Second Order Numerical Methods for First Order Hamilton-Jacobi Equations [1]

Adam Szpiro and Paul Dupuis

Lefschetz Center for Dynamical Systems
Brown University
Division of Applied Mathematics
Providence, R.I. 02912

May 3, 1999

[1]This research was supported in part by the National Science Foundation (NSF-DMS-9704426), the Army Research Office (DAAH04-96-1-0075), and the Office of Naval Research (ONR-N000014-96-1-0276).

Abstract

We present practical numerical methods which produce provably second order approximations for a class of stationary first order Hamilton-Jacobi partial differential equations. Using probabilistic methods, we derive high order asymptotic expansions for a first order method, and then use those results to design second order methods. We prove second order convergence for the solution and for its gradient on a subset of the domain where the solution is smooth. Although we limit our attention to second order schemes, in principle the techniques in this paper can be extended to arbitrarily high order methods. Examples illustrate the rate of convergence as well as global sharp resolution of discontinuities. The Hamilton-Jacobi equations we consider correspond to deterministic optimal control problems, and our rate of convergence results are valid for the value functions and for the optimal feedback controls.

1 Introduction

This paper is concerned with the numerical solution of a class of first order stationary Hamilton-Jacobi partial differential equations. We give a detailed asymptotic analysis of a first order Markov chain based numerical method and then describe two approaches to obtaining second order convergence. The PDE's that we study arise in connection with a variety of applications in computer vision, large deviations, and robust control [4], and they can be interpreted as dynamic programming equations for a class of deterministic optimal control problems on a finite domain. Under our assumptions, the value function $V^0(x)$ for a given control problem is the unique viscosity solution to the corresponding Hamilton-Jacobi PDE, and the optimal feedback control function $u^0(x)$ is obtained by taking the argument in the optimization part of the Hamilton-Jacobi equation on those parts of the domain where it is uniquely defined.

Much of the difficulty in solving equations of this type derives from the fact that the solutions need not be smooth on the entire domain. Under our assumptions, $V^0(x)$ is globally Lipschitz, but it need not be differentiable at all points. The control function $u^0(x)$ is closely related to the gradient of $V^0(x)$. It need not be uniquely defined at all points, and it is often discontinuous across those points where it is not uniquely defined. Typically, however, there are regions of strong regularity which are open and dense in the domain, and in those regions both $V^0(x)$ and $u^0(x)$ are as smooth as the problem data. Our asymptotic results hold on certain subsets of the regions of strong regularity, and we propose numerical methods which provide second order estimates for $V^0(x)$ and $u^0(x)$ on those regions, without sacrificing global convergence properties.

The typical numerical methods are of finite difference type, or the probabilistic equivalent. Derivatives in the Hamilton-Jacobi PDE are replaced by finite difference approximations with a parameter $h > 0$ to obtain approximations $V^h(x)$ to the value function. There are two principal approaches to proving that these approximations converge to the limit value. A relatively recent approach is analytic and exploits the fact that $V^0(x)$ is the unique viscosity solution to the Hamilton-Jacobi PDE. Since the $V^h(x)$ are solutions to difference equations with operators converging to the limit differential operator, this uniqueness can be used to show that the $V^h(x)$ converge uniformly to $V^0(x)$; see [1, 7]. While the viscosity solution method analyzes the approximations at the level of the PDE, the approach we employ uses probabilistic methods to interpret the approximations at the level of the control problem. In this framework, the first numerical methods were developed by Kushner [26]. We regard the equation satisfied by $V^h(x)$ as the dynamic programming equation for a control problem with Markov chain dynamics which approximate the deterministic dynamics in the limit control problem. Thus, the $V^h(x)$ are value functions for approximating control problems, and we can use probabilistic arguments to show that they converge uniformly to $V^0(X)$ on the entire domain. See references [4, 26, 27]. In addition, probabilistic arguments have been employed in [12] to prove that the prelimit optimal feedback controls $u^h(x)$ converge to $u^0(x)$, uniformly on compact subsets of the regions of strong regularity where $u^0(x)$ is smooth.

All of the numerical methods which have been rigorously analyzed using either of the above approaches are intrinsically first order. That is because the finite differences are only first order approximations to the corresponding derivatives, or equivalently because the generators of the Markov chains approximate the generator for the limit dynamics to only first order when applied to smooth functions. The most natural approach to obtaining second order convergence would be to use second order finite difference operators in the approximations. This can sometimes be made to work in practice, but current techniques cannot be used to analyze the convergence properties, or even to show that the resulting approximations are stable. The problem is that the use of higher order finite difference operators results in the loss of critical monotonicity properties in the prelimit difference equations. From the point of view of Markov chain approximations, this gives rise to negative transition probabilities, so it is impossible to construct the dynamics. Our approach to obtaining second order convergence does not involve the use of higher order finite differences. However, in a different context (see the proof of Lemma 4.6), we introduce a representation which circumvents the problem of negative transition probabilities, at the cost of adding an exponential growth term. The growth term clearly exposes the stability issue which is critical in numerical methods based upon higher order finite differences, and this representation may prove useful in their analysis.

Our second order approximation methods are based upon an asymptotic expansion of the form

$$V^h(x) = V^0(x) + \sum_{m=1}^{M} h^m e^m(x) + o(h^M), \tag{1.1}$$

holding on a subset of the regions of strong regularity, with smooth functions $e^m(x)$. In Section 3, we obtain the first term in this expansion using techniques similar to those in [11]. Then, under the more restrictive conditions of Section 4, we use probabilistic arguments like those in [16] to establish the full asymptotic expansion. The situation in [16] involves the approximation of a deterministic optimal control problem by a stochastic problem with a small noise Brownian motion added to the dynamics. In that case, the asymptotic expansion is alternatively obtained by viscosity solution methods in reference [18]. In Section 5, we exploit equation (1.1) and an analogous asymptotic expansion for the optimal feedback control to obtain new approximations which are second order accurate in those regions where the expansions are valid. One approach is to define $V^{h,*}(x)$ by subtracting from $V^h(x)$ a correction term which approximates $he^1(x)$ to first order, and then to exploit the relationship between $u^0(x)$ and the gradient of $V^0(x)$ to obtain $u^{h,*}(x)$. A second method is to apply Richardson extrapolation, putting $V^{h,*}(x) = 2V^h(x) - V^{2h}(x)$, and defining $u^{h,*}(x)$ analogously. Given an expansion of the form (1.1), it is straightforward to see that approximations defined this way can be expected to be second order accurate. Our experimental results in Section 6 confirm this in both the L^1 and the L^∞ norms on the appropriate subsets of the regions of strong regularity, while first order convergence properties are maintained elsewhere.

Other approaches to obtaining higher order convergence for Hamilton-Jacobi PDE's have been suggested, especially for the finite time case. In reference [13], high order time stepping methods are explored, but the spatial discretizations are not considered. While the full discretization is discussed in [14], critical stability issues are not resolved for the general case. A theoretically interesting second order scheme is presented in [30], but as the authors note, it can not be practically implemented. High order essentially non-oscillatory (ENO) [31], weighted ENO [25], Godunov [29], and finite element [23] schemes originally developed for hyperbolic conservation laws have been successfully applied to Hamilton-Jacobi equations. While these methods have produced very convincing numerical results for finite time problems, complete convergence proofs are not available. To our knowledge, ours are the only practical schemes which have been proved to yield second order convergence of the value functions or of the optimal feedback controls on any part of the domain for multidimensional stationary Hamilton-Jacobi equations.

Conservation laws are closely related to the Hamilton-Jacobi equations that we study, and there is an extensive body of literature dealing with their numerical solution. In fact, the two types of equation are equivalent in one dimension, and it is interesting that in [30] the proof of convergence for a second order scheme for scalar conservation laws is obtained by means of viscosity solution methods for Hamilton-Jacobi equations. The scheme proposed in [22] for conservation laws is similar in philosophy to one of our second order methods, as it involves the use of an analogous correction term. An excellent overview of the available techniques can be found in reference [28]. While many effective higher order methods have been developed for time dependent problems, rigorous proofs of convergence in more than one dimension are not available. Furthermore, it has proven quite challenging to extend these methods to stationary problems; see reference [21] for some results in this direction.

We end this section with some notation. Let $I\!R^n$ be n-dimensional Euclidian space, and let $Z\!\!Z^n$ be the subset of $I\!R^n$ consisting of n-tuples of integers. For vectors $x, y \in I\!R^n$, $\langle x, y\rangle$ is the scalar product, $\|x\| = \sqrt{\langle x, x\rangle}$ is the Euclidean norm, $\|x\|_1 = \Sigma_{i=0}^n |x_i|$ is the l^1-vector norm, and $|x| = (|x_1|, \ldots, |x_n|)$ is the componentwise absolute value. For a process $X(\cdot)$ taking values in $I\!R^n$ and for $S < +\infty$, $|\!|\!|\, X(\cdot)\, |\!|\!|_S = \sup_{0\leq t\leq S} \|X(t)\|$ is the uniform L^2-norm. For any two subsets A and A' of $I\!R^n$, $d(A, A')$ denotes minimum Euclidean distance between $\bar{A}$ and $\bar{A}'$. The positive part of a scalar is $a^+ = \max(a, 0)$, and its negative part is $a^- = -\min(a, 0)$. For a vector, the positive and negative parts are taken componentwise, so that $x^\pm = (x_1^\pm, \ldots, x_n^\pm)$. In general, we use subscripts to denotes the components of a vector, while superscripts index possibly vector valued quantities. Thus, x_i is always a scalar quantity, while x^i may denote a vector.

For a smooth function f mapping $I\!R^n$ to $I\!R$ and for a positive integer r, put $D_i^r f(x) = \frac{\partial^r}{\partial^r x_i} f(x)$. The diagonal of the array of rth order partial derivatives is denoted by

$$D_\backslash^r f(x) = \Big(D_1^r f(x), \ldots, D_n^r f(x)\Big).$$

In the case $r = 1$, we use the standard notation $Df(x) = D_\backslash^1 f(x)$ for the gradient of f. For $h > 0$, the operators $D^{h,\pm}$ are one sided finite difference approximations to the gradient operator. The ith component of $D^{h,+} f(x)$ is given by

$$D_i^{h,+} f(x) = \frac{f(x + he_i) - f(x)}{h},$$

while the ith component of $D^{h,-} f(x)$ is given by

$$D_i^{h,-} f(x) = \frac{f(x) - f(x - he_i)}{h}.$$

In addition, $D^{h,c} f(x)$ is the centered difference approximation to the gradient vector $Df(x)$, while $D_\backslash^{2,h} f(x)$ is the centered difference approximation to the diagonal second derivative vector $D_\backslash^2 f(x)$.

2 Statement of Problem and First Order Numerical Approximation

We describe a deterministic optimal control problem and set forth a first order Markov chain based numerical method to approximate its solution. We then recall results which guarantee convergence of the numerical value function and of the numerical feedback control to their respective counterparts in the problem being approximated.

Let $G \subset I\!R^n$ be open with compact closure, and assume that G satisfies uniform interior and exterior cone conditions (see [4] for definitions). Let b and c be C^∞ functions from $I\!R^n$ to $I\!R$, and let a be a C^∞ function from $I\!R^n$ to the space of symmetric positive definite $n \times n$ matrices. Notice that a is uniformly positive definite on G. Assume that $c(x) \geq c_0 > 0$ on G. For a control $\underline{u}^0(t)$ which is in $L^2([0, S]; I\!R^n)$ for all $S < +\infty$ and for an initial condition $x \in G$, we define $\underline{X}^0(t)$ by the dynamics

$$\underline{X}^0(t) = x + \int_0^t \underline{u}^0(s)ds, \tag{2.1}$$

up to the time when it exits from the domain G. The corresponding generator $\mathcal{L}_u^0$ is given by

$$\mathcal{L}_u^0 f = \langle u, Df \rangle, \tag{2.2}$$

for any smooth function f mapping $\mathbb{R}^n$ to $\mathbb{R}$. We define the exit time $\underline{\tau}^0 = \inf[t : \underline{X}^0(t) \notin G]$. For the running cost

$$L(x,u) = \frac{1}{2}\Big\langle (u - b(x)), a^{-1}(x)(u - b(x)) \Big\rangle + c(x),$$

we define the payoff functional

$$J^0(x, \underline{u}^0) = \int_0^{\underline{\tau}^0} L(\underline{X}^0(t), \underline{u}^0(t))dt.$$

The problem is to minimize the payoff by choosing a suitable control. Define the value function,

$$V^0(x) = \inf_{\underline{u}^0} J^0(x, \underline{u}^0),$$

where the infimum is over controls $\underline{u}^0$ which are in $L^2([0,S];\mathbb{R}^n)$ for all $S < +\infty$. We employ the underscore notation here to indicate trajectories which are obtained from an arbitrary control. The same notations, without the underscores, will be used later to refer to trajectories which are obtained through the application of an optimal control.

The dynamics in (2.1) involve an open loop control $\underline{u}^0(t)$ which is defined for all $t > 0$. It is generally desirable, from the point of view of robustness and for convenience of implementation, to consider controls which can be represented in the feedback form

$$\underline{X}^0(t) = x + \int_0^t \underline{u}^0(\underline{X}^0(s))ds. \tag{2.3}$$

A key feature of the regions of strong regularity is that the optimal open loop controls for all initial conditions in a region of strong regularity correspond to a unique smooth feedback function $u^0(x)$. The following lemma is proved by elementary arguments in [12, Lemma 2.1]. We will use the $T < +\infty$ from this lemma frequently in our analysis.

Lemma 2.1 *$V^0(x)$ is bounded and uniformly Lipschitz on G, and there exists $T < +\infty$ such that every optimal trajectory exits from G by time $T-1$. Furthermore, there exists a compact set $U \subset \mathbb{R}^n$ such that every optimal open loop control is contained in the interior of U for each $0 \le t \le T-1$.*

The value function V^0 need not be differentiable on the entire domain G, but given Lemma 2.1, it follows from [2, Theorem 6.1] that it is the unique non-negative viscosity solution on G to the Hamilton-Jacobi dynamic programming equation (DPE)

$$\inf_u \Big[\mathcal{L}_u^0 V^0(x) + L(x,u)\Big] = 0, \tag{2.4}$$

with the continuous boundary condition $V^0(x) = 0$ on ∂G, where the generator $\mathcal{L}_u^0$ is defined in (2.2). See references [1] and [17] for a thorough account of the relationship between viscosity solutions of Hamilton-Jacobi PDE's and the value functions for various types of optimal control problems.

It turns out that V^0 is smooth on most of the domain G. Let Q be a subset of $\bar{G}$ which is open relative to G. We call Q a region of strong regularity if the following hold:

1. For each initial condition $x \in Q$, there is a unique optimal open loop control, and the corresponding trajectory $X_x^0(t)$ is contained in Q up to its exit time τ_x^0. The optimal trajectory meets ∂G nontangentially at a point z_x^0.

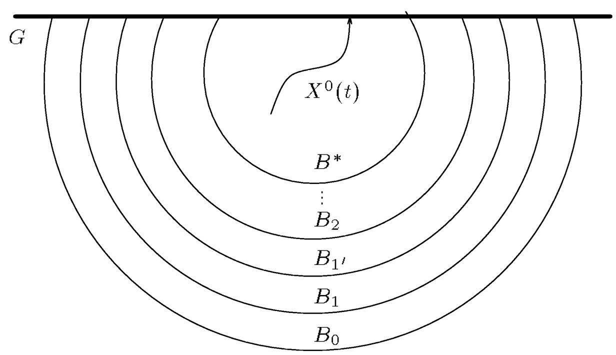


Figure 1: Regions of Strong Regularity

2. $V^0 \in C^\infty(Q)$.

3. There is a unique $u^0 \in C^\infty(Q)$ such that the optimal control can be represented in feedback form and is given by $u^0(x)$ for each $x \in Q$.

4. For $\xi \in Q \cap \partial G$, let T_ξ be the supremum of the exit times for optimal trajectories beginning in Q and exiting at ξ. The bijective map $\Xi(s,\xi)$ from

$$\{(s,\xi) : \xi \in Q \cap \partial G, 0 \le s < T_\xi\}$$

to Q given by $\Xi(\tau^0_x - t, z^0_x) = X^0_x(t)$ is non-singular in the sense that the quantities

$$\frac{\partial \Xi}{\partial s}(s,\xi), \frac{\partial \Xi}{\partial \nu_1}(s,\xi), \ldots, \frac{\partial \Xi}{\partial \nu_{n-1}}(s,\xi)$$

are linearly independent, where the ν_i are linearly independent tangent vectors to ∂G at the boundary point ξ.

The rather technical fourth condition in the above definition is the requirement that points in the regions of strong regularity not be conjugate. The uniqueness of optimal trajectories in the first condition is equivalent to the statement that the map $\Xi(s,\xi)$ is one-to-one. The non-degeneracy required in the fourth condition is a slightly stronger statement which precludes "almost" non-uniqueness in an infinitesimal sense. The classical method of characteristics and its application to proving the existence of regions of strong regularity for the present problem is discussed in the appendices of references [16, 18]. Detailed information on the structure of the regions of strong regularity for closely related problems can be found in references [5, 6, 15]. In general, the union of the regions of strong regularity is open and dense in the domain.

Let B^* be a subset of $\bar{G}$ such that $\bar{B}^* \subset Q$. Our asymptotic results will hold on sets which include $\bar{B}^*$. Let $M \ge 1$ be a fixed integer. While our choice of M here is arbitrary, it corresponds to the order of the asymptotic expansion that we will obtain in Section 4. Consider a region of strong regularity B_0, and for $i = 1, \ldots, M$, regions of strong regularity B_i and $B_{i'}$ such that

$$\bar{B}^* \subset B_{M'} \subset \ldots \subset B_1 \subset \bar{B}_1 \subset B_0 \subset \bar{B}_0 \subset Q, \tag{2.5}$$

and such that

$$\bar{B}_{i+1} \subset B_{i'} \subset \bar{B}_{i'} \subset B_i \tag{2.6}$$

for each $i = 1, \ldots, M$, where B_{M+1} is taken to be B^* in (2.6). The existence of such nested regions of strong regularity follows from Theorem 2.4 in reference [18]. These relationships are illustrated in Figure 1.

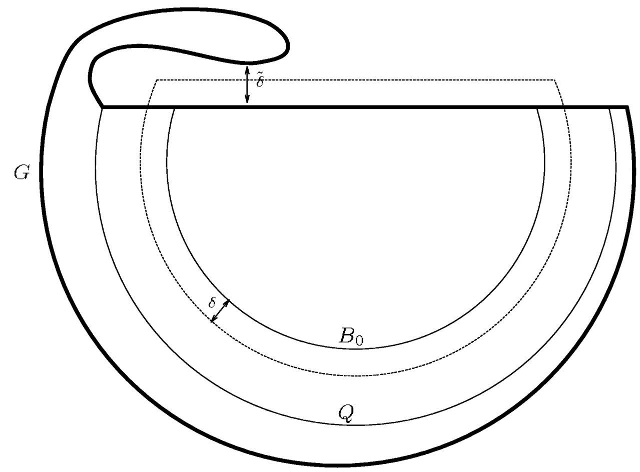


Figure 2: Region for Smooth Extension of u^0

Since V^0 is a classical solution to the DPE (2.4) on the region of strong regularity Q, the optimal feedback control can be explicitly evaluated there,

$$u^0(x) = -a(x)DV^0(x) + b(x). \tag{2.7}$$

We assume the following:

Assumption 2.2 *The boundary section $Q \cap \partial G$ is parallel to one of the coordinate hyperplanes. Furthermore, the minimum distance in the outward normal direction from $Q \cap \partial G$ to $\partial G / Q$ is equal to $\tilde{\delta} > 0$.*

We now present a first order Markov chain based numerical approximation to the deterministic optimal control problem discussed above, along with known convergence results. The method of approximation by Markov chains was first described by Kushner [26], and an up to date treatment can be found in the book of Kushner and Dupuis [27]. Our approximation is essentially the one used in [4] and in [12]. In the sections which follow, we will present more detailed asymptotic results for this approximation and then describe a new numerical method which yields a qualitatively better rate of convergence.

Let $h > 0$ be a discretization parameter and define the discrete domain $G^h = h\mathbb{Z}^n \cap G$. For any $A \subset \mathbb{R}^n$, we define $A^h = h\mathbb{Z}^n \cap A^o$, where A^o is the interior of A. We consider limits as $h \to 0$, with the h chosen such that the hyperplane in which the boundary section $Q \cap \partial G$ lies lines up with the lattice $h\mathbb{Z}^n$ (see Assumption 2.2 and Figure 2). We will construct a continuous time controlled jump Markov process on G^h which approximates the deterministic dynamics in (2.3).

Let $\underline{u}^h$ be any feedback control on G^h. Let $\underline{X}^h$ be the Markov process with controlled generator $\mathcal{L}_u^h$ given by

$$\mathcal{L}_u^h f = \langle u^+, D^{h,+} f \rangle - \langle u^-, D^{h,-} f \rangle$$

for any smooth function f mapping $\mathbb{R}^n$ to $\mathbb{R}$. See Section 1 for the notation in this definition. The corresponding stochastic dynamics will be called the h-dynamics. As in the description of the limit problem, we employ the underscore notation to indicate objects which are obtained from the application of an arbitrary possibly suboptimal feedback control.

Since we consider only feedback controls, it is straightforward to construct $\underline{X}^h$, as in Section 4.3 of [27] and in [8]. We define a sequence of i.i.d. exponential random fields parameterized by u, with mean values specified as follows:

$$\overline{\Delta t}^h(u) = \begin{cases} \frac{h}{\|u\|_1} & u \neq 0 \\ h & u = 0. \end{cases} \tag{2.8}$$

Suppose that after $m-1$ jumps, $\underline{X}^h(s)$ is defined for $0 \leq s \leq t$ and that $\underline{X}^h(t) = x$. Then we take $\underline{X}^h(s) = x$ for all $t \leq s < t + \eta$, where the waiting time η is the exponential random variable obtained by evaluating the mth random field with the parameter value $u = \underline{u}^h(\underline{X}^h(s))$. If $u = 0$ then $\underline{X}^h(t+\eta) = x$, but otherwise it is conditionally distributed according to the jump probabilities

$$p^h(x, y|u) = \begin{cases} \frac{u_i^\pm}{\|u\|_1} & \text{if } y = x \pm he_i \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

It is easy to verify that the mean velocity of $\underline{X}^h$ at time t conditioned on $\underline{X}^h(t) = x$ is equal to $\underline{u}^h(x)$, so this is a consistent approximation to the limit dynamics in (2.3) if $\underline{u}^0(x)$ is replaced by $\underline{u}^h(x)$ there.

We now formulate the discrete approximation to the optimal control problem discussed above. Define the value function

$$V^h(x) = \inf_{\underline{u}^h} E_x \int_0^{\underline{\tau}^h} L(\underline{X}^h(t), \underline{u}^h(\underline{X}^h(t)))dt,$$

where the exit time is $\underline{\tau}^h = \inf[t : \underline{X}^h(t) \notin G^h]$, and the infimum is over feedback controls $\underline{u}^h$. Using standard methods [27, Section 4.3] it can be shown that V^h is the unique solution on G^h to the DPE

$$\inf_u \left[\mathcal{L}_u^h V^h(x) + L(x, u)\right] = 0, \tag{2.10}$$

with zero boundary condition on $h\mathbb{Z}^n/G^h$. It is straightforward to verify that (2.10) is equivalent to

$$V^h(x) = \inf_u \left[\sum_{y \in \mathbb{R}^n} p^h(x, y|u)V^h(y) + \overline{\Delta t}^h(u)L(x, u)\right], \tag{2.11}$$

and that the minimizing values of u are the same for these two equations. As suggested by the form of (2.11), the fixed point and an optimal feedback control can be found numerically using either Jacobi or Gauss-Seidel iteration schemes. We note that equation (2.11) is the DPE for a different approximating control problem, where instead of the continuous time Markov chain described above, a discrete time Markov chain is used to approximate the deterministic dynamics. That is the approach taken in [4], where the time step $\overline{\Delta t}^h(u)$ is used to interpolate the Markov chain into continuous time. As discussed in [4], the choice of one-sided transition probabilities and of a control-dependent time step facilitates rapid convergence of the iterative schemes used to solve (2.11), and the required infima at each step can be evaluated analytically.

The DPE (2.10) gives rise to an approximate value function $V^h(x)$ on G^h as well as an approximate optimal feedback control $u^h(x)$ on G^h. Typical convergence results for numerical methods dealing with optimal control problems concern the convergence of the value functions $V^h(x)$ to the limit value $V^0(x)$. The following theorem is proved for the present problem in [4, Theorem 5.4].

Theorem 2.3 *For any $\varepsilon > 0$, there exists $h_0 > 0$ such that*

$$|V^h(x) - V^0(x)| < \varepsilon$$

for all $0 < h \leq h_0$ and all $x \in G^h$.

For applications it is often important to have a good numerical approximation to the optimal feedback control $u^0(x)$. Typically, this quantity is not uniquely defined on the entire domain, and until recently no general approximation theorems were available. The following theorem [12, Corollary 5.6] establishes convergence of the approximate optimal controls $u^h(x)$ to $u^0(x)$ on the regions of strong regularity for the present problem. Aside from its intrinsic value, this result plays a pivotal role in the detailed asymptotic analysis that we carry out in the this paper.

Theorem 2.4 *For any $\varepsilon > 0$, there exists $h_0 > 0$ such that*

$$\|u^h(x) - u^0(x)\| < \varepsilon$$

for all $0 < h \leq h_0$ and for all $x \in B_0^h$.

It is convenient to have $u^0(x)$ defined and Lipschitz on all of $\mathbb{R}^n$ and to have $u^h(x)$ defined on all of $h\mathbb{Z}^n$. We abuse notation by extending $u^0(x)$ to $\mathbb{R}^n$ and changing its values on the complement of $\bar{B}_0$. Let $\delta > 0$ be such that $\delta < d(B_0, \partial Q \cap G)$ and such that $\delta \leq \tilde{\delta}$, where $\tilde{\delta}$ is as in Assumption 2.2; see Figure 2. We define a Lipschitz function $\tilde{u}^0(x)$ on $B_\delta(\bar{B}_0)$ by setting $\tilde{u}^0(x) = u^0(x)$ on $B_\delta(\bar{B}_0) \cap G$, and by extending it to $B_\delta(\bar{B}_0) \cap G^c$ as follows. For $x \in B_\delta(\bar{B}_0) \cap \partial G$ and for $0 \leq \gamma \leq \delta$, let $\tilde{u}^0(x + \gamma n) = u^0(x)$, where n is the outward normal vector at x; see Figure 2. Now let $\phi(x)$ be a C^∞ function on $\mathbb{R}^n$ taking values in $[0,1]$ such that $\phi(x) = 1$ on $B_{\delta/2}(\bar{B}_0)$ and $\phi(x) = 0$ outside of $B_\delta(\bar{B}_0)$. Such a function can be constructed by standard methods using a smooth convolution kernel [19, Theorem 0.17]. We can now redefine $u^0(x)$ to be equal to $\phi(x)\tilde{u}^0(x)$ on $B_\delta(\bar{B}_0)$ and zero everywhere else. This new $u^0(x)$ is Lipschitz on $\mathbb{R}^n$ and satisfies equation (2.7) on the region B_0. Finally, we put $u^h(x) = u^0(x)$ for all $x \in h\mathbb{Z}^n / B_0$. Notice that by Lemma 2.1 and by Theorem 2.4, we have $u^0(x)$ and $u^h(x)$ contained in the compact set U for all x at which they are defined, for $h > 0$ sufficiently small.

For most of what follows, we will be concerned only with initial conditions x in the region of strong regularity B_0. For $x \in B_0$, we define the optimal trajectory $X_x^0(t)$ for all $t \geq 0$ by applying the extended feedback control $u^0(x)$. Since B_0 is a region of strong regularity, $X_x^0(t)$ is optimally controlled until time τ_x^0, which is its first exit time from G and from the interior of B_0. We also define the exit location $z_x^0 = X_x^0(\tau_x^0)$. For $x \in B_0^h$, there is a unique process $X_x^h(t)$ defined for all $t \geq 0$ which is optimally controlled by $u^h(x)$ until it exits from B_0^h. We define the exit time $\tau_{x,B_0}^h = \inf[t : X_x^h(t) \notin B_0^h]$ and the exit location $z_{x,B_0}^h = X_x^h(\tau_{x,B_0}^h)$. We will often suppress the initial conditions in the subscripts of all of these notations.

The following lemma is proved just like [12, Lemma 2.3]. We use it to parlay information about convergence of trajectories into information about convergence of the corresponding exit times and locations.

Lemma 2.5 *Let M_2,M_1 be regions of strong regularity such that $\bar{M}_1 \subset M_2 \subset B_0$. For each sufficiently small $\varepsilon > 0$, there exists $\eta > 0$ such that the following holds. Let X be a trajectory with initial condition in M_2, and let τ_{M_2} and z_{M_2} be its exit time and location from the interior of M_2. If $x \in M_2$ is such that $|||\, X - X_x^0 \,|||_T \leq \eta$ holds, then $\tau_{M_2} \leq \tau_x^0 + \varepsilon$. If, in addition, $x \in M_1$, then it also follows that $|\tau_{M_2} - \tau_x^0| \leq \varepsilon$ and $\|z_{M_2} - z_x^0\| \leq \varepsilon$.*

Our proofs of the detailed asymptotic results in this paper depend upon having sharp estimates for the rate of convergence of the prelimit processes to the corresponding limit trajectories. We derive exponential rates of convergence in probability from the large deviations upper bound in Theorem B.1. For a region of strong regularity $M_2 \subset B_0$, consider a sequence of feedback control functions $\hat{u}^h(x)$ defined on $h\mathbb{Z}^n$ which satisfy $\hat{u}^h(x) = u^0(x)$ for $x \notin M_2^h$. Suppose that for each $\varepsilon > 0$, there exists $h_0 > 0$ such that

$$\|\hat{u}^h(x) - u^0(x)\| < \varepsilon,$$

for all $0 < h \leq h_0$ and for all $x \in M_2^h$. In particular, given Theorem 2.4, we can take for $\hat{u}^h(x)$ either $u^h(x)$ or $u^0(x)$. For $x \in M_2^h$, let $\hat{X}_x^h(t)$ be the process defined for $t \geq 0$ by applying the control $\hat{u}^h$ in the h-dynamics. Define $\hat{\tau}_{x,M_2}^h$ to be the exit time of $\hat{X}_x^h$ from M_2^h, and let $\hat{z}_{x,M_2}^h$ be its exit location.

Lemma 2.6 *Let M_2, M_1 be regions of strong regularity such that $\bar{M}_1 \subset M_2 \subset B_0$, and let $\hat{u}^h(x)$ and its corresponding trajectories be as above. For any $\varepsilon > 0$, there exists $K > 0$ such that*

$$(i) \qquad P_x\left[||| \hat{X}^h - X_x^0 |||_T \geq \varepsilon \right] < \frac{1}{K} e^{-K/h}$$

$$(ii) \qquad P_x\left[\hat{\tau}_{M_2}^h > \tau_x^0 + \varepsilon \right] < \frac{1}{K} e^{-K/h}$$

holds for for all $x \in M_2^h$ and for all sufficiently small $h > 0$. In addition,

$$(iii) \quad P_x\left[|\hat{\tau}_{M_2}^h - \tau_x^0| \geq \varepsilon \right] < \frac{1}{K} e^{-K/h}$$

$$(iv) \quad P_x\left[\|\hat{z}_{M_2}^h - z_x^0\| \geq \varepsilon \right] < \frac{1}{K} e^{-K/h}$$

holds for for all $x \in M_1^h$ and for all sufficiently small $h > 0$.

Proof. Part (i) is obtained by applying the large deviations upper bound in Theorem B.1 with $n_1 = n$, where we take $F = \{\phi : ||| \phi - X_x^0 |||_T \geq \varepsilon\}$, and we note that $I_x(\phi)$ is uniformly bounded away from zero for all $\phi \in F$. Parts (ii)-(iv) follow from part (i) by Lemma 2.5. ■

3 First Term in the Asymptotic Expansion

In this section, we derive a first order asymptotic expansion for $V^h(x)$ around $V^0(x)$ in the region of strong regularity B_1. Our methodology is essentially that employed in [11], and Theorem 2.4 plays a key role in the proof. The first order expansion is of independent interest, as it establishes a rate of convergence for $V^h(x)$ as an approximation to $V^0(x)$ in the regions of strong regularity. Our primary interest in this paper, however, is in the fact that the first order expansion suggests that it is possible to obtain higher order convergence through a modification of the Markov chain approximation. In the sections which follow, we prove, under some additional assumptions, that such schemes can be implemented to yield second order convergence.

For each $x \in B_0$ and for each $u \in U$, we define the difference between the prelimit and the limit generators applied to $V^0(x)$ by

$$\begin{aligned} r^h(x,u) &= \mathcal{L}_u^h V^0(x) - \mathcal{L}_u^0 V^0(x) \\ &= \frac{1}{2}\, h \left\langle |u|, D_\backslash^2 V^0(x) \right\rangle + O(h^2), \end{aligned} \tag{3.1}$$

where the $O(h^2)$ term is uniform on $B_0 \times U$. We can now write the DPE (2.4) for $V^0(x)$ in the alternative form

$$\inf_u \left[\mathcal{L}_u^h V^0(x) + L(x,u) - r^h(x,u)\right] = 0. \tag{3.2}$$

Equation (3.2) is the dynamic programming equation for an optimal control problem with Markov chain dynamics and with running cost equal to $L - r^h$. A comparison with the DPE for $V^h(x)$ in equation (2.10) suggests that the difference between $V^h(x)$ and $V^0(x)$ should be approximately equal to an integral of r^h along the optimal trajectory with initial condition x. The theorem below verifies that this intuition is correct, at least to a first order approximation. It is useful in what follows to define a compact notation for the running cost under a feedback control $u(x)$ by

$$L_u(x) = L(x, u(x)).$$

Similarly, we define

$$r_u^h(x) = r^h(x, u(x))$$

for all $x \in B_0$ and for all $h > 0$.

For $x \in B_0$, we define the error function

$$e^1(x) = \frac{1}{2}\int_0^{\tau_x^0} \left\langle \left|u^0(X_x^0(t))\right|,\ D^2 V^0(X_x^0(t)) \right\rangle dt, \tag{3.3}$$

and we note that the integrand is a scaled version of the first order approximation to r^h given in (3.1) above. Recall the region of strong regularity B_1, and the relationships specified in (2.5) and illustrated in Figure 1.

Theorem 3.1 *The asymptotic expansion*

$$V^h(x) = V^0(x) + he^1(x) + o(h) \tag{3.4}$$

holds as $h \to 0$, uniformly for $x \in B_1^h$.

Proof. We prove this theorem in two steps, first considering the upper bound on $V^h(x)$ and then the lower bound.

Upper Bound:

It is useful here to identify the suboptimal trajectories obtained by applying the limit optimal feedback control $u^0(x)$ in the h-dynamics. For an initial condition x in B_1^h, let $X^{h,0}$ be the process obtained by taking $\underline{u}^h = u^0$ in the Markov chain dynamics of Section 2 with parameter h. Define the exit time $\tau_{B_0}^{h,0} = \inf[t : X^{h,0}(t) \notin B_0^h]$ and the exit location $z_{B_0}^{h,0} = X^{h,0}(\tau_{B_0}^{h,0})$. Since the infimum in equation (3.2) is achieved at $u^0(x)$, we can use a standard verification argument to establish for all $x \in B_1^h$ the representation

$$V^0(x) = E_x\left[\int_0^{\tau_{B_0}^{h,0}} L_{u^0}(X^{h,0}) - r_{u^0}^h(X^{h,0})dt + V^0(z_{B_0}^{h,0})\right]. \tag{3.5}$$

We use part (ii) of Lemma 2.6 and the strong Markov property to obtain the uniform integrability of the $\tau_{B_0}^{h,0}$ needed for the right hand side of (3.5) to be finite. We obtain the following series of relations holding for x in B_1^h:

$$\begin{aligned} V^h(x) &\leq E_x\left[\int_0^{\tau_{B_0}^{h,0}} L_{u^0}(X^{h,0})dt + V^h(z_{B_0}^{h,0})\right] \\ &= V^0(x) + E_x\left[\int_0^{\tau_{B_0}^{h,0}} r_{u^0}^h(X^{h,0})dt\right] \\ &\quad + E_x\left[V^h(z_{B_0}^{h,0}) - V^0(z_{B_0}^{h,0})\right] \\ &= V^0(x) + \frac{1}{2}hE_x\left[\int_0^{\tau_{B_0}^{h,0}} \left\langle |u^0(X^{h,0})|, D_\backslash^2 V^0(X^{h,0})\right\rangle dt\right] \\ &\quad + E_x\left[V^h(z_{B_0}^{h,0}) - V^0(z_{B_0}^{h,0})\right] + O(h^2). \end{aligned} \tag{3.6}$$

The first line follows from the definition of $V^h(x)$ and from the strong Markov property; the second line is a consequence of the representation in (3.5); and the third line is obtained from the estimate in (3.1), where the $O(h^2)$ term is uniform on B_1^h.

Using parts (i) and (iii) of Lemma 2.6, it is straightforward to see that the integral term in the last line of (3.6) is equal to $he^1(x) + o(h)$, uniformly for x in B_1^h. It remains to estimate the boundary term in the last line of (3.6). Let $\varepsilon > 0$ be equal to $d(B_1, \partial B_0 \cap G)$; see Figure 3 in the next section. Since $z_x^0 \in B_{q+1} \cap \partial G$, it follows from the fact that $h > 0$ is chosen so that the lattice $h\mathbb{Z}^n$ lines up with $Q \cap \partial G$ that if $\|z_{B_0}^{h,0} - z_x^0\| < \varepsilon$ then $z_{B_0}^{h,0} \in \partial G$, and in that case

$$V^h(z_{B_0}^{h,0}) = V^0(z_{B_0}^{h,0}) = 0.$$

Since $V^0(x)$ and $V^h(x)$ are uniformly bounded, part (iv) of Lemma 2.6 implies that the boundary term in the last line of (3.6) is equal to $O(e^{-K/h})$, uniformly for x in B_1^h. Combining these estimates, we have

$$V^h(x) \leq V^0(x) + he^1(x) + o(h),$$

holding uniformly for x in B_1^h.

Lower Bound:

Similarly to (3.6) in the proof of the upper bound, we obtain the following series of relations holding for x in B_1^h:

$$\begin{aligned} V^0(x) &\le E_x\left[\int_0^{\tau_{B_0}^h} L_{u^h}(X^h) - r_{u^h}^h(X^h)dt + V^0(z_{B_0}^h)\right] \\ &= V^h(x) - E_x\left[\int_0^{\tau_{B_0}^h} r_{u^h}^h(X^h)dt\right] \\ &\quad + E_x\left[V^0(z_{B_0}^h) - V^h(z_{B_0}^h)\right] \\ &= V^h(x) - \frac{1}{2}hE_x\left[\int_0^{\tau_{B_0}^h} \left\langle |u^h(X^h)|, D_\backslash^2 V^0(X^h)\right\rangle dt\right] \\ &\quad + E_x\left[V^0(z_{B_0}^h) - V^h(z_{B_0}^h)\right] + O(h^2). \end{aligned} \tag{3.7}$$

The first line is a consequence of the fact that $u^h(x)$ is suboptimal in the control problem corresponding to equation (3.2) and of the strong Markov property; the second line follows from the definition of $V^h(x)$; and the third line is obtained from the estimate in (3.1), where the $O(h^2)$ term is uniform on B_1^h.

As in the proof of the upper bound, we estimate the terms in the last line of (3.7) to obtain

$$V^0(x) \le V^h(x) - he^1(x) + o(h),$$

holding uniformly for x in B_1^h. This time, we need to use the convergence of u^h to u^0 from Theorem 2.4 in the application of Lemma 2.6 and then again to show that the integrand in the last line of (3.7) converges to the integrand in the definition of $e^1(x)$. ■

Theorem 3.1 establishes the rate of convergence of $V^h(x)$ to $V^0(x)$ in the regions of strong regularity. It is not surprising that the convergence is first order, since the Markov chain approximation we consider gives rise to a discrete DPE (2.10) which could be obtained by replacing the derivatives in (2.4) with first order finite difference approximations. In fact, all numerical methods for problems of this type which have been proved to converge are intrinsically first order accurate.

What is promising about Theorem 3.1 is that equation (3.4) suggests that we can modify the present first order Markov chain method in order to obtain second order convergence in the regions of strong regularity. One approach is to approximate $e^1(x)$ and to subtract this approximation from $V^h(x)$. For this approach to be successful, two issues need to be addressed. First, we need to verify that the $o(h)$ error estimate in (3.4) can be sharpened to $O(h^2)$. Second, we need a first order approximation to $e^1(x)$. To get this we require first order estimates to the unknown quantities $u^0(x)$ and $D_\backslash^2 V^0(x)$. Both of these issues would be addressed if we could establish a higher order asymptotic expansion of the form

$$V^h(x) = V^0(x) + he^1(x) + h^2e^2(x) + O(h^3), \tag{3.8}$$

where both $e^1(x)$ and $e^2(x)$ are smooth functions of x. If (3.8) holds, then we can obtain first order approximations to the first and second derivatives of $V^0(x)$ by applying standard finite difference operators to $V^h(x)$. Since $u^0(x)$ is closely related to $DV^0(x)$, this information can be used to obtain a first order approximation to $u^0(x)$. Another approach to obtaining second convergence is to apply

Richardson extrapolation directly to the approximate value functions. If equation (3.8) is verified, then it is trivial that

$$2V^{h/2}(x) - V^h(x)$$

differs from $V^0(x)$ by an error which is $O(h^2)$, uniformly on the relevant region of strong regularity. In the sections which follow, we will specify conditions under which (3.8) can be verified, and then we will describe in detail how this information can be used to construct global numerical methods which provide second order approximation for both $V^0(x)$ and $u^0(x)$ in the regions of strong regularity.

4 Full Asymptotic Expansion

In this section, we establish an asymptotic expansion of $V^h(x)$ around $V^0(x)$ to order M, where M is an arbitrary positive integer. The results which follow require that we impose the following additional restriction on the region B_0.

Assumption 4.1 *There exists $\kappa > 0$ such that*

$$|u_i^0(x)| > \kappa$$

holds for each $i = 1, \ldots, n$ and for all $x \in B_0$.

Since $u^0(x)$ is continuous on B_0, Assumption 4.1 implies that each component of the optimal control $u^0(x)$ has a fixed sign in the region B_0. Thus, without loss of generality, we can assume that there exists $\kappa > 0$ such that

$$u_i^0(x) > \kappa,$$

for each $i = 1, \ldots, n$ and for all $x \in B_0$. Theorem 2.4 guarantees that for $h > 0$ sufficiently small, we also have $u_i^h(x) > 0$ holding for each $i = 1, \ldots, n$ and for all $x \in B_0^h$. Thus, as long as we restrict the analysis which follows to sufficiently small $h > 0$, we can deal exclusively with forward differences in the numerical approximation. As such, we redefine the generator for the h-dynamics to be such that

$$\mathcal{L}_u^h f = \langle u, D^h f \rangle$$

for any smooth function f mapping $\mathbb{R}^n$ to $\mathbb{R}$. The operator D^h is taken to be $D^{h,+}$, the forward finite difference approximation to the gradient defined in Section 1.

For convenience, we record here the DPE's for the limit problem and for the prelimit problem, with the generators fully written out. The limit DPE (2.4) takes the form

$$\begin{aligned} 0 &= \inf_u \left[\langle u, DV^0 \rangle + L \right] \\ &= -\frac{1}{2}\langle DV^0, aDV^0 \rangle + \langle b, DV^0 \rangle + c, \end{aligned} \tag{4.1}$$

and it holds for all $x \in B_0$, with the minimizer $u^0(x)$ given by (2.7). Similarly, the DPE (2.10) for the problem with h-dynamics can now be written in the the form

$$\begin{aligned} 0 &= \inf_u \left[\langle u, D^hV^h \rangle + L \right] \\ &= -\frac{1}{2}\langle D^hV^h, aD^hV^h \rangle + \langle b, D^hV^h \rangle + c, \end{aligned} \tag{4.2}$$

where the minimizer $u^h(x)$ can be explicitly evaluated

$$u^h(x) = -a(x)D^hV^h(x) + b(x), \tag{4.3}$$

for all $x \in B_0^h$ and for all $h > 0$ sufficiently small.

In light of the discussion at the end of Section 3, the condition in Assumption 4.1 is quite natural. In order for our approach to analyzing a second order numerical approximation to be successful, we require that both both $e^1(x)$ and $e^2(x)$ in the asymptotic expansion (3.8) be smooth functions of x. In general, this fails to be true even for $e^1(x)$, owing to the $|u^0|$ term in the integrand of (3.3). Under the present assumptions, however, we can write $e^1(x)$ in the form

$$e^1(x) = \frac{1}{2}\int_0^{\tau_x^0} \Big\langle\, u^0(X_x^0(t)),\ D_\backslash^2 V^0(X_x^0(t))\,\Big\rangle dt, \tag{4.4}$$

which implies that $e^1(x)$ is a smooth function of x, and it satisfies the equation

$$\langle u^0, De^1\rangle + \frac{1}{2}\langle u^0, D_\backslash^2 V^0\rangle = 0, \tag{4.5}$$

for all x in B_0, with zero boundary condition on $\partial G \cap B_0$. The increased regularity can be accounted for by the fact that the Hamiltonian in the DPE (4.2) is smooth, whereas the Hamiltonian in the DPE (2.10) fails to be a continuously differentiable function of $D^{h,\pm}V^h(x)$. We remark that even in those regions where Assumption 4.1 fails to hold, our methods seem to work quite well; see the examples in Section 6.

Before stating the main theorem of this section, we recall our standing assumptions. The domain $G \subset I\!\!R^n$ is open with compact closure and satisfies uniform interior and exterior cone conditions. We assume that b and c are C^∞ functions from $I\!\!R^n$ to $I\!\!R$ with $c(x) \geq c_0 > 0$ on G, and that a is a C^∞ function from $I\!\!R^n$ to the space of symmetric positive definite $n \times n$ matrices. Additionally, we assume that we have regions of strong regularity as specified by equations (2.5)-(2.6) such that the flat boundary condition of Assumption 2.2 and the condition on the optimal controls given in Assumption 4.1 both hold.

It is useful for what follows to introduce some notation. For $h > 0$, we define a new approximate feedback control

$$\bar{u}^h(x) = \frac{1}{2}(u^0(x) + u^h(x)). \tag{4.6}$$

Combining equations (4.1) and (4.2), and noting that

$$\bar{u}^h(x) = -\frac{1}{2}a(x)D^hV^h(x) - \frac{1}{2}DV^0(x) + b(x) \tag{4.7}$$

holds for all x in B_0^h, we conclude that

$$\Big\langle \bar{u}^h, D^h(V^h - V^0)\Big\rangle + \Big\langle \bar{u}^h, D^hV^0 - DV^0\Big\rangle = 0, \tag{4.8}$$

holds on the region B_0^h.

Theorem 4.2 *Recall the region B^* from equation (2.5). With $M \geq 1$ the arbitrary constant chosen in Section 2 and with the $e^m(x)$ as given below, the asymptotic expansion*

$$V^h(x) = V^0(x) + \sum_{m=1}^{M} h^m e^m(x) + o(h^M), \tag{4.9}$$

holds as $h \to 0$, uniformly for $x \in (B^*)^h$.

The following corollary establishes an analogous asymptotic expansion for the optimal feedback control. It is a direct consequence of Theorem 4.2 and of the expressions for $u^0(x)$ and $u^h(x)$ given in equations (2.7) and (4.3).

Corollary 4.3 *With $M \geq 1$ the arbitrary constant chosen in Section 2 and with the $e^m(x)$ as given below, the asymptotic expansion*

$$\begin{aligned} u^h(x) &= u^0(x) - a(x) \sum_{m=1}^{M-1} h^m D^h e^m(x) \\ &\quad - a(x) \sum_{m=1}^{M-1} h^m \frac{1}{(m+1)!} D_\backslash^{m+1} V^0(x) + o(h^{M-1}), \end{aligned} \tag{4.10}$$

holds as $h \to 0$, uniformly for $x \in (B^*)^h$.

In Section 3, we were able to determine the value of $e^1(x)$ by considering the formal difference between the generators $\mathcal{L}_u^0$ and $\mathcal{L}_u^h$, and then through the new dynamic programming equation (3.2), considering $V^0(x)$ as the value function for a modified optimal control problem with the h-dynamics. We cannot use this approach to determine the form of $e^m(x)$ for $m \geq 2$ because, in addition to the higher order terms in the expansion of $r^h(x)$, there are approximation effects which result from the difference between the controls $u^h(x)$ and $u^0(x)$. Instead, we use a formal recursive procedure, illustrated below for the case $m = 2$.

Assume that the expansion in (4.10) holds to first order, so it follows that

$$\bar{u}^h(x) = u^0(x) - \frac{1}{2} h a(x) De^1(x) - \frac{1}{4} h a(x) D_\backslash^2 V^0(x) + o(h). \tag{4.11}$$

We now combine equations (4.5) and (4.8) to obtain

$$\begin{aligned} &\left\langle \bar{u}^h, h^{-2} D^h [V^h - V^0 - he^1] \right\rangle \\ &+ h^{-2} \left\langle \bar{u}^h, D^h V^0 - DV^0 \right\rangle - \frac{1}{2} h^{-1} \left\langle u^0, D_\backslash^2 V^0 \right\rangle \\ &+ h^{-1} \left\langle \bar{u}^h - u^0, D^h e^1 \right\rangle + h^{-1} \left\langle u^0, D^h e^1 - De^1 \right\rangle = 0. \end{aligned}$$

Using the expression in (4.11), along with the fact that Taylor's theorem implies

$$D^h f(x) = \sum_{m=1}^{M} \frac{1}{m!} h^{m-1} D_\backslash^m f(x) + o(h^M) \tag{4.12}$$

for all smooth functions $f(x)$, we conclude that

$$\left\langle \bar{u}^h, h^{-2} D^h [V^h - V^0 - he^1] \right\rangle + r^2 + o(1) = 0, \tag{4.13}$$

where we define $r^2(x)$ by

$$\begin{aligned} r^2(x) &= \frac{1}{2} \left\langle u^0(x), D_\backslash^2 e^1(x) \right\rangle + \frac{1}{6} \left\langle u^0(x), D_\backslash^3 e^0(x) \right\rangle \\ &\quad - \frac{1}{2} \left\langle De^1(x), a(x) D_\backslash^2 e^0(x) \right\rangle - \frac{1}{2} \left\langle De^1(x), a(x) De^1(x) \right\rangle \\ &\quad - \frac{1}{8} \left\langle D_\backslash^2 e^0(x), a(x) D_\backslash^2 e^0(x) \right\rangle. \end{aligned}$$

Heuristically, equation (4.13) suggests that we should set

$$e^2(x) = \frac{1}{2}\int_0^{\tau_x^0} r^2(X_x^0(t))dt.$$

The formal argument outlined above will be made rigorous when we prove Theorem 4.2 later in this section.

In order to record the general form of $e^m(x)$ for $m = 1, \ldots, M$, we adopt the convention $e^0(x) = V^0(x)$ and define

$$\begin{aligned} r^m(x) \quad &= \quad \sum_{k=2}^{m+1} \frac{1}{k!}\Big\langle u^0(x)\ ,\ D_\backslash^k e^{m+1-k}(x)\Big\rangle \\ &\quad - \frac{1}{2}\sum_{l=2}^{m}\ \sum_{k=1}^{m+2-l}\ \frac{1}{k!l!}\Big\langle D_\backslash^l e^0(x)\ ,\ a(x)D_\backslash^k e^{m+2-l-k}(x)\Big\rangle \\ &\quad - \frac{1}{2}\sum_{j=1}^{m-1}\ \sum_{l=1}^{m-j}\ \sum_{k=1}^{m+2-j-l}\ \frac{1}{k!l!}\Big\langle D_\backslash^l e^j(x)\ ,\ a(x)D_\backslash^k e^{m+2-j-l-k}(x)\Big\rangle, \end{aligned}$$

for all x in B_0. Then recursively applying the heuristic outlined above suggests that we should define $e^m(x)$ by

$$e^m(x) = \frac{1}{2}\int_0^{\tau_x^0} r^m(X_x^0(t))dt, \tag{4.14}$$

for each $m = 1, \ldots, M$ and for all x in B_0. By the method of characteristics, we then have that

$$\langle u^0, De^m\rangle + r^m = 0, \tag{4.15}$$

holds for each $m = 1, \ldots, M$ and for all x in B_0, with zero boundary conditions on $\partial G \cap B_0$. Since each $r^m(x)$ depends upon the $e^i(x)$ with $i < m$, we define the $e^m(x)$ recursively beginning with $m = 1$. We note that the conclusion of Lemma 4.7 serves to verify that we have properly defined the $e^m(x)$ for the problem at hand.

An alternative heuristic method to derive equation (4.15) is to assume that (4.9) holds and that the error term $o(h^M)$ is a smooth function of x. We apply the operator D^h to both sides of equation (4.9) and then use (4.12) to express $D^hV^h(x)$ in terms of derivatives of smooth functions. We substitute this expression into the DPE (4.2), formally differentiate the resulting equation m times with respect to the parameter h, and then set $h = 0$ to obtain the equation for $e^m(x)$.

The proof of Theorem 4.2 can be broken down into three lemmas to be applied recursively. To make the notation consistent, we define a new region $B_{0'}$ to be equal to B_0. See equations (2.5) and (2.6) along with Figure 1 for the relationships between the regions of strong regularity in the following lemmas. Given an asymptotic expansion for $D^hV^h(x)$ to q terms, the first lemma establishes the expansion for $V^h(x)$ to $q+1$ terms. The needed expansion for $D^hV^h(x)$ with $q = 0$ has been established in Theorem 2.4. Although it is not needed for the development, we also note that in Section 3 we have already proved Lemma 4.4 for the case $q = 0$.

Lemma 4.4 *Let $0 \le q \le M-1$ be an integer, and suppose that*

$$D^hV^h(x) = D^hV^0(x) + \sum_{m=1}^{q} h^m D^h e^m(x) + o(h^q) \tag{4.16}$$

holds as $h \to 0$, uniformly for x in $B^h_{q'}$. Then

$$V^h(x) = V^0(x) + \sum_{m=1}^{q+1} h^m e^m(x) + o(h^{q+1}) \tag{4.17}$$

holds as $h \to 0$, uniformly for x in B^h_{q+1}.

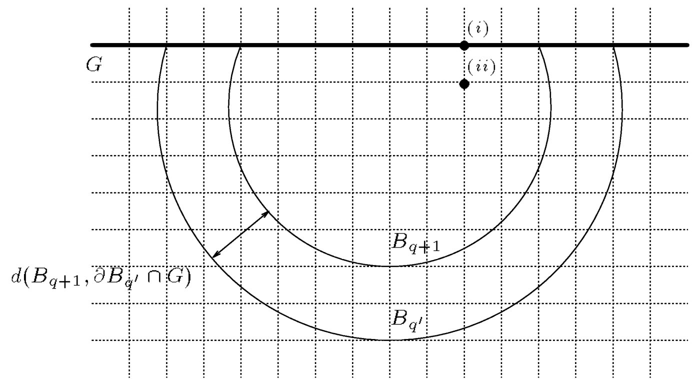


Figure 3: Boundary Points

The next lemma establishes the expansion for $D^hV^h(x)$ to $q+1$ terms, but only in a neighborhood of the boundary. The two cases in the lemma are illustrated in Figure 3.

Lemma 4.5 *Let $0 \le q \le M-1$ be an integer, and suppose that (4.16) holds. Then for each $p = 1, \ldots, n$,*

$$D^h_pV^h(x) = D^h_pV^0(x) + \sum_{m=1}^{q+1} h^m D^h_p e^m(x) + o(h^{q+1}) \tag{4.18}$$

holds as $h \to 0$, uniformly for $x \in \mathbb{R}^n$ such that either **(i)** *$x \in \partial G$ and $x + he_p \in B^h_{q+1}$, or* **(ii)** *$x \in B^h_{q+1}$ and $x + he_p \in \partial G$.*

Finally, given the conclusion of Lemma 4.5, the third lemma carries the asymptotic expansion for $D^hV^h(x)$ to $q+1$ terms on a smaller region of strong regularity.

Lemma 4.6 *Let $0 \le q \le M-1$ be an integer, and suppose that (4.16) and (4.18) hold. Then*

$$D^hV^h(x) = D^hV^0(x) + \sum_{m=1}^{q+1} h^m D^h e^m(x) + o(h^{q+1}) \tag{4.19}$$

holds as $h \to 0$, uniformly for x in $B^h_{(q+1)'}$.

Given the nesting of the regions of strong regularity, these three lemmas complete one step of a recursive process. We can apply Lemma 4.4 with q replaced by $q+1$, and iterate until we have established the full asymptotic series to order M. In preparation for the proofs of these lemmas, we introduce some more notation. For $q = 1, \ldots, M$ and for $x \in B_0$, we define

$$\Phi^{h,q}(x) = \frac{1}{h^q}\left[V^h(x) - V^0(x) - \sum_{m=1}^{q} h^m e^m(x)\right],$$

and for each $p = 1, \ldots, n$,

$$\begin{aligned} \Psi_p^{h,q}(x) &= D_p^h \Phi^{h,q}(x) \\ &= \frac{1}{h^q}\left[D_p^h V^h(x) - D_p^h V^0(x) - \sum_{m=1}^{q} h^m D_p^h e^m(x) \right], \end{aligned}$$

and then we put $\Psi^{h,q}(x) = (\Psi_1^{h,q}(x), \ldots, \Psi_n^{h,q}(x))$. The conclusions in the above lemmas can now be formulated in terms of the convergence as $h \to 0$ of the $\Phi^{h,q+1}(x)$ and the $\Psi^{h,q+1}(x)$ to zero.

The proofs of Lemmas 4.4 – 4.6 involve some elementary but rather lengthy algebraic calculations. We summarize the results of these calculations in the following lemma, the proof of which is deferred to Appendix A. It should be emphasized that, while the proof of Lemma 4.7 is elementary, it is in fact the key step which confirms that we have properly defined the $e^m(x)$ for the particular numerical approximation that we are studying. We recall our convention that subscripts refer to the components of a vector, while superscripts act as indices for possibly vector valued quantities.

Lemma 4.7 *Let $0 \leq q \leq M-1$ be an integer, and suppose that (4.16) holds. Then*

$$\langle \bar{u}^h, D^h \Phi^{h,q+1} \rangle + \phi^{h,q+1} = 0, \tag{4.20}$$

where $\phi^{h,q+1} = o(1)$ as $h \to 0$, uniformly for x in $B_{q'}^h$. Furthermore, for each $p = 1, \ldots, n$

$$\langle \hat{u}^{h,q+1,p}, D^h \Psi_p^{h,q+1} \rangle + \langle \lambda^{h,q+1,p}, \Psi^{h,q+1} \rangle + \psi_p^{h,q+1} = 0, \tag{4.21}$$

where $\hat{u}^{h,q+1,p} = u^0 + o(1)$, $\lambda^{h,q+1,p} = D_p u^0 + o(1)$, and $\psi_p^{h,q+1} = o(1)$ as $h \to 0$, uniformly for x in $B_{q'}^h$.

Proof of Lemma 4.4. The conclusion in (4.17) is equivalent to the statement that $\Phi^{h,q+1}(x)$ converges to zero as $h \to 0$, uniformly for x in B_{q+1}^h. For such x, let $\bar{X}_x^h(t)$ be the process defined for $t \geq 0$ by applying the control $\bar{u}^h$ in the h-dynamics. Define $\bar{\tau}_{x,B_{q'}}^h$ to be the exit time of $\bar{X}_x^h$ from $B_{q'}$ and let $\bar{z}_{x,B_{q'}}^h$ be its exit location. In light of (4.20), a standard verification argument shows that $\Phi^{h,q+1}(x)$ satisfies

$$\Phi^{h,q+1}(x) = E_x \left[\int_0^{\bar{\tau}_{B_{q'}}^h} \phi^{h,q+1}(\bar{X}^h) dt + \Phi^{h,q+1}(\bar{z}_{B_{q'}}^h) \right], \tag{4.22}$$

for each x in B_{q+1}^h. Part (ii) of Lemma 2.6 and the strong Markov property imply that the $\bar{\tau}_{B_{q'}}^h$ are uniformly integrable, which guarantees that the right hand side in the above expression is finite. We use that $\phi^{h,q+1}(x)$ converges to zero uniformly on $B_{q'}$ and that the $E_x[\bar{\tau}_{B_{q'}}^h]$ are uniformly bounded, along with the fact that $\Phi^{h,q+1}(x)$ is equal to zero on ∂G, to conclude that

$$\Phi^{h,q+1}(x) = o(1) + P_x\left[\bar{z}_{B_{q'}}^h \notin \partial G\right] O(h^{-(q+1)}),$$

holds uniformly for x in B_{q+1}^h. Let $\varepsilon > 0$ be equal to $d(B_{q+1}, \partial B_{q'} \cap G)$; see Figure 3. Since $z_x^0 \in B_{q+1} \cap \partial G$, it follows from the fact that $h > 0$ is chosen so that the lattice $h\mathbb{Z}^n$ lines up with $Q \cap \partial G$ that if $\|\bar{z}_{x,B_{q'}}^h - z_x^0\| < \varepsilon$ then $\bar{z}_{x,B_{q'}}^h \in \partial G$. Thus, we can apply the exponential bound from part (iv) of Lemma 2.6 to conclude that $\Phi^{h,q+1}(x)$ converges to zero as $h \to 0$, uniformly for x in B_{q+1}^h. That completes the proof of the lemma. ■

Proof of Lemma 4.5. For simplicity, we fix p and treat only case (ii), where $x \in B^h_{(q+1)'}$ and $x + he_p \in \partial G$. The proof for case (i) differs only in notation. Since V^h, V^0, and all of the e^m satisfy a zero boundary condition on ∂G, we have for each x satisfying condition (ii)

$$\begin{aligned}
&\frac{1}{h^{q+1}}\left| D^h_p V^h(x) - D^h_p V^0(x) - \sum_{m=1}^{q+1} h^m D^h_p e^m(x)\right| \\
&\quad = \frac{1}{h^{q+1}}\left[\frac{1}{h}|V^h(x) - V^0(x) - \sum_{m=1}^{q+1} h^m e^m(x)|\right] \\
&\quad = \frac{1}{h}|\Phi^{h,q+1}(x)|.
\end{aligned}$$

Thus, it suffices to prove that $\frac{1}{h}|\Phi^{h,q+1}(x)|$ converges to zero, uniformly for x satisfying condition (ii). Equation (4.22) implies that the relation

$$\frac{1}{h}|\Phi^{h,q+1}(x)| = \frac{1}{h}E_x\left[\bar\tau^h_{B_{q'}}\right] o(1) + P_x\left[\bar z^h_{B_{q'}} \notin \partial G\right] O(h^{-(q+2)}) \tag{4.23}$$

holds uniformly for x satisfying condition (ii). As in the proof of Lemma 4.4, the exponential bound from part (iv) of Lemma 2.6 implies that the second term in the right hand side of (4.23) converges to zero as $h \to 0$, uniformly for x satisfying condition (ii). Thus, it remains to show that $\frac{1}{h}E_x[\bar\tau^h_{B_{q'}}]$ is uniformly bounded, so that the first term in the right hand side of (4.23) also converges to zero. We observe that

$$E_x\left[\bar\tau^h_{B_{q'}}\right] \le E_x\left[\bar\tau^h_{B_{q'}}; A_1\right] + E_x\left[\bar\tau^h_{B_{q'}}; A_2\right] + E_x\left[\bar\tau^h_{B_{q'}}; A_3\right],$$

where we define the events

$$\begin{aligned}
A_1 &= \left[\bar X^h(t) + he_p \in \partial G\ ,\ 0 \le t \le \bar\tau^h_{B_{q'}}\right] \\
A_2 &= \left[\bar\tau^h_{B_{q'}} > T\right] \\
A_3 &= \left[\bar\tau^h_{B_{q'}} \le T\ ,\ \exists\, t \in [0, \bar\tau^h_{B_{q'}}) \text{ s.t. } \bar X^h(t) + he_p \notin \partial G\right].
\end{aligned}$$

Given the nontangential exit property for the limit optimal trajectories and the convergence of $\bar u^h$ to u^0, there exists $\eta > 0$ such that the following holds. For sufficiently small $h > 0$, if $x \in B^h_{q+1}$ and if $x + he_p \in \partial G$, the probability of exiting from G in one jump is at least equal to η; see Figure 3. Furthermore, since $\|\bar u^h\|_1$ is bounded away from zero, the conditional expected value of each inter-jump waiting time is bounded by δh for some fixed $\delta > 0$. Thus,

$$E_x\left[\bar\tau^h_{B_{q'}}; A_1\right] \le \delta h \eta \left(\sum_{i=0}^{\infty} i(1-\eta)^{i-1}\right) P_x\left[A_1\right] = O(h),$$

so $\frac{1}{h}E_x[\bar\tau^h_{B_{q'}}; A_1]$ is uniformly bounded for all x satisfying condition (ii). It follows from part (ii) of Lemma 2.6 and from the strong Markov property that $E_x[\bar\tau^h_{B_{q'}}|A_2]$ is uniformly bounded. Again applying part (ii) of Lemma 2.6, we obtain the uniform bound

$$P_x[A_2] = O(e^{-K/h}), \tag{4.24}$$

so that $\frac{1}{h}E_x[\bar{\tau}^h_{B_{q'}}; A_2]$ converges to zero uniformly for all x satisfying condition (ii). Since $E_x[\bar{\tau}^h_{B_{q'}}|A_3]$ is clearly bounded by T, we can complete the proof by obtaining a bound similar to the last display for the probability of the event A_3. On account of the fact that we use one-sided transition probabilities and given the nontangential exit property for the limit optimal trajectories, it follows from the fact that the boundary segment $B_{q'} \cap \partial G$ is parallel to one of the coordinate hyperplanes that $\bar{X}^h$ with initial condition x satisfying condition (ii) must move parallel to the boundary segment $B_{q'} \cap \partial G$ up to the time when it exits from $B_{q'}$; see Assumption 2.2 and Figure 3. Consequently, event A_2 can only occur if the exit location $\bar{z}^h_{B_{q'}}$ is not in ∂G. Thus, since $z^0_x \in B_{q+1} \cap G$, the fact that $d(B_{q+1}, \partial B_{q'} \cap G) > 0$ implies that we can obtain a bound for $P_x[A_3]$ analogous to (4.24) by applying part (iv) of Lemma 2.6. ■

Proof of Lemma 4.6. The conclusion in (4.19) is equivalent to the statement that $\Psi^{h,q+1}(x)$ converges to zero as $h \to 0$, uniformly for x in $B^h_{(q+1)'}$. Our plan is to proceed as in the proof of Lemma 4.4, first obtaining a representation for $\Psi^{h,q+1}(x)$ analogous to the representation for $\Phi^{h,q+1}(x)$ given in (4.22), and then using that representation to show that $\Psi^{h,q+1}(x)$ converges to zero. In this case, however, we cannot use an entirely standard representation because $\Psi^{h,q+1}(x)$ is a vector quantity, and the equations for its components are coupled. We develop the representation fairly carefully in order to highlight what we believe to be a novel aspect of our approach that may prove useful in developing high order numerical methods based directly upon the use of higher order finite difference approximations; see the discussion in Section 1. First, along the lines suggested in [20, Section 5.4], we expand the state space so that $(x,p) \in I\!R^n \times \{1,\ldots,n\}$ is the state variable, rather than just x. To that end, we recall the quantities defined in Lemma 4.7 and abuse notation by defining

$$\Psi^{h,q+1}(x,p) = \Psi^{h,q+1}_p(x), \qquad \psi^{h,q+1}(x,p) = \psi^{h,q+1,p}(x),$$

$$\hat{u}^{h,q+1}(x,p) = \hat{u}^{h,q+1,p}(x), \qquad \lambda^{h,q+1}(x,p) = \lambda^{h,q+1,p}(x),$$

for each (x,p), and we note that $\Psi^{h,q+1}$ and $\psi^{h,q+1}$ are scalar valued, while $\hat{u}^{h,q+1}$ and $\lambda^{h,q+1}$ are vector valued. We can now regard p as an argument rather as an index in equation (4.21) and obtain the scalar equation

$$\begin{aligned} &\left\langle \hat{u}^{h,q+1}(x,p), D^h\Psi^{h,q+1}(x,p)\right\rangle \\ &+\sum_{i=1}^n \lambda_i^{h,q+1}(x,p)\left[\Psi^{h,q+1}(x,i) - \Psi^{h,q+1}(x,p)\right] \\ &+\sum_{i=1}^n \lambda_i^{h,q+1}(x,p)\Psi^{h,q+1}(x,p) + \psi^{h,q+1}(x,p) = 0, \end{aligned} \tag{4.25}$$

holding for all $x \in B^h_{q+1}$ and for all $p = 1,\ldots,n$. If the $\lambda_i^{h,q+1}(x,p)$ were known to be non-negative, it would be possible to regard them as probabilities of jumping from (x,p) to (x,i), and then to deduce a representation for $\Psi^{h,q+1}(x,p)$ in terms of a Markov chain directly from (4.25). Since the signs of the $\lambda_i^{h,q+1}(x,p)$ are indefinite, however, we must further expand the state space in a novel way to eliminate the possibility of negative transition probabilities. We consider state variable $(x,p,\sigma) \in I\!R^n \times \{1,\ldots,n\} \times \{-1,1\}$ and again abuse notation by defining

$$\Psi^{h,q+1}(x,p,\sigma) = \sigma\Psi^{h,q+1}(x,p), \qquad \psi^{h,q+1}(x,p,\sigma) = \sigma\psi^{h,q+1}(x,p),$$

$$\lambda^{h,q+1}(x,p,\sigma) = \lambda^{h,q+1}(x,p),$$

for each (x,p,σ). Notice that $\lambda^{h,q+1}(x,p,\sigma)$ is actually independent of σ; we introduce the dependence only for convenience. In light of this notation, the linear structure of equation (4.25) implies that we can write

$$\begin{aligned}
&\left\langle \hat{u}^{h,q+1}(x,p), D^h\Psi^{h,q+1}(x,p,\sigma)\right\rangle \\
&+\sum_{i=1}^{n}\left(\lambda_i^{h,q+1}(x,p,\sigma)\right)^+\left[\Psi^{h,q+1}(x,i,\sigma)-\Psi^{h,q+1}(x,p,\sigma)\right] \\
&+\sum_{i=1}^{n}\left(\lambda_i^{h,q+1}(x,p,\sigma)\right)^-\left[\Psi^{h,q+1}(x,i,-\sigma)-\Psi^{h,q+1}(x,p,\sigma)\right] \\
&+\|\lambda^{h,q+1}(x,p,\sigma)\|_1\Psi^{h,q+1}(x,p,\sigma)+\psi^{h,q+1}(x,p,\sigma)=0,
\end{aligned} \tag{4.26}$$

for all $x \in B^h_{q+1}$, $p = 1,\ldots,n$, and $\sigma \in \{-1,1\}$. The left hand side of equation (4.26) indicates the action on $\Psi^{h,q+1}(x)$ of a legitimate generator with positive transition probabilities, so we can use it to derive a representation for $\Psi^{h,q+1}(x)$ in terms of a Markov chain $\Xi^{h,q+1}(t)$ taking values in $I\!R^n \times \{1,\ldots,n\} \times \{-1,1\}$. We construct the Markov chain $\Xi^{h,q+1}$ with exponentially distributed waiting times with mean

$$\overline{\Delta t}^{h,q+1}(x,p,\sigma) = \frac{h}{\Gamma^{h,q+1}(x,p,\sigma)},$$

where

$$\Gamma^{h,q+1}(x,p,\sigma) = \|\hat{u}^{h,q+1}(x,p)\|_1 + h\|\lambda^{h,q+1}(x,p,\sigma)\|_1,$$

and with transition probabilities given by

$$p^{h,q+1}(x,p,\sigma;\xi) = \begin{cases} \frac{\hat{u}_i^{h,q+1}(x,p)}{\Gamma^{h,q+1}(x,p,\sigma)} & \xi = (x+he_i,p,\sigma) \\[2ex] h\frac{\left(\lambda_i^{h,q+1}(x,p,\sigma)\right)^+}{\Gamma^{h,q+1}(x,p,\sigma)} & \xi = (x,i,\sigma) \\[2ex] h\frac{\left(\lambda_i^{h,q+1}(x,p,\sigma)\right)^-}{\Gamma^{h,q+1}(x,p,\sigma)} & \xi = (x,i,-\sigma) \\[2ex] 0 & \text{otherwise.} \end{cases}$$

For details on the construction of such a process, see the discussion in Section 2. Let $X^{h,q+1}$ be the first component of $\Xi^{h,q+1}$. For initial conditions (x,p,σ) such that x is in $B_{(q+1)'}$, let $\tau^{h,q+1}_{B_{q+1}}$ be the first exit time of $X^{h,q+1}$ from B^h_{q+1}, and let $z^{h,q+1}_{B_{q+1}}$ be equal to its exit location. Although the full Markov chain $\Xi^{h,q+1}$ does not converge to a deterministic limit, Theorem B.1 implies that a large deviations upper bound holds for the component process $X^{h,q+1}$. Combined with Lemma 2.5, the large deviations upper bound implies the following analogue of Lemma 2.6. For any $\varepsilon > 0$, there exists $K > 0$ such that

$$(i) \qquad P_{x,p,\sigma}\left[\|\!| X^{h,q+1} - X^0_x \|\!|_T \geq \varepsilon\right] < \frac{1}{K}e^{-K/h}$$

$$(ii) \qquad P_{x,p,\sigma}\left[\tau^{h,q+1}_{B_{q+1}} > \tau^0_x + \varepsilon\right] < \frac{1}{K}e^{-K/h}$$

holds for all initial conditions such that x is in B_{q+1}, while

$$(iii) \quad P_{x,p,\sigma}\left[|\tau^{h,q+1}_{B_{q+1}} - \tau^0_x| \geq \varepsilon\right] < \frac{1}{K}e^{-K/h}$$

$$(iv) \quad P_{x,p,\sigma}\left[\|z^{h,q+1}_{B_{q+1}} - z^0_x\| \geq \varepsilon\right] < \frac{1}{K}e^{-K/h}$$

holds for all initial conditions such that x is in $B_{(q+1)'}$. The representation for $\Psi^{h,q+1}(x)$ now takes the form

$$\begin{aligned} &\Psi^{h,q+1}(x,p,\sigma) \\ &= \left[E_{x,p,\sigma}\int_0^{\tau^{h,q+1}_{B_{q+1}}} e^{\int_0^t \|\lambda^{h,q+1}(\Xi^{h,q+1}(s))\|_1 ds}\psi^{h,q+1}(\Xi^{h,q+1}(t))dt \right. \\ &\quad \left. + \ \Psi^{h,q+1}(\Xi^{h,q+1}(\tau^{h,q+1}_{B_{q+1}}))\right], \end{aligned} \tag{4.27}$$

for all (x,p,σ) such that x is in $B^h_{(q+1)'}$. Property (ii) above and the strong Markov property guarantee that

$$P_{x,p,\sigma}\left[\tau^{h,q+1}_{B_{q+1}} \geq kT\right] \leq \left(\frac{1}{K}\right)^k e^{-kK/h},$$

holds for all initial conditions such that x is in $B^h_{(q+1)'}$ where $h > 0$ is sufficiently small and for all integers $k \geq 1$. Since the $\lambda^{h,q+1}$ are uniformly bounded, this is sufficient to guarantee that the right hand side of (4.27) is finite for sufficiently small $h > 0$, so that a standard verification argument can be used to establish that the indicated representation holds. In fact, since by Lemma 4.7 the $\psi^{h,q+1}(x,p,\sigma)$ converge uniformly to zero, this reasoning also implies that the first term on the right hand side converges to zero as $h \to 0$. Thus, we can complete the proof of the lemma by showing that the second term also converges to zero. Since $\Psi^{h,q+1}(x,p,\sigma)$ is equal to zero whenever $x \in \partial G$, the second term in (4.27) is bounded by

$$P_{x,p,\sigma}\left[z^{h,q+1}_{B_{q+1}} \notin \partial G\right] O(h^{-(q+2)}).$$

Let $\varepsilon > 0$ be equal to $d(B_{(q+1)'}, \partial B_{q+1} \cap G)$. As in the proof of Lemma 4.4, Assumption 2.2 implies that if $\|z^{h,q+1}_{B_{q+1}} - z^0_x\| < \varepsilon$ then $z^{h,q+1}_{B_{q+1}} \in \partial G$. Thus, we can apply the exponential bound from property (iv) above to conclude that the last display converges to zero as $h \to 0$, uniformly for all (x,p,σ) such that x is in B_{q+1}. This completes the proof that $\Psi(x,p,\sigma)$ converges uniformly to zero, and so establishes the lemma. ■

5 Second Order Numerical Approximations

The numerical method described in Section 2 yields approximations to the value function and to the optimal control which are first order accurate as $h \to 0$ on the regions of strong regularity. This notion is made precise in Theorem 3.1 and in the stronger results in Theorem 4.2 and Corollary 4.3 which hold under Assumption 4.1. In this section, we exploit the detailed asymptotic information from Section 4 to propose two different second order numerical methods. We prove in Theorem 5.1

that each of the modified numerical methods produces approximations to $V^0(x)$ and to $u^0(x)$ on the region B^* which are second order convergent as $h \to 0$. While these theoretical results apply only on regions of strong regularity where each component of the limit optimal control is bounded away from zero, in practice we do not know the locations of such regions. As such, our methods are of necessity applied on the entire domain. As the numerical examples in the next section illustrate, we obtain second order convergence even in large parts of the domain where the theoretical results do not apply.

Approximation Method I. For the purpose of defining a practical algorithm on the entire domain, we abuse notation by reverting to the optimal feedback control function $u^h(x)$ which is obtained on all of G^h by taking the maximizing argument in the DPE (2.10), before it is redefined on $h\mathbb{Z}^n/B_0$ in Section 2. Additionally, we consider versions of X^h, τ^h, and z^h which are defined using that control up to the exit time from G^h. When we prove the convergence properties on B^*, we will again employ the modified control and trajectories which were defined at the end of Section 2 and which were used in the analysis up to now.

The first step in the present algorithm is to obtain $V^h(x)$ and $u^h(x)$ by applying the first order method from Section 2. By Theorem 4.2, we have the asymptotic result

$$V^h(x) = V^0(x) + he^1(x) + h^2e^2(x) + O(h^3), \tag{5.1}$$

holding for all $x \in (B^*)^h$, where the $e^m(x)$ are as defined in equation (4.14). For each x in G^h, we define an approximation to $e^1(x)$ by

$$e^{1,h}(x) = \frac{1}{2}E_x\left[\int_0^{\tau^h} \Big\langle |u^h(X^h)|, D^{2,h}_{\setminus}V^h(X^h)\Big\rangle dt\right],$$

and we note that $e^{1,h}(x)$ can be obtained by applying an iterative method to solve the linear equation

$$\langle u^{h,+}, D^{h,+}e^{1,h}\rangle - \langle u^{h,-}, D^{h,-}e^{1,h}\rangle + \frac{1}{2}\langle |u^h|, D^{2,h}_{\setminus}V^h\rangle = 0, \tag{5.2}$$

with zero boundary condition on the complement of G^h. To implement an iterative solver, we express equation (5.2) in the form

$$e^{1,h}(x) = \sum_{y\in\mathbb{R}^n} p^h(x,y|u^h(x))e^{1,h}(y) + \overline{\Delta t}^h(u^h(x))\frac{1}{2}\langle |u^h(x)|, D^{2,h}_{\setminus}V^h(x)\rangle, \tag{5.3}$$

which is analogous to the expression for the discrete dynamic programming equation given in (2.11). Since the value function $V^h(x)$ is finite and since the running cost is bounded away from zero, every state in the optimally controlled Markov chain corresponding to equations (2.11) and (5.3) must communicate with the boundary. Thus, the right hand side is a contraction, and the equation is guaranteed to have a unique fixed point. We now define a new approximation to $V^0(x)$ by subtracting a correction term from $V^h(x)$

$$V^{h,*}(x) = V^h(x) - he^{1,h}(x), \tag{5.4}$$

and a new approximation to the optimal control by

$$u^{h,*}(x) = -a(x)D^{h,c}V^{h,*}(x) + b(x), \tag{5.5}$$

for all x in G^h, where $D^{h,c}$ is the centered difference approximation to the gradient operator. We will prove that $e^{1,h}(x)$ is a first order approximation to $e^1(x)$ on B^*, so it will follow directly from equation (5.1) that $V^{h,*}(x)$ is a second order approximation to $V^0(x)$. Additionally, a more detailed analysis of the asymptotic properties of $e^{1,h}(x)$ will be used to establish that $u^{h,*}(x)$ is a second order approximation to $u^0(x)$ on the region B^*.

We recall that the rate of convergence results that we will prove in Theorem 5.1 apply only on regions of strong regularity where each component of the limit optimal feedback control is bounded away from zero. A potential drawback of the method just described is that it may degrade the convergence properties of $V^h(x)$ and of $u^h(x)$ on those parts of the domain where second order convergence is not guaranteed. Our numerical experiments suggest that this is largely a theoretical concern, but nonetheless we mention here a possible approach to avoiding the problem. If we knew that the $e^{1,h}(x)$ were uniformly bounded, then it would follow from the definition of $V^{h,*}(x)$ that it has convergence properties at least as good as those proved for $V^h(x)$. That is, in addition to second order convergence on the regions of strong regularity which satisfy Assumption 4.1, the $V^{h,*}(x)$ would exhibit uniform convergence to $V^0(x)$ on the entire domain and first order convergence on the regions of strong regularity. In general, the $e^{1,h}(x)$ need not be uniformly bounded, but a sensible approach to guaranteeing this type of robustness would be to limit the norm of $D^{2,h}_{\setminus}V^h(x)$ in the definition of the $e^{1,h}(x)$. Similarly, if we could uniformly bound the $D^{h,c}e^{1,h}(x)$ then it would also follow that the $u^{h,*}(x)$ converge uniformly to the optimal feedback control on those parts of the regions of strong regularity where second order convergence is not assured. In any case, our numerical experiments have been done without any type of limiter, and we have not observed significant difficulties of this type.

Approximation Method II. Another approach to obtaining second order convergence on the regions of strong regularity is to apply Richardson extrapolation to the approximations $V^h(x)$ and $u^h(x)$ obtained from the first order method in Section 2. As in the description of Method I, we abuse notation by considering the optimal feedback control function $u^h(x)$ which is obtained on all of G^h by taking the maximizing argument in the DPE (2.10). For x in the grid $2h\mathbb{Z}^n$, we define a new approximation to the value function by

$$V^{h,*}(x) = 2V^h(x) - V^{2h}(x),$$

and a new approximate optimal feedback control by

$$u^{h,*}(x) = 2u^h(x) - u^{2h}(x).$$

It is a trivial consequence of Theorem 4.2 and of Corollary 4.3 that these are second order approximations to $V^0(x)$ and $u^0(x)$ on the region B^*.

An advantage of Richardson extrapolation over Method I is that there is no concern that the $V^{h,*}(x)$ or the $u^{h,*}(x)$ will exhibit qualitatively poorer convergence than their first order counterparts on those parts of the domain where Theorem 5.1 does not apply. This is a consequence of the fact that the new approximations are obtained by taking linear combinations of the first order approximations $V^h(x)$ and $u^h(x)$. This does not seem to be a practical problem even for Method I, and Method I has the advantages of producing somewhat smaller errors in our experiments and of being more straightforward to code because it involves only a single grid. Still, in circumstances where it is vital to maintain first order convergence outside of the regions where second order convergence is guaranteed, the present approach may be preferred.

Theorem 5.1 *Let $V^{h,*}(x)$ and $u^{h,*}(x)$ be obtained by either Method I or II above. Then the estimates*

$$V^{h,*}(x) = V^0(x) + O(h^2)$$

and

$$u^{h,*}(x) = u^0(x) + O(h^2),$$

hold uniformly for all $x \in (B^*)^h$.

For the purposes of proving this theorem, we return to using the feedback control $u^h(x)$ and the corresponding trajectories which were defined at the end of Section 2 and which were used for the analysis in the previous sections. Recall that $u^h(x)$ is defined for all x in $h\mathbb{Z}^n$ but is optimal only on the region B_0^h. Furthermore, Assumption 4.1 implies that each component of $u^h(x)$ is positive on the region B_0^h.

Proof of Theorem 5.1. As noted above, the conclusion of the theorem for the case of Method II is a trivial consequence of Theorem 4.2 and of Corollary 4.3. Thus, we proceed to prove the result for the case where $V^{h,*}(x)$ and $u^{h,*}(x)$ are obtained by Method I. We define the function

$$\Theta^h(x) = \frac{1}{h}\left[e^1(x) - e^{1,h}(x)\right]$$

for all x in $(B_0)^h$. Then it follows from equation (5.1) that

$$V^{h,*}(x) = V^0(x) + h^2\Theta^h(x) + h^2 e^2(x) + O(h^3), \tag{5.6}$$

so the estimate for $V^{h,*}(x)$ will be verified if we show that the $\Theta^h(x)$ are bounded uniformly on $(B^*)^h$. Define the function $\theta^h(x)$ by

$$\begin{aligned}\theta^h \quad &= \quad \Big\langle \frac{1}{h}(u^0 - u^h), D^h e^1 \Big\rangle + \Big\langle u^0, \frac{1}{h}(De^1 - D^h e^1) \Big\rangle \\ &\quad + \frac{1}{2}\Big\langle \frac{1}{h}(u^0 - u^h), D^{2,h}V^h \Big\rangle + \frac{1}{2}\Big\langle u^0, \frac{1}{h}(D^2V^0 - D^{2,h}V^h) \Big\rangle.\end{aligned}$$

Given the fact that all of the components of $u^h(x)$ are assumed to be positive, equations (4.5) and (5.2) imply that $\Theta^h(x)$ satisfies

$$\langle u^h, D^h\Theta^h \rangle + \theta^h = 0 \tag{5.7}$$

for all x in $(B^*)^h$, with a zero boundary condition on $\partial G \cap (B^*)^h$. So, we have the representation

$$\Theta^h(x) = E_x\left[\int_0^{\tau^h_{B^*}} \theta^h(X^h)dt + \Theta^h(z^h_{B^*})\right]. \tag{5.8}$$

The asymptotic expansions in Theorem 4.2 and in Corollary 4.3 imply that the $\theta^h(x)$ are uniformly bounded on $(B^*)^h$, so part (ii) of Lemma 2.6 implies that the integral term in (5.8) is bounded. Since the $\Theta^h(x)$ are uniformly bounded by $O(h^{-2})$ on B_0^h, we can apply the exponential estimate from part (iv) of Lemma 2.6 to bound the last term in (5.8). See the proof of Lemma 4.4, where details are given for a similar argument. We have shown that the $\Theta^h(x)$ are uniformly bounded on $(B^*)^h$, and this completes the proof of the first part of the theorem.

In light of the formula for $u^0(x)$ in equation (2.7) and the definition of $u^{h,*}(x)$, it is sufficient for the second part of the theorem to show that $D^{h,c}V^{h,*}(x)$ is a second order approximation to $D^{h,c}V^0(x)$, which is in turn a second order approximation to $DV^0(x)$. Since we do not know that $\Theta^h(x)$ satisfies a Lipschitz type bound, we cannot derive such an estimate directly from equation (5.6). Instead, we will derive a first order asymptotic expansion of the form

$$\Theta^h(x) = \Theta(x) + O(h), \tag{5.9}$$

where $\Theta(x)$ is a smooth function. We can then substitute this expression into (5.6) and apply the second order centered difference operator $D^{h,c}$ to both sides of that equation in order to verify that $D^{h,c}V^{h,*}(x)$ is a second order approximation to $D^{h,c}V^0(x)$.

We define the function

$$\Theta(x) = \int_0^{\tau_x^0} \theta(X_x^0)dt,$$

for all x in B_0, where $\theta(x)$ is given by

$$\begin{aligned}\theta &= \Big\langle \frac{1}{2}aD_\backslash^2 V^0 + aDe^1, De^1\Big\rangle + \Big\langle u^0, -\frac{1}{2}D_\backslash^2 e^1\Big\rangle \\ &\quad + \frac{1}{2}\Big\langle \frac{1}{2}aD_\backslash^2 V^0 + aDe^1, D^2V^0\Big\rangle + \frac{1}{2}\Big\langle u^0, -\frac{1}{2}D^2 e^1\Big\rangle,\end{aligned}$$

Thus, $\Theta(x)$ satisfies the equation

$$\langle u^0, D\Theta\rangle + \theta = 0 \tag{5.10}$$

for all x in B^*, with a zero boundary condition on $\partial G \cap B^*$. In light of Theorem 4.2 and Corollary 4.3, a term by term comparison of the definitions of $\theta^h(x)$ and $\theta(x)$ reveals that

$$\theta^h(x) = \theta(x) + O(h),$$

uniformly for x in $(B^*)^h$. We can now combine equations (5.7) and (5.10) to obtain the relation

$$\Big\langle u^h, D^h[\Theta^h - \Theta]\Big\rangle + O(h) = 0,$$

holding for all x in $(B^*)^h$, with a zero boundary condition on $\partial G \cap (B^*)^h$. Just as in the proof of the first part of the theorem, we can derive a representation for the quantity $[\Theta^h(x) - \Theta(x)]$ and use the exponential estimates from Lemma 2.6 to establish that it is equal to $O(h)$. This implies that equation (5.9) holds uniformly on $(B^*)^h$ and completes the proof of the theorem. ■

6 Examples

In this section, we present the results of experiments which illustrate the rates of convergence for the numerical methods described in this paper. For each example, we compute solutions using the first order method of Section 2 and then using the two second order methods described in Section 5. We solve the DPE (2.10), and for the second order Method I a linear version of that equation, by Gauss-Seidel iteration with a tolerance of 10^{-8}. Error values are indicated in the L^1 and L^∞ norms on the entire domain and on regions where the solution is known to be smooth. Approximate rates of convergence are determined by taking $\log_2(E_{m/2+1}/E_{m+1})$, where E_k is the error obtained using k gridpoints. In the case of the results obtained by Method II using Richardson extrapolation, the indicated number of gridpoints corresponds to the more refined of the two grids used in the calculation. In general, we observe second order or near second order convergence of the value functions in the L^1 norm on the entire domain and in the L^1 and L^∞ norms on the regions of strong regularity. In the case of the optimal controls, we generally observe second order convergence only in the L^1 norm on the regions of strong regularity. Results of this type are to be expected, since Assumption 4.1 is typically violated on some parts of the regions of strong regularity, so the theoretical rate of convergence results in Theorem 5.1 which would guarantee second order convergence in the L^∞ norm do not apply everywhere. In addition to the error values, we indicate the total number of iterative steps required to obtain each solution. As expected with Gauss-Seidel iteration, the number of iterations is essentially independent of the number of gridpoints. This last observation is very important in terms of the efficiency of practical calculations.

Example 1: One Dimensional Problem
We begin with a one dimensional example on the domain $[-1, 1]$. The running cost is taken to be

$$L(x,u) = (2 + 3x^2)^2 + \frac{1}{4}u^2,$$

and we can analytically evaluate the solution

$$V(x) = \begin{cases} 2x + x^3 + 3 & x \leq 0 \\ -2x - x^3 + 3 & x > 0. \end{cases}$$

Approximations to the value function and to the optimal control are shown in Figure 4. The value function is approximated quite well by the first order method and by both second order methods, with only a slight overshoot appearing at the singularity when we use Method I. This is an apparent manifestation of our observation in Section 5 that it is possible for this method to result in the loss of convergence at points of discontinuity. The discontinuity in the optimal control is resolved quite sharply, although sizable overshoots occur with both second order methods. We note, however, that these overshoots appear to remain bounded as the grid is refined, so they do not indicate a lack of numerical stability.

In Table 1, we indicate the errors in computing the value function using the three numerical schemes considered in this paper. We discuss the results for the optimal feedback control in the following paragraph. Errors are given for the entire domain and for those points which are at least a distance of 0.1 away from the singularity. Using the first order scheme, we obtain clear first order convergence of the value function in both the L^1 and the L^∞ norms on the entire domain and on the regions of strong regularity. Both second order schemes result in second order convergence of the value function in the L^1 norm on the entire domain as well as in the L^1 and L^∞ norms on the regions of strong regularity. Results of this quality are expected for a one dimensional problem, since the optimal control cannot change sign within a region of strong regularity, and Theorem 5.1 is valid on every such region. Somewhat surprisingly, Method II also yields second order convergence of the value function in the L^∞ norm on the entire domain.

The situation with the optimal feedback control is unusual due to the fact that the problem is one dimensional. Using the first order scheme, the approximate feedback controls on the regions of strong regularity are exactly (to machine accuracy) equal to the optimal feedback control for the limit problem. The following simple observation clarifies why this behavior is to be expected. Equations (4.1) and (4.2) imply for each x in the regions of strong regularity that

$$\frac{1}{2}\Big\langle DV^0(x), a(x)DV^0(x)\Big\rangle + \Big\langle b(x), DV^0(x)\Big\rangle + c(x) = 0$$

and

$$\frac{1}{2}\Big\langle D^hV^h(x), a(x)DV^h(x)\Big\rangle + \Big\langle b(x), DV^h(x)\Big\rangle + c(x) = 0.$$

In one dimension, these quadratic equations imply that there are only two possible values for $DV^0(x)$ and $D^hV^h(x)$. Since Theorem 2.4 guarantees that the $D^hV^h(x)$ converge to $DV^0(x)$, it must be that they are equal for sufficiently small $h > 0$. Given the relationships between the optimal feedback controls and $DV^0(x)$ and $D^hV^h(x)$, it follows that the approximate optimal feedback controls should be exactly correct. This observation applies only in one dimension, and our two dimensional examples exhibit more typical convergence of the optimal feedback controls. For the present problem, Method II inherits the exact convergence of the optimal feedback controls since it involves taking linear combinations of results obtained using the first order method, while Method I produces second order convergence in the L^1 and L^∞ norms in the regions of strong regularity.

Example 2: Perturbed Escape Time

Our next example is a two dimensional problem which is obtained by perturbing the escape time problem on the unit square. The escape time problem has the simple running cost $\tilde{L}(x,u) = \frac{1}{2}\|u\|^2 + 1/2$, and its value function is known analytically to be given by the shortest distance to any of the four edges of the square. Thus, the complement of the diagonals $x_1 = x_2$ and $x_1 = -x_2$ is a maximal region of strong regularity, and the value function is linear in each connected component of that region. As in reference [12], we modify the data for this problem to obtain one with smooth data and a more interesting solution that can still be determined analytically. To that end, we introduce the C^∞ double bump function defined by

$$\chi(\xi) = \begin{cases} e^{-\lambda((\xi-m)^2-\sigma^2)^{-2}} & \xi \in [m-\sigma, m+\sigma] \\ e^{-\lambda((-\xi-m)^2-\sigma^2)^{-2}} & \xi \in [-m-\sigma, -m+\sigma] \\ 0 & \text{otherwise} \end{cases}$$

where we use the parameter values $m = 0.7$, $\sigma = 0.5$, and $\lambda = 0.07$. Now we define a mollifier by

$$\Phi(x) = \chi(x_1 + x_2)\chi(x_1 - x_2)$$

for all $x = (x_1, x_2)$ in the unit square, and then define the value function $V^0(x)$ by multiplying the value function for the escape time problem by $1 + \Phi(x)$. The resulting function has the same regions of strong regularity as the escape time problem, and it maintains the simple structure in a neighborhood of the singularities. In a similar spirit, we define

$$a(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 3\sin(2\pi x_1)^2 \begin{bmatrix} 2 & 5 \\ 5 & 18 \end{bmatrix} \Phi(x)$$

and

$$b(x) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} x_1 \\ x_2 \sin((x_1^2 + x_2^2)^{1/2} - 1/2) \end{bmatrix} \Phi(x),$$

so that $a(x)$ is the identity and $b(x)$ is the zero vector in a neighborhood of the singularities. Now, we define $c(x)$ on the regions of strong regularity by

$$c(x) = (1/2)\Big\langle DV^0(x), a(x)DV^0(x)\Big\rangle - \Big\langle b(x), DV^0(x)\Big\rangle.$$

Our use of a mollifier in defining all of the relevant functions ensures that the cost function $c(x)$ extends smoothly to $c(x) = 1/2$ at the singularities, and it turns out that $V^0(x)$ solves the limit control problem for the indicated cost structure.

In Table 2, we indicate the errors in computing the value function and the optimal controls using the three numerical schemes considered in this paper. Errors are given for the entire domain and for those regular points which are at least a distance of 0.1 away from the singularities. The first order method consistently produces first order convergence of the value function and of the optimal controls, except when the control errors are measured in the L^∞ norm on the entire domain. This is expected, since the convergence results for the optimal controls do not apply at the singular points. The qualitative behaviors of the two second order methods are essentially the same, although the absolute error values for the feedback control are somewhat smaller using Method I. We obtain second order convergence of the value functions in the L^1 norm on the entire domain, while first order convergence is maintained in the L^∞ norm. In the regions of strong regularity, convergence in the L^1 norm is second order, and convergence in the L^∞ norm is significantly better than first order, but perhaps not second order. We note that the lack of clear second order convergence in the L^∞ norm is to be expected, since Assumption 4.1 does not hold everywhere in the regions of strong regularity for this problem.

In the case of the optimal controls, first order convergence is maintained in the L^1 norm on the entire domain, and we observe second order convergence in the L^1 norm in the regions of strong regularity. The errors on the regions of strong regularity measured in the L^∞ norm appear to decay only at a first order rate. This can be accounted for by the fact that the present problem does not satisfy Assumption 4.1 on the maximal region of strong regularity, as the optimal controls for the limit problem are not of fixed sign. By considering subsets of the regions of strong regularity where the controls do not change signs, we can obtain more convincing second order convergence in the L^∞ norm for both the value function and the optimal control. The results in Table 3 are obtained by restricting our attention to those parts of the domain where $x > 0.5$ and $y > 0$ and where $x < -0.5$ and $y < 0$. This is a somewhat arbitrary choice, but it singles out a set of points where the sign changes do not seem to interfere with second order convergence.

In Figure 5, we display the exact values of $V^0(x)$ and of the first component of $u^0(x)$, as well as approximations obtained by using 41 points in the two second order numerical methods. The graphs of the feedback controls illustrate that our methods resolve discontinuities quite sharply. It is also worth noting the slight overshoot in the center of the value function obtained by Method I. We observed in Section 5 that it is theoretically possible for convergence of the value function at singular points to be compromised when this second order method is employed. The overshoot is the only apparent manifestation of that possibility, and as predicted by the theory it does not appear when we compute the approximations using Method II.

Example 3: Quadratic Running Cost

Our final example is a two dimensional problem on the unit square where the boundaries of the maximal region of strong regularity are curved. The running cost is defined to be $L(x, u) = \frac{1}{2}\|u\|^2 + 6x_1^2 + 1$. We do not have an analytic expression for the solution, but by looking at approximate solutions, it is easy to visually identify the boundaries of the regions of strong regularity. In Figure 6a, we indicate those boundaries, as well as the subset of the domain which we utilize for the purpose of calculating rates of convergence on the regions of strong regularity, and in Figure 6b we show approximations to the characteristics.

In Table 4, we give relative errors for the the first order scheme and for the two second order schemes, both on the entire domain and on a subset of the regions of strong regularity. The relative errors are computed by comparing the solution computed using $m + 1$ gridpoints to the solution computed using $m/2 + 1$ gridpoints. These values are not as reliable as absolute errors for determining rates of convergence, but they do give a reasonable indication, and the results are consistent with our expectations. As in the previous example, both second order methods produce second order convergence of the value functions when measured in the L^1 norm on the entire domain and on the regions of strong regularity. When these errors are measured in the L^∞ norm on the regions of strong regularity, Method I produces second order convergence while the convergence with Method II is somewhat better than first order. In the case of the optimal controls, both methods apparently give rise to second order convergence in the L^1 norm on the regions of strong regularity, while preserving first order convergence in the L^1 norm on the entire domain and in the L^∞ norm on the regions of strong regularity. Again, since Assumption 4.1 does not appear to hold everywhere on the regions of strong regularity, results of this type are not surprising.

Figures 6c-6f show approximations with 41 points to the value function and to the first component of the optimal feedback control. There are no apparent qualitative differences in the two approximations to the value function. With Method I, the discontinuities in the control are resolved quite sharply. Method II seems to produce large overshoots, but we note that these are bounded as the grid is refined and do not indicate lack of numerical stability.

7 Conclusion

We have exhibited two global second order numerical methods for the solution of a class of Hamilton-Jacobi partial differential equations which are related to deterministic optimal control problems. Our analysis establishes that these methods produce provably second order approximations to the value function and to the optimal feedback control on certain subsets of the regions of strong regularity. The rate of convergence results apply only on those regions where each component of the limit optimal feedback control is bounded away from zero, so that the sign remains constant. In practice, we observe second order convergence in the L^1 norm on the regions of strong regularity even where this assumption is violated, but the results are less favorable when the L^∞ norm is used to measure the errors.

The reason that our theory can not accommodate changes in the signs of the optimal controls is that the transition probabilities given in (2.9) are singular as functions of the control u wherever any of the components of u are equal to zero. It is likely that we could eliminate the theoretical restrictions on the present method by applying our techniques to a first order method with smooth transition probabilities. In order for such an approach to be practical, however, it would be necessary that the new transition probabilities preserve the highly desirable qualitative properties of those considered here. The one sided nature of the present transition probabilities is important in terms of achieving sharp resolution of the discontinuities, and their simple form allows us to evaluate the minima analytically at each step in the iterative process used to solve the discrete dynamic programming equation. Additionally, the control dependent mean jump times in (2.8) ensure that the number of Gauss-Seidel iterations needed to solve the dynamic programming equation is small and essentially bounded as the grid is refined. We do not know of smooth transition probabilities which have these qualities, so it is not presently possible for us to improve upon the methods in this paper.

We conclude by suggesting other possible extensions to our work. In principle, the high order asymptotic analysis carried out in Section 4 can be used to formulate numerical methods of arbitrarily high order, either by using a more refined correction term in Method I or by taking the linear combination of several approximate solutions in Method II. As a practical matter, since even the second order convergence is not as consistent as we might have hoped in the typical situation where Assumption 4.1 is violated, we do not pursue this avenue. Our analysis considers only homogeneous boundary conditions, but this should not be an essential restriction. Homogeneity at the boundary is used quite strongly in our proof of Lemma 4.5 and in an analogous step in the proof of Theorem 2.4 [12], so it would be necessary to find an alternative approach to estimating the gradient at points near the boundary. Next, the quadratic structure of the running cost as a function of u is probably not needed for the type of asymptotic analysis that we carry out. In practice, this structure is essential for the efficient solution of the discrete dynamic programming equations as it enables us to evaluate the minima analytically, but it would also be interesting to construct higher order numerical methods for problems where the running cost does not have this form. Finally, we remark that it is worth considering the possibility of applying methods like ours to construct higher order numerical methods for the solution of the second order Hamilton-Jacobi partial differential equations which arise from stochastic control problems with dynamics given by controlled diffusions.

A Proof of Lemma 4.7

The purpose of this appendix is to indicate the calculations which are used to prove Lemma 4.7. The argument essentially consists of elementary algebraic manipulations, but they are rather involved, so it is worthwhile to set out some of the key steps. We combine equation (4.8) and equation (4.15) with $m = 1, \ldots, q+1$ to obtain the following relation holding on the region B_0^h:

$$
\begin{aligned}
&\left\langle \bar{u}^h, D^h \Phi^{h,q+1} \right\rangle \\
&\left.\begin{aligned}
&+ \sum_{i=0}^{q+1} h^{-(q+1-i)} \left\langle u^0, D^h e^i - D e^i \right\rangle \\
&- \sum_{m=1}^{q+1} \sum_{k=2}^{m+1} \frac{1}{k!} h^{-(q+1-m)} \left\langle u^0, D_{\backslash}^k e^{m+1-k} \right\rangle
\end{aligned}\right\}(a) \\
&\left.\begin{aligned}
&+ h^{-(q+1)} \left\langle \bar{u}^h - u^0, D^h e^0 - D e^0 \right\rangle \\
&+ \frac{1}{2} \sum_{m=1}^{q+1} \sum_{l=2}^{m} \sum_{k=1}^{m+2-l} \frac{1}{l!k!} h^{-(q+1-m)} \left\langle D_{\backslash}^l e^0, a D_{\backslash}^k e^{m+2-l-k} \right\rangle
\end{aligned}\right\}(b) \qquad (A.1) \\
&\left.\begin{aligned}
&+ \sum_{j=1}^{q+1} h^{-(q+1-j)} \left\langle \bar{u}^h - u^0, D^h e^j \right\rangle \\
&+ \frac{1}{2} \sum_{m=1}^{q+1} \sum_{j=1}^{m-1} \sum_{l=1}^{m-j} \sum_{k=1}^{m+2-j-l} \frac{1}{l!k!} h^{-(q+1-m)} \left\langle D_{\backslash}^l e^j, a D_{\backslash}^k e^{m+2-j-l-k} \right\rangle
\end{aligned}\right\}(c) \\
&= 0
\end{aligned}
$$

The first part of Lemma 4.7 will be established if we show that each of the expressions (a) – (c) converges to zero, uniformly for x in $B_{q'}^h$. A key observation is that we can use equation (4.16) along with (2.7) and (4.7) to obtain the relation

$$
\bar{u}^h - u^0 = -\frac{1}{2} a \sum_{k=2}^{q+1} \frac{1}{k!} h^{k-1} D_{\backslash}^k e^0 - \frac{1}{2} a \sum_{i=1}^{q} \sum_{k=1}^{q+1} \frac{1}{k!} h^{i+k-1} D_{\backslash}^k e^i + o(h^q), \qquad (A.2)
$$

holding unformly on $B_{q'}^h$. We indicate the details of the manipulations for (b), and note that the calculations are similar for (a) and (c). Consider the following series of relations which is easily seen to imply that (b) converges to zero, uniformly on $B_{q'}$. Each line is explained after the display.

$$
\begin{aligned}
&h^{-(q+1)} \left\langle \bar{u}^h - u^0, D^h e^0 - D e^0 \right\rangle \\
&= -\frac{1}{2} \sum_{l=2}^{q+1} \sum_{k=2}^{q+3-l} \frac{1}{l!k!} h^{-(q+3-k-l)} \left\langle a D_{\backslash}^k e^0, D_{\backslash}^l e^0 \right\rangle
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{2}\sum_{i=1}^{q}\sum_{l=2}^{q+2-i}\sum_{k=1}^{q+3-l-i}\frac{1}{l!k!}h^{-(q+3-k-l-i)}\Big\langle aD_\backslash^k e^i, D_\backslash^l e^0\Big\rangle + o(1) \\
= \quad & -\frac{1}{2}\sum_{l=2}^{q+1}\sum_{k=2}^{q+3-l}\frac{1}{l!k!}h^{-(q+3-k-l)}\Big\langle D_\backslash^l e^0, aD_\backslash^k e^0\Big\rangle \\
& -\frac{1}{2}\sum_{l=2}^{q+1}\sum_{k=1}^{q+2-l}\sum_{m=l+k-1}^{q+1}\frac{1}{l!k!}h^{-(q+1-m)}\Big\langle D_\backslash^l e^0, aD_\backslash^k e^{m+2-l-k}\Big\rangle + o(1) \\
= \quad & -\frac{1}{2}\sum_{l=2}^{q+1}\sum_{k=2}^{q+3-l}\sum_{m=l+k-2}^{m=l+k-2}\frac{1}{l!k!}h^{-(q+1-m)}\Big\langle D_\backslash^l e^0, aD_\backslash^k e^{m+2-l-k}\Big\rangle \\
& -\frac{1}{2}\sum_{l=2}^{q+1}\sum_{k=1}^{q+2-l}\sum_{m=l+k-1}^{q+1}\frac{1}{l!k!}h^{-(q+1-m)}\Big\langle D_\backslash^l e^0, aD_\backslash^k e^{m+2-l-k}\Big\rangle + o(1) \\
= \quad & -\frac{1}{2}\sum_{m=2}^{q+1}\sum_{l=2}^{m}\sum_{k=1}^{m+2-l}\frac{1}{l!k!}h^{-(q+1-m)}\Big\langle D_\backslash^l e^0, aD_\backslash^k e^{m+2-l-k}\Big\rangle + o(1)
\end{aligned}
$$

The first equality is obtained by applying expression (A.2) and by using the Taylor expansion for $e^0(x)$, as in equation (4.12); to get the second equality, we rearrange the orders of summation and make the change of variables $m = i + l + k - 2$; we get the third equality by inroducing m as a dummy variable in the first summation; the final equality is obtained by rearranging the orders of summation in each of the two terms in the previous line and then combining them into a single summation.

We now turn to verifying the second part of Lemma 4.7. It is easy to see that the following discrete product rule is valid for $p = 1, \ldots, n$:

$$
\begin{aligned}
D_p^h\Big\langle f(x), g(x)h(x)\Big\rangle \quad = \quad & \Big\langle D_p^h f(x), g(x+he_p)h(x+he_p)\Big\rangle \\
& +\Big\langle f(x), g(x+he_p)D_p^h h(x)\Big\rangle \\
& +\Big\langle f(x), D_p^h g(x)h(x)\Big\rangle.
\end{aligned}
$$

Our first application of this product rule to prove the second part of Lemma 4.7 actually involves the simpler case where $g(x)$ is the identity matrix. We state the rule for general $g(x)$ because it will be used later in dealing with the special case of proving the second part of Lemma 4.7 for $q = 0$. For each $p = 1, \ldots, n$, we apply the operator D_p^h to equation (A.1) and obtain

$$
\Big\langle \bar{u}^h(x+he_p), D^h\Psi_p^{h,q+1}(x)\Big\rangle + \Big\langle D_p^h\bar{u}^h(x), \Psi^{h,q+1}(x)\Big\rangle + \cdots = 0,
$$

where the unspecified quantities are those obtained by applying the product rule with $g(x)$ the identity matrix to the terms (a) – (c) in equation (A.1). This yields six new terms, and as in the treatment of (b) above, each of these can be expanded and rearranged to obtain coefficients converging to zero. In the case $q \geq 1$, the condition in (4.16) implies that $D_p^h\bar{u}^h$ converges to $D_p u^0$, so these manipulations suffice to establish the lemma.

Unfortunately, this general calculation is not sufficiently detailed to prove the second part of Lemma 4.7 when $q = 0$, so we treat that as a special case. We do not generalize the manipulations which work for $q = 0$, as they are considerably more involved than the ones indicated above. By using the product rule to apply the operator D_p^h directly to equations (4.1), (4.2), and (4.5) and then taking a linear combination of the resulting expressions, we obtain after some manipulations

$$\begin{aligned}
&\Big\langle - a(x+he_p)(\frac{1}{2}D^hV^h(x+he_p) + \frac{1}{2}D^hV^h(x)) \\
&\quad + b(x+he_p) - \frac{1}{2}ha(x+he_p)D_p^hDV^0(x)\ ,\ D^h\Psi_p^{h,1}(x)\Big\rangle \\
&+ \Big\langle - a(x+he_p)D_p^hDV^0(x) - D_p^ha(x)(\frac{1}{2}D^hV^h(x) + \frac{1}{2}DV^0(x)) \\
&\quad + D_p^hb(x)\ ,\ \Psi^{h,1}(x)\Big\rangle \\
&+ o(1) = 0,
\end{aligned}$$

holding uniformly for x in $B_{1'}^h$. Given the assumed convergence of the D^hV^h to DV^0 and the representation for u^0 in (2.7), it is easy to see that the above expression is consistent with (4.21), so it completes the proof of the lemma.

B Large Deviations Upper Bound

In this appendix, we state a general large deviations upper bound result for a broad class of Markov processes with possibly discontinuous statistics. Our result is essentially an extension of [10, Theorem 1.1]. The generalization is in two directions. First, we specify conditions under which the result holds for processes with statistics which, before the standard scalings, vary with the parameter ε but converge to limit values as $\varepsilon \to 0$. Second, and more interesting, we allow the situation where a large deviations upper bound holds only for some components of the processes under consideration; see reference [3] for results in a related direction.. A simple example illustrates these two points. Let $w_i(t), i = 1, 2$, be standard Brownian motions, and consider stochastic processes $X^\varepsilon = (X_1^\varepsilon, X_2^\varepsilon)$, where

$$\begin{cases} dX_1^\varepsilon = b^\varepsilon(X_2^\varepsilon)dt + \varepsilon^{1/2}dw_1 & X_1^\varepsilon(0) = 0 \\ dX_2^\varepsilon = dw_2 & X_2^\varepsilon(0) = 0, \end{cases}$$

and $b^\varepsilon(\cdot)$ converges uniformly to a constant b. While no large deviations upper bound holds for the full processes X^ε, the following result guarantees that the X_1^ε converge in the sense of a large deviations upper bound to the law of large numbers limit $\phi(t) = bt$.

We state carefully the new large deviations result that is used in the present paper, and then we outline the proof. The proof is similar to the one given for [10, Theorem 1.1], so we only indicate the key steps and the details where they differ from those given in [10]. Note that the notation in this appendix is independent of the notation in the preceding sections.

Let $\mathbb{R}^n = R^{n_1+n_2}$, and consider a sequence of Markov processes $X^\varepsilon = (X_1^\varepsilon, X_2^\varepsilon)$ with trajectories in $\mathcal{D}([0,\infty) : \mathbb{R}^n)$ and generators $\mathcal{L}^\varepsilon$ such that for any smooth function $f(x)$ mapping $\mathbb{R}^n$ to $\mathbb{R}$,

$$\begin{aligned}
\mathcal{L}^\varepsilon f(x) &= \langle b^\varepsilon(x), Df(x)\rangle + \frac{\varepsilon}{2}\mathrm{tr}\Big[a^\varepsilon(x)D^2f(x)\Big] \\
&\quad + \frac{1}{\varepsilon}\int_{\mathbb{R}^n}\Big[f(x+\varepsilon\nu) - f(x) - \varepsilon\langle \nu, Df(x)\rangle\Big]\mu^\varepsilon(x)(d\nu),
\end{aligned}$$

For each $\varepsilon > 0$, $a^\varepsilon(x)$ and $b^\varepsilon(x)$ are uniformly bounded functions from $I\!R^n$ to the spaces of $n \times n$ matrices and n-vectors, respectively, and $\mu^\varepsilon(x)$ is a function from $I\!R^n$ to the space of non-negative measures on $I\!R^n$ such that $\mu^\varepsilon(x)$ is uniformly bounded and has uniformly compact support for $x \in I\!R^n$. We consider block decompositions

$$a^\varepsilon(x) = \begin{bmatrix} a^\varepsilon_{11}(x) & a^\varepsilon_{12}(x) \\ a^\varepsilon_{21}(x) & a^\varepsilon_{22}(x) \end{bmatrix}, \qquad b^\varepsilon(x) = \begin{bmatrix} b^\varepsilon_1(x) \\ b^\varepsilon_2(x) \end{bmatrix},$$

where $a^\varepsilon_{11}(x)$ is an $n_1 \times n_1$ matrix valued function, $b^\varepsilon_1(x)$ is an n_1-vector valued function, and the other blocks are of appropriate sizes to complete the decompositions. Also, we let $\mu^\varepsilon_1(x)$ be the marginal measure of $\mu^\varepsilon(x)$ on $I\!R^{n_1}$. In general, we will employ without comment the notation $x = (x_1, x_2)$, where $x_i \in I\!R^{n_i}$ for $i = 1, 2$.

For the space of finite measures on $I\!R^{n_1}$, we define a metric $d(\cdot,\cdot)$ as follows. For two such measures η_1 and η_2, the distance $d(\eta_1, \eta_2)$ is defined to be the supremum of

$$\left| \int_{I\!R^{n_1}} f(\xi)\eta_1(d\xi) - \int_{I\!R^{n_1}} f(\xi)\eta_2(d\xi) \right|$$

over all functions $f(\xi)$ for which $\|f\|_\infty \le 1$ and which satisfy a Lipschitz condition with constant no greater than 1. We note that $d(\cdot,\cdot)$ metrizes weak convergence of probability measures [9, Proposition 11.3.2]. Assume the following:

1. The $a^\varepsilon_{11}(x)$ and $b^\varepsilon_1(x)$ are uniformly bounded for all $x \in I\!R^n$ and for all $\varepsilon > 0$.

2. There exists a compact set $K \subset I\!R^{n_1}$ and a constant $M < +\infty$ such that $\mu^\varepsilon_1(x)(K^c) = 0$ and $\mu^\varepsilon_1(x)(K) \le M$ for all $x \in I\!R^n$ and for all $\varepsilon > 0$.

3. There exist functions $a_{11}(x)$ and $b_1(x)$ such that

$$a^\varepsilon_{11}(x) \to a_{11}(x), \qquad b^\varepsilon_1(x) \to b_1(x).$$

   hold as $\varepsilon \to 0$, uniformly for x in $I\!R^n$ with x_1 in compact subsets of $I\!R^{n_1}$.

4. There exists a measure valued function $\mu_1(x)$ such that

$$d\Big(\mu^\varepsilon_1(x), \mu_1(x)\Big) \to 0.$$

   holds as $\varepsilon \to 0$, uniformly for x in $I\!R^n$ with x_1 in compact subsets of $I\!R^{n_1}$.

We now proceed to define the rate function for our large deviations upper bound. For each $x \in I\!R^n$ and for each vector $\alpha \in I\!R^{n_1}$, define the convex function

$$\begin{aligned} H(x,\alpha) &= \langle b_1(x), \alpha\rangle + \frac{1}{2}\mathrm{tr}\Big[a_{11}(x)\alpha\alpha^t\Big] \\ &\quad + \int_{I\!R^{n_1}} \Big[e^{\langle \nu,\alpha\rangle} - 1 - \langle \nu, \alpha\rangle\Big]\, \mu_1(x)(d\nu), \end{aligned}$$

and then for each $x_1 \in I\!R^{n_1}$

$$H_1(x_1, \alpha) = \sup_{x_2 \in I\!R^{n_2}} H((x_1, x_2), \alpha).$$

We further define the upper semicontinuous regularization,

$$h_1(x_1, \alpha) = \lim_{\delta \to 0} h_{1,\delta}(x_1, \alpha),$$

where

$$h_{1,\delta}(x_1,\alpha) = \sup_{\|y_1 - x_1\| \le \delta} H(y_1,\alpha).$$

Consider the Legendre-Fenchel transform given by

$$l(x_1,\beta) = \sup_{\alpha \in \mathbb{R}^{n_1}} \left[\langle \beta, \alpha\rangle - h_1(x_1,\alpha)\right]$$

for each $\beta \in \mathbb{R}^{n_1}$. Given $T < +\infty$ and for each $x \in \mathbb{R}^n$, we define the rate function $I_{x_1}(\phi_1)$ by

$$I_{x_1}(\phi_1) = \int_0^T l\Big(\phi_1(s), \dot{\phi}_1(s)\Big)\, ds$$

for absolutely continuous functions ϕ_1 taking values in $\mathbb{R}^{n_1}$ which satisfy $\phi_1(0) = x_1$, and we set $I_{x_1}(\phi_1)$ to be $+\infty$ otherwise.

Theorem B.1 *Given a compact set $C \subset \mathbb{R}^{n_1}$, the following hold:*

(i) Given $L < +\infty$ and for each $x_1 \in \mathbb{R}^{n_1}$, define

$$\Phi_{x_1}(L) = \{\phi_1 \in \mathcal{D}([0,T] : \mathbb{R}^{n_1}) : I_{x_1}(\phi_1) \le L\}.$$

Then the set $\bigcup_{x_1 \in C} \Phi_{x_1}(L)$ is compact.

(ii) For each closed set $F \subset \mathcal{D}([0,T] : \mathbb{R}^{n_1})$,

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_x\{X_1^\varepsilon \in F\} \le - \inf_{\phi_1 \in F} I_{x_1}(\phi_1),$$

holds uniformly for $x \in \mathbb{R}^n$ such that $x_1 \in C$.

As in [10], we refer to the corollary to Theorem 1 of Section 9.1.3 and to Theorem 3 of Section 9.1.4 of [24] for the proof of part (i) of Theorem B.1. For the sake of notational simplicity, we take $T = 1$ in what follows. In order to prove part (ii) of Theorem B.1, we consider the following approximations to the functions $H(x,\alpha)$, $H_1(x_1,\alpha)$, and $h_1(x,\alpha)$, corresponding to the generator of X^ε. For each $x \in \mathbb{R}^n$ and for each vector $\alpha \in \mathbb{R}^{n_1}$, define

$$\begin{aligned} H^\varepsilon(x,\alpha) &= \langle b_1^\varepsilon(x), \alpha\rangle + \frac{1}{2}\mathrm{tr}\Big[a_{11}^\varepsilon(x)\alpha\alpha^t\Big] \\ &\quad + \int_{\mathbb{R}^{n_1}} \Big[e^{\langle \nu, \alpha\rangle} - 1 - \langle \nu, \alpha\rangle\Big]\, \mu_1^\varepsilon(x)(d\nu) \end{aligned}$$

and for each $x_1 \in \mathbb{R}^{n_1}$ define

$$H_1^\varepsilon(x_1,\alpha) = \sup_{x_2 \in \mathbb{R}^{n_2}} H^\varepsilon((x_1,x_2),\alpha).$$

We further define the upper semicontinuous regularization,

$$h_1^\varepsilon(x_1,\alpha) = \lim_{\delta \to 0} h_{1,\delta}^\varepsilon(x_1,\alpha),$$

where

$$h_{1,\delta}^\varepsilon(x_1,\alpha) = \sup_{\|y_1 - x_1\| \le \delta} H_1^\varepsilon(y_1,\alpha).$$

The following lemma is a simple consequence of assumptions 1-4.

Lemma B.2 *The $H^\varepsilon(x,\alpha)$ are uniformly bounded for α in bounded sets. Furthermore, for α in bounded sets and for x in $I\!\!R^n$ such that x_1 is in compact subsets of $I\!\!R^{n_1}$, they converge uniformly to $H(x,\alpha)$ as $\varepsilon \to 0$.*

We also note that there exists a constant $C_1 < +\infty$ such that for all $x \in I\!\!R^n$, the bound

$$H^\varepsilon(x,\alpha) \le \bar{h}(\|\alpha\|) = C_1 e^{C_1\|\alpha\|} \tag{B.1}$$

holds for each $\alpha \in I\!\!R^{n_1}$. Now for $b > 0$, we define

$$\begin{aligned}\bar{l}(b) &= \sup_{a\in I\!\!R}\left[ab - \bar{h}(a)\right] \\ &= \frac{1}{C_1}\left(\log(b/C_1^2) - 1\right),\end{aligned}$$

and note that $\bar{l}(b)$ grows superlinearly as $b \to +\infty$. The following is a special case of [10, Lemma 2.2].

Lemma B.3 *For all $\alpha \in I\!\!R^{n_1}$, all $x \in I\!\!R^n$, and all $s \ge 0$,*

$$N_\alpha^\varepsilon(t) = \exp\frac{1}{\varepsilon}\left[\left\langle X_1^\varepsilon(t) - X_1^\varepsilon(s), \alpha\right\rangle - \int_s^t H^\varepsilon(X^\varepsilon(u),\alpha)du\right]$$

is a P_x-martingale for $t \ge s$.

As in [10], the bound in the next lemma follows from Lemma B.3 by an application of the bound (B.1) and a standard submartingale inequality.

Lemma B.4 *Given $\delta > 0$ and $s \ge 0$,*

$$P_x\left\{\sup_{s\le u\le t}\|X_1^\varepsilon(u) - X_1^\varepsilon(s)\| \ge \delta\right\} \le 2n\exp - \left[(t-s)\bar{l}(\delta/2n^{1/2}(t-s))/\varepsilon\right]$$

holds for all $x \in I\!\!R^n$, whenever $\delta > 2C_1^2 n^{1/2}(t-s)$.

Define the piecewise linear process

$$Y^\varepsilon(t) = \frac{X^\varepsilon(i\varepsilon)(i\varepsilon + \varepsilon - t) + X^\varepsilon(i\varepsilon+\varepsilon)(t - i\varepsilon)}{\varepsilon}$$

for $t \in [i\varepsilon, i\varepsilon + \varepsilon)$ and $i = 1, 2, \ldots, [1/\varepsilon]$, and consider the standard decomposition $Y^\varepsilon = (Y_1^\varepsilon, Y_2^\varepsilon)$. The next lemma establishes that Y_1^ε is exponentially close to X_1^ε, so it suffices to prove the upper bound in Theorem B.1 for Y_1^ε. This is convenient, as it is easier to prove an exponential tightness estimate (c.f. Lemma B.6) for Y_1^ε than for X_1^ε.

Lemma B.5 *Given $\delta > 0$ and uniformly for $x \in I\!\!R^n$,*

$$\limsup_{\varepsilon\to 0}\varepsilon\log P_x\left\{\sup_{0\le t\le 1}\|X_1^\varepsilon(t) - Y_1^\varepsilon(t)\| \ge \delta\right\} = -\infty.$$

Proof. As in the proof of [10, Lemma 2.4], apply Lemma B.4 on each of the intervals $[i\varepsilon, i\varepsilon + \varepsilon)$. ■

The following lemma establishes that the processes Y_1^ε are exponentially tight. Its proof is identical to the proof of [10, Lemma 2.5].

Lemma B.6 *Given a compact set $C \subset I\!R^n$ and a constant $B < +\infty$, there exists a compact set $K \subset \mathcal{C}([0,1] : I\!R^{n_1})$ such that*

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_x \{Y_1^\varepsilon \notin K\} \leq -B$$

holds for all $x \in C$.

We now state the key lemma for the proof of part (ii) of Theorem B.1. It is the analogue of [10, Lemma 2.7].

Lemma B.7 *Given $\delta > 0$ and a compact set $C \subset I\!R^{n_1}$,*

$$\limsup_{\varepsilon \to 0} \log E_x \exp \frac{1}{\varepsilon} \Big\langle Y_1^\varepsilon(\varepsilon) - Y_1^\varepsilon(0), \alpha \Big\rangle \leq h_{1,\delta}(x_1, \alpha)$$

holds uniformly for x such that $x_1 \in C$ and for α in bounded sets.

Proof. Just as in the proof of [10, Lemma 2.7], we obtain a constant $C_2 < +\infty$ such that

$$\begin{aligned} E_x \exp \frac{1}{\varepsilon} \Big\langle Y_1^\varepsilon(\varepsilon) - Y_1^\varepsilon(0), \alpha \Big\rangle &\leq e^{h_\delta^\varepsilon(x,\alpha)} + C_2 e^{-C_2/\varepsilon} \\ &\leq e^{h_{1,\delta}^\varepsilon(x_1,\alpha)} + C_2 e^{-C_2/\varepsilon}, \end{aligned}$$

holds for all x such that $x_1 \in C$, for all α in a fixed bounded set, and for all sufficiently small $\varepsilon > 0$. Lemma B.2 implies that the $h_{1,\delta}^\varepsilon(x_1, \alpha)$ are uniformly bounded below by a finite constant and that they converge uniformly to $h_{1,\delta}(x_1, \alpha)$ as $\varepsilon \to 0$. Thus, the result follows by taking the logarithm of each side and then taking limits. ■

Using the compactness result from part (i) of Theorem B.1, part (ii) of that theorem can now be proved just as in Proposition 2.9 and Proposition 2.10 in [10].

References

[1] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, Boston, 1997.

[2] M. Bardi and P. Soravia, *Hamilton-Jacobi equations with singular boundary conditions on a free boundary and applications to differential games*, Transactions of the American Mathematical Society, 325 (1991), pp. 205–229.

[3] C. Bezuidenhout, *A large deviations principle for small perturbations of random evolution equations*, Annals of Probability, 15 (1987), pp. 646–658.

[4] M. Boué and P. Dupuis, *Markov chain approximations for deterministic control problems with affine dynamics and quadratic cost in the control*, SIAM Journal on Numerical Analysis. To appear.

[5] P. Cannarsa, A. Mennucci, and C. Sinestrari, *Regularity results for solutions of a class of Hamilton-Jacobi equations*, Archive for Rational Mechanics and Analysis, 140 (1997), pp. 197–223.

[6] P. Cannarsa and C. Sinestrari, *Convexity properties of the minimum time function*, Calculus of Variations and Partial Differential Equations, 3 (1995), pp. 273–298.

[7] M. G. Crandall and P.-L. Lions, *Two approximations of solutions of Hamilton-Jacobi equations*, Mathematics of Computation, 43 (1984), pp. 1–19.

[8] M. H. A. Davis, *Piecewise deterministic markov processes: A general class of non-diffusion stochastic models*, Journal of the Royal Statistical Society, Series B, 46 (1984), pp. 353–388.

[9] R. M. Dudley, *Real Analysis and Probability*, Wadsworth and Brooks/ Cole, California, 1989.

[10] P. Dupuis, R. S. Ellis, and A. Weiss, *Large deviations for Markov processes with discontinuous statistics I: General upper bounds*, Annals of Probability, 19 (1991), pp. 1280–1297.

[11] P. Dupuis and M. James, *Rates of convergence for approximation schemes in optimal control*, SIAM Journal on Control and Optimization, 36 (1998), pp. 719–741.

[12] P. Dupuis and A. Szpiro, *Convergence if the optimal feedback policies in a numerical method for a class of deterministic optimal control problems.* Preprint.

[13] M. Falcone and R. Ferretti, *Discrete time high-order schemes for viscosity solutions of Hamilton-Jacobi-Bellman equations*, Numerische Mathematik, 67 (1994), pp. 315–344.

[14] ——, *Convergence analysis for a class of high-order semi-lagrangian advection schemes*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 909–940.

[15] W. Fleming, *The Cauchy problem for a nonlinear first order partial differential equation*, Journal of Differential Equation, 5 (1967), pp. 515–530.

[16] W. H. Fleming, *Stochastic control for small noise intensities*, SIAM Journal on Control and Optimization, 9 (1971), pp. 473–517.

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

[18] W. H. Fleming and P. E. Souganidis, *Asymptotic series and the method of vanishing viscosity*, Indiana University Mathematics Journal, 35 (1986), pp. 425–447.

[19] G. Folland, *Introduction to Partial Differential Equations*, Princeton University Press, New Jersey, 1976.

[20] M. Freidlin, *Functional Integration and Partial Differential Equations*, Princeton University Press, New Jersey, 1985.

[21] S. Gottlieb, *Convergence to Steady-State of Weighted ENO Schemes, Norm Preserving Runge-Kutta Methods and a Modified Conjugate Gradient Method*, PhD thesis, Brown University, Providence, 1998.

[22] A. Harten, *High resolution schemes for hyperbolic conservation laws*, Journal of Computational Physics, 49 (1983), pp. 357–393.

[23] C. Hu and C.-W. Shu, *A discontinuous Galerkin finite element method for Hamilton-Jacobi equations*. ICASE Report #98-2.

[24] A. D. Ioffe and V. M. Tikhomirov, *Theory of Extremal Problems. Studies in Mathematics and its Aplications*, North-Holland, Amsterdam, 1979.

[25] G. Jiang and D. Peng, *Weighted ENO schemes for Hamilton-Jacobi equation*, SIAM Journal on Scientific Computation. Submitted.

[26] H. J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977.

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

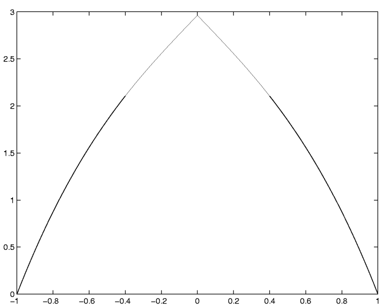
[28] R. J. Levesque, *Numerical Methods for Conservation Laws*, Birkhauser, Basel, 1992.

[29] C.-T. Lin and E. Tadmor, *High-resolution non-oscillatory central schemes for Hamilton-Jacobi equations*. UCLA CAM Report #98-38.

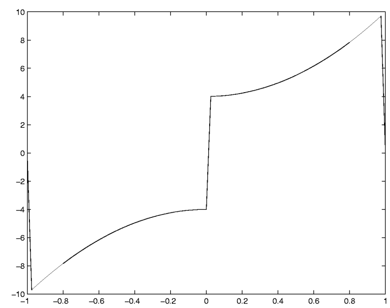
[30] P.-L. Lions and P. Souganidis, *Convergence of MUSCL and filtered schemes for scalar conservation law and Hamilton-Jacobi equations*, Numerische Mathematik, 69 (1995), pp. 441–470.

[31] S. Osher and C.-W. Shu, *High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 907–922.

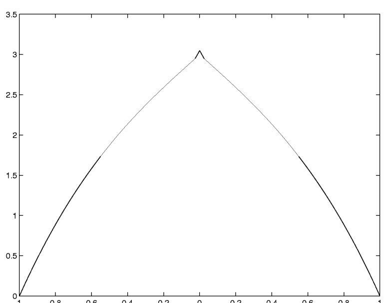
[32] A. Szpiro, *Asymptotic Analysis and High Order Markov Chain Based Numerical Methods for Optimal Control Problems and Related Hamilton-Jacobi Partial Differential Equations*, PhD thesis, Brown University, Providence, 1999.



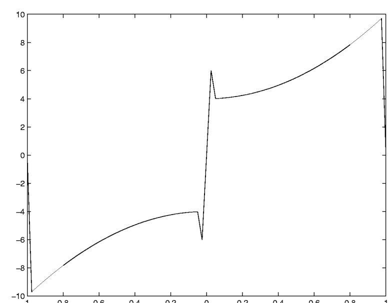
(a) Value Function
First Order



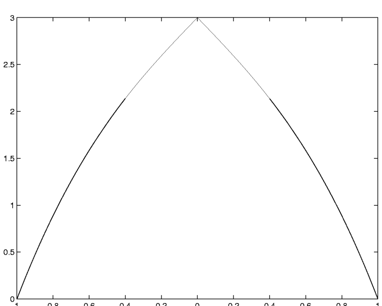
(b) Feedback Control
First Order



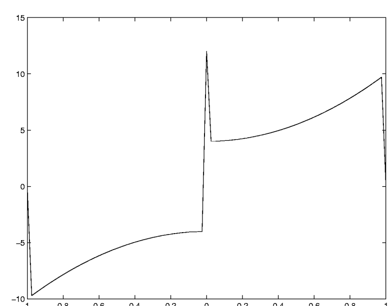
(c) Value Function
Method I



(d) Feedback Control
Method I

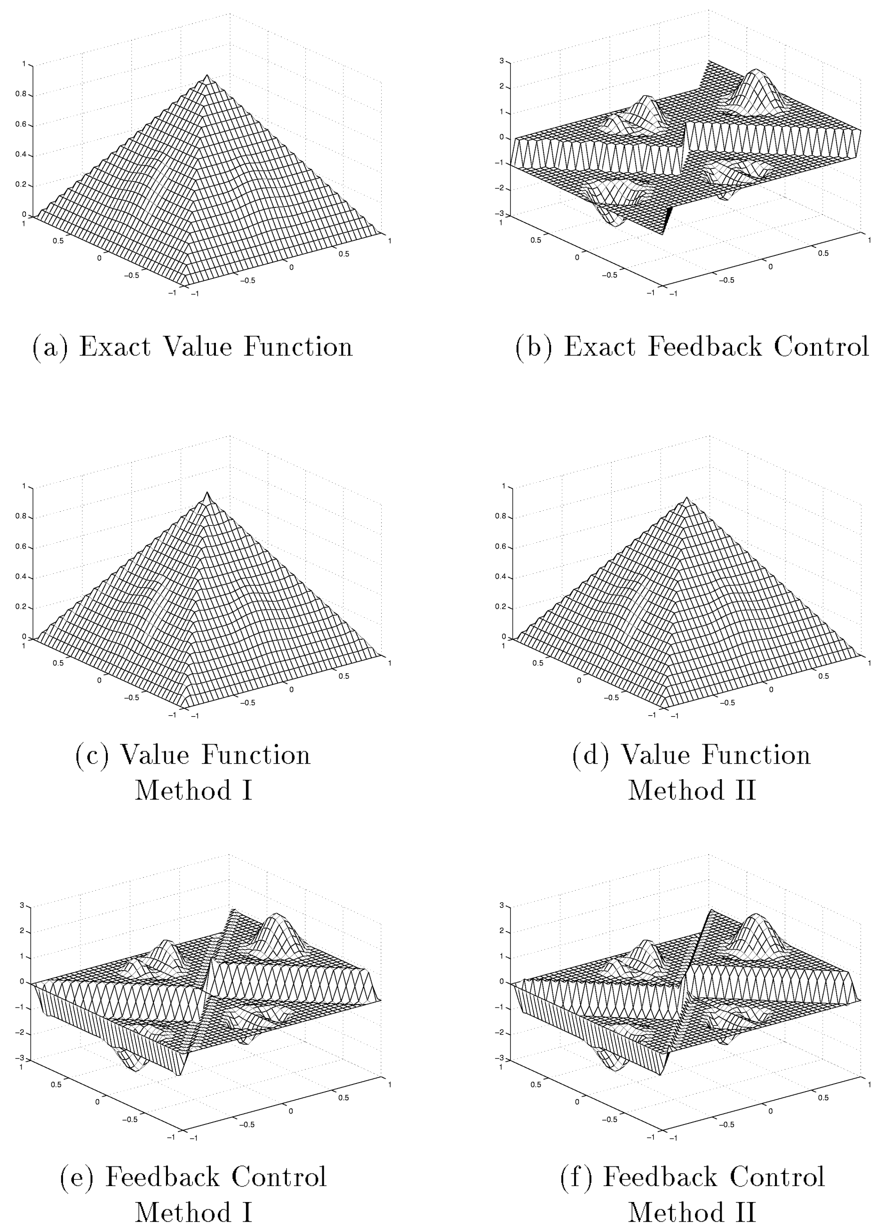


(e) Value Function
Method II



(f) Feedback Control
Method II

Figure 4: One Dimensional Problem Solutions



(a) Exact Value Function

(b) Exact Feedback Control

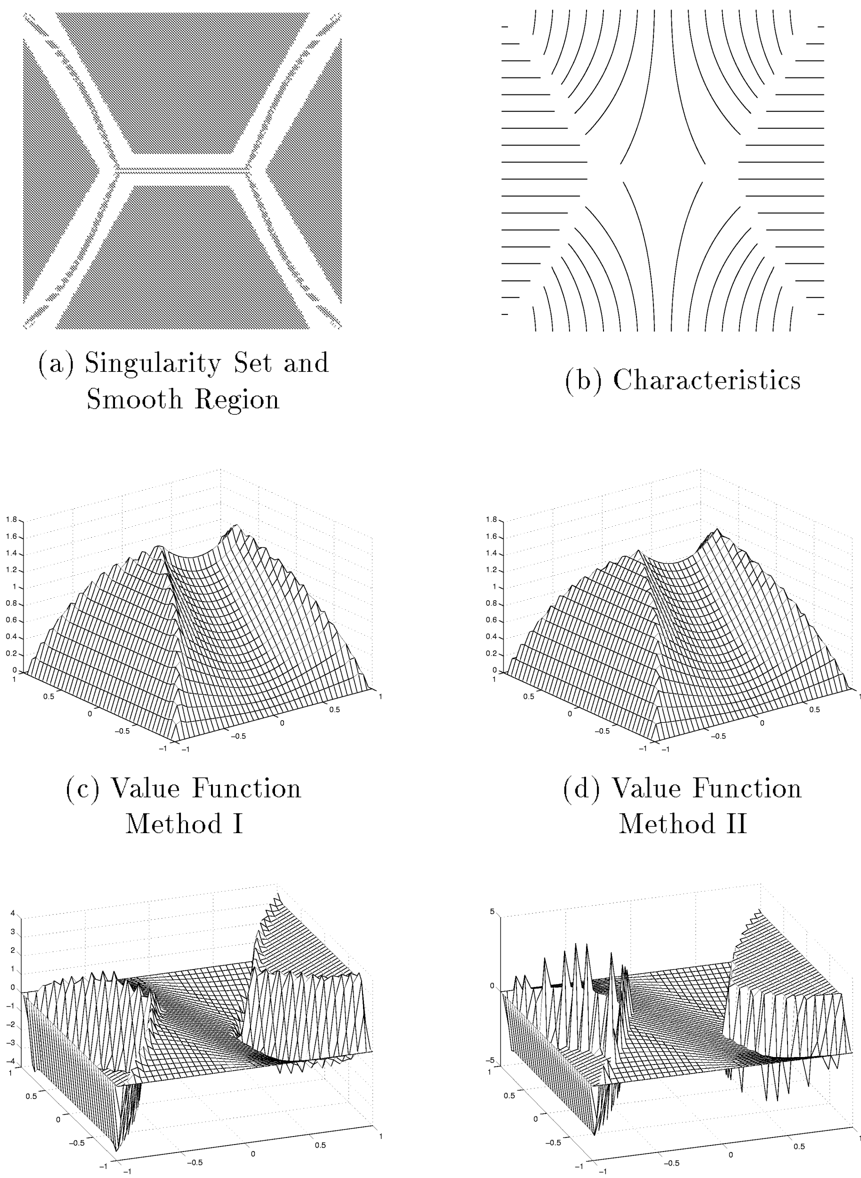
(c) Value Function
Method I

(d) Value Function
Method II

(e) Feedback Control
Method I

(f) Feedback Control
Method II

Figure 5: Perturbed Escape Time Problem Solutions



(a) Singularity Set and Smooth Region

(b) Characteristics

(c) Value Function Method I

(d) Value Function Method II

(e) Feedback Control Method I

(f) Feedback Control Method II

Figure 6: Quadratic Running Cost Problem Solutions

	Pts	Iter	L^1 Error	L^1 Ord	L^∞ Error	L^∞ Ord	L^1 RSR Error	L^1 RSR Ord	L^∞ RSR Error	L^∞ RSR Ord
1st Order	21	4	$1.95\text{ e}-01$	–	$1.45\text{ e}-01$	–	$1.66\text{ e}-01$	–	$1.44\text{ e}-01$	–
	41	4	$9.87\text{ e}-02$	1.0	$7.38\text{ e}-02$	1.0	$8.40\text{ e}-02$	1.0	$7.31\text{ e}-02$	1.0
	81	4	$4.97\text{ e}-02$	1.0	$3.72\text{ e}-02$	1.0	$4.23\text{ e}-02$	1.0	$3.68\text{ e}-02$	1.0
	161	4	$2.49\text{ e}-02$	1.0	$1.87\text{ e}-02$	1.0	$2.12\text{ e}-02$	1.0	$1.85\text{ e}-02$	1.0
	321	4	$1.25\text{ e}-02$	1.0	$9.36\text{ e}-03$	1.0	$1.06\text{ e}-02$	1.0	$9.26\text{ e}-03$	1.0
	641	4	$6.25\text{ e}-03$	1.0	$4.68\text{ e}-03$	1.0	$5.31\text{ e}-03$	1.0	$4.64\text{ e}-03$	1.0
2nd Order I	21	8	$4.02\text{ e}-02$	–	$1.77\text{ e}-01$	–	$2.03\text{ e}-02$	–	$2.25\text{ e}-02$	–
	41	8	$1.06\text{ e}-02$	1.9	$9.39\text{ e}-02$	0.9	$5.06\text{ e}-03$	2.0	$5.63\text{ e}-03$	2.0
	81	8	$2.73\text{ e}-03$	2.0	$4.85\text{ e}-02$	1.0	$1.27\text{ e}-03$	2.0	$1.41\text{ e}-03$	2.0
	161	8	$6.93\text{ e}-04$	2.0	$2.46\text{ e}-02$	1.0	$3.16\text{ e}-04$	2.0	$3.52\text{ e}-04$	2.0
	321	8	$1.75\text{ e}-04$	2.0	$1.24\text{ e}-02$	1.0	$7.91\text{ e}-05$	2.0	$8.79\text{ e}-05$	2.0
	641	8	$4.38\text{ e}-05$	2.0	$6.23\text{ e}-03$	1.0	$1.98\text{ e}-05$	2.0	$2.20\text{ e}-05$	2.0
2nd Order II	21	8	$1.00\text{ e}-02$	–	$1.00\text{ e}-02$	–	$8.00\text{ e}-03$	–	$8.00\text{ e}-03$	–
	41	8	$2.50\text{ e}-03$	2.0	$2.50\text{ e}-03$	2.0	$2.03\text{ e}-03$	2.0	$2.25\text{ e}-03$	1.8
	81	8	$6.25\text{ e}-04$	2.0	$6.25\text{ e}-04$	2.0	$5.06\text{ e}-04$	2.0	$5.63\text{ e}-04$	2.0
	161	8	$1.56\text{ e}-04$	2.0	$1.56\text{ e}-04$	2.0	$1.27\text{ e}-04$	2.0	$1.40\text{ e}-04$	2.0
	321	8	$3.91\text{ e}-05$	2.0	$3.91\text{ e}-05$	2.0	$3.16\text{ e}-05$	2.0	$3.52\text{ e}-05$	2.0
	641	8	$9.77\text{ e}-06$	2.0	$9.77\text{ e}-06$	2.0	$7.91\text{ e}-06$	2.0	$8.79\text{ e}-06$	2.0

Table 1: One Dimensional Problem Value Function Errors

		Pts	Iter	L^1 Error	L^1 Ord	L^∞ Error	L^∞ Ord	L^1 RSR Error	L^1 RSR Ord	L^∞ RSR Error	L^∞ RSR Ord
Value Function	1st Order	21	12	2.14 e − 02	–	3.80 e − 02	–	6.34 e − 03	–	1.90 e − 02	–
		41	12	7.46 e − 03	1.5	1.95 e − 02	1.0	3.27 e − 03	1.0	1.02 e − 02	0.9
		81	13	2.90 e − 03	1.4	1.01 e − 02	0.9	1.69 e − 03	1.0	5.28 e − 03	1.0
		161	16	1.24 e − 03	1.2	5.19 e − 03	1.0	8.65 e − 04	1.0	2.72 e − 03	1.0
		321	16	5.69 e − 04	1.1	2.64 e − 03	1.0	4.40 e − 04	1.0	1.38 e − 03	1.0
		641	16	2.71 e − 04	1.1	1.33 e − 03	1.0	2.21 e − 04	1.0	6.91 e − 04	1.0
	2nd Order I	21	24	2.71 e − 02	–	6.58 e − 02	–	7.93 e − 03	–	1.18 e − 02	–
		41	28	7.28 e − 03	1.9	3.31 e − 02	1.0	1.55 e − 03	2.4	3.16 e − 03	1.9
		81	36	1.92 e − 03	1.9	1.68 e − 02	1.0	4.04 e − 04	1.9	9.31 e − 04	1.8
		161	39	4.96 e − 04	2.0	8.52 e − 03	1.0	1.05 e − 04	1.9	2.63 e − 04	1.8
		321	35	1.26 e − 04	2.0	4.27 e − 03	1.0	2.68 e − 05	2.0	9.93 e − 05	1.4
		641	35	3.18 e − 05	2.0	2.12 e − 03	1.0	6.77 e − 06	2.0	2.79 e − 05	1.8
	2nd Order II	21	20	1.66 e − 02	–	1.76 e − 02	–	1.40 e − 02	–	1.76 e − 02	–
		41	24	3.80 e − 03	2.1	5.23 e − 03	1.8	8.04 e − 04	4.1	2.85 e − 03	2.6
		81	25	1.17 e − 03	1.7	2.79 e − 03	0.9	2.99 e − 04	1.4	1.01 e − 03	1.5
		161	29	3.27 e − 04	1.8	1.40 e − 03	1.0	9.08 e − 05	1.7	3.51 e − 04	1.5
		321	32	8.67 e − 05	1.9	8.01 e − 04	0.8	2.48 e − 05	1.9	1.06 e − 04	1.7
		641	32	2.24 e − 05	2.0	4.07 e − 04	1.0	6.61 e − 06	1.9	3.42 e − 05	1.6

		Pts	Iter	L^1 Error	L^1 Ord	L^∞ Error	L^∞ Ord	L^1 RSR Error	L^1 RSR Ord	L^∞ RSR Error	L^∞ RSR Ord
Feedback Control	1st Order	21	12	5.94 e − 01	–	1.86 e − 00	–	1.03 e − 01	–	2.66 e − 01	–
		41	12	3.24 e − 01	0.9	7.99 e − 01	0	5.39 e − 02	0.9	1.59 e − 01	0.7
		81	13	1.73 e − 01	0.9	1.86 e − 00	0	3.09 e − 02	0.8	8.94 e − 02	0.8
		161	16	8.94 e − 02	1.0	8.08 e − 01	0	1.67 e − 02	0.9	4.84 e − 02	0.9
		321	16	4.56 e − 02	1.0	8.13 e − 01	0	8.73 e − 03	0.9	2.54 e − 02	0.9
		641	16	2.30 e − 02	1.0	1.87 e − 00	0	4.45 e − 03	1.0	1.31 e − 02	1.0
	2nd Order I	21	24	4.89 e − 01	–	1.00 e − 00	–	1.73 e − 01	–	2.46 e − 01	–
		41	28	2.50 e − 01	1.0	1.00 e − 00	0	3.43 e − 02	2.3	9.58 e − 02	1.4
		81	36	1.24 e − 01	1.0	1.00 e − 00	0	8.38 e − 03	2.0	2.44 e − 02	2.0
		161	39	6.11 e − 02	1.0	1.00 e − 00	0	2.42 e − 03	1.8	1.22 e − 02	1.0
		321	35	3.03 e − 02	1.0	1.00 e − 00	0	6.86 e − 04	1.8	6.99 e − 03	0.8
		641	35	1.50 e − 02	1.0	1.00 e − 00	0	1.79 e − 04	1.9	3.85 e − 03	0.9
	2nd Order II	21	20	9.88 e − 01	–	3.79 e − 00	–	3.46 e − 01	–	3.38 e − 01	–
		41	24	5.07 e − 01	0.9	2.37 e − 00	0	5.72 e − 02	2.6	2.46 e − 02	3.8
		81	25	2.67 e − 01	1.0	3.80 e − 00	0	1.72 e − 02	1.7	1.20 e − 01	1.0
		161	29	1.29 e − 01	1.0	1.39 e − 00	0	5.12 e − 03	1.7	2.59 e − 02	2.2
		321	32	6.43 e − 02	1.0	8.09 e − 01	0	1.49 e − 03	1.8	3.03 e − 02	0
		641	32	3.21 e − 02	1.0	3.30 e − 00	0	4.09 e − 04	1.9	1.26 e − 02	1.3

Table 2: Perturbed Escape Time Problem Errors

	Value Function				Optimal Control			
	Method I		Method II		Method I		Method II	
Pts	L^∞ Error	Ord	L^∞ Error	Ord	L^∞ Error	Ord	L^∞ Error	Ord
21	7.87 e − 03	–	8.21 e − 03	–	1.49 e − 01	–	2.02 e − 01	–
41	2.27 e − 03	1.8	9.39 e − 04	3.1	3.27 e − 02	2.2	4.37 e − 02	2.2
81	6.45 e − 04	1.8	6.34 e − 04	0.6	1.24 e − 02	1.4	3.26 e − 02	0.4
161	1.74 e − 04	1.9	2.55 e − 04	1.3	3.49 e − 03	1.8	1.07 e − 02	1.6
321	4.55 e − 05	1.9	8.18 e − 05	1.6	9.97 e − 04	1.8	3.14 e − 03	1.8
641	1.16 e − 05	2.0	2.32 e − 05	1.8	2.65 e − 04	1.9	8.65 e − 04	1.9

Table 3: Perturbed Escape Time Problem Errors on Partial Domain

			L^1		L^∞		L^1 RSR		L^∞ RSR	
	Pts	Iter	Rel Error	Ord	Rel Error	Ord	Rel Error	Ord	Rel Error	Ord
Value Function, 1st Order	21	7	9.54 e − 02	–	9.73 e − 02	–	2.37 e − 02	–	6.05 e − 02	–
	41	7	4.14 e − 02	1.2	5.47 e − 02	0.8	1.47 e − 02	0.7	3.05 e − 02	1.0
	81	7	1.86 e − 02	1.2	2.90 e − 02	0.9	7.56 e − 03	1.0	1.70 e − 02	0.8
	161	7	8.85 e − 03	1.1	1.50 e − 02	1.0	3.81 e − 03	1.0	8.97 e − 03	0.9
	321	7	4.29 e − 03	1.0	7.67 e − 03	1.0	1.92 e − 03	1.0	4.49 e − 03	1.0
	641	7	2.11 e − 02	1.0	3.89 e − 03	1.0	9.67 e − 04	1.0	2.27 e − 03	1.0
Value Function, 2nd Order I	21	14	1.51 e − 01	–	2.23 e − 01	–	2.61 e − 02	–	7.47 e − 02	–
	41	14	4.92 e − 02	1.6	1.37 e − 01	0.7	2.59 e − 03	3.3	1.08 e − 02	2.8
	81	14	1.44 e − 02	1.8	7.64 e − 02	0.8	5.17 e − 04	2.3	9.36 e − 04	3.5
	161	14	3.78 e − 03	1.9	4.15 e − 02	0.9	1.43 e − 04	1.9	2.62 e − 04	1.8
	321	14	9.99 e − 04	1.9	2.23 e − 02	0.9	3.78 e − 05	1.9	7.10 e − 05	1.9
	641	14	2.50 e − 04	2.0	1.11 e − 02	1.0	9.71 e − 06	2.0	1.84 e − 05	1.9
Value Function, 2nd Order II	21	13	8.53 e − 02	–	6.14 e − 02	–	2.76 e − 02	–	4.29 e − 02	–
	41	14	2.63 e − 02	1.7	5.47 e − 02	0.2	1.01 e − 03	8.1	1.76 e − 03	7.9
	81	14	8.46 e − 03	1.6	4.09 e − 02	0.4	5.36 e − 04	0.9	1.26 e − 03	0.5
	161	14	1.79 e − 03	2.2	2.26 e − 02	0.9	1.76 e − 04	1.6	4.28 e − 04	1.6
	321	14	5.04 e − 04	1.8	9.51 e − 03	1.2	5.19 e − 05	1.8	1.24 e − 04	1.8
	641	14	1.27 e − 04	2.0	6.66 e − 03	0.5	1.41 e − 05	1.9	3.52 e − 05	1.8

			L^1		L^∞		L^1 RSR		L^∞ RSR	
	Pts	Iter	Rel Error	Ord	Rel Error	Ord	Rel Error	Ord	Rel Error	Ord
Feedback Control, 1st Order	21	7	1.20 e − 00	–	2.57 e − 00	–	1.35 e − 01	–	1.45 e − 01	–
	41	7	6.37 e − 01	0.9	3.29 e − 00	0	7.52 e − 02	0.8	8.69 e − 02	0.7
	81	7	3.28 e − 01	1.0	3.29 e − 00	0	3.98 e − 02	0.9	5.43 e − 02	0.7
	161	7	1.49 e − 01	1.1	3.45 e − 00	0	2.08 e − 02	0.9	3.03 e − 02	0.8
	321	7	8.31 e − 02	0.8	3.40 e − 00	0	1.06 e − 02	1.0	1.62 e − 02	0.9
	641	7	3.76 e − 02	1.1	3.57 e − 00	0	5.33 e − 03	1.0	8.41 e − 03	0.9
Feedback Control, 2nd Order I	21	14	1.52 e − 00	–	1.33 e − 00	–	6.53 e − 01	–	1.07 e − 00	–
	41	14	9.49 e − 01	0.7	1.33 e − 00	0	2.01 e − 01	1.7	1.18 e − 00	0
	81	14	5.10 e − 01	0.9	1.36 e − 00	0	4.35 e − 03	5.5	6.70 e − 02	4.1
	161	14	2.68 e − 01	0.9	1.35 e − 00	0	9.65 e − 04	2.2	2.03 e − 03	5.0
	321	14	1.36 e − 01	1.0	1.40 e − 00	0	2.58 e − 04	1.9	7.48 e − 04	1.4
	641	14	6.87 e − 02	1.0	1.39 e − 00	0	6.71 e − 05	1.9	3.68 e − 04	1.0
Feedback Control, 2nd Order II	21	13	3.79 e − 00	–	4.81 e − 00	–	6.31 e − 01	–	9.02 e − 01	–
	41	14	1.07 e − 00	1.8	2.48 e − 00	0	3.10 e − 02	1.0	1.04 e − 01	3.1
	81	14	7.54 e − 01	0.5	5.71 e − 00	0	6.03 e − 03	5.7	1.26 e − 02	3.0
	161	14	3.12 e − 01	1.3	3.27 e − 00	0	2.02 e − 03	1.6	5.53 e − 03	1.2
	321	14	2.14 e − 01	0.5	6.56 e − 00	0	6.11 e − 04	1.7	1.78 e − 03	1.6
	641	14	9.85 e − 02	1.1	6.95 e − 00	0	1.67 e − 04	1.9	8.33 e − 04	1.1

Table 4: Quadratic Running Cost Problem Relative Errors