

Continuity of cost in Borkar control topology and implications on discrete space and time approximations for controlled diffusions under several criteria

Somnath Pradhan* Serdar Yüksel†

Abstract

We first show that the discounted cost, cost up to an exit time, and ergodic cost involving controlled non-degenerate diffusions are continuous on the space of stationary control policies when the policies are given a topology introduced by Borkar [V. S. Borkar, A topology for Markov controls, *Applied Mathematics and Optimization* 20 (1989), 55–62]. The same applies for finite horizon problems when the control policies are Markov and the topology is revised to include time also as a parameter. We then establish that finite action/piecewise constant stationary policies are dense in the space of stationary Markov policies under this topology and the same holds for continuous policies. Using the above mentioned continuity and denseness results we establish that finite action/piecewise constant policies approximate optimal stationary policies with arbitrary precision. This gives rise to the applicability of many numerical methods such as policy iteration and stochastic learning methods for discounted cost, cost up to an exit time, and ergodic cost optimal control problems in continuous-time. For the finite-horizon setup, we establish additionally near optimality of time-discretized policies by an analogous argument. We thus present a unified and concise approach for approximations directly applicable under several commonly adopted cost criteria.

Keywords: controlled diffusions; near optimality; piecewise constant policy; finite actions; Hamilton-Jacobi-Bellman equation.

MSC2020 subject classifications: Primary 93E20; 60J60, Secondary 35Q93.

Submitted to EJP on November 10, 2022, final version accepted on January 30, 2024.

*Department of Mathematics and Statistics, Queen’s University, Kingston, ON, Canada.
E-mail: sp165@queensu.ca

†Department of Mathematics and Statistics, Queen’s University, Kingston, ON, Canada.
E-mail: yuksel@queensu.ca

1 Introduction

In this paper, we study regularity properties of induced cost (under several criteria) on a controlled diffusion process with respect to a control topology defined by Borkar [15], and implications of these properties on existence and, in particular, approximations for optimal controlled diffusions. We will arrive at very general approximation results for optimal control policies by quantized (finite action / piecewise constant) stationary control policies for a general class of controlled diffusions in the whole space \mathbb{R}^d as well as time-discretizations for the criteria with finite horizons.

Such a problem is of significant practical consequence, and accordingly has been studied extensively in a variety of setups. Due to its wide range of applications in domains that spans from mathematical finance, large deviations and robust control, vehicle and mobile robot control and several other fields, the stochastic optimal control problems for controlled diffusions have been studied extensively in literature see, e.g., [14, 40] (finite horizon cost) [13, 12] (discounted cost) [2, 3, 18, 17, 19, 5] (ergodic cost) and references therein. Typically, there are two main approaches to deal with these problems. The first one is the Bellman's Dynamic Programming Principal (DPP). The DPP approach allows one to characterize the value function of the optimal control problem as the unique solution of the associated Hamilton-Jacobi-Bellman (HJB) equation [14, 40, 5, 37, 38]. The second one is Pontryagin maximum principal (in the stochastic framework) [41].

For numerical methods as well as learning theoretic methods, it is imperative to arrive at rigorous approximation results.

In the continuous-time literature, most of the approximation results are build on time-discretization and mainly focused on finite horizon or discounted cost criteria see, e.g., [35, 31, 32, 27, 29, 7, 8], though the ergodic control and control up to an exit time criteria have also been studied [35, 34].

For finite horizon criteria, a commonly adopted approach of approximating controlled diffusions by a sequence of discrete time Markov chain via weak convergence methods was studied by Kushner and Kushner and Dupuis, see [35, 31, 32]. These works deal with numerical procedures to construct near optimal control policies for controlled diffusion models by approximating the space of (open-loop adapted) relaxed control policies with those that are piece-wise constant, and by considering the weak convergence of approximating probability measures on the path space to the measure on the continuous-time limit. It is shown in [35, 31, 32] that if the constructed controlled Markov chain satisfies a certain "consistency" condition at the discrete-time sampling instants, then the state process and the corresponding value function asymptotically approximates the continuous time state process and the associated value function. This approach has been referred to as the *weak convergence* approach.

In an alternative program, building on finite difference approximations for Bellman's equations utilizing their regularity properties, Krylov [27, 29] established the convergence rate of for such approximation techniques, where finite difference approximations are studied to arrive at stability results. In particular, some estimates for the error bound of the finite-difference approximation schemes in the problem of finding viscosity or probabilistic solutions to degenerate Bellman's equations are established. The proof technique is based on mean value theorems for stochastic integrals (as in [30]), obtained on the basis of elementary properties of the associated Bellman's equations. Also, for controlled non-degenerate diffusion processes, it is shown in [28] that using policies which are constant on intervals of length h^2 , one can approximate the value function with errors of order $h^{\frac{1}{3}}$. In [7, 8] Barles et. al. improved the error bounds obtained in [27, 29, 28].

Borkar [15, 14], for the finite-horizon cost case pursued an alternative approach to show continuity (when only stationary state feedback policies are considered for finite horizon problems) in his newly introduced topology; he studied the dependence of the strategic measures (on the path space) on the control policy, via regularity properties of generator functions. Additionally, Borkar [15] did not study the implications in approximations.

Instead of the approaches adopted in the aforementioned studies, in this paper, utilizing regularity results of the associated Poisson equations via PDE theory, we arrive at continuity results under a relatively weaker set of assumptions on the diffusion coefficients (with the exception of Krylov's method, which is tailored for finite horizon problems). Our approach allows one to arrive at a unification of approximation methods for finite horizon criterion, infinite discounted criterion, control up to an exit time, and ergodic cost criterion problems. Accordingly, our primary approach is to utilize the regularity properties of the partial differential equations directly, first via uniqueness of solutions, and then via regularity properties of the solutions to establish consistency of optimality equations satisfied by the limits of solutions (as policies converge). We will see that one can obtain rather concise, direct, and general results.

Additionally, our results can be used to present weaker conditions under which the weak convergence methods can be applicable or when discretized approximations can be shown to be near optimal: For example it will be a consequence of our analysis that for many of the criteria one can utilize piece-wise continuous or continuous control policies for near optimality, which implies [35, Assumption A2.3, p. 322] used for approximations under ergodic cost criteria (where invariant measures under sampled chains can be shown to converge to the invariant measure of a continuous-time limit as discretization gets finer). Furthermore, we do not impose uniform boundedness conditions on the drift term or (uniform) Lipschitz continuity conditions, a common assumption in [35, 31, 32, 27, 29].

As noted above, the study of the finite action/piecewise constant approximation problem plays important role in computing near optimal policies and learning algorithms for controlled diffusions in \mathbb{R}^d . As it is pointed out in [42, 25], piecewise constant policies are also useful in numerical methods for solving HJB equations. The computational advantage comes from the fact that over the intervals in which the policy is constant, we have to only solve linear PDEs. In the continuous time setup learning problems become much more involved due to the complex structure of the dynamics and the optimality equation. One common approach to overcome these difficulties is to construct simpler models by discretizing time, space and action spaces which approximates the original continuous time model. In a recent work [9], the authors studied an approximate Q -learning algorithm for controlled diffusion models by discretizing the time, space and action spaces. Under mild assumptions, they produced a learning algorithm which converges to some approximately optimal control policy for a discounted cost problem. They assumed that the discretization is uniform in time but the discretization in state and action can be non-uniform. Similar learning algorithm for controlled diffusions is proposed in [39], this result is based on the finite difference and finite element approximations (as in [35]). Thus, if one can establish that learning a control model with finitely many control actions is sufficient for approximate optimality, then it will be easier to produce efficient learning algorithms for the original model.

In the literature of discrete time Markov decision processes (MDPs), various techniques are available to address the approximation problems, e.g., approximate dynamic programming, approximate value or policy iteration, approximate linear programming, simulation based techniques, neuro-dynamic programming (or reinforcement learning), state aggregation, etc. (see [11, 44, 22, 45] and the references therein). For discrete

time controlled models near optimality of quantized policies has been studied extensively in the literature see, e.g., [26, 47, 46, 48, 45]. In [47, 46], authors studied the finite state, finite action approximation (respectively) of fully observed MDPs with Borel state and action spaces, for both discounted and average costs criteria. In the compact state space case explicit rate of convergence is also established in [47]. Later, these results have been extended to partially observed Markov decision process setup in [26, 48], also see the references therein. Recently, [6, Section 4] established the denseness of the performance of deterministic policies with finite action spaces, among the performance values attained by the set of all randomized stationary policies.

Contributions and main results

In this manuscript our main goal is to study the following approximation problem: for a general class of controlled diffusions in \mathbb{R}^d under what conditions one can approximate the optimal control policies for both finite/infinite horizon cost criteria by policies with finite actions/ piecewise constant/continuous policies? While the time discretization approximation results for finite horizon problems, studied extensively by Krylov [27, 29, 28] (for degenerate diffusions), we will discuss this (for the non-degenerate case) as an application of our results.

In order to address these questions, we first show that both finite horizon and infinite horizon (discounted/ergodic) costs are continuous as a function of control policies under Borkar topology [15]. We establish these results by exploiting the existence and uniqueness results of the associated Poisson equations (see, Theorem 6.1 (finite horizon), Theorem 3.1 (discounted), Theorem 3.2 (control up to an exit time), Theorem 3.5, 3.8 (ergodic)). The analysis of ergodic cost case is relatively more involved. One of the major issues in analyzing the ergodic cost criteria under the near-monotone hypothesis is the non-uniqueness/restricted uniqueness of the solution of the associated HJB/Poisson equation (see, [5, Example 3.8.3], [3]). In [5, Example 3.8.3], [3] it is shown that under near-monotone hypothesis the associated HJB/Poisson equation may admit uncountably many solutions. In this paper, we have shown that under the near-monotone hypothesis, the associated Poisson equation admits a unique solution in the space of compatible solution pairs (see, [3, Definition 1.1]). Continuity results obtained in the paper will be also useful in establishing the existence of optimal policies of the corresponding optimal control problems.

Next, utilizing Lusin's theorem and Tietze's extension theorem we show that under the Borkar topology, quantized (finite actions/ piecewise constant) stationary policies are dense in the space of stationary Markov policies (see, Section 4). Also, following the analogous proof technique, we establish the denseness of space of continuous stationary policies in the space of stationary policies (see Theorem 4.3).

Following and briefly modifying the proof technique of the denseness of stationary policies, including time also as a parameter we establish that piecewise constant Markov policies are dense in the space of Markov policies under the Borkar topology (see, Theorem 6.2).

Then, using our continuity and denseness results, we deduce that for both finite and infinite horizon cost criteria, the optimal control policies can be approximated by quantized (finite actions/ piecewise constant) policies with arbitrary precision (see, Theorem 6.3 (finite horizon), Theorem 5.2 (control up to an exit time), Theorem 5.3, 5.4 (infinite horizon)).

The remaining part of the paper is organized as follows. In Section 2 we provide the problem formulation. The continuity of the discounted cost/cost up to an exit time as a function of control policy are proved in Section 3.1. A similar continuity result for ergodic cost is presented in Section 3.2, where we establish these results under two

types of conditions; stability or near-monotonicity. Section 4 is devoted to establish the denseness of finite action/piecewise constant or continuous stationary policies under Borkar topology. Then using the denseness and continuity results we establish near optimality of finite models for cost up to an exit time and discounted/ ergodic cost criteria in Section 5. Finally, in Section 6, we analyze the denseness of piecewise constant Markov policies under the Borkar topology and then exploiting the denseness result we prove the near optimality of continuous or piecewise constant Markov policies for the finite horizon cost criterion.

Notation:

- For any set $A \subset \mathbb{R}^d$, by $\tau(A)$ we denote *first exit time* of the process $\{X_t\}$ from the set $A \subset \mathbb{R}^d$, defined by

$$\tau(A) := \inf \{t > 0 : X_t \notin A\}.$$

- \mathcal{B}_r denotes the open ball of radius r in \mathbb{R}^d , centered at the origin,
- $\tau_r, \check{\tau}_r$ denote the first exist time from $\mathcal{B}_r, \mathcal{B}_r^c$ respectively, i.e., $\tau_r := \tau(\mathcal{B}_r)$, and $\check{\tau}_r := \tau(\mathcal{B}_r^c)$.
- By $\text{Tr } S$ we denote the trace of a square matrix S .
- For any domain $\mathcal{D} \subset \mathbb{R}^d$, the space $C^k(\mathcal{D})$ ($C^\infty(\mathcal{D})$), $k \geq 0$, denotes the class of all real-valued functions on \mathcal{D} whose partial derivatives up to and including order k (of any order) exist and are continuous.
- $C_c^k(\mathcal{D})$ denotes the subset of $C^k(\mathcal{D})$, $0 \leq k \leq \infty$, consisting of functions that have compact support. This denotes the space of test functions.
- $C_b(\mathbb{R}^d)$ denotes the class of bounded continuous functions on \mathbb{R}^d .
- $C_0^k(\mathcal{D})$, denotes the subspace of $C^k(\mathcal{D})$, $0 \leq k < \infty$, consisting of functions that vanish in \mathcal{D}^c .
- $C^{k,r}(\mathcal{D})$, denotes the class of functions whose partial derivatives up to order k are Hölder continuous of order r .
- $L^p(\mathcal{D})$, $p \in [1, \infty)$, denotes the Banach space of (equivalence classes of) measurable functions f satisfying $\int_{\mathcal{D}} |f(x)|^p dx < \infty$.
- $\mathcal{W}^{k,p}(\mathcal{D})$, $k \geq 0, p \geq 1$ denotes the standard Sobolev space of functions on \mathcal{D} whose weak derivatives up to order k are in $L^p(\mathcal{D})$, equipped with its natural norm (see, [1]).
- If $\mathcal{X}(Q)$ is a space of real-valued functions on Q , $\mathcal{X}_{\text{loc}}(Q)$ consists of all functions f such that $f\varphi \in \mathcal{X}(Q)$ for every $\varphi \in C_c^\infty(Q)$. In a similar fashion, we define $\mathcal{W}_{\text{loc}}^{k,p}(\mathcal{D})$.
- For $\mu > 0$, let $e_\mu(x) = e^{-\mu\sqrt{1+|x|^2}}$, $x \in \mathbb{R}^d$. Then $f \in L^{p,\mu}((0,T) \times \mathbb{R}^d)$ if $f e_\mu \in L^p((0,T) \times \mathbb{R}^d)$. Similarly, $\mathcal{W}^{1,2,p,\mu}((0,T) \times \mathbb{R}^d) = \{f \in L^{p,\mu}((0,T) \times \mathbb{R}^d) \mid f, \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x_i}, \frac{\partial^2 f}{\partial x_i \partial x_j} \in L^{p,\mu}((0,T) \times \mathbb{R}^d)\}$ with natural norm (see [10])

$$\begin{aligned} \|f\|_{\mathcal{W}^{1,2,p,\mu}} &= \left\| \frac{\partial f}{\partial t} \right\|_{L^{p,\mu}((0,T) \times \mathbb{R}^d)} + \|f\|_{L^{p,\mu}((0,T) \times \mathbb{R}^d)} \\ &\quad + \sum_i \left\| \frac{\partial f}{\partial x_i} \right\|_{L^{p,\mu}((0,T) \times \mathbb{R}^d)} + \sum_{i,j} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_{L^{p,\mu}((0,T) \times \mathbb{R}^d)}. \end{aligned}$$

Also, we use the following convention $\|f\|_{\mathcal{W}^{1,2,p,\mu}} = \|f\|_{1,2,p,\mu}$.

2 The Borkar topology on control policies, cost criteria, and the problem statement

Let \mathbb{U} be a compact metric space and $V = \mathcal{P}(\mathbb{U})$ be the space of probability measures on \mathbb{U} with topology of weak convergence. Let

$$b : \mathbb{R}^d \times \mathbb{U} \rightarrow \mathbb{R}^d,$$

$$\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}, \sigma = [\sigma_{ij}(\cdot)]_{1 \leq i, j \leq d},$$

be given functions. We consider a stochastic optimal control problem whose state is evolving according to a controlled diffusion process given by the solution of the following stochastic differential equation (SDE)

$$dX_t = b(X_t, U_t)dt + \sigma(X_t)dW_t, \quad X_0 = x \in \mathbb{R}^d. \quad (2.1)$$

Where

- W is a d -dimensional standard Wiener process, defined on a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$.
- We extend the drift term $b : \mathbb{R}^d \times V \rightarrow \mathbb{R}^d$ as follows:

$$b(x, \nu) = \int_{\mathbb{U}} b(x, \zeta) \nu(d\zeta),$$

for $\nu \in V$.

- U is a V valued adapted process satisfying following non-anticipativity condition: for $s < t$, $W_t - W_s$ is independent of

$$\tilde{\mathfrak{F}}_s := \text{the completion of } \sigma(X_0, U_r, W_r : r \leq s) \text{ relative to } (\mathfrak{F}, \mathbb{P}).$$

The process U is called an *admissible* control, and the set of all admissible controls is denoted by \mathfrak{U} (see, [18]). By a Markov control we mean an admissible control of the form $U_t = v(t, X_t)$ for some Borel measurable function $v : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow V$. The space of all Markov controls is denoted by \mathfrak{U}_m . If the function v is independent of t , i.e., $U_t = v(X_t)$ then U or by an abuse of notation v itself is called a stationary Markov control. The set of all stationary Markov controls is denoted by \mathfrak{U}_{sm} .

To ensure existence and uniqueness of strong solutions of (2.1), we impose the following assumptions on the drift b and the diffusion matrix σ .

- (A1) *Local Lipschitz continuity:* The function $\sigma = [\sigma^{ij}] : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$, $b : \mathbb{R}^d \times \mathbb{U} \rightarrow \mathbb{R}^d$ are locally Lipschitz continuous in x (uniformly with respect to the other variables for b). In other words, for some constant $C_R > 0$ depending on $R > 0$, we have

$$|b(x, \zeta) - b(y, \zeta)|^2 + \|\sigma(x) - \sigma(y)\|^2 \leq C_R |x - y|^2$$

for all $x, y \in \mathcal{B}_R$ and $\zeta \in \mathbb{U}$, where $\|\sigma\| := \sqrt{\text{Tr}(\sigma\sigma^T)}$. Also, we are assuming that b is jointly continuous in (x, ζ) .

- (A2) *Affine growth condition:* b and σ satisfy a global growth condition of the form

$$\sup_{\zeta \in \mathbb{U}} \langle b(x, \zeta), x \rangle^+ + \|\sigma(x)\|^2 \leq C_0(1 + |x|^2) \quad \forall x \in \mathbb{R}^d,$$

for some constant $C_0 > 0$.

- (A3) *Nondegeneracy:* For each $R > 0$, it holds that

$$\sum_{i, j=1}^d a^{ij}(x) z_i z_j \geq C_R^{-1} |z|^2 \quad \forall x \in \mathcal{B}_R,$$

and for all $z = (z_1, \dots, z_d)^T \in \mathbb{R}^d$, where $a := \frac{1}{2} \sigma \sigma^T$.

2.1 The Borkar topology on control policies

We now introduce the Borkar topology on stationary or Markov controls [15]

- *Topology of Stationary Policies:* From [5, Section 2.4], we have that the set \mathfrak{U}_{sm} is metrizable with compact metric.

Definition 2.1 (Borkar topology of stationary Markov policies). *A sequence $v_n \rightarrow v$ in \mathfrak{U}_{sm} if and only if*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v_n(x)(d\zeta) dx = \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x)(d\zeta) dx \quad (2.2)$$

for all $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$ (for more details, see [5, Lemma 2.4.1], [15]).

- *Topology of Markov Policies:* In the proof of [15, Theorem 3.1, Lemma 3.1], replacing A_n by $\hat{A}_n = A_n \times [0, n]$ and following the arguments as in the proof of [15, Theorem 3.1, Lemma 3.1], we have the following topology on the space of Markov policies \mathfrak{U}_m .

Definition 2.2 (Borkar topology of Markov policies). *A sequence $v_n \rightarrow v$ in \mathfrak{U}_m if and only if*

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^\infty \int_{\mathbb{R}^d} f(t, x) \int_{\mathbb{U}} g(x, t, \zeta) v_n(t, x)(d\zeta) dx dt \\ = \int_0^\infty \int_{\mathbb{R}^d} f(t, x) \int_{\mathbb{U}} g(x, t, \zeta) v(x, t)(d\zeta) dx dt \end{aligned} \quad (2.3)$$

for all $f \in L^1(\mathbb{R}^d \times [0, \infty)) \cap L^2(\mathbb{R}^d \times [0, \infty))$ and $g \in C_b(\mathbb{R}^d \times [0, \infty) \times \mathbb{U})$.

It is well known that under the hypotheses (A1)–(A3), for any admissible control (2.1) has a unique weak solution [5, Theorem 2.2.11], and under any stationary Markov strategy (2.1) has a unique strong solution which is a strong Feller (therefore strong Markov) process [5, Theorem 2.2.12].

Remark 2.3. If we replace $\sigma(x)$ by $\sigma(x, \zeta)$, in the relaxed control framework if $\sigma(\cdot, v(\cdot))$ is Lipschitz continuous for $v \in \mathfrak{U}_{sm}$ then (2.1) admits unique strong solution. But in general stationary policies $v \in \mathfrak{U}_{sm}$ are just measurable functions, and existence of suitable strong solutions in our setting is more intricate (see, [5, Remarks 2.3.2], [16]). However, under stationary Markov policies one can prove the existence of weak solutions which may not be unique [49], [5, Remarks 2.3.2] (note though that uniqueness is established for $d = 1, 2$ in [49, pp. 192–194] under some conditions). The existence of a suitable strong solution under stationary Markov policies is essential to obtain stochastic representation of the solutions of the associated HJB equations (by applying the Itô-Krylov formula).

2.2 Cost criteria

Let $c: \mathbb{R}^d \times \mathbb{U} \rightarrow \mathbb{R}_+$ be the *running cost* function. We assume that c is bounded, jointly continuous in (x, ζ) and locally Lipschitz continuous in its first argument uniformly with respect to $\zeta \in \mathbb{U}$. We extend $c: \mathbb{R}^d \times \mathbb{V} \rightarrow \mathbb{R}_+$ as follows: for $v \in \mathbb{V}$

$$c(x, v) := \int_{\mathbb{U}} c(x, \zeta) v(d\zeta).$$

In this article, we consider the problem of minimizing finite horizon cost, α -discounted cost and ergodic cost, respectively:

2.2.1 Finite horizon cost

For $U \in \mathfrak{U}$, the associated *finite horizon cost* is given by

$$\mathcal{J}_T(x, U) = \mathbb{E}_x^U \left[\int_0^T c(X_s, U_s) ds + H(X_T) \right], \quad (2.4)$$

and the optimal value is defined as

$$\mathcal{J}_T^*(x) := \inf_{U \in \mathfrak{U}} \mathcal{J}_T(x, U). \quad (2.5)$$

Then a policy $U^* \in \mathfrak{U}$ is said to be optimal if we have

$$\mathcal{J}_T(x, U^*) = \mathcal{J}_T^*(x). \quad (2.6)$$

2.2.2 Discounted cost criterion

For $U \in \mathfrak{U}$, the associated α -*discounted cost* is given by

$$\mathcal{J}_\alpha^U(x, c) := \mathbb{E}_x^U \left[\int_0^\infty e^{-\alpha s} c(X_s, U_s) ds \right], \quad x \in \mathbb{R}^d, \quad (2.7)$$

where $\alpha > 0$ is the discount factor and $X(\cdot)$ is the solution of (2.1) corresponding to $U \in \mathfrak{U}$ and \mathbb{E}_x^U is the expectation with respect to the law of the process $X(\cdot)$ with initial condition x . The controller tries to minimize (2.7) over his/her admissible policies \mathfrak{U} . Thus, a policy $U^* \in \mathfrak{U}$ is said to be optimal if for all $x \in \mathbb{R}^d$

$$\mathcal{J}_\alpha^{U^*}(x, c) = \inf_{U \in \mathfrak{U}} \mathcal{J}_\alpha^U(x, c) \quad (= V_\alpha(x)), \quad (2.8)$$

where $V_\alpha(x)$ is called the optimal value.

2.2.3 Ergodic cost criterion

For $U \in \mathfrak{U}$, the associated *ergodic cost* is given by

$$\mathcal{E}_x(c, U) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^U \left[\int_0^T c(X_s, U_s) ds \right], \quad (2.9)$$

and the optimal value is defined as

$$\mathcal{E}^*(c) := \inf_{x \in \mathbb{R}^d} \inf_{U \in \mathfrak{U}} \mathcal{E}_x(c, U). \quad (2.10)$$

Then a policy $U^* \in \mathfrak{U}$ is said to be optimal if we have

$$\mathcal{E}_x(c, U^*) = \mathcal{E}^*(c). \quad (2.11)$$

2.2.4 Control up to an exit time

For each $U \in \mathfrak{U}$ the associated cost is given as

$$\hat{\mathcal{J}}_e^U(x) := \mathbb{E}_x^U \left[\int_0^{\tau(O)} e^{-\int_0^t \delta(X_s, U_s) ds} c(X_t, U_t) dt + e^{-\int_0^{\tau(O)} \delta(X_s, U_s) ds} h(X_{\tau(O)}) \right], \quad x \in \mathbb{R}^d,$$

where $O \subset \mathbb{R}^d$ is a smooth bounded domain, $\tau(O) := \inf\{t \geq 0 : X_t \notin O\}$, $\delta(\cdot, \cdot) : \bar{O} \times \mathbb{U} \rightarrow [0, \infty)$ is the discount function and $h : \bar{O} \rightarrow \mathbb{R}_+$ is the terminal cost function. The optimal value is defined as

$$\hat{\mathcal{J}}_e^*(x) = \inf_{U \in \mathfrak{U}} \hat{\mathcal{J}}_e^U(x).$$

We assume that $\delta \in C(\bar{O} \times \mathbb{U})$, $h \in \mathcal{W}^{2,p}(O)$.

2.3 Problems studied

The main purpose of this manuscript will be to address the following problems:

- **Continuity of finite and infinite horizon costs.** Suppose $\{v_n\}_{n \in \mathbb{N}}$ is a sequence of control policies which converge to another control policy v in some sense (in particular, under Borkar topology, see Subsection 2.1). Does this imply that
 - for finite horizon cost: $\mathcal{J}_T(x, v_n) \rightarrow \mathcal{J}_T(x, v)$?
 - for discounted cost: $\mathcal{J}_\alpha^{v_n}(x, c) \rightarrow \mathcal{J}_\alpha^v(x, c)$?
 - for ergodic cost: $\mathcal{E}_x(c, v_n) \rightarrow \mathcal{E}_x(c, v)$?
 - for cost up to an exit time: $\hat{\mathcal{J}}_e^{v_n}(x) \rightarrow \hat{\mathcal{J}}_e^v(x)$?
- **Near optimality of quantized policies.** For any given $\epsilon > 0$, whether it is possible to construct a quantized (finite action/ piecewise constant) policy v_ϵ such that
 - for finite horizon cost: $\mathcal{J}_T(x, v_\epsilon) \leq \mathcal{J}_T^*(x) + \epsilon$?
 - for discounted cost: $\mathcal{J}_\alpha^{v_\epsilon}(x, c) \leq V_\alpha(x) + \epsilon$?
 - for ergodic cost: $\mathcal{E}_x(c, v_\epsilon) \leq \mathcal{E}^*(c) + \epsilon$?
 - for cost up to an exit time: $\hat{\mathcal{J}}_e^{v_\epsilon}(x) \leq \hat{\mathcal{J}}_e^*(x) + \epsilon$?

In this manuscript, we have shown that under a mild set of assumptions the answers to the above mentioned questions are affirmative. For the finite horizon case, we also study the time-discretization approximations as a further implication of our analysis.

Let us introduce a parametric family of elliptic operators, which will be useful in our analysis. With $\zeta \in \mathbb{U}$ treated as a parameter, we define a family of operators \mathcal{L}_ζ mapping $C^2(\mathbb{R}^d)$ to $C(\mathbb{R}^d)$ by

$$\mathcal{L}_\zeta f(x) := \text{Tr}(a(x)\nabla^2 f(x)) + b(x, \zeta) \cdot \nabla f(x), \quad (2.12)$$

where $f \in C^2(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ and for $v \in \mathbb{V}$ we extend \mathcal{L}_ζ as follows:

$$\mathcal{L}_v f(x) := \int_{\mathbb{U}} \mathcal{L}_\zeta f(x) v(d\zeta). \quad (2.13)$$

Also, for each $v \in \mathfrak{U}_{\text{sm}}$, we define

$$\mathcal{L}_v f(x) := \text{Tr}(a\nabla^2 f(x)) + b(x, v(x)) \cdot \nabla f(x). \quad (2.14)$$

3 Continuity of expect cost under various criteria in control policies under the Borkar topology

3.1 Continuity for discounted cost/cost up to an exit time

Since the proof techniques are almost similar, in this section, we analyze the continuity of both discounted cost as well as the cost up to an exit time with respect to the policies in the space of stationary policies under the Borkar topology (see Definition 2.1), i.e., we show that the maps $v \rightarrow \mathcal{J}_\alpha^v$ and $v \rightarrow \hat{\mathcal{J}}_e^v$ are continuous on \mathfrak{U}_{sm} .

3.1.1 Continuity of discounted cost

Now we prove the continuity of the discounted cost as a function of the control policies.

Theorem 3.1. *Suppose Assumptions (A1)-(A3) hold. Then the map $v \mapsto \mathcal{J}_\alpha^v(x, c)$ from \mathfrak{U}_{sm} to \mathbb{R} is continuous.*

Proof. Let $\{v_n\}_n$ be a sequence in \mathfrak{U}_{sm} such that $v_n \rightarrow v$ in \mathfrak{U}_{sm} . It is known that $\mathcal{J}_\alpha^{v_n}(x, c)$ is a solution to the Poisson's equation (see, [5, Lemma A.3.7])

$$\mathcal{L}_{v_n} \mathcal{J}_\alpha^{v_n}(x, c) - \alpha \mathcal{J}_\alpha^{v_n}(x, c) = -c(x, v_n(x)). \tag{3.1}$$

Now by standard elliptic p.d.e. estimates as in [24, Theorem 9.11], for any $p \geq d + 1$ and $R > 0$, we deduce that

$$\|\mathcal{J}_\alpha^{v_n}(x, c)\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \kappa_1 (\|\mathcal{J}_\alpha^{v_n}(x, c)\|_{L^p(\mathcal{B}_{2R})} + \|c(x, v_n(x))\|_{L^p(\mathcal{B}_{2R})}), \tag{3.2}$$

for some positive constant κ_1 which is independent of n . Since

$$\|c\|_\infty := \sup_{(x,u) \in \mathbb{R}^d \times \mathbb{U}} c(x, u) \leq M < \infty, \quad \text{and} \quad \mathcal{J}_\alpha^{v_n}(x, c) \leq \frac{\|c\|_\infty}{\alpha},$$

from (3.1) we obtain

$$\|\mathcal{J}_\alpha^{v_n}(x, c)\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \kappa_1 M \left(\frac{|\mathcal{B}_{2R}|^{\frac{1}{p}}}{\alpha} + |\mathcal{B}_{2R}|^{\frac{1}{p}} \right). \tag{3.3}$$

We know that for $1 < p < \infty$, the space $\mathcal{W}^{2,p}(\mathcal{B}_R)$ is reflexive and separable, hence, as a corollary of Banach Alaoglu theorem, we have that every bounded sequence in $\mathcal{W}^{2,p}(\mathcal{B}_R)$ has a weakly convergent subsequence (see, [20, Theorem 3.18.]). Also, we know that for $p \geq d + 1$ the space $\mathcal{W}^{2,p}(\mathcal{B}_R)$ is compactly embedded in $C^{1,\beta}(\bar{\mathcal{B}}_R)$, where $\beta < 1 - \frac{d}{p}$ (see [5, Theorem A.2.15 (2b)]), which implies that every weakly convergent sequence in $\mathcal{W}^{2,p}(\mathcal{B}_R)$ will converge strongly in $C^{1,\beta}(\bar{\mathcal{B}}_R)$. Thus, in view of estimate (3.3), by a standard diagonalization argument and Banach Alaoglu theorem, we can extract a subsequence $\{V_\alpha^{n_k}\}$ such that for some $V_\alpha^* \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d)$

$$\begin{cases} \mathcal{J}_\alpha^{v_{n_k}}(x, c) \rightarrow V_\alpha^* & \text{in } \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \text{ (weakly)} \\ \mathcal{J}_\alpha^{v_{n_k}}(x, c) \rightarrow V_\alpha^* & \text{in } C_{\text{loc}}^{1,\beta}(\mathbb{R}^d) \text{ (strongly)}. \end{cases} \tag{3.4}$$

In the following we will show that $V_\alpha^* = \mathcal{J}_\alpha^v(x, c)$. Note that

$$\begin{aligned} b(x, v_{n_k}(x)) \cdot \nabla \mathcal{J}_\alpha^{v_{n_k}}(x, c) - b(x, v(x)) \cdot \nabla V_\alpha^*(x) &= b(x, v_{n_k}(x)) \cdot \nabla (\mathcal{J}_\alpha^{v_{n_k}}(x, c) - V_\alpha^*)(x) \\ &\quad + (b(x, v_{n_k}(x)) - b(x, v(x))) \cdot \nabla V_\alpha^*(x). \end{aligned}$$

Since $\mathcal{J}_\alpha^{v_{n_k}}(x, c) \rightarrow V_\alpha^*$ in $C_{\text{loc}}^{1,\beta}(\mathbb{R}^d)$ and b is locally bounded, on any compact set $b(x, v_{n_k}(x)) \cdot \nabla (\mathcal{J}_\alpha^{v_{n_k}}(x, c) - V_\alpha^*)(x) \rightarrow 0$ strongly. Also, since $\nabla V_\alpha^* \in C_{\text{loc}}^{0,\beta}(\mathbb{R}^d)$, in view of the topology of \mathfrak{U}_{sm} , for any $\phi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} b(x, v_{n_k}(x)) \cdot \nabla V_\alpha^*(x) \phi(x) dx = \int_{\mathbb{R}^d} b(x, v(x)) \cdot \nabla V_\alpha^*(x) \phi(x) dx.$$

Hence, as $k \rightarrow \infty$, we obtain

$$b(x, v_{n_k}(x)) \cdot \nabla \mathcal{J}_\alpha^{v_{n_k}}(x, c) + c(x, v_{n_k}(x)) \rightarrow b(x, v(x)) \cdot \nabla V_\alpha^*(x) + c(x, v(x)) \quad \text{weakly.} \tag{3.5}$$

Now, multiplying by a test function $\phi \in C_c^\infty(\mathbb{R}^d)$, from (3.1), it follows that

$$\begin{aligned} \int_{\mathbb{R}^d} \text{Tr}(a(x) \nabla^2 \mathcal{J}_\alpha^{v_{n_k}}(x, c)) \phi(x) dx + \int_{\mathbb{R}^d} \{b(x, v_{n_k}(x)) \cdot \nabla \mathcal{J}_\alpha^{v_{n_k}}(x, c) + c(x, v_{n_k}(x))\} \phi(x) dx \\ = \alpha \int_{\mathbb{R}^d} \mathcal{J}_\alpha^{v_{n_k}}(x, c) \phi(x) dx. \end{aligned}$$

Hence, using (3.4), (3.5), and letting $k \rightarrow \infty$ (in the sense of distributions), we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} \text{Tr}(a(x)\nabla^2 V_\alpha^*(x))\phi(x)dx + \int_{\mathbb{R}^d} \{b(x, v(x)) \cdot \nabla V_\alpha^*(x) + c(x, v(x))\}\phi(x)dx \\ = \alpha \int_{\mathbb{R}^d} V_\alpha^*(x)\phi(x)dx. \end{aligned} \tag{3.6}$$

Since $\phi \in C_c^\infty(\mathbb{R}^d)$ is arbitrary and $V_\alpha^* \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d)$ from (3.6), we deduce that the function $V_\alpha^* \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap C_b(\mathbb{R}^d)$ satisfies

$$\text{Tr}(a(x)\nabla^2 V_\alpha^*(x)) + b(x, v(x)) \cdot \nabla V_\alpha^*(x) + c(x, v(x)) = \alpha V_\alpha^*(x). \tag{3.7}$$

Let X be the solution of the SDE (2.1) corresponding to v . Now applying Itô-Krylov formula, we obtain the following

$$\begin{aligned} \mathbb{E}_x^v [e^{-\alpha T} V_\alpha^*(X_T)] - V_\alpha^*(x) \\ = \mathbb{E}_x^v \left[\int_0^T e^{-\alpha s} \{ \text{Tr}(a(X_s)\nabla^2 V_\alpha^*(X_s)) + b(X_s, v(X_s)) \cdot \nabla V_\alpha^*(X_s) - \alpha V_\alpha^*(X_s) \} ds \right]. \end{aligned}$$

Hence, by (3.7), we get

$$\mathbb{E}_x^v [e^{-\alpha T} V_\alpha^*(X_T)] - V_\alpha^*(x) = - \mathbb{E}_x^v \left[\int_0^T e^{-\alpha s} c(X_s, v(X_s)) ds \right]. \tag{3.8}$$

Since V_α^* is bounded and

$$\mathbb{E}_x^v [e^{-\alpha T} V_\alpha^*(X_T)] = e^{-\alpha T} \mathbb{E}_x^v [V_\alpha^*(X_T)],$$

letting $T \rightarrow \infty$, it follows that

$$\lim_{T \rightarrow \infty} \mathbb{E}_x^v [e^{-\alpha T} V_\alpha^*(X_T)] = 0.$$

Thus, letting $T \rightarrow \infty$ by monotone convergence theorem, from (3.8), we obtain

$$V_\alpha^*(x) = \mathbb{E}_x^v \left[\int_0^\infty e^{-\alpha s} c(X_s, v(X_s)) ds \right] = \mathcal{J}_\alpha^v(x, c). \tag{3.9}$$

This completes the proof. □

3.1.2 Continuity of cost up to an exit time

Following the proof technique of Theorem 3.1, now we show that the cost up to an exit time (defined in Subsection 2.2.4) is continuous as a function of the control policies.

Theorem 3.2. *Suppose Assumptions (A1)-(A3) hold. Then the map $v \mapsto \hat{\mathcal{J}}_e^v(x)$ from \mathfrak{U}_{sm} to \mathbb{R} is continuous.*

Proof. Let $\{v_n\}_n$ be a sequence in \mathfrak{U}_{sm} such that $v_n \rightarrow v$ in \mathfrak{U}_{sm} . From [24, Theorem 9.15], it follows that there exist a unique function $\psi_n(x) \in \mathcal{W}^{2,p}(O)$ satisfying the following Poisson's equation

$$\mathcal{L}_{v_n} \psi_n(x) - \delta(x, v_n(x))\psi_n(x) + c(x, v_n(x)) = 0 \quad \text{with} \quad \psi_n = h \quad \text{on} \quad \partial O. \tag{3.10}$$

Applying Itô-Krylov formula, one can show that $\psi_n(x) = \hat{\mathcal{J}}_e^{v_n}(x)$ (this stochastic representation also ensures the uniqueness of the solution of (3.10)).

Now following the argument as in Theorem 3.1, by standard elliptic p.d.e. estimates [24, Theorem 9.11], we deduce that there exists $\psi(x) \in \mathcal{W}^{2,p}(O)$ such that $\psi_n \rightarrow \psi$ weakly in $\mathcal{W}^{2,p}(O)$. Thus, closely following the proof of Theorem 3.1, letting $n \rightarrow \infty$, from (3.10) it follows that

$$\mathcal{L}_v \psi(x) - \delta(x, v(x))\psi(x) + c(x, v(x)) = 0 \quad \text{with} \quad \psi = h \quad \text{on} \quad \partial O. \quad (3.11)$$

Again, by Itô-Krylov formula, using (3.11) we deduce that $\psi(x) = \hat{\mathcal{J}}_e^v(x)$. This completes the proof of the theorem. \square

3.2 Continuity for ergodic cost

In this section we study the continuity of the ergodic costs with respect to policies under Borkar topology in the space of stationary Markov policies. We will study this problem under two sets of assumptions: the first is so called near-monotonicity assumption on the running cost function and other one is Lyapunov stability assumption on the system. Our proof strategies will be slightly different under these two setups: In the former we will build on regularity properties of invariant probability measures, in the latter we will build more directly on regularity properties of solutions to HJB equations.

3.2.1 Under a near-monotonicity assumption

We assume that the running cost function c is near-monotone with respect to $\mathcal{E}^*(c)$, i.e.,

(A4) It holds that

$$\liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta) > \mathcal{E}^*(c). \quad (3.12)$$

This condition penalizes the escape of probability mass to infinity. Since our running cost c is bounded it is easy to see that $\mathcal{E}^*(c) \leq \|c\|_\infty$. It is known that under (3.12), optimal control exists in the space of stable stationary Markov controls (see, [5, Theorem 3.4.5]).

Remark 3.3. It may seem difficult at first to verify (3.12) unless $\mathcal{E}^*(c)$ is known. However there are important cases where (3.12) always holds, e.g.,

- When $\inf_{\zeta \in \mathbb{U}} c(x, \zeta)$ is inf-compact (i.e., the sub-level sets $\{x : \inf_{\zeta \in \mathbb{U}} c(x, \zeta) \leq k\}$, $k \in \mathbb{R}$, are either empty or compact sets) and $\mathcal{E}^*(c)$ is finite, then (3.12) holds trivially.
- It is easy to see that if there exists $v \in \mathfrak{U}_{\text{sm}}$ such that

$$\liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta) > c(x, v(x)) \quad \text{for all} \quad x \in \mathbb{R}^d, \quad (3.13)$$

then (3.12) holds. For example, in dimension $d = 1$, let $\mathbb{U} = \{0, 1\}$ and $c(x, \zeta) = \frac{x^2}{1+x^2} + \zeta$ then for suitable choice of $v \in \mathfrak{U}_{\text{sm}}$, (3.13) holds.

First, we prove that for each stable stationary Markov policy $v \in \mathfrak{U}_{\text{sm}}$ the associated Poisson's equation admits a unique solution in a certain function space. This uniqueness result will be useful in establishing the continuity and near optimality of quantized policies. For the following supporting result, we closely follow [5].

Theorem 3.4. *Suppose that Assumptions (A1) - (A4) hold. Let $v \in \mathfrak{U}_{\text{sm}}$ be a stable control with unique invariant measure η_v , such that*

$$\liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta) > \inf_{x \in \mathbb{R}^d} \mathcal{E}_x(c, v). \quad (3.14)$$

Then, there exists a unique pair $(V^v, \rho_v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \times \mathbb{R}$, $1 < p < \infty$, with $V^v(0) = 0$, $\inf_{\mathbb{R}^d} V^v > -\infty$ and $\rho_v = \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, u)v(x)(du)\eta_v(dx)$, satisfying

$$\rho_v = [\mathcal{L}_v V^v(x) + c(x, v(x))] \tag{3.15}$$

Moreover, we have

- (i) $\rho_v = \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v)$.
- (ii) for all $x \in \mathbb{R}^d$

$$V^v(x) = \lim_{r \downarrow 0} \mathbb{E}_x^v \left[\int_0^{\tilde{\tau}_r} (c(X_t, v(X_t)) - \rho_v) dt \right]. \tag{3.16}$$

Proof. Since c is bounded, we have $(\rho^v :=) \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, u)v(x)(du)\eta_v(dx) \leq \|c\|_\infty$. In view of (3.14), by writing $\rho^v = \alpha \int \mathcal{J}_\alpha^v(x, c)\eta_v(dx)$ from [5, Lemma 3.6.1], we have

$$\inf_{\kappa(\rho^v)} \mathcal{J}_\alpha^v(x, c) = \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) \leq \frac{\rho^v}{\alpha}, \tag{3.17}$$

where $\kappa(\rho^v) := \{x \in \mathbb{R}^d \mid \min_{\zeta \in \mathbb{U}} c(x, \zeta) \leq \rho^v\}$ and $\mathcal{J}_\alpha^v(x, c)$ is the α -discounted cost defined as in (2.7). As earlier, we have that $\mathcal{J}_\alpha^v(x, c)$ is a solution to the Poisson's equation (see, [5, Lemma A.3.7])

$$\mathcal{L}_v \mathcal{J}_\alpha^v(x, c) - \alpha \mathcal{J}_\alpha^v(x, c) = -c(x, v(x)). \tag{3.18}$$

Since $x \rightarrow \min_{\zeta \in \mathbb{U}} c(x, \zeta)$ is continuous, we have $\kappa(\rho^v)$ is closed and (3.14) implies that it is bounded. Therefore $\kappa(\rho^v)$ is compact and hence for some $R_0 > 0$, we have $\kappa(\rho^v) \subset \mathcal{B}_{R_0}$. This gives us $\inf_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) = \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c)$.

Thus, following the arguments as in [5, Lemma 3.6.3], we deduce that for each $R > R_0$ there exist constants $\tilde{C}_2(R), \tilde{C}_2(R, p)$ depending only on d, R_0 such that

$$\left(\text{osc}_{\mathcal{B}_{2R}} \mathcal{J}_\alpha^v(x, c) := \sup_{\mathcal{B}_{2R}} \mathcal{J}_\alpha^v(x, c) - \inf_{\mathcal{B}_{2R}} \mathcal{J}_\alpha^v(x, c) \leq \tilde{C}_2(R) \left(1 + \alpha \inf_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) \right) \right), \tag{3.19}$$

$$\|\mathcal{J}_\alpha^v(\cdot, c) - \mathcal{J}_\alpha^v(0, c)\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \tilde{C}_2(R, p) \left(1 + \alpha \inf_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) \right). \tag{3.20}$$

Hence, by following the arguments as in [5, Lemma 3.6.6], we conclude that there exists $(V^v, \tilde{\rho}^v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \times \mathbb{R}$ such that along a subsequence (as $\alpha \rightarrow 0$), $\mathcal{J}_\alpha^v(\cdot, c) - \mathcal{J}_\alpha^v(0, c) \rightarrow V^v(\cdot)$ and $\alpha \mathcal{J}_\alpha^v(0, c) \rightarrow \tilde{\rho}^v$ and the pair $(V^v, \tilde{\rho}^v)$ satisfies

$$\mathcal{L}_v V^v(x) + c(x, v(x)) = \tilde{\rho}^v. \tag{3.21}$$

We will show that the subsequential limits are unique.

From (3.17), we get $\tilde{\rho}^v \leq \rho^v$. Now, in view of estimates (3.17) and (3.20), it is easy to see that

$$\|V^v\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \tilde{C}_2(R, p) (1 + M). \tag{3.22}$$

Also, for each $x \in \mathbb{R}^d$, we have

$$\begin{aligned} V^v(x) &= \lim_{\alpha \rightarrow 0} (\mathcal{J}_\alpha^v(x, c) - \mathcal{J}_\alpha^v(0, c)) \\ &\geq \liminf_{\alpha \rightarrow 0} \left(\mathcal{J}_\alpha^v(x, c) - \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) + \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) - \mathcal{J}_\alpha^v(0, c) \right) \\ &\geq -\limsup_{\alpha \rightarrow \infty} \left(\mathcal{J}_\alpha^v(0, c) - \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) \right) + \liminf_{\alpha \rightarrow \infty} \left(\mathcal{J}_\alpha^v(x, c) - \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) \right) \end{aligned}$$

Near optimality of quantized policies

$$\begin{aligned} &\geq -\limsup_{\alpha \rightarrow \infty} \left(\mathcal{J}_\alpha^v(0, c) - \inf_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) \right) + \liminf_{\alpha \rightarrow \infty} \left(\mathcal{J}_\alpha^v(x, c) - \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) \right) \\ &\geq -\limsup_{\alpha \rightarrow \infty} \left(\text{osc}_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) \right); \quad \left(\text{since } \mathcal{J}_\alpha^v(x, c) - \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c) \geq 0 \right), \end{aligned} \quad (3.23)$$

where in the third inequality we have used the fact that $\inf_{\mathcal{B}_{R_0}} \mathcal{J}_\alpha^v(x, c) = \inf_{\mathbb{R}^d} \mathcal{J}_\alpha^v(x, c)$. Thus, from (3.19), we deduce that

$$V^v \geq -\tilde{C}_2(R_0)(1 + M). \quad (3.24)$$

This shows that $\inf_{\mathbb{R}^d} V^v > -\infty$.

Now, applying Itô-Krylov formula and using (3.21) we obtain

$$\mathbb{E}_x^v [V^v(X_{T \wedge \tau_R})] - V^v(x) = \mathbb{E}_x^v \left[\int_0^{T \wedge \tau_R} (\tilde{\rho}_v - c(X_t, v(X_t))) dt \right].$$

This implies

$$\inf_{y \in \mathbb{R}^d} V^v(y) - V^v(x) \leq \mathbb{E}_x^v \left[\int_0^{T \wedge \tau_R} (\tilde{\rho}_v - c(X_t, v(X_t))) dt \right].$$

Since v is stable, letting $R \rightarrow \infty$, we get

$$\inf_{y \in \mathbb{R}^d} V^v(y) - V^v(x) \leq \mathbb{E}_x^v \left[\int_0^T (\tilde{\rho}_v - c(X_t, v(X_t))) dt \right].$$

Now dividing both sides of the above inequality by T and letting $T \rightarrow \infty$, it follows that

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^v \left[\int_0^T c(X_t, v(X_t)) dt \right] \leq \tilde{\rho}_v.$$

Thus, $\rho^v \leq \tilde{\rho}_v$. This indeed implies that $\rho^v = \tilde{\rho}_v$. The representation (3.16) of V^v follows by closely mimicking the argument of [5, Lemma 3.6.9]. Therefore, we have a solution pair (V^v, ρ_v) to (3.15) satisfying (i) and (ii).

Next we want to prove that the solution pair is unique. To this end, let $(\hat{V}^v, \hat{\rho}_v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \times \mathbb{R}$, $1 < p < \infty$, with $\hat{V}^v(0) = 0$, $\inf_{\mathbb{R}^d} \hat{V}^v > -\infty$ and $\hat{\rho}_v = \int_{\mathbb{R}^d} \int_{\mathcal{U}} c(x, u)v(x)(du)\eta_v(dx)$, satisfying

$$\hat{\rho}_v = \left[\mathcal{L}_v \hat{V}^v(x) + c(x, v(x)) \right] \quad (3.25)$$

Since \hat{V}^v is bounded from below, applying Itô-Krylov formula and using (3.25) we get

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^v \left[\int_0^T c(X_t, v(X_t)) dt \right] \leq \hat{\rho}_v \quad (3.26)$$

Hence, from (3.26), it follows that

$$\hat{\rho}_v = \int_{\mathbb{R}^d} \int_{\mathcal{U}} c(x, u)v(x)(du)\eta_v(dx) \leq \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^v \left[\int_0^T c(X_t, v(X_t)) dt \right] \leq \hat{\rho}_v \quad (3.27)$$

This implies that $\hat{\rho}^v = \rho_v$. Now, applying Itô-Krylov formula and using (3.25), we obtain

$$\hat{V}^v(x) = \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r \wedge \tau_R} (c(X_t, v(X_t)) - \hat{\rho}_v) dt + \hat{V}^v(X_{\check{\tau}_r \wedge \tau_R}) \right]. \quad (3.28)$$

Since v is stable and \hat{V}^v is bounded from below, for all $x \in \mathbb{R}^d$ we obtain

$$\liminf_{R \rightarrow \infty} \mathbb{E}_x^v \left[\hat{V}^v (X_{\tau_R}) \mathbb{1}_{\{\check{\tau}_r \geq \tau_R\}} \right] \geq \liminf_{R \rightarrow \infty} \left(\inf_{\mathbb{R}^d} \hat{V}^v \right) \mathbb{P}_x (\check{\tau}_r \geq \tau_R) = 0.$$

In the above we have used the fact that $\tau_R \rightarrow \infty$ as $R \rightarrow \infty$ and $\mathbb{E}_x^v [\check{\tau}_r] < \infty$.

Again, since v is stable we have $\mathbb{E}_x^v [\check{\tau}_r] < \infty$ (see [5, Theorem 2.6.10]). Hence, letting $R \rightarrow \infty$ by Fatou's lemma from (3.28), it follows that

$$\begin{aligned} \hat{V}^v(x) &\geq \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \hat{\rho}_v) dt + \hat{V}^v (X_{\check{\tau}_r}) \right] \\ &\geq \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \hat{\rho}_v) dt \right] + \inf_{\mathbb{B}_r} \hat{V}^v. \end{aligned}$$

Since $\hat{V}^v(0) = 0$, letting $r \rightarrow 0$, we deduce that

$$\hat{V}^v(x) \geq \limsup_{r \downarrow 0} \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \hat{\rho}_v) dt \right]. \tag{3.29}$$

From (3.16) and (3.29), it is easy to see that $V^v - \hat{V}^v \leq 0$ in \mathbb{R}^d . On the other hand by (3.15) and (3.25) one has $\mathcal{L}_v (V^v - \hat{V}^v) (x) \geq 0$ in \mathbb{R}^d . Hence, applying the strong maximum principle [24, Theorem 9.6], one has $V^v = \hat{V}^v$. This proves uniqueness. \square

Now we prove the continuity of ergodic cost under near-monotonicity assumption on the running cost function.

Theorem 3.5. *Suppose that Assumptions (A1)-(A4) hold. Let $\{v_n\}_n$ be a sequence of stable policies such that $v_n \rightarrow v$ in \mathfrak{U}_{sm} and $\{\eta_{v_n}\}_n$ tight. If*

$$\sup_n \inf_{x \in \mathbb{R}^d} \mathcal{E}_x(c, v_n) < \liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta),$$

then we have the following

$$\inf_{\mathbb{R}^d} \mathcal{E}_x(c, v_n) \rightarrow \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v) \quad \text{as } n \rightarrow \infty. \tag{3.30}$$

Proof. From Theorem 3.4, we know that for each $n \in \mathbb{N}$ there exists $(V^{v_n}, \rho_{v_n}) \in \mathcal{W}_{loc}^{2,p}(\mathbb{R}^d) \times \mathbb{R}$, $1 < p < \infty$, with $V^{v_n}(0) = 0$ and $\inf_{\mathbb{R}^d} V^{v_n} > -\infty$, satisfying

$$\rho_{v_n} = \mathcal{L}_{v_n} V^{v_n}(x) + c(x, v_n(x)), \tag{3.31}$$

where $\rho_{v_n} = \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v_n(x) (d\zeta) \eta_{v_n}(dx) = \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v_n)$. Now from [4, Lemma 4.4], since we impose tightness a priori, we deduce that $\eta_{v_n} \rightarrow \eta_v$ in total variation topology. Hence the associated densities $\varphi_{v_n} \rightarrow \varphi$ in $L^1(\mathbb{R}^d)$ (see the proof of [5, Lemma 3.2.5]). Note that

$$\begin{aligned} &\int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v_n(x) (d\zeta) \eta_{v_n}(dx) - \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v(x) (d\zeta) \eta_v(dx) \\ &= \left(\int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v_n(x) (d\zeta) \varphi_{v_n}(x) dx - \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v_n(x) (d\zeta) \varphi(x) dx \right) \\ &\quad + \left(\int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v_n(x) (d\zeta) \varphi(x) dx - \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, \zeta) v(x) (d\zeta) \varphi(x) dx \right) \end{aligned} \tag{3.32}$$

Since c is bounded, the first term of the right hand side converges to zero since $\varphi_{v_n} \rightarrow \varphi$ in $L^1(\mathbb{R}^d)$ and the second term converges to zero by the convergence of $v_n \rightarrow v$ (see Definition 2.1). Hence, it follows that $\int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, u) v_n(x) (du) \eta_{v_n}(dx) \rightarrow \int_{\mathbb{R}^d} \int_{\mathbb{U}} c(x, u) v(x) (du) \eta_v(dx)$. Thus, in view of Theorem 3.4, we obtain $\inf_{\mathbb{R}^d} \mathcal{E}_x(c, v_n) \rightarrow \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v)$ as $n \rightarrow \infty$. This completes the proof. \square

Remark 3.6. The tightness assumption is not superfluous. In view of [3], we know that the map $v \mapsto \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v)$ in general may not be continuous on \mathfrak{U}_{sm} under near-monotone cost criterion (of the form, (3.30)). The reason is the following: for each $n \in \mathbb{N}$ let (V^{v_n}, ρ_{v_n}) be the unique compatible solution pair (see, [3, Definition 1.1]) of the equation (3.31), if (V^{v_n}, ρ_{v_n}) converge to a solution pair $(\bar{V}, \bar{\rho})$ of the limiting equation of (3.31) as $n \rightarrow \infty$, the solution pair $(\bar{V}, \bar{\rho})$ may not necessarily be compatible (see, [3]). One sufficient condition which ensure this continuity is the tightness of the space of corresponding invariant measures $\{\eta_{v_n} : n \in \mathbb{N}\}$.

3.2.2 Under Lyapunov stability

In this section we study the continuity of ergodic cost criterion under Lyapunov stability assumption. We assume the following Lyapunov stability condition on the dynamics.

(A5) There exists a positive constant \widehat{C}_0 , and a pair of inf-compact functions $(\mathcal{V}, h) \in C^2(\mathbb{R}^d) \times C(\mathbb{R}^d \times \mathbb{U})$ (i.e., the sub-level sets $\{\mathcal{V} \leq k\}, \{h \leq k\}$ are compact or empty sets in $\mathbb{R}^d, \mathbb{R}^d \times \mathbb{U}$ respectively for each $k \in \mathbb{R}$) such that

$$\mathcal{L}_\zeta \mathcal{V}(x) \leq \widehat{C}_0 - h(x, \zeta) \quad \text{for all } (x, \zeta) \in \mathbb{R}^d \times \mathbb{U}, \tag{3.33}$$

where $h (> 0)$ is locally Lipschitz continuous in its first argument uniformly with respect to the second and $\mathcal{V} > 1$.

A function $f \in \mathcal{O}(\mathcal{V})$ if $f \leq \widehat{C}_1 \mathcal{V}$ for some positive constant \widehat{C}_1 and $f \in \mathfrak{o}(\mathcal{V})$ if $\limsup_{|x| \rightarrow \infty} \frac{|f|}{\mathcal{V}} = 0$.

Now following [5, Lemma 3.7.8], we want to prove that a certain equation admits a unique solution in some suitable function space. This uniqueness result is crucial to obtain continuity of the map $v \rightarrow \mathcal{E}_x(c, v)$ on \mathfrak{U}_{sm} .

Theorem 3.7. Suppose that Assumptions (A1)-(A3) and (A5) hold. Then for each $v \in \mathfrak{U}_{\text{sm}}$ there exists a unique solution pair $(\widehat{V}^v, \widehat{\rho}^v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap \mathfrak{o}(\mathcal{V}) \times \mathbb{R}$ for any $p > 1$ satisfying

$$\widehat{\rho}^v = \mathcal{L}_v \widehat{V}^v(x) + c(x, v(x)) \quad \text{with } \widehat{V}^v(0) = 0. \tag{3.34}$$

Furthermore, we have

- (i) $\widehat{\rho}^v = \mathcal{E}_x(c, v)$
- (ii) for all $x \in \mathbb{R}^d$, we have

$$\widehat{V}^v(x) = \lim_{r \downarrow 0} \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \mathcal{E}_x(c, v)) dt \right]. \tag{3.35}$$

Proof. Existence of a solution pair $(\widehat{V}^v, \widehat{\rho}^v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap \mathfrak{o}(\mathcal{V}) \times \mathbb{R}$ for any $p > 1$ satisfying (i) and (ii) follows from [5, Lemma 3.7.8]. Also, it is known that along a subsequence $\alpha \mathcal{J}_\alpha^v(0, c) \rightarrow \widehat{\rho}^v$ and $\mathcal{J}_\alpha^v(x, c) - \mathcal{J}_\alpha^v(0, c) \rightarrow \widehat{V}^v$ uniformly over compact subsets of \mathbb{R}^d (see [5, Lemma 3.7.8 (i)]).

Next we show that the sub-sequential limits are unique. This indeed imply the uniqueness of the solutions. Let $(\bar{V}^v, \bar{\rho}^v) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap \mathfrak{o}(\mathcal{V}) \times \mathbb{R}$ for any $p > 1$ be any other solution pair of (3.34) with $\bar{V}^v(0) = 0$. Thus, by Itô-Krylov formula, for $R > 0$ we obtain

$$\begin{aligned} \mathbb{E}_x^v [\bar{V}^v(X_{T \wedge \tau_R})] - \bar{V}^v(x) &= \mathbb{E}_x^v \left[\int_0^{T \wedge \tau_R} \mathcal{L}_v \bar{V}^v(X_s) ds \right] \\ &= \mathbb{E}_x^v \left[\int_0^{T \wedge \tau_R} (\bar{\rho}^v - c(X_s, v(X_s))) ds \right]. \end{aligned} \tag{3.36}$$

Note that

$$\int_0^{T \wedge \tau_R} (\bar{\rho}^v - c(X_s, v(X_s))) ds = \int_0^{T \wedge \tau_R} \bar{\rho}^v - \int_0^{T \wedge \tau_R} c(X_s, v(X_s)) ds$$

Thus, letting $R \rightarrow \infty$ by monotone convergence theorem, we get

$$\lim_{R \rightarrow \infty} \mathbb{E}_x^v \left[\int_0^{T \wedge \tau_R} (\bar{\rho}^v - c(X_s, v(X_s))) ds \right] = \mathbb{E}_x^v \left[\int_0^T (\bar{\rho}^v - c(X_s, v(X_s))) ds \right].$$

Since $\bar{V}^v \in \mathfrak{o}(\mathcal{V})$, in view of [5, Lemma 3.7.2 (ii)], letting $R \rightarrow \infty$, we deduce that

$$\mathbb{E}_x^v [\bar{V}^v(X_T)] - \bar{V}^v(x) = \mathbb{E}_x^v \left[\int_0^T (\bar{\rho}^v - c(X_s, v(X_s))) ds \right]. \quad (3.37)$$

Also, from [5, Lemma 3.7.2 (ii)], we have

$$\lim_{T \rightarrow \infty} \frac{\mathbb{E}_x^v [\bar{V}^v(X_T)]}{T} = 0.$$

Hence, dividing both sides of (3.37) by T and letting $T \rightarrow \infty$, we obtain

$$\bar{\rho}^v = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^v \left[\int_0^T (c(X_s, v(X_s))) ds \right].$$

This implies that $\bar{\rho}^v = \hat{\rho}^v$. Again, applying Itô-Krylov formula and using (3.34), we have

$$\bar{V}^v(x) = \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r \wedge \tau_R} (c(X_t, v(X_t)) - \bar{\rho}^v) dt + \bar{V}^v(X_{\check{\tau}_r \wedge \tau_R}) \right]. \quad (3.38)$$

Also, from (3.33), by Itô-Krylov formula it follows that

$$\mathbb{E}_x^v [\mathcal{V}(X_{\check{\tau}_r \wedge \tau_R})] - \mathcal{V}(x) = \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r \wedge \tau_R} \mathcal{L}_v \mathcal{V}(X_t) dt \right] \leq \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r \wedge \tau_R} (\widehat{C}_0 - h(X_t, v(X_t))) dt \right].$$

This gives us the following (since $h(x, \zeta) > 0$)

$$\mathbb{E}_x^v [\mathcal{V}(X_{\tau_R}) \mathbb{1}_{\{\check{\tau}_r \geq \tau_R\}}] \leq \widehat{C}_0 \mathbb{E}_x^v [\check{\tau}_r] + \mathcal{V}(x) \quad \text{for all } r < |x| < R.$$

Now, it is easy to see that

$$-\sup_{\partial \mathcal{B}_R} \frac{|\widehat{V}^v|}{\mathcal{V}} \left(\widehat{C}_0 \mathbb{E}_x^v [\check{\tau}_r] + \mathcal{V}(x) \right) \leq \mathbb{E}_x^v \left[\widehat{V}^v(X_{\tau_R}) \mathbb{1}_{\{\check{\tau}_r \geq \tau_R\}} \right] \leq \sup_{\partial \mathcal{B}_R} \frac{|\widehat{V}^v|}{\mathcal{V}} \left(\widehat{C}_0 \mathbb{E}_x^v [\check{\tau}_r] + \mathcal{V}(x) \right).$$

Since $\bar{V}^v \in \mathfrak{o}(\mathcal{V})$, from the above estimate, we get

$$\liminf_{R \rightarrow \infty} \mathbb{E}_x^v \left[\widehat{V}^v(X_{\tau_R}) \mathbb{1}_{\{\check{\tau}_r \geq \tau_R\}} \right] = 0.$$

Thus, letting $R \rightarrow \infty$ by Fatou's lemma from (3.38), it follows that

$$\begin{aligned} \bar{V}^v(x) &\geq \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \bar{\rho}^v) dt + \bar{V}^v(X_{\check{\tau}_r}) \right] \\ &\geq \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \bar{\rho}^v) dt \right] + \inf_{\mathcal{B}_r} \bar{V}^v. \end{aligned}$$

Since $\bar{V}^v(0) = 0$, letting $r \rightarrow 0$, we deduce that

$$\bar{V}^v(x) \geq \limsup_{r \downarrow 0} \mathbb{E}_x^v \left[\int_0^{\check{\tau}_r} (c(X_t, v(X_t)) - \bar{\rho}^v) dt \right]. \tag{3.39}$$

Since $\hat{\rho}^v = \bar{\rho}^v$, from (3.35) and (3.39), it follows that $\hat{V}^v - \bar{V}^v \leq 0$ in \mathbb{R}^d . Also, since $(\hat{V}^v, \hat{\rho}^v)$ and $(\bar{V}^v, \bar{\rho}^v)$ are two solution pairs of (3.34), we have $\mathcal{L}_v(\hat{V}^v - \bar{V}^v)(x) = 0$ in \mathbb{R}^d . Hence, by strong maximum principle [24, Theorem 9.6], one has $\hat{V}^v = \bar{V}^v$. This proves uniqueness \square

Next we prove that the map $v \rightarrow \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v)$ is continuous on \mathfrak{U}_{sm} under the Borkar topology.

Theorem 3.8. *Suppose that Assumptions (A1)-(A3) and (A5) hold. Let $\{v_n\}_n$ be a sequence of policies in \mathfrak{U}_{sm} such that $v_n \rightarrow v$ in \mathfrak{U}_{sm} . Then we have*

$$\inf_{\mathbb{R}^d} \mathcal{E}_x(c, v_n) \rightarrow \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v) \quad \text{as } n \rightarrow \infty. \tag{3.40}$$

Proof. From Theorem 3.7, we know that for each $n \in \mathbb{N}$ there exists unique solution pair $(\hat{V}^{v_n}, \hat{\rho}^{v_n}) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap \mathfrak{o}(\mathcal{V}) \times \mathbb{R}$ for any $p > 1$ satisfying

$$\hat{\rho}^{v_n} = \mathcal{L}_{v_n} \hat{V}^{v_n}(x) + c(x, v_n(x)) \quad \text{with } \hat{V}^{v_n}(0) = 0, \tag{3.41}$$

where

- (i) $\hat{\rho}^{v_n} = \mathcal{E}_x(c, v_n)$
- (ii) for all $x \in \mathbb{R}^d$, we have

$$\hat{V}^{v_n}(x) = \lim_{r \downarrow 0} \mathbb{E}_x^{v_n} \left[\int_0^{\check{\tau}_r} (c(X_t, v_n(X_t)) - \mathcal{E}_x(c, v_n)) dt \right].$$

In view of (3.33), it is easy to see that, each $v \in \mathfrak{U}_{\text{sm}}$ is stable and $\inf_{v \in \mathfrak{U}_{\text{sm}}} \eta_v(\mathcal{B}_R) > 0$ for any $R > 0$ (see, [5, Lemma 3.3.4] and [5, Lemma 3.2.4(b)]). Thus, from [5, Theorem 3.7.4], it follows that

$$\|\mathcal{J}_\alpha^{v_n}(\cdot, c) - \mathcal{J}_\alpha^{v_n}(0, c)\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \frac{\hat{C}_2(R, p)}{\eta_{v_n}(\mathcal{B}_{2R})} \left(\frac{\hat{\rho}^{v_n}}{\eta_{v_n}(\mathcal{B}_{2R})} + \sup_{\mathcal{B}_{4R} \times \mathcal{U}} c(x, \zeta) \right), \tag{3.42}$$

where the positive constant $\hat{C}_2(R, p)$ depends only on R and p . Since the running cost is bounded we have $\|c\|_\infty \leq M$ for some positive constant M . Thus, we have $\hat{\rho}^{v_n} \leq M$. Hence from (3.42), we deduce that

$$\|\mathcal{J}_\alpha^{v_n}(\cdot, c) - \mathcal{J}_\alpha^{v_n}(0, c)\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \frac{M \hat{C}_2(R, p)}{\inf_n \eta_{v_n}(\mathcal{B}_{2R})} \left(\frac{1}{\inf_n \eta_{v_n}(\mathcal{B}_{2R})} + 1 \right).$$

This implies that

$$\|\hat{V}^{v_n}\|_{\mathcal{W}^{2,p}(\mathcal{B}_R)} \leq \hat{C}_3(R, p), \tag{3.43}$$

where $\hat{C}_3(R, p)$ is a positive constant which depends only on R and p . Hence, by a standard diagonalization argument and Banach Alaoglu theorem (see, (3.4)), one can extract a subsequence $\{\hat{V}^{v_{n_k}}\}$ such that for some $\hat{V}^* \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d)$ we have

$$\begin{cases} \hat{V}^{v_{n_k}} \rightarrow \hat{V}^* & \text{in } \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \text{ (weakly)} \\ \hat{V}^{v_{n_k}} \rightarrow \hat{V}^* & \text{in } C_{\text{loc}}^{1,\beta}(\mathbb{R}^d) \text{ (strongly)}. \end{cases} \tag{3.44}$$

Also, since $\widehat{\rho}^{v_n} \leq M$, along a further subsequence $\widehat{\rho}^{v_{n_k}} \rightarrow \widehat{\rho}^*$ (without loss of generality denoting by the same sequence). Now, by similar argument as in Theorem 3.1, multiplying by test function on the both sides of (3.41) and letting $k \rightarrow \infty$, we deduce that $(\widehat{V}^*, \widehat{\rho}^*) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \times \mathbb{R}$ satisfies

$$\widehat{\rho}^* = \mathcal{L}_v \widehat{V}^*(x) + c(x, v(x)). \tag{3.45}$$

Since $\widehat{V}^{v_n}(0) = 0$ for each n , we get $\widehat{V}^*(0) = 0$

Next we want to show that $\widehat{V}^* \in \mathfrak{o}(\mathcal{V})$. Following the proof of [5, Lemma 3.7.8] (see, eq.(3.7.47) or eq.(3.7.50)), it is easy to see that

$$\widehat{V}^{v_n}(x) \leq \mathbb{E}_x^{v_n} \left[\int_0^{\tilde{\tau}_r} (c(X_t, v_n(X_t)) - \mathcal{E}_x(c, v_n)) dt + \widehat{V}^{v_n}(X_{\tilde{\tau}_r}) \right].$$

This gives us the following estimate

$$|\widehat{V}^{v_n}(x)| \leq M \sup_n \mathbb{E}_x^{v_n} \left[\int_0^{\tilde{\tau}_r} (c(X_t, v_n(X_t)) + 1) dt + \sup_{\mathcal{B}_r} |\widehat{V}^{v_n}| \right]. \tag{3.46}$$

We know that, for $d < p < \infty$, the space $\mathcal{W}^{2,p}(\mathcal{B}_R)$ is compactly embedded in $C^{1,\beta}(\overline{\mathcal{B}}_R)$, where $\beta < 1 - \frac{d}{p}$ (see [5, Theorem A.2.15 (2b)]). Thus, from (3.43), we obtain $\sup_n \sup_{\mathcal{B}_r} |\widehat{V}^{v_n}| < \widehat{M}$ for some positive constant \widehat{M} . Therefore, in view of [5, Lemma 3.7.2 (i)], from (3.46), we deduce that $\widehat{V}^* \in \mathfrak{o}(\mathcal{V})$. Since the pair $(\widehat{V}^*, \widehat{\rho}^*) \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d) \cap \mathfrak{o}(\mathcal{V}) \times \mathbb{R}$ satisfies (3.45), by uniqueness of solution of (3.45) (see, Theorem 3.7) it follows that $(\widehat{V}^*, \widehat{\rho}^*) \equiv (\widehat{V}^v, \widehat{\rho}^v)$. This completes the proof of the theorem. \square

4 Denseness of finite action/piecewise constant stationary policies

4.1 Denseness of policies with finite actions

Let $d_{\mathbb{U}}$ be the metric on the action space \mathbb{U} . Since \mathbb{U} is compact, we have \mathbb{U} is totally bounded. Thus, one can find a sequence of finite grids $\{\{\zeta_{n,i}\}_{i=1}^{k_n}\}_{n \geq 1}$ such that

$$\min_{i=1,2,\dots,k_n} d(\zeta, \zeta_{n,i}) < \frac{1}{n} \quad \text{for all } \zeta \in \mathbb{U}.$$

Let $\Lambda_n := \{\zeta_{n,1}, \zeta_{n,2}, \dots, \zeta_{n,k_n}\}$ and define a function $Q_n : \mathbb{U} \rightarrow \Lambda_n$ by

$$Q_n(\zeta) = \arg \min_{\zeta_{n,i} \in \Lambda_n} d(\zeta, \zeta_{n,i}),$$

where ties are broken so that Q_n is measurable. The function Q_n is often known as nearest neighborhood quantizer (see, [47]).

For each n the function Q_n induces a partition $\{\mathbb{U}_{n,i}\}_{i=1}^{k_n}$ of the action space \mathbb{U} given by

$$\mathbb{U}_{n,i} = \{\zeta \in \mathbb{U} : Q_n(\zeta) = \zeta_{n,i}\}.$$

By triangle inequality, it follows that $\text{diam}(\mathbb{U}_{n,i}) := \sup_{\zeta_1, \zeta_2 \in \mathbb{U}_{n,i}} d_{\mathbb{U}}(\zeta_1, \zeta_2) < \frac{2}{n}$. Now, for each $v \in \mathfrak{U}_{\text{sm}}$ define a sequence of policies with finite actions as follows:

$$v_n(\zeta_{n,i}|x) = v(Q_n^{-1}(\zeta_{n,i})|x) = v(\mathbb{U}_{n,i}|x). \tag{4.1}$$

In the next lemma we prove that the space of stationary policies with finite actions are dense in \mathfrak{U}_{sm} with respect to the *Borkar topology* (see, Definition 2.1).

Lemma 4.1. For each $v \in \mathfrak{U}_{sm}$ there exists a sequence of policies $\{v_n\}_n$ (defined as in (4.1)) with finite actions, satisfying

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v_n(x)(d\zeta) dx = \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x)(d\zeta) dx \quad (4.2)$$

for all $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$

Proof. Let $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$. Then from the construction of the sequence $\{v_n\}_n$, it is easy to see that

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, u) v_n(x)(du) dx - \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x)(d\zeta) dx \right| \\ & \leq \int_{\mathbb{R}^d} |f(x)| \sum_{i=1}^{k_n} \int_{\mathbb{U}_{n,i}} |g(x, \zeta_{n,i}) - g(x, \zeta)| v(x)(d\zeta) dx. \end{aligned}$$

Since $g \in C_b(\mathbb{R}^d \times \mathbb{U})$ and $\text{diam}(\mathbb{U}_{n,i}) < \frac{2}{n}$, it follows that

$$|f(x)| \sum_{i=1}^{k_n} \int_{\mathbb{U}_{n,i}} |g(x, \zeta_{n,i}) - g(x, \zeta)| v(x)(d\zeta) \rightarrow 0 \quad \text{for all } x \in \mathbb{R}^d.$$

As we know that g is bounded, for some positive constant M_1 we have $|g| \leq M_1$. Thus, we deduce that

$$|f(x)| \sum_{i=1}^{k_n} \int_{\mathbb{U}_{n,i}} |g(x, \zeta_{n,i}) - g(x, \zeta)| v(x)(d\zeta) \leq 2M_1 |f(x)| \quad \text{for all } x \in \mathbb{R}^d.$$

Since $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, by dominated convergence theorem, we obtain

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, u) v_n(x)(du) dx = \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x)(d\zeta) dx.$$

This completes the proof of the lemma. □

4.2 Denseness of piecewise constant policies

Let $d_{\mathcal{P}}$ be the Prokhorov metric on \mathcal{V} . Since $(\mathbb{U}, d_{\mathbb{U}})$ is separable (being a compact metric space) thus convergence in $(\mathcal{V}, d_{\mathcal{P}})$ is equivalent to weak convergence of probability measures.

Theorem 4.2. For each $v \in \mathfrak{U}_{sm}$ there exists a sequence of piecewise constant policies $\{v_m\}_m$ in \mathfrak{U}_{sm} such that

$$\lim_{m \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v_m(x)(d\zeta) dx = \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x)(d\zeta) dx \quad (4.3)$$

for all $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$

Proof. Let $\mathcal{B}_0 = \emptyset$ and define $D_n = \mathcal{B}_n \setminus \mathcal{B}_{n-1}$ for $n \in \mathbb{N}$. Thus it is easy to see that $\mathbb{R}^d = \cup_{n=1}^{\infty} D_n$. Since each $v \in \mathfrak{U}_{sm}$ is a measurable map $v : \mathbb{R}^d \rightarrow \mathcal{V}$, it follows that $\hat{v}_n := v|_{D_n} : D_n \rightarrow \mathcal{V}$ is a measurable map. Hence, by Lusin's theorem (see [21, Theorem 7.5.2]), for any $\epsilon_n > 0$ there exists a compact set $K_n^{\epsilon_n} \subset D_n$ and a continuous function $\hat{v}_n^{\epsilon_n} : K_n^{\epsilon_n} \rightarrow \mathcal{V}$ such that (the Lebesgue measure of the set $D_n \setminus K_n^{\epsilon_n}$) $|(D_n \setminus K_n^{\epsilon_n})| < \epsilon_n$ and $\hat{v}_n \equiv \hat{v}_n^{\epsilon_n}$ on $K_n^{\epsilon_n}$. Again, Tietze's extension theorem (see [23, Theorem 4.1]) there exists a continuous function $\tilde{v}_n^{\epsilon_n} : D_n \rightarrow \mathcal{V}$ such that $\tilde{v}_n^{\epsilon_n} \equiv \hat{v}_n^{\epsilon_n}$ on $K_n^{\epsilon_n}$.

Step1 Therefore for any $\hat{f} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $\hat{g} \in C(\mathbb{U})$, we have

$$\begin{aligned} & \left| \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_n(x)(d\zeta) dx - \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) dx \right| \\ & \leq \left| \int_{D_n \setminus K_n^{\epsilon_n}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_n(x)(d\zeta) dx - \int_{D_n \setminus K_n^{\epsilon_n}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) dx \right| \\ & \leq \left| \int_{D_n \setminus K_n^{\epsilon_n}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_n(x)(d\zeta) dx \right| + \left| \int_{D_n \setminus K_n^{\epsilon_n}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) dx \right| \\ & \leq \|\hat{g}\|_{\infty} \int_{D_n \setminus K_n^{\epsilon_n}} |\hat{f}(x)| dx + \|\hat{g}\|_{\infty} \int_{D_n \setminus K_n^{\epsilon_n}} |\hat{f}(x)| dx \\ & \leq 2\|\hat{g}\|_{\infty} \|\hat{f}\|_{L^2(\mathbb{R}^d)} \sqrt{|(D_n \setminus K_n^{\epsilon_n})|} \leq 2\sqrt{\epsilon_n} \|\hat{g}\|_{\infty} \|\hat{f}\|_{L^2(\mathbb{R}^d)} \quad (\text{by Hölder's inequality}). \end{aligned} \tag{4.4}$$

Now, since $(V, d_{\mathcal{P}})$ is compact, for each $m \in \mathbb{N}$ there exists a finite set $\hat{\Lambda}_m = \{\mu_{m,1}, \mu_{m,2}, \dots, \mu_{m,k_m}\}$ such that

$$\inf_{\mu_{m,i} \in \hat{\Lambda}_m} d_{\mathcal{P}}(\mu, \mu_{m,i}) < \frac{1}{m} \quad \text{for any } \mu \in V.$$

Let $\hat{Q}_m : V \rightarrow \hat{\Lambda}_m$ be defined as

$$\hat{Q}_m(\mu) = \arg \min_{\mu_{m,i} \in \hat{\Lambda}_m} d_{\mathcal{P}}(\mu, \mu_{m,i}).$$

Ties are broken so that \hat{Q}_m is a measurable map. Hence, it induces a partition $\{\hat{U}_{m,i}\}_{i=1}^{k_m}$ of the space V which is given by

$$\hat{U}_{m,i} = \{\mu \in V : \hat{Q}_m(\mu) = \mu_{m,i}\}.$$

By triangle inequality it is easy to see that

$$\text{diam}(\hat{U}_{m,i}) := \sup_{\mu_1, \mu_2} d_{\mathcal{P}}(\mu_1, \mu_2) < \frac{2}{m}.$$

Now, for $v \in \mathfrak{U}_{\text{sm}}$ define $D_{n,i}^m = (\tilde{v}_n^{\epsilon_n})^{-1}(\hat{U}_{m,i})$. This implies that $D_n = \cup_{i=1}^{k_m} D_{n,i}^m$. Define

$$\hat{v}_{n,m}^{\epsilon_n}(x) := \sum_{i=1}^{k_m} \mu_{m,i} \mathbb{1}_{\{D_{n,i}^m\}}(x) \quad \text{for all } x \in D_n \text{ and } m \in \mathbb{N}.$$

Therefore, we deduce that

$$\begin{aligned} & \left| \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) dx - \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_{n,m}^{\epsilon_n}(x)(d\zeta) dx \right| \\ & \leq \sum_{i=1}^{k_m} \left| \int_{D_{n,i}^m} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) dx - \int_{D_{n,i}^m} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \mu_{m,i}(d\zeta) dx \right| \\ & \leq \sum_{i=1}^{k_m} \int_{D_{n,i}^m} |\hat{f}(x)| \left| \int_{\mathbb{U}} \hat{g}(\zeta) \tilde{v}_n^{\epsilon_n}(x)(d\zeta) - \int_{\mathbb{U}} \hat{g}(\zeta) \mu_{m,i}(d\zeta) \right| dx \\ & \leq \|\hat{f}\|_{L^1(\mathbb{R}^d)} \epsilon_n \quad (\text{for large enough } m). \end{aligned} \tag{4.5}$$

If we choose $\epsilon_n = \min\left\{\frac{\epsilon_n \|\hat{f}\|_{L^2(\mathbb{R}^d)} \|\hat{g}\|_{\infty}}{4}, \frac{\epsilon_n \|\hat{f}\|_{L^1(\mathbb{R}^d)}}{2}\right\}$, combining (4.4), (4.5), there exists $\bar{M}_0 > 0$ (depending on \hat{f}, \hat{g} and ϵ_n) such that

$$\left| \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_n(x)(d\zeta) dx - \int_{D_n} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_{n,m}^{\epsilon_n}(x)(d\zeta) dx \right| \leq \epsilon_n, \tag{4.6}$$

for all $m \geq \bar{M}_0$.

Step2 Let $\epsilon > 0$ be a small number. Now define

$$\bar{v}_m^\epsilon := \sum_{n=1}^{\infty} \hat{v}_{n,m}^{\epsilon_n} \quad \text{for } m \in \mathbb{N}. \quad (4.7)$$

Since $\hat{f} \in L^1(\mathbb{R}^d)$ there exists $N_0 \in \mathbb{N}$ such that $\int_{\mathbb{B}_{N_0}^c} |\hat{f}(x)| dx < \frac{\epsilon}{4\|\hat{g}\|_\infty}$

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) v(x) (d\zeta) dx - \int_{\mathbb{R}^d} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \bar{v}_m^\epsilon(x) (d\zeta) dx \right| \\ & \leq \left| \int_{\mathbb{B}_{N_0}^c} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \\ & \quad + \left| \int_{\mathbb{B}_{N_0}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \\ & \leq \frac{\epsilon}{2} + \left| \int_{\mathbb{B}_{N_0}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \end{aligned}$$

Now, choose $\epsilon_i > 0$ for $i = 1, \dots, N_0$ such that $\sum_{i=1}^{N_0} \epsilon_i < \frac{\epsilon}{2}$. Thus, in view of (4.6) there exists $M_i > 0$ such that for each $i = 1, \dots, N_0$

$$\left| \int_{D_i} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_i(x) (d\zeta) dx - \int_{D_i} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \hat{v}_{i,m}^{\epsilon_i}(x) (d\zeta) dx \right| \leq \epsilon_i,$$

for all $m \geq M_i$. Hence, for $m \geq \max\{M_i, i = 1, \dots, N_0\}$, we get

$$\begin{aligned} & \left| \int_{\mathbb{B}_{N_0}} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \\ & \leq \sum_{i=1}^{N_0} \left| \int_{D_i} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) (\hat{v}_i(x) - \hat{v}_{i,m}^{\epsilon_i}(x)) (d\zeta) dx \right| \leq \sum_{i=1}^{N_0} \epsilon_i < \frac{\epsilon}{2}. \end{aligned} \quad (4.8)$$

Therefore, for each $\epsilon > 0$ we deduce that there exists a positive constant $\hat{M}_0 (= \max\{M_i, i = 1, \dots, N_0\})$ such that for $m \geq \hat{M}_0$ (where \hat{M}_0 depends on $\hat{f}, \hat{g}, \epsilon$)

$$\left| \int_{\mathbb{R}^d} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) v(x) (d\zeta) dx - \int_{\mathbb{R}^d} \hat{f}(x) \int_{\mathbb{U}} \hat{g}(\zeta) \bar{v}_m^\epsilon(x) (d\zeta) dx \right| \leq \epsilon. \quad (4.9)$$

Step3 Let $\{\hat{f}_k\}_{k \in \mathbb{N}}$ and $\{h_j\}_{j \in \mathbb{N}}$ be countable dense set in $L^1(\mathbb{R}^d)$ and $C(\mathbb{U})$ respectively. Thus (4.9) holds true for each \hat{f}_k and h_j .

Let $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$. Since $f \in L^1(\mathbb{R}^d)$ for $\epsilon > 0$ there exists $N_1 \in \mathbb{N}$ such that $\int_{\mathbb{B}_{N_1}^c} |f(x)| dx \leq \frac{\epsilon}{4\|g\|_\infty}$. This implies

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x) (d\zeta) dx - \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) \bar{v}_m^\epsilon(x) (d\zeta) dx \right| \\ & \leq \left| \int_{\mathbb{B}_{N_1}^c} f(x) \int_{\mathbb{U}} g(x, \zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \\ & \quad + \left| \int_{\mathbb{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right| \\ & \leq \frac{\epsilon}{2} + \left| \int_{\mathbb{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta) (v(x) - \bar{v}_m^\epsilon(x)) (d\zeta) dx \right|. \end{aligned} \quad (4.10)$$

Near optimality of quantized policies

It is well known that in $C_b(\bar{\mathcal{B}}_{N_1} \times \mathbb{U})$ the functions of the form $\{\sum_i^m r_i(x)p_i(\zeta)\}_{m \in \mathbb{N}}$ forms an algebra which contains constants, where $r_i \in C(\bar{\mathcal{B}}_{N_1})$ and $p_i \in C(\mathbb{U})$. Thus by Stone-Weierstrass theorem there exists \hat{m} (large enough) such that

$$\sup_{\mathcal{B}_{N_1} \times \mathbb{U}} |g(x, \zeta) - \sum_i^{\hat{m}} r_i(x)p_i(\zeta)| \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d)}}. \quad (4.11)$$

Since $p_i \in C(\mathbb{U})$ we can find $h_{j(i)} \in C(\mathbb{U})$ such that

$$\sup_{\zeta \in \mathbb{U}} |p_i(\zeta) - h_{j(i)}(\zeta)| \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d)}\|r_i\|_\infty}. \quad (4.12)$$

Also, since $fr_i \in L^1(\mathbb{R}^d)$ there exists $\hat{f}_{k(i)}$ such that

$$\int_{\mathcal{B}_{N_1}} |f(x)r_i(x) - \hat{f}_{k(i)}(x)|dx \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d)}\|h_i\|_\infty}. \quad (4.13)$$

Now, using (4.11), (4.12), (4.13) we have the following

$$\begin{aligned} & \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta)v(x) - \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \\ & \leq \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta)v(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} \sum_i^{\hat{m}} r_i(x)p_i(\zeta)v(x)(d\zeta)dx \right| \\ & \quad + \sum_{i=1}^{\hat{m}} \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)p_i(\zeta)v(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)h_{j(i)}(\zeta)v(x)(d\zeta)dx \right| \\ & \quad + \sum_{i=1}^{\hat{m}} \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)h_{j(i)}(\zeta)v(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)v(x)(d\zeta)dx \right| \\ & \quad + \sum_{i=1}^{\hat{m}} \left| \int_{\mathcal{B}_{N_1}} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)v(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \\ & \quad + \sum_{i=1}^{\hat{m}} \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)h_{j(i)}(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \\ & \quad + \sum_{i=1}^{\hat{m}} \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)p_i(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} r_i(x)h_{j(i)}(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \\ & \quad + \left| \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} g(x, \zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx - \int_{\mathcal{B}_{N_1}} f(x) \int_{\mathbb{U}} \sum_i^{\hat{m}} r_i(x)p_i(\zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \\ & \leq \frac{\epsilon}{4} + \sum_{l=1}^{N_1} \sum_{i=1}^{\hat{m}} \left| \int_{D_l} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)v(x)(d\zeta)dx - \int_{D_l} \hat{f}_{k(i)}(x) \int_{\mathbb{U}} h_{j(i)}(\zeta)\bar{v}_{l,m}^{\epsilon_l}(x)(d\zeta)dx \right| \end{aligned} \quad (4.14)$$

Now, choose $\epsilon_{l,i}$ for $l = 1, \dots, N_1$ and $i = 1, \dots, \hat{m}$ in such a way that $\sum_{l=1}^{N_1} \sum_{i=1}^{\hat{m}} \epsilon_{l,i} \leq \frac{\epsilon}{4}$. Thus, in view of (4.6) there exists $\hat{M}_2 := \max\{M_{k(i),j(i)}^l : i = 1, \dots, \hat{m}; l = 1, \dots, N_1\}$ (where $M_{k(i),j(i)}^l \in \mathbb{N}$ is the constant obtained as in (4.6) for $i = 1, \dots, \hat{m}; l = 1, \dots, N_1$). Therefore, from (4.10) and (4.14), we conclude that

$$\left| \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta)v(x)(d\zeta)dx - \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta)\bar{v}_m^\epsilon(x)(d\zeta)dx \right| \leq \epsilon, \quad (4.15)$$

for all $m \geq \hat{M}_2$. This completes the proof of the theorem. □

4.3 Denseness of continuous policies

Following the discussions above, one can show that the space of continuous stationary policies are also dense in the space of stationary policies under Borkar topology. This is a useful result as continuity allows for many approximation results to be invoked with little effort (see e.g. [35, Assumption A2.3, p. 322] where convergence properties of invariant measures corresponding to time-discretizations are facilitated).

Theorem 4.3. *For each $v \in \mathfrak{U}_{\text{sm}}$ there exists a sequence of continuous policies $\{v_m\}_m$ in \mathfrak{U}_{sm} such that*

$$\lim_{m \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v_m(x) (d\zeta) dx = \int_{\mathbb{R}^d} f(x) \int_{\mathbb{U}} g(x, \zeta) v(x) (d\zeta) dx \quad (4.16)$$

for all $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^d \times \mathbb{U})$

Proof. As earlier we have $\{f_i\}_{i \in \mathbb{N}}$ is a countable dense set in $L^1(\mathbb{R}^d)$. Now for each $i \in \mathbb{N}$, define a finite measure ν_i on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, given by

$$\nu_i(A) = \int_A |f_i(x)| dx \quad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

Let $v \in \mathfrak{U}_{\text{sm}}$. Then, as in the proof of Theorem 4.2, by successive application of Lusin's theorem (see [21, Theorem 7.5.2]) and Tietze's extension theorem (see [23, Theorem 4.1]), for any $\epsilon_i > 0$ there exists a closed set $K_i \in \mathbb{R}^d$ and a continuous function $v^i : \mathbb{R}^d \rightarrow \mathbb{V}$ such that $v^i \equiv v$ on K_i and $\nu_i(\mathbb{R}^d \setminus K_i) < \epsilon_i$. Hence, for any $g \in C_b(\mathbb{R}^d \times \mathbb{U})$, we have

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} f_i(x) \int_{\mathbb{U}} g(x, \zeta) v(x) (d\zeta) dx - \int_{\mathbb{R}^d} f_i(x) \int_{\mathbb{U}} g(x, \zeta) v^i(x) (d\zeta) dx \right| \\ & \leq \left| \int_{\mathbb{R}^d \setminus K_i} f_i(x) \int_{\mathbb{U}} g(x, \zeta) v(x) (d\zeta) dx - \int_{\mathbb{R}^d \setminus K_i} f_i(x) \int_{\mathbb{U}} g(x, \zeta) v^i(x) (d\zeta) dx \right| \\ & \leq 2 \|g\|_{\infty} \int_{\mathbb{R}^d \setminus K_i} |f_i(x)| dx \\ & = 2 \|g\|_{\infty} \nu_i(\mathbb{R}^d \setminus K_i) \leq 2 \|g\|_{\infty} \epsilon_i. \end{aligned}$$

Since $\{f_i\}_{i \in \mathbb{N}}$ is dense in $L^1(\mathbb{R}^d)$, by choosing ϵ_i appropriately, we obtain our result. \square

5 Near optimality of finite models for controlled diffusions

First we prove the near optimality of quantized policies for the α -discounted cost.

Theorem 5.1. *Suppose Assumptions (A1)-(A3) hold. Then for each $\epsilon > 0$ there exists a policy $v_{\epsilon}^* \in \mathfrak{U}_{\text{sm}}$ with finite actions and piecewise constant policies $\bar{v}_{\epsilon}^* \in \mathfrak{U}_{\text{sm}}$ such that*

$$\mathcal{J}_{\alpha}^{v_{\epsilon}^*}(x, c) \leq \inf_{U \in \mathfrak{U}} \mathcal{J}_{\alpha}^U(x, c) + \epsilon \quad \text{and} \quad \mathcal{J}_{\alpha}^{\bar{v}_{\epsilon}^*}(x, c) \leq \inf_{U \in \mathfrak{U}} \mathcal{J}_{\alpha}^U(x, c) + \epsilon \quad \text{for all } x \in \mathbb{R}^d. \quad (5.1)$$

Proof. From [5, Theorem 3.5.6], it follows that there exists $v^* \in \mathfrak{U}_{\text{sm}}$ such that $\mathcal{J}_{\alpha}^{v^*}(x, c) = \inf_{U \in \mathfrak{U}} \mathcal{J}_{\alpha}^U(x, c)$ for all $x \in \mathbb{R}^d$. Since the map $v \mapsto \mathcal{J}_{\alpha}^v(x, c)$ is continuous on \mathfrak{U}_{sm} (see, Theorem 3.1) and the space of quantized stationary policies are dense in \mathfrak{U}_{sm} (see, Lemma 4.1), it follows that for each $\epsilon > 0$ there exists a quantized policy $v_{\epsilon}^* \in \mathfrak{U}_{\text{sm}}$ satisfying (5.1). Similarly, since the piecewise constant policies are dense in \mathfrak{U}_{sm} (see, Theorem 4.2), we conclude that for any $\epsilon > 0$ there exists $\bar{v}_{\epsilon}^* \in \mathfrak{U}_{\text{sm}}$ which satisfies (5.1). This completes the proof. \square

We now show that for the cost up to an exit time, the quantized (finite action/piecewise constant) policies are near optimal.

Theorem 5.2. *Suppose Assumptions (A1)-(A3) hold. Then for each $\epsilon > 0$ there exists a policy $v_\epsilon^* \in \mathfrak{U}_{sm}$ with finite actions and piecewise constant policies $\bar{v}_\epsilon^* \in \mathfrak{U}_{sm}$ such that*

$$\hat{\mathcal{J}}_\epsilon^{v_\epsilon^*}(x) \leq \inf_{U \in \mathfrak{U}} \hat{\mathcal{J}}_e^U(x) + \epsilon \quad \text{and} \quad \hat{\mathcal{J}}_\epsilon^{\bar{v}_\epsilon^*}(x) \leq \inf_{U \in \mathfrak{U}} \hat{\mathcal{J}}_e^U(x) + \epsilon \quad \text{for all } x \in \mathbb{R}^d. \quad (5.2)$$

Proof. From [16, p. 229], we know that there exists $v^* \in \mathfrak{U}_{sm}$ such that $\hat{\mathcal{J}}_e^{v^*}(x) = \inf_{U \in \mathfrak{U}} \hat{\mathcal{J}}_e^U(x)$. Now from the continuity of the map $v \rightarrow \hat{\mathcal{J}}_e^v(x)$ (see Theorem 3.2) and the density results (see Section 4), it is easy to see that for any given $\epsilon > 0$ there exists policies $v_\epsilon^* \in \mathfrak{U}_{sm}$ with finite actions and piecewise constant policies $\bar{v}_\epsilon^* \in \mathfrak{U}_{sm}$ satisfying (5.2). This completes the proof of the theorem. \square

Next we prove the near optimality of the quantized policies for the ergodic cost under near-monotonicity assumption on the running cost. Let

$$\Theta_v := \{v_n \mid v_n \text{ is the quantized policy defined as in (4.1) corresponding to } v\}$$

and

$$\bar{\Theta}_v := \{\bar{v}_n \mid \bar{v}_n \text{ is the quantized policy defined as in (4.7) corresponding to } v\}.$$

In order to establish our result we are assuming that the invariant measures set

$$\Gamma_{v^*} := \{\eta_{v_n^*} \mid \eta_{v_n^*} \text{ is the invariant measure corresponding to } v_n^* \in \Theta_{v^*}\}$$

and

$$\bar{\Gamma}_{v^*} := \{\eta_{\bar{v}_n^*} \mid \eta_{\bar{v}_n^*} \text{ is the invariant measure corresponding to } \bar{v}_n^* \in \bar{\Theta}_{v^*}\}$$

are tight, where $v^* \in \mathfrak{U}_{sm}$ is an ergodic optimal control. The sufficient condition which assures the required tightness is the following: if there exists a non-negative inf-compact function $f \in C^2(\mathbb{R}^d)$ such that

$$\mathcal{L}_{v_n^*} f(x) \leq \kappa_0 - f(x) \quad \text{and} \quad \mathcal{L}_{\bar{v}_n^*} f(x) \leq \kappa_0 - f(x)$$

for some constant $\kappa_0 > 0$.

Theorem 5.3. *Suppose that Assumptions (A1) - (A4) hold. Also, suppose that corresponding to the optimal policy $v^* \in \mathfrak{U}_{sm}$, the following set of invariant measures Γ_{v^*} and $\bar{\Gamma}_{v^*}$ are tight and the running cost c is near monotone with respect to $\sup_{v_n^* \in \Theta_{v^*}} \mathcal{E}_x(c, v_n^*)$ and $\sup_{\bar{v}_n^* \in \bar{\Theta}_{v^*}} \mathcal{E}_x(c, \bar{v}_n^*)$, that is,*

$$\sup_{v_n^* \in \Theta_{v^*}} \inf_{x \in \mathbb{R}^d} \mathcal{E}_x(c, v_n^*) < \liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta) \quad \text{and} \quad \sup_{\bar{v}_n^* \in \bar{\Theta}_{v^*}} \inf_{x \in \mathbb{R}^d} \mathcal{E}_x(c, \bar{v}_n^*) < \liminf_{\|x\| \rightarrow \infty} \inf_{\zeta \in \mathbb{U}} c(x, \zeta).$$

Then for any given $\epsilon > 0$ there exists a policy $v_\epsilon \in \mathfrak{U}_{sm}$ with finite actions and a piecewise constant policy $\bar{v}_\epsilon \in \mathfrak{U}_{sm}$ such that

$$\mathcal{E}_x(c, v_\epsilon) \leq \mathcal{E}^*(c) + \epsilon \quad \text{and} \quad \mathcal{E}_x(c, \bar{v}_\epsilon) \leq \mathcal{E}^*(c) + \epsilon. \quad (5.3)$$

Proof. From [5, Theorem 3.6.10], we know there exists a stable $v^* \in \mathfrak{U}_{sm}$ such that $\mathcal{E}_x(c, v^*) = \mathcal{E}^*(c)$. Since, by our assumption, the set of invariant measures Γ_{v^*} and $\bar{\Gamma}_{v^*}$ are tight. Thus by the continuity result (see Theorem 3.5) and the density results (see Lemma 4.1, Theorem 4.2), we deduce that for each $\epsilon > 0$ there exists $v_\epsilon \in \mathfrak{U}_{sm}$ with finite actions and piecewise constant policy $\bar{v}_\epsilon \in \mathfrak{U}_{sm}$ such that (5.3) holds. This completes the proof. \square

Now for the ergodic cost criterion, under the Lyapunov type stability assumption we prove near optimality of quantized policies.

Theorem 5.4. *Suppose that assumptions (A1) - (A3) and (A5) hold. Then for any given $\epsilon > 0$ there exists a quantized policy $v_\epsilon \in \mathfrak{U}_{\text{sm}}$ with finite actions and a piecewise constant policy $\bar{v}_\epsilon^* \in \mathfrak{U}_{\text{sm}}$ such that*

$$\mathcal{E}_x(c, v_\epsilon) \leq \mathcal{E}^*(c) + \epsilon \quad \text{and} \quad \mathcal{E}_x(c, \bar{v}_\epsilon^*) \leq \mathcal{E}^*(c) + \epsilon. \quad (5.4)$$

Proof. From [5, Theorem 3.7.14], we know that there exists $v^* \in \mathfrak{U}_{\text{sm}}$ such that $\mathcal{E}_x(c, v^*) = \mathcal{E}^*(c)$. Now, since the space of quantized policies and piecewise constant policies are dense in \mathfrak{U}_{sm} (see, Lemma 4.1 and Theorem 4.2) and the map $v \rightarrow \inf_{\mathbb{R}^d} \mathcal{E}_x(c, v)$ is continuous on \mathfrak{U}_{sm} (see, Theorem 3.8). For any given $\epsilon > 0$, one can find a quantized policy $v_\epsilon \in \mathfrak{U}_{\text{sm}}$ with finite actions and a piecewise constant policy $\bar{v}_\epsilon^* \in \mathfrak{U}_{\text{sm}}$ such that (5.4) holds. \square

Remark 5.5. In view of the continuity (see Section 3.1, Section 3.2) and the denseness (see Theorem 4.3) results, we have the near optimality of continuous stationary policies.

6 Finite horizon cost: time discretization of Markov policies and near optimality of piecewise constant policies

Recall (2.4) as our cost criterion for the finite horizon setup. We will present three results in this section, where the ultimate goal is to arrive at near optimality of piecewise constant policies. While this approximation problem is a well-studied problem [35, 33, 43], our proof method is rather direct and appears to be new. Under uniform Lipschitz continuity and uniform boundedness assumptions on the diffusion coefficients and running cost function, in [35, 33, 43] the authors have established similar approximation results using numerical procedures.

Continuity of finite horizon cost on Markov policies under the Borkar topology

For simplicity, in this subsection we are assuming that a, b, c are uniformly bounded (it is possible to relax these boundedness assumptions). In particular we are assuming that

(B1) The functions a, b, c are uniformly bounded, i.e.,

$$\sup_{(x, \zeta) \in \mathbb{R}^d \times \mathbb{U}} \left[|b(x, \zeta)| + \|a(x)\| + \sum_i^d \left\| \frac{\partial a}{\partial x_i}(x) \right\| + |c(x, \zeta)| \right] \leq K.$$

for some positive constant K . Moreover, $H \in \mathcal{W}^{2,p,\mu}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, $p \geq 2$.

In view of [10, Theorem 3.3, p. 235], the optimality equation (or, the HJB equation)

$$\begin{aligned} \frac{\partial \psi}{\partial t} + \inf_{\zeta \in \mathbb{U}} [\mathcal{L}_\zeta \psi + c(x, \zeta)] &= 0 \\ \psi(T, x) &= H(x) \end{aligned}$$

admits a unique solution $\psi \in \mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d) \cap L^\infty((0, T) \times \mathbb{R}^d)$, $p \geq 2$. Thus, by Itô-Krylov formula (see the verification results as in [40, Theorem 3.5.2]), we know the existence of an optimal Markov policy, that is, there exists $v^* \in \mathfrak{U}_{\text{m}}$ such that $\mathcal{J}_T(x, v^*) = \mathcal{J}_T^*(x)$.

In the following theorem, we show that the finite horizon cost is continuous in \mathfrak{U}_{m} with respect to the Borkar topology (see Definition 2.2).

Theorem 6.1. *Suppose Assumptions (A1), (A3) and (B1) hold. Then the map $v \mapsto \mathcal{J}_T(x, v)$ from \mathfrak{U}_{m} to \mathbb{R} is continuous.*

Near optimality of quantized policies

Proof. Let v_n be a sequence in \mathfrak{U}_m such that $v_n \rightarrow v$ in \mathfrak{U}_m , for some $v \in \mathfrak{U}_m$. From [10, Theorem 3.3, p. 235], we have that for each $n \in \mathbb{N}$ there exists a unique solution $\psi_n \in \mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d) \cap L^\infty((0, T) \times \mathbb{R}^d)$, $p \geq 2$ to the following Poisson equation

$$\begin{aligned} \frac{\partial \psi_n}{\partial t} + [\mathcal{L}_{v_n} \psi_n + c(x, v_n(t, x))] &= 0 \\ \psi_n(T, x) &= H(x). \end{aligned} \tag{6.1}$$

By Itô-Krylov formula, we deduce that

$$\psi_n(t, x) = \mathbb{E}_x^{v_n} \left[\int_t^T c(X_s, v_n(s, X_s)) ds + H(X_T) \right] \tag{6.2}$$

This gives us

$$\|\psi_n\|_\infty \leq T \|c\|_\infty + \|H\|_\infty. \tag{6.3}$$

Rewriting (6.1), we get

$$\begin{aligned} \frac{\partial \psi_n}{\partial t} + \mathcal{L}_{v_n} \psi_n + \lambda_0 \psi_n &= \lambda_0 \psi_n - c(x, v_n(t, x)) \\ \psi_n(T, x) &= H(x), \end{aligned}$$

for some fixed $\lambda_0 > 0$. Thus, by parabolic pde estimate [10, eq. (3.8), p. 234], we deduce that

$$\|\psi_n\|_{\mathcal{W}^{1,2,p,\mu}} \leq \kappa_1 \|\lambda_0 \psi_n - c(x, v_n(t, x))\|_{L^{p,\mu}}. \tag{6.4}$$

Hence, from (6.3), (6.4), it follows that $\|\psi_n\|_{\mathcal{W}^{1,2,p,\mu}} \leq \kappa_2$ for some positive constant κ_2 (independent of n). Since $\mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d)$ is a reflexive Banach space, as a corollary of Banach Alaoglu theorem, there exists $\psi^* \in \mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d)$ such that along a subsequence (without loss of generality denoting by same sequence)

$$\begin{cases} \psi_n \rightarrow \psi^* & \text{in } \mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d) \quad (\text{weakly}) \\ \psi_n \rightarrow \psi^* & \text{in } \mathcal{W}^{0,1,p,\mu}((0, T) \times \mathbb{R}^d) \quad (\text{strongly}). \end{cases} \tag{6.5}$$

Since $v_n \rightarrow v$ in \mathfrak{U}_m , multiplying both sides of the (6.1) by test function $\phi \in C_c^\infty((0, T) \times \mathbb{R}^d)$ and integrating, we get

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \frac{\partial \psi_n}{\partial t} \phi(t, x) dt dx + \int_0^T \int_{\mathbb{R}^d} \text{Tr}(a(x) \nabla^2 \psi_n) \phi(t, x) dt dx \\ + \int_0^T \int_{\mathbb{R}^d} \{b(x, v_n(t, x)) \cdot \nabla \psi_n + c(x, v_n(t, x))\} \phi(t, x) dt dx = 0. \end{aligned} \tag{6.6}$$

In view of (6.5), letting $n \rightarrow \infty$, from (6.6) we obtain that

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \frac{\partial \psi^*}{\partial t} \phi(t, x) dt dx + \int_0^T \int_{\mathbb{R}^d} \text{Tr}(a(x) \nabla^2 \psi^*) \phi(t, x) dt dx \\ + \int_0^T \int_{\mathbb{R}^d} \{b(x, v(t, x)) \cdot \nabla \psi^* + c(x, v(t, x))\} \phi(t, x) dt dx = 0. \end{aligned}$$

This implies that $\psi^* \in \mathcal{W}^{1,2,p,\mu}((0, T) \times \mathbb{R}^d)$ satisfies

$$\begin{aligned} \frac{\partial \psi^*}{\partial t} + [\mathcal{L}_v \psi^* + c(x, v(t, x))] &= 0 \\ \psi(T, x) &= H(x). \end{aligned} \tag{6.7}$$

Again, by Itô-Krylov formula, it follows that

$$\psi^*(t, x) = \mathbb{E}_x^v \left[\int_t^T c(X_s, v(s, X_s)) ds + H(X_T) \right]. \tag{6.8}$$

Therefore, from (6.2) and (6.8), we conclude that $v \mapsto \mathcal{J}_T(x, v)$ from \mathfrak{U}_m to \mathbb{R} is continuous. \square

6.1 Time discretization of Markov policies

Following, and briefly modifying, our approach so far involving stationary policies, in this section we show that piece-wise constant Markov policies are dense in the space of Markov policies \mathfrak{U}_m . Also, using this result we deduce the near optimality of piece-wise constant Markov policies.

Theorem 6.2. *For any $v \in \mathfrak{U}_m$ there exists a sequence of piecewise constant policies $\{v_m\}_m$ such that*

$$\begin{aligned} \lim_{m \rightarrow \infty} \int_0^\infty \int_{\mathbb{R}^d} f(x, t) \int_{\mathbb{U}} g(x, t, \zeta) v_m(x, t) (d\zeta) dx dt \\ = \int_0^\infty \int_{\mathbb{R}^d} f(x, t) \int_{\mathbb{U}} g(x, t, \zeta) v(x, t) (d\zeta) dx dt \end{aligned} \tag{6.9}$$

for all $f \in L^1(\mathbb{R}^d \times [0, \infty)) \cap L^2(\mathbb{R}^d \times [0, \infty))$ and $g \in C_b(\mathbb{R}^d \times [0, \infty) \times \mathbb{U})$.

Proof. Let $\hat{\mathcal{B}}_0 = \emptyset$ and $\hat{\mathcal{B}}_n = \mathcal{B}_n \times [0, n)$. Then, define $\hat{D}_n = \hat{\mathcal{B}}_n \setminus \hat{\mathcal{B}}_{n-1}$ for $n \in \mathbb{N}$. Now, it is clear that $\mathbb{R}^d \times [0, \infty) = \cup_{n=1}^\infty \hat{D}_n$. Since $\bar{v}_n := v|_{\hat{D}_n} : \hat{D}_n \rightarrow V$ is a measurable map. As in Theorem 4.2, by Lusin's theorem and Tietze's extension theorem, for any $\epsilon_n > 0$ there exists a compact set $\hat{K}_n^{\epsilon_n} \subset \hat{D}_n$ and a continuous function $\bar{v}_n^{\epsilon_n} : \hat{D}_n \rightarrow V$ such that $\bar{v}_n^{\epsilon_n} \equiv \bar{v}_n$ on $\hat{K}_n^{\epsilon_n}$ and $|\hat{D}_n \setminus \hat{K}_n^{\epsilon_n}| < \epsilon_n$.

Also, as in Theorem 4.2, since $(V, d_{\mathcal{P}})$ is compact, for each $m \in \mathbb{N}$ there exists a finite set $\hat{\Lambda}_m = \{\mu_{m,1}, \mu_{m,2}, \dots, \mu_{m,k_m}\}$ and a quantizer $\hat{Q}_m : V \rightarrow \hat{\Lambda}_m$ which induces a partition $\{\hat{U}_{m,i}\}_{i=1}^{k_m}$ of the space V .

Now, for any $v \in \mathfrak{U}_m$ define $\hat{D}_{n,i}^m = (\bar{v}_n^{\epsilon_n})^{-1}(\hat{U}_{m,i})$. It is easy to see that $\hat{D}_n = \cup_{i=1}^{k_m} \hat{D}_{n,i}^m$. Define

$$\bar{v}_{n,m}^{\epsilon_n}(x, t) := \sum_{i=1}^{k_m} \mu_{m,i} \mathbb{1}_{\{\hat{D}_{n,i}^m\}}(x) \quad \text{for all } (x, t) \in \hat{D}_n \text{ and } m \in \mathbb{N}.$$

Hence, as in the proof of Theorem 4.2 (see Step 1), for any $\hat{f} \in L^1(\mathbb{R}^d \times [0, \infty)) \cap L^2(\mathbb{R}^d \times [0, \infty))$, $\hat{g} \in C_b(\mathbb{U})$, there exists a positive constant \bar{M}_0 (depending on \hat{f}, \hat{g} and ϵ_n) such that

$$\left| \int_{\hat{D}_n} \hat{f}(x, t) \int_{\mathbb{U}} \hat{g}(\zeta) \bar{v}_n(x, t) (d\zeta) dx dt - \int_{\hat{D}_n} \hat{f}(x, t) \int_{\mathbb{U}} \hat{g}(\zeta) \bar{v}_{n,m}^{\epsilon_n}(x, t) (d\zeta) dx dt \right| \leq \epsilon_n, \tag{6.10}$$

for all $m \geq \bar{M}_0$.

Now, for any given $\epsilon > 0$, define

$$\tilde{v}_m^\epsilon := \sum_{n=1}^\infty \bar{v}_{n,m}^{\epsilon_n} \quad \text{for } m \in \mathbb{N}. \tag{6.11}$$

Since $\hat{f} \in L^1(\mathbb{R}^d \times [0, \infty))$ there exists $N_0 \in \mathbb{N}$ such that $\int_{\hat{\mathcal{B}}_{N_0}^c} |\hat{f}(x, t)| dx dt < \frac{\epsilon}{4\|\hat{g}\|_\infty}$. Then closely mimicking the argument of Theorem 4.2 (see Step 2), we have that for each $\epsilon > 0$

there exists a positive constant \hat{M}_0 (depending on $\hat{f}, \hat{g}, \epsilon$) such that for all $m \geq \hat{M}_0$

$$\left| \int_{[0, \infty)} \int_{\mathbb{R}^d} \hat{f}(x, t) \int_{\mathbb{U}} \hat{g}(\zeta) v(x, t) (d\zeta) dx dt - \int_{[0, \infty)} \int_{\mathbb{R}^d} \hat{f}(x, t) \int_{\mathbb{U}} \hat{g}(\zeta) \bar{v}_m^\epsilon(x, t) (d\zeta) dx dt \right| \leq \epsilon. \tag{6.12}$$

Let $\{\hat{f}_k\}_{k \in \mathbb{N}}$ and $\{h_j\}_{j \in \mathbb{N}}$ be countable dense set in $L^1(\mathbb{R}^d \times [0, \infty))$ and $C(\mathbb{U})$ respectively. Suppose that $f \in L^1(\mathbb{R}^d \times [0, \infty)) \cap L^2(\mathbb{R}^d \times [0, \infty))$ and $g \in C_b(\mathbb{R}^d \times [0, \infty) \times \mathbb{U})$. Since $f \in L^1(\mathbb{R}^d \times [0, \infty))$, for given $\epsilon > 0$ there exists $N_1 \in \mathbb{N}$ such that $\int_{\hat{\mathbb{B}}_{N_1}^\epsilon} |f(x, t)| dx dt \leq \frac{\epsilon}{4\|g\|_\infty}$. We know that in $C_b(\hat{\mathbb{B}}_{N_1} \times \mathbb{U})$ the functions of the form $\{\sum_i^m r_i(x, t) p_i(\zeta)\}_{m \in \mathbb{N}}$ forms an algebra which contains constants, where $r_i \in C(\hat{\mathbb{B}}_{N_1})$ and $p_i \in C(\mathbb{U})$. Thus by Stone-Weierstrass theorem there exists \hat{m} (large enough) such that

$$\sup_{\hat{\mathbb{B}}_{N_1} \times \mathbb{U}} \left| g(x, t, \zeta) - \sum_i^{\hat{m}} r_i(x, t) p_i(\zeta) \right| \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d \times [0, \infty))}}. \tag{6.13}$$

Since $p_i \in C(\mathbb{U})$ one can choose $h_{j(i)} \in C(\mathbb{U})$ such that

$$\sup_{\zeta \in \mathbb{U}} |p_i(\zeta) - h_{j(i)}(\zeta)| \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d \times [0, \infty))} \|r_i\|_\infty}. \tag{6.14}$$

Also, since $f r_i \in L^1(\mathbb{R}^d \times [0, \infty))$ there exists $\hat{f}_{k(i)}$ such that

$$\int_{\hat{\mathbb{B}}_{N_1}} |f(x, t) r_i(x, t) - \hat{f}_{k(i)}(x, t)| dx dt \leq \frac{\epsilon}{24\|f\|_{L^1(\mathbb{R}^d \times [0, \infty))} \|h_i\|_\infty}. \tag{6.15}$$

Thus, in view of (6.13), (6.14), (6.15), following the steps of Theorem 4.2 (see Step 3) we conclude that

$$\left| \int_0^\infty \int_{\mathbb{R}^d} f(x, t) \int_{\mathbb{U}} g(x, t, \zeta) v(x, t) (d\zeta) dx dt - \int_0^\infty \int_{\mathbb{R}^d} f(x, t) \int_{\mathbb{U}} g(x, t, \zeta) \bar{v}_m^\epsilon(x, t) (d\zeta) dx dt \right| \leq \epsilon,$$

for all $m \geq \hat{M}_1$, for some positive constant \hat{M}_1 . This completes the proof of the theorem. \square

Near optimality of piecewise constant policies for finite horizon cost

Now, from Theorem 6.2 and Theorem 6.1, we have the following near-optimality results.

Theorem 6.3. *Suppose that assumptions (A1), (A3) and (B1) hold. Then for any given $\epsilon > 0$ there exists a piecewise constant policy $\bar{v}_\epsilon^* \in \mathfrak{U}_m$ such that*

$$\mathcal{J}_T(x, \bar{v}_\epsilon^*) \leq \mathcal{J}_T^*(x) + \epsilon \quad \text{for all } x \in \mathbb{R}^d. \tag{6.16}$$

Proof. From our previous discussion, we know that there exists $v^* \in \mathfrak{U}_m$ such that $\mathcal{J}_T(x, v^*) = \mathcal{J}_T^*(x)$. Since the space of piecewise constant policies are dense in \mathfrak{U}_m (see Theorem 6.2) and the map $v \mapsto \mathcal{J}_T(x, v)$ is continuous on \mathfrak{U}_m (see Theorem 6.1), for any given $\epsilon > 0$, one can find a piecewise constant policy $\bar{v}_\epsilon^* \in \mathfrak{U}_m$ such that (6.16) holds. \square

Remark 6.4. In view of the existence results as in [36, Chapter 4], in obtaining the near optimality of piecewise constant Markov policies for finite horizon costs, one can relax the uniform boundedness assumption (B1), in particular, under (A1)-(A3) we can deduce similar results. Which extends the results of [35, 33, 43] to a more general control model, where the drift term b and the diffusion matrix σ are allowed to be unbounded.

Conclusion

We studied regularity properties of induced cost (under several criteria) on a controlled diffusion process with respect to a control policy space defined by Borkar [15]. We then studied implications of these properties on existence and, in particular, approximations for optimal controlled diffusions. Via such a unified approach, we arrived at very general approximation results for optimal control policies by quantized (finite action / piecewise constant) stationary control policies for a general class of controlled diffusions in the whole space \mathbb{R}^d as well as time-discretizations for the criteria with finite horizons.

References

- [1] R. A. Adams, *Sobolev spaces*, Academic Press, New York, 1975. MR0450957
- [2] A. Arapostathis, *On the policy iteration algorithm for nondegenerate controlled diffusions under the ergodic criterion*, in Optimization, control, and applications of stochastic systems, 1–20, Systems Control Found. Appl., pp. 1–12, Birkhäuser/Springer, New York, 2012. MR2961376
- [3] A. Arapostathis, *On the non-uniqueness of solutions to the average cost HJB for controlled diffusions with near-monotone costs*, CoRR **abs/1309.6307** (2013), 55–62.
- [4] A. Arapostathis and V. S. Borkar, *Uniform recurrence properties of controlled diffusions and applications to optimal control*, SIAM Journal on Control and Optimization **48** (2010), no. 7, 4181–4223. MR2665463
- [5] A. Arapostathis, V. S. Borkar, and M. K. Ghosh, *Ergodic control of diffusion processes*, Encyclopedia of Mathematics and its Applications, vol. 143, Cambridge University Press, Cambridge, 2012. MR2884272
- [6] A. Arapostathis and S. Yüksel, *Convex analytic method revisited: Further optimality results and performance of deterministic policies in average cost stochastic control*, Journal of Mathematical Analysis and Applications **517** (2023), no. 2, 126567. MR4477061
- [7] G. Barles and E. R. Jakobsen, *On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique **36** (2002), no. 1, 33–54. MR1916291
- [8] G. Barles and E. R. Jakobsen, *Error bounds for monotone approximation schemes for hamilton-jacobi-bellman equations*, SIAM Journal on Numerical Analysis **43** (2006), no. 2, 540–558. MR2177879
- [9] E. Bayraktar and A. D. Kara, *An approximate reinforcement learning algorithm for controlled diffusion processes*, SIAM Journal on Mathematics of Data Science **5** (2023), 615–638.
- [10] A. Bensoussan and J. L. Lions, *Impulse control and quasi-variational inequalities*, Bristol: Gauthier-Villars, 1984. MR0673169
- [11] D. P. Bertsekas, *Convergence of discretization procedures in dynamic programming*, IEEE Transactions on Automatic Control **20** (1975), 415–419. MR0446700
- [12] A. G. Bhatt and V. S. Borkar, *Occupation measures for controlled Markov processes: characterization and optimality*, Annals of Probability **24** (1996), 1531–1562. MR1411505
- [13] V. S. Borkar, *A remark on the attainable distributions of controlled diffusions*, Stochastics **18** (1986), 17–23. MR0854653
- [14] V. S. Borkar, *Optimal control of diffusion processes*, Pitman Research Notes in Mathematics Series, vol. 203, Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989. MR1005532
- [15] V. S. Borkar, *A topology for Markov controls*, Applied Mathematics and Optimization **20** (1989), 55–62. MR0989430
- [16] V. S. Borkar, *Controlled diffusion processes*, Probab. Surveys **2** (2005), 213–244. MR2178045
- [17] V. S. Borkar and M. K. Ghosh, *Ergodic control of multidimensional diffusions. i. the existence results*, SIAM J. Control Optim. **26** (1988), 112–126. MR0923306

- [18] V. S. Borkar and M. K. Ghosh, *Controlled diffusions with constraints*, Journal of Mathematical Analysis and Applications **152** (1990), no. 1, 88–108. MR1072929
- [19] V. S. Borkar and M. K. Ghosh, *Ergodic control of multidimensional diffusions ii. adaptive control*, Appl. Math. Optim. **21** (1990b), 191–220. MR1019400
- [20] H. Brezis, *Functional analysis, sobolev spaces and partial differential equations*, Universitext, Springer-Verlag, New York, 2010. MR2759829
- [21] R. M. Dudley, *Real analysis and probability*, Cambridge University Press, Cambridge, 2002. MR1932358
- [22] F. Dufour and T. Prieto-Rumeau, *Finite linear programming approximations of constrained discounted Markov decision processes*, SIAM Journal on Control and Optimization **51** (2013), no. 2, 1298–1324. MR3036990
- [23] J. Dugundji, *An extension of tietze's theorem*, Pacific Journal of Mathematics **1** (1951), no. 3, 353–367. MR0044116
- [24] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, second ed., Grundlehren der Mathematischen Wissenschaften, vol. 224, Springer-Verlag, Berlin, 1983. MR737190
- [25] E. R. Jakobsen, A. Picarelli, and C. Reisinger, *Improved order 1/4 convergence for piecewise constant policy approximation of stochastic control problems*, Electronic Communications in Probability **24** (2019), 1–10. MR4017133
- [26] A. D. Kara and S. Yüksel, *Near optimality of finite memory feedback policies in partially observed Markov decision processes*, J. Mach. Learn. Res. **23** (2022), 11:1–11:46. MR4420736
- [27] N. Krylov, *On the rate of convergence of finite-difference approximations for bellman's equations*, St. Petersburg Math. J. **9** (1998), 639–650. MR1466804
- [28] N. Krylov, *Approximating value functions for controlled degenerate diffusion processes by using piece-wise constant policies*, Electron. J. Probab. **4** (1999), 1–19. MR1668597
- [29] N. Krylov, *On the rate of convergence of finite-difference approximations for bellmans equations with variable coefficients*, Probab Theory Relat Fields **117** (2000), 1–16. MR1759507
- [30] N. Krylov, *Mean value theorems for stochastic integrals*, Ann. Probab. **29** (2001), 385–410. MR1825155
- [31] H. J. Kushner, *Probability methods for approximations in stochastic control and for elliptic equations*, Math. Sci. Eng., vol. 129, Academic Press, New York, 1977. MR0469468
- [32] H. J. Kushner, *Heavy traffic analysis of controlled queueing and communication networks*, Stoch. Model. Appl. Probab., vol. 47, Springer-Verlag, New York, 2001.
- [33] H. J. Kushner, *Numerical approximations for stochastic differential games*, SIAM Journal on Control and Optimization **41** (2002), no. 2, 457–486. MR1920267
- [34] H. J. Kushner, *A partial history of the early development of continuous-time nonlinear stochastic systems theory*, Automatica **50** (2014), no. 2, 303–334. MR3163780
- [35] H. J. Kushner and P. G. Dupuis, *Numerical methods for stochastic control problems in continuous time*, Springer-Verlag, Berlin, New York,, 2001. MR1800098
- [36] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ura l' ceva, *Linear and quasi-linear equations of parabolic type*, Translated from the Russian by S. Smith. Translations of Mathematical Monographs, vol. 23, Providence, RI: Americal Mathematical Society, 1967. MR0241822
- [37] P. L. Lions, *Optimal control of diffusion processes and hamilton-jacobi-bellman equations. i. the dynamic programming principle and applications*, Comm. Partial Differential Equations **8** (1983), no. 10, 1101–1174. MR0709164
- [38] P. L. Lions, *Optimal control of diffusion processes and hamilton-jacobi-bellman equations. ii. viscosity solutions and uniqueness*, Comm. Partial Differential Equations **8** (1983), no. 11, 1229–1276. MR0709162
- [39] R. Munos and P. Bourguine, *Reinforcement learning for continuous stochastic control problems*, vol. 10, MIT Press, 1997.
- [40] H. Pham, *Continuous-time stochastic control and applications with financial applications*, Stochastic Modelling and Applied Probability, vol. 61, Springer, 2009. MR2533355

- [41] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko, *The mathematical theory of optimal processes*, Wiley Interscience, New York,, 1962, Neustadt. MR0166037
- [42] C. Reisinger and P. A. Forsyth, *Piecewise constant policy approximations to hamilton-jacobi-bellman equations*, Applied Numerical Mathematics **103** (2015), 27–47. MR3458022
- [43] C. Reisinger and P.A. Forsyth, *Piecewise constant policy approximations to hamilton-jacobi-bellman equations*, Applied Numerical Mathematics **103** (2016), 27–47. MR3458022
- [44] B. Van Roy, *Performance loss bounds for approximate value iteration with state aggregation*, Mathematics of Operations Research **31** (2006), no. 2, 234–244. MR2233994
- [45] N. Saldi, T. Linder, and S. Yüksel, *Finite approximations in discrete-time stochastic control*, Systems & Control: Foundations & Applications, Springer, Cham, 2018. MR3791811
- [46] N. Saldi, S. Yüksel, and T. Linder, *Near optimality of quantized policies in stochastic control under weak continuity conditions*, Journal of Mathematical Analysis and Applications **435** (2016), no. 1, 321–337. MR3423399
- [47] N. Saldi, S. Yüksel, and T. Linder, *On the asymptotic optimality of finite approximations to Markov decision processes with Borel spaces*, Mathematics of Operations Research **42** (2017), no. 4, 945–978. MR3722422
- [48] N. Saldi, S. Yüksel, and T. Linder, *Asymptotic optimality of finite model approximations for partially observed Markov decision processes with discounted cost*, IEEE Transactions on Automatic Control **65** (2020), no. 1, 130–142. MR4052863
- [49] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, vol. 233, Springer Science & Business Media, 1997. MR0532498

Acknowledgments. We wish to thank the anonymous referees for the careful reading of the manuscript and proposed improvements. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).