Journal of Dynamics and Games
©American Institute of Mathematical Sciences
Volume **1**, Number **4**, October **2014**
doi:10.3934/jdg.2014.1.639
pp. **639–659**

A LIMIT THEOREM FOR MARKOV DECISION PROCESSES

Mathias Staudigl

Center for Mathematical Economics, University Bielefeld
Postfach 100131, 33501 Bielefeld, Germany

(Communicated by Onesimo Hernandez-Lerma)

Abstract. In this paper I prove a deterministic approximation theorem for a sequence of Markov decision processes with finitely many actions and general state spaces as they appear frequently in economics, game theory and operations research. Using viscosity solution methods no a-priori differentiabililty assumptions are imposed on the value function.

1. **Introduction.** In this paper I study the following standard sequential decision problem. Consider a controlled Markov chain $\{X_n^\varepsilon\}_{n\in\mathbb{N}_0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and taking values in $\mathbb{R}^d$. The evolution of this process is controlled by an action process $\{A_n^\varepsilon\}_{n\in\mathbb{N}_0}$, which is assumed to take values in a finite set $\mathcal{A}$. The controlled evolution of the state is assumed to follow the system equation

$$\begin{cases} X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon f_{n+1}^\varepsilon(X_n^\varepsilon, A_n^\varepsilon) & \forall n \in \mathbb{N}_0 \\ X_0^\varepsilon = x \in \mathcal{X} \subset \mathbb{R}^d. \end{cases} \tag{1}$$

Assume that real time is a continuous variable, taking values in the set of non-negative real numbers $t \in \mathbb{R}_+$. Fitting the discrete process $\{(X_n^\varepsilon, \hat{A}_n^\varepsilon)\}_{n\in\mathbb{N}}$ into continuous time by defining càdlàg processes

$$\hat{X}^\varepsilon(t) = X_n^\varepsilon, \text{ and } \hat{A}^\varepsilon(t) = A_n^\varepsilon \qquad n\varepsilon \leq t < (n+1)\varepsilon, n \geq 0,$$

we obtain a jump process, with deterministic periods between consecutive jumps of length ε. Consider a decision maker, whose objective is to maximize his total sum of stage payoffs over an infinite time horizon and discount factor $\lambda_\varepsilon := e^{-r\varepsilon}$. Assume that the decision maker is an expected utility maximizer so that his preferences have a numerical representation as

$$U(x,\sigma) := E_x^\sigma \left[\sum_{n=0}^\infty (1-\lambda_\varepsilon)\lambda_\varepsilon^n u(X_n^\varepsilon, A_n^\varepsilon) \right].$$

Using the continuous-time interpolations, this preference relation can be equivalently represented as

$$E_x^\sigma \left[\int_0^\infty r e^{-rt} u(\hat{X}^\varepsilon(t), \hat{A}^\varepsilon(t)) \mathrm{d}t \right].$$

2010 *Mathematics Subject Classification.* Primary: 49L20, 49L25, 90C40; Secondary: 60J20, 60F17.

Key words and phrases. Markov decision processes, optimal control, viscosity solutions, weak convergence methods.

The author is supported by U.S. Air Force OSR Grant FA9550-09-0538 and the Vienna Science and Technology Fund (WWTF) under project fund MA 09-017.

The mapping σ is a (behavior) *strategy* for the decision maker, essentially describing a probability distribution over actions at each decision node. Formally, it is a collection of functions from the observation process of the decision maker to the set of probability distributions over the available actions, denoted by $\Delta(\mathcal{A})$. Precise definitions will be given in Section 2.1.

As a comparison problem, consider the deterministic optimal control problem

$$\sup_{\alpha\in\mathcal{S}} \int_0^\infty re^{-rt}u(y_x(t,\alpha),\alpha(t))\mathrm{d}t \tag{2}$$

$$\text{s.t. } \dot{y}_x(t,\alpha) = b(y_x(t,\alpha),\alpha(t)),\ y_x(0,\alpha) = x \tag{3}$$

where $\mathcal{S}$ is the set of measurable functions $\alpha : \mathbb{R}_+ \to \Delta(\mathcal{A})$, and b is a suitably defined Lipschitz continuous and bounded vector field. In this paper we address the question under which conditions solutions (i.e. value function and the strategies) of the stochastic sequential decision model, with decisions made on the discrete time grid $\{0,\varepsilon,2\varepsilon,\ldots\}$, converge to solutions of the deterministic optimal control problem described above. The motivation for studying this question are two-fold. The first motivation is guided by practical considerations. There are some arguments in favor of using deterministic continuous optimal control problems over the stochastic discrete decision processes. Solving the stochastic decision problem numerically is often a computationally very intensive task, due to the "curse of dimensionality" of dynamic programming.[1] The deterministic optimal control problem is often amenable to efficient numerical methods which seem to perform better than algorithms based on dynamic programming (see [11] for illustrations). Second, in some situations, the continuous deterministic formulation allows for an analytic treatment of the decision problem, using either dynamic programming methods, or the Pontryagin maximum principle (see [27, 23, 24] for fruitful applications of this idea). Hence, if one has the theoretical justification to replace the stochastic decision problem by a deterministic one, there are some good reasons to do that. My second motivation for investigating this question is in establishing convergence results for dynamic games in discrete time to dynamic games in continuous time. The present paper is therefore the basis for a model in which the limit dynamic game is characterized by a deterministic ordinary differential equation (i.e. a differential game).

1.1. **Related literature.** Related convergence and approximation questions are at the core of optimal control theory. Indeed, the present study is heavily influenced by the Markov chain method developed by [20]. This is a powerful numerical approximation tool to obtain feedback controls in stochastic and deterministic optimal control problems. Similar approaches can be found in [5, 13, 9] and [2]. The difference between these papers and the present one is the nature of the question I am addressing. While the above mentioned literature is interested to construct a numerical approximation scheme in order to approximate a given optimal control problem, I instead ask the question, given a discrete controlled Markov chain model, what is the limit as the discretization becomes arbitrarily fine?

While writing this paper I have learned from the paper by [11]. They establish a limit result for a finite-horizon Markov decision process converging to a deterministic optimal control problem. This paper differs from [11] in the problem formulation as well as in the proof techniques. First I study infinite horizon problems with

[1]Note that for numerical implementation of the decision problem one needs to discretize the state space somehow. Usually at this stage the curse of dimensionality kicks in.

discounting. Second, my proof techniques are based on a combination of weak convergence arguments and viscosity solution techniques, whereas [11] rely on ideas from stochastic approximation theory. Third, the paper by [11] studies a class of optimization problems where the controlled dynamics depends on the empirical measure of the behavior of N small sub-entities. This number N is used as the mesh-size, and it is shown that as the number of entities grows to infinity, the family of discrete-time problems converges to a "mean-field" limit model, which is characterized by a deterministic differential equation. This deterministic limit dynamics can be interpreted as the mean-field dynamic, describing the evolution of the aggregate behavior of the population.[2]

In game theory there is a burgeoning literature on continuous-time limits of discrete-time games. [6] and [12] study various versions of zero-sum repeated games with incomplete information in the spirit of [1]. Using PDE techniques, [6] prove the existence of a limit value as the frequency of play increases. A similar result, in a different model setting, is reported in [12]. [21] performs a limit analysis for stochastic games with parameterized transition probabilities. Again, results for the convergence of the values for the family of discrete-time games are obtained under various assumptions on the problem data. All these studies have in common that only the convergence of the value is investigated. In this paper I do not only demonstrate the convergence of the value of the Markov decision process, but also prove the convergence of the decision maker's strategy. I belief that this is important, because it sheds light on conceptual problems we are facing in any convergence analysis: Usually the space of of discrete-time strategies is not weakly compact, and thus some compactification method has to be used when passing to the continuous-time limit [29, 7]. Once this is done, the limit objects are usually not interpretable as strategies in the original sense, but rather as measure-valued processes, also called *relaxed controls* (see [28]). Relaxed controls have no equivalent in continuous-time games, though can be interpreted as a version of a correlated strategy ([22]). We show in this note that this problem is avoidable in single-player decision problems, by interpreting the limit objects as open-loop controls.

1.2. **Examples.**

1.2.1. *Dynamic pricing policy of a monopoly.* Consider an infinitely lived monopoly, who sets prices $a \in \mathcal{A} = \{1, 2, \dots, m\}$. The monopolist can announce prices at the periods $\{0, \varepsilon, 2\varepsilon, \dots\}$. It faces a stochastic market demand, following a Markovian dynamics $\{X_n^\varepsilon\}_{n\in\mathbb{N}_0}$ with sample paths given by (1). The vector field $f_n^\varepsilon(x, a)$ capture the random changes in market demand, given the current demand is x and the quoted price is $a \in \mathcal{A}$. The probability measures $\mu_a^\varepsilon(\cdot|x)$ define the law of the random changes in demand, given the current demand is x and the monopolist announces a price a. The monopolist has a flow profit function $u(x, a)$. A strategy for the monopolist is to design an optimal pricing strategy $\{\sigma_n\}_{n=0}^\infty$, where σ_n is a function of the demand history to probability distributions over prices. Hence, the monopolists' problem is to maximize

$$U(x, \sigma) = E_x^\sigma \left[\sum_{n=0}^{\infty} (1 - \lambda_\varepsilon)\lambda_\varepsilon^n u(X_n^\varepsilon, A_n^\varepsilon) \right]$$

[2]This model setting is typical in stochastic dynamics consisting of a large number of identical and weakly interacting random processes ("particles").

where $x \in \mathbb{R}$ is the initially given demand, assumed to be known to the monopolist. As $\varepsilon \to 0$ the monopolist is able to post prices in arbitrary short time spans, and thus can react arbitrarily fast to the random market demand. If the market is sufficiently stable where random fluctuations over very small time spans are negligible, a deterministic approximation to this model seems to the sensible.

1.2.2. *Optimal stopping.* A firm has to decide when to exit an industry. The state of the market is modeled by a discrete-time Markov chain $\{X_n^\varepsilon\}_{n\in\mathbb{N}_0}$ which lives on $\mathbb{R}_+$. For concreteness think of X_n^ε as the market price in period n. Real time t takes values in the set of non-negative reals $\mathbb{R}_+$ and the firm receives information on the prevailing market price only at discrete points in time contained in the grid $\{0, \varepsilon, 2\varepsilon, \ldots\}$. The firm is small, and therefore cannot influence the evolution of the price dynamics. However, it has a model for the time series of prices, which is the AR(1) process given by eq. (1).

In each period the firm can decide whether to stay or exit the market. This is modeled by a binary action set $\mathcal{A} = \{0, 1\}$, where action 0 means to exit the market and 1 means to stay in the market. In each period in which the firm stays in the market it has to pay a random fee $-r(X_n^\varepsilon) < 0$, and the state evolves according to an uncontrolled Markov chain with transition function q^ε on a set of possible prices $\mathcal{K} \subseteq \mathbb{R}_+$. If the firm decides to exit the market in period $N \in \mathbb{N}$ it gets a terminal reward $g(X_N^\varepsilon)$ and the evolution of prices stops (or the firm does simply not monitor the price evolution anymore). The function $g(\cdot)$ is non-negative (otherwise the firm would want to exit immediately) and bounded. This problem is contained in our model setup by specifying the following data. The transition dynamics are $\mu_0^\varepsilon(\cdot|x) = \delta_{\mathbf{0}}$ and $\mu_1^\varepsilon(\cdot|x) = q^\varepsilon(\cdot|x) \in \mathsf{M}_1^+(\mathbb{R})$, where $q^\varepsilon(\cdot|x)$ is a given probability law modeling the uncontrolled evolution of the price time series. The utility rate function is given by

$$u(x,a) = \begin{cases} -r(x) & \text{if } a = 1, \\ g(x) & \text{if } a = 0. \end{cases}$$

The objective function of the decision maker is

$$U^\varepsilon(x,\sigma) = E_x^\sigma \left[\sum_{n=0}^{\infty} (1-\lambda_\varepsilon)\lambda_\varepsilon^n u(X_n^\varepsilon, A_n^\varepsilon) \right]$$

where σ is a measurable function mapping histories of the state process into probability distributions over actions (i.e. a *strategy*). Now suppose that the information about current prices appears in periods of length ε. In real time, the price time series evolves therefore according to the step process $\hat{X}^\varepsilon$, and the decision whether to exit the market or stay in the market can be made at all time points which are multiples of the step size ε. In the limit as ε approaches 0 the firm monitors the price evolution with more and more accuracy, and can also react to the price dynamics at virtually any point in real time. The results reported in this paper investigate such a scenario where in the limit as $\varepsilon \to 0$ the limit price dynamics can be modeled by a deterministic differential equation.

2. **Problem formulation.**

2.1. **The discrete-time problem.** Let $\{(X_n^\varepsilon, A_n^\varepsilon)\}_{n\in\mathbb{N}_0}$ be a stochastic process taking values in the set $\mathbb{R}^d \times \mathcal{A}$, whose sample paths satisfy the dynamical systems equation (1). Each A_n^ε is an $\mathcal{A}$-valued random variable, adapted to the filtration $\mathcal{F}_n^\varepsilon = \sigma(X_0^\varepsilon, \ldots, X_n^\varepsilon)$, and controlling the evolution of the state process. The law of

the random variables A_n^ε for $n = 0, 1, 2, \ldots$ are determined by a (behavior) *strategy*. A *strategy* is a collection of functions $\sigma = \{\sigma_n\}_{n\in\mathbb{N}_0}$, where each $\sigma_n(\cdot)$ is a probability distribution over the finite set of actions $\mathcal{A} := \{1, 2, \ldots, m\}$, adapted to the sigma-algebra $\mathcal{F}_n^\varepsilon$.[3] A strategy is *Markov* if, for every n, we can express the behavior strategy σ_n in terms of a single function $\underline{\alpha} : \mathbb{R}^d \to \Delta(\mathcal{A})$, so that

$$\sigma_n(a|x_0, \ldots, x_n) = \underline{\alpha}(a|x_n) \quad \forall n \geq 0, a \in \mathcal{A}. \tag{4}$$

Markov strategies are of fundamental importance in Markov decision processes, as we will see in due course.[4] $\{f_n^\varepsilon(x, a)\}_{n\in\mathbb{N}}$ is a sequence of i.i.d random variables with common law $\mu_a^\varepsilon(\cdot|x)$ on $\mathbb{R}^d$. The collection of probability distributions $\mu_1^\varepsilon(\cdot|x), \ldots, \mu_m^\varepsilon(\cdot|x)$, defined on the Borel sets of $\mathbb{R}^d$, are the control measures of the Markov decision process.Let $\Omega = (\mathbb{R}^d \times \mathcal{A})^{\mathbb{N}_0}$ denote the sample path space of the controlled Markov chain, and let $\mathcal{F}$ denote the σ-algebra generated by the finite cylinder sets. By the Ionescu-Tulcea Theorem (see e.g. [3]), each strategy σ defines a unique probability measure P_x^σ on $(\Omega, \mathcal{F})$ with the characteristics

$$\begin{aligned}
&P_x^\sigma(X_0^\varepsilon \in \Gamma) = \delta_x(\Gamma),\\
&P_x^\sigma(X_{n+1}^\varepsilon \in \Gamma | X_n^\varepsilon = x, A_n^\varepsilon = a) = Q^\varepsilon(\Gamma|x, a),\\
&P_x^\sigma(A_n^\varepsilon = a | X_0^\varepsilon, \ldots, X_n^\varepsilon) = \sigma(a|X_0^\varepsilon, \ldots, X_n^\varepsilon),
\end{aligned}$$

where Γ is Borel measurable subset of $\mathbb{R}^d$, and

$$Q^\varepsilon(\Gamma|x, a) = \mu_a^\varepsilon\left(\frac{1}{\varepsilon}(\Gamma - x)|x\right) \quad \forall (x, a) \in \mathbb{R}^d \times \mathcal{A}.$$

Under this (canonical) construction of the controlled Markov chain we think of the random variables X_n^ε and A_n^ε as the coordinate processes $X_n^\varepsilon(\omega) = x_n$ and $A_n^\varepsilon(\omega) = a_n$, for every $\omega = (x_0, a_0, \ldots, x_n, a_n, \ldots) \in \Omega$.Given a strategy σ let E_x^σ denote expectations with respect to the probability measure P_x^σ. The objective of the decision maker is to maximize his normalized expected infinite horizon discounted utility

$$U^\varepsilon(x, \sigma) = E_x^\sigma\left[\sum_{n=0}^{\infty}(1 - \lambda_\varepsilon)\lambda_\varepsilon^n u(X_n^\varepsilon, A_n^\varepsilon)\right].$$

The discount factor per unit time λ_ε is defined as $\lambda_\varepsilon = e^{-r\varepsilon}$. $r > 0$ is the discount rate. The factor $(1-\lambda_\varepsilon)$ provides the correct normalization of the stream of utilities. The maximized utility of the decision maker, or the *value function*, is defined as

$$V^\varepsilon(x) = \sup_\sigma U^\varepsilon(x, \sigma), \tag{5}$$

where the supremum is taken over all strategies available to the decision maker. A standard result in Markov decision processes is that the decision maker does not gain by using more complicated strategies than Markov strategies. Indeed, for every fixed $\varepsilon > 0$, it is well known (see e.g. [15]) that the decision maker can choose a stationary Markov strategy $\underline{\alpha}^\varepsilon : \mathbb{R}^d \to \Delta(\mathcal{A})$ which solves the decision problem, i.e.

$$V^\varepsilon(x) = U^\varepsilon(x, \underline{\alpha}^\varepsilon) \quad \forall x \in \mathbb{R}^d.$$

[3]Technically speaking, each σ_n is a stochastic kernel on $\mathcal{A}$ given $(\mathbb{R}^d)^{n+1}$. See [3] for the precise measure-theoretic definition of stochastic kernels.

[4]In the literature on stochastic games such strategies are often referred to as stationary.

2.1.1. *Standing hypothesis.* This section provides a collection of all the technical assumptions we impose on the problem data. The controlled stochastic dynamics defining the optimization problem in discrete time is given by the random walk model

$$\begin{cases} X^\varepsilon_{n+1} = X^\varepsilon_n + \varepsilon f^\varepsilon_{n+1}(X^\varepsilon_n, A^\varepsilon_n) & \forall n \in \mathbb{N}_0, \\ X^\varepsilon_0 = x \in \mathcal{X} \subset \mathbb{R}^d, \end{cases}$$

where $\mathcal{X}$ is a given compact subset of possible initial conditions. We start with a uniform tightness condition on the distributions of the random vector fields $\{f_n(x,a)\}_{n\in\mathbb{N}}$.

Assumption 1. *The control measures $\mu^\varepsilon_1(\cdot|x), \dots, \mu^\varepsilon_m(\cdot|x)$ are supported on a common compact subset $\mathcal{K} \subset \mathbb{R}^d$ for each $x \in \mathbb{R}^d$.*

This assumption implies that the vector fields $b^\varepsilon(x,a)$ are all contained in the closed convex hull of the compact set $\mathcal{K}$.

The next assumption is a continuity assumption on the *drift* of the state process $\{X^\varepsilon_n\}_{n\in\mathbb{N}_0}$, defined as the conditional mean increment of the process of the controlled random walk. We denote the *drift* $b^\varepsilon : \mathbb{R}^d \times \mathcal{A} \to \mathbb{R}^d$ by

$$b^\varepsilon(x,a) := E^\sigma_x\left[\frac{1}{\varepsilon}\left(X^\varepsilon_{n+1} - X^\varepsilon_n\right) \mid X^\varepsilon_n = x, A^\varepsilon_n = a\right] = \int_{\mathbb{R}^d} z\mu^\varepsilon_a(\mathrm{d}z|x). \tag{6}$$

Assumption 2. *The function $b^\varepsilon : \mathbb{R}^d \times \mathcal{A} \to \mathbb{R}^d$ is Lipschitz continuous and converges to a Lipschitz continuous function $b : \mathbb{R}^d \times \mathcal{A} \to \mathbb{R}^d$ locally uniformly on compact sets.*

Since the drift b^ε is contained in the closed convex hull of the compact set $\mathcal{K}$, Assumption 2 implies that the limit drift b takes also values in this set. Hence, the controlled vector field of the limit dynamics (3) is uniformly bounded by some constant $M_b > 0$:

$$\sup_{x\in\mathbb{R}^d} ||b(x,a)|| \leq M_b \quad \forall (x,a) \in \mathbb{R}^d \times \mathcal{A}, \tag{7}$$

and Lipschitz continuous for every control parameter $a \in \mathcal{A}$.

Now we impose some restriction on the utility flow function of the decision maker.

Assumption 3. *The utility flow function $u : \mathbb{R}^d \times \mathcal{A} \to \mathbb{R}$ is uniformly bounded and Hölder continuous for each action $a \in \mathcal{A}$:*

$$\sup_{x\in\mathbb{R}^d} |u(x,a)| \leq M_u \quad \forall a \in \mathcal{A}, \text{ and} \tag{8}$$

$$|u(x,a) - u(y,a)| \leq M_u||x-y||^\gamma \quad \forall x,y \in \mathbb{R}^d, a \in \mathcal{A} \tag{9}$$

for some constants $M_u > 0$ and $\gamma \in [0,1]$.

The final assumption we make concerns the scaling relationship between the variance of the increments of the state process and the step size ε. This assumption is essential in making the deterministic approximation result work, as it says that in the limit of small step sizes, sample paths of the state process look like solutions of an ordinary differential equation with drift b. This will be made precise in Section 5, where the technical details are provided.

Assumption 4. *The covariance matrix of the increments of the state process $\{X^\varepsilon_n\}_{n\in\mathbb{N}_0}$ satisfies the scaling relationship*

$$\mathrm{Var}^\sigma_x\left[X^\varepsilon_{n+1} - X^\varepsilon_n | X^\varepsilon_n = x, A^\varepsilon_n = a\right] \leq \varepsilon^2 M_v \tag{10}$$

for every $(x,a) \in \mathbb{R}^d \times \mathcal{A}$, for some uniform constant $M_v \geq 0$.

2.2. **The limit problem.** The limit problem is a deterministic optimal control problem where the decision maker wants to maximize his total discounted utility over an infinite time horizon. To formulate the limit problem, recall that $\Delta(\mathcal{A})$ denotes the set of probability distributions (mixed actions) over the set $\mathcal{A}$. As usual, extend the utility flow function to the domain $\mathbb{R}^d \times \Delta(\mathcal{A})$ linearly, and write (with the usual abuse of notation)

$$u(x,\alpha) := \sum_{a\in\mathcal{A}} u(x,a)\alpha(a).$$

Similarly, extend the drift b to $\mathbb{R}^d \times \Delta(\mathcal{A})$ by $b(x,\alpha) := \sum_{a\in\mathcal{A}} b(x,a)\alpha(a)$. The value function of the optimal control problem is defined as

$$v(x) := \sup_{\alpha\in\mathcal{S}} U(x,\alpha). \tag{OC}$$

The functional $U(x,\alpha)$, defined as

$$U(x,\alpha) := \int_0^\infty re^{-rt}u(y_x(t,\alpha),\alpha(t))\mathrm{d}t,$$

is the payoff of the decision maker under the deterministic strategy $\alpha \in \mathcal{S}$. Strategies induce the state dynamics

$$\dot{y}_x(t,\alpha) = b(y_x(t,\alpha),\alpha(t)),\ y_x(0,\alpha) = x. \tag{11}$$

Existence and uniqueness to solutions of the differential equation (11) is guaranteed by Assumption 2. The set of strategies of the decision maker is the set of measurable functions $\alpha : \mathbb{R}_+ \to \Delta(\mathcal{A})$,

$$\mathcal{S} := \{\alpha : \mathbb{R}_+ \to \Delta(\mathcal{A}) | \alpha(\cdot) \text{ measurable}\}.$$

Note that these functions are defined without any reference to the current state and hence are *open-loop* controls. Therefore, the set of admissible strategies in the limit problem, has no a-priori connection to the set of discrete-time strategies, as these have been defined as processes adapted to the filtration generated by the state process X^ε. Nevertheless, the convergence analysis in the forthcoming sections, will establish an interesting connection between these two strategy sets.

The following technical lemma establishes that the value function of the deterministic optimal control problem (OC) is an element of the space of continuous bounded functions $v \in \mathcal{C}_b(\mathbb{R}^d : \mathbb{R})$.

Lemma 2.1. *Under Assumptions 2, 1 and 3 the value function $v : \mathbb{R}^d \to \mathbb{R}$ satisfies*

$$|v(x)| \le M_u \qquad \forall x \in \mathbb{R}^d, \tag{12}$$

and it is Hölder continuous with coefficient $\gamma \in (0, \min\{\frac{r}{M_b}, 1\})$.

Proof. The proof of this Lemma is based upon standard arguments, which can be found in [2]. The uniform boundedness of the value function is a trivial consequence of the uniform boundedness of the utility flow function u, stated in Assumption 3. Indeed, for any strategy $\alpha \in \mathcal{S}$, we have

$$U(x,\alpha) = \int_0^\infty re^{-rt}u(y_x(t,\alpha),\alpha(t))\mathrm{d}t \le M_u r\int_0^\infty e^{-rt}\mathrm{d}t = M_u.$$

For the second statement, pick two points $x_1, x_2 \in \mathbb{R}^d$ and fix a strategy $\alpha \in \mathcal{S}$ such that

$$v(x_1) - \delta \le U(x_1,\alpha).$$

Such a strategy exists by definition of the supremum. Now $v(x_2) \geq U(x_2, \alpha)$, and w.l.o.g we assume that $v(x_1) > v(x_2)$. Then

$$\begin{aligned} |v(x_1) - v(x_2)| &\leq |U(x_1, \alpha) + \delta - U(x_2, \alpha)| \\ &= |\int_0^\infty re^{-rt}[u(y_{x_1}(t, \alpha), \alpha(t)) - u(y_{x_2}(t, \alpha), \alpha(t))]\mathrm{d}t + \delta|. \end{aligned}$$

By eq. (9) and standard estimates on solutions to ordinary differential equations, we see that

$$\begin{aligned} |u(y_{x_1}(t, \alpha), \alpha(t)) - u(y_{x_2}(t, \alpha), \alpha(t))| &\leq M_u||y_{x_1}(t, \alpha(t)) - y_{x_2}(t, \alpha(t))||^\gamma \\ &\leq M_u||x_1 - x_2||^\gamma e^{M_b \gamma t}. \end{aligned}$$

Using this estimate in the previous display shows that

$$|v(x_1) - v(x_2)| \leq M_u||x_1 - x_2||^\gamma| \int_0^\infty e^{(-r+\gamma M_b)t}\mathrm{d}t| + 2\delta.$$

To ensure that the integral on the right-hand side of this estimate converges, we consider three cases. If $r > M_b$ then the condition $\gamma < r/M_b$ is sufficient for convergence. In particular $\gamma = 1$ can be chosen, which shows that the value function is Lipschitz in this case. If $r = M_b$ any choice $\gamma \in (0, 1)$ can be made. Finally if $r < M_b$ then we need to pick $0 \leq \gamma < r/M_b$. This completes the proof the Lemma. □

The dynamic programming approach to deterministic optimal control theory allows us to characterize the value function as a solution to a partial differential equation of the first-order, known as the Hamilton-Jacobi-Bellman equation. The Hamiltonian associated to the optimal control problem (OC) is given by

$$H(x, p) = \max_{a \in \mathcal{A}} \{\langle p, b(x, a)\rangle + ru(x, a)\}.$$

Note that here we have already used the fact that the maximum value of the Hamiltonian expression will be attained at a pure action. It is well-known that, under the technical assumptions made in this paper, the value function v is the unique viscosity solution of the Hamilton-Jacobi-Bellman equation

$$rv(x) - H(x, Dv(x)) = 0 \quad \forall x \in \mathbb{R}^d. \tag{HJB}$$

See [2], chapters II and III. Since the Hamiltonian maximization condition can be formulated to optimize over elements in the finite action set $\mathcal{A}$, it follows that

$$v(x) = \sup_{\alpha \in \mathcal{S}^\#} \int_0^\infty re^{-rt}u(y_x(t, \alpha), \alpha(t))\mathrm{d}t$$

where $\mathcal{S}^\# \subset \mathcal{S}$ is the space of measurable $\mathcal{A}$-valued open-loop strategies. In order to be able to connect the discrete time control problem with the continuous one, we will have to restrict the class of strategies even further. In particular, we exhibit controls which may only be δ-optimal (for arbitrary tolerance bound δ), but be at least piecewise constant.[5] This offers are large enough class of continuous-time strategies which can easily be adapted to the discrete-time problem. To construct piecewise constant suboptimal strategies we replace the optimal control problem by a deterministic dynamic programming problem, which can be interpreted as the

[5]Note that if $\alpha \in \mathcal{S}^\#$ is piecewise continuous it must be piecewise constant on the intervals of continuity.

"expected deterministic" version (as in [26]) of the Markov decision process we are studying.[6] For each $\varepsilon > 0$ let

$$\mathcal{S}_\varepsilon^\# := \{\alpha \in \mathcal{S} | \alpha(\cdot) \text{ is piecewise constant on } [n\varepsilon, (n+1)\varepsilon), n \in \mathbb{N}_0\}. \tag{13}$$

For each strategy $\alpha \in \mathcal{S}_\varepsilon^\#$ define a controlled trajectory recursively on the time grid $\{0, \varepsilon, 2\varepsilon, \ldots\}$ by

$$\begin{aligned} y_x^\varepsilon(n\varepsilon, \alpha) &= x + \varepsilon \sum_{k=0}^{n-1} b(y_x(k\varepsilon, \alpha), \alpha(k\varepsilon)), \\ y_x^\varepsilon(0, \alpha) &= x. \end{aligned}$$

Interpolate the state trajectory by setting $y^\varepsilon(t, \alpha) = y^\varepsilon(n\varepsilon, \alpha)$ for each $t \in [n\varepsilon, (n+1)\varepsilon), n \in \mathbb{N}_0$. In terms of this continuous time interpolation it is easily seen, recalling the identity $\lambda_\varepsilon = e^{-r\varepsilon}$, that

$$\begin{aligned} U(x, \alpha) &= \sum_{n=0}^{\infty} (1 - \lambda_\varepsilon) \lambda_\varepsilon^n u(y_x^\varepsilon(n\varepsilon, \alpha), \alpha(n\varepsilon)) \\ &= r \int_0^\infty e^{-rt} u(y_x^\varepsilon(t, \alpha), \alpha(t)) \mathrm{d}t \leq v(x) \end{aligned}$$

where the last inequality follows from the maximality of the value function. This holds for every piecewise constant strategy $\alpha \in \mathcal{S}_\varepsilon^\#$.[7] Let

$$v^\varepsilon(x) := \sup_{\alpha \in \mathcal{S}_\varepsilon^\#} U(x, \alpha), \tag{OC$_\varepsilon$}$$

and let us put to record that, for each $\varepsilon > 0$, we have $v^\varepsilon \leq v$ pointwise. We now establish some simple, but useful, general properties of the value function v^ε.

Lemma 2.2. *The dynamic programming problem (OC$_\varepsilon$) has a solution, and the value function v^ε is unique. Moreover, it is uniformly bounded by the constant M_u and Hölder continuous with exponent $\gamma \in \left(0, \min\{\frac{r}{M_b}, 1\}\right)$.*

Proof. The proof of this Lemma is fairly standard, and so we only provide a sketch of the proof (for all details see e.g. [2], Section VI.4). First we show existence and uniqueness of solutions to (OC$_\varepsilon$). Define the operator T_ε, acting on bounded functions $v : \mathbb{R}^d \to \mathbb{R}$, by

$$T_\varepsilon v(x) := \max_{a \in \mathcal{A}} \{(1 - \lambda_\varepsilon) u(x, a) + \lambda_\varepsilon v(x + \varepsilon b(x, a))\}$$

Since $\lambda_\varepsilon \in (0, 1)$ for each $\varepsilon > 0$, it is easy to see that T_ε defines a contraction mapping on the space of bounded functions on $\mathbb{R}^d$. With the supremum norm this is a Banach space, and the Banach fixed point theorem states that there exists a unique function v^ε such that $T_\varepsilon v^\varepsilon = v^\varepsilon$ pointwise. Standard arguments then show that v^ε is the value function of the restricted problem (OC$_\varepsilon$). The uniform boundedness and Hölder-continuity of the function v^ε follow directly from the proof of Lemma 2.1. □

[6] Similar ideas appeared already in [10].

[7] A decision maker cannot obtain a higher utility by constraining himself to the smaller set of strategies $\mathcal{S}_\varepsilon^\#$.

Next, we construct a deterministic Markov strategy $\underline{a}^\varepsilon : \mathbb{R}^d \to \mathcal{A}$ which solves the problem (OC_ε). For each $x \in \mathbb{R}^d$ let

$$\underline{a}^\varepsilon(x) := \max\left\{a \in \mathcal{A} | v^\varepsilon(x) = (1-\lambda_\varepsilon)u(x,a) + \lambda_\varepsilon v^\varepsilon(x + \varepsilon b(x,a))\right\}. \tag{14}$$

Based on this Markov strategy, we define a piecewise constant strategy in continuous time by setting

$$y_x^\varepsilon(t) = y_x^\varepsilon(n\varepsilon) = x + \varepsilon \sum_{k=0}^{n-1} b(y_x^\varepsilon(k\varepsilon), \underline{a}^\varepsilon(y_x^\varepsilon(k\varepsilon))) \quad \forall t \in [n\varepsilon, (n+1)\varepsilon), n \geq 0,$$

and, for fixed initial condition $x \in \mathbb{R}^d$,

$$\alpha^\varepsilon(t) := \underline{a}^\varepsilon(y_x^\varepsilon(t)) \quad \forall t \geq 0. \tag{15}$$

It follows ([2], Theorem VI.4.6) that

$$v^\varepsilon(x) = \int_0^\infty re^{-rt} u(y_x^\varepsilon(t), \alpha^\varepsilon(t)) \mathrm{d}t = U(x, \alpha^\varepsilon).$$

It remains to check the consistency of the approximation procedure as $\varepsilon \to 0^+$.

Lemma 2.3. *$v^\varepsilon \to v$ as $\varepsilon \to 0^+$, where v is the unique viscosity solution to (HJB).*

Proof. For each $\varepsilon > 0$ the value function v^ε is uniformly bounded and Hölder continuous. By the Arzelà-Ascoli theorem we can assume that there exists a subsequence $\{v^{\varepsilon_j}\}_{j\in\mathbb{N}}$ such that $\varepsilon_j \to 0^+$ as $j \to \infty$, and along which $v^{\varepsilon_j} \to v$ locally uniformly on $\mathbb{R}^d$. To complete the proof, we will show that v is a viscosity solution of (HJB). This is done by showing that v is simultaneously a viscosity sub and supersolution of (HJB). Let $\phi \in \mathcal{C}^1(\mathbb{R}^d : \mathbb{R})$ be a given map. The function $v \in \mathcal{C}_b(\mathbb{R}^d : \mathbb{R})$ is a viscosity subsolution of (HJB) if, whenever $v - \phi$ has a local maximum at a point x, then

$$rv(x) - H(x, \nabla\phi(x)) \leq 0. \tag{16}$$

$v \in \mathcal{C}_b(\mathbb{R}^d : \mathbb{R})$ is a viscosity supersolution of (HJB) if, whenever $v - \phi$ has a local minimum at a point x, then

$$rv(x) - H(x, \nabla\phi(x)) \geq 0. \tag{17}$$

We now come to the verification. Take $\phi \in \mathcal{C}^1(\mathbb{R}^d : \mathbb{R})$ and $x_0 \in \mathbb{R}^d$ a local maximum point for $v - \phi$. Then there exists a closed ball $\mathbb{B}$ centered at x_0 such that

$$(v-\phi)(x_0) \geq (v-\phi)(x) \quad \forall x \in \mathbb{B}. \tag{18}$$

For each $j \in \mathbb{N}$ pick $x_0^j \in \arg\max_{x\in\mathbb{B}}(v^{\varepsilon_j} - \phi)(x)$. By the continuity of the value function v^{ε_j} and the local uniform convergence to v it follows that $x_0^j \to x_0$. Then, for j sufficiently large, the boundedness of the drift eq. (7) implies that $x_0^j + \varepsilon^j b(x_0^j, a) \in \mathbb{B}$ for all $a \in \mathcal{A}$. Therefore, eq. (18) implies that

$$v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j, a)) - v^{\varepsilon_j}(x_0^j) \leq \phi(x_0^j + \varepsilon^j b(x_0^j, a)) - \phi(x_0^j) \qquad \forall a \in \mathcal{A}. \tag{19}$$

The discrete dynamic programming equation corresponding to problem (OC_ε) states that

$$0 = \max_{a\in\mathcal{A}} \left\{(1-\lambda_{\varepsilon^j})u(x_0^j, a) + \lambda_{\varepsilon^j} v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j, a)) - v^{\varepsilon_j}(x_0^j)\right\}$$

for every $j \in \mathbb{N}$. This, together with eq. (19), implies that

$$0 = \max_{a\in\mathcal{A}} \left\{(1-\lambda_{\varepsilon^j})[u(x_0^j,a) - v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a))] + v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a)) - v^{\varepsilon_j}(x_0^j)\right\}$$
$$\leq \left\{(1-\lambda_{\varepsilon^j})[u(x_0^j,a) - v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a))] + \phi^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a)) - \phi^{\varepsilon_j}(x_0^j)\right\}.$$

Since $\phi \in \mathcal{C}^1(\mathbb{R}^d : \mathbb{R})$, the mean-value theorem implies that

$$\phi^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a)) - \phi^{\varepsilon_j}(x_0^j) = \varepsilon^j \langle \nabla\phi(x_0^j + \theta^j \varepsilon^j b(x_0^j,a)), b(x_0^j,a)\rangle$$

for every $j \in \mathbb{N}$ and some $\theta^j \in [0,1]$. Hence,

$$0 \leq \max_{a\in\mathcal{A}} \left\{(1-\lambda_{\varepsilon^j})[u(x_0^j,a) - v^{\varepsilon_j}(x_0^j + \varepsilon^j b(x_0^j,a))] + \varepsilon^j \langle \nabla\phi(x_0^j + \theta^j \varepsilon^j b(x_0^j,a)), b(x_0^j,a)\rangle\right\}$$

Dividing by ε^j and observing that

$$\frac{1}{\varepsilon^j}(1-\lambda_{\varepsilon^j}) = \frac{1}{\varepsilon^j}(1-e^{-r\varepsilon^j}) \to r$$

as $j \to \infty$, we conclude that

$$0 \leq -rv(x_0) + H(x, \nabla\phi(x_0)) \Leftrightarrow rv(x_0) - H(x_0, \nabla\phi(x_0)) \leq 0.$$

This shows that v satisfies the viscosity subsolution condition (16). The proof that v also satisfies the viscosity supersolution condition (17) is done, mutatis mutandis, in the same way, and is omitted. □

Proposition 1. *The sequence of strategies* $\{\alpha^\varepsilon\}_{\varepsilon\in(0,1)}$ *is a maximizing sequence:*

$$U(x,\alpha^\varepsilon) \to \sup_{\alpha\in\mathcal{S}^\#} U(x,\alpha) = v(x).$$

as $\varepsilon \to 0^+$.

Proof. For each $\varepsilon > 0$ we know that $v^\varepsilon(x) = U(x,\alpha^\varepsilon)$. By the arguments of the previous Lemma, the value function v^ε converges locally uniformly to the viscosity solution v. By uniqueness of solutions it follows that v is the value function of the optimal control problem (OC). □

This proposition shows that the strategies α^ε, obtained from the approximation procedure (15), guarantee the decision maker a suboptimal payoff which approximates the maximal payoff when ε is sufficiently small. In particular, for every $\delta > 0$ there exists a $\varepsilon_\delta > 0$ such that

$$U(x,\alpha^\varepsilon) \geq v(x) - \delta \quad \forall\varepsilon \in (0,\varepsilon_\delta).$$

A related result has been obtained by [11] in a different model setting.

3. The main result.

Having described the Markov decision process and its limit problem in detail, we come now to the main result of this paper.

Theorem 3.1. *Under assumptions 2-4 we have* $V^\varepsilon \to v$ *as* $\varepsilon \to 0$.

The main steps of the proof are as follows. First we define continuous-time interpolations of the controlled Markov chain and the action process which will provide the approximation of the controlled pairs for the limit problem. Consider the step-functions

$$\hat{X}^\varepsilon(t) = X_n^\varepsilon,\ \hat{A}^\varepsilon(t) = A_n^\varepsilon \quad \forall t \in [n\varepsilon,(n+1)\varepsilon), n \in \mathbb{N}_0. \tag{20}$$

$\hat{X}^\varepsilon$ is a random element of the space of right-continuous functions with left limits, denoted by $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$, and $\hat{A}^\varepsilon$ is a random element of $\mathcal{D}(\mathbb{R}_+ : \mathcal{A})$. Both these

spaces are complete separable metric spaces, when endowed with the Skorokhod metric (see e.g. [4]). In terms of these step functions, the utility to the decision maker under the strategy σ is given by

$$U^\varepsilon(x,\sigma) = E_x^\sigma\left[\int_0^\infty re^{-rt}u(\hat{X}^\varepsilon(t),\hat{A}^\varepsilon(t))\mathrm{d}t\right]$$
$$= E_x^\sigma\left[\sum_{n=0}^\infty(1-\lambda_\varepsilon)\lambda_\varepsilon^n u(\hat{X}^\varepsilon(n\varepsilon),\hat{A}^\varepsilon(n\varepsilon))\right].$$

In Section 5.1 we show that the sequence of interpolated processes $\{(\hat{X}^\varepsilon(t),\hat{A}^\varepsilon(t)), t \geq 0\}$ are tight in their respective function spaces. By Prohorov's theorem this guarantees that every sequence has a convergent subsequence. Using a suitable representation of the action process in terms of mixed actions (made precise in section 5), this relative compactness result allows me to prove that there exists a well defined limit process $(\bar{X},\nu)$, where $\bar{X}$ is a stochastic process taking values in the space of continuous functions $\mathcal{C}(\mathbb{R}^d:\mathbb{R})$ and ν is a stochastic process taking values in $\mathcal{S}$.[8] The two are coupled by the stochastic integral equation

$$\bar{X}(t) = x + \int_0^t b(\bar{X}(s),\nu(s))\mathrm{d}s. \tag{21}$$

For every element of the probability space variable, the pair $(\bar{X}(\omega),\nu(\omega))$ defines an admissible control pair for the deterministic optimal control problem (OC). Consequently, the strategy $\nu(\omega)$ is an element of the set $\mathcal{S}$, and therefore cannot give the decision maker a larger utility as he could obtain by solving the deterministic problem directly. This forms the basis for the proof that $\limsup_{\varepsilon\to 0} V^\varepsilon(x) \leq v(x)$. To show equality of the value functions, we need to show that also $\liminf_{\varepsilon\to 0} V^\varepsilon(x) \geq v(x)$. This will be shown by adapting the deterministic piecewise constant strategy α^ε constructed in eq. (15), and using this strategy as a strategy for the Markov decision process. The details of all these arguments are provided in Section 5.

4. **Conclusion.** We have focused in this paper on a standard stochastic optimal control problem, and studied the convergence of the value to the value of a related deterministic continuous-time problem. The key assumption which allowed us to prove this deterministic limit result is, of course, the "asymptotically vanishing" variance of the increments of the state process. Without this assumption a diffusion limit should be expected. Second, we have focused in this paper on the theoretically important case in which the decision maker has only finitely many actions among which he can choose. It should be not too difficult to adapt the arguments to general action spaces, imposing eventually additional technical assumptions on the problem data. A further interesting extension of the present analysis is to remove the boundedness assumption on the stage-utility function u. Allowing for unbounded utility functions is potentially important in applications to queuing networks and telecommunication networks. For an interesting study in the (more general) setting of continuous-time Markov games, see [14].

A more challenging question, and one which actually motivated me to look at this problem, is to extend the current result to stochastic games with imperfect public monitoring. In this extended setting the state process $\{X_n^\varepsilon\}_{n\in\mathbb{N}_0}$ is interpreted as

[8] In the control-theoretic literature this relaxation procedure is standard since the classical works of [29]. See section 5.1 for the precise definition of the relaxed representation of the action process.

the public signal process the players can observe, and public strategies are adapted processes with respect to the filtration generated by this process. The deterministic limit case is then only one of many scenarios one could study, and in fact might not be the most interesting one. A challenging problem is to prove a limit theorem where the limit signal process evolves according to a jump diffusion process. In the setting of repeated games with imperfect public monitoring where the limit dynamics is a continuous diffusion process (such as in [25]), we refer the reader to [28]. The analysis found there generalizes to stochastic games with imperfect public monitoring as in [16]. An open important question is to generalize the present study to jump-diffusions, as these models gain much importance in contract theory and mathematical finance. This is left for future research.

5. **Proofs.** Let $\{(X_n^\varepsilon, A_n^\varepsilon)\}_{n\in\mathbb{N}_0}$ be the data of the Markov decision process. For each $\varepsilon > 0$ the law of the action process is described by a feedback strategy σ^ε, being a stochastic kernel on $\mathcal{A}$ given $\mathbb{R}^d$. In the following, we assume that the initial condition of the state process is a given point $x \in \mathcal{X} \subset \mathbb{R}^d$. The pair process $\{(X_n^\varepsilon, A_n^\varepsilon)\}_{n\in\mathbb{N}_0}$ induces the law $P_x^{\sigma^\varepsilon} \equiv P_x^\varepsilon$ on $(\Omega, \mathcal{F})$, where $\Omega := (\mathbb{R}^d \times \mathcal{A})^{\mathbb{N}_0}$, and $\mathcal{F}$ denotes the sigma-algebra generated by the finite cylinder sets. The expectation operator with respect to this measure is denoted as E_x^ε. Our proof method is based on weak convergence arguments. Recall that a sequence of probability measures $\{P^\varepsilon\}_{\varepsilon>0}$ converges weakly to a limit measure P if

$$\lim_{\varepsilon\to 0} \int_\Omega f(\omega) \mathrm{d}P^\varepsilon(\omega) = \int_\Omega f(\omega) \mathrm{d}P(\omega)$$

for all bounded continuous random variables $f : \Omega \to \mathbb{R}$. We will use this notion of convergence to speak about limits of suitably interpolated versions of the data $\{(X_n^\varepsilon, A_n^\varepsilon)\}_{n\in\mathbb{N}_0}$. Once we have settled the convergence issue, we will be able to determine the limit of the family of value functions $\{V^\varepsilon\}_{\varepsilon\in(0,1)}$.

5.1. **Convergence of interpolated processes.** By definition, we have

$$X_n^\varepsilon(\omega) = x + \varepsilon \sum_{k=0}^{n-1} f_{k+1}^\varepsilon(X_k^\varepsilon(\omega), A_k^\varepsilon(\omega)).$$

Denote by $Z_{n+1}^\varepsilon := f_{n+1}^\varepsilon(X_n^\varepsilon, A_n^\varepsilon)$ the random (normalized) increment of the state process in stage n of the algorithm, and $\hat{Z}^\varepsilon(t) = Z_{n+1}^\varepsilon$ for $t \in [n\varepsilon, (n+1)\varepsilon)$ its corresponding step process. By construction, the interpolated process $\hat{Z}^\varepsilon$ is predictable with respect to the filtration $\{\hat{\mathcal{G}}_t^\varepsilon\}_{t\geq 0}$, where $\hat{\mathcal{G}}_t^\varepsilon := \sigma(\hat{X}^\varepsilon(s), \hat{A}^\varepsilon(s); s \leq t)$. Using the step processes $\hat{X}^\varepsilon$ and $\hat{A}^\varepsilon$ introduced in eq. (20), we can write the above recursive relation as an integral equation

$$\hat{X}^\varepsilon(t,\omega) = x + \int_0^{n\varepsilon} \hat{Z}^\varepsilon(s,\omega)\mathrm{d}s \qquad \forall t \in [n\varepsilon, (n+1)\varepsilon), n \geq 0.$$

Introducing the random variable

$$M_n^\varepsilon := \varepsilon \left(Z_{n+1}^\varepsilon - b^\varepsilon(X_n^\varepsilon, A_n^\varepsilon) \right),$$

we obtain, for $n\varepsilon \leq t < (n+1)\varepsilon$, the representation

$$\hat{X}^\varepsilon(t,\omega) = x + \int_0^{n\varepsilon} b^\varepsilon(\hat{X}^\varepsilon(s,\omega), \hat{A}^\varepsilon(s,\omega))\mathrm{d}s + \sum_{k=0}^{n-1} M_k^\varepsilon(\omega).$$

Given the definition of the function b^ε, the following Lemma is very simple.

Lemma 5.1. *The process* $\{\sum_{k=0}^{n} M_k^\varepsilon\}_{n\in\mathbb{N}_0}$ *is a martingale with respect to the filtration* $\mathcal{G}_n^\varepsilon = \sigma\left(X_0^\varepsilon, A_0^\varepsilon, \dots, X_n^\varepsilon, A_n^\varepsilon\right)$.

It follows that $\{||\sum_{k=0}^{n} M_k^\varepsilon||^2\}_{n\in\mathbb{N}_0}$ is a submartingale with respect to $\mathcal{G}_n^\varepsilon$. This translates in a straightforward way to the continuous-time submartingale $t \mapsto ||\hat{M}^\varepsilon(t)||^2$, where

$$\hat{M}^\varepsilon(t) := \int_0^{n\varepsilon} (\hat{Z}^\varepsilon(s) - b^\varepsilon(\hat{X}^\varepsilon(s), \hat{A}^\varepsilon(s))\mathrm{d}s \qquad \forall t \in [n\varepsilon, (n+1)\varepsilon).$$

An application of the submartingale inequality ([18], Theorem 3.8) gives the bound

$$P_x^\varepsilon\left[\sup_{0\le t\le T} ||\hat{M}^\varepsilon(t)||^2 \ge \lambda\right] \le \frac{1}{\lambda} E_x^\varepsilon ||\hat{M}^\varepsilon(T)||^2$$

for every $\lambda > 0$ and $T < \infty$. Using assumptions 4 and 1, the expectation on the right-hand side of this inequality is $o(1)$. Therefore,

$$\lim_{\varepsilon\to 0} P_x^\varepsilon\left[\sup_{0\le t\le T} ||\hat{M}^\varepsilon(t)||^2 \ge \lambda\right] = 0 \tag{22}$$

for every strategy σ and initial state $x \in \mathcal{X}$.

The action process $\{\hat{A}^\varepsilon(t,\omega), t \ge 0\}$ is, for each $\omega \in \Omega$, a deterministic right-continuous step function taking values in the discrete set $\mathcal{A}$. Given its discrete nature we cannot talk about function convergence in an ordinary sense, because of chattering. To speak about convergence of this process we interpret the pure action $\hat{A}^\varepsilon(t)$ as a behavior strategy taking values in the simplex $\Delta(\mathcal{A})$. To achieve this, we define the mixed action process by

$$\nu_a^\varepsilon(t,\omega) = \delta_{\hat{A}^\varepsilon(t,\omega)}(a) := \begin{cases} 1 & \text{if } \hat{A}^\varepsilon(t,\omega) = a, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

Clearly the random variable $\nu^\varepsilon(t,\omega)$ is an element of the mixed action simplex $\Delta(\mathcal{A})$ and the map $t \mapsto \nu^\varepsilon(t,\omega)$ is an element of the space of open-loop controls for the limit problem $\mathcal{S} = \{\alpha : \mathbb{R}_+ \to \Delta(\mathcal{A}) | \alpha(\cdot) \text{ measurable}\}$ for each fixed $\omega \in \Omega$.[9] Denote by $\mathcal{S}|_{[0,T]}$ the subspace of open loop controls restricted to the domain $[0,T]$. The embedding of the strategy process $t \mapsto \nu_\cdot^\varepsilon(t,\omega)$ allows us to work on a relatively compact space. Indeed, say that a sequence $\{\nu^j\}_{j\in\mathbb{N}} \subset \mathcal{S}|_{[0,T]}$ converges weak* to a limit $\nu \in \mathcal{S}|_{[0,T]}$ if for every integrable function $f : \mathcal{A} \times [0,T] \to \mathbb{R}$ we have

$$\lim_{j\to\infty} \int_0^T \sum_{a\in\mathcal{A}} f(a,t)\nu_a^j(t)\mathrm{d}t = \int_0^T \sum_{a\in\mathcal{A}} f(a,t)\nu_a(t)\mathrm{d}t. \tag{24}$$

The following result follows from general functional analytic facts (essentially Alaoglu's theorem).

Lemma 5.2. *For every* $T > 0$, *the set* $\mathcal{S}|_{[0,T]}$ *is sequentially compact in the weak** *topology. Hence every sequence* $\{\alpha^j\} \subset \mathcal{S}|_{[0,T]}$ *has a weak** *convergent subsequence with limit in* $\mathcal{S}|_{[0,T]}$.

Proof. See e.g. Lemma 5.1. in [5]. □

Defining a topology on $\mathcal{S}$, by saying that a sequence of open-loop controls $\{\alpha^j\}$ converges weak* to a limit α if and only if each restriction $\alpha^j|_{[0,T]}$ converges to the restriction $\alpha|_{[0,T]}$, shows that $\mathcal{S}$ is a weak* compact subset of $L^\infty(\mathbb{R}_+, \Delta(\mathcal{A}))$. To

[9] $t \mapsto \nu^\varepsilon(t,\omega)$ is a step function, thus trivially measurable.

summarize, for every $\omega \in \Omega$ and every subsequence $\{\nu^{\varepsilon_j}\}_{j\in\mathbb{N}}$ with $\varepsilon_j \to 0$ as $j \to \infty$, there exists a weak* converging subsequence with limit $\nu \in \mathcal{S}$. Therefore, we can state the following technical fact.

Lemma 5.3. *The family of open-loop strategies $\{\nu^\varepsilon\}_{\varepsilon\in(0,1)}$ is sequentially compact in $\mathcal{S}$ with respect to the weak* convergence. Therefore, for every subseqeuence of $\{\nu^\varepsilon\}_{\varepsilon\in(0,1)}$ there exists a subsubsequence $\{\nu^{\varepsilon_j}\}, \varepsilon_j \in (0,1), \varepsilon_j \to 0$ as $j \to \infty$, which converges weakly to a random element ν of the space of open loop strategies $\mathcal{S}$.*

Proof. Given a family of controls ν^ε as defined in (23), define the measure-valued random variable

$$m^\varepsilon(a,T) := \int_0^T \nu_a^\varepsilon(t)\mathrm{d}t.$$

This random variable takes values in the space of Borel measures on $\mathcal{A}\times[0,\infty)$ with the property that $m^\varepsilon(\mathcal{A},T) = T$ for all $T \geq 0$. Measures with this property are known as relaxed controls, and the space of deterministic relaxed controls is known to be tight in our setting. (see e.g. [19] and the references therein). Let us call the space of deterministic relaxed controls as $\mathcal{R}$. Every such measure can be disintegrated in the form $m(a,T) = \int_0^T m_t(a)\mathrm{d}t$, which, in our present context, leads to the identification

$$m_t^\varepsilon(a) = \nu_a^\varepsilon(t)$$

almost surely and almost everywhere with respect to Lebesgue measure. There exists a topology on this space under which it is a complete separable metric space (again see [19] for details). Thus, by Prohorov's theorem ([4]), from every subsequence of the family $\{m^\varepsilon\}$ we can extract a weakly converging subsequence. By [17], Theorem 14.16, this can be characterized in terms of "test functions", by requiring that for every bounded and continuous function $f : \mathcal{A}\times[0,\infty) \to \mathbb{R}_+$ with compact support $\mathcal{A}\times[0,T]$, we have

$$\int_{[0,T)} \sum_{a\in\mathcal{A}} f(a,t)m_t^\varepsilon(a)\mathrm{d}t \to \int_{[0,T)} \sum_{a\in\mathcal{A}} f(a,t)m_t(a)\mathrm{d}t \qquad \text{in distribution.}$$

Comparing this definition of weak convergence of these measures-valued random variables with eq. (24) gives the result. □

We now finalize the proof of the convergence of the interpolated sample paths by showing that the family of $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$ valued random variables $\{\hat{X}^\varepsilon(t); t \geq 0\}_{\varepsilon\in(0,1)}$ is relatively compact. This relative compactness allows us to focus on subsequences which convergence in distribution to a random element $\bar{X}$, which is shown to have almost surely sample paths in $\mathcal{C}(\mathbb{R}_+ : \mathbb{R}^d)$. This fact is established in Lemma 5.4 and Theorem 5.5, respectively.

Lemma 5.4. *The family of $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$-valued processes $\{\hat{X}^\varepsilon\}_{\varepsilon\in(0,1)}$ is relatively compact, i.e. for every subsequence of $\{\hat{X}^\varepsilon\}$, there exists a subsubsequence $\{\hat{X}^{\varepsilon_j}\}_{j\in\mathbb{N}}$ converging in distribution to a $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$-valued random variable $\bar{X}$.*

Proof. For $\mathbf{x} \in \mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$ define the modulus of continuity by

$$w(\mathbf{x},\delta,T) = \inf_{\{t_i\}} \max_{1\leq i\leq n} \sup_{s,t\in[t_{i-1},t_i)} ||\mathbf{x}(s) - \mathbf{x}(t)||_\infty$$

where the sequence $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \ldots < t_{n-1} < T \leq t_n$ with $\min_{1\leq i\leq n}(t_i - t_{i-1}) > \delta$. For every fixed $\varepsilon > 0$ pick $\delta = \frac{\varepsilon}{2}$.

Then the sequence $t_i = i\varepsilon, i = 0, 1, \ldots, \lceil T/\varepsilon \rceil$ is admissible, and we see that

$$\max_i \max_{s,t\in[(i-1)\varepsilon, i\varepsilon)} ||\hat{X}^\varepsilon(t) - \hat{X}^\varepsilon(s)||_\infty = \max_{1\leq i\leq n} ||\varepsilon f_i^\varepsilon(X_{i-1}^\varepsilon, A_{i-1}^\varepsilon)||_\infty.$$

By Assumption 1, the random vector fields $\{f_n^\varepsilon\}_n$ take values in the compact set $\mathcal{K}$ and can therefore be uniformly embedded in a compact cube $\Gamma \subset \mathbb{R}^d$. It follows that for every $\omega \in \Omega, \varepsilon > 0$ and $T > 0$ we have

$$\lim_{\delta\to 0} w(\hat{X}^\varepsilon(\omega), \delta, T) = 0.$$

Using Assumption 1 once again, we see that for every $T > 0$ the sample paths of the step process $\hat{X}^\varepsilon$ are contained in a compact cube $\Gamma_T \subset \mathbb{R}^d$ with probability 1. Theorem 7.2 in [8] states that under these two conditions the family of processes $\{\hat{X}^\varepsilon\}_{\varepsilon\in(0,1)}$ is relatively compact in $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$. Under the Skorokhod topology, this space is complete and separable. By Prohorov's theorem, a family of probability measures on $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$ is relatively compact if and only if they are tight. Since convergence in distribution means that the induced laws of the processes $\hat{X}^{\varepsilon_j}$ on the space $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$ converge weakly in the sense of measures, the tightness of the processes means that the laws are tight. This completes the proof of the Lemma. □

Lemma 5.4 implies that every subsequence of the sequence of interpolated process $\hat{X}^\varepsilon$ has a further subsubsequence which converges in distribution to a random variable $\bar{X}$ taking values in the space $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$, and the same holds true for the sequence of controls ν^ε. The next Lemma characterizes the structure of the limit process $\bar{X}$ given a weakly converging subsequence of the pair $(\hat{X}^\varepsilon, \nu^\varepsilon)$.

Theorem 5.5. *Let $(\hat{X}^\varepsilon, \nu^\varepsilon)$ be a sequence of interpolated process obtained from the discrete Markov decision process which converges in distribution to a $\mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$-valued random variable $\bar{X}$, and a random element of the set $\mathcal{S}$, respectively. Then $\bar{X}$ is almost surely a random element of $\mathcal{C}(\mathbb{R}_+ : \mathbb{R}^d)$. Moreover, it has almost surely absolutely continuous sample paths, whose derivative with respect to Lebesgue measure is almost surely given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{X}(t) = b(\bar{X}(t), \nu(t)).$$

Proof. For every $\mathbf{x} \in \mathcal{D}(\mathbb{R}_+ : \mathbb{R}^d)$ define

$$J(\mathbf{x}) := \int_0^\infty e^{-s} \min\{J(\mathbf{x}, s), 1\}\mathrm{d}s,$$

with

$$J(\mathbf{x}, s) := \sup_{0\leq t\leq s} ||\mathbf{x}(t) - \mathbf{x}(t-)||_\infty,$$

and $\mathbf{x}(t-) \equiv \lim_{\tau\to t^-} \mathbf{x}(\tau)$. Then, for every $s > 0$, it follows that

$$J(\hat{X}^\varepsilon(\omega), s) \leq \varepsilon \sup_{k\in\mathcal{K}} ||k||_\infty \qquad \forall \omega \in \Omega,$$

and therefore

$$J(\hat{X}^\varepsilon(\omega)) \leq \varepsilon \sup_{k\in\mathcal{K}} ||k||_\infty \to 0 \text{ for } \varepsilon \to 0$$

for every $\omega \in \Omega$. Hence, $J(\hat{X}^\varepsilon)$ converse to the 0 process in distribution (and hence in probability). By hypothesis, $\hat{X}^\varepsilon \to \bar{X}$ in distribution. Theorem 10.2 in [8] implies that $\bar{X}$ is almost surely a random process in $\mathcal{C}(\mathbb{R}_+ : \mathbb{R}^d)$, and the convergence to

this process is uniform on compact intervals. We will now characterize the sample paths of this process $\bar{X}$. Define the process

$$\begin{aligned} Y^\varepsilon(t,\omega) :&= x + \int_0^t b^\varepsilon(\hat{X}^\varepsilon(s,\omega), \hat{A}^\varepsilon(s,\omega))\mathrm{d}s \\ &= x + \int_0^t b^\varepsilon(\hat{X}^\varepsilon(s,\omega), \nu^\varepsilon(s,\omega))\mathrm{d}s. \end{aligned}$$

Here we have extended the domain of the drift b to $\mathbb{R}^d \times \Delta(\mathcal{A})$ in the obvious way. Then, for every $T > 0$ and $t \in [0,T]$, there exists a $n \in \mathbb{N}_0$ such that $n\varepsilon \leq t < (n+1)\varepsilon$. Then, we see that

$$\begin{aligned} \left\| Y^\varepsilon(t) - \hat{X}^\varepsilon(t) \right\|^2 &\leq 2 \left\| \int_{n\varepsilon}^t b^\varepsilon(\hat{X}^\varepsilon(s), \nu^\varepsilon(s))\mathrm{d}s \right\|^2 + 2 \left\| \hat{M}^\varepsilon(t) \right\|^2 \\ &\leq 2C\varepsilon + 2 \left\| \hat{M}^\varepsilon(t) \right\|^2 \end{aligned}$$

for some constant C, which can be chosen independently of t, T and ε, by (7). Hence, for $\lambda > 2C\varepsilon$, which can be made arbitrary small by making ε small, we conclude from equation (22) that

$$\begin{aligned} P_x^\varepsilon \left(\sup_{0\leq t\leq T} \left\| Y^\varepsilon(t) - \hat{X}_t^\varepsilon \right\|^2 \geq \lambda \right) &\leq P_x^\varepsilon \left(\sup_{0\leq t\leq T} \left\| \hat{M}^\varepsilon(t) \right\|^2 \geq \frac{\lambda}{2} - C\varepsilon \right) \\ &\leq \frac{1}{\frac{\lambda}{2} - C\varepsilon} E_x^\varepsilon \left\| \hat{M}^\varepsilon(T) \right\|^2 \to 0 \text{ as } \varepsilon \to 0. \end{aligned}$$

Hence,

$$\sup_{0\leq t\leq T} \left\| Y^\varepsilon(t) - \hat{X}^\varepsilon(t) \right\| \to 0 \text{ as } \varepsilon \to 0$$

in probability. By [4], Theorem 1.3.1, this implies that $Y^\varepsilon \to \bar{X}$ in distribution.

By assumption 2 the drift converges locally uniformly to a Lipschitz continuous function b. Together with the continuous mapping theorem ([17], Theorem 3.27), this implies that

$$\lim_{\varepsilon\to 0} \int_0^t b^\varepsilon(\hat{X}^\varepsilon(s), \nu^\varepsilon(s))\mathrm{d}s = \int_0^t b(\bar{X}(s), \nu(s))\mathrm{d}s$$

for every $t > 0$ and in distribution. Hence, $Y^\varepsilon \to \bar{X}$ as $\varepsilon \to 0$ in distribution, with

$$\bar{X}(t) = x + \int_0^t b(\bar{X}(s), \nu(s))\mathrm{d}s. \tag{25}$$

The w.p.1 absolute continuity of the sample paths of the process $\bar{X}$ is now immediate. □

5.2. **Convergence of values.** We now complete the proof of the main result we show that $V^\varepsilon \to V$ for a compact set of initial conditions $\mathcal{X} \subset \mathbb{R}^d$. Lemma 5.3 and Lemma 5.4 implies that from every subsequence of $\{(\hat{X}^\varepsilon, \nu^\varepsilon)\}_{\varepsilon\in(0,1)}$, there exists a further subsequence, still denoted by $\{(\hat{X}^\varepsilon, \nu^\varepsilon)\}_{\varepsilon\in(0,1)}$ with some abuse of notation, which converges in distribution to a pair $(\bar{X}, \nu)$. The limit ν is almost surely a random element of the space of open loop control, and $\bar{X}$ is a random element of the set of continuous functions, absolutely continuous w.r.t. Lebesgue measure. We will now show the following Lemma.

Lemma 5.6. $\limsup_{\varepsilon\to 0} V^\varepsilon(x) \leq v(x)$.

Proof. Let $(\hat{X}^\varepsilon, \nu^\varepsilon)$ be interpolated date obtained by solving the discrete-time Markov decision process, and using the mesh size ε, and fix the initial condition x. By the above said, we can find a subsequence converging in distribution to the pair $(\bar{X}, \nu)$. By the Skorokhod representation theorem ([4]) there exists a probability space $(\bar{\Omega}, \bar{\mathcal{G}}, \bar{\mathsf{P}}_x)$, on which we can define random variables $(\bar{X}^\varepsilon, \bar{\nu}^\varepsilon)$, which have the same distribution as the pair $(\hat{X}^\varepsilon, \nu^\varepsilon)$, but which converge almost surely to the processes $(\bar{X}, \nu)$. Hence, there is a set $N \in \bar{\mathcal{G}}$ with $\bar{\mathsf{P}}_x(N) = 0$ such that for every $\omega \in \bar{\Omega} \setminus N$ the function $\bar{X}^\varepsilon(\omega)$ converges in the Skorokhod metric to the limit $\bar{X}(\omega)$ and $\bar{\nu}^\varepsilon(\omega)$ converges weak* to a limit $\nu(\omega)$. The random variables $\bar{X}, \nu$ have the properties described in Theorem 5.5. We will henceforth not distinguish between the random elements $(\hat{X}^\varepsilon, \nu^\varepsilon)$ and its Skorokhod representation $(\bar{X}^\varepsilon, \bar{\nu}^\varepsilon)$, as they describe the same processes in distribution.

For each $\varepsilon \in (0,1)$ we have

$$V^\varepsilon(x) := \bar{\mathsf{E}}_x \left[\int_0^\infty re^{-rt} u(\hat{X}^\varepsilon(t), \hat{A}^\varepsilon(t)) \mathrm{d}t \right] = \bar{\mathsf{E}}_x \left[\int_0^\infty re^{-rt} u(\hat{X}^\varepsilon(t), \nu^\varepsilon(t)) \mathrm{d}t \right],$$

where $\bar{\mathsf{E}}_x$ represents the expectation operator with respect to the probability measure $\bar{\mathsf{P}}_x$. Define the function $g : \mathcal{D}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{S} \to \mathbb{R}$ by

$$g(\phi, \alpha) := \int_0^\infty re^{-rt} u(\phi(t), \alpha(t)) \mathrm{d}t.$$

Then, for each $\omega \in \bar{\Omega}$, the number $g(\hat{X}^\varepsilon(\omega), \nu^\varepsilon(\omega))$ is the payoff of the decision maker under the control pair $(\hat{X}^\varepsilon, \nu^\varepsilon)$. Since g is continuous at the limit $(\bar{X}, \nu)$, it follows from the continuous mapping theorem ([17], Theorem 3.27) that, for each $\omega \in \bar{\Omega} \setminus N$

$$\begin{aligned} \lim_{\varepsilon \to 0} g(\hat{X}^\varepsilon(\omega), \nu^\varepsilon(\omega)) &= g(\bar{X}(\omega), \nu(\omega)) \\ &= \int_0^\infty re^{-rt} u(\bar{X}(t, \omega), \nu(t, \omega)) \mathrm{d}t \\ &= U(x, \nu(\omega)) \\ &\leq v(x). \end{aligned} \tag{26}$$

Since the map $\omega \mapsto g(\hat{X}^\varepsilon(\omega), \nu^\varepsilon(\omega))$ is bounded (Assumption 3 and Lemma 2.1) it follows from Lebesgue's dominated convergence theorem that

$$\begin{aligned} \limsup_{\varepsilon \to 0} V^\varepsilon(x) &= \lim_{\varepsilon \to 0} \bar{\mathsf{E}}_x \left[\int_0^\infty re^{-rt} u(\hat{X}^\varepsilon(t), \nu^\varepsilon(t)) \mathrm{d}t \right] \\ &= \bar{\mathsf{E}}_x \left[\int_0^\infty re^{-rt} u(\bar{X}(t), \nu(t)) \mathrm{d}t \right] \\ &\leq v(x). \end{aligned}$$

The last inequality follows from the relation (26), and completes the proof of the Lemma. □

To finish the proof of Theorem 3.1 it remains to show the validity of the following result.

Lemma 5.7. $\liminf_{\varepsilon \to 0^+} V^\varepsilon(x) \geq v(x)$.

Proof. We make use of the explicit approximation procedure, described in section 2.2. For each $\varepsilon > 0$ let α^ε denote the piecewise constant control, taking values in

$\mathcal{A}$, constructed in eq. (15). From Proposition 1, we know that for every $\delta > 0$ there exists $\varepsilon_\delta > 0$ sufficiently small so that

$$U(x, \alpha^\varepsilon) \geq v(x) - \delta \quad \forall \varepsilon \in (0, \varepsilon_\delta).$$

We adapt this strategy for the controlled Markov chain as follows. For each $n \in \mathbb{N}_0$ we define a deterministic action process $A_n^\varepsilon := \alpha^\varepsilon(n\varepsilon)$. Hence, independent of the probability space variable ω, we always implement the same action process $\{A_n^\varepsilon\}_{n\in\mathbb{N}}$. Denote by P_x^ε the resulting probability measure, and E_x^ε the corresponding expectations operator. The so constructed strategy is admissible and gives guarantees the decision maker the payoff

$$E_x^\varepsilon \left[\sum_{n=0}^{\infty} (1 - \lambda_\varepsilon) \lambda_\varepsilon^n u(X_n^\varepsilon, A_n^\varepsilon) \right] \leq V^\varepsilon(x^\varepsilon).$$

With a slight abuse of notation, denote the left-hand side of this equation by $U^\varepsilon(x, \alpha^\varepsilon)$. Set $\hat{X}^\varepsilon(t) = X_n^\varepsilon$ and $\nu^\varepsilon(t) = \delta_{A_n^\varepsilon}$ for each $t \in [n\varepsilon, (n+1)\varepsilon)$. It follows from the sequential compactness of relaxed controls (Lemma 5.3) that, passing if necessary to a subsequence, the deterministic limit

$$\lim_{\varepsilon\to 0} \nu^\varepsilon = \lim_{\varepsilon\to 0} \delta_{\alpha^\varepsilon(\cdot)} = \nu$$

exists and defines an open-loop control in $\mathcal{S}$ (see also [5] for a related argument). Along the same subsequence, it follows from arguments used in section 5.1 that $\hat{X}^\varepsilon \Rightarrow \bar{X}$ in distribution, where

$$\bar{X}(t) = x + \int_0^t b(\bar{X}(s), \nu(s)) \mathrm{d}s.$$

$\bar{X}$ is a deterministic process which, by uniqueness of solutions to the controlled ODE $\dot{y} = b(y, \nu(t))$ ([2], Theorem III.5.5), corresponds to the limit process of the controlled pair $(y_x^\varepsilon, \alpha^\varepsilon)$. Therefore, Proposition 1 implies that

$$\liminf_{\varepsilon\to 0} U^\varepsilon(x, \alpha^\varepsilon) = \int_0^\infty r e^{-rt} u(\bar{X}(t), \nu(t)) \mathrm{d}t = v(x).$$

As $V^\varepsilon(x) \geq U^\varepsilon(x, \alpha^\varepsilon)$ for every ε, we conclude that

$$\liminf_{\varepsilon\to 0} V^\varepsilon(x) \geq \liminf_{\varepsilon\to 0} U^\varepsilon(x, \alpha^\varepsilon) = v(x).$$

□

Combining Lemma 5.6 with Lemma 5.7 completes the proof of Theorem 3.1.

Acknowledgments. This paper was written while I was visiting Nuffield College at the University of Oxford, and finished at the Department of Mathematics of the University of Vienna. I thank both institutions for their hospitality and, in particular, my sponsor Peyton Young, for his support. I also would like to thank Frank Riedel, Giorgio Ferrari, Jan-Henrik Steg, Immanuel Bomze, two anonymous referees, the editor in charge and Bill Sandholm for useful comments.

REFERENCES

[1] R. J. Aumann and M. Maschler, *Repeated Games with Incomplete Information*, MIT Press, Cambridge, MA, 1995.

[2] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser - Systems & Control: Foundations & Applications, 1997.

[3] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, 1978.

[4] P. Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, 1999.

[5] I. Capuzzo-Dolcetta and H. Ishii, Approximate solutions of the Bellman equation of deterministic control theory, *Applied Mathematics & Optimization*, **11** (1984), 161–181.

[6] P. Cardaliaguet, C. Rainer, D. Rosenberg and N. Vieille, Markov games with frequent actions and incomplete information, 2013. arXiv:1307.3365v1 [math.OC].

[7] N. El Karoui, D. Nguyen and M. Jeanblanc-Picqué, Compactification methods in the control of degenerate diffusions: Existence of an optimal control, *Stochastics*, **20** (1987), 169–219.

[8] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York, 1986.

[9] M. Falcone, A numerical approach to the infinite horizon problem of deterministic control theory, *Applied Mathematics & Optimization*, **15** (1987), 1–13.

[10] W. H. Fleming, The convergence problem for differential games, *Journal of Mathematical Analysis and Applications*, **3** (1961), 102–116.

[11] N. Gast, B. Gaujal and J.-Y. Le Boudec, Mean field for markov decision processes: From discrete to continuous optimization, *IEEE Transactions on Automatic Control*, **57** (2012), 2266–2280.

[12] F. Gensbittel, *Continuous-time Limit of Dynamic Games with Incomplete Information and a More Informed Player,* hal-00910970, version 1, 2013.

[13] R. Gonzalez and E. Rofman, On deterministic control problems: An approximation procedure for the optimal cost I. the stationary problem, *SIAM Journal on Control and Optimization*, **23** (1985), 242–266.

[14] X. Guo and O. Hernández-Lerma, Nonzero-sum games for continuous-time markov chains with unbounded discounted payoffs, *Journal of Applied Probability*, **42** (2005), 303–320.

[15] O. Hernández-Lerma and J. B. Laserre, *Discrete-Time Markov Control Processes: Basic Optimality criteria*, Springer-Verlag, 1996.

[16] J. Hörner, T. Sugaya, S. Takahashi and N. Vieille, Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem, *Econometrica*, **79** (2011), 1277–1318.

[17] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York [u.a.], 2nd edition, 2002.

[18] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, 2nd edition, 2000.

[19] H. J. Kushner, *Weak convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser - Systems & Control: Foundations & Applications, Boston, MA, 1990.

[20] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, New York, 2nd edition, 2001.

[21] A. Neyman, Stochastic games with short-stage duration, *Dynamic Games and Applications*, **3** (2013), 236–278.

[22] A. S. Nowak and T. E. S. Raghavan, Existence of stationary correlated equilibria with symmetric information for discounted stochastic games, *Mathematics of Operations Research*, **17** (1992), 519–526.

[23] W. H. Sandholm and M. Staudigl, Stochastic stability in the small noise double limit, I: Theory, Unpublished manuscript, University of Wisconsin and Bielefeld University, 2014.

[24] W. H. Sandholm and M. Staudigl, Stochastic stability in the small noise double limit, II: The logit model, Unpublished manuscript, University of Wisconsin and Bielefeld University, 2014.

[25] Y. Sannikov, Games with imperfectly observable actions in continuous time, *Econometrica*, **75** (2007), 1285–1329.

[26] S. Soulaimani, M. Quincampoix and S. Sorin, Repeated games and qualitative differential games: Approachability and comparison of strategies, *SIAM Journal on Control and Optimization*, **48** (2009), 2461–2479.

[27] M. Staudigl, Stochastic stability in asymmetric binary choice coordination games, *Games and Economic Behavior*, **75** (2012), 372–401.

[28] M. Staudigl and J.-H. Steg, On repeated games with imperfect public monitoring: From discrete to continuous time, Bielefeld University, unpublished manuscript, 2014.

[29] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York-London, 1972.

Received January 2014; revised June 2014.

E-mail address: mathias.staudigl@uni-bielefeld.de