POLICY ITERATION FOR THE DETERMINISTIC CONTROL PROBLEMS – A VISCOSITY APPROACH

WENPIN TANG, HUNG VINH TRAN, YUMING PAUL ZHANG

Abstract. This paper is concerned with the convergence rate of policy iteration for (deterministic) optimal control problems in continuous time. To overcome the problem of ill-posedness due to lack of regularity, we consider a semi-discrete scheme by adding a viscosity term via finite differences in space. We prove that PI for the semi-discrete scheme converges exponentially fast, and provide a bound on the error induced by the semi-discrete scheme. We also consider the discrete space-time scheme, where both space and time are discretized. Convergence rate of PI and the discretization error are studied.

1. Introduction

Optimal control is ubiquitous in science and engineering with a variety of applications including aerospace engineering [6, 10], chemical engineering [31], economy [24], operations research [36, 37] and robotics [2, 12]. Dynamic programming (DP) has proved to be an efficient tool to solve multistage optimal control problems since its inception by Bellman [5]. In recent years, reinforcement learning (RL) has shown great success in resolving complex decision making problems, notably AlphaGo [38] and humanoid tasks [18]. Policy iteration (PI), as a class of approximate or adaptive dynamic programming (ADP), is instrumental in many RL algorithms [40].

The idea of PI dates back to Howard [20] in a stochastic environment known as the Markov decision process (MDP). Subsequent works [7, 33, 34] explored PI for MDPs in discrete time and space; recently, [8, 30] considered PI for (deterministic) optimal control problems in discrete time and continuous space. In these works, PIs are proved to converge to the optimal control under suitable conditions on the model parameters. On the other hand, many real-world problems are modeled by dynamical systems evolving in continuous time, and it is known that DP for optimal control in continuous time and space entails the Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE). Despite its importance, PI for optimal control problems in continuous time and space has not been rigorously studied until recently. [1, 43] proved the convergence of PI for continuous-time linear quadratic optimal control problems; more general cases were settled in [28] under a fixed point assumption. For the stochastic control problems, [26, 35] showed that PI converges exponentially fast in the case where controls are only exercised on the drift term of the state process. Similar results were derived for the corresponding entropy-regularized problems [22, 41]. We also mention, in a closely related direction, [9, 45]

Date: January 1, 2023.

Key words and phrases. Finite differences, Hamilton-Jacobi-Bellman equations, optimal control, policy iteration.

The work of W. Tang is supported by NSF grants DMS-2113779 and DMS-2206038, and a start-up grant at Columbia University. The work of H. Tran is partially supported by NSF CAREER grant DMS-1843320, a Simons Fellowship, and a Vilas Faculty Early-Career Investigator Award.

studied value iteration for optimal control problems. See [29, 44] for recent progress on theory and applications of ADP for optimal control and RL.

In this paper, we study the convergence rate of PI for optimal control problems in continuous time and their discretization under fairly general conditions on the model parameters. Note that the convergence analysis in [1, 43] relies on the linear quadratic structure of the problem, while [28] assumed that the HJB operator enjoys a fixed point, or a contraction property which is hard to verify. None of these works quantified the convergence of PI to the optimal control. Moreover, PI for continuous-time control problems may even be ill-posed due to lack of regularity. Our idea is to introduce a viscosity term "$h\Delta^h$" in the policy evaluation, where h is the mesh size and Δ^h is the discrete Laplacian in space. We call it a *semi-discrete scheme*. Essentially the viscosity term is of order 1, which assures that the finite difference scheme is monotone. In fact, a monotone scheme is commonly desirable for numerical implementation so the addition of the finite difference viscosity term is natural. On the other hand, the viscosity term in the semi-discrete scheme mimics the vanishing viscosity approximation to first-order PDEs [16], which forces PI to converge exponentially fast (Theorem 3.1) as for the stochastic control problems. We also prove that the discrepancy between the optimal control problem and its semi-discrete scheme is of order $\sqrt{h}$ as $h \to 0$ (Theorem 3.3). Further we consider the time-discretization, called a *discrete space-time scheme*. The same results hold for PI for the discrete space-time scheme (Theorem 4.1 and Theorem 4.2). Our results echo recent work [19], which asserts that noise enhances the convergence of finite-horizon RL algorithms. In our setting, noise corresponds to the viscosity term, and the importance of finite-horizon is seen from various bounds with exponential dependence in time. Our analysis relies on PDE techniques, and may carry over to the study of differential games in solving Hamilton-Jacobi-Bellman-Issacs (HJBI) equations.

The rest of the paper is organized as follows. In Section 2, we provide background, and present the semi-discrete and the discrete space-time schemes. In Section 3 we study the semi-discrete scheme, and in Section 4 we analyze the discrete space-time scheme. We provide further PDE perspectives in Section 5. We conclude with Section 6.

2. Setup and preliminary results

In this section, we present the semi-discrete and the discrete space-time schemes. Consider a system whose state is governed by the ordinary differential equation:

$$\frac{dx_t}{dt} = f(t, x_t, \alpha_t), \tag{2.1}$$

where for $0 \le t \le T$, $x_t \in \mathbb{R}^d$ is the system state, and $\alpha_t \in A \subset \mathbb{R}^m$ is the control or policy. Here, A is a given compact subset of $\mathbb{R}^m$. The objective is

$$J(t, x, \alpha) := \int_t^T c(s, x_s, \alpha_s)\, ds + q(x_T) \quad \text{given } x_t = x, \tag{2.2}$$

and the goal is to minimize this objective function. Denote by

$$v_*(t, x) := \inf_{\alpha \in \mathcal{A}_t} J(t, x, \alpha), \tag{2.3}$$

where $\mathcal{A}_t$ is the standard admissible policy defined as $\mathcal{A}_t = \{\alpha : [t, T] \to A : \alpha \text{ is measurable}\}$. It is known that under suitable conditions on $c(\cdot)$ and $q(\cdot)$ (see [17, Chapter 2] or [42, Chapter

2]), v_* defined by (2.3) is the viscosity solution to

$$\begin{cases} \partial_t v + H(t,x,\nabla v) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ v(T,x) = q(x) & \text{on } \mathbb{R}^d, \end{cases} \tag{2.4}$$

where the Hamiltonian H is given by $H(t,x,p) := \inf_{a\in A}\left[c(t,x,a) + p\cdot f(t,x,a)\right]$. The optimal policy is given by

$$\alpha_*(t,x) = \alpha(t,x,\nabla v_*), \tag{2.5}$$

where

$$\alpha(t,x,p) := \arg\min_{a\in A}\left[c(t,x,a) + p\cdot f(t,x,a)\right], \tag{2.6}$$

is assumed for simplicity to be unique throughout this paper.

Policy iteration is an approximate dynamic programming, which alternates between policy evaluation to get the value function with the current control and policy improvement to optimize the value function. More precisely, for $n = 0,1,\cdots$, the iterative procedure is:

- Given $\alpha_n(t,x)$, solve the PDE

$$\begin{cases} \partial_t v_n + c(t,x,\alpha_n) + \nabla v_n\cdot f(t,x,\alpha_n) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ v_n(T,x) = q(x) & \text{on } \mathbb{R}^d. \end{cases} \tag{2.7}$$

- Set

$$\alpha_{n+1}(t,x) = \alpha(t,x,\nabla v_n) = \arg\min_{a\in A}\left[c(t,x,a) + \nabla v_n(t,x)\cdot f(t,x,a)\right]. \tag{2.8}$$

The key is to understand how the sequence $\{v_n\}$ approximates the optimal value v_*, and how $\{\alpha_n\}$ approximates the optimal policy α_*.

On the other hand, it is not clear whether the policy iteration scheme (2.7)–(2.8) is well-posed. Intuitively, to make sense of α_{n+1} we need v_n to be Lipschitz continuous, for which we then need α_n to be Lipschitz. This in turn requires ∇v_{n-1} to be Lipschitz. After iterations, this means that we need v_0 to be smooth which is in general not true.

2.1. **Semi-discrete schemes.** For $T \geq 1$, $h \in (0,1)$, $N \geq \max\{1, \|f\|_\infty/2\}$ and a given continuous function $\alpha_0 : \mathbb{R}\times\mathbb{R}^d \to A$, we solve for $n = 0,1,\cdots$

$$\begin{cases} \partial_t v_n^h + c(t,x,\alpha_n) + \nabla^h v_n^h\cdot f(t,x,\alpha_n) = -Nh\Delta^h v_n^h & \text{in } (0,T)\times\mathbb{R}^d \\ v_n^h(T,x) = q(x) & \text{on } \mathbb{R}^d. \end{cases} \tag{2.9}$$

Then set

$$\alpha_{n+1}(t,x) = \alpha(t,x,\nabla^h v_n^h) \qquad \text{in } (0,T)\times\mathbb{R}^d. \tag{2.10}$$

Here, for any $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $h \in \mathbb{R}\setminus\{0\}$, we use the notations

$$\nabla^h\varphi(x) := \left(\frac{\varphi(x+he_1) - \varphi(x-he_1)}{2h}, \cdots, \frac{\varphi(x+he_d) - \varphi(x-he_d)}{2h}\right),$$

$$\Delta^h\varphi(x) := \sum_{i=1}^d \frac{\varphi(x+he_i) - 2\varphi(x) + \varphi(x-he_i)}{h^2}.$$

Later we will also write $D^h\varphi(x) := \left(\frac{\varphi(x+he_1)-\varphi(x)}{h}, \cdots, \frac{\varphi(x+he_d)-\varphi(x)}{h}\right)$. It is clear that

$$\nabla^h\varphi(x) = \frac{1}{2}(D^h\varphi(x) - D^{-h}\varphi(x)). \tag{2.11}$$

The assumption $N \geq \|f\|_\infty/2$ guarantees that the numerical Hamiltonian is monotone and, as a consequence of this, the following comparison principle holds (see e.g., [13, 32, 42]).

Lemma 2.1. *Let v_0^h and $\tilde{v}_0^h$ be, respectively, a bounded continuous super- and sub- solution to* (2.9) *with $n = 0$, and satisfy $\tilde{v}_0^h \leq v_0^h$ at $t = T$. Then $\tilde{v}_0^h \leq v_0^h$ in $[0, T] \times \mathbb{R}^d$. Here by a supersolution (resp. subsolution), we mean that it satisfies* (2.9) *with the first equality replaced by $\leq$ (resp. $\geq$).*

First, we show that the scheme (2.9)–(2.10) is well-posed. We need the following assumptions:

(A1) $c(\cdot,\cdot,\cdot), f(\cdot,\cdot,\cdot), q(\cdot)$ are uniformly bounded and Lipschitz continuous in all of their dependences.
(A2) $\alpha(\cdot,\cdot,\cdot)$ and $\alpha_0(\cdot,\cdot)$ are uniformly Lipschitz continuous in all of their dependences.

Throughout this paper, we write C as various universal constants that only depend on d, N, and the constants in (A1)–(A2) unless otherwise stated. Specifically, since T is not a universal constant, we keep track of the dependence on T in most estimates. The constants C might vary from one line to another. By C_X or $C(X)$ we mean a constant that depend on universal constants and X.

Proposition 2.2. *Assume* (A1)–(A2) *and $N \geq \max\{1, \|f\|_\infty/2\}$. Then the iterative process* (2.9)–(2.10) *is well-defined, that is, there are Lipschitz continuous functions v_n^h, α_n satisfying* (2.9)–(2.10) *and v_n^h are uniformly bounded for all $n \geq 0$ and $h > 0$.*

Proof. Since α_0 is Lipschitz continuous, the unique solvability of (2.9) for $n = 0$ follows from [27, Theorem 2.4]. If one can show that v_0^h is uniformly bounded and Lipschitz continuous with Lipschitz constant C_h, then α_1 is Lipschitz continuous with Lipschitz constant C_h/h by the assumption that α is Lipschitz. From the same argument, we obtain a unique bounded and Lipschitz solution v_1^h. The existence of solutions then follows from iterations.

First we prove the boundedness of v_0^h. Since $c(\cdot,\cdot,\cdot), q(\cdot)$ are uniformly bounded, we have that $\pm\left[\|q\|_\infty + \|c\|_\infty(T - t)\right]$ are a supersolution and a subsolution to (2.9) with $n = 0$, respectively. Hence

$$-\|q\|_\infty - \|c\|_\infty(T - t) \leq v_0^h(t, x) \leq \|q\|_\infty + \|c\|_\infty(T - t),$$

for all $(t, x) \in [0, T] \times \mathbb{R}^d$. Actually the same bound holds for all v_n^h by this argument.

Next we show that v_0^h is Lipschitz continuous with Lipschitz constant independent of h when T is sufficiently small depending only on the assumption (A1) (but not on q). The general result follows immediately by iterations. For simplicity of notations, write

$$G(t, x, p) := c(t, x, \alpha_0(t, x)) + p \cdot f(t, x, \alpha_0(t, x)).$$

Then for $M := 2\|\nabla q\|_\infty + 1$, define

$$\tilde{G}(t, x, p) := \begin{cases} G(t, x, p) & \text{if } |p| \leq M, \\ G(t, x, Mp/|p|) & \text{if } |p| > M. \end{cases}$$

It follows from (A1) and the choice of N that

$$|\tilde{G}_t(t, x, p)|,\ |\tilde{G}_x(t, x, p)| \leq C(1 + M), \quad |\tilde{G}_p(t, x, p)| \leq 2N. \tag{2.12}$$

Now let $\tilde{v}^h$ be the solution to

$$\begin{cases} \partial_t \tilde{v}^h + \tilde{G}(t,x,\nabla^h \tilde{v}^h) = -Nh\Delta^h \tilde{v}^h, \\ \tilde{v}^h(T,x) = q(x). \end{cases}$$

The goal is to show that $\tilde{v}^h$ is Lipschitz continuous, and $\tilde{v}^h = v_0^h$ in $[0,T]\times\mathbb{R}^d$.

Note that for any $e \in \mathbb{S}^{d-1}$ and $s \in (0,1)$, $p_s := \frac{\tilde{v}^h(t,x+se)-\tilde{v}^h(t,x)}{s}$ satisfies

$$\begin{cases} \partial_t p_s + G_1(t,x) + G_2(t,x)\cdot\nabla^h p_s = -Nh\Delta^h p_s & \text{in } (0,T)\times\mathbb{R}^d, \\ p_s(T,x) = \dfrac{q(x+se)-q(x)}{s} & \text{on } \mathbb{R}^d, \end{cases} \tag{2.13}$$

where

$$G_1(t,x) := \frac{1}{s}\int_0^s \tilde{G}_x\left(t, x+ze, \nabla^h \tilde{v}^h(t,x+se)\right)\cdot e\, dz,$$
$$G_2(t,x) := \int_0^1 \tilde{G}_p\left(t,x,\nabla^h\tilde{v}^h(t,x) + z(\nabla^h\tilde{v}^h(t,x+se) - \nabla^h\tilde{v}^h(t,x))\right) dz.$$

It is clear from (2.12) that $|G_1| \le C(1+M)$ and $|G_2| \le 2N$. This yields that the comparison principle for (2.13) holds. Thus, by comparing p_s with $\pm(\|\nabla q\|_\infty + C(1+M)(T-t))$, we obtain $|p_s(t,x)| \le \|\nabla q\|_\infty + C(1+M)(T-t)$. Sending $s\to 0$ yields for some C depending only on (A1), $|\nabla_e \tilde{v}^h(t,x)| \le \|\nabla q\|_\infty + C(1+M)(T-t)$. Thus, if $t \le T \le (2C)^{-1}$, we have that $\tilde{v}^h(t,x)$ is Lipschitz and

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\nabla \tilde{v}^h(t,x)| \le \|\nabla q\|_\infty + 1/2 + M/2 = M.$$

From the definition of ∇^h, we get $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\nabla^h\tilde{v}^h(t,x)| \le M$. Hence, $\tilde{v}^h$ is a solution to (2.9) for $n=0$. The uniqueness of the solution to (2.9) yields that $v_0^h \equiv \tilde{v}^h$. So we obtain the uniform Lipschitz continuity of v_0^h in space, with Lipschitz constant of the form $C\exp(CT)$. The Lipschitz regularity in time follows from the equation. □

We point out that the Lipschitz constant of v_n^h may depend on both n and h for $n \ge 1$. Another consequence of the comparison principle is that the functions v_n^h are monotone decreasing in n.

Proposition 2.3. *Under the assumptions of Proposition 2.2, we have for all $n \ge 0$,*

$$v_{n+1}^h \le v_n^h \qquad \text{in } [0,T]\times\mathbb{R}^d.$$

Proof. By the definition of α_n,

$$\begin{aligned} &c(t,x,\alpha_{n+1}(t,x)) + \nabla^h v_n^h \cdot f(t,x,\alpha_{n+1}(t,x)) \\ &\qquad \le c(t,x,\alpha_n(t,x)) + \nabla^h v_n^h\cdot f(t,x,\alpha_n(t,x)). \end{aligned}$$

Thus v_n^h is a supersolution to (2.9) with subscripts $n+1$ as it satisfies

$$\partial_t v_n^h + c(t,x,\alpha_{n+1}) + \nabla^h v_n^h \cdot f(t,x,\alpha_{n+1}) \le -Nh\Delta^h v_n^h \qquad \text{in } (0,T)\times\mathbb{R}^d.$$

Therefore, the comparison principle yields $v_n^h \le v_{n+1}^h$ in $[0,T]\times\mathbb{R}^d$ for each $n\ge 0$. □

Since v_n^h is uniformly bounded for all $n \geq 0$, the monotonicity property yields that v_n^h converges locally uniformly as $n \to \infty$. Let us denote the limit as v^h. Then by the stability property of viscosity solutions, v^h solves

$$\begin{cases} \partial_t v^h + H(t, x, \nabla^h v^h) = -Nh\Delta^h v^h & \text{in } (0,T) \times \mathbb{R}^d, \\ v^h(T, x) = q(x) & \text{on } \mathbb{R}^d, \end{cases} \tag{2.14}$$

where

$$\begin{aligned} H(t, x, p) &:= c(t, x, \alpha(t, x, p)) + p \cdot f(t, x, \alpha(t, x, p)) \\ &= \min_{a \in A} \left[c(t, x, a) + p \cdot f(t, x, a) \right]. \end{aligned} \tag{2.15}$$

Since $\alpha(t, x, p)$ is assumed to be uniformly Lipschitz continuous in all of its dependences, there exists $C > 0$ such that for all $(t, x, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$,

$$|H_t(t, x, p)|, \, |H_x(t, x, p)| \leq C(1 + |p|), \quad |H_p(t, x, p)| \leq C. \tag{2.16}$$

Moreover, one can expect that v^h converges as $h \to 0$, and the limit v should be the unique solution to (2.4).

It was proved in Proposition 2.2 that v_0^h is uniformly bounded and Lipschitz continuous in $[0, T] \times \mathbb{R}^d$. By the same proof, we have that v^h and v are also uniformly bounded and Lipschitz continuous for all $h > 0$.

Lemma 2.4. *Under the assumptions of Proposition 2.2, let v_0^h, v^h and v be, respectively, solutions to* (2.9) *(for $n = 0$),* (2.14) *and* (2.4). *Then in $[0, T] \times \mathbb{R}^d$, v_0^h, v^h and v are bounded by $C(1 + T)$ and are Lipschitz continuous with Lipschitz constant $C \exp(CT)$ for some universal constant $C > 0$.*

For a general class of first-order Hamilton-Jacobi (continuous) equations, we refer to [3, 4] for the regularity results.

2.2. **Discrete space-time schemes.** Now we consider the scheme that is discrete in both space and time. Let $\tau, h \in (0, 1)$, and N such that

$$\max\{1, \|f\|_\infty / 2\} \leq N \leq h/(2\tau). \tag{2.17}$$

Assuming that $T/\tau \in \mathbb{N}^+$, we denote

$$\mathbb{N}_T^\tau := \{0, \tau, 2\tau, \cdots, T\}, \quad \mathbb{Z}_h^d := h\mathbb{Z}^d,$$

$$\Omega_T^{\tau,h} := \mathbb{N}_T^\tau \times \mathbb{Z}_h^d \quad \text{and} \quad \Omega_T' := (\mathbb{N}_T^\tau \backslash \{0\}) \times \mathbb{Z}_h^d.$$

Given a Lipschitz continuous function $\alpha_0(t, x)$, let $V_n^{\tau,h} : \Omega_T^{\tau,h} \to \mathbb{R}$ be defined iteratively for $n = 0, 1, \cdots$ as follows:

$$\begin{cases} \partial_t^\tau V_n^{\tau,h}(t, x) + c(t, x, \alpha_n) + \nabla^h V_n^{\tau,h} \cdot f(t, x, \alpha_n) = -Nh\Delta^h V_n^{\tau,h} & \text{in } \Omega_T', \\ V_n^{\tau,h}(T, x) = q(x) & \text{on } \mathbb{Z}_h^d \end{cases} \tag{2.18}$$

with

$$\alpha_{n+1}(t, x) := \alpha(t, x, \nabla^h V_n^{\tau,h}) \qquad \text{in } \Omega_T'. \tag{2.19}$$

Here we used the notation $\partial_t^\tau V_n^{\tau,h}(t, x) := \frac{V_n^{\tau,h}(t,x) - V_n^{\tau,h}(t-\tau,x)}{\tau}$.

We also consider the following equation

$$\begin{cases} \partial_t^\tau V^{\tau,h} + H(t,x,\nabla^h V^{\tau,h}) = -Nh\Delta^h V^{\tau,h} & \text{in } \Omega_T', \\ V^{\tau,h}(T,x) = q(x) & \text{on } \mathbb{Z}_h^d. \end{cases} \tag{2.20}$$

where H is given by (2.15). The goal is to show that $V_n^{\tau,h}$ converges to $V^{\tau,h}$ as $n \to \infty$, and $V^{\tau,h}$ converges to v as $\tau, h \to 0$, where v is given by (2.4).

Similarly as before, we write

$$c := c(t,x,\alpha(t,x,\nabla^h V^{\tau,h})) \quad \text{and} \quad f := f(t,x,\alpha(t,x,\nabla^h V^{\tau,h}))$$
$$c_n := c(t,x,\alpha_n(t,x)) \quad \text{and} \quad f_n := f(t,x,\alpha_n(t,x)))$$

for simplicity, when there is no confusion.

We will use the following operator. For each $t \in \mathbb{N}_T^\tau$, let $\mathcal{F}_t : L^\infty(\mathbb{Z}_h^d) \to L^\infty(\mathbb{Z}_h^d)$ be defined as

$$\mathcal{F}_t(U)(x) := U(x) + \tau H(t,x,\nabla^h U(x)) + Nh\tau\Delta^h U(x). \tag{2.21}$$

Then the equation in (2.20) can be rewritten as $V_n^{\tau,h}(t-\tau,x) = \mathcal{F}_t(V_n^{\tau,h}(t,\cdot))(x)$. We need

$$\max\{1, \|H_p\|_\infty/2\} \leq N \leq h/(2\tau),$$

(which corresponds (2.17) as $\|H_p\|_\infty = \|f\|_\infty$) to guarantee a monotonicity property of the operator $\mathcal{F}_t$. That is, for all $t \in \mathbb{N}_T^\tau$ and $U, V \in L^\infty(\mathbb{Z}_h^d)$ satisfying $U \leq V$, we have $\mathcal{F}_t(U) \leq \mathcal{F}_t(V)$, see e.g., [13, 42]. It is easy to see that the same holds if we replace $H(t,x,p)$ by $c_n(t,x) + p \cdot f_n(t,x)$.

The monotonicity property is important because it immediately implies the comparison principle of (2.20) and the scheme (2.18)–(2.19), in the sense that is similar to Lemma 2.1. As a consequence of this, one can show the following properties.

Proposition 2.5. *Assume* (A1)–(A2) *and* (2.17). *Then in $\Omega_T^{\tau,h}$, the solutions $V_n^{\tau,h}, V^{\tau,h}$ are bounded by $C(1+T)$, and are Lipschitz continuous with Lipschitz constant $C\exp(CT)$ for some universal constant $C > 0$. Moreover for all $n \geq 0$ we have $V_{n+1}^{\tau,h} \leq V_n^{\tau,h}$ in $\Omega_T^{\tau,h}$.*

The proof of Proposition 2.5 is similar to those of Propositions 2.2, 2.3 and Lemma 2.4, and hence we skip it.

3. Analysis of semi-discrete schemes

3.1. Convergence of PI.

We show that for each fixed $h \in (0,1)$, $v_n^h \to v^h$ as $n \to \infty$ exponentially fast in L^2_{loc} norm.

Theorem 3.1. *Assume* (A1)–(A2) *and $N \geq 1$. Let v_n^h and v^h be, respectively, continuous solutions to* (2.9) *and* (2.14). *Then there exists a universal constant $C > 0$ such that for all $n \geq 1$, $R \geq 1$ and $t \in [0,T]$ we have*

$$\int_{B_R} \left| v_n^h(t,x) - v^h(t,x) \right|^2 dx \leq \frac{h}{2^{n+1}} \int_t^T \int_{\mathbb{R}^d} \exp\left[C(1 + \|\nabla^h v^h\|_\infty^2)(s-t)/h \right] \times$$
$$\left| D^h(v_0^h(s,x) - v^h(s,x)) \right|^2 \min\left\{1, e^{-|x|+R+1}\right\} dxds.$$

In particular, we have $\sup_{t\in[0,T]} \int_{B_R} \left| v_n^h(t,x) - v^h(t,x) \right|^2 dx \leq C2^{-n} \exp\left[C\exp(CT)/h\right] R^d$.

Proof. In this proof, let us write $v_n := v_n^h$ and $v := v^h$, and assume $T \geq 1$ for simplicity. For any fixed $R \geq 1$, let $\varphi = \varphi_R : [0, \infty) \to (0, 1]$ be C^1 and satisfying

$$\begin{aligned} &\varphi(r) = 1 \quad \text{on } [0, R], \qquad \varphi(r) = e^{-r+R} \quad \text{on } [R+1, \infty), \\ &-\varphi'(r) \in [0, 4\varphi(r)] \quad \text{for all } r > 0. \end{aligned} \tag{3.1}$$

It is clear that such φ exists.

Next, for some $A \geq 1$ to be determined, set

$$E_{t,n} := \frac{1}{2} e^{At} \int_{\mathbb{R}^d} |v_n(t,x) - v(t,x)|^2 \varphi(|x|)\, dx \tag{3.2}$$

which is finite since v_n, v are uniformly bounded. Direct computation yields

$$\frac{d}{dt} E_{t,n} = A E_{t,n} + e^{At} \underbrace{\int_{\mathbb{R}^d} (v_n(t,x) - v(t,x)) \left(\partial_t v_n(t,x) - \partial_t v(t,x)\right) \varphi(|x|)\, dx}_{=: X_{t,n}} . \tag{3.3}$$

Recall from (2.15) that $H(t, x, \nabla^h v) = c(t, x, \alpha(t, x, \nabla^h v)) + \nabla v \cdot f(t, x, \alpha(t, x, \nabla^h v))$. Write $c := c(t, x, \alpha(t, x, \nabla^h v))$, $f := f(t, x, \alpha(t, x, \nabla^h v))$, $c_n := c(t, x, \alpha_n(t, x))$ and $f_n := f(t, x, \alpha_n(t, x))$ for simplicity. Then, using

$$\begin{aligned} \int_{\mathbb{R}^d} \Delta^h v\, v \varphi\, dx = &- \int_{\mathbb{R}^d} |D^h v|^2 \varphi\, dx + \frac{1}{h^2} \sum_{i=1}^d \int_{\mathbb{R}^d} v(t, x + he_i)(v(t, x+he_i) - v(t,x)) \varphi(t,x)\, dx \\ &- \frac{1}{h^2} \sum_{i=1}^d \int_{\mathbb{R}^d} v(t,x)(v(t,x) - v(t, x - he_i)) \varphi(t,x)\, dx \\ = &- \int_{\mathbb{R}^d} |D^h v|^2 \varphi\, dx - \int_{\mathbb{R}^d} v D^{-h} v \cdot D^{-h} \varphi\, dx, \end{aligned}$$

we obtain from the equation that

$$\begin{aligned} X_{t,n} &= - \int_{\mathbb{R}^d} (v_n - v)(\nabla^h v_n \cdot f_n + c_n + Nh \Delta^h v_n - \nabla^h v \cdot f - c - Nh \Delta^h v) \varphi\, dx \\ &\geq Nh \int_{\mathbb{R}^d} |D^h(v_n - v)|^2 \varphi\, dx - Nh \int_{\mathbb{R}^d} |v_n - v| |D^{-h}(v_n - v)| |D^{-h} \varphi|\, dx \\ &\quad - \int_{\mathbb{R}^d} |v_n - v| \Big(|\nabla^h(v_n - v)| |f_n| + |f_n - f| |\nabla^h v| + |c_n - c| \Big) \varphi\, dx. \end{aligned} \tag{3.4}$$

Due to (3.1), $|D^{-h} \varphi(|x|)| \leq C \varphi(|x|)$ for some constant $C > 0$. Also, using $\|f\|_\infty < \infty$ and (2.11), we have $|\nabla^h(v_n - v)||f_n| \leq C(|D^h(v_n - v)| + |D^{-h}(v_n - v)|)$. Since v is Lipschitz continuous, $|\nabla^h v| \leq M$ for some $M \geq 1$. So, by (2.10) and the uniform Lipschitz continuity of f, c and α, we have for some $C > 0$,

$$|f_n - f| |\nabla^h v| + |c_n - c| \leq CM(|D^h(v_{n-1} - v)| + |D^{-h}(v_{n-1} - v)|). \tag{3.5}$$

With all these, if denoting

$$G_{t,n}^h := \int_{\mathbb{R}^d} |D^h(v_n(t,x) - v(t,x))|^2 \varphi(|x|)\, dx,$$

it follows from (3.4) that for some $C>0$,

$$\begin{aligned} X_{t,n} \geq NhG^h_{t,n} - C\int_{\mathbb{R}^d} |v_n - v|(|D^h(v_n-v)| + |D^{-h}(v_n-v)|)\varphi\, dx \\ - CM\int_{\mathbb{R}^d} |v_n - v|(|D^h(v_{n-1}-v)| + |D^{-h}(v_{n-1}-v)|)\varphi\, dx. \end{aligned}$$

By the choice of φ, there exists $C>0$ such that

$$G^{-h}_{t,n} \leq (1+Ch)G^h_{t,n} \tag{3.6}$$

Then, using (3.2) and Young's inequality, we get for any $\sigma_1,\sigma_2>0$,

$$\begin{aligned} X_{t,n} &\geq NhG^h_{t,n} - \frac{\sigma_1}{2+Ch}\int_{\mathbb{R}^d}(|D^h(v_n-v)|^2 + |D^{-h}(v_n-v)|^2)\varphi\, dx \\ &\quad - \frac{\sigma_2}{2+Ch}\int_{\mathbb{R}^d}(|D^h(v_{n-1}-v)|^2 + |D^{-h}(v_{n-1}-v)|^2)\varphi\, dx \\ &\quad - C(2+Ch)(\sigma_1^{-1} + M^2\sigma_2^{-1})\int_{\mathbb{R}^d}|v_n-v|^2\varphi\, dx \\ &\geq (Nh-\sigma_1)G^h_{t,n} - \sigma_2 G^h_{t,n-1} - C(\sigma_1^{-1}+M^2\sigma_2^{-1})e^{-At}E_{t,n}. \end{aligned}$$

Using this and $E_{T,n}=0$, and integrating (3.3) over $[t,T]$, we obtain for some universal $C>0$,

$$\begin{aligned} -E_{t,n} &\geq (A - C\sigma_1^{-1} - CM^2\sigma_2^{-1})\int_t^T E_{s,n}\, ds \\ &\quad + (Nh-\sigma_1)\int_t^T e^{As}G^h_{s,n}\, ds - \sigma_2\int_t^T e^{As}G^h_{s,n-1}\, ds. \end{aligned} \tag{3.7}$$

Now taking $\sigma_1 := h/2$, $\sigma_2 := h/4$ and $A := 6CM^2/h$, then (3.7) and $N\geq 1$ yield

$$\int_t^T e^{As}G^h_{s,n}\, ds \leq \frac{1}{2}\int_t^T e^{As}G^h_{s,n-1}\, ds \leq \ldots \leq 2^{-n}\int_t^T e^{As}G^h_{s,0}\, ds.$$

With this, (3.7) also shows that $E_{t,n} \leq \frac{h}{4}\int_t^T e^{As}G^h_{s,n-1}\, ds \leq \frac{h}{2^{n+1}}\int_t^T e^{As}G^h_{s,0}\, ds$. Therefore, for all $n\geq 0$ and $t\in[0,T]$, we obtain

$$\int_{B_R}|v_n(t,x)-v(t,x)|^2\, dx \leq \frac{h}{2^{n+1}}\int_t^T\int_{\mathbb{R}^d} e^{A(s-t)}|D^h(v_0(s,x)-v(s,x))|^2\varphi(|x|)\, dxds,$$

which, combined with Lemma 2.4, concludes the proof. $\square$

Remark 3.1. *In the proof of Theorem 3.1, we only used the following: uniform Lipschitz continuity of f, c and α, and uniform boundedness of f and $|\nabla^h v^h|$. In particular, the solutions v^h_n and v^h are allowed to have certain growth at $x=\infty$, and the comparison principle is not needed.*

By Theorem 3.1, we immediately have the convergence of the policies.

Theorem 3.2. *Assume* (A1)–(A2) *and $N\geq 1$. Then there exists a universal constant $C>0$ such that for all $n\geq 0$ and $R\geq 1$ we have*

$$\sup_{t\in[0,T]}\int_{B_R}\left|\alpha(t,x,\nabla^h v^h_n(t,x)) - \alpha(t,x,\nabla^h v^h(t,x))\right|^2\, dx \leq C2^{-n}\exp\left[C\exp(CT)/h\right]R^d.$$

Proof. Since α is Lipschitz continuous,

$$\int_{B_R} \left|\alpha(t,x,\nabla^h v_n^h(t,x)) - \alpha(t,x,\nabla^h v^h(t,x))\right|^2 dx$$
$$\leq \frac{C}{h^2}\sum_{i=1}^d \int_{B_R} \left|v_n^h(t,x+he_i) - v^h(t,x+he_i) - v_n^h(t,x-he_i) + v^h(t,x-he_i)\right|^2 dx.$$

We can then conclude the proof from Theorem 3.1. $\square$

3.2. **Convergence of v^h as $h \to 0$.** Let v^h and v be, respectively, solutions to (2.14) and (2.4). We show $|v^h - v| \leq C_T\sqrt{h}$, where the rate is sharp (we refer to a simple example given in [14]).

Theorem 3.3. *Assume* (A1)–(A2) *and* $N \geq \max\{1, \|f\|_\infty/2\}$. *Then there exists a universal constant* $C > 0$ *such that*

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |v(t,x) - v^h(t,x)| \leq C(1+T)(1+\|\nabla v\|_\infty)\sqrt{h}.$$

In particular, we have $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |v(t,x) - v^h(t,x)| \leq C\exp(CT)\sqrt{h}$.

Remark 3.2. *This rate was obtained in* [15, 27] *for a large class of parabolic Bellman equations with Lipschitz coefficients. We apply a different argument – the classical doubling variable method that is used in* [13] *in which a discrete space-time homogeneous Hamilton-Jacobi equation is discussed. This argument allows us to obtain the same sharp estimate for the scheme* (2.18)*, while it seems that the method in* [15, 27] *cannot (see Remark 4.1). See also* [11] *for a different proof of this convergence rate via the nonlinear adjoint method.*

Proof. Let us assume that $T \geq 1$. Suppose for some $(t_0,x_0) \in [0,T]\times\mathbb{R}^d$ such that

$$8\sigma := v(t_0,x_0) - v^h(t_0,x_0) \geq \frac{1}{2}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left[v(t,x) - v^h(t,x)\right] > 0. \tag{3.8}$$

Below we will show $\sigma \leq CT(1+\|\nabla v\|_\infty)\sqrt{h}$.

Consider a smooth function $g : \mathbb{R}^{d+1} \to [0,1]$ such that

(g1) $g(t,x) = 1 - t^2 - |x|^2$ if $t^2 + |x|^2 < 1/2$,

(g2) $0 \leq g(t,x) \leq 1/2$ if $t^2 + |x|^2 > 1/2$, and $g(t,x) = 0$ if $t^2 + |x|^2 > 1$.

For $\varepsilon > 0$, denote $g_\varepsilon(t,x) := g(t/\varepsilon, x/\varepsilon)$, and

$$L := \sup\left\{v(t,x), -v^h(t,x) \,:\, (t,x) \in [0,T]\times\mathbb{R}^d\right\} + 1 \geq 1,$$

By Lemma 2.4, $\sigma \leq L \leq CT$ for some universal constant $C > 0$. Next, for $\phi(x) := (1+|x|^2)^{1/2}$ and $R \geq |x_0| + T$, we define $\Phi^h : [0,T]^2 \times \mathbb{R}^{2d} \to \mathbb{R}$ by

$$\Phi^h(t,s,x,y) := v(t,x) - v^h(s,y) - \frac{\sigma}{T}(2T - t - s)$$
$$- \frac{\sigma}{R}(\phi(x) + \phi(y)) + (8L + 2\sigma)g_\varepsilon(t-s, x-y).$$

Since v, v^h are bounded, there exists $(t_1,s_1,x_1,y_1) \in [0,T]^2\times\mathbb{R}^{2d}$ such that

$$\Phi^h(t_1,s_1,x_1,y_1) = \max_{[0,T]^2\times\mathbb{R}^{2d}} \Phi^h(t,s,x,y). \tag{3.9}$$

Due to $\phi(x_0) \le R$, by (3.8),

$$\Phi^h(t_1, s_1, x_1, y_1) \ge \Phi^h(t_0, t_0, x_0, x_0) \ge 8L + 6\sigma. \tag{3.10}$$

Since $\max\{v(t_1, x_1), -v^h(s_1, y_1)\} \le L$, we deduce $\Phi^h(t_1, s_1, x_1, y_1) \le 2L + (8L + 2\sigma)g_\varepsilon(t_1 - s_1, x_1 - y_1)$, which, together with (3.10), implies $g_\varepsilon(t_1 - s_1, x_1 - y_1) \ge 3/4$. Then by (g1), we get that for some $C > 0$,

$$g_\varepsilon(t - s, x - y) = 1 - \varepsilon^{-2}|t - s|^2 - \varepsilon^{-2}|x - y|^2, \tag{3.11}$$

whenever $|t - t_1|, |s - s_1|, |x - x_1|, |y - y_1| \le \varepsilon/C$.

Now, by (3.9), the mapping

$$(t, x) \mapsto v(t, x) + \frac{\sigma}{T}t - \frac{\sigma}{R}\phi(x) + (8L + 2\sigma)g_\varepsilon(t - s_1, x - y_1). \tag{3.12}$$

is maximized at $(t, x) = (t_1, x_1)$. Together with the fact that v is Lipschitz continuous (taking $M := 1 + \|\nabla v\|_\infty$) and $|\nabla\phi| \le 1$, we find that $|\nabla_x g_\varepsilon(t_1 - s_1, x_1 - y_1)| \le (M + \sigma R^{-1})(8L + 2\sigma)^{-1}$ and, $|\partial_t g_\varepsilon(t_1 - s_1, x_1 - y_1)| \le (M + \sigma T^{-1})(8L + 2\sigma)^{-1}$. By (3.11), $\sigma \le L \le CT$ and $R \ge T$, these yield

$$|x_1 - y_1| \le C\varepsilon^2(M + \sigma R^{-1})(L + \sigma)^{-1} \le C\varepsilon^2 ML^{-1}, \tag{3.13}$$

and

$$|t_1 - s_1| \le C\varepsilon^2(M + \sigma T^{-1})(L + \sigma)^{-1} \le C\varepsilon^2 ML^{-1}. \tag{3.14}$$

Now, we firstly assume that $t_1, s_1 < T$. In view of (3.12), we apply the viscosity solution test for v to get

$$\begin{aligned} &-\frac{\sigma}{T} - (8L + 2\sigma)\,\partial_t g_\varepsilon(t_1 - s_1, x_1 - y_1) \\ &\qquad + H\left(t_1, x_1, \frac{\sigma}{R}\nabla\phi(x_1) - (8L + 2\sigma)\nabla_x g_\varepsilon(t_1 - s_1, x_1 - y_1))\right) \ge 0. \end{aligned} \tag{3.15}$$

Similarly, since $(s, y) \to v^h(s, y) - \frac{\sigma}{T}s + \frac{\sigma}{R}\phi(y) - (8L + 2\sigma)g_\varepsilon(t_1 - s, x_1 - y)$ is minimized at (s_1, y_1), the comparison principle yields

$$\begin{aligned} &\frac{\sigma}{T} - (8L + 2\sigma)\,\partial_t g_\varepsilon(t_1 - s_1, x_1 - y_1) \\ &\qquad + H\left(s_1, y_1, -\frac{\sigma}{R}\nabla^h\phi(y_1) - (8L + 2\sigma)\nabla^h_x g_\varepsilon(t_1 - s_1, x_1 - y_1)\right) \\ &\qquad - Nh\Delta^h\left[\frac{\sigma}{R}\phi(y_1) - (8L + 2\sigma)g_\varepsilon(t_1 - s_1, x_1 - y_1))\right] \le 0. \end{aligned}$$

Thus we get

$$\begin{aligned} \frac{2\sigma}{T} &\le H\left(t_1, x_1, \frac{\sigma}{R}\nabla\phi(x_1) - (8L + 2\sigma)\nabla_x g_\varepsilon(t_1 - s_1, x_1 - y_1))\right) \\ &\qquad - H\left(s_1, y_1, -\frac{\sigma}{R}\nabla\phi(y_1) - (8L + 2\sigma)\nabla^h_x g_\varepsilon(t_1 - s_1, x_1 - y_1)\right) \\ &\qquad + Nh\Delta^h\left[\frac{\sigma}{R}\phi(y_1) - (8L + 2\sigma)g_\varepsilon(t_1 - s_1, x_1 - y_1))\right]. \end{aligned} \tag{3.16}$$

It follows from (3.11) that for $h \ll \varepsilon$, we have at point $(t_1 - s_1, x_1 - y_1)$,

$$\nabla^h_x g_\varepsilon = \nabla_x g_\varepsilon = 2\varepsilon^{-2}(x_1 - y_1), \quad \Delta^h g_\varepsilon = -2d\varepsilon^{-2}. \tag{3.17}$$

Due to $|\nabla\phi| \le 1$ and $\Delta^h\phi \le C$, we get

$$Nh\Delta^h\left[\frac{\sigma}{R}\phi(y_1) - (8L + 2\sigma)g_\varepsilon(t_1 - s_1, x_1 - y_1))\right] \le CL\varepsilon^{-2}h. \tag{3.18}$$

Using (3.16)–(3.18) and the regularity of H (see (2.16)), we obtain for some universal C,

$$\begin{aligned}2\sigma T^{-1} &\le C\sigma R^{-1}(|\nabla\phi(x_1)| + |\nabla\phi(y_1)|) + CL\varepsilon^{-2}h\\ &\qquad + C(|t_1 - s_1| + |x_1 - y_1|)\,[1 + (8L + 2\sigma)|\nabla_x\, g_\varepsilon(t_1 - s_1, x_1 - y_1)|]\\ &\le C\sigma R^{-1} + CL\varepsilon^{-2}h + C(|t_1 - s_1| + |x_1 - y_1|)\left(1 + L\varepsilon^{-2}|x_1 - y_1|\right)\end{aligned}$$

which, by (3.13) and (3.14), yields $\sigma T^{-1} \le C\sigma R^{-1} + CL\varepsilon^{-2}h + C\varepsilon^2 M^2 L^{-1}$. Now we take $\varepsilon := M^{-1/2}L^{1/2}h^{1/4}$ and pass $R \to \infty$. Then when h is sufficiently small, we obtain $\sigma \le CTM\sqrt{h}$ for some universal $C > 0$. This finishes the proof of the upper bound of $\sup_{[0,T]\times\mathbb{R}^d}(v - v^h)$ in the case when $t_1, s_1 < T$.

Next, suppose that one of t_1 and s_1 equals to T. Let us only prove for the case when $t_1 = T$. By (3.10) and the definition of Φ^h,

$$8L + 6\sigma \le v(t_1, x_1) - v^h(s_1, y_1) + (8L + 2\sigma)g_\varepsilon(t_1 - s_1, x_1 - y_1).$$

It follows from the proof Lemma 2.4 that v^h is Lipschitz continuous with unit Lipschitz constant when $|T - t| \le C$. Note that $\varepsilon^2 ML^{-1} \le C$. Hence (3.13) and (3.14) yield

$$\begin{aligned}8L + 6\sigma &\le |v(T, x_1) - q(y_1)| + |q(y_1) - v^h(s_1, y_1)| + (8L + 2\sigma)g_\varepsilon(T - s_1, x_1 - y_1)\\ &\le C(|x_1 - y_1| + |T - s_1|) + 8L + 2\sigma \le C\varepsilon^2 ML^{-1} + 8L + 2\sigma.\end{aligned}$$

This yields $\sigma \le C\sqrt{h}$ for some universal $C > 0$.

Finally, the upper bound estimate for $\sup_{[0,T]\times B_R}(v^h - v)$ follows by using the same argument as the above. Applying Lemma 2.4 permits to conclude. □

3.3. **Almost everywhere convergence of the policy.** It was proved in [23] that the solution v to (2.4) is semi-concave in space. From this, we are able to derive the almost everywhere convergence of the policies.

Below we say that a function $g : \mathbb{R}^d \to \mathbb{R}$ is uniformly semi-concave if there exists $C > 0$ such that for all $x, y \in \mathbb{R}^d$ we have $g(x+y) + g(x-y) - 2g(x) \le C|y|^2$. If g is uniformly bounded and Lipschitz continuous, and both $\pm g$ are uniformly semi-concave, then g is bounded in $W^{2,\infty}(\mathbb{R}^d)$. We make the following assumption:

(A3) $q(\cdot)$ is uniformly semi-concave, and $c(t, \cdot, a), f(t, \cdot, a)$ are bounded in $W^{2,\infty}(\mathbb{R}^d)$ uniformly in $t \in [0, T]$ and $a \in A$.

Theorem 3.4. *Under the assumptions of Theorem 3.3, further assume* (A3). *Then $v(t, \cdot)$ is uniformly semi-concave for all $t \in [0, T]$. Moreover, for each $t \in [0, T]$ we have for a.e. $x \in \mathbb{R}^d$,*

$$\alpha(t_h, x_h, \nabla^h v^h(t_h, x_h)) \to \alpha(t, x, \nabla v(t, x)) \quad \text{as } h \to 0$$

where $[0, T] \times \mathbb{R}^d \ni (t_h, x_h) \to (t, x)$ as $h \to 0$.

We next show a weak type of semi-concavity of v^h.

Theorem 3.5. *Under the assumptions of Theorem 3.4, there exists $C > 0$ (also depending on* (A3)*) such that for all $h \in (0, 1)$, $t \in [0, T]$ and $x, y \in \mathbb{R}^d$,*

$$v^h(t, x + y) + v^h(t, x - y) - 2v^h(t, x) \le C\exp(CT)\,(|y|^2 + \sqrt{h}).$$

The proofs of the two theorems are similar to those of Theorem 4.3 and Theorem 4.4, and we choose to write the full details down there (as it is slightly more complicated there).

4. Analysis of discrete space-time schemes

4.1. **Convergence of PI.** The parallel result of Theorem 3.1 on the convergence of $V_n^{\tau,h} \to V^{\tau,h}$ holds the same. However the proof is more involved due to the discretization in the time direction. In it, we will emphasize the difference.

Theorem 4.1. *Assume* (A1)–(A2) *and* $N \geq 1$. *Let* $V_n := V_n^{\tau,h}$ *and* $V := V^{\tau,h}$ *be, respectively, continuous solutions to* (2.18) *and* (2.20). *Then there exists a universal constant* $C > 0$ *such that if* $C(1 + \|\nabla^h V\|_\infty^2)\tau \leq h$, *we have for all* $n \geq 1$, $R \geq 1$ *and* $t \in \mathbb{N}_T^\tau$,

$$\sum_{x \in \mathbb{Z}_h^d, |x| \leq R} |V_n(t,x) - V(t,x)|^2 \leq \frac{h\tau}{2^{n+1}} \sum_{t \leq s \in \mathbb{N}_T^\tau} \sum_{x \in \mathbb{Z}_h^d} \exp\left[C \exp(1 + \|\nabla^h V\|_\infty^2)(s-t)/h\right] \left|D^h(V_0(s,x) - V(s,x))\right|^2 \min\left\{1, e^{-|x|+R+1}\right\}.$$

In particular, we have

$$\max_{t \in \mathbb{N}_T^\tau} \sum_{x \in \mathbb{Z}_h^d, |x| \leq R} |V_n(t,x) - V(t,x)|^2 \leq C 2^{-n} \exp\left[C \exp(CT)/h\right] R^d.$$

$$\max_{t \in \mathbb{N}_T^\tau} \sum_{x \in \mathbb{Z}_h^d, |x| \leq R} \left|\alpha(t,x,\nabla^h V_n(t,x)) - \alpha(t,x,\nabla^h V(t,x))\right|^2 \leq C 2^{-n} \exp\left[C \exp(CT)/h\right] R^d.$$

Proof. Assume $T \geq 1$. Let $\varphi = \varphi_R : [0,\infty) \to [0,1]$ be C^1 and satisfying (3.1), and let $A := CT^2/h$ for some $C > 0$ to be determined. Then for $t \in \mathbb{N}_T^\tau$ set

$$E_{t,n} := \frac{1}{2} e^{At} \sum_{x \in \mathbb{Z}_h^d} |V_n(t,x) - V(t,x)|^2 \varphi(|x|)$$

which is finite. Direct computation yields

$$\begin{aligned}
\frac{E_{t,n} - E_{t-\tau,n}}{\tau} &\geq A e^{-A\tau} E_{t,n} + \frac{1}{2} e^{A(t-\tau)} \sum_{x \in \mathbb{Z}_h^d} (V_n(t,x) + V_n(t-\tau,x) - V(t,x) - V(t-\tau,x)) \times \\
&\qquad (\partial_t^\tau V_n(t,x) - \partial_t^\tau V(t,x)) \varphi(|x|) \\
&= A e^{-A\tau} E_{t,n} + e^{A(t-\tau)} \sum_{x \in \mathbb{Z}_h^d} (V_n(t,x) - V(t,x))(\partial_t^\tau V_n(t,x) - \partial_t^\tau V(t,x)) \varphi(|x|) \\
&\qquad - \frac{\tau}{2} e^{A(t-\tau)} \sum_{x \in \mathbb{Z}_h^d} \left|\partial_t^\tau V_n(t,x) - \partial_t^\tau V(t,x)\right|^2 \varphi(|x|) \\
&=: A e^{-A\tau} E_{t,n} + e^{A(t-\tau)} X_{t,n} - \frac{\tau}{2} e^{A(t-\tau)} Y_{t,n}.
\end{aligned} \tag{4.1}$$

First, we consider the term $Y_{t,n}$ (which does not appear in the semi-discretization problem in Theorem 3.1). It follows from the equations (2.18) and (2.20) that

$$Y_{t,n} = \sum_{x \in \mathbb{Z}_h^d} \left|c_n + \nabla^h V_n \cdot f_n + Nh\Delta^h V_n - H(t,x,\nabla^h V) - Nh\Delta^h V\right|^2 \varphi(|x|)$$

Recall that $\alpha = \alpha(t, x, \nabla^h V)$, $H(t, x, \nabla^h V) = c(t, x, \alpha) + f(t, x, \alpha) \cdot \nabla^h V$, and $|\nabla^h V| \le M$ for some $M \ge 1$. Since $\alpha_n = \alpha(t, x, \nabla^h V_{n-1})$, the regularity assumptions and (2.11) yield

$$Y_{t,n} \le C \sum_{x \in \mathbb{Z}_h^d} \Big(M^2 |D^h V_{n-1} - D^h V|^2 + M^2 |D^{-h} V_{n-1} - D^{-h} V|^2 + \\ |D^h V_n - D^h V|^2 + |D^{-h} V_n - D^{-h} V|^2 \Big) \varphi(|x|) \le C \left(M^2 G^h_{t,n-1} + G^h_{t,n} \right),$$

where in the last inequality we used the notation $G^h_{t,n} := \sum_{x \in \mathbb{Z}_h^d} |D^h V_n(t,x) - D^h V(t,x)|^2 \varphi(|x|)$. and (3.6) with the above defined $G^h_{t,n}$ (which clearly holds the same).

Next, we consider the term $X_{t,n}$. Note that for any $v \in L^\infty(\mathbb{Z}_h^d)$, $\sum_{x \in \mathbb{Z}_h^d} \Delta^h v\, v\varphi = -\sum_{x \in \mathbb{Z}_h^d} |D^h v|^2 \varphi - \sum_{x \in \mathbb{Z}_h^d} v D^{-h} v \cdot D^{-h} \varphi$. So, similarly as before (also using the equation, (3.1), (3.6), the uniform Lipschitz assumptions, and Young's inequality), we have for any $\sigma_1, \sigma_2 > 0$,

$$\begin{aligned} X_{t,n} &\ge Nh G^h_{t,n} - C \sum_{x \in \mathbb{Z}_h^d} |V_n - V| (|D^h (V_n - V)| + |D^{-h}(V_n - V)|) \varphi \\ &\quad - CM \sum_{x \in \mathbb{Z}_h^d} |V_n - V| (|D^h (V_{n-1} - V)| + |D^{-h}(V_{n-1} - V)|) \varphi \\ &\ge (Nh - \sigma_1) G^h_{t,n} - \sigma_2 G^h_{t,n-1} - C(\sigma_1^{-1} + M^2 \sigma_2^{-1}) e^{-At} E_{t,n}. \end{aligned}$$

Since $E_{T,n} \equiv 0$, putting the above together and summing up (4.1) with respect to t yield

$$\begin{aligned} -E_{t,n}/\tau &\ge (Nh - \sigma_1 - C\tau) \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,n} - (\sigma_2 + CM^2\tau) \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,n-1} \\ &\quad + (A - C\sigma_1^{-1} - CM^2 \sigma_2^{-1}) e^{-A\tau} \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} E^h_{s,n} \end{aligned} \tag{4.2}$$

for some universal constant $C > 0$.

Finally we take $\sigma_1 := h/4$, $\sigma_2 := h/8$, $A := 12CM^2/h$. Then if $\tau \le h/(8CM^2)$, (4.2) yields

$$\sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,n} \le \frac{1}{2} \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,n-1} \le \ldots \le 2^{-n} \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,0},$$

and then $E_{t,n} \le \frac{h\tau}{4} \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,n-1} \le \frac{h\tau}{2^{n+1}} \sum_{t+\tau \le s \in \mathbb{N}_T^\tau} e^{A(s-\tau)} G^h_{s,0}$. This, together with Proposition 2.5, concludes the proof of the first claim as before.

The second claim follows similarly as in Theorem 3.2. □

By shifting the solutions, we actually obtain uniform pointwise exponential convergence of $V_n^{\tau,h}$ to $V^{\tau,h}$ and $\alpha(\cdot, \cdot, \nabla^h V_n^{\tau,h})$ to $\alpha(\cdot, \cdot, \nabla^h V^{\tau,h})$ as $n \to \infty$, in $\Omega_T^{\tau,h}$.

4.2. **Convergence of $V^{\tau,h}$ as $\tau, h \to 0$.** Let $V^{\tau,h}$ and v be, respectively, solutions to (2.20) and (2.4). The following theorem proves that the difference between $V^{\tau,h}$ and v is at most of order $\sqrt{h}$. The argument follows the idea of [13, Theorem 1], which considered the discrete space-time scheme for the homogeneous Hamilton-Jacobi equation $v_t + H(Dv) = 0$.

Theorem 4.2. *Assume* (A1)–(A2) *and* (2.17)*. Then there exists a universal $C > 0$ such that*

$$\sup_{(t,x)\in\Omega_T^{\tau,h}} |v(t,x) - V^{\tau,h}(t,x)| \le C(1+T)(1+\|\nabla v\|_\infty)\sqrt{h}.$$

In particular, we have $\sup_{(t,x)\in\Omega_T^{\tau,h}} |v(t,x) - V^{\tau,h}(t,x)| \le C\exp(CT)\sqrt{h}$.

Remark 4.1. *It was shown in* [14, 15, 27] *that*

$$\sup_{(t,x)\in\Omega_T^{\tau,h}} |v(t,x) - V^{\tau,h}(t,x)| \le C(\tau^{1/4} + h^{1/2}) \quad \text{for some } C = C(T) > 0,$$

where v solves a general degenerate parabolic Bellman equation and $V^{\tau,h}$ is its space-time finite difference approximation. For the first order equations, our Theorem 4.2 obtains a better convergence rate.

Proof. Assume $T \ge 1$. And suppose for some $(t_0, x_0) \in \Omega_T^{\tau,h}$ such that

$$8\sigma := v(t_0,x_0) - V^{\tau,h}(t_0,x_0) \ge \frac{1}{2}\sup_{(t,x)\in\Omega_T^{\tau,h}} \Big[v(t,x) - V^{\tau,h}(t,x)\Big] > 0. \tag{4.3}$$

Let $D_{T,\tau,h} := [0,T] \times \mathbb{N}_T^\tau \times \mathbb{R}^d \times \mathbb{Z}_h^d$, and

$$L := \sup\Big\{v(t,x), -V^{\tau,h}(t,x) \,:\, (t,x) \in \Omega_T^{\tau,h}\Big\} + 1.$$

Then $\sigma \le L \le CT$ for some universal constant $C > 0$. Moreover, let R, g and g_ε with $\varepsilon \in (0,1)$, and ϕ be from the proof of Theorem 3.3, and define $\Phi^h : D_{T,\tau,h} \to \mathbb{R}$ by

$$\begin{aligned}\Phi^h(t,s,x,y) := v(t,x) - V^{\tau,h}(s,y) - \frac{\sigma}{T}(2T - t - s)\\ - \frac{\sigma}{R}(\phi(x) + \phi(y)) + (8L + 2\sigma)g_\varepsilon(t-s, x-y).\end{aligned}$$

Suppose

$$\Phi^h(t_1,s_1,x_1,y_1) = \max_{D_{T,\tau,h}} \Phi^h(t,s,x,y). \tag{4.4}$$

It is clear that (3.10)–(3.14) hold the same. By (3.14) if $\tau \ll \varepsilon^2 M/L$, we get

$$|t_1 - s_1 - \tau| \le C\varepsilon^2 M/L \quad \text{with } M = 1 + \|\nabla v\|_\infty. \tag{4.5}$$

First, assume $t_1, s_1 < T$. The viscosity solution test for v shows (3.15) by (3.12). Next since $\Omega_T^{\tau,h} \ni (s,y) \to V^{\tau,h}(s,y) - \frac{\sigma}{T}s + \frac{\sigma}{R}\phi(y) - (8L+2\sigma)g_\varepsilon(t_1 - s, x_1 - y)$ is minimized at (s_1,y_1), then for all $(s,y) \in \Omega_T^{\tau,h}$,

$$\begin{aligned}V^{\tau,h}(s,y) \ge V^{\tau,h}(s_1,y_1) - \frac{\sigma}{T}(s_1 - s) + \frac{\sigma}{R}(\phi(y_1) - \phi(y))\\ - (8L+2\sigma)\left[g_\varepsilon(t_1 - s_1, x_1 - y_1) - g_\varepsilon(t_1 - s, x_1 - y)\right] =: \tilde{V}(s,y).\end{aligned}$$

Recall that $s_1 + \tau \le T$ and $\mathcal{F}_t$ from (2.21) satisfies the monotonicity property. We obtain

$$V^{\tau,h}(s_1,y_1) = \mathcal{F}_{s_1+\tau}(V^{\tau,h}(s_1+\tau,\cdot))(y_1) \ge \mathcal{F}_{s_1+\tau}(\tilde{V}(s_1+\tau,\cdot))(y_1),$$

which gives

$$
\begin{aligned}
0 \geq \frac{\sigma}{T} &- (8L+2\sigma)\, \partial_t^{\tau} g_\varepsilon(t_1-s_1, x_1-y_1) \\
&+ H\left(s_1+\tau, y_1, -\frac{\sigma}{R}\nabla^h\phi(y_1) - (8L+2\sigma)\nabla_x^h\, g_\varepsilon(t_1-s_1-\tau, x_1-y_1)\right) \\
&- Nh\Delta^h\left[\frac{\sigma}{R}\phi(y_1) - (8L+2\sigma) g_\varepsilon(t_1-s_1-\tau, x_1-y_1))\right].
\end{aligned} \tag{4.6}
$$

By (3.11), if $\tau, h \ll \varepsilon^2$,

$$
|\partial_t^{\tau}\, g_\varepsilon(t_1-s_1, x_1-y_1) - \partial_t g_\varepsilon(t_1-s_1, x_1-y_1)| \leq C\varepsilon^{-2}\tau, \tag{4.7}
$$

$$
\nabla_x^h\, g_\varepsilon(t_1-s_1-\tau, x_1-y_1) = \nabla_x\, g_\varepsilon(t_1-s_1, x_1-y_1) = 2\varepsilon^{-2}(x_1-y_1). \tag{4.8}
$$

Combining (4.6) with (3.15), and using (4.7) and (4.8) yield

$$
\begin{aligned}
\frac{2\sigma}{T} \leq H&\left(t_1, x_1, \frac{\sigma}{R}\nabla\phi(x_1) - (8L+2\sigma)2\varepsilon^{-2}(x_1-y_1)\right) \\
&- H\left(s_1+\tau, y_1, -\frac{\sigma}{R}\nabla\phi(y_1) - (8L+2\sigma)2\varepsilon^{-2}(x_1-y_1)\right) \\
&+ Nh\Delta^h\left[\frac{\sigma}{R}\phi(y_1) - (8L+2\sigma) g_\varepsilon(t_1-s_1-\tau, x_1-y_1))\right] + CL\varepsilon^{-2}\tau.
\end{aligned} \tag{4.9}
$$

The definitions of ϕ and g_ε show (3.18). Then, applying (2.16) and (3.18) into (4.9), if $(\tau \leq\ ) h \ll \varepsilon^2$ we deduce for some $C > 0$ that

$$
\begin{aligned}
\sigma T^{-1} &\leq C\sigma R^{-1}(|\nabla\phi(x_1)| + |\nabla\phi(y_1)|) + CL\varepsilon^{-2}h + CL\varepsilon^{-2}\tau \\
&\qquad + C(|t_1-s_1-\tau| + |x_1-y_1|)\left[1 + (8L+2\sigma)2\varepsilon^{-2}|x_1-y_1|\right] \\
&\leq C\sigma R^{-1} + CL\varepsilon^{-2}h + C\varepsilon^2 M^2 L^{-1}
\end{aligned} \tag{4.10}
$$

where in the second inequality we also used (3.13) and (4.5).

Now we take $\varepsilon := M^{-1/2}L^{1/2}h^{1/4}$, and send $R \to \infty$. It is clear that $\tau \ll \varepsilon^2 M/L$ is satisfied when h is small. We obtain from (4.10) that $\sigma \leq CTM\sqrt{h}$, which finishes the proof of the upper bound of $\sup_{\Omega_T^{\tau,h}}(v - V^{\tau,h})$ in the case when $t_1, s_1 < T$.

Next, if at least one of t_1 and s_1 equals to T, the argument of Theorem 3.3 applies the same except that we need to use Proposition 2.5 in place of Lemma 2.4. Finally, the proof for the upper bound of $\sup_{\Omega_T^{\tau,h}}(V^{\tau,h} - v)$ is the same. □

4.3. **Almost everywhere convergence of the policy.** We show the almost everywhere convergence of the policy, and some semi-concavity property of the solution.

Theorem 4.3. *Under the assumptions of Theorem 4.2, further assume* (A3). *Then v is uniformly semi-concave for all $t \in [0,T]$. Moreover, for each $t \in [0,T]$ we have for a.e. $x \in \mathbb{R}^d$,*

$$
\alpha(t_h, x_h, \nabla^h V^{\tau_h,h}(t_h, x_h)) \to \alpha(t, x, \nabla v(t,x)) \quad \text{as } h \to 0
$$

where $\Omega_T^{\tau_h,h} \ni (t_h, x_h) \to (t,x)$ as $h \to 0$ and τ_h satisfies $0 < 2N\tau_h \leq h$.

Proof. The semi-concavity of $v(t,\cdot)$ follows from [23].

For the second claim, it suffices to prove that for a fixed $t \in [0,T)$, and for a.e. $x \in \mathbb{R}^d$,

$$
\nabla^h V^{\tau_h,h}(t_h, x_h) \to \nabla v(t,x) \quad \text{as } h \to 0. \tag{4.11}
$$

For any function $g : \mathbb{R}^d \to \mathbb{R}$, let us denote by $D^+ g(x)$ the set of subdifferential of g:

$$D^+ g(x) := \left\{ p \in \mathbb{R}^d \,\Big|\, \limsup_{y \to x} \frac{g(y) - g(x) - p \cdot (y - x)}{|y - x|} \le 0 \right\}.$$

Due to $v(t, \cdot)$ is semi-concave, $D^+ v(t, x)$ is non-empty for all $x \in \mathbb{R}^d$.

Because $v(t, \cdot)$ is Lipschitz continuous, $\nabla_x v(t, x)$ exists for a.e. $x \in \mathbb{R}^d$. We fix one such x. Since $V^{\tau_h,h}$ are Lipschitz continuous uniformly in h, after passing to a subsequence of $h \to 0$, we can assume that $\nabla^h V^{\tau_h,h}(t_h, x_h) \to p$ for some $p \in \mathbb{R}^d$. Since $V^{\tau_h,h}(t_h, x_h) \to v(t, x)$ as $h \to 0$, the stability of subdifferential yields that $p \in D^+ v(t, x)$. While because $\nabla_x v(t, x)$ exists, we get $p = \nabla_x v(t, x)$. Note that this is for any convergent subsequence of $\nabla^h V^{\tau_h,h}(t_h, x_h)$, and so we obtain (4.11). □

Below, we show a weak type of semi-concavity of $V^{\tau,h}(t, \cdot)$. We adopt the "doubling variable" method, see e.g., [23].

Theorem 4.4. *Under the assumptions of Theorem 4.3, there exists $C > 0$ (also depending on* (A3)*) such that for all $t \in \mathbb{N}^\tau_T$ and $x, y \in \mathbb{Z}^d_h$,*

$$V^{\tau,h}(t, x + y) + V^{\tau,h}(t, x - y) - 2V^{\tau,h}(t, x) \le C \exp(CT)\, (|y|^2 + \sqrt{h}).$$

Proof. It suffices to show that there exist $C_T, C'_T > 0$ depending on the assumptions such that

$$V^{\tau,h}(t, x) + V^{\tau,h}(t, z) - 2V^{\tau,h}(t, y) \le C_T \left(|x - y|^2 + |z - y|^2 + |x + z - 2y| \right) + C'_T \sqrt{h} \tag{4.12}$$

for all $t \in \mathbb{N}^\tau_T$ and $x, y, z \in \mathbb{Z}^d_h$. By the assumption on q, the inequality holds for $t = T$ with $C_T = \|q\|_{W^{2,\infty}} =: C_0$, and $C'_T = 0$.

Suppose for contradiction that (4.12) fails. Then we have for some $C_1 \ge 1$ to be determined, and some $C \ge 2$,

$$\begin{aligned} &V^{\tau,h}(t, x) + V^{\tau,h}(t, z) - 2V^{\tau,h}(t, y) \\ &\qquad - 2C_0 e^{C_1(T-t)} \left(|x - y|^4 + |z - y|^4 + |x + z - 2y|^2 \right)^{1/2} \ge C e^{C_1(T-t)} \sqrt{h} \end{aligned} \tag{4.13}$$

for some $(t, x, y, z) = (t', x', y', z') \in \mathbb{N}^\tau_T \times \mathbb{Z}^d_h$. Since $V^{\tau,h}(t, \cdot)$ is Lipschitz continuous (with Lipschitz constant bounded by $C \exp(C(T - t))$ by Proposition 2.5 with a shift in time), after possibly enlarging the constant C in (4.13), we can assume that

$$|x' + z' - 2y'| \ge \sqrt{h}. \tag{4.14}$$

Let us denote $\psi(x, y, z) := |x - y|^4 + |z - y|^4 + |x + z - 2y|^2$, and by (4.14), $\delta := \psi(x', y', z')^{1/2} \ge \sqrt{h}$. Then for all $\varepsilon > 0$ sufficiently small, we obtain from (4.13) that

$$\Phi(t, x, y, z) := e^{C_1 t} \left(V^{\tau,h}(t, x) + V^{\tau,h}(t, z) - 2V^{\tau,h}(t, y) \right) - C_0 e^{C_1 T} \left(\delta + \delta^{-1} \psi(x, y, z) \right) - \varepsilon |y|^2$$

satisfies $\Phi(t', x', y', z') \ge e^{C_1 T} \sqrt{h}$. With the positive ε-term, Φ obtains its positive maximum that is at least $e^{C_1 T} \sqrt{h}$ in $\Omega^{\tau,h}_T$ at some point $(t_0, x_0, y_0, z_0) \in \mathbb{N}^\tau_T \times \mathbb{Z}^d_h$, where (t_0, x_0, y_0, z_0) depends on ε and δ. It is clear that $t_0 \le T - \tau$ by the choice of C_0. Moreover, for $\gamma_0 := \delta + \delta^{-1} \psi(x_0, y_0, z_0)$, we have

$$V^{\tau,h}(t_0, x_0) + V^{\tau,h}(t_0, z_0) - 2V^{\tau,h}(t_0, y_0) \ge C_0 e^{C_1(T-t_0)} \gamma_0 + e^{C_1(T-t_0)} \sqrt{h}. \tag{4.15}$$

Due to uniform boundedness of $V^{\tau,h}$, by further taking ε to be small enough depending on C, T and h, it is easy to get $\varepsilon |y_0| \le h$.

Now since $\Omega_T^{\tau,h} \ni (t,x) \to e^{C_1 t} V^{\tau,h}(t,x) - C_0 e^{C_1 T}\delta^{-1}\left(|x-y_0|^4 + |x+z_0-2y_0|^2\right)$ is maximized at (t_0,x_0), we get for all $(t,x) \in \Omega_T^{\tau,h}$ that

$$\begin{aligned} V^{\tau,h}(t,x) \le e^{C_1(t_0-t)} V^{\tau,h}(t_0,x_0) + C_0 e^{C_1(T-t)}\delta^{-1}\left(|x-y_0|^4 + |x+z_0-2y_0|^2\right) \\ - C_0 e^{C_1(T-t_0)}\delta^{-1}\left(|x_0-y_0|^4 + |x_0+z_0-2y_0|^2\right) =: \tilde V(t,x). \end{aligned}$$

Due to the equation and the monotonicity property of $\mathcal{F}_t$ (defined in (2.21)), $V^{\tau,h}(t_0,x_0) = \mathcal{F}_{t_0+\tau}(V^{\tau,h}(t_0+\tau,\cdot))(x_0) \le \mathcal{F}_{t_0+\tau}(\tilde V(t_0+\tau,\cdot))(x_0)$. By direct computation,

$$\nabla_x^h(|x-y_0|^4 + |x+z_0-2y_0|^2) = 4(|x-y_0|^2+h^2)(x-y_0) + 2(x+z_0-2y_0),$$

$$\Delta_x^h(|x-y_0|^4 + |x+z_0-2y_0|^2) = (8+4d)|x-y_0|^2 + 2dh^2 + 2d.$$

We then get

$$\begin{aligned} \frac{(1-e^{-C_1\tau})}{\tau} V^{\tau,h}(t_0,x_0) &\le H\left(t_0+\tau, x_0, \nabla_x^h \tilde V(t_0+\tau,x_0)\right) + Nh\Delta_x^h \tilde V(t_0+\tau,x_0) \\ &\le H\left(t_0+\tau, x_0, 2C_{T,\delta}(q_{x_0}+p_0)\right) + CC_{T,\delta}h(|x_0-y_0|^2+1) \end{aligned} \tag{4.16}$$

where $q_{x_0} := 2(|x_0-y_0|^2+h^2)(x_0-y_0)$,

$$C_{T,\delta} := C_0 e^{C_1(T-t_0-\tau)}/\delta \quad \text{and} \quad p_0 := x_0+z_0-2y_0. \tag{4.17}$$

Similarly, since $\Omega_T^{\tau,h} \ni (t,z) \to e^{C_1 t}V^{\tau,h}(t,z) - C_0 e^{C_1 T}\delta^{-1}(|z-y_0|^4 + |x_0+z-2y_0|^2)$ is maximized at (t_0,z_0), we get

$$\frac{(1-e^{-C_1\tau})}{\tau} V^{\tau,h}(t_0,z_0) \le H\left(t_0+\tau, z_0, 2C_{T,\delta}(q_{z_0}+p_0)\right) + CC_{T,\delta}h(|z_0-y_0|^2+1). \tag{4.18}$$

where $q_{z_0} := 2(|z_0-y_0|^2+h^2)(z_0-y_0)$.

Next, note that $\Omega_T^{\tau,h} \ni (t,y) \to 2e^{C_1 t}V^{\tau,h}(t,y) + C_0 e^{C_1 T}\delta^{-1}\psi(x_0,y,z_0) + \varepsilon|y|^2$ is minimized at (t_0,y_0). Hence we get $V^{\tau,h}(t_0,y_0) \ge \mathcal{F}_{t_0+\tau}(\hat V(t_0+\tau,\cdot))(y_0)$ where

$$\begin{aligned} \hat V(t,y) := e^{C_1(t_0-t)}V^{\tau,h}(t_0,y_0) - (\varepsilon/2)|y|^2 + (\varepsilon/2)|y_0|^2 \\ - (C_0/2)e^{C_1(T-t)}\delta^{-1}\psi(x_0,y,z_0) + (C_0/2)e^{C_1(T-t_0)}\delta^{-1}\psi(x_0,y_0,z_0). \end{aligned}$$

From this we obtain

$$\begin{aligned} -\frac{(1-e^{-C_1\tau})}{\tau} V^{\tau,h}(t_0,y_0) \le -H\left(t_0+\tau, y_0, 2C_{T,\delta}(q_{y_0}+p_0) - \varepsilon y_0\right) \\ + CC_{T,\delta}h(|x_0-y_0|^2 + |z_0-y_0|^2 + 1) + Ch\varepsilon \end{aligned}$$

where $q_{y_0} := (|x_0-y_0|^2+h^2)(x_0-y_0) + (|z_0-y_0|^2+h^2)(z_0-y_0)$, and $C_{T,\delta}$ and p_0 are given in (4.17). Using $|H_p| \le C$ and $\varepsilon|y_0| \le h$ yields

$$\begin{aligned} -\frac{(1-e^{-C_1\tau})}{\tau} V^{\tau,h}(t_0,y_0) \le -H\left(t_0+\tau, y_0, 2C_{T,\delta}(q_{y_0}+p_0)\right) \\ + CC_{T,\delta}h(|x_0-y_0|^2 + |z_0-y_0|^2 + 1) + Ch \end{aligned} \tag{4.19}$$

Now let $\alpha \in \mathcal{A}$ be such that

$$H\left(t_0+\tau, y_0, 2C_{T,\delta}(q_{y_0}+p_0)\right) = c(t_0+\tau, y_0, \alpha) + 2C_{T,\delta} f(t_0+\tau, y_0, \alpha)\cdot(q_{y_0}+p_0).$$

By (2.15), denoting $c_\alpha(\cdot) := c(t_0+\tau,\cdot,\alpha)$ and $f_\alpha(\cdot) := f(t_0+\tau,\cdot,\alpha)$, we have

$$\begin{aligned}
&H\left(t_0+\tau, x_0, 2C_{T,\delta}(q_{x_0}+p_0)\right) + H\left(t_0+\tau, z_0, 2C_{T,\delta}(q_{z_0}+p_0)\right)\\
&\qquad - 2H\left(t_0+\tau, y_0, 2C_{T,\delta}(q_{y_0}+p_0)\right)\\
&\le c_\alpha(x_0)+c_\alpha(z_0)-2c_\alpha(y_0)+2C_{T,\delta}\left[f_\alpha(x_0)\cdot(q_{x_0}+p_0)\right.\\
&\qquad \left.+\ f_\alpha(z_0)\cdot(q_{x_0}+p_0)-2f_\alpha(y_0)\cdot(q_{y_0}+p_0)\right]\\
&= c_\alpha(x_0)+c_\alpha(z_0)-2c_\alpha(y_0)+2C_{T,\delta}\left[(f_\alpha(x_0)-f_\alpha(y_0))\cdot q_{x_0}+(f_\alpha(z_0)-f_\alpha(y_0))\cdot q_{z_0}+\right.\\
&\qquad \left.+\ (f_\alpha(x_0)+f_\alpha(z_0)-2f_\alpha(y_0))\cdot p_0\right]\\
&\le \|c_\alpha\|_{W^{2,\infty}}(|x_0-y_0|^2+|z_0-y_0|^2+|x_0+z_0-2y_0|)\\
&\qquad + 2C_{T,\delta}\|f_\alpha\|_{\mathrm{Lip}}(|x_0-y_0||q_{x_0}|+|z_0-y_0||q_{z_0}|)\\
&\qquad + 2C_{T,\delta}\|f_\alpha\|_{W^{2,\infty}}(|x_0-y_0|^2+|z_0-y_0|^2+|x_0+z_0-2y_0|)|x_0+z_0-2y_0|,
\end{aligned} \tag{4.20}$$

where we used $2q_{y_0} = q_{x_0}+q_{z_0}$ and that for any $x,y,z\in\mathbb{R}^d$ and $g\in W^{2,\infty}(\mathbb{R}^d)$, $|g(x)+g(z)-2g(y)| \le \|g\|_{W^{2,\infty}}(|x-y|^2+|z-y|^2+|x+z-2y|)$. By Young's inequality, we get $|x_0-y_0||q_{x_0}|+|z_0-y_0||q_{z_0}| \le 2|x_0-y_0|^4+2|z_0-y_0|^4+h^4$. Also using the definitions of $C_{T,\delta}$ and ψ, we get the left-hand side of (4.20) $\le Ce^{C_1(T-t_0)}(\delta+\delta^{-1}\psi(x_0,y_0,z_0)) = Ce^{C_1(T-t_0)}\gamma_0+Ce^{C_1(T-t_0)}h^4/\delta$ with $C>0$ only depending on $\|q\|_{W^{2,\infty}}$, $\|c_\alpha\|_{W^{2,\infty}}$ and $\|f_\alpha\|_{W^{2,\infty}}$.

Now summing up (4.16), (4.18) and twice of (4.19), we get

$$\begin{aligned}
&\frac{(1-e^{-C_1\tau})}{\tau}\left[V^{\tau,h}(t_0,x_0)+V^{\tau,h}(t_0,z_0)-2V^{\tau,h}(t_0,y_0)\right]\\
&\qquad\le Ce^{C_1(T-t_0)}\gamma_0+Ce^{C_1(T-t_0)}h^4/\delta+CC_{T,\delta}h(|x_0-y_0|^2+|z_0-y_0|^2+1)+Ch\\
&\qquad\le Ce^{C_1(T-t_0)}\gamma_0+Ce^{C_1(T-t_0)}\delta^{-1}(|x_0-y_0|^4+|z_0-y_0|^4)+Ce^{C_1(T-t_0)}\sqrt{h}\\
&\qquad\le Ce^{C_1(T-t_0)}\gamma_0+Ce^{C_1(T-t_0)}\sqrt{h},
\end{aligned}$$

where in the second inequality, we used $\delta\ge\sqrt{h}$. Finally, this and (4.15) yield

$$C_1(C_0e^{C_1(T-t_0)}\gamma_0+e^{C_1(T-t_0)}\sqrt{h}) \le Ce^{C_1(T-t_0)}\gamma_0+Ce^{C_1(T-t_0)}\sqrt{h},$$

with $C>0$ depending only on d, N and the regularity assumptions of q, c, f. Thus, if C_1 is sufficiently large depending only on the assumptions, we get a contradiction which finishes the proof of (4.12), which finishes the proof. □

5. Generalization: a PDE perspective

In this section, we consider PI for HJB equations with a general Hamiltonian. For convenient use of the Legendre transform, we write the system in the forward-in-time setting. It is easy to carry over to the backward-in-time setting.

Suppose $\mathcal{H} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ is continuous such that $\mathcal{H}(t,x,p)$ is convex in p. Let $\mathcal{L}(t,x,\mu)$ be the Legendre transform of $\mathcal{H}$, that is,

$$\mathcal{L}(t,x,\mu) := \sup_{p\in\mathbb{R}^d}\left[p\cdot\mu-\mathcal{H}(t,x,p)\right] \qquad \text{for } (t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}^d.$$

We always have the following inequality $\mathcal{L}(t,x,\mu)+\mathcal{H}(t,x,p)\ge p\cdot\mu$, with equality holds if and only if $\mu=\nabla_p\mathcal{H}(t,x,p)$, and if and only if $p=\nabla_\mu\mathcal{L}(t,x,\mu)$.

The HJB equation is

$$\begin{cases} \partial_t v + \mathcal{H}(t,x,\nabla v) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ v(0,x) = q(x) & \text{on } \mathbb{R}^d. \end{cases} \tag{5.1}$$

Under some assumptions (see [3, 42]), it is a classical result that v is uniformly Lipschitz continuous if q is Lipschitz continuous. So we can assume

$$\|\nabla v\|_{L^\infty([0,T]\times\mathbb{R}^d)} \le M \quad \text{for some } M > 0. \tag{5.2}$$

Now we take $m_1 := \min_{\substack{|p|=2M,\\ t\in[0,T],x\in\mathbb{R}^d}} \mathcal{H}(t,x,p)$ and $m_2 \ge \max_{\substack{|p|=3M,\\ t\in[0,T],x\in\mathbb{R}^d}} [\mathcal{H}(t,x,p) - m_1]/M$, and we can assume that $m_2 \ge 2$. Then define

$$\tilde{\mathcal{H}}(t,x,p) := \begin{cases} \mathcal{H}(t,x,p) & \text{if } |p|\le 2M, \\ \max\{\mathcal{H}(t,x,p), m_1 + m_2(|p|-2M)\} & \text{if } 2M < |p| \le 3M, \\ m_1 + m_2(|p|-2M) & \text{if } |p| > 3M. \end{cases}$$

It is not hard to verify that $\tilde{\mathcal{H}}$ is continuous in all its dependencies, and is convex in p. Due to (5.2), v is also a solution of (5.1) with $\mathcal{H}$ replaced by $\tilde{\mathcal{H}}$. Moreover for $N := m_2/2 \ge 1$, we have

$$|\tilde{\mathcal{H}}_p(t,x,p)| \le 2N \quad \text{in } [0,T]\times\mathbb{R}^d\times\mathbb{R}^d. \tag{5.3}$$

We define $\tilde{\mathcal{L}}$ as the Legendre transform of $\tilde{\mathcal{H}}$. Since the goal is to approximate v, it suffices to study $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{L}}$ instead of $\mathcal{H}$ and $\mathcal{L}$. From now on, with a slight abuse of notation, we write $\mathcal{H}$ and $\mathcal{L}$ as $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{L}}$, respectively.

With the modified operators, we can consider the semi-discretization. For $h > 0$,

$$\begin{cases} \partial_t v^h + \mathcal{H}(t,x,\nabla^h v^h) = Nh\Delta^h v^h & \text{in } (0,T)\times\mathbb{R}^d, \\ v^h(0,x) = q(x) & \text{on } \mathbb{R}^d. \end{cases} \tag{5.4}$$

As before, $N \ge \|\nabla_p \mathcal{H}\|_\infty/2$ guarantees that the finite difference scheme is monotone. Let us also assume that there exists $C > 0$ such that for all $t,x,p \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d$,

$$|\mathcal{H}_t(t,x,p)|,\ |\mathcal{H}_x(t,x,p)| \le C(1+|p|), \quad |\mathcal{H}(t,x,0)| \le C. \tag{5.5}$$

Actually, we can replace $C(1+|p|)$ by just C for the modified operator. We will not discuss the space-time discretization of (5.1) since it is similar.

Now we present the iteration scheme for (5.4). Fixing small $h > 0$, we start with a uniformly bounded and Lipschitz continuous function $v_0^h(t,x)$, and then iteratively compute v_n^h as follows. For $n \ge 1$, let $v_n^h(t,x)$ be the solution to

$$\begin{cases} \partial_t v_n^h + \nabla_p \mathcal{H}(t,x,\nabla^h v_{n-1}^h)\cdot\nabla^h v_n^h - \mathcal{L}(t,x,\mu_{n-1}^h) = Nh\Delta^h v_n^h & \text{in } (0,T)\times\mathbb{R}^d, \\ v_n^h(0,x) = q(x) & \text{on } \mathbb{R}^d \end{cases} \tag{5.6}$$

where we denoted $\mu_n^h(t,x) := \nabla_p\mathcal{H}(t,x,\nabla^h v_n^h)$. Note $\mathcal{L}(t,x,\mu_n^h)$ is finite due to $\mu_n^h \le 2N$. Essentially, v_n^h solves a linearized equation of (5.4).

Let v_n^h (for each $n \ge 1$ with given v_0^h), v^h and v be, respectively, Lipschitz continuous solutions to (5.6), (5.4) and (5.1). We have the following monotonicity property.

Proposition 5.1. *Suppose $N \geq \max\{1, \|\nabla_p \mathcal{H}\|_\infty/2\}$, and $\mathcal{H}(t,x,p)$ is convex in p and satisfies (5.3) and (5.5). Let q and v_0^h be uniformly bounded and Lipschitz continuous for all $h > 0$. Then the solutions v_n^h are uniformly bounded for all $n \geq 1$ and $h > 0$. Moreover, we have for all $n \geq 0$,*

$$v_{n+1}^h \leq v_n^h \quad \text{in } [0,T] \times \mathbb{R}^d.$$

We also have the following convergence results.

Theorem 5.2. *Under the assumptions of Proposition 5.1, for all $R \geq 1$ there exists a constant C depending only on T and the assumptions such that we have for all $t \in [0,T]$,*

$$\int_{B_R} \left| v_n^h(t,x) - v^h(t,x) \right|^2 dx \leq C 2^{-n} h e^{Ct/h} R^d,$$

$$\int_{B_R} \left| \alpha(t,x,\nabla^h v_n^h(t,x)) - \alpha(t,x,\nabla^h v^h(t,x)) \right|^2 dx \leq C 2^{-n} e^{Ct/h} R^d / h.$$

Moreover, we have $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |v^h(t,x) - v(t,x)| \leq C\sqrt{h}$.

Next, let $\mathcal{H}$ take the form $\mathcal{H}(t,x,p) := \sup_{a\in A} [c(t,x,a) + p \cdot f(t,x,a)]$, where A is some set, $c : [0,T] \times \mathbb{R}^d \times A \to \mathbb{R}$ and $f : [0,T] \times \mathbb{R}^d \times A \to \mathbb{R}^d$.

Theorem 5.3. *Under the assumptions of Theorem 5.2, assume that $c(t,\cdot,a), f(t,\cdot,a)$ are bounded in $W^{2,\infty}(\mathbb{R}^d)$ uniformly for all $t \in [0,T]$ and $a \in A$. Then for each $t \in [0,T]$, we have for a.e. $x \in \mathbb{R}^d$,*

$$\alpha(t_h, x_h, \nabla^h v^h(t_h, x_h)) \to \alpha(t, x, \nabla v(t,x)) \quad \text{as } h \to 0,$$

where $[0,T] \times \mathbb{R}^d \ni (t_h, x_h) \to (t,x)$ as $h \to 0$.

Moreover, there exists $C > 0$ depending only on the assumptions such that for all $h \in (0,1)$, $t \in [0,T]$ and $x, y \in \mathbb{R}^d$, $v^h(t,x+y) + v^h(t,x-y) - 2v^h(t,x) \leq C\exp(CT)(|y|^2 + \sqrt{h})$.

6. Conclusion

In this paper, we study the convergence rate of PI for optimal control problems in continuous time. To overcome the problem of ill-posedness, we consider a semi-discrete scheme by adding a viscosity term using finite differences. We prove that PI for the semi-discrete scheme converges exponentially fast, and provide a bound on the discrepancy between the semi-discrete scheme and the optimal control. We also study the discrete space-time scheme, where both space and time are discretized.

There are a few directions to extend this work. First, under what conditions on the model parameters does PI (2.7)–(2.8) converge exponentially fast? For instance, for $f(t,x,a) = a$, $c(t,x,a) = \frac{1}{2}|a|^2$ and $q \equiv 0$, the HJB equation is $\partial_t v - \frac{1}{2}|\nabla v|^2 = 0$ and $v(T,x) = 0$, which has the solution $v_* \equiv 0$. On the other hand, PI yields $v_n(t,x) = c_n(t)x^2$ with $c_1(t) = \frac{1}{2}$ for a suitable initialization. It is easy to check that $c_n(t) \leq 2^{-n}$ for $n \geq 1$, and thus we get the exponential convergence of v_n to v_* on any compact set. However, it is not clear what are the right conditions to impose on the model parameters so that PI converges exponentially fast. It is also interesting to adapt PI to the differential game setting and design efficient numerical schemes (see e.g. [21]). We refer to [25, 39] for the use of PI to solve numerically fully nonlinear HJB and HJBI equations.

References

[1] M. Abu-Khalaf and F. L. Lewis. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 41(5):779–791, 2005.

[2] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin. Hamilton-Jacobi reachability: a brief overview and recent advances. In *56th Annual Conference on Decision and Control (CDC)*, pages 2242–2253, 2017.

[3] G. Barles. Regularity results for first order Hamilton-Jacobi equations. *Differential Integral Equations*, 3(1):103–125, 1990.

[4] G. Barles. Uniqueness and regularity results for first-order Hamilton-Jacobi equations. *Indiana Univ. Math. J.*, 39(2):443–466, 1990.

[5] R. Bellman. *Dynamic programming.* Princeton University Press, 1957.

[6] J. Z. Ben-Asher. *Optimal Control Theory with Aerospace Applications.* AIAA Education Series. AIAA, 2010.

[7] D. P. Bertsekas. *Dynamic programming and optimal control. Vol. II. Approximate dynamic programming.* Athena Scientific, fourth edition, 2012.

[8] D. P. Bertsekas. Value and policy iterations in optimal control and adaptive dynamic programming. *IEEE Trans. Neural Netw. Learn. Syst.*, 28(3):500–509, 2017.

[9] T. Bian and Z.-P. Jiang. Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design. *Automatica*, 71:348–360, 2016.

[10] L. Blackmore, B. Açikmeşe, and D. P. Scharf. Minimum-landing-error powered-descent guidance for mars landing using convex optimization. *J. Guid. Control Dyn*, 33(4):1161–1171, 2010.

[11] F. Cagnetti, D. Gomes, and H. V. Tran. Convergence of a semi-discretization scheme for the Hamilton-Jacobi equation: a new approach with the adjoint method. *Appl. Numer. Math.*, 73:2–15, 2013.

[12] M. Chen and C. J. Tomlin. Hamilton-Jacobi reachability: some recent theoretical advances and applications in unmanned airspace management. *Annu. Rev. Control Robot. Auton. Syst.*, 1(1):333–358, 2018.

[13] M. G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comp.*, 43(167):1–19, 1984.

[14] H. Dong and N. V. Krylov. Rate of convergence of finite-difference approximations for degenerate linear parabolic equations with C^1 and C^2 coefficients. *Electron. J. Differential Equations*, pages No. 102, 25, 2005.

[15] H. Dong and N. V. Krylov. The rate of convergence of finite-difference approximations for parabolic Bellman equations with Lipschitz coefficients in cylindrical domains. *Appl. Math. Optim.*, 56(1):37–66, 2007.

[16] L. C. Evans. On solving certain nonlinear partial differential equations by accretive operator methods. *Israel J. Math.*, 36(3-4):225–247, 1980.

[17] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability.* Springer, second edition, 2006.

[18] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *35th International Conference on Machine Learning (ICML)*, pages 1861–1870, 2018.

[19] B. Hambly, R. Xu, and H. Yang. Policy gradient methods for the noisy linear quadratic regulator over a finite horizon. *SIAM J. Control Optim.*, 59(5):3359–3391, 2021.

[20] R. A. Howard. *Dynamic programming and Markov processes.* Technology Press of M.I.T.; John Wiley & Sons, Inc., 1960.

[21] K. Huang, X. Di, Q. Du, and X. Chen. A game-theoretic framework for autonomous vehicles velocity control: bridging microscopic differential games and macroscopic mean field games. *Discrete Contin. Dyn. Syst. Ser. B*, 25(12):4869–4903, 2020.

[22] Y.-J. Huang, Z. Wang, and Z. Zhou. Convergence of policy improvement for entropy-regularized stochastic control problems. 2022. arXiv:2209.07059.

[23] H. Ishii and P.-L. Lions. Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. *J. Differential Equations*, 83(1):26–78, 1990.

[24] M. I. Kamien and N. L. Schwartz. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, volume 31 of *Advanced Textbooks in Economics.* Elsevier, second edition, 1991.

[25] E. L. Kawecki and T. Sprekeler. Discontinuous Galerkin and C^0-IP finite element approximation of periodic Hamilton-Jacobi-Bellman-Isaacs problems with application to numerical homogenization. *ESAIM Math. Model. Numer. Anal.*, 56(2):679–704, 2022.

[26] B. Kerimkulov, D. Šiška, and L. Szpruch. Exponential convergence and stability of Howard's policy improvement algorithm for controlled diffusions. *SIAM J. Control Optim.*, 58(3):1314–1340, 2020.
[27] N. V. Krylov. The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients. *Appl. Math. Optim.*, 52(3):365–399, 2005.
[28] J. Lee and R. S. Sutton. Policy iterations for reinforcement learning problems in continuous time and space–fundamental theory and methods. *Automatica*, 126:Paper No. 109421, 15, 2021.
[29] F. L. Lewis and D. Vrabie. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits Syst. Mag.*, 9(3):32–50, 2009.
[30] Y. Li, K. H. Johansson, J. Mårtensson, and D. P. Bertsekas. Data-driven rollout for deterministic optimal control, 2021. arXiv:2105.03116.
[31] E. Nolasco, V. S. Vassiliadis, W. Kähm, S. D. Adloor, R. Al Ismaili, R. Conejeros, T. Espaas, N. Gangadharan, V. Mappas, and F. Scott. Optimal control in chemical engineering: Past, present and future. *Comput. Chem. Eng.*, 155:107528, 2021.
[32] S. Osher and C.-W. Shu. High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations. *SIAM J. Numer. Anal.*, 28(4):907–922, 1991.
[33] W. B. Powell. *Approximate dynamic programming: solving the curses of dimensionality*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc, 2007.
[34] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1994.
[35] C. Reisinger, W. Stockinger, and Y. Zhang. Linear convergence of a policy gradient method for finite horizon continuous time stochastic control problems. 2022. arXiv:2203.11758.
[36] I. M. Ross, M. Karpenko, and R. J. Proulx. A nonsmooth calculus for solving some graph-theoretic control problems. *IFAC-PapersOnLine*, 49(18):462–467, 2016.
[37] S. Sethi and X. Y. Zhou. Stochastic dynamic job shops and hierarchical production planning. *IEEE Trans. Automat. Control*, 39(10):2061–2076, 1994.
[38] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, and A. Bolton. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.
[39] I. Smears and E. Süli. Discontinuous Galerkin finite element approximation of Hamilton-Jacobi-Bellman equations with Cordes coefficients. *SIAM J. Numer. Anal.*, 52(2):993–1016, 2014.
[40] R. S. Sutton and A. G. Barto. *Reinforcement learning: an introduction*. Adaptive Computation and Machine Learning. MIT Press, second edition, 2018.
[41] W. Tang and X. Y. Zhou. Regret of Q-learning in continuous time. 2023+. In preparation.
[42] H. V. Tran. *Hamilton-Jacobi equations—theory and applications*, volume 213 of *Graduate Studies in Mathematics*. American Mathematical Society, 2021.
[43] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.
[44] D. Wang, H. He, and D. Liu. Adaptive critic nonlinear robust control: A survey. *IEEE Trans. Cybern.*, 47(10):3429–3451, 2017.
[45] Q. Wei, D. Liu, and H. Lin. Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems. *IEEE Trans. Cybern.*, 46(3):840–853, 2015.

(W. Tang) Department of Industrial Engineering and Operations Research, Columbia University, S.W. Mudd Building, 500 W 120th St, New York, NY 10027

Email address: wt2319@columbia.edu

(H. V. Tran) Department of Mathematics, University of Wisconsin Madison, Van Vleck Hall, 480 Lincoln Drive, Madison, WI 53706

Email address: hung@math.wisc.edu

(Y. P. Zhang) Department of Mathematics and Statistics, Auburn University, Parker Hall, 221 Roosevelt Concourse, Auburn, AL 36849

Email address: yzhangpaul@auburn.edu