

ON THE STABILITY OF LIPSCHITZ CONTINUOUS CONTROL PROBLEMS AND ITS APPLICATION TO REINFORCEMENT LEARNING

NAMKYEONG CHO AND YEONEUNG KIM

ABSTRACT. We address the crucial yet underexplored stability properties of the Hamilton–Jacobi–Bellman (HJB) equation in model-free reinforcement learning contexts, specifically for Lipschitz continuous optimal control problems. We bridge the gap between Lipschitz continuous optimal control problems and classical optimal control problems in the viscosity solutions framework, offering new insights into the stability of the value function of Lipschitz continuous optimal control problems. By introducing structural assumptions on the dynamics and reward functions, we further study the rate of convergence of value functions. Moreover, we introduce a generalized framework for Lipschitz continuous control problems that incorporates the original problem and leverage it to propose a new HJB-based reinforcement learning algorithm. The stability properties and performance of the proposed method are tested with well-known benchmark examples in comparison with existing approaches.

1. INTRODUCTION

Reinforcement learning (RL) is known to be an effective approach for sequential decision-making problems or optimal control problems, particularly for discrete-in-time settings [19, 15]. One of the well-known data-driven approaches to tackle such problems is Q -learning [19], which is built upon Bellman’s dynamic programming principle [3]. However, the classical Q -learning and its variants such as the Deep- Q -learning algorithm are often limited to discrete-in-time problems. For the extension of RL algorithm to the continuous-in-time problem, various methods have been proposed such as [10, 11, 20]. In [10], authors provide a rigorous justification of Q -function for continuous stochastic optimal control problem under the presence of entropy-regularizer in the reward function, which is motivated by [18]. On the other hand, [11] proposed a model-free reinforcement learning algorithm called Hamilton–Jacobi deep Q -learning (HJDQN) to tackle continuous-in-time deterministic optimal control problems based on the viscosity solution framework. Their idea is to introduce a value function that contains both initial state and action restricting the action to be Lipschitz continuous while the value function only includes the state variable and is defined to be an optimal reward or cost corresponding to the initial state given in standard optimal control theory. This way, one can admit the viscosity solution to the corresponding Hamilton–Jacobi equation as a Q -function which consists of a pair of an initial state and an action, denoted by $Q^L(x, a)$ for some $L > 0$ given. However, such an extension is meaningful only if we restrict the class of admissible controls to be L -Lipschitz continuous, and hence, the value function depends on L . It is also studied in the paper that the choice of hyperparameter L is essential for better performance in practical applications while the optimal choice of L is unknown. Nevertheless, their approach is still a stepping stone to understanding continuous time Q -learning through the lens of the theory of viscosity solution.

In this paper, we rigorously analyze the stability property of $Q^L(x, a)$ on the Lipschitz constraint as well as the convergence of $Q^L(x, a)$ as L grows to infinity. We also demonstrate the rate of convergence under some structural assumptions on the dynamics and reward function. Furthermore, we introduce a slightly general landscape for admitting Q -function as a solution to a Hamilton–Jacobi–Bellman equation by considering a general norm and extending the existing HJDQN algorithm accordingly. Our idea is to consider different metric when defining the Lipschitz continuity of action, which allows flexibility for choosing actions.

Organization of paper. The paper is organized as follows. In Section 2, the problem setup and motivation are presented. We then provide some stability and regularity results on the viscosity solution to the Hamilton–Jacobi–Bellman equations via the theory of viscosity solutions in Section 3. In the following section, a rate of convergence is discussed. Finally, we introduce a different class of compact control to generalize the Lipschitz continuous optimal control problem and present empirical results in Section 5 and Section 6 respectively. We then conclude with Section 7.

Notations. We need a set of notations used throughout the paper. Let $n, k \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $p \geq 1$.

- We denote $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$.
- $D_x f(x) := (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$ and $\Delta f := \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$.
- Define k -dimensional ball as

$$B_{a,k}(y) := \{x \in \mathbb{R}^k : \|x - y\|_2 \leq a\},$$

for $y \in \mathbb{R}^k$.

- For k -continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\|f\|_{C^k(\mathbb{R}^n)} := \sum_{\alpha_1 + \dots + \alpha_n \leq k} \sup_{\mathbb{R}^n} \|D^\alpha f(x)\|_2.$$

- We denote

$$\eta^\varepsilon(x) := \frac{1}{\varepsilon} \eta\left(\frac{x}{\varepsilon}\right), \quad (1.1)$$

where

$$\eta(x) := \begin{cases} e^{-1/(1-\|x\|_2^2)} & \text{for } \|x\|_2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2023-00219980) and the second author is partially supported by National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2023-00211503). Authors are thankful to Hung Vinh Tran at the University of Wisconsin-Madison for sharing ideas and discussions.

2. SETUP AND PRELIMINARIES

Let us introduce the collection of control

$$\mathcal{A} := \{a(\cdot) : [0, \infty) \rightarrow \mathbb{R}^m, a(\cdot) \text{ is measurable}\}. \quad (2.1)$$

Given $a(\cdot) \in \mathcal{A}$, we consider a dynamics evolving under

$$\begin{cases} x'(s) = f(x(s), a(s)), \\ x(0) = x, \end{cases}$$

where $x(t)$ is the system state and $f(x, a) : \mathbb{R}^n \mapsto \mathbb{R}^m \rightarrow \mathbb{R}^n$ is smooth. Hence, it is known that $x(t)$ is absolutely continuous. Then, the standard infinite-horizon discounted optimal control problem is formulated as

$$Q(x) = \sup_{a \in \mathcal{A}} \left\{ \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds : x(0) = x \in \mathbb{R}^n \right\},$$

where $\gamma > 0$ denotes the discount factor, and $r(x, a) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a smooth reward function.

Given $L > 0$, we introduce the class of the Lipschitz controls denoted by

$$\mathcal{A}^L := \{a(\cdot) \in \mathcal{A} : \|a(s_1) - a(s_2)\|_2 \leq L|s_1 - s_2| \text{ for all } s_1, s_2 \in [0, \infty)\}, \quad (2.2)$$

and define the value function as

$$Q^L(x, a) = \sup_{a \in \mathcal{A}^L} \left\{ \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds : x(0) = x, a(0) = a \right\}. \quad (2.3)$$

Assumption 2.1. Throughout the paper, we assume that f and r are Lipschitz continuous, that is, there exists $C > 0$ such that

$$\|f\|_{L^\infty(\mathbb{R}^n \times \mathbb{R}^m)} + \|r\|_{L^\infty(\mathbb{R}^n \times \mathbb{R}^m)} + \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} + \|r\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \leq C.$$

Under the assumption above, it is known that $Q(x)$ and $Q^L(x, a)$ solve

$$\gamma Q - \sup_{a \in \mathbb{R}^m} (D_x Q \cdot f(x, a) + r(x, a)) = 0,$$

and

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - L \|D_a Q^L\|_2 - r(x, a) = 0, \quad (2.4)$$

respectively in viscosity sense [17, 2, 11].

We end this section by reminding of the regularity property of Q . A further investigation on Q^L will be presented in the following section. In the classical infinite-horizon optimal control problem, it is often

assumed that the control takes values in a compact set as opposed to our setting. Nevertheless, we still have the same regularity property.

Proposition 2.2. *Under Assumption 2.1, the unique viscosity solution Q to*

$$\gamma Q - \sup_{a \in \mathbb{R}^m} (D_x Q \cdot f(x, a) + r(x, a)) = 0$$

is Lipschitz continuous if $\gamma > \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$.

Proof. See [17]. □

In the following section, it is shown that Q^L is also Lipschitz continuous. In addition to that we study the stability of Q^L in L as well as the convergence as L grows to infinity.

3. SOME QUANTITATIVE PROPERTIES OF Q^L

In this section, we explore some quantitative behaviors of Q^L by employing standard methods [17] used for analyzing the regularity properties of viscosity solutions to Hamilton-Jacobi-Bellman (HJB) equations. We first show that Q^L is uniformly Lipschitz in the state and action variable and derive the estimate for the rate of change in L .

3.1. Lipschitz Regularity. Let us begin by providing a result on Lipschitz continuity of Q^L in state and action variables. The following lemma suggests that the Lipschitz constant is independent of L . To show this, as demonstrated in [11], we introduce a new state variable $z := (x, a)$ and control variable $b(\cdot) := \dot{a}(\cdot)$ such that $\|b(\cdot)\|_2 \leq L$.

Lemma 3.1. *Let $L > 0$ be given and suppose that Q^L is the solution of (2.4), and f and r satisfy Assumption 2.1. In addition, we assume that*

$$\gamma > \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}.$$

Then, $Q^L(x, a)$ is Lipschitz continuous in x and a . Furthermore, we have the estimate

$$|Q^L| + \|D_x Q^L\|_2 + L \|D_a Q^L\|_2 \leq C \tag{3.1}$$

for some $C > 0$.

Proof. Let $\mathcal{B}^L := \{b(\cdot) : [0, \infty) \rightarrow \mathbb{R}^m, b(\cdot) \text{ is measurable and } \|b(\cdot)\|_2 \leq L\}$. Given $z := (x, a) \in \mathbb{R}^{n+m}$ and $b(\cdot) \in \mathcal{B}^L$, let us define

$$V_{b(\cdot)}^L(x, a) := \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds,$$

subject to

$$\begin{cases} x'(s) = f(x(s), a(s)) & \text{for } s > 0, \\ a'(s) = b(s) & \text{for } s > 0, \\ x(0) = x, \\ a(0) = a. \end{cases}$$

Fixing $\tilde{z} = (\tilde{x}, \tilde{a}) \in \mathbb{R}^{n+m}$ and $b(\cdot) \in \mathcal{B}^L$ as well as the trajectory $\tilde{z}(s) := (\tilde{x}(s), \tilde{a}(s))$ satisfying

$$\begin{cases} \tilde{x}'(s) = f(\tilde{x}(s), \tilde{a}(s)) & \text{for } s > 0, \\ \tilde{a}'(s) = b(s) & \text{for } s > 0, \\ \tilde{x}(0) = \tilde{x}, \\ \tilde{a}(0) = \tilde{a}, \end{cases}$$

and letting $z(s) := (x(s), a(s))$, we have that

$$\|z'(s) - \tilde{z}'(s)\|_2 \leq \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \|z(s) - \tilde{z}(s)\|_2.$$

For simplicity, let $\gamma_0 := \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$. Invoking Gronwall's inequality, we have that

$$\|z(s) - \tilde{z}(s)\|_2 \leq e^{\gamma_0 s} \|z(0) - \tilde{z}(0)\|_2 = e^{\gamma_0 s} (\|x - \tilde{x}\|_2 + \|a - \tilde{a}\|_2).$$

Now observing

$$\begin{aligned}
|V_{b(\cdot)}^L(x, a) - V_{\tilde{b}(\cdot)}^L(\tilde{x}, \tilde{a})| &= \left| \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds - \int_0^\infty e^{-\gamma s} r(\tilde{x}(s), \tilde{a}(s)) ds \right| \\
&\leq \|r\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \int_0^\infty e^{-\gamma s} (\|x(s) - \tilde{x}(s)\|_2 + \|a(s) - \tilde{a}(s)\|_2) ds \\
&\leq \|r\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} (\|x - \tilde{x}\|_2 + \|a - \tilde{a}\|_2) \int_0^\infty e^{(-\gamma + \gamma_0)s} ds \\
&\leq C(\|x - \tilde{x}\|_2 + \|a - \tilde{a}\|_2),
\end{aligned}$$

for some $C > 0$ since $\gamma > \gamma_0$, we deduce that

$$|Q^L(x, a) - Q^L(\tilde{x}, \tilde{a})| \leq C(\|x - \tilde{x}\|_2 + \|a - \tilde{a}\|_2)$$

by taking supremum over $b \in \mathcal{B}^L$.

The inequality (3.1) follows from the equation (2.4) and we finish the proof. \square

3.2. Error Bounds of $Q^{L+\ell} - Q^L$ for $L, \ell \geq 0$. Using the Lipschitz estimate above, one can estimate the rate of change of Q^L with respect to L based on the standard doubling variable argument [17, 12, 8, 7]. The result we propose implies that the value function Q^L does increase abruptly as L increases. Numerical justification for this property is also presented in Section 6.

Lemma 3.2. *Let Assumption 2.1 be enforced and $\gamma > \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$. Given $L > 0$ and $\ell > 0$, let $Q^L(x, a)$ and $Q^{L+\ell}(x, a)$ be solutions to (2.4) respectively where the latter is associated with $L + \ell$ instead of L . Then, there exists a constant C satisfying*

$$0 \leq Q^{L+\ell}(x, a) - Q^L(x, a) \leq \frac{C\ell}{L + \ell}, \quad (3.2)$$

and hence,

$$\limsup_{\ell \rightarrow 0} \frac{Q^{L+\ell}(x, a) - Q^L(x, a)}{\ell} \leq \frac{C}{L}.$$

Proof. Since Q^L is defined as (2.3) and \mathcal{A}^L , the set of Lipschitz continuous controls, increases in L , it follows that

$$Q^L(x, a) \leq Q^{L+\ell}(x, a) \quad \text{for every } (x, a) \in \mathbb{R}^n \times \mathbb{R}^m.$$

Therefore, the left side of inequality (3.2) holds.

To show the right side of inequality (3.2), we use the doubling variable technique [17, 1, 7, 12]. Let $\mu(x) \in C^1(\mathbb{R}^n)$ $\tilde{\mu}(a) \in C^1(\mathbb{R}^m)$ satisfy

$$\|D_x \mu\|_2 \leq 1 \quad \text{and} \quad \|D_a \tilde{\mu}\|_2 \leq 1. \quad (3.3)$$

For $\delta, \varepsilon > 0$ and $(x, y, a, b) \in \mathbb{R}^{2n} \times \mathbb{R}^{2m}$, we define an auxiliary function as follows:

$$\Psi(x, y, a, b) := Q^{L+\ell}(x, a) - Q^L(y, b) - \frac{\|x - y\|_2^2 + \|a - b\|_2^2}{2\varepsilon} - \delta(\mu(x) + \mu(y) + \tilde{\mu}(a) + \tilde{\mu}(b)).$$

We choose a point $(\bar{x}, \bar{y}, \bar{a}, \bar{b}) \in \mathbb{R}^{2n} \times \mathbb{R}^{2m}$ satisfying

$$\Psi(\bar{x}, \bar{y}, \bar{a}, \bar{b}) = \sup_{(x, y, a, b) \in \mathbb{R}^{2n} \times \mathbb{R}^{2m}} \{\Psi(x, y, a, b) : (x, y, a, b) \in \mathbb{R}^{2n} \times \mathbb{R}^{2m}\}.$$

From the inequality $\Psi(\bar{y}, \bar{y}, \bar{a}, \bar{b}) \leq \Psi(\bar{x}, \bar{y}, \bar{a}, \bar{b})$, we see that

$$Q^{L+\ell}(\bar{y}, \bar{a}) - \delta\mu(\bar{y}) \leq Q^{L+\ell}(\bar{x}, \bar{a}) - \frac{\|\bar{x} - \bar{y}\|_2^2}{2\varepsilon} - \delta\mu(\bar{x}).$$

Recalling Lemma 3.1 and (3.3), we establish that

$$\begin{aligned}
\frac{\|\bar{x} - \bar{y}\|_2^2}{2\varepsilon} &\leq Q^L(\bar{x}, \bar{a}) - Q^L(\bar{y}, \bar{a}) - \delta(\mu(\bar{y}) - \mu(\bar{x})) \\
&\leq C(1 + \delta)\|\bar{x} - \bar{y}\|_2
\end{aligned}$$

for some constant $C > 0$. Rearranging inequality above gives

$$\|\bar{x} - \bar{y}\|_2 \leq C(1 + \delta)\varepsilon. \quad (3.4)$$

Similarly, we start with the inequality $\Psi(\bar{x}, \bar{y}, \bar{b}, \bar{b}) \leq \Psi(\bar{x}, \bar{y}, \bar{a}, \bar{b})$. Invoking Lemma 3.1 and (3.3) once again,

$$\frac{\|\bar{a} - \bar{b}\|_2^2}{2\varepsilon} \leq Q^{L+\ell}(\bar{x}, \bar{a}) - Q^{L+\ell}(\bar{x}, \bar{b}) + \delta(\tilde{\mu}(\bar{b}) - \tilde{\mu}(\bar{a}))$$

$$\leq \frac{C\|\bar{a} - \bar{b}\|_2}{L + \ell} + \delta\|\bar{a} - \bar{b}\|_2.$$

Dividing both side by $\|\bar{a} - \bar{b}\|_2$, we derive that

$$\|\bar{a} - \bar{b}\|_2 \leq \left(\frac{C}{L + \ell} + 2\delta \right) \varepsilon. \quad (3.5)$$

Since the map

$$(x, a) \mapsto Q^{L+\ell}(x, a) - \frac{\|x - \bar{y}\|_2^2 + \|a - \bar{b}\|_2^2}{2\varepsilon} - \delta(\mu(x) + \tilde{\mu}(a))$$

obtains a local maximum value at (\bar{x}, \bar{a}) , we have that

$$\gamma Q^{L+\ell}(\bar{x}, \bar{a}) - f(\bar{x}, \bar{a}) \cdot \left(\frac{\bar{x} - \bar{y}}{\varepsilon} + \delta D_x \mu(\bar{x}) \right) - (L + \ell) \left\| \frac{\bar{a} - \bar{b}}{\varepsilon} + \delta D_a \tilde{\mu}(\bar{a}) \right\|_2 - r(\bar{x}, \bar{a}) \leq 0, \quad (3.6)$$

by the subsolution test.

Similarly, the map

$$(y, b) \mapsto Q^L(y, b) - \left(-\frac{\|y - \bar{x}\|_2^2 + \|b - \bar{a}\|_2^2}{2\varepsilon} - \delta(\mu(y) + \tilde{\mu}(b)) \right)$$

obtains a local minimum value at (\bar{y}, \bar{b}) . From the definition of viscosity supersolution, we see that

$$\gamma Q^L(\bar{y}, \bar{b}) + f(\bar{y}, \bar{b}) \cdot \left(\frac{\bar{y} - \bar{x}}{\varepsilon} - \delta D_x \mu(\bar{y}) \right) - L \left\| \frac{\bar{a} - \bar{b}}{\varepsilon} + \delta D_a \tilde{\mu}(\bar{b}) \right\|_2 - r(\bar{y}, \bar{b}) \geq 0. \quad (3.7)$$

Subtracting (3.7) from (3.6), we have

$$\begin{aligned} & \gamma (Q^{L+\ell}(\bar{x}, \bar{a}) - Q^L(\bar{y}, \bar{b})) \\ & \leq (f(\bar{x}, \bar{a}) - f(\bar{y}, \bar{b})) \cdot \left(\frac{\bar{x} - \bar{y}}{\varepsilon} \right) + (r(\bar{x}, \bar{a}) - r(\bar{y}, \bar{b})) + \ell \left\| \frac{\bar{a} - \bar{b}}{\varepsilon} \right\|_2 \\ & \quad + \delta (\|f(\bar{x}, \bar{a})\|_2 \|D_x \mu(\bar{x})\|_2 + \|f(\bar{y}, \bar{b})\|_2 \|D_x \mu(\bar{y})\|_2 + (L + \ell) \|D_a \tilde{\mu}(\bar{a})\|_2 + L \|D_a \tilde{\mu}(\bar{b})\|_2). \end{aligned} \quad (3.8)$$

Combining (3.3), (3.4) and (3.5), there exists a constant C satisfying

$$\begin{aligned} & \gamma (Q^{L+\ell}(\bar{x}, \bar{a}) - Q^L(\bar{y}, \bar{b})) \\ & \leq C (\|\bar{x} - \bar{y}\|_2 + \|\bar{a} - \bar{b}\|_2) \frac{\|\bar{x} - \bar{y}\|_2}{\varepsilon} + C (\|\bar{x} - \bar{y}\|_2 + \|\bar{a} - \bar{b}\|_2) + \ell \left\| \frac{\bar{a} - \bar{b}}{\varepsilon} \right\|_2 + \delta (C + 2L + \ell) \\ & \leq C \left(\frac{\|\bar{x} - \bar{y}\|_2}{\varepsilon} + 1 \right) (\|\bar{x} - \bar{y}\|_2 + \|\bar{a} - \bar{b}\|_2) + \ell \left\| \frac{\bar{a} - \bar{b}}{\varepsilon} \right\|_2 + \delta (C + 2L + \ell) \\ & \leq C\varepsilon + \frac{C\ell}{L + \ell} + \delta (C + 2L + \ell). \end{aligned}$$

Finally, invoking the inequality

$$\Psi(x, x, a, a) \leq \Psi(\bar{x}, \bar{y}, \bar{a}, \bar{b}) \quad \forall (x, a) \in \mathbb{R}^n \times \mathbb{R}^m,$$

we have that

$$\gamma (Q^{L+\ell}(x, a) - Q^L(x, a)) - 2\delta(\mu(x) + \tilde{\mu}(a)) \leq C\varepsilon + \frac{C\ell}{L + \ell} + \delta (C + 2L + \ell).$$

We complete the proof by taking $\delta, \varepsilon \rightarrow 0$. □

4. CONVERGENCE OF $Q^L(x, a)$ TO $Q(x)$

4.1. General convergence result. We provide the general convergence result based on the stability properties of viscosity solutions. Let us recall the following definition of upper or lower half-relaxed limit [14, 1]. It is crucial to note that the limit functions are subsolution and supersolution to the limiting equation.

Definition 4.1. For a family of locally bounded functions on \mathbb{R}^n denoted by $\{u_\alpha\}_{\alpha \in \mathbb{R}}$. We define the upper and lower half-relaxed limits u^* and u_* of u_α as

$$u^*(x) = \limsup_{\alpha \rightarrow \infty}^* u_\alpha(x) := \lim_{\alpha \rightarrow \infty} \sup \{u_\beta(y) : \|x - y\|_2 \leq \frac{1}{\beta}, \beta \geq \alpha\},$$

and

$$u_*(x) = \liminf_{\alpha \rightarrow \infty}^* u_\alpha(x) := \lim_{\alpha \rightarrow \infty} \inf \{u_\beta(y) : \|x - y\|_2 \leq \frac{1}{\beta}, \beta \geq \alpha\}.$$

Let us state one of the main results, the convergence of Q^L to Q as L grows to infinity.

Theorem 4.2. *Let $Q^L := Q^L(x, a) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a viscosity solution to*

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - L \|D_a Q^L\|_2 - r(x, a) = 0.$$

Then we have that

$$Q^L(x, a) \rightarrow Q(x) \quad \text{locally uniformly as } L \rightarrow \infty,$$

where $Q := Q(x)$ is a unique viscosity solution to

$$\gamma Q - \sup_{a \in \mathbb{R}^m} (D_x Q \cdot f(x, a) + r(x, a)) = 0.$$

Proof. By the comparison principle, we have that

$$-\frac{\|r\|_{L^\infty(\mathbb{R}^n \times \mathbb{R}^m)}}{\gamma} \leq Q^L(x, a) \leq \frac{\|r\|_{L^\infty(\mathbb{R}^n \times \mathbb{R}^m)}}{\gamma}.$$

Noticing that $\|D_a Q(x)\|_2 = 0$, $Q(x)$ satisfies

$$\gamma Q(x) - D_x Q \cdot f(x, a) - r(x, a) - L \|D_a Q(x)\|_2 \geq 0,$$

which implies that

$$Q^L(x, a) \leq Q(x).$$

Since Q^L is bounded, we also get

$$Q^* := \limsup_{L \rightarrow \infty} Q^L(x, a) \leq Q(x).$$

Similarly,

$$Q_* := \liminf_{L \rightarrow \infty} Q^L(x, a).$$

Our goal is to show $Q^* = Q_* = Q(x)$, which is equivalent to

$$\lim_{L \rightarrow \infty} Q^L(x, a) = Q(x).$$

Rewriting (2.4), one can see that Q^L is a viscosity solution to

$$-\|D_a Q^L\|_2 + \frac{1}{L}(\gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a)) = 0.$$

Taking $L \rightarrow \infty$ and invoking the stability result on the viscosity solution [1], the lower half-relaxed limit Q_* is indeed a supersolution to

$$-\|D_a Q_*\|_2 \geq 0,$$

and hence, $|D_a Q_*| \leq 0$. Therefore Q_* is independent of a , that is,

$$Q_*(x, a) \equiv Q_*(x).$$

Additionally, Q^L also satisfies

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a) = L \|D_a Q^L\|_2 \geq 0.$$

Again by the stability property of viscosity solutions, taking the limit of L , we derive that Q_* is a supersolution to

$$\gamma Q_* - D_x Q_* \cdot f(x, a) - r(x, a) \geq 0,$$

and

$$\gamma Q_* - \sup_{a \in \mathbb{R}^m} \{D_x Q_* \cdot f(x, a) - r(x, a)\} \geq 0,$$

since Q_* is independent of a . Therefore, again by the comparison principle,

$$u_* \geq u,$$

and we finish the proof as $Q_* = Q^* = Q$ implying

$$\lim_{L \rightarrow \infty} Q^L(x, a) = Q(x).$$

□

The convergence property can be achieved only under the boundedness assumptions on the dynamics and reward functions, f and r . One can show further how fast this convergence occurs introducing some structural assumptions on f and r .

4.2. Rate of convergence for some special cases. In the previous section, we show that $Q^L(x, a)$ converges to $Q(x)$ using the classical stability result on viscosity solutions. We now further investigate the rate of convergence of $Q^L(x, a) \rightarrow Q(x)$ as $L \rightarrow \infty$ under appropriate assumptions. We begin by showing the existence of optimal control for Q and Q^L based on [2]. To prove the existence of an optimal control achieving the optimal value Q , we enforce the following assumption.

Assumption 4.3. A Borel function $\alpha(x, p) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ is uniquely determined to satisfy

$$f(x, \alpha(x, p)) \cdot p + r(x, \alpha(x, p)) = \max_{a \in \mathbb{R}^m} (f(x, a) \cdot p + r(x, a)).$$

We further assume that there exists $C > 0$, satisfying

$$\begin{aligned} \|f(x+h) - 2f(x, a) + f(x-h, a)\|_2 &\leq C\|h\|_2^2, \\ r(x+h) - 2r(x, a) + r(x-h, a) &\geq -C\|h\|_2^2. \end{aligned}$$

Proposition 4.4. Let $\gamma > 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$ and Assumption 4.3 be enforced. For any given $x \in \mathbb{R}^n$, there exists $a(\cdot) \in \mathcal{A}$ from (2.1) such that

$$Q(x) = \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds, \quad (4.1)$$

where

$$\begin{cases} x'(s) = f(x(s), a(s)) & \text{for } s > 0, \\ x(0) = x. \end{cases} \quad (4.2)$$

Proof. By [17][Theorem 2.8], we have $Q(x)$ is bounded and Lipschitz continuous in \mathbb{R}^n . Denoting $\tilde{Q}(x) := -Q(x)$, we have

$$\tilde{Q}(x) + H_\gamma(x, D_x \tilde{Q}(x)) = 0,$$

where

$$H_\gamma(x, p) := \sup_{a \in \mathbb{R}^m} \left(-\frac{f(x, a)}{\gamma} \cdot p + \frac{r(x, a)}{\gamma} \right).$$

We will show that $\tilde{Q}^L(x)$ is semiconcave. By Assumption 4.3, for all $x, p, q \in \mathbb{R}^n$, we find

$$\begin{aligned} H_\gamma(x, p) - H_\gamma(x, q) &= \sup_{a \in \mathbb{R}^m} \left(-\frac{f(x, a)}{\gamma} \cdot p + \frac{r(x, a)}{\gamma} \right) - \sup_{a \in \mathbb{R}^m} \left(-\frac{f(x, a)}{\gamma} \cdot q + \frac{r(x, \alpha(x, p))}{\gamma} \right) \\ &\leq \left(-\frac{f(x, \alpha(x, p))}{\gamma} \cdot p + \frac{r(x, \alpha(x, p))}{\gamma} \right) - \left(-\frac{f(x, \alpha(x, p))}{\gamma} \cdot q + \frac{r(x, \alpha(x, p))}{\gamma} \right) \\ &\leq \left\| -\frac{f(x, \alpha(x, p))}{\gamma} \right\|_2 \|p - q\|_2. \end{aligned}$$

We may assume that the left-hand side is positive. Otherwise, we change p and q . Hence,

$$|H_\gamma(x, p) - H_\gamma(x, q)| \leq C\|p - q\|_2$$

for some $C > 0$. We claim that there exists $\tilde{C} > 0$

$$H_\gamma(x+h, p+\tilde{C}h) - 2H_\gamma(x, p) + H_\gamma(x-h, p-\tilde{C}h) \geq -\tilde{C}\|h\|_2^2, \quad (4.3)$$

for all $x, h \in \mathbb{R}^n$ and $p \in \mathbb{R}^n$ satisfying $\|p\|_2 \leq 3\|Q(x, a)\|_{\text{Lip}(\mathbb{R}^n)}$. We observe that

$$\begin{aligned} &H_\gamma(x+h, p+\tilde{C}h) + H_\gamma(x-h, p-\tilde{C}h) \\ &\geq \sup_{a \in \mathbb{R}^m} \left(-\frac{f(x+h, a)}{\gamma} \cdot (p+\tilde{C}h) - \frac{f(x-h, a)}{\gamma} \cdot (p-\tilde{C}h) + \frac{r(x+h, a)}{\gamma} + \frac{r(x-h, a)}{\gamma} \right) \\ &= \sup_{a \in \mathbb{R}^m} \left(-\frac{f(x+h, a) + f(x-h, a) - 2f(x, a)}{\gamma} \cdot p + \frac{r(x+h, a) + r(x-h, a) - 2r(x, a)}{\gamma} \right. \\ &\quad \left. - \frac{f(x+h, a) - f(x-h, a)}{\gamma} \cdot (\tilde{C}h) - 2\frac{f(x, a)}{\gamma} \cdot p + 2\frac{r(x, a)}{\gamma} \right) =: I. \end{aligned}$$

Note that we have

$$\frac{\|f(x+h, a) - f(x-h, a)\|_2}{\gamma} \leq \frac{2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}\|h\|_2}{\gamma}.$$

From the inequality above and the Assumption 4.3, we estimate I by

$$I \geq - \left(C_1(\|Q(x)\|_{\text{Lip}(\mathbb{R}^n)} + 1) + -\frac{2\tilde{C}}{\gamma}\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \right) \|h\|_2^2 + 2H_\gamma(x, p),$$

for some constant $C_1 > 0$. By choosing \tilde{C} large enough to satisfy

$$\tilde{C} > \frac{C_1 \gamma (\|Q(x)\|_{\text{Lip}(\mathbb{R}^n)} + 1)}{\gamma - 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}},$$

the inequality (4.3) holds. Note that such \tilde{C} can be chosen since $\gamma > 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$. Recalling [2][Theorem 4.9, Chapter II], $\tilde{Q}(x)$ is semiconcave in x .

By [2][Proposition 4.7, Chapter II], we have $D_x^+ \tilde{Q}(x) = \partial_x \tilde{Q}(x) \supset D_x^* \tilde{Q}(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$.¹ Let us define a function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ satisfying

$$\psi(x) = \begin{cases} D_x \tilde{Q}(x) & \text{if } \tilde{Q} \text{ is differentiable at } x \in \mathbb{R}^n, \\ \text{some } p \in D_x^+ \tilde{Q}(x) \cap D_x^* \tilde{Q}(x) & \text{otherwise.} \end{cases} \quad (4.4)$$

For a given state $x(s)$, we choose a action $a(s) = \alpha(x(s), -\psi(x(s)))$. For the moment, let us assume that we can choose $p \in D_x^+ \tilde{Q}(x) \cap D_x^* \tilde{Q}(x) \cap D_x^* \tilde{Q}(x)$ so that $\psi(x)$ is a Borel function. Then clearly, we have $a(\cdot) \in \mathcal{A}$. The details of the proof can be found in Appendix 8.1. Then from Assumption 4.3 and the definition of ψ , we have

$$\begin{aligned} & f(x(s), a(s)) \cdot \psi(x(s)) - r(x(s), a(s)) \\ &= -(f(x(s), a(s)) \cdot (-\psi(x(s))) + r(x(s), a(s))) \\ &= -\max_{a \in \mathbb{R}^m} (f(x(s), a) \cdot (-\psi(x(s))) + r(x(s), a)) \\ &= \min_{a \in \mathbb{R}^m} (f(x(s), a) \cdot \psi(x(s)) - r(x(s), a)). \end{aligned}$$

Since $\psi(x) \in D_x^+ \tilde{Q}(x)$, the assumption of [2][Theorem 2.52, Chapter III] is satisfied.² Therefore, we have

$$\tilde{Q}(x) = \int_0^\infty e^{-\gamma s} (-r(x(s), a(s))) ds,$$

which implies

$$Q(x) = \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds.$$

This completes the proof. \square

In the following proposition, we show the existence of optimal control for Q^L . Some of the arguments in the proof overlap with the previous proposition, but we shall provide details for the convenience of the reader.

Proposition 4.5. *Assume further that $\gamma > 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$ and let $(x, a) \in \mathbb{R}^n \times \mathbb{R}^m$ be given. Then, there exists $a_L(s) \in \mathcal{A}^L$ from (2.2) such that*

$$Q^L(x, a) = \int_0^\infty e^{-\gamma s} r(x_L(s), a_L(s)) ds, \quad (4.5)$$

where

$$\begin{cases} x_L'(s) = f(x_L(s), a_L(s)) & \text{for } s > 0, \\ x_L(0) = x, \\ a_L(0) = a. \end{cases} \quad (4.6)$$

Proof. We first prove the theorem under the assumption

$$f(x, a), r(x, a) \in C^2(\mathbb{R}^n \times \mathbb{R}^m) \quad \text{and} \quad \|f\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)} + \|r\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)} \quad \text{is bounded.} \quad (4.7)$$

Applying Lemma 3.1, we have $Q(x, a)$ is bounded and Lipschitz continuous in $\mathbb{R}^n \times \mathbb{R}^m$. We introduce extended state variable $z(s)$ and dynamics $G : (\mathbb{R}^n \times \mathbb{R}^m) \times \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ satisfying

$$z(s) = (x_L(s), a_L(s)) \quad \text{and} \quad G(z, b) = (f(z), b).$$

Then (4.6) can be written as

$$\begin{cases} z'(s) &= G(z(s), b(s)) \quad \text{for } s > 0, \\ z(0) &= (x, a). \end{cases}$$

¹Here, $D_x^+ Q(x)$ and $D_x^- Q(x)$ denote the super-differential and sub-differential of the function $Q(x)$. The notation $\partial_x Q(x)$ refers to a generalized gradient or Clarke's gradient and $D_x^* u = \{p \in \mathbb{R}^n : p = \lim D_x u(x_n), x_n \rightarrow x\}$. For details about super-, sub-differential, and Clarke's gradient, see [2][Chapter II.4.1].

²Although the action spaces are non-compact set, with the same argument, the results of [2][Theorem 2.52, Chapter III] still hold.

Rewriting (2.4), in terms of $z(t)$ and $G(z, b)$ and $\tilde{Q}^L(x, a) := -Q^L(x, a)$, we have

$$\tilde{Q}^L(x, a) + H_\gamma(z, D_z \tilde{Q}^L(z)) = 0,$$

where

$$H_\gamma(z, p) := \sup_{\|b\|_2 \leq L} \left(-\frac{G(z, b)}{\gamma} \cdot p + \frac{r(z)}{\gamma} \right).$$

We will show that $\tilde{Q}^L(z)$ is semiconcave.

It is clear that

$$|H_\gamma(z, p) - H_\gamma(z, q)| \leq C\|p - q\|_2,$$

for some $C > 0$. We claim that there exists $\tilde{C} > 0$ such that

$$H_\gamma(z + h, p + \tilde{C}h) - 2H_\gamma(z, p) + H_\gamma(z - h, p - \tilde{C}h) \geq -\tilde{C}\|h\|_2^2 \quad (4.8)$$

for all $z, h \in \mathbb{R}^n \times \mathbb{R}^m$ and $p \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying $\|p\|_2 \leq 3\|Q^L(x, a)\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$. For justification, let us first observe that

$$\begin{aligned} & H_\gamma(z + h, p + \tilde{C}h) + H_\gamma(z - h, p - \tilde{C}h) \\ & \geq \sup_{\|b\|_2 \leq L} \left(-\frac{G(z + h, b)}{\gamma} \cdot (p + \tilde{C}h) - \frac{G(z - h, b)}{\gamma} \cdot (p - \tilde{C}h) + \frac{r(z + h)}{\gamma} + \frac{r(z - h)}{\gamma} \right) \\ & = \sup_{\|b\|_2 \leq L} \left(-\frac{G(z + h, b) + G(z - h, b) - 2G(z, b)}{\gamma} \cdot p + \frac{r(z + h) + r(z - h) - 2r(z)}{\gamma} \right. \\ & \quad \left. - \frac{G(z + h, b) - G(z - h, b)}{\gamma} \cdot (\tilde{C}h) - 2\frac{G(z, b)}{\gamma} \cdot p + 2\frac{r(z)}{\gamma} \right) =: I. \end{aligned}$$

By the mean value theorem, we find

$$\begin{aligned} \frac{\|G(z + h, b) - 2G(z, b) + G(z - h, b)\|_2}{\gamma} & \leq \frac{C\|f\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)}\|h\|_2^2}{\gamma}, \\ \frac{|r(z + h) - 2r(z) + r(z - h)|}{\gamma} & \leq \frac{C\|r\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)}\|h\|_2^2}{\gamma}, \\ \frac{\|G(z + h) - G(z - h)\|_2}{\gamma} & \leq \frac{2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}\|h\|_2}{\gamma}. \end{aligned}$$

From the inequalities above, we estimate I by

$$\begin{aligned} I & \geq - \left(\frac{C_1}{\gamma} (\|f\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)}\|Q^L(x, a)\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} + \|r\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)}) + \frac{2\tilde{C}}{\gamma}\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \right) \|h\|_2^2 \\ & \quad + 2H_\gamma(z, p). \end{aligned}$$

By choosing \tilde{C} large enough to satisfy

$$\tilde{C} > \frac{C_1(\|f\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)}\|Q^L(x, a)\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} + \|r\|_{C^2(\mathbb{R}^n \times \mathbb{R}^m)})}{\gamma - 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}},$$

the inequality (4.8) holds. Note that such \tilde{C} can be chosen since $\gamma > 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$. Recalling [2][Theorem 4.9, Chapter II], \tilde{Q}^L is semiconcave on \mathbb{R}^n . By [2][Proposition 4.7, Chapter II], we have $D_z^+ \tilde{Q}^L(z) = \partial_z \tilde{Q}^L(z)$ for all $z \in \mathbb{R}^n \times \mathbb{R}^m$. Finally, by [11][Theorem 1], there exists an optimal control $a_L(\cdot) \in \mathcal{A}^L$ and a corresponding state $x_L(s)$ that satisfy (4.5) and (4.6).

Next, we remove the assumption (4.7). For all $\varepsilon > 0$, let $\tilde{\eta}^\varepsilon$ be a mollifier, as defined in (1.1), but with the domain \mathbb{R}^n replaced by $\mathbb{R}^n \times \mathbb{R}^m$. By defining mollified functions

$$f_i^\varepsilon(z) := f_i * \tilde{\eta}^\varepsilon(z) \quad \text{and} \quad r_i^\varepsilon(z) := r_i * \tilde{\eta}^\varepsilon(z),$$

where the subscript i denotes the i th component.

From [8][Theorem 6, Appendix C.4], we have $f^\varepsilon(x, a), r^\varepsilon(x, a) \in C^2(\mathbb{R}^n \times \mathbb{R}^m)$ and

$$f^\varepsilon(z) \rightarrow f(z), \quad r^\varepsilon(z) \rightarrow r(z) \quad \text{locally uniformly.}$$

For all $z_1, z_2 \in \mathbb{R}^n \times \mathbb{R}^m$, we find

$$\begin{aligned} \|f^\varepsilon(z_1) - f^\varepsilon(z_2)\|_2 & \leq \int_{\mathbb{R}^n \times \mathbb{R}^m} \|(f(z_1 - z) - f(z_2 - z))\tilde{\eta}^\varepsilon(z)\|_2 dz \\ & \leq \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}|z_1 - z_2| \int_{\{\|z\|_2 \leq \varepsilon\}} \tilde{\eta}^\varepsilon(z) dz \end{aligned}$$

$$\leq \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} |z_1 - z_2|.$$

Therefore, $\|f^\varepsilon\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \leq \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$ and as a result, $\gamma > 2\|f^\varepsilon\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$ holds for all $\varepsilon > 0$. As a direct consequence of the previous argument, there exist $x_L^\varepsilon(s)$ and $a_L^\varepsilon(\cdot) \in \mathcal{A}_L$ satisfying (4.5) and (4.6) replacing f, r and Q^L by $f^\varepsilon, r^\varepsilon$ and $Q^{L,\varepsilon}$, respectively.

For all $s \in [0, s_0]$ with any fixed $s_0 > 0$, we have the following estimates:

$$\begin{aligned} & \|x_L^\varepsilon(s)\| + \left\| \frac{dx_L^\varepsilon(s)}{ds} \right\|_2 + \|a_L^\varepsilon(s)\|_2 + \left\| \frac{da_L^\varepsilon(s)}{ds} \right\|_2 \\ & \leq \|x\|_2 + \int_0^{s_0} \left\| \frac{dx_L^\varepsilon(s)}{ds} \right\|_2 ds + \left\| \frac{dx_L^\varepsilon(s)}{ds} \right\|_2 + \|a\|_2 + \int_0^{s_0} \left\| \frac{da_L^\varepsilon(s)}{ds} \right\|_2 ds + \left\| \frac{da_L^\varepsilon(s)}{ds} \right\|_2 \\ & \leq \|x\|_2 + (s_0 + 1)\frac{\gamma}{2} + \|a\|_2 + (s_0 + 1)L. \end{aligned}$$

Given $n \in \mathbb{N}$, we can find a sequence $\{\varepsilon_k^n\}_{k=1}^\infty \searrow 0$ such that

$$a_L^{\varepsilon_k^n} \rightarrow \hat{a}_L^n \quad \text{and} \quad x_L^{\varepsilon_k^n} \rightarrow \hat{x}_L^n \quad \text{locally uniformly on } [0, n] \quad \text{as } k \rightarrow \infty,$$

by Arzelà–Ascoli theorem. By taking a further subsequence, we may assume that

$$a_L^{\varepsilon_k} \rightarrow \hat{a}_L \quad \text{and} \quad x_L^{\varepsilon_k} \rightarrow \hat{x}_L \quad \text{locally uniformly on } [0, \infty) \quad \text{as } k \rightarrow \infty.$$

Clearly, \hat{a}_L is L -Lipschitz continuous on $[0, \infty)$. From

$$x_L^\varepsilon(t) = \int_0^t f(\hat{x}_L^\varepsilon(s), \hat{a}_L^\varepsilon(s)) ds + x,$$

and the local uniform convergence, we also have that

$$\hat{x}_L(t) = \int_0^t f(\hat{x}_L(s), \hat{a}_L(s)) ds + x.$$

Finally, we will show that

$$Q^L(x, a) = \int_0^\infty e^{-\gamma s} r(\hat{x}_L(s), \hat{a}_L(s)) ds$$

using the stability property of the viscosity solution. To this end, let us define $\hat{Q}^L(x, a)$ as

$$\hat{Q}^L(x, a) := \int_0^\infty e^{-\gamma s} r(\hat{x}_L(s), \hat{a}_L(s)) ds.$$

and $g_L(x, a, t) : (\mathbb{R}^n \times \mathbb{R}^m) \times [0, \infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ satisfy $g_L(x, a, t) = (\tilde{x}_L(t), \tilde{a}_L(t))$, where $\tilde{x}(t)$ and $\tilde{a}(t)$ satisfies

$$\begin{cases} \tilde{x}_L'(s) = f(\tilde{x}_L(s), \tilde{a}_L(s)), \\ \tilde{x}_L(0) = x, \\ \tilde{a}_L(0) = a, \end{cases}$$

together with $\|\dot{\tilde{a}}(s)\|_2 \leq L$ for all $s > 0$.

For any compact set $K \subset \mathbb{R}^n \times \mathbb{R}^m$ and $s_0 > 0$ to be choose later, we define $U_L := g_L(K, [0, s_0])$. Then for all $(x, a) \in K$, we find that

$$\begin{aligned} |Q^{L, \varepsilon_k}(x, a) - \hat{Q}^L(x, a)| & \leq \int_0^\infty e^{-\gamma s} |r^{\varepsilon_k}(x_L^{\varepsilon_k}(s), a_L^{\varepsilon_k}(s)) - r(\hat{x}_L(s), \hat{a}_L(s))| ds \\ & \leq \int_0^{s_0} e^{-\gamma s} |r^{\varepsilon_k}(x_L^{\varepsilon_k}(s), a_L^{\varepsilon_k}(s)) - r(\hat{x}_L(s), \hat{a}_L(s))| ds + Ce^{-\gamma s_0} \\ & \leq \int_0^{s_0} e^{-\gamma s} |r^{\varepsilon_k}(x_L^{\varepsilon_k}(s), a_L^{\varepsilon_k}(s)) - r^{\varepsilon_k}(\hat{x}_L(s), \hat{a}_L(s))| ds \\ & \quad + \int_0^{s_0} e^{-\gamma s} |r^{\varepsilon_k}(\hat{x}_L(s), \hat{a}_L(s)) - r(\hat{x}_L(s), \hat{a}_L(s))| ds + Ce^{-\gamma s_0} \\ & \leq C_1 |s_0| (\|x^{\varepsilon_k} - \hat{x}\|_{L^\infty([0, s_0])} + \|a^{\varepsilon_k} - \hat{a}\|_{L^\infty([0, s_0])}) \\ & \quad + |s_0| \|r^{\varepsilon_k} - r\|_{L^\infty(U_L)} + C_2 e^{-\gamma s_0}, \end{aligned}$$

where the last inequality comes from

$$\|f^\varepsilon\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \leq \|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \quad \text{and} \quad \|r^\varepsilon\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)} \leq \|r\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}.$$

For small $\delta > 0$ given, we choose $s_0 > 0$ such that

$$C_2 e^{-\gamma s_0} < \frac{\delta}{2}.$$

Then U is determined and we choose $\varepsilon_k > 0$ sufficiently small to satisfy

$$C_1 |s_0| (\|x^{\varepsilon_k} - \hat{x}\|_{L^\infty([0, s_0])} + \|a^{\varepsilon_k} - \hat{a}\|_{L^\infty([0, s_0])}) + |s_0| \|r^{\varepsilon_k} - r\|_{L^\infty(U_L)} < \frac{\delta}{2}.$$

This implies that $Q^{L, \varepsilon_k}(x, a)$ converges to $\hat{Q}^L(x, a)$ uniformly on compact sets. Moreover, $Q^{L, \varepsilon}$ is the unique viscosity solution of

$$\gamma Q^{L, \varepsilon} - f^\varepsilon(x, a) \cdot D_x Q^{L, \varepsilon} - L \|D_a Q^{L, \varepsilon}\|_2 - r^\varepsilon(x, a) = 0.$$

Then, by the stability of the viscosity solution [2][Proposition 2.2, Chapter II], we find that $\hat{Q}^L(x, a)$ is the viscosity solution of (2.4). Consequently, $\hat{Q}^L = Q^L$. Since $\hat{a}(\cdot) \in \mathcal{A}^L$ and $\hat{x}(\cdot)$, $\hat{a}(\cdot)$ satisfy the dynamic equation (4.6), it follows that $\hat{a}^L(\cdot)$ is the desired optimal policy. This completes the proof. \square

We now get back to our original interest, the rate of convergence of Q^L to Q as $L \rightarrow \infty$. To understand such a quantitative property, we introduce structural assumptions on the dynamics and reward function.

Assumption 4.6. There exists $C > 0$ such that

$$\|f(x, a)\|_2 + |r(x, a)| \leq \frac{C}{\|a\|_2^\sigma}$$

for some $\sigma > 0$.

The following assumption indicates that an optimal control from Proposition 4.4 is not changing abruptly.

Assumption 4.7. Given $x \in \mathbb{R}^n$, let $a(\cdot) \in \mathcal{A}$ be an optimal control obtained in Proposition 4.4, that is,

$$Q(x) = \int_0^\infty e^{-\gamma s} r(x(s), a(s)) ds.$$

For this control, we assume that there exists $\beta > 0$,

$$\|\tilde{a}(s) - a^\varepsilon(s)\|_{L^2(\mathbb{R})} \leq O(\varepsilon^\beta).$$

where $a^\varepsilon(s) = \tilde{a}(s) * \eta^\varepsilon$ and $\tilde{a}(s)$ is odd extension of $a(s)$ ³.

Remark 4.8. If $a(\cdot)$ is piecewise continuous and $|a(\cdot)| \leq M$ for some $M > 0$, then $\beta = 1$. In general if $\tilde{a} \in W^{k, 2}(\mathbb{R})$ for some $k \geq 1$, then we can choose $\beta = k$.

We now state the result on a rate of convergence of $Q^L \rightarrow Q$. The idea of the proof is that we truncate the optimal control $a(s)$ and consider the mollification of this function to relate it with the feasible class \mathcal{A}^L .

Theorem 4.9. Let $\gamma \geq \|f\|_{Lip(\mathbb{R}^n \times \mathbb{R}^m)} + 1$ and $(x, a) \in \mathbb{R}^n \times \mathbb{R}^m$ be given. Under Assumption 4.7, we have that

$$|Q^L(x, a) - Q(x)| = O\left(\frac{\log L}{L^{\sigma/2}} + \left(\frac{\log L}{\sqrt{L}}\right)^\beta + \frac{1}{\sqrt{L}}\right).$$

Proof. Throughout the proof, we will use the mollifier η_ε defined for $n = 1$ in (1.1). Let $(x(s), a(s))$ and $(x_L(s), a_L(s))$ be optimal state trajectory and control pairs associated with $Q(x)$ and $Q^L(x, a)$ respectively.

Given $R > 0$ and $a(s) = (a_1(s), \dots, a_m(s))$, define

$$a_i^R(s) := a_i(s) \mathbf{1}_{|a_i(s)| \leq R} + R \mathbf{1}_{\{a_i(s) \geq R\}} - R \mathbf{1}_{\{a_i(s) \leq -R\}}$$

and let $a^R(s) := (a_1^R(s), \dots, a_m^R(s))$.

³The measurable function $a : [0, \infty) \rightarrow \mathbb{R}^m$ can be extended to $\tilde{a} : (-\infty, \infty) \rightarrow \mathbb{R}^m$ in a way that $a(-s) = -a(s)$ and mollified with η_ε with $n = 1$ in (1.1).

Since $Q(x) \geq Q^L(x, a)$, we only need to estimate $Q(x) - Q^L(x, a)$. Fixing $T > 0$, we have that

$$\begin{aligned} Q(x) - Q^L(x, a) &\leq \int_0^\infty e^{-\gamma s} \left(r(x(s), a(s)) - r(x_L(s), a_L(s)) \right) ds \\ &\leq \int_0^T e^{-\gamma s} \left(r(x(s), a(s)) - r(x(s), a^R(s)) \right) ds \\ &\quad + \int_0^T e^{-\gamma s} \left(r(x(s), a^R(s)) - r(x_L(s), a_L(s)) \right) ds + Ce^{-T} \end{aligned} \quad (4.9)$$

for some C depending only on γ and C_1 from Assumption 2.1. We notice that the first term of (4.9) can be bounded as

$$\begin{aligned} \int_0^T e^{-\gamma s} \left(r(x(s), a(s)) - r(x(s), a^R(s)) \right) ds &\leq CT \min\left\{ \frac{1}{R^\sigma}, \|a(s) - a^R(s)\|_2 \right\} \\ &\leq \frac{CT}{R^\sigma}. \end{aligned}$$

For the second part of (4.9), we assume for the moment that R and ε satisfy

$$\left| \frac{da^{R,\varepsilon}(s)}{ds} \right| \leq \frac{eR}{\varepsilon} = L,$$

where $a_i^{R,\varepsilon}(s) := a_i^R(s) * \eta_\varepsilon$, and hence, $a^{R,\varepsilon}(s) \in \mathcal{A}^L$. Therefore, the second term in (4.9) is estimated as

$$\begin{aligned} \int_0^T e^{-\gamma s} \left(r(x(s), a^R(s)) - r(x_L(s), a_L(s)) \right) ds &= \int_0^T e^{-\gamma s} \left(r(x(s), a^R(s)) - r(x_L(s), a_L(s)) \right) ds \\ &\leq \underbrace{\int_0^T e^{-\gamma s} \left(r(x(s), a^R(s)) - r(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s)) \right) ds}_{(a)}, \end{aligned}$$

where $x^{R,\varepsilon}(s)$ satisfies

$$\begin{cases} (x^{R,\varepsilon})'(s) = f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s)) & \text{for } s > 0, \\ x^{R,\varepsilon}(0) = x, \\ a^{R,\varepsilon}(0) = a. \end{cases}$$

Now observing (a),

$$\begin{aligned} (a) &\leq \int_0^T e^{-\gamma s} |r(x(s), a^R(s)) - r(x(s), a^{R,\varepsilon}(s))| ds + \int_0^T e^{-\gamma s} |r(x(s), a^{R,\varepsilon}(s)) - r(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))| ds \\ &\leq C \|a^R(s) - a^{R,\varepsilon}(s)\|_{L^2([0,\infty))}^2 + \int_0^T e^{-\gamma s} |r(x(s), a^{R,\varepsilon}(s)) - r(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))| ds \\ &\leq C \|a(s) - a^\varepsilon(s)\|_{L^2([0,\infty))}^2 + \int_0^T e^{-\gamma s} |r(x(s), a^{R,\varepsilon}(s)) - r(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))| ds, \end{aligned} \quad (4.10)$$

where $a^\varepsilon(s) := (a_1(s) * \eta_\varepsilon, \dots, a_m(s) * \eta_\varepsilon)$.

Let $w(s) := \frac{1}{2}\|x(s) - x^{R,\varepsilon}(s)\|_2^2$ and $\tilde{C} = 2\|f\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$.

$$\begin{aligned}
w'(s) &= (x(s) - x^{R,\varepsilon}(s)) \cdot \left(f(x(s), a(s)) - f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s)) \right) \\
&= (x(s) - x^{R,\varepsilon}(s)) \cdot \left(f(x(s), a(s)) - f(x(s), a^{R,\varepsilon}(s)) + f(x(s), a^{R,\varepsilon}(s)) - f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s)) \right) \\
&\leq \|x(s) - x^{R,\varepsilon}(s)\|_2 \times \\
&\quad \left(\|f(x(s), a(s)) - f(x(s), a^{R,\varepsilon}(s))\|_2 + \|f(x(s), a^{R,\varepsilon}(s)) - f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))\|_2 + \frac{\tilde{C}}{2}\|x(s) - x^{R,\varepsilon}(s)\|_2 \right) \\
&\leq \tilde{C}w + \|x(s) - x^{R,\varepsilon}(s)\|_2 \left(\|f(x(s), a(s)) - f(x(s), a^{R,\varepsilon}(s))\|_2 + \|f(x(s), a^{R,\varepsilon}(s)) - f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))\|_2 \right) \\
&\leq \tilde{C}w + 2w + \frac{1}{2} \left(\|f(x(s), a(s)) - f(x(s), a^{R,\varepsilon}(s))\|_2^2 + \|f(x(s), a^{R,\varepsilon}(s)) - f(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))\|_2^2 \right) \\
&\leq (\tilde{C} + 2)w + \frac{C}{R^{2\sigma}} + \frac{\tilde{C}}{4}\|a^R(s) - a^{R,\varepsilon}(s)\|_2^2.
\end{aligned}$$

By the Gronwall's inequality, for any $t \in [0, T]$, we have that

$$\begin{aligned}
e^{-(\tilde{C}+2)t}\|x(t) - x^{R,\varepsilon}(t)\|_2^2 &\leq C \left(\frac{1}{R^{2\sigma}} + \int_0^t \|a^R(s) - a^{R,\varepsilon}(s)\|_2^2 ds \right) \\
&\leq C \left(\frac{1}{R^{2\sigma}} + \|a(s) - a^\varepsilon(s)\|_{L^2([0,\infty))}^2 \right).
\end{aligned}$$

Noticing that $2\gamma \geq \tilde{C} + 2$, we have

$$\begin{aligned}
\int_0^T e^{-\gamma s} |r(x(s), a^{R,\varepsilon}(s)) - r(x^{R,\varepsilon}(s), a^{R,\varepsilon}(s))| ds &\leq C \int_0^T e^{-\gamma s} \|x(s) - x^{R,\varepsilon}(s)\|_2 ds \\
&\leq C \left(\int_0^T 1 ds \right)^{1/2} \left(\int_0^T e^{-2\gamma s} \|x(s) - x^{R,\varepsilon}(s)\|_2^2 ds \right)^{1/2} \\
&\leq CT \left(\frac{1}{R^{2\sigma}} + \|a(s) - a^\varepsilon(s)\|_{L^2([0,\infty))}^2 \right).
\end{aligned}$$

Finally, we set $R = \sqrt{L}$, $T = \log R$, and $\varepsilon = \frac{eR}{L}$,

$$\begin{aligned}
Q(x) - Q^L(x, a) &\leq C \left(\frac{T}{R^\sigma} + T\|a(s) - a^\varepsilon(s)\|_{L^2([0,\infty))}^2 + e^{-T} \right) \\
&\leq O\left(\frac{T}{R^\sigma} + T\varepsilon^{2\beta} + \frac{1}{\sqrt{L}}\right) \\
&\leq O\left(\frac{\log L}{L^{\sigma/2}} + \left(\frac{\log L}{\sqrt{L}}\right)^\beta + \frac{1}{\sqrt{L}}\right).
\end{aligned}$$

□

5. GENERALIZED HAMILTON-JACOBI BASED Q-LEARNING

The core idea of Hamilton-Jacobi based Q-learning for the Lipschitz continuous control problem proposed by [11] is to restrict the derivative of the action $a(\cdot)$ to be bounded, that is, $|\dot{a}(\cdot)| \leq L$ for some $L > 0$, which results in

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - L\|D_a Q^L\|_2 - r(x, a) = 0$$

for Q^L defined as (2.3). We now generalize the constraint by introducing

$$\mathcal{A}_p^L := \{a(\cdot) \in \mathcal{A}^L : \|\dot{a}(\cdot)\|_p \leq L \text{ for all } s \in [0, \infty)\}$$

for $p \in [1, \infty]$. Here, we note that $\mathcal{A}_2^L = \mathcal{A}^L$ when $p = 2$.

The value function is defined as

$$Q_p^L(x, a) = \sup_{a \in \mathcal{A}_p^L} \left\{ \int_0^\infty e^{-\gamma s} r(x(s), a(s)) dt : x(0) = x, a(0) = a \right\}, \quad (5.1)$$

and we obtain a slightly different Hamilton-Jacobi equation

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a) - \sup_{\|b\|_p \leq L} b \cdot D_a Q = \gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a) - L\|D_a Q\|_q = 0,$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

5.1. Demonstration of optimal action. Given L and p , let us examine the structural property of optimal action when the value function $Q_p^L(x, a)$ is differentiable everywhere.

Let us illustrate some special cases $p = 1$ or $p = \infty$ first and consider general $p \in (1, \infty)$. When $p = 1$, $Q^L := Q_1^L$ is a viscosity solution to

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a) - L \max_{i \in [1, m]} \{|D_{a_i} Q^L|\} = 0.$$

The optimal action $a(\cdot) = (a_1(\cdot), \dots, a_m(\cdot))$ satisfies that

$$\dot{a}_j(\cdot) = \begin{cases} \pm L & \text{if } j = \max_{i \in [1, m]} |D_{a_i} Q^L(x(\cdot), a(\cdot))|, \\ 0 & \text{if } D_{a_i} Q^L(x(\cdot), a(\cdot)) \neq 0 \text{ and } i \neq j. \end{cases}$$

On the other hand, if $p = \infty$, we also derive that $Q^L := Q_\infty^L(x, a)$ solves

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - r(x, a) - L \sum_{i=1}^m |D_{a_i} Q^L| = 0$$

in the viscosity sense. Similarly, the optimal action satisfies that

$$\dot{a}_i(\cdot) = \begin{cases} L & \text{if } D_{a_i} Q^L(x(\cdot), a(\cdot)) > 0, \\ -L & \text{if } D_{a_i} Q^L(x(\cdot), a(\cdot)) < 0. \end{cases}$$

One interesting feature of implementing ℓ_∞ constraint is that the rate of change of control $\dot{a}_i(\cdot)$ is independent of the coordinate while that under ℓ_2 constraint is not. Specifically, $\dot{a}_1(t) = L$ for some t fixed does not affect the choice of $\dot{a}_2(t)$ under ℓ_∞ constraint but the condition forces $\dot{a}_i(t) = 0$ for $i \neq 0$.

We now characterize the optimal control associated with $p \in (1, \infty)$.

Proposition 5.1. *Let $p \in (1, \infty)$ and $Q_p^L(x, a)$ from (5.1) be differentiable. If $D_{a_i} Q_p^L(x, a) \neq 0$ for all i , then optimal action $a(\cdot)$ satisfies*

$$\dot{a}_i(\cdot) = \pm L \frac{(D_{a_i} Q^L)^{1/(p-1)}}{\|D_a Q^L\|_q^{q/p}},$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

5.2. Convergence result. Regardless of the choice of p , invoking the same argument presented in Theorem 4.2, we still have the general convergence result.

Corollary 5.2. *Let $p \in [1, \infty]$ and $Q^L := Q_p^L(x, a) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a viscosity solution to*

$$\gamma Q^L - D_x Q^L \cdot f(x, a) - L \|D_a Q^L\|_q - r(x, a) = 0 \quad (5.2)$$

for q satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Then we have that

$$Q^L(x, a) \rightarrow Q(x) \quad \text{locally uniformly as } L \rightarrow \infty,$$

where $Q := Q(x)$ is a unique viscosity solution to

$$\gamma Q - \sup_{a \in \mathbb{R}^m} (D_x Q \cdot f(x, a) + r(x, a)) = 0.$$

6. NUMERICAL EXPERIMENT

This section presents numerical experiments conducted using the HJDQN learning framework, with different values of L and p . We assess performance in the Hopper-v2, HalfCheetah-v2, and Walker2d-v2 environments from the OpenAI Gym suite [6], simulated using the MuJoCo engine [16]. Additionally, we include a 20D linear quadratic regulator (LQR) control problem as a benchmark task. Our experiments are based on the code provided by [11]⁴. Unless otherwise stated, the hyperparameters remain consistent with those reported in [11].

6.1. Proposed algorithm. Yet almost identical to the procedure presented in [11], we include the modified algorithm for completeness.

⁴<https://github.com/HJDQN/HJQ>

Algorithm 1 Modified Hamilton–Jacobi DQN (p -HJDQN)

```

1:  $L, p$  and  $h$  are given;
2: Initialization: Initialize  $Q$ -function by a neural network with random weight  $\theta$  as  $Q_\theta$ , and the
   target  $Q$ -function as  $Q_{\theta^-}$  with weights  $\theta^- = \theta$ ;
3: Initialize replay buffer with fixed capacity;
4: for episode=1 to  $M$  do
5:   Sample initial state-action pair  $(x_0, a_0)$ ;
6:   for  $k=0$  to  $K$  do
7:     Execute action  $a_k$  and observe reward  $r_k$  and the next state  $x_{k+1}$ ;
8:     Store  $(x_k, a_k, r_k, x_{k+1})$  in buffer;
9:     Sample the random mini-batch  $\{(x_j, a_j, r_j, x_{j+1})\}$  from buffer;
10:    Compute  $\eta$  via Proposition 5.1;
11:    Set  $y_j^- := hr_j + (1 - \gamma h)Q_{\theta^-}(x_{j+1}, a'_j)$  for all  $j$  where  $a'_j := a_j \pm hL \frac{(D_{a_j} Q^L)^{1/(p-1)}}{\|D_a Q^L\|_q^{q/p}}$ .
12:    Update  $\theta$  by minimizing  $\sum_j (y_j^- - Q_\theta(x_j, a_j))^2$ ;
13:     $\theta^- \leftarrow (1 - \alpha)\theta^- + \alpha\theta$  for  $\alpha$  small positive;
14:    Set the next action as  $a_{k+1} := a_k + h\eta + \varepsilon$ , where  $\varepsilon \sim N(0, \sigma^2 I_m)$ ;
15:   end for
16: end for

```

6.2. Numerical analysis on convergence result. This subsection provides empirical evidence on the convergence result in Lemma 3.2. To this end, we respectively sample 500 states and actions and then compute $\max_{1 \leq i, j \leq 500} |Q_L(x_i, a_j) - Q_{L+10}(x_i, a_j)|$ for different values of L from $L = 10, \dots, 150$.

The numerical values of $\max_{1 \leq i, j \leq 500} |Q_L(x_i, a_j) - Q_{L+10}(x_i, a_j)|$ for four different tasks are presented in Table 1. Figure 2 illustrates the corresponding line graphs derived from the data in Table 1. A decreasing trend of $\|Q_{L+10} - Q_L\|_{\text{Lip}(\mathbb{R}^n \times \mathbb{R}^m)}$ is observed as L increases, providing empirical support for Lemma 3.2. However, this convergence trend exhibits fluctuations and instability, particularly in the HalfCheetah-v2 and 20-dimensional LQ problem environments. Further investigation is warranted to understand the causes of instability and to refine the HJDQN learning framework.

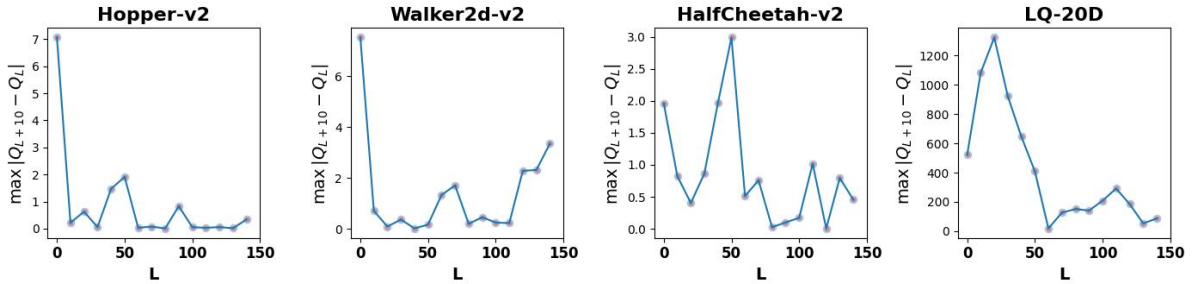


FIGURE 1. Plot for the Table 1

6.3. Numerical analysis on different values of $1 \leq p \leq \infty$. In this section, we explore the impact of varying p values on the results. We examine p values of 1, 2, 10, 100 and ∞ . As a benchmark, we utilize the DDPG algorithm, as introduced in [13]. We present the learning curves for different tasks and p values below.

It's noteworthy that in the HalfCheetah-v2 environment, DDPG consistently outperforms all HJDQN variants for $p = 1, 2, 10, 100$ and ∞ . Also, the standard HJDQN setting ($p = 2$) demonstrates superior performance over other p values.

Conversely, the HJDQN framework exhibits enhanced performance in the Walker2d-v2 and 20-dimensional LQ problems. Specifically, in Walker2d-v2, we observe an upward trend in the average return as p increases, highlighting the influence of p can differ across various tasks.

Subsequently, we compare action trajectories between DDPG and HJDQN across different p values, as illustrated in Figure 3, 4 and 5, 6 and 7. A notable observation is the increased oscillation frequency in the actions with rising p values, indicating a potential impact on system stability or efficiency.

	Hopper-v2	Walker2d-v2	HalfCheetah-v2	LQR-20D
$ Q_{20} - Q_{10} $	7.0607	7.5138	1.9608	522.4116
$ Q_{30} - Q_{20} $	0.2177	0.7117	0.8246	1082.9729
$ Q_{40} - Q_{30} $	0.6256	0.0996	0.4034	1323.8116
$ Q_{50} - Q_{40} $	0.0577	0.3741	0.8640	922.2965
$ Q_{60} - Q_{50} $	1.4704	0.0268	1.9657	646.7611
$ Q_{70} - Q_{60} $	1.9120	0.1756	2.9905	408.9908
$ Q_{80} - Q_{70} $	0.0351	1.3410	0.5117	18.6145
$ Q_{90} - Q_{80} $	0.0746	1.7137	0.7575	126.5660
$ Q_{100} - Q_{90} $	0.0093	0.2049	0.0248	151.0665
$ Q_{110} - Q_{100} $	0.8314	0.4649	0.1005	140.7522
$ Q_{120} - Q_{110} $	0.0557	0.2555	0.1706	208.1184
$ Q_{130} - Q_{120} $	0.0317	0.2358	1.0160	291.1623
$ Q_{140} - Q_{130} $	0.0586	2.2847	0.0063	184.8106
$ Q_{150} - Q_{140} $	0.0155	2.3188	0.7942	53.1037
$ Q_{160} - Q_{150} $	0.3373	3.3487	0.4554	86.3219

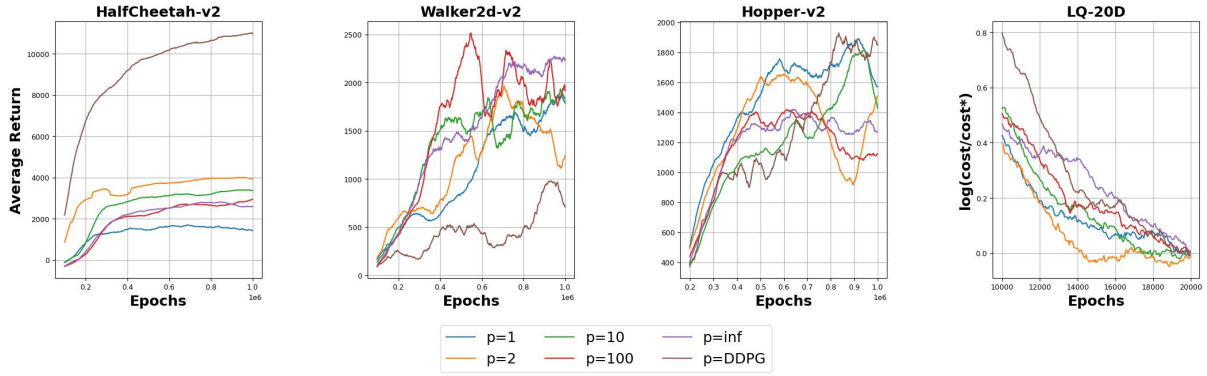
TABLE 1. Numerical calculation of $\|Q^{L+10} - Q^L\|_{L^\infty(\mathbb{R}^n \times \mathbb{R}^m)}$ for different tasks.

FIGURE 2. Learning curves for different tasks.

7. DISCUSSIONS

We study the behavior of the value function for optimal control problems where the control is restricted to be L -Lipschitz continuous, which arises in a novel HJB-based Q -learning. By augmenting the control variable, we formulate the problem in the standard optimal control problem where the control takes value in a compact set and investigates the effect of L , particularly, stability properties and the rate of convergence. We extend the theory by introducing p -norm and discover that different choice of p indeed leads to different features of optimal control. We observe that the larger p induces a more frequent change of control.

The choice of p is subtle. HJDQN with $p = 2$ and DDPG are not always leading to.

Conflict of interest The authors declare that they have no conflict of interest.

Data availability Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

REFERENCES

1. Yves Achdou, Guy Barles, Hitoshi Ishii, Grigory L Litvinov, and Guy Barles, *An introduction to the theory of viscosity solutions for first-order Hamilton–Jacobi equations and applications*, Hamilton–Jacobi Equations: Approximations, Numerical Analysis and Applications: Cetraro, Italy 2011, Editors: Paola Loreti, Nicoletta Anna Tchou (2013), 49–109.
2. Martino Bardi, Italo Capuzzo Dolcetta, et al., *Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations*, vol. 12, Springer, 1997.
3. Richard Bellman, *Dynamic programming*, Science **153** (1966), no. 3731, 34–37.
4. Kim C Border, *Introduction to correspondences*, 2013.
5. Jonathan M Borwein, Jon D Vanderwerff, et al., *Convex functions: constructions, characterizations and counterexamples*, vol. 109, Cambridge University Press Cambridge, 2010.

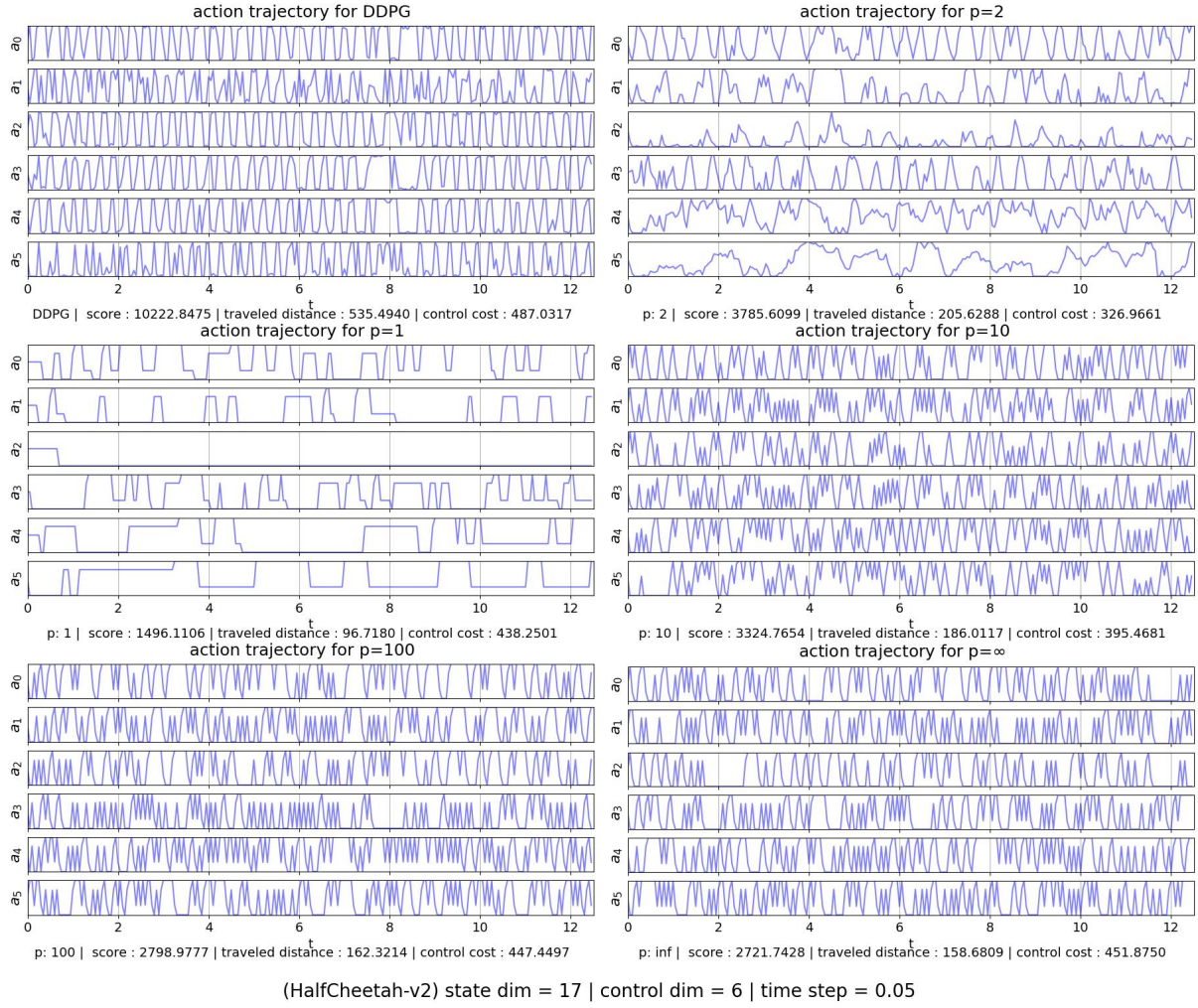


FIGURE 3. Action trajectories obtained by different values of p in HJDQN and DDPG for HalfCheetah-v2.

6. Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba, *Openai gym*, arXiv preprint arXiv:1606.01540 (2016).
7. Michael G Crandall and Pierre-Louis Lions, *Viscosity solutions of Hamilton–Jacobi equations*, Transactions of the American Mathematical Society **277** (1983), no. 1, 1–42.
8. Lawrence C Evans, *Partial Differential Equations*, vol. 19, American Mathematical Society, 2022.
9. A Hitchhiker’s Guide, *Infinite dimensional analysis*, Springer, 2006.
10. Yanwei Jia and Xun Yu Zhou, *q-learning in continuous time*, Journal of Machine Learning Research **24** (2023), no. 161, 1–61.
11. Jeongho Kim, Jaek Shin, and Insoon Yang, *Hamilton–Jacobi deep Q-learning for deterministic continuous-time systems with Lipschitz continuous controls*, The Journal of Machine Learning Research **22** (2021), no. 1, 9363–9396.
12. Yoonseung Kim, Hung V Tran, and Son N Tu, *State-constraint static Hamilton–Jacobi equations in nested domains*, SIAM Journal on Mathematical Analysis **52** (2020), no. 5, 4161–4184.
13. Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra, *Continuous control with deep reinforcement learning*, arXiv preprint arXiv:1509.02971 (2015).
14. Hiroyoshi Mitake and Hung V Tran, *Dynamical properties of Hamilton–Jacobi equations via the nonlinear adjoint method: large time behavior and discounted approximation*, Dynamical and geometric aspects of Hamilton–Jacobi and linearized Monge–Ampère equations—VIASM **2183** (2016), 125–228.
15. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al., *Human-level control through deep reinforcement learning*, Nature **518** (2015), no. 7540, 529–533.
16. Emanuel Todorov, Tom Erez, and Yuval Tassa, *Mujoco: A physics engine for model-based control*, 2012 IEEE/RSJ international conference on intelligent robots and systems, IEEE, 2012, pp. 5026–5033.
17. Hung V Tran, *Hamilton–Jacobi equations: Theory and Applications*, vol. 213, American Mathematical Society, 2021.
18. Haoran Wang, Thaleia Zariphopoulou, and Xun Yu Zhou, *Reinforcement learning in continuous time and space: A stochastic control approach*, Journal of Machine Learning Research **21** (2020), no. 198, 1–34.
19. Christopher JCH Watkins and Peter Dayan, *Q-learning*, Machine Learning **8** (1992), 279–292.

20. Cagatay Yildiz, Markus Heinonen, and Harri Lähdesmäki, *Continuous-time model-based reinforcement learning*, International Conference on Machine Learning, PMLR, 2021, pp. 12009–12018.

8. APPENDIX

8.1. Existence of ψ in (4.4). In this subsection, we prove the existence a Borel function satisfying (4.4) using Borel selection theorem for set-valued functions. Since \tilde{Q} is a semiconcave, there exists a constant $C > 0$ and a convex function $u : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$\tilde{Q}(x) = C|x|^2 - u(x).$$

It is enough to show that there exists a Borel function satisfying $g(x) \in D_x^*u(x)$. We denote by $2^{\mathbb{R}^n}$ the power set of \mathbb{R}^n . Let us define a set-valued function $T : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$ by $T(x) = D_x^-u(x)$ and $T_{D_x^*u}(x) = D_x^*u(x)$.

By [5][Proposition 6.1.1], for all closed set $F \subset \mathbb{R}^n$, we have

$$T^{-1}(F) := \{x \in \mathbb{R}^n : T(x) \cap F \neq \emptyset\} \text{ is closed.}$$

From the definition of D_x^*u , it is clear that $T_{D_x^*u}(x)$ is a closed multi-valued function, see [4][Definition 16], for the details. Since \tilde{Q} is Lipschitz for each $x \in \mathbb{R}^n$, so is u which implies that $T(x)$ is compact-valued for each $x \in \mathbb{R}^n$. By the definition, for $\tilde{T}(x) := T(x) \cap T_{D_x^*u}(x)$, $\tilde{T}(x) = T_{D_x^*u}(x)$ and it also satisfies the condition

$$\tilde{T}^{-1}(F) = \{x \in \mathbb{R}^n : \tilde{T}(x) \cap F \neq \emptyset\} \text{ is closed for any closed set } F,$$

by [9][Theorem 17.25]. Next, according to the Kuratowski–Ryll–Nardzewski Selection Theorem, as referenced in [9][18.13], there exists a Borel function such that $g(x) \in D_x^*u$, which completes the proof. For more details, we refer to [9][Chapters 17 and 18], [5][Chapter 6] and [4], which discuss selection theory related to set-valued functions and convex functions.

8.2. Action plots on Walker2d-v2 and 20-dimensional LQ problems. We provide action trajectories plots for Walker-2D and 20-dimensional LQ problems in this subsection.

CENTER FOR MATHEMATICAL MACHINE LEARNING AND ITS APPLICATIONS, POSTECH, 77 CHEONGAM-RO, NAM-GU, POHANG, GYEONGBUK, KOREA

Email address: namkyeong.cho@gmail.com

DEPARTMENT OF APPLIED ARTIFICIAL INTELLIGENCE, SEOULTECH, 232 GONGNEUNG-RO, NOWON-GU, SEOUL, KOREA

Email address: kimyeoneung@gmail.com

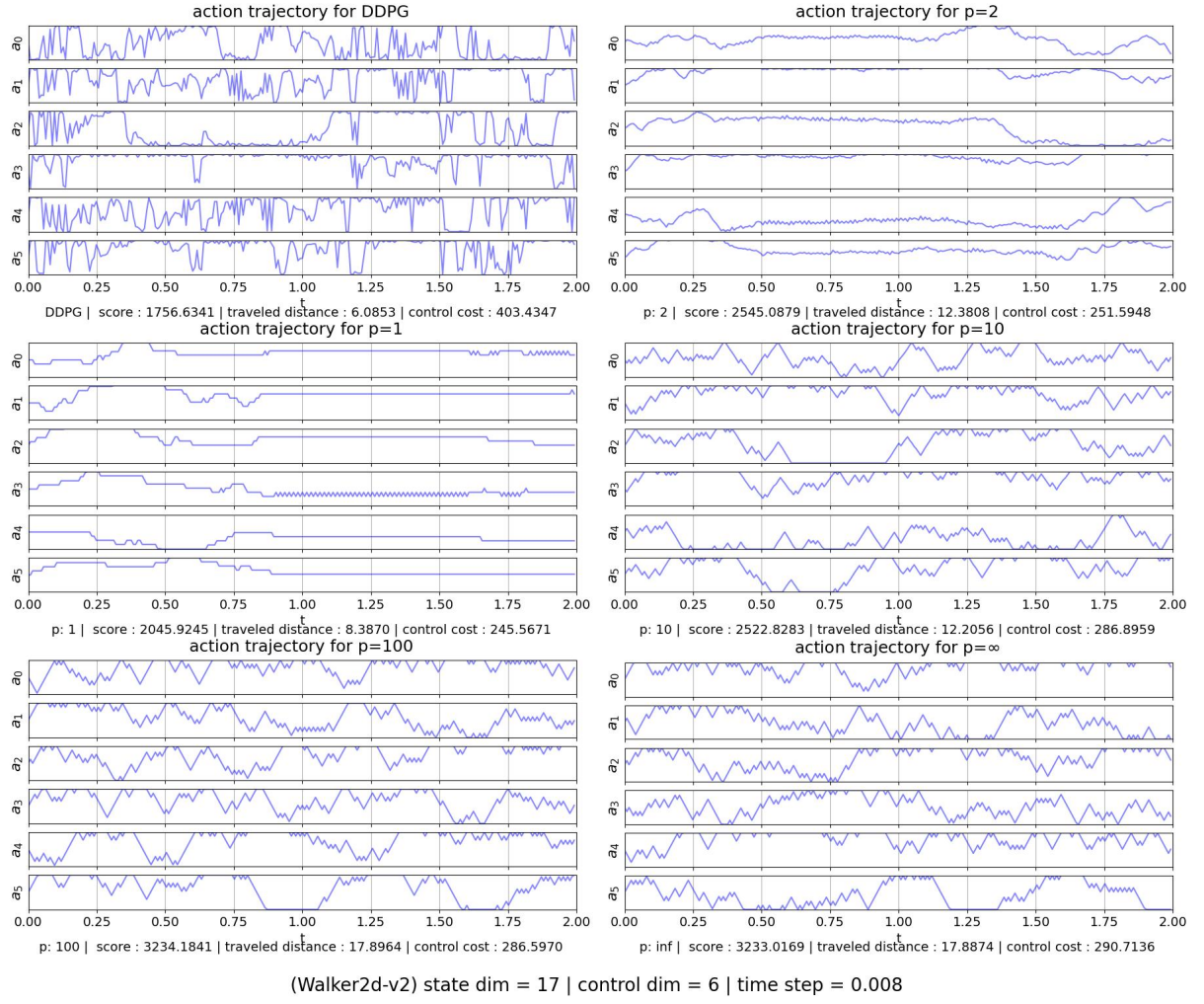


FIGURE 4. Action trajectories obtained by different values of p in HJDQN and DDPG for Walker2d-v2.

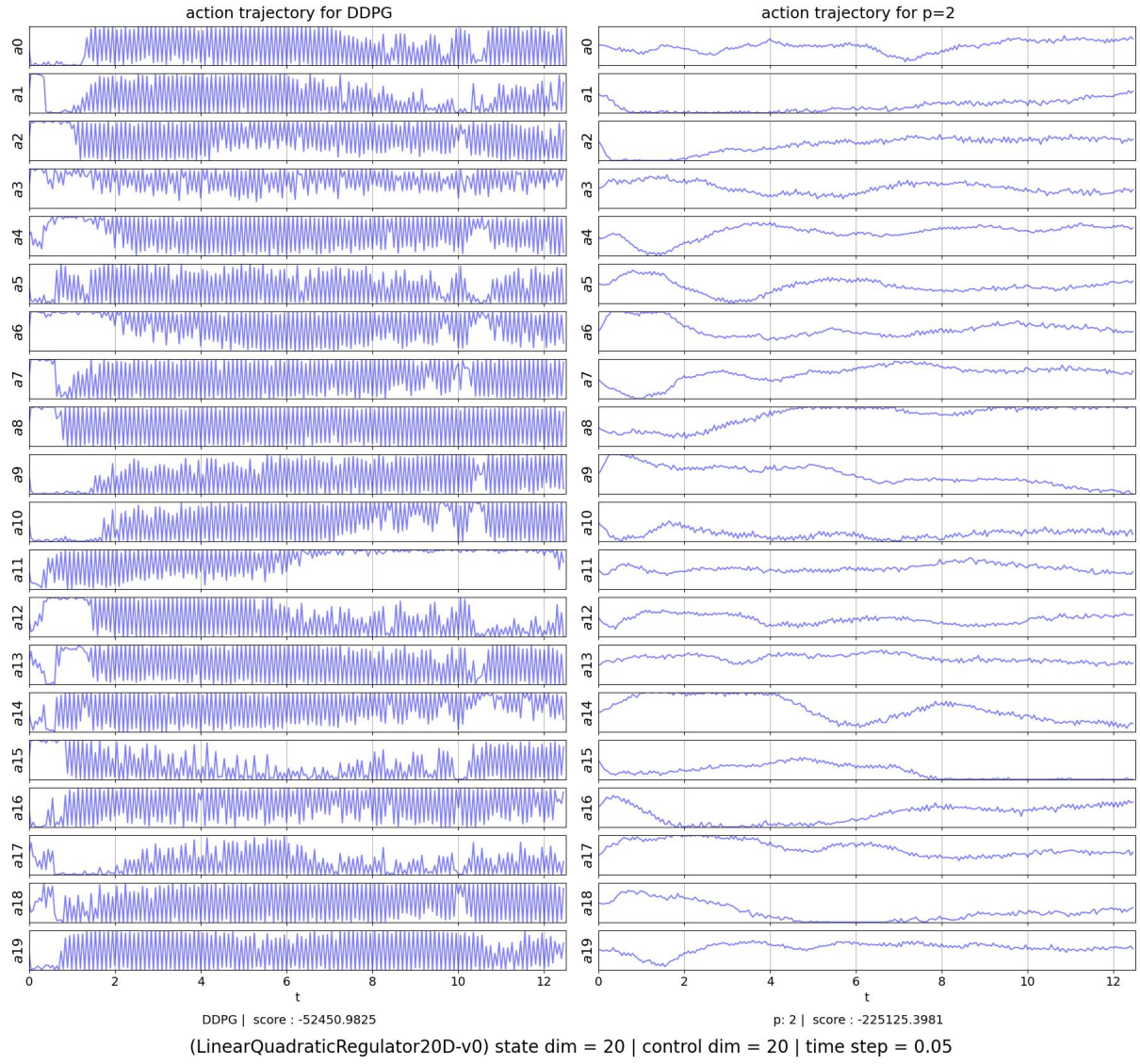


FIGURE 5. Action trajectories obtained by HJDQN with $p = 2$ and DDPG for 20D LQR problem.

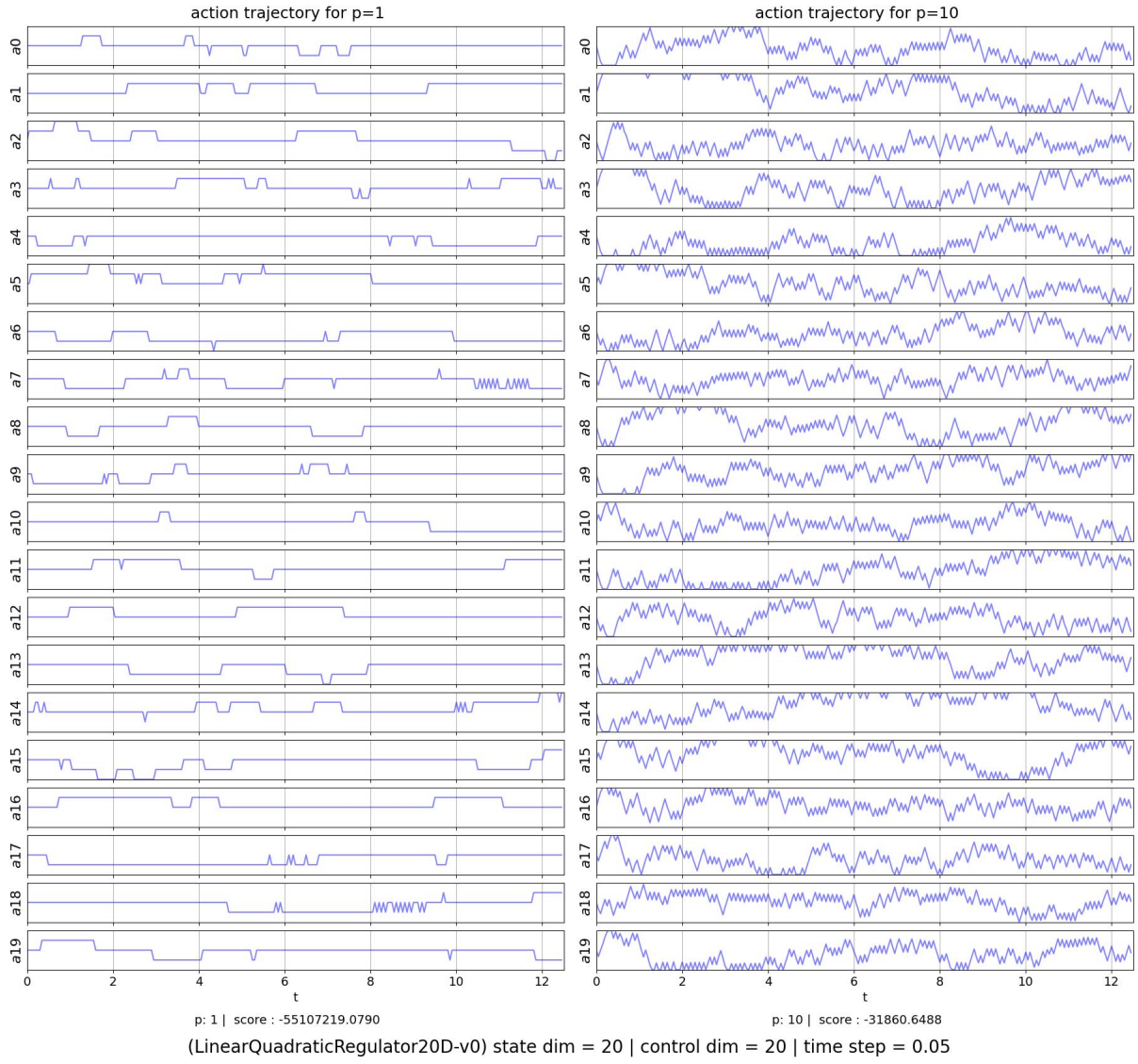


FIGURE 6. Action trajectories obtained by HJDQN with $p = 1$ and $p = 10$ for 20D LQR problem.

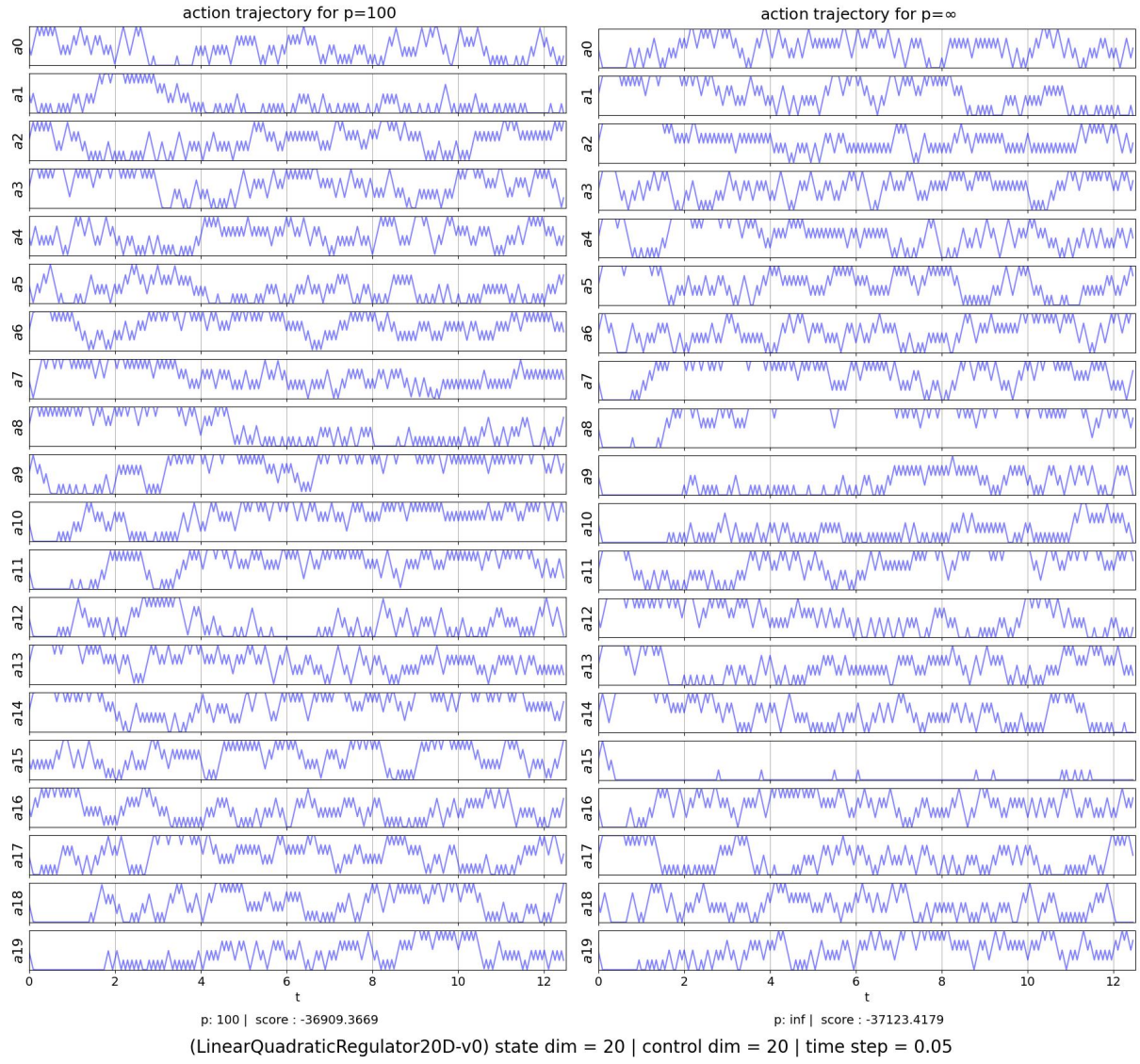


FIGURE 7. Action trajectories obtained by HJDQN with $p = 100$ and ∞ for LQR problem with 20 dimension.