$ and all $\widehat{r}_{\{m_i\}_{i=1}^k}^k = 0$.) Thus, at each step, k , the semiconvex dual of \overline{V}^k , \overline{a}^k , is represented as the finite set of triples

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

(We remark that in actual implementation, it is simpler to index the entries of each \widehat{Q}_k with integers $i \in \{1, 2, \dots, \#\widehat{Q}_k\}$ rather than the sequences $\{m_i\}_{i=1}^k$.)

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

$$\overline{V}^k(x) \doteq \max_{\{m_i\}_{i=1}^k} \frac{1}{2} (x - \widehat{x}_{\{m_i\}_{i=1}^k}^k)' \widehat{P}_{\{m_i\}_{i=1}^k}^k (x - \widehat{x}_{\{m_i\}_{i=1}^k}^k) + \widehat{\rho}_{\{m_i\}_{i=1}^k}^k$$

where

$$\begin{aligned}
\widehat{P}_{\{m_i\}_{i=1}^k}^k &= C \widehat{F} \widehat{Q}_{\{m_i\}_{i=1}^k}^k \widehat{F} C + (\widehat{F} C - I)' C (\widehat{F} C - I) \\
\widehat{x}_{\{m_i\}_{i=1}^k}^k &= - (\widehat{P}_{\{m_i\}_{i=1}^k}^k)^{-1} \left[C \widehat{F} \widehat{Q}_{\{m_i\}_{i=1}^k}^k \widehat{G} + (\widehat{F} C - I)' C \widehat{F} \widehat{Q}_{\{m_i\}_{i=1}^k}^k \right] \widehat{z}_{\{m_i\}_{i=1}^k}^k \\
\widehat{\rho}_{\{m_i\}_{i=1}^k}^k &= \widehat{r}_{\{m_i\}_{i=1}^k}^k + \frac{1}{2} (\widehat{z}_{\{m_i\}_{i=1}^k}^k)' \left[\widehat{G}' \widehat{Q}_{\{m_i\}_{i=1}^k}^k \widehat{G} + \widehat{Q}_{\{m_i\}_{i=1}^k}^k \widehat{F} C \widehat{F} \widehat{Q}_{\{m_i\}_{i=1}^k}^k \right] \widehat{z}_{\{m_i\}_{i=1}^k}^k \\
\widehat{F} &\doteq \left(\widehat{Q}_{\{m_i\}_{i=1}^k}^k + C \right)^{-1}
\end{aligned} \tag{33}$$

and

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

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

$$\mathcal{P}_k \doteq \left\{ \left(\widehat{P}_{\{m_i\}_{i=1}^k}^k, \widehat{x}_{\{m_i\}_{i=1}^k}^k, \widehat{\rho}_{\{m_i\}_{i=1}^k}^k \right) \mid m_i \in \mathcal{M} \ \forall i \in \{1, 2, \dots, k\} \right\}. \tag{34}$$

We note that the triples which comprise \mathcal{P}_k are analytically obtained from the triples $(\widehat{Q}_{\{m_i\}_{i=1}^k}^k, \widehat{z}_{\{m_i\}_{i=1}^k}^k, \widehat{r}_{\{m_i\}_{i=1}^k}^k)$ by matrix multiplications and an inverse. The transference from $(\widehat{Q}_{\{m_i\}_{i=1}^k}^k, \widehat{z}_{\{m_i\}_{i=1}^k}^k, \widehat{r}_{\{m_i\}_{i=1}^k}^k)$ to $(\widehat{P}_{\{m_i\}_{i=1}^k}^k, \widehat{x}_{\{m_i\}_{i=1}^k}^k, \widehat{\rho}_{\{m_i\}_{i=1}^k}^k)$ need only be done once which is at the termination of the algorithm propagation. We note that (34) is our approximate solution of the original control problem/HJB PDE.

The curse-of-dimensionality is replaced by another type of rapid computational cost growth. We refer to this as the curse-of-complexity. If $\#\mathcal{M} = 1$, then all the computations

for our algorithm (excepting the solution of a Riccati equation) are unnecessary, and we *informally* refer to this as complexity one. When there are $M = \#\mathcal{M}$ such quadratics in the Hamiltonian, \widetilde{H} , we say it has complexity M . Note that

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

For large k , this is indeed a large number. In order for the computations to be practical, one must reduce this by pruning and other techniques; see Section 9.

Lastly, note that in the above, we assumed \bar{a}^0 to consist of a single quadratic \hat{a}^0 . In general, we take $\bar{a}^0 = \bigoplus_{j \in \mathcal{J}_0} \hat{a}_j^0(x)$ with $\mathcal{J}_0 \doteq \{1, 2, \dots, J_0\}$, where each

$$\hat{a}_j^0(x) = \frac{1}{2}(x - \hat{z}_j^0)' \widehat{Q}_j^0(x - \hat{z}_j^0) + \hat{r}_j^0.$$

This increases the size of each \widehat{Q}_k by a factor of J_0 . Denote the elements of $\mathcal{M}^k \doteq \{\{m_i\}_{i=1}^k \mid m_i \in \mathcal{M} \forall i\}$ by $m^k \in \mathcal{M}^k$. With the additional quadratics in \bar{a}^0 and the reduced-complexity indexes, one sees that at each step, we would have \widehat{Q}_k in the form

$$\widehat{Q}_k = \left\{ \left(\widehat{Q}_{(j, m^k)}^k, \hat{z}_{(j, m^k)}^k, \hat{r}_{(j, m^k)}^k \right) \mid j \in \mathcal{J}_0, m^k \in \mathcal{M}^k \right\}. \quad (35)$$

5 Error bounds and convergence

There are two error sources with this curse-of-dimensionality-free approach to solution of the HJB PDE.

For concreteness, we suppose that the algorithm is initialized with $\bar{V}^0 = \bigoplus_{m \in \mathcal{M}} V^m$. Letting $T \doteq N\tau$ where N is the number of iterations, the first error source is

$$\varepsilon_{FT}(x, N, \tau) = \widetilde{S}_T \left[\bar{V}^0 \right] (x) - \bar{S}_T^\tau \left[\bar{V}^0 \right] (x). \quad (36)$$

This is the error due to the time-discretization of the μ process. The second error source is

$$\varepsilon_{IT}(x, N, \tau) = \widetilde{V}(x) - \widetilde{S}_T \left[\bar{V}^0 \right] (x). \quad (37)$$

This is the error due to approximating the infinite time-horizon problem by the finite-time horizon problem with horizon T . The total error is obviously $\varepsilon_{TE} = \varepsilon_{FT}(x, N, \tau) + \varepsilon_{IT}(x, N, \tau)$. We begin the error analysis with the former of these two error sources in Section 6. In Section 7, we consider the latter, and in Section 8, the two are combined.

6 Errors from time-discretization

We henceforth assume that all the σ^m used in the dynamics (4) are the same, i.e., $\sigma^m = \sigma$ for all $m \in \mathcal{M}$. The authors do not know if this is required, but were unable to obtain the (already technically difficult) proofs of the estimates in this section without this assumption. The proof of the following is quite technical. However, we assure the reader that the proof of the time-horizon error estimate in Section 7 is much less tedious.

Theorem 6.1 *There exists $K_\delta < \infty$ such that, for all sufficiently small $\tau > 0$,*

$$0 \leq \tilde{S}_T[V^m](x) - \bar{S}_T^\tau[V^m](x) \leq K_\delta(1 + |x|^2)(\tau + \sqrt{\tau})$$

for all $T \in (0, \infty)$, $x \in \mathbb{R}^n$, and $m \in \mathcal{M}$.

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

$$\tilde{S}_T[V^m](x) - \left[\int_0^T l^{\mu^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + V^m(\xi_T^\varepsilon) \right] \leq \varepsilon = \frac{\hat{\varepsilon}}{2}(1 + |x|^2) \quad (38)$$

where ξ^ε satisfies (12) with inputs w^ε , μ^ε .

We will let $\bar{\xi}^\varepsilon$ satisfy (12) with inputs w^ε and a $\bar{\mu}^\varepsilon \in \mathcal{D}_\infty^\tau$ (where τ has yet to be chosen). Integrating (12), one has

$$\xi_t^\varepsilon = \int_0^t \left[A^{\mu_r^\varepsilon} \xi_r^\varepsilon + \sigma w_r^\varepsilon \right] dr + x, \quad \text{and} \quad \bar{\xi}_t^\varepsilon = \int_0^t \left[A^{\bar{\mu}_r^\varepsilon} \bar{\xi}_r^\varepsilon + \sigma w_r^\varepsilon \right] dr + x. \quad (39)$$

Differencing these, one has

$$\begin{aligned} \xi_t^\varepsilon - \bar{\xi}_t^\varepsilon &= \int_0^t A^{\mu_r^\varepsilon} \xi_r^\varepsilon - A^{\bar{\mu}_r^\varepsilon} \bar{\xi}_r^\varepsilon dr \\ &= \int_0^t (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon dr + \int_0^t A^{\bar{\mu}_r^\varepsilon} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon) dr. \end{aligned} \quad (40)$$

Letting $z_t \doteq \frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2$, and using (40), yields

$$\dot{z} = (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon + (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^{\bar{\mu}_r^\varepsilon} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)$$

which, by using Assumption Block $(A.m)$,

$$\leq (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon - 2c_A z_r.$$

Noting that $z_0 = 0$, and solving this differential inequality, one obtains

$$\frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 \leq \int_0^t e^{-2c_A(t-r)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon dr. \quad (41)$$

Consequently, we now seek a bound on the right-hand side of (41), which will go to zero as $\tau \downarrow 0$ independent of $0 \leq t \leq T < \infty$. We will use the boundedness of $\|\xi^\varepsilon\|$, $\|\bar{\xi}^\varepsilon\|$, and $\|w^\varepsilon\|$, which is independent of T for this class of systems [37]. (Precise statements are given in Lemmas 6.6 and 6.7 below.)

For any given $\tau > 0$, we build $\bar{\mu}_r^\varepsilon$ from μ_r^ε over $[0, T]$ in the following manner. Fix $\tau > 0$. Let N^t be the largest integer such that $N^t \tau \leq t$. For any Lebesgue measurable subset of \mathbb{R} , \mathcal{I} , let $\mathcal{L}(\mathcal{I})$ be the measure of \mathcal{I} . For $t \in [0, T]$, $m \in \mathcal{M}$, let

$$\begin{aligned}\mathcal{I}_t^m &= \{r \in [0, t) \mid \mu_r^\varepsilon = m\}, & \bar{\mathcal{I}}_t^m &= \{r \in [0, t) \mid \bar{\mu}_r^\varepsilon = m\} \\ \lambda_t^m &= \mathcal{L}(\mathcal{I}_t^m), & \bar{\lambda}_t^m &= \mathcal{L}(\bar{\mathcal{I}}_t^m).\end{aligned}\tag{42}$$

At the end of any time step $n\tau$ (with $n \leq N^T$), we pick one of the $m \in \mathcal{M}$ with the largest (positive) error so far committed and we correct it, i.e., let

$$\bar{m} \in \operatorname{argmax}_{m \in \mathcal{M}} \{\lambda_{n\tau}^m - \bar{\lambda}_{(n-1)\tau}^m\},\tag{43}$$

and we set

$$\bar{\mu}_r^\varepsilon = \bar{m} \quad \forall r \in [(n-1)\tau, n\tau).\tag{44}$$

Finally, we simply set

$$\bar{\mu}_r^\varepsilon \in \operatorname{argmax}_{m \in \mathcal{M}} \{\lambda_r^m - \bar{\lambda}_{N^T\tau}^m\} \quad \forall r \in [N^T\tau, T].\tag{45}$$

Obviously, for all $0 \leq r \leq T$ the sum of the errors in measure is null, that is,

$$\sum_{m \in \mathcal{M}} (\lambda_r^m - \bar{\lambda}_r^m) = \sum_{m \in \mathcal{M}} \lambda_r^m - \sum_{m \in \mathcal{M}} \bar{\lambda}_r^m = r - r = 0.$$

With this construction we also get:

Lemma 6.2 *For any $t \in [0, T]$, and any $m \in \mathcal{M}$, one has $\lambda_t^m - \bar{\lambda}_t^m \geq -\tau$.*

PROOF. Let us assume that at time $n\tau$, it is true (which is certainly true for $n = 0$). We first consider the case $(n+1)\tau \leq T$. Since $\sum_{m \in \mathcal{M}} \lambda_{n\tau}^m - \bar{\lambda}_{n\tau}^m = 0$, there exists \hat{m} such that $\lambda_{n\tau}^{\hat{m}} - \bar{\lambda}_{n\tau}^{\hat{m}} \geq 0$. Hence $\max_{m \in \mathcal{M}} (\lambda_{(n+1)\tau}^m - \bar{\lambda}_{n\tau}^m) \geq \max_{m \in \mathcal{M}} (\lambda_{n\tau}^m - \bar{\lambda}_{n\tau}^m) \geq 0$. Now, choosing \bar{m} by (43) (for time step $(n+1)\tau$), we have

$$\lambda_{(n+1)\tau}^{\bar{m}} - \bar{\lambda}_{n\tau}^{\bar{m}} \geq 0.\tag{46}$$

Now let $t \in [n\tau, (n+1)\tau]$. Then, by (46),

$$\lambda_t^{\bar{m}} - \bar{\lambda}_t^{\bar{m}} \geq -(\lambda_{(n+1)\tau}^{\bar{m}} - \lambda_t^{\bar{m}}) - (\bar{\lambda}_t^{\bar{m}} - \bar{\lambda}_{n\tau}^{\bar{m}})$$

which by (44) and the choice of \bar{m} ,

$$= -(\lambda_{(n+1)\tau}^{\bar{m}} - \lambda_t^{\bar{m}}) - (t - n\tau),$$

and since $\lambda_t^m - \lambda_s^m \leq (t - s)$ for all m and $t \geq s$,

$$\begin{aligned}&\geq (t - (n+1)\tau) - (t - n\tau) \\ &= -\tau.\end{aligned}\tag{47}$$

Now, since $\bar{\mu}_r^\varepsilon = \bar{m}$ for all $r \in [n\tau, (n+1)\tau)$, $\bar{\lambda}_r^m = \bar{\lambda}_{n\tau}^m$ for all $r \in [n\tau, (n+1)\tau)$ and all $m \neq \bar{m}$. Consequently, for $t \in [n\tau, (n+1)\tau]$ and $m \neq \bar{m}$,

$$\lambda_t^m - \bar{\lambda}_t^m = \lambda_t^m - \bar{\lambda}_{n\tau}^m \geq \lambda_{n\tau}^m - \bar{\lambda}_{n\tau}^m$$

which by the induction assumption,

$$\geq -\tau. \quad (48)$$

Combining (47) and (48) completes the proof for the case $(n+1)\tau \leq T$.

We must still consider the case $n\tau < T < (n+1)\tau$. The proof is essentially identical to that above with the exception that $(n+1)\tau$ is replaced by T , $\bar{m} \in \operatorname{argmax}_{m \in \mathcal{M}} \{\lambda_T^m - \bar{\lambda}_{n\tau}^m\}$ is used in place of (43), and (45) is used in place of (44). We do not repeat the details. \square

Using this, we will find:

Lemma 6.3 *For any $t \in [0, T]$, and any $m \in \mathcal{M}$, one has $\lambda_t^m - \bar{\lambda}_t^m \leq (M-1)\tau$.*

PROOF. Since $\sum_{m \in \mathcal{M}} \lambda_t^m - \bar{\lambda}_t^m = 0$, for all $\tilde{m} \in \mathcal{M}$,

$$\lambda_t^{\tilde{m}} - \bar{\lambda}_t^{\tilde{m}} = - \sum_{m \neq \tilde{m}} \lambda_t^m - \bar{\lambda}_t^m$$

which by Lemma 6.2,

$$\leq (M-1)\tau. \quad \square$$

We now develop some more-delicate machinery, which will allow us to make fine estimates of the difference between ξ_t^ε and $\bar{\xi}_t^\varepsilon$. For each m we divide \mathcal{I}_t^m into pieces $\widehat{\mathcal{I}}_k^{m,t}$ of length τ as follows. Let $\widehat{K}_t^m = \max\{k \in \mathbf{N} \cup \{0\} \mid \exists \text{ integer } n \leq t/\tau \text{ s.t. } \bar{\lambda}_{n\tau}^m = k\tau\}$. Then, for $k \leq \widehat{K}_t^m$, let $n_k^m \doteq \min\{n \in \mathbf{N} \cup \{0\} \mid \bar{\lambda}_{n\tau}^m = k\tau\}$ and $\widehat{\mathcal{I}}_k^{m,t} = [(n_k^m - 1)\tau, n_k^m\tau]$. Let $K_t^m \doteq]1, \widehat{K}_t^m[$ where for any integers $m \leq n$, $]m, n[$ denotes $\{m, m+1, \dots, n\}$. Loosely speaking, we will now let $\widehat{\mathcal{I}}_k^{m,t}$ denote a subset of \mathcal{I}_t^m of measure τ , corresponding to $\widehat{\mathcal{I}}_k^{m,t}$. More specifically, we define $\widehat{\mathcal{I}}_k^{m,t}$ as follows. Introduce the functions $\Phi_k^{m,t}(r)$ which are monotonically increasing (hence measurable) functions (that will match $\widehat{\mathcal{I}}_k^{m,t} = [(n_k^m - 1)\tau, n_k^m\tau]$ with $\widehat{\mathcal{I}}_k^{m,t}$) given by

$$\begin{aligned} \Phi_k^{m,t}(r) &= \inf \left\{ \rho \in [0, t] \mid \lambda_\rho^m = (k-1)\tau + [r - (n_k^m - 1)\tau] \right\} \\ &= \inf \left\{ \rho \in [0, t] \mid \lambda_\rho^m = r + (k - n_k^m)\tau \right\}, \end{aligned}$$

where, in particular, we take $\Phi_k^{m,t}(r) = t$ if there does not exist $\rho \in [0, t]$ such that $\lambda_\rho^m = r + (k - n_k^m)\tau$. We note that the $\Phi_k^{m,t}(r)$ are translations by part. Then, (neglecting the point $r = t$ which has measure zero anyway) $\widehat{\mathcal{I}}_k^{m,t} = \Phi_k^{m,t}(\widehat{\mathcal{I}}_k^{m,t})$.

We also define $\widehat{\mathcal{I}}_f^{m,t}$ as the last part of $\widehat{\mathcal{I}}_t^m$, with length $\mathcal{L}(\widehat{\mathcal{I}}_f^{m,t}) \leq \tau$, and $\widehat{\mathcal{I}}_f^{m,t}$ as the last part of \mathcal{I}_t^m not corresponding to an interval of length τ of $\widehat{\mathcal{I}}_t^m$. That is, $\widehat{\mathcal{I}}_f^{m,t} = \widehat{\mathcal{I}}_t^m \setminus \bigcup_{k \in K_t^m} \widehat{\mathcal{I}}_k^{m,t}$ and $\widehat{\mathcal{I}}_f^{m,t} = \mathcal{I}_t^m \setminus \bigcup_{k \in K_t^m} \widehat{\mathcal{I}}_k^{m,t}$.

With this additional machinery, we now return to obtaining the bound on $|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|$. Note that

$$\begin{aligned} & e^{-2cAt} \int_0^t e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon dr \\ &= e^{-2cAt} \sum_{m \in \mathcal{M}} \left\{ \int_{\mathcal{I}_t^m} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\bar{\mathcal{I}}_t^m} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\}. \end{aligned}$$

Combining this with (41) yields

$$\begin{aligned} \frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 &\leq e^{-2cAt} \sum_{m \in \mathcal{M}} \left\{ \int_{\mathcal{I}_t^m} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\bar{\mathcal{I}}_t^m} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\} \\ &= e^{-2cAt} \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_t^m} \left[\int_{\widehat{\mathcal{I}}_k^{m,t}} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right. \right. \\ &\quad \left. \left. - \int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right] \right. \\ &\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\bar{\mathcal{I}}_f^{m,t}} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\}, \end{aligned}$$

which by the definition of $\Phi_k^{m,t}$,

$$\begin{aligned} &= e^{-2cAt} \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_t^m} \left[\int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} e^{2cA\Phi_k^{m,t}(r)} (\xi_{\Phi_k^{m,t}(r)}^\varepsilon - \bar{\xi}_{\Phi_k^{m,t}(r)}^\varepsilon)' A^m \xi_{\Phi_k^{m,t}(r)}^\varepsilon \right. \right. \\ &\quad \left. \left. - e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right] \right. \\ &\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\bar{\mathcal{I}}_f^{m,t}} e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\}. \end{aligned} \tag{49}$$

Note that

$$\begin{aligned} & e^{2cA\Phi_k^{m,t}(r)} (\xi_{\Phi_k^{m,t}(r)}^\varepsilon - \bar{\xi}_{\Phi_k^{m,t}(r)}^\varepsilon)' A^m \xi_{\Phi_k^{m,t}(r)}^\varepsilon - e^{2cAr} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon \\ &= \int_r^{\Phi_k^{m,t}(r)} \left[2cA e^{2cA\rho} (\xi_\rho^\varepsilon - \bar{\xi}_\rho^\varepsilon)' A^m \xi_\rho^\varepsilon + e^{2cA\rho} (\xi_\rho^\varepsilon - \bar{\xi}_\rho^\varepsilon)' A^m (A^{\mu_\rho^\varepsilon} \xi_\rho^\varepsilon + \sigma w_\rho^\varepsilon) \right. \\ &\quad \left. + e^{2cA\rho} (A^{\mu_\rho^\varepsilon} \xi_\rho^\varepsilon - A^{\bar{\mu}_\rho^\varepsilon} \bar{\xi}_\rho^\varepsilon)' A^m \xi_\rho^\varepsilon \right] d\rho \\ &\leq \int_r^{\Phi_k^{m,t}(r)} e^{2cAt} K_1 \left(|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2 \right) d\rho, \end{aligned} \tag{50}$$

for proper choice of K_1 independent of r, t, x, m, k and $\varepsilon > 0$. Substituting (50) into (49) yields

$$\frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 \leq \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_t^m} \left[\int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_r^{\Phi_k^{m,t}(r)} K_1 \left(|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2 \right) d\rho \right] \right\}$$

$$+ \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \Big\}. \quad (51)$$

Define

$$\Phi_k^{m,t,+}(r) = \begin{cases} \Phi_k^{m,t}(r) & \text{if } \Phi_k^{m,t}(r) \geq r, \\ r & \text{otherwise,} \end{cases}, \quad \Phi_k^{m,t,-}(r) = \begin{cases} \Phi_k^{m,t}(r) & \text{if } \Phi_k^{m,t}(r) \leq r, \\ r & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} & \int_r^{\Phi_k^{m,t}(r)} K_1 \left(|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2 \right) d\rho \\ &= \int_r^{\Phi_k^{m,t,+}(r)} K_1 \left(|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2 \right) d\rho + \int_{\Phi_k^{m,t,-}(r)}^r K_1 \left(|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2 \right) d\rho \end{aligned} \quad (52)$$

where at most one of the integrals on the right is nonzero. We need to evaluate the distance $|\Phi_k^{m,t}(r) - r|$ for r in $[(n_k^{m,t} - 1)\tau, n_k^{m,t}\tau]$.

Lemma 6.4 *For all $m \in \mathcal{M}$, $\Phi_k^{m,t,+}(r) < n_{k+2}^{m,t}\tau + (r - n_k^{m,t}\tau)$ for all $r \in [(n_k^{m,t} - 1)\tau, n_k^{m,t}\tau]$ and $k \in]1, \widehat{K}_t^m - 2[$, and $\Phi_k^{m,t,-}(r) > n_{k-M}^{m,t}\tau + (r - n_k^{m,t}\tau)$ for all $r \in [(n_k^{m,t} - 1)\tau, n_k^{m,t}\tau]$ and $k \in]M + 1, \widehat{K}_t^m[$.*

PROOF. We prove each of the two inequalities separately. Suppose there exist $m \in \mathcal{M}$, $r \in [(n_k^{m,t} - 1)\tau, n_k^{m,t}\tau]$ and $k \in]1, \widehat{K}_t^m[\subseteq K_t^m$ such that

$$\Phi_k^{m,t,+}(r) \geq n_{k+2}^{m,t}\tau + (r - n_k^{m,t}\tau). \quad (53)$$

By definition,

$$k\tau + r - n_k^{m,t}\tau = \lambda_{\Phi_k^{m,t}(r)}^m$$

which by Lemma 6.2,

$$\begin{aligned} & \geq \bar{\lambda}_{\Phi_k^{m,t}(r)}^m - \tau \\ & = \bar{\lambda}_{\Phi_k^{m,t}(r)}^m - \bar{\lambda}_{n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau}^m + \bar{\lambda}_{n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau}^m - \tau \end{aligned}$$

which by (53) and the monotonicity of $\bar{\lambda}^m$,

$$\geq \bar{\lambda}_{n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau}^m - \tau. \quad (54)$$

However, note that

$$\bar{\lambda}_\rho^m = j\tau + \rho - n_j^{m,t}\tau \quad \text{for } \rho \in [(n_j^{m,t} - 1)\tau, n_j^{m,t}\tau], \quad (55)$$

and that $n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau \in [(n_{k+2}^{m,t} - 1)\tau, n_{k+2}^{m,t}\tau]$, and so

$$\bar{\lambda}_{n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau}^m = (k+2)\tau + n_{k+2}^{m,t}\tau + r - n_k^{m,t}\tau - n_{k+2}^{m,t}\tau = (k+2)\tau + r - n_k^{m,t}\tau.$$

Consequently, (54) yields

$$k\tau + r - n_k^{m,t}\tau \geq (k+2)\tau + r - n_k^{m,t}\tau - \tau > k\tau + r - n_k^{m,t}\tau,$$

which is a contradiction.

Now we turn to the second inequality. Suppose there exist $m \in \mathcal{M}$, $r \in [(n_k^{m,t} - 1)\tau, n_k^{m,t}\tau]$ and $k \in]M+1, \widehat{K}_t^m[\subset K_t^m$ such that

$$\Phi_k^{m,t,-}(r) \leq n_{k-M}^{m,t}\tau + (r - n_k^{m,t}\tau). \quad (56)$$

Again, by definition,

$$k\tau + r - n_k^{m,t}\tau = \lambda_{\Phi_k^{m,t}(r)}^m$$

which by Lemma 6.3,

$$\leq \bar{\lambda}_{\Phi_k^{m,t}(r)}^m + (M-1)\tau,$$

which by (56) and the monotonicity of $\bar{\lambda}^m$,

$$\leq \bar{\lambda}_{n_{k-M}^{m,t}\tau + r - n_k^{m,t}\tau}^m + (M-1)\tau. \quad (57)$$

Noting that $n_{k-M}^{m,t}\tau + r - n_k^{m,t}\tau \in [(n_{k-M}^{m,t} - 1)\tau, n_{k-M}^{m,t}\tau]$, and again appealing to (55), one sees that $\bar{\lambda}_{n_{k-M}^{m,t}\tau + r - n_k^{m,t}\tau}^m = (k-M)\tau + r - n_k^{m,t}\tau$, and so (57) implies

$$k\tau + r - n_k^{m,t}\tau \leq (k-M)\tau + r - n_k^{m,t}\tau + (M-1)\tau = (k-1)\tau + r - n_k^{m,t}\tau,$$

which is a contradiction. \square

Lemma 6.5 *Suppose $f(\cdot)$ is nonnegative and integrable over $[0, T]$. For any $t \in [0, T]$ and any $m \in \mathcal{M}$,*

$$\sum_{k \in K_t^m} \left[\int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_r^{\Phi_k^{m,t}(r)} f(\rho) d\rho dr \right] \leq (M+2)\tau \int_0^t f(r) dr.$$

PROOF. Using the first inequality of Lemma 6.4, one has

$$\begin{aligned} & \sum_{k \in K_t^m} \int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_r^{\Phi_k^{m,t,+}(r)} f(\rho) d\rho dr \\ & \leq \sum_{k \in K_t^m} \int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_r^{[(\tilde{n}_{k+2}^{m,t} - n_k^{m,t})\tau + r] \wedge t} f(\rho) d\rho dr \end{aligned}$$

where $\tilde{n}_k^{m,t} = n_k^{m,t}$ if $k \leq \widehat{K}_t^m$ and $\tilde{n}_k^{m,t} = +\infty$ otherwise, and with a change of variables, this is

$$\begin{aligned}
&= \sum_{k \in K_t^m} \int_0^\tau \int_{(n_k^{m,t}-1)\tau+\rho}^{[(\tilde{n}_{k+2}^{m,t}-1)\tau+\rho] \wedge t} f(s) ds d\rho \\
&= \sum_{k \in K_t^m} \int_0^\tau \left\{ \int_{[(\tilde{n}_{k+1}^{m,t}-1)\tau+\rho] \wedge t}^{[(\tilde{n}_{k+2}^{m,t}-1)\tau+\rho] \wedge t} f(s) ds + \int_{(n_k^{m,t}-1)\tau+\rho}^{[(\tilde{n}_{k+1}^{m,t}-1)\tau+\rho] \wedge t} f(s) ds \right\} d\rho \\
&= \int_0^\tau \left\{ \sum_{k \in K_t^m} \int_{[(\tilde{n}_{k+1}^{m,t}-1)\tau+\rho] \wedge t}^{[(\tilde{n}_{k+2}^{m,t}-1)\tau+\rho] \wedge t} f(s) ds + \sum_{k \in K_t^m} \int_{(n_k^{m,t}-1)\tau+\rho}^{[(\tilde{n}_{k+1}^{m,t}-1)\tau+\rho] \wedge t} f(s) ds \right\} d\rho \\
&\leq \int_0^\tau 2 \int_0^t f(s) ds d\rho = 2\tau \int_0^t f(s) ds.
\end{aligned} \tag{58}$$

To handle the $\Phi_k^{m,t,-}$ terms, one can employ the second inequality of Lemma 6.4. In particular, one finds

$$\begin{aligned}
&\sum_{k \in K_t^m} \int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_{\Phi_k^{m,t,-}(r)}^r f(\rho) d\rho dr \\
&\leq \sum_{k \in K_t^m} \int_{(n_k^{m,t}-1)\tau}^{n_k^{m,t}\tau} \int_{[(\hat{n}_{k-M}^{m,t}\tau+r-n_k^{m,t}\tau] \vee 0}^r f(\rho) d\rho dr
\end{aligned}$$

where $\hat{n}_k^{m,t} = n_k^{m,t}$ if $k \geq 1$ and $\hat{n}_k^{m,t} = -\infty$ otherwise, and again with a change of variables, this becomes

$$\begin{aligned}
&= \sum_{k \in K_t^m} \int_0^\tau \int_{[(\hat{n}_{k-M}^{m,t}-1)\tau+\rho] \vee 0}^{(n_k^{m,t}-1)\tau+\rho} f(s) ds d\rho \\
&= \sum_{k \in K_t^m} \int_0^\tau \sum_{m'=-M}^{-1} \int_{[(\hat{n}_{k+m'}^{m,t}-1)\tau+\rho] \vee 0}^{[(\hat{n}_{k+m'+1}^{m,t}-1)\tau+\rho] \vee 0} f(s) ds d\rho \\
&= \int_0^\tau \sum_{m'=-M}^{-1} \sum_{k \in K_t^m} \int_{[(\hat{n}_{k+m'}^{m,t}-1)\tau+\rho] \vee 0}^{[(\hat{n}_{k+m'+1}^{m,t}-1)\tau+\rho] \vee 0} f(s) ds d\rho \\
&\leq \int_0^\tau \sum_{m'=-M}^{-1} \int_0^t f(s) ds d\rho \\
&= M \int_0^\tau \int_0^t f(s) ds d\rho = M\tau \int_0^t f(s) ds. \square
\end{aligned} \tag{59}$$

Applying Lemma 6.5 to (51) yields

$$\begin{aligned}
\frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 &\leq \sum_{m \in \mathcal{M}} \left\{ (M+2)\tau \int_0^t K_1 (|\xi_\rho^\varepsilon|^2 + |\bar{\xi}_\rho^\varepsilon|^2 + |w_\rho^\varepsilon|^2) d\rho \right. \\
&\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\}.
\end{aligned} \tag{60}$$

Now, by the system structure given by Assumption Block (A.m) and by the fact that the V^m are in \mathcal{G}_δ , one obtains the following lemmas exactly as in [37].

Lemma 6.6 *For any $t < \infty$ and $\varepsilon > 0$,*

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

Lemma 6.7 *For any $t < \infty$ and $\varepsilon > 0$,*

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

and

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

Applying Lemmas 6.6 and 6.7 to (60), one obtains

$$\begin{aligned} \frac{1}{2} |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 \leq & \sum_{m \in \mathcal{M}} \left\{ K_2 (1 + |x|^2) \tau \right. \\ & \left. + \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr \right\} \end{aligned} \quad (61)$$

for proper choice of K_2 independent of x, t and $\hat{\varepsilon} \leq 1$.

Now we deal with the end parts (the integrals over $\widehat{\mathcal{I}}_f^{m,t}$ and $\widehat{\mathcal{I}}_f^{m,t}$). By Lemma 6.3, for any $m \in \mathcal{M}$, $\lambda_{N^t \tau}^m - \bar{\lambda}_{N^t \tau}^m \leq (M-1)\tau$. By the monotonicity of λ^m , this implies

$$\lambda_t^m - \bar{\lambda}_{N^t \tau}^m \leq M\tau,$$

and this implies $\mathcal{L}(\widehat{\mathcal{I}}_f^{m,t}) \leq M\tau$ for all $t \in [0, T]$. Also, by the definition of $\widehat{\mathcal{I}}_f^{m,t}$, $\mathcal{L}(\widehat{\mathcal{I}}_f^{m,t}) \leq t - N^t \tau \leq \tau$ for all $m \in \mathcal{M}$ and $t \in [0, T]$. Also, note that

$$\frac{d}{dt} [|\xi^\varepsilon|^2] \leq -2c_A |\xi^\varepsilon|^2 + 2c_\sigma |\xi^\varepsilon| |w^\varepsilon| \leq -c_A |\xi^\varepsilon|^2 + \frac{c_\sigma^2}{c_A} |w^\varepsilon|^2$$

which implies

$$|\xi_t^\varepsilon|^2 \leq |x|^2 + \frac{c_\sigma^2}{c_A} \|w^\varepsilon\|^2 \quad (62)$$

for all $t \in [0, T]$. Combining this with Lemma 6.6 implies

$$|\xi_t^\varepsilon|^2 \leq K_3 (1 + |x|^2) \quad \forall t \geq 0 \quad (63)$$

for appropriate choice of K_3 independent of $\hat{\varepsilon} \leq 1$ and $t \in [0, \infty)$. Similarly, one obtains

$$|\bar{\xi}_t^\varepsilon|^2 \leq K_3(1 + |x|^2) \quad \forall t \geq 0, \quad (64)$$

independent of $\hat{\varepsilon} \leq 1$ and $t \in [0, \infty)$. Now note that

$$\begin{aligned} & \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \xi_r^\varepsilon dr - \int_{\widehat{\mathcal{I}}_f^{m,t}} e^{2c_A(r-t)} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^m \bar{\xi}_r^\varepsilon dr \\ & \leq \int_{\widehat{\mathcal{I}}_f^{m,t}} \frac{d_A}{2} [3|\xi_r^\varepsilon|^2 + |\bar{\xi}_r^\varepsilon|^2] dr + \int_{\widehat{\mathcal{I}}_f^{m,t}} \frac{d_A}{2} [3|\xi_r^\varepsilon|^2 + |\bar{\xi}_r^\varepsilon|^2] dr \end{aligned}$$

(where we recall $d_A = \max_{m \in \mathcal{M}} |A^m|$), which by (63) and (64),

$$\leq \int_{\widehat{\mathcal{I}}_f^{m,t}} 2d_A K_3(1 + |x|^2) dr + \int_{\widehat{\mathcal{I}}_f^{m,t}} 2d_A K_3(1 + |x|^2) dr,$$

which by the remarks just above,

$$\leq 2(M+1)d_A K_3(1 + |x|^2)\tau. \quad (65)$$

Combining (61) and (65) yields

$$\frac{1}{2}|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 \leq K_4(1 + |x|^2)\tau \quad (66)$$

for proper choice of K_4 independent of x, t and $\hat{\varepsilon} \leq 1$.

Now that we have a bound on $\frac{1}{2}|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2$, we also obtain a bound on $\int_0^t \frac{1}{2}|\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr$ in a similar fashion. Note that

$$\frac{d}{dt} \int_0^t \frac{1}{2}|\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr = \frac{1}{2}|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2,$$

which, using $\xi_0^\varepsilon - \bar{\xi}_0^\varepsilon = 0$,

$$\begin{aligned} & = \int_0^t (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} \xi_r^\varepsilon - A^{\bar{\mu}_r^\varepsilon} \bar{\xi}_r^\varepsilon) dr \\ & = \int_0^t (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' (A^{\mu_r^\varepsilon} - A^{\bar{\mu}_r^\varepsilon}) \xi_r^\varepsilon dr + \int_0^t (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^{\bar{\mu}_r^\varepsilon} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon) dr \\ & \leq -2c_A \int_0^t \frac{1}{2}|\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr + \int_0^t (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^{\bar{\mu}_r^\varepsilon} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon) dr. \end{aligned}$$

In other words, letting $z_t \doteq \int_0^t \frac{1}{2}|\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr$, we have $\dot{z} \leq -2c_A z + \int_0^t (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon)' A^{\bar{\mu}_r^\varepsilon} (\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon) dr$ where $z_0 = 0$. Solving this differential inequality yields

$$\int_0^t \frac{1}{2}|\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr \leq \int_0^t e^{-2c_A(t-r)} \int_0^r (\xi_\rho^\varepsilon - \bar{\xi}_\rho^\varepsilon)' (A^{\mu_\rho^\varepsilon} - A^{\bar{\mu}_\rho^\varepsilon}) \xi_\rho^\varepsilon d\rho dr. \quad (67)$$

Now, proceeding exactly as above, but without an $e^{-2c_A(r-\rho)}$ term in the integral (which was irrelevant in the above bound on $\frac{1}{2}|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2$, one finds

$$\int_0^r (\xi_\rho^\varepsilon - \bar{\xi}_\rho^\varepsilon)' (A^{\mu_\rho^\varepsilon} - A^{\bar{\mu}_\rho^\varepsilon}) \xi_\rho^\varepsilon d\rho \leq K_4(1 + |x|^2)\tau \quad (68)$$

for all $x \in \mathbb{R}^n$, $\hat{\varepsilon} \leq 1$ and $r \in [0, \infty)$. Substituting (68) into (67) yields

$$\int_0^t \frac{1}{2} |\xi_r^\varepsilon - \bar{\xi}_r^\varepsilon|^2 dr \leq \int_0^t e^{-2c_A(t-r)} K_4(1 + |x|^2) \tau dr \leq \frac{K_4}{2c_A} (1 + |x|^2) \tau. \quad (69)$$

Now that we have the above bounds on these differences between ξ_t^ε and $\bar{\xi}_t^\varepsilon$, we turn to bounding the difference between $\tilde{S}_T[V^m]$ and $\bar{S}_T^\tau[V^m]$. Note that

$$\begin{aligned} & \int_0^T l^{\mu_i^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + V^m(\xi_T^\varepsilon) - \left[\int_0^T l^{\bar{\mu}_i^\varepsilon}(\bar{\xi}_t^\varepsilon) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + V^m(\bar{\xi}_T^\varepsilon) \right] \\ &= \frac{1}{2} \int_0^T \xi_t^\varepsilon D^{\mu_i^\varepsilon} \xi_t^\varepsilon - \bar{\xi}_t^\varepsilon D^{\bar{\mu}_i^\varepsilon} \bar{\xi}_t^\varepsilon dt + \frac{1}{2} (\xi_T^\varepsilon)' P^m \xi_T^\varepsilon - \frac{1}{2} (\bar{\xi}_T^\varepsilon)' P^m \bar{\xi}_T^\varepsilon. \end{aligned} \quad (70)$$

We work first with the integral terms on the right-hand side. One has

$$\int_0^T \xi_t^\varepsilon D^{\mu_i^\varepsilon} \xi_t^\varepsilon - \bar{\xi}_t^\varepsilon D^{\bar{\mu}_i^\varepsilon} \bar{\xi}_t^\varepsilon dt = \int_0^T (\xi_t^\varepsilon)' (D^{\mu_i^\varepsilon} - D^{\bar{\mu}_i^\varepsilon}) \xi_t^\varepsilon dt + \int_0^T (\xi_t^\varepsilon)' D^{\bar{\mu}_i^\varepsilon} \xi_t^\varepsilon - (\bar{\xi}_t^\varepsilon)' D^{\bar{\mu}_i^\varepsilon} \bar{\xi}_t^\varepsilon dt \quad (71)$$

Also,

$$\begin{aligned} & \int_0^T (\xi_t^\varepsilon)' D^{\bar{\mu}_i^\varepsilon} \xi_t^\varepsilon - (\bar{\xi}_t^\varepsilon)' D^{\bar{\mu}_i^\varepsilon} \bar{\xi}_t^\varepsilon dt \\ &= \int_0^T (\xi_t^\varepsilon + \bar{\xi}_t^\varepsilon)' D^{\bar{\mu}_i^\varepsilon} (\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon) dt \\ &\leq \int_0^T |D^{\bar{\mu}_i^\varepsilon}| \left[|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 + 2|\xi_t^\varepsilon| |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon| \right] dt \\ &\leq c_D \left[\int_0^T |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 dt + 2\|\xi^\varepsilon\| \left\{ \int_0^T |\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|^2 dt \right\}^{1/2} \right], \end{aligned}$$

which by (69) and Lemma 6.7,

$$\begin{aligned} & \leq 2c_D M^2 K_5(1 + |x|^2) \tau + 2 \left\{ K_6(1 + |x|^2) \right\}^{1/2} \left\{ 2M^2 K_5(1 + |x|^2) \tau \right\}^{1/2} \\ & \leq K_7(1 + |x|^2)(\tau + \sqrt{\tau}), \end{aligned} \quad (72)$$

for appropriate K_5, K_6, K_7 independent of x, T, τ and $\hat{\varepsilon} \leq 1$.

Now we turn to the first integral on the right-hand side of (71). Bounding this term is more difficult, and we use techniques similar to those used in bounding $|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|$ above. Note that

$$\begin{aligned} \int_0^T (\xi_t^\varepsilon)' (D^{\mu_i^\varepsilon} - D^{\bar{\mu}_i^\varepsilon}) \xi_t^\varepsilon dt &= \sum_{m \in \mathcal{M}} \left\{ \int_{\mathcal{I}_T^m} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\bar{\mathcal{I}}_T^m} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\} \\ &= \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_T^m} \left[\int_{\widehat{\mathcal{I}}_k^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{(n_k^{m,T} - 1)\tau}^{n_k^{m,T} \tau} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right] \right. \\ & \quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\bar{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\} \end{aligned}$$

which as in (49)–(51),

$$\begin{aligned}
&= \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_T^m} \left[\int_{(n_k^{m,T}-1)\tau}^{n_k^{m,T}\tau} (\xi_{\Phi_k^{m,T}}^\varepsilon(t))' D^m \xi_{\Phi_k^{m,T}}^\varepsilon(t) - (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right] \right. \\
&\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\} \\
&= \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_T^m} \left[\int_{(n_k^{m,T}-1)\tau}^{n_k^{m,T}\tau} \int_t^{\Phi_k^{m,T}(t)} 2(\xi_r^\varepsilon)' D^m (A^{\mu_r^\varepsilon} \xi_r^\varepsilon + \sigma w_r^\varepsilon) dr dt \right] \right. \\
&\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\} \\
&\leq \sum_{m \in \mathcal{M}} \left\{ \sum_{k \in K_T^m} \left[\int_{(n_k^{m,T}-1)\tau}^{n_k^{m,T}\tau} \int_t^{\Phi_k^{m,T}(t)} 2K_8(|\xi_r^\varepsilon|^2 + |w_r^\varepsilon|^2) dr dt \right] \right. \\
&\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\},
\end{aligned}$$

for appropriate choice of K_8 independent of x, T, τ and $\hat{\varepsilon} \leq 1$. By Lemma 6.5, this implies

$$\begin{aligned}
&\int_0^T (\xi_t^\varepsilon)' (D^{\mu_i^\varepsilon} - D^{\bar{\mu}_i^\varepsilon}) \xi_t^\varepsilon dt \\
&\leq \sum_{m \in \mathcal{M}} \left\{ (M+2)\tau \int_0^T K_8(|\xi_r^\varepsilon|^2 + |w_r^\varepsilon|^2) dr \right. \\
&\quad \left. + \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\},
\end{aligned}$$

which, by Lemmas 6.6 and 6.7,

$$\leq K_9(1 + |x|^2)\tau + \sum_{m \in \mathcal{M}} \left\{ \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt - \int_{\widehat{\mathcal{I}}_f^{m,T}} (\xi_t^\varepsilon)' D^m \xi_t^\varepsilon dt \right\},$$

for appropriate choice of K_9 independent of x, T, τ and $\hat{\varepsilon} \leq 1$. Then, applying the same steps as used in bounding $|\xi_t^\varepsilon - \bar{\xi}_t^\varepsilon|$, this is

$$\leq K_{10}(1 + |x|^2)\tau \tag{73}$$

for appropriate choice of K_{10} independent of x, T, τ and $\hat{\varepsilon} \leq 1$. Combining (71), (72) and (73)

$$\int_0^T \xi_t^\varepsilon D^{\mu_i^\varepsilon} \xi_t^\varepsilon - \bar{\xi}_t^\varepsilon D^{\bar{\mu}_i^\varepsilon} \bar{\xi}_t^\varepsilon dt \leq K_7(1 + |x|^2)(\tau + \sqrt{\tau}) + K_{10}(1 + |x|^2)\tau. \tag{74}$$

Now we turn to the last two terms in (70). As with the above integral terms,

$$(\xi_T^\varepsilon)' P^m \xi_T^\varepsilon - (\bar{\xi}_T^\varepsilon)' P^m \bar{\xi}_T^\varepsilon \leq |P^m| \left[|\xi_T^\varepsilon - \bar{\xi}_T^\varepsilon|^2 + 2|\xi_T^\varepsilon| |\xi_T^\varepsilon - \bar{\xi}_T^\varepsilon| \right]$$

which by (66) and (64),

$$\leq c_P \left[2K_4(1 + |x|^2)\tau + 2\sqrt{K_3}\sqrt{1 + |x|^2}\sqrt{2K_4(1 + |x|^2)\tau} \right]$$

where $c_P = \max_{m \in \mathcal{M}} |P^m|$,

$$\leq K_{11}(1 + |x|^2)(\tau + \sqrt{\tau}) \quad (75)$$

for appropriate choice of K_{11} independent of x, T, τ, m and $\hat{\varepsilon} \leq 1$.

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

$$\begin{aligned} & \int_0^T l^{\mu_t^\varepsilon}(\xi_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2 dt + V^m(\xi_T^\varepsilon) - \left[\int_0^T l^{\bar{\mu}_t^\varepsilon}(\bar{\xi}_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2 dt + V^m(\bar{\xi}_T^\varepsilon) \right] \\ & \leq K_\delta(1 + |x|^2)(\tau + \sqrt{\tau}), \end{aligned} \quad (76)$$

for appropriate choice of K_δ independent of x, T, τ, m and $\hat{\varepsilon} \leq 1$.

Combining (38) and (76), one has

$$\tilde{S}_T[V^m](x) - \int_0^T l^{\bar{\mu}_t^\varepsilon}(\bar{\xi}_t^\varepsilon) - \frac{\gamma^2}{2}|w_t^\varepsilon|^2 dt + V^m(\bar{\xi}_T^\varepsilon) \leq \frac{\hat{\varepsilon}}{2}(1 + |x|^2) + K_\delta(1 + |x|^2)(\tau + \sqrt{\tau}).$$

This implies

$$\tilde{S}_T[V^m](x) - \bar{S}_T^\tau[V^m](x) \leq \frac{\hat{\varepsilon}}{2}(1 + |x|^2) + K_\delta(1 + |x|^2)(\tau + \sqrt{\tau}),$$

and since this is true for all $\hat{\varepsilon} \in (0, 1]$, we finally obtain

$$\tilde{S}_T[V^m](x) - \bar{S}_T^\tau[V^m](x) \leq K_\delta(1 + |x|^2)(\tau + \sqrt{\tau}),$$

which completes the proof of Theorem 6.1. \square

Recall that a reasonable initialization is $\bar{V}^0 = \bigoplus_{m \in \mathcal{M}} V^m$. Using the max-plus linearity of \tilde{S}_τ , one easily sees that Theorem 6.1 implies

Corollary 6.8 *There exists $K_\delta < \infty$ such that, for all sufficiently small $\tau > 0$,*

$$0 \leq \tilde{S}_T[\bar{V}^0](x) - \bar{S}_T^\tau[\bar{V}^0](x) \leq K_\delta(1 + |x|^2)(\tau + \sqrt{\tau})$$

for all $T \in (0, \infty)$, $x \in \mathbb{R}^n$, and $m \in \mathcal{M}$.

7 Finite-time truncation errors

Going back to our main subject, we want to get an estimate of the convergence speed of $\bar{S}_T^\tau[\bar{V}^0](x)$ toward $\tilde{V}(x)$ as $T \rightarrow \infty$ and $\tau \downarrow 0$. In the previous section we have already obtained an estimate of the convergence of $\bar{S}_T^\tau[\bar{V}^0](x)$ to $\tilde{S}_T[\bar{V}^0](x)$. We now need to evaluate the difference between $\tilde{S}_T[\bar{V}^0](x)$ and $\tilde{V}(x)$ as $T \rightarrow \infty$.

Theorem 7.1 For all $\delta > 0$ satisfying $V \in \mathcal{G}_\delta, \tilde{V} \in \mathcal{G}_\delta$, there exists \bar{K}_δ such that for all $T > 0$ and all $x \in \mathbb{R}^n$,

$$|\tilde{S}_T[V](x) - \tilde{V}(x)| \leq \frac{\bar{K}_\delta}{T}(1 + |x|^2). \quad (77)$$

PROOF. The proof is similar to results in [29], [30], [37]. In particular, let $\delta > 0$ be such that with $\hat{\gamma}^2 \doteq (\gamma - \delta)^2$, one has $\frac{\hat{\gamma}^2 c_A^2}{c_\sigma^2 c_D} > 1$, and suppose $V \in \mathcal{G}_\delta$, i.e., such that

$$0 \leq V(x) \leq \frac{c_A \hat{\gamma}^2}{2c_\sigma^2} |x|^2. \quad (78)$$

Recall the integral bound from Lemma 6.7 (which held for all $t < \infty$ and $\varepsilon \in (0, 1]$). As in [29] and [37], applying the bound to the integral between $\bar{T}/2$ and \bar{T} , one finds that for any $\bar{T} > 0$, there exists $T \in [\bar{T}/2, \bar{T}]$ such that

$$\begin{aligned} |\xi_T^\varepsilon|^2 &\leq \frac{2\varepsilon}{\bar{T}} \left\{ \frac{2\varepsilon}{\delta} \frac{c_\sigma^2}{c_A^2} + \left[\frac{c_\sigma^2}{\delta c_A} \left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2} \right) + \frac{1}{c_A} \right] |x|^2 \right\} \\ &\leq \frac{2\varepsilon}{\bar{T}} \left\{ \frac{2\varepsilon}{\delta} \frac{c_\sigma^2}{c_A^2} + \left[\frac{c_\sigma^2}{\delta c_A} \left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2} \right) + \frac{1}{c_A} \right] \right\} (1 + |x|^2). \end{aligned} \quad (79)$$

Also, for all $t \geq \bar{T}$ one has by (62),

$$|\xi_t^\varepsilon|^2 \leq |\xi_T^\varepsilon|^2 + \frac{c_\sigma^2}{c_A} \|w^\varepsilon\|_{L^2[T, t]}^2$$

which by Lemma 6.6,

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

Replacing $|\xi_T^\varepsilon|^2$ by its bound in (79) we find,

$$\begin{aligned} |\xi_t^\varepsilon|^2 &\leq \frac{2\varepsilon}{\delta} \frac{c_\sigma^2}{c_A} \\ &\quad + \frac{2\varepsilon}{\bar{T}} \left\{ \frac{2\varepsilon}{\delta} \frac{c_\sigma^2}{c_A^2} + \underbrace{\left[\frac{c_\sigma^2}{\delta c_A} \left(\frac{c_D}{c_A^2} + \frac{\gamma^2}{c_\sigma^2} \right) + \frac{1}{c_A} \right]}_{C_1} \right\} \underbrace{\left\{ 1 + \frac{c_\sigma^2}{\delta c_A} \left[\frac{c_D}{c_A} + \frac{c_A \gamma^2}{c_\sigma^2} \right] \right\}}_{C_2} (1 + |x|^2). \end{aligned} \quad (80)$$

Finally, for any $T > 0$ given, and any $w^\varepsilon, \mu^\varepsilon$ ε -optimal on $[0, T]$,

$$\begin{aligned} \tilde{S}_T[V](x) &\leq \tilde{J}(x, T, w^\varepsilon, \mu^\varepsilon) + \varepsilon \\ &\leq \int_0^T \mu^{\mu^\varepsilon}(\xi_t) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + V(\xi_T^\varepsilon) + \varepsilon \end{aligned}$$

which by (78), (80) and $\tilde{V} \geq 0$,

$$\begin{aligned}
&\leq \int_0^T l^{\mu^\varepsilon}(\xi_t) - \frac{\gamma^2}{2} |w_t^\varepsilon|^2 dt + \underbrace{\tilde{V}(\xi_T^\varepsilon)}_{\geq 0} + \varepsilon \\
&\quad + \underbrace{\frac{c_A \hat{\gamma}^2}{2c_\sigma^2} \left\{ \frac{2\varepsilon c_\sigma^2}{\delta c_A} + \frac{2}{T} \left[\frac{2\varepsilon c_\sigma^2}{\delta c_A^2} + C_1 \right] C_2(1 + |x|^2) \right\}}_{\geq V(\xi_T^\varepsilon)}
\end{aligned}$$

where the lower bracket bound follows from Assumption Block (A.m) and (80). This is

$$\begin{aligned}
&\leq \tilde{S}_T[\tilde{V}](x) + \varepsilon + \frac{c_A \hat{\gamma}^2}{2c_\sigma^2} \left\{ \frac{2\varepsilon c_\sigma^2}{\delta c_A} + \frac{2}{T} \left[\frac{2\varepsilon c_\sigma^2}{\delta c_A^2} + C_1 \right] C_2(1 + |x|^2) \right\} \\
&= \tilde{V}(x) + \varepsilon + \frac{c_A \hat{\gamma}^2}{2c_\sigma^2} \left\{ \frac{2\varepsilon c_\sigma^2}{\delta c_A} + \frac{2}{T} \left[\frac{2\varepsilon c_\sigma^2}{\delta c_A^2} + C_1 \right] C_2(1 + |x|^2) \right\}.
\end{aligned}$$

Since this is true for all $\varepsilon > 0$, one sees that for all $T > 0$,

$$\tilde{S}_T[V](x) \leq \tilde{V}(x) + \frac{2}{T} \frac{c_A \hat{\gamma}^2}{2c_\sigma^2} C_1 C_2 (1 + |x|^2),$$

which for proper choice of \bar{K}_δ ,

$$\leq \tilde{V}(x) + \frac{\bar{K}_\delta}{T} (1 + |x|^2).$$

We note that we can repeat exactly the same reasoning with $\tilde{V}(x) = \tilde{S}_T[\tilde{V}](x)$ on the left side of the inequalities and $\tilde{S}_T[V](x)$ on the right side. Hence, we obtain the result,

$$|\tilde{S}_T[V](x) - \tilde{V}(x)| \leq \frac{\bar{K}_\delta}{T} (1 + |x|^2). \quad \square$$

8 Combined errors

We are now able to give a precise estimate of the convergence of $\bar{S}_T^\tau[\bar{V}^0]$ towards \tilde{V} as $\tau \downarrow 0$ and $T \rightarrow \infty$. Indeed we have,

$$\begin{aligned}
\bar{S}_T^\tau[\bar{V}^0](x) &\leq \tilde{V}(x) = \tilde{S}_T[\tilde{V}](x) \\
&\leq \tilde{S}_T[\bar{V}^0](x) + \frac{\bar{K}_\delta}{T} (1 + |x|^2) \\
&\leq \bar{S}_T^\tau[\bar{V}^0](x) + \frac{\bar{K}_\delta}{T} (1 + |x|^2) + K_\delta (1 + |x|^2) (\tau + \sqrt{\tau}).
\end{aligned}$$

For example, if we want $0 \leq \tilde{V}(x) - \bar{S}_T^\tau[\bar{V}^0](x) \leq 2\varepsilon(1 + |x|^2)$, we can choose,

1. $T \geq \bar{K}_\delta/\varepsilon \doteq \hat{T}$

2. τ such that $\tau + \sqrt{\tau} \leq \varepsilon/K_\delta$,

Suppose that such a τ satisfies $\tau \leq 1$. Then 2. becomes $\tau \leq [\varepsilon/(2K_\delta)]^2$. Hence to get an approximation of order ε , it is sufficient to have $N = T/\tau \propto \varepsilon^{-3}$.

9 The theoretical and the practical

There are two aspects to this work. One is the theoretical result that there exists a numerical method (for this class of problems) which is not subject to the curse-of-dimensionality. The second is the question of the practicality of the new approach. As with the interior point methods developed for linear programming, where construction of increasingly fast methods required further advances over the initial algorithm concept, it is clear that this will be the case here as well. We will discuss the main theoretical computational cost bound, and then turn to some remarks on the issue of practical implementation.

Suppose one begins the algorithm with an initial \bar{V}^0 , or equivalently, an initial \bar{a}^0 , of the form $\bar{V}^0 = \bigoplus_{j=1}^{J_0} \hat{V}_j^0$ for some set of initial quadratic forms, $\mathcal{V}^0 = \{\hat{V}_j^0\}_{j=1}^{J_0}$. Let $\mathcal{J}_0 \doteq]1, J_0[$. From Section 8, one sees that we can obtain an approximate solution, V^ε , with error

$$0 \leq \tilde{V} - V^\varepsilon \leq \varepsilon(1 + |x|^2)$$

where $V^\varepsilon = \bar{S}_{N\tau}^{\bar{S}}[\bar{V}^0]$ in $N_\varepsilon = \lceil (\bar{K}/\varepsilon^3) \rceil$ steps where \bar{K} may depend on $c_A, \bar{A}, c_D, c_\sigma, \gamma$ and choice of \mathcal{V}^0 . Then, the number of elements in the initial set of triples is $\#\hat{\mathcal{Q}}_0 = J_0$. Recalling the algorithm from Section 4, at the $k+1$ step, one generates each of the triples in $\hat{\mathcal{Q}}_{k+1}$ from the triples in $\hat{\mathcal{Q}}_k$ by matrix/vector operations requiring $\tilde{C}(1+n^3)$ operations where we recall that n is the space dimension, and \tilde{C} is a universal constant. Note that there are $J_0 M^k$ triples in $\hat{\mathcal{Q}}_k$ at each step k . Consequently, we have the following result, which is of theoretical importance as it clearly states the freedom from the curse-of-dimensionality.

Theorem 9.1 *Suppose one initializes with $\mathcal{V}^0 = \{\hat{V}_j^0\}_{j=1}^{J_0}$ (where $\bar{V}^0 = \bigoplus_{j=1}^{J_0} \hat{V}_j^0$). Then number of arithmetic operations required to obtain a solution with error no greater than $\varepsilon(1 + |x|^2)$ over the entire space is bounded by*

$$\tilde{C} J_0 \left[\sum_{k=0}^{N_\varepsilon} M^k \right] (1 + n^3)$$

where $N_\varepsilon = \lceil (\bar{K}/\varepsilon^3) \rceil$, \bar{K} may depend on $c_A, \bar{A}, c_D, c_\sigma, \gamma$ and choice of \mathcal{V}^0 , and \tilde{C} is universal.

Two remarks regarding the practicality are appropriate here. First, this approach is clearly most appropriate when one is not attempting to obtain a solution with extremely

small errors, due to the curse-of-complexity (i.e., the exponential growth with base M). Second, direct implementation of the algorithm of Section 4 is likely not reasonable without some technique for mitigation of the exponential growth rate. We note the following techniques for reduction of the computational complexity, without which the algorithm is not practical.

9.1 Initialization

In practice, we have found that the initialization $\bar{V}^0 = \bigoplus_{m=1}^M V^m$ (where we recall that each $V^m(x) = \frac{1}{2}x'P^m x$ and P^m is the solution of the Riccati equation for the m^{th} constituent Hamiltonian) greatly reduces the computation over the use of initialization $\bar{V}^0 \equiv 0$. The saving are due to the reduction of N_ε .

9.2 Pruning

We have also found that quite often the overwhelming majority of the $J_0 M^k$ triples at the k^{th} step do not contribute at all to \bar{V}^k . That is, they never achieve the maximum value over \mathcal{M} at any point $x \in \mathbb{R}^n \setminus \{0\}$, and this provides an opportunity for reduction of computational cost.

Recall that with initialization $\bar{a}^0 = \bigoplus_{j \in \mathcal{J}_0} \hat{a}^0$, the solution at each step had the representation given by (35). We now introduce the additional notation

$$\hat{q}_{(j,m^k)}^k \doteq \left(\hat{Q}_{(j,m^k)}^k, \hat{z}_{(j,m^k)}^k, \hat{r}_{(j,m^k)}^k \right)$$

for all $j \in \mathcal{J}_0$ and $m^k \in \mathcal{M}^k$. Given any $\hat{q}_{(\bar{j},\bar{m}^k)}^k \in \hat{\mathcal{Q}}_k$, let the set of all its progeny at step $k + \kappa$ be denoted by

$$\hat{\mathcal{Q}}_{(\bar{j},\bar{m}^k)}^{k,\kappa} \doteq \left\{ \hat{q}_{(\bar{j},\{m^k, m^\kappa\})}^{k+\kappa} \mid m^\kappa \in \mathcal{M}^\kappa \right\}.$$

Let

$$\hat{a}_{(j,m^k)}^k(x) \doteq \frac{1}{2}(x - \hat{z}_{(j,m^k)}^k)' \hat{Q}_{(j,m^k)}^k (x - \hat{z}_{(j,m^k)}^k) + \hat{r}_{(j,m^k)}^k.$$

The following is obvious from the definition of the propagation.

Theorem 9.2 *Suppose $(\bar{j}, \bar{m}^k) \in \mathcal{J}_0 \times \mathcal{M}^k$ is such that*

$$\hat{a}_{(\bar{j},\bar{m}^k)}^k(x) < \bar{a}^k(x) \quad \forall x \in \mathbb{R}^n \setminus \{0\}. \quad (81)$$

Then, for all $\kappa > 0$ and all $m^\kappa \in \mathcal{M}^\kappa$,

$$\hat{a}_{(j,\{m^k, m^\kappa\})}^k(x) < \bar{a}^{k+\kappa}(x) \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

Let $\hat{\mathcal{Q}}_k^* \subset \hat{\mathcal{Q}}_k$. We say $\hat{\mathcal{Q}}_k^* \prec \hat{\mathcal{Q}}_k$ if (81) holds for all $\hat{q}_{(j,m^k)}^k \in \hat{\mathcal{Q}}_k^*$.

Corollary 9.3 Suppose $\widehat{\mathcal{Q}}_k^* \subset \widehat{\mathcal{Q}}$ satisfies $\widehat{\mathcal{Q}}_k^* \prec \widehat{\mathcal{Q}}_k$. Let

$$\widehat{\mathcal{Q}}_{k+\kappa}^* = \bigcup_{\widehat{q}_{(\bar{j}, \bar{m}^k)}^k \in \widehat{\mathcal{Q}}_k^*} \widehat{\mathcal{Q}}_{(\bar{j}, \bar{m}^k)}^{k, \kappa},$$

and $\widehat{\mathcal{Q}}_{k+\kappa}^1 \doteq \widehat{\mathcal{Q}}_{k+\kappa} \setminus \widehat{\mathcal{Q}}_{k+\kappa}^*$. Let $\overline{V}^{1, k+\kappa}$ be the solution generated by $\widehat{\mathcal{Q}}_{k+\kappa}^1$. Then

$$\overline{V}^{1, k+\kappa}(x) = \overline{V}^{k+\kappa}(x) \quad \forall x \in \mathbb{R}^n.$$

Let us say that $\widehat{q}_{(\bar{j}, \bar{m}^k)}^k$ is *strictly inactive* if (81) holds. Then, the corollary implies that one may prune any and all strictly inactive elements at any and all steps k with no repercussions. To give some idea of the computational savings with such pruning, suppose that one happened to remove the same fraction, f at each step of the iteration. Then the fractional computational reduction from that given by Theorem 9.1 would be

$$\frac{\sum_{k=0}^{N_\varepsilon} [(1-f)M]^k}{\sum_{k=0}^{N_\varepsilon} M^k},$$

which is (very) roughly $(1-f)^{N_\varepsilon}$.

In some of the examples tested so far, $1-f$ was small, and so this type of pruning has been useful in practical implementation of the approach. It should be noted that condition (81) is not obviously easy to check. Instead we utilize the simple, conservative, test where we prune $\widehat{q}_{(\bar{j}, \bar{m}^k)}^k$ if there exists (j, m^k) such that

$$\widehat{a}_{(\bar{j}, \bar{m}^k)}^k(x) < \widehat{a}_{(j, m^k)}^k(x) \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

(Note that this test may be done analytically.) In spite of the apparent conservativeness of this test, it has produced excellent computational savings over implementation without pruning in some (but not all) cases.

9.3 Over-pruning

Note that the algorithm converges to \overline{V} from any $\overline{V}^0 = \bigoplus_{j \in \mathcal{J}_0} \widehat{V}_j^0 \in \mathcal{G}_\delta$ for sufficiently small δ (with quadratic \widehat{V}_j^0 of course). Suppose that for steps $k \in \{1, 2, \dots, K_p\}$ (for some finite K_p), one pruned elements of $\widehat{\mathcal{Q}}_k$ which were not necessarily strictly inactive. By viewing the resulting \overline{V}^{K_p} as a new initialization, we see that the algorithm will nonetheless converge to the correct solution. Noting the curse-of-complexity, we see that this *over-pruning* (removing potentially useful triples) may be an attractive approach. In practice, we have employed this approach in a heuristic fashion by, for a fixed number of steps, removing all triples whose corresponding $\widehat{a}_{(j, m^k)}^k$ did not achieve the maximum over $(j, m^k) \in \mathcal{J}_0 \times \mathcal{M}^k$ on some fixed, pre-specified, finite set, $\mathcal{X}_p \subset \mathbb{R}^n$.

For example, in some tests, we chose \mathcal{X}_p to consist of the corners of the unit hypercube. Note that in this case, $\#\mathcal{X}_p = 2^n$, and so by performing this over-pruning for a fixed number of steps, we are introducing a curse-of-dimensionality-dependent component to the computations. (However, the growth rate of 2^n is extremely slow relative to that for grid-based techniques.) In the examples so far tested, employing this purely heuristic pruning for a quite a few steps led to *tremendous improvements* in computation time, or equivalently, multiple orders of magnitude reduction in solution error, $\varepsilon(1 + |x|^2)$.

This leads to the question of whether some approach that over-pruned the $\widehat{\mathcal{Q}}_k$, reducing the number of the (not strictly inactive) elements of $\widehat{\mathcal{Q}}_k$ by some fraction going to zero as k increased, might be a highly effective approach. Of course, one would use a pruning criterion that was not curse-of-dimensionality dependent.

Acknowledgements

The authors thank the referees for their substantial comments, which have led to a greatly improved paper.

References

- [1] M. Akian, S. Gaubert and A. Lakhoua, *The max-plus finite element method for solving deterministic optimal control problems: basic properties and convergence analysis*, SIAM J. Control and Optim., **47** (2008), 817–848.
- [2] M. Akian, S. Gaubert and A. Lakhoua, *A max-plus finite element method for solving finite horizon deterministic optimal control problems*, Proc. 16th International Symposium on Mathematical Theory of Networks and Systems (2004).
- [3] F.L. Baccelli, G. Cohen, G.J. Olsder and J.-P. Quadrat, *Synchronization and Linearity*, John Wiley, 1992.
- [4] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, Boston, 1997.
- [5] F. Camilli, M. Falcone, P. Lanucara and A. Seghini, *A Domain Decomposition Method for Bellman Equations*, Contemp. Math., **180** (1994), 477–483.
- [6] E. Carlini, M. Falcone and R. Ferretti, *An Efficient Algorithm for Hamilton-Jacobi Equations in High Dimensions*, Computing and Visualization Sci., **7** (2004), 15–29.
- [7] G. Cohen, S. Gaubert and J.-P. Quadrat, *Duality and Separation Theorems in Idempotent Semimodules*, Linear Algebra and Applications, **379** (2004), 395–422.

- [8] G. Collins and W.M. McEneaney, *Min-Plus Eigenvector Methods for Nonlinear H_∞ Problems with Active Control*, Optimal Control, Stabilization and Nonsmooth Analysis, LNCIS, Vol. 301, M.S. de Queiroz; M. Malisoff; P. Wolenski (Eds.), Springer (2004), 101–120.
- [9] M. G. Crandall, H. Ishii and P.-L. Lions, *Uniqueness of viscosity solutions of Hamilton–Jacobi equations revisited*, J. Math. Society Japan, **39** (1987) 581–595.
- [10] M. G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton–Jacobi Equations*, Trans. of the Amer. Math. Soc., **277** (1983), 1–42.
- [11] M. G. Crandall and P.-L. Lions, *On existence and uniqueness of solutions of Hamilton–Jacobi equations*, Nonlinear Anal. Theory and Appl., **10** (1986), 353–370.
- [12] R.A. Cuninghame-Green, *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems 166, Springer, 1979.
- [13] P. Dupuis and M. Boué, *Markov chain approximations for deterministic control problems with affine dynamics and quadratic cost in the control*, SIAM J. on Numerical Analysis, **36** (1999), 667–695.
- [14] P. Dupuis and A. Szpiro, *Convergence of the optimal feedback policies in a numerical method for a class of deterministic optimal control problems*, SIAM J. Control and Optim. **40** (2001), 393–420.
- [15] P. Dupuis and A. Szpiro, *Second order numerical methods for first order Hamilton–Jacobi equations*, SIAM J. on Num. Analysis, **40** (2002), 113–118.
- [16] L.C. Evans, *Partial Differential Equations*, Springer–Verlag, 1998.
- [17] M. Falcone and R. Ferretti, *Convergence analysis for a class of high-order semi-lagrangian advection schemes*, SIAM J. Num. Analysis, **35** (1998), 909–940.
- [18] W.H. Fleming, *Max-Plus Stochastic Processes*, Applied Math. and Optim., **49** (2004), 159–181.
- [19] W.H. Fleming, *Functions of Several Variables*, Springer-Verlag, 1977.
- [20] W.H. Fleming and W.M. McEneaney, *A max-plus based algorithm for an HJB equation of nonlinear filtering*, SIAM J. Control and Optim., **38** (2000), 683–710.
- [21] W.H. Fleming and H.M. Soner, *Controlled Markov Process and Viscosity Solutions*, Springer-Verlag, New York, 1993.
- [22] F. John, *Partial Differential Equations*, Springer-Verlag, New York, 1978.
- [23] J.W. Helton and M.R. James, *Extending H_∞ Control to Nonlinear Systems, Control of Nonlinear Systems to Achieve Performance Objectives*, Advances in Design and Control, SIAM, Philadelphia, 1999.

- [24] V.N. Kolokoltsov and V.P. Maslov, *Idempotent Analysis and Its Applications*, Kluwer, 1997.
- [25] H.J. Kushner and P. Dupuis, *Numerical methods for Stochastic Control Problems in Continuous Time*, Springer Verlag, New York, 1992.
- [26] G.L. Litvinov, V.P. Maslov and G.B. Shpiz, *Idempotent Functional Analysis: An Algebraic Approach*, Mathematical Notes, **69**, No. 5 (2001), 696–729.
- [27] V.P. Maslov, *On a new principle of superposition for optimization problems*, Russian Math. Surveys, **42** (1987), 43–54.
- [28] W.M. McEneaney, A. Deshpande and S. Gaubert, *Curse-of-Complexity Attenuation in the Curse-of-Dimensionality-Free Method for HJB PDEs*, Proc. ACC 2008, Seattle (2008).
- [29] W.M. McEneaney, *Max-Plus Methods for Nonlinear Control and Estimation*, Birkhauser Systems and Control Series, 2006.
- [30] W.M. McEneaney, *A Curse-of-Dimensionality-Free Numerical Method for Solution of Certain HJB PDEs*, SIAM J. on Control and Optim., **46** (2007), 1239–1276.
- [31] W.M. McEneaney, *Curse-of-Dimensionality Free Method for Bellman PDEs with Hamiltonian Written as Maximum of Quadratic Forms*, Proc. 44th IEEE Conf. on Decision and Control (2005).
- [32] W.M. McEneaney, *A Curse-of-Dimensionality-Free Numerical Method for a Class of HJB PDEs*, Proc. 16th IFAC World Congress (2005).
- [33] W.M. McEneaney, *Max-Plus Summation of Fenchel-Transformed Semigroups for Solution of Nonlinear Bellman Equations*, Systems and Control Letters, **56** (2007), 255-264.
- [34] W.M. McEneaney and P.M. Dower, *A max-plus affine power method for approximation of a class of mixed l_∞/l_2 value functions*, Proc. 42nd IEEE Conf. on Dec. and Control, Maui (2003), 2573–2578.
- [35] W.M. McEneaney, *Max-Plus Eigenvector Representations for Solution of Nonlinear H_∞ Problems: Error Analysis*, SIAM J. Control and Optim., **43** (2004), 379–412.
- [36] W.M. McEneaney *Error Analysis of a Max-Plus Algorithm for a First-Order HJB equation*, Stochastic Theory and Control, Proceedings of a Workshop held in Lawrence, Kansas, October 18-20, 2001, Lecture Notes in Control and Information Sciences, Springer-Verlag, B.Pasik-Duncan (Ed.) (2002), 335–352.
- [37] W.M. McEneaney, *A Uniqueness result for the Isaacs equation corresponding to non-linear H_∞ control*, Math. Controls, Signals and Systems, **11** (1998), 303–334.

- [38] R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional conference Series in Applied Mathematics 16, SIAM, 1974.
- [39] R.T. Rockafellar and R.J. Wets, *Variational Analysis*, Springer–Verlag, 1997.
- [40] P. Soravia, *H_∞ control of nonlinear systems: differential games and viscosity solutions*, SIAM J. Control and Optim., **34** (1996), 1071–1097.
- [41] A.J. van der Schaft, *L_2 -Gain analysis of nonlinear systems and nonlinear state feedback H_∞ control*, IEEE Trans. Auto. Control, **37** (1992), 770–784.