Abstract. We analyse a version of the policy iteration algorithm for the discounted infinite-horizon problem for controlled multidimensional diffusion processes, where both the drift and the diffusion coefficient can be controlled. We prove that, under assumptions on the problem data, the payoffs generated by the algorithm converge monotonically to the value function and an accumulation point of the sequence of policies is an optimal policy. The algorithm is stated and analysed in continuous time and state, with discretisation featuring neither in theorems nor the proofs. A key technical tool used to show that the algorithm is well-defined is the mirror coupling of Lindvall and Rogers.

1. Introduction

Howard's policy iteration algorithm (PIA) \cite{13} is a well-known tool for solving control problems for Markov decision processes (see e.g. \cite{4} for a recent survey of approximate policy iteration methods for finite state, discrete time, stochastic dynamic programming problems). The algorithm can be recast for general state-spaces continuous-time control problems, where allowed actions can be chosen from a Polish space. In this paper we investigate the convergence of the PIA for an infinite horizon discounted cost problem in the context of controlled diffusion processes in $\mathbb{R}^d$, where the control takes place in an arbitrary compact metric space. The main aim of the paper is to analyse the convergence of a sequence of policies and the corresponding payoff functions produced by the PIA under assumptions that are at least in principle verifiable in terms of the model data.

Our control setting is similar to that of \cite{1}, where an ergodic cost criterion was considered. The main differences, beyond the cost criterion, are as follows: (1) we allow the controller to modulate the drift as well as the diffusion coefficient; (2) we consider a generalised version of the PIA where an arbitrary scaling function can be used to simplify the algorithm; (3) we investigate the convergence not only of payoffs but also of policies produced by the PIA; (4) we work with Markov policies that are defined for every $x \in \mathbb{R}^d$, not almost everywhere, and obtain a locally uniform convergence of a subsequence to the optimal policy.

This last point in this list is particularly important in our setting, as our aim is to design and analyse an algorithm that can in principle at least be used in to construct an optimal policy. This requirement forces us to work in the context of classical solutions of PDEs, rather than...
relying on generalised solutions of the Poisson equation in appropriate Sobolev spaces. The latter approach, followed in [1], is based on the fact that there exists a canonical solution to the Poisson equation in $W^{2,p}_{loc}(\mathbb{R}^d)$, see [2]. The analysis of the PIA can than be performed using Krylov’s extension [15] of Itô’s formula to functions in the Sobolev space $W^{2,p}_{loc}(\mathbb{R}^d)$.

Our method for solving the Poisson equation in the classical sense is based on the coupling of Lindvall and Rogers [18]. This coupling plays a crucial role in the proof of Proposition 1, guaranteeing that a payoff function for a locally Lipschitz Markov policy is the classical solution of the corresponding Poisson equation. The main technical contribution of the paper is the result in Lemma 7 which shows that the mirror coupling from [18] is successful with very high probability, when the diffusion processes are started sufficiently close together. Interestingly, the condition in [18] for the coupling to be successful is not satisfied in our setting in general. The proof of Lemma 7 is based on a local path-wise comparison of a time-change of the distance between the coupled diffusions and an appropriately chosen Bessel process.

The convergence of the policies and payoffs in the PIA is obtained in several steps. First we show that the PIA always improves the payoff. Then we prove, using a “diagonalisation” argument and an Arzela-Ascoli type result, that a subsequence of the policies produced by the PIA converges locally uniformly to a locally Lipschitz limiting policy. The final stage of the argument shows that this limiting policy is indeed an optimal policy with payoff equal to the pointwise limit of the payoffs produced by the PIA. These steps are detailed in Section 2 and proved in Section 5.

The literature on the PIA for Markov decision processes in various settings is extensive (see e.g. [7], [11], [12], [17], [20], [16], [19], [22], [23] and [24] and the references therein). Our approach is in some sense closest to the continuous time analysis in general state spaces in [7], where the convergence of the subsequence of the policies is established in the case of finite action space. In [24] this restriction is removed, but the controlled processes considered do not include diffusions. A recent application of the PIA to impulse control in continuous time is given in [3].

Finally, as mentioned above, we observe that the PIA can be generalised by multiplying the expression to be minimised by an arbitrary positive scaling function that can depend both on the state and the control action (see (gPIA) below). A choice of the scaling function clearly influences the sequence of policies produced by the algorithm. In particular, in the one-dimensional case, the scaling function can be used to eliminate the second derivative of the payoff from the algorithm. This idea is described in Section 3. A numerical example of the PIA is reported in Section 4. In this examples at least, the convergence of the PIA is very fast as the algorithm finds an optimal policy in fewer than six iterations.

2. Multidimensional controlled diffusion processes

Let $(A, d_A)$ be a compact metric space of available control actions and, for some $d, m \in \mathbb{N}$, let $\sigma : \mathbb{R}^d \times A \to \mathbb{R}^{d \times m}$ and $\mu : \mathbb{R}^d \times A \to \mathbb{R}^d$ be measurable functions. Let the set $\mathcal{A}(x)$ of admissible policies at $x \in \mathbb{R}^d$ consist of processes $\Pi = (\Pi_t)_{t \geq 0}$ satisfying the following: $\Pi$ is $A$-valued, adapted to a filtration $(\mathcal{F}_t)_{t \geq 0}$ and there exists an $(\mathcal{F}_t)$-adapted process $X^{\Pi,x} = (X_t^{\Pi,x})_{t \geq 0}$ satisfying the SDE

$$X_t^{\Pi,x} = x + \int_0^t \sigma (X_s^{\Pi,x}, \Pi_s) \, dB_s + \int_0^t \mu (X_s^{\Pi,x}, \Pi_s) \, ds, \quad t \geq 0,$$
where $B = (B_t)_{t \geq 0}$ is an $m$-dimensional $(\mathcal{F}_t)$-Brownian motion. Note that the filtration $(\mathcal{F}_t)_{t \geq 0}$ (and indeed the entire filtered probability space) depends on the policy $\Pi$ in $\mathcal{A}(x)$. Pick measurable functions $\alpha : \mathbb{R}^d \times A \to \mathbb{R}$ and $f : \mathbb{R}^d \times A \to \mathbb{R}$. For any $x \in \mathbb{R}^d$ and $\Pi \in \mathcal{A}(x)$, define the payoff of the policy $\Pi$ by
\[
V_\Pi(x) := \mathbb{E} \left( \int_0^\infty e^{-\int_0^t \alpha(X_{\pi,x}^t,\Pi_s)ds} f \left( X_{\pi,x}^t, \Pi_t \right) dt \right).
\]

**Control problem.** Construct the value function $V$, defined by
\[
V(x) := \inf_{\Pi \in \mathcal{A}(x)} V_\Pi(x), \quad x \in \mathbb{R}^d,
\]
and, if it exists, an optimal control $\{\Pi^x \in \mathcal{A}(x) : x \in \mathbb{R}^d\}$, satisfying $V(x) = V_{\Pi^x}(x)$.

Note first that the problem is specified completely by the deterministic data $\sigma$, $\mu$, $\alpha$, and $f$. In order to define an algorithm that solves this problem, the functions $\sigma$, $\mu$, $\alpha$, $f$ are assumed to satisfy Assumption 1 below throughout this section.

**Assumption 1.** The functions $\sigma$, $\mu$, $\alpha$ and $f$ are bounded, and Lipschitz on compacts in $\mathbb{R}^d \times A$, i.e. for every compact set $K \subseteq \mathbb{R}^d \times A$ there exists a constant $C_K > 0$ such that
\[
\|h(x,y) - h(y,r)\| \leq C_K (\|x - y\|^2 + d_A(p,r))^\frac{1}{2}
\]
holds for every $(x,y), (y,r) \in K$, and $h \in \{\sigma, \mu, \alpha, f\}$. In addition, $\alpha(x,p) > \epsilon_0 > 0$ for all $(x,p) \in \mathbb{R}^d \times A$, and there exists $\lambda > 0$ such that
\[
\langle \sigma(x,p)\sigma(x,p)^T, v, v \rangle \geq \lambda \|v\|^2 \quad \text{for all } x \in \mathbb{R}^d, p \in A, v \in \mathbb{R}^d.
\]

**Remark 1.** Here, and throughout the paper, $\| \cdot \|$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean norm and inner product respectively. The norm $\|M\| = \sup\{\|Mv\|/\|v\| : v \in \mathbb{R}^m \setminus \{0\}\} = \sqrt{\lambda_{\max}(MM^T)}$, for a matrix $M \in \mathbb{R}^{d \times m}$, is used in $(2)$ for $h = \sigma$, where $\lambda_{\max}(MM^T)$ is the largest eigenvalue of the non-negative definite matrix $MM^T \in \mathbb{R}^{d \times d}$ and $M^T \in \mathbb{R}^{m \times d}$ denotes the transpose of $M$.

**Remark 2.** The uniform ellipticity condition in $(3)$ is the multidimensional analogue of the volatility being bounded away from 0. Hence, for all $x \in \mathbb{R}^d$ and $p \in A$, the smallest eigenvalue of $\sigma(x,p)\sigma(x,p)^T \in \mathbb{R}^{d \times d}$ is at least of size $\lambda$ and, in particular, $m \geq d$ (cf. Remark 3 below).

A measurable function $\pi : \mathbb{R}^d \to A$ is a Markov policy (or synonymously Markov control) if for $x \in \mathbb{R}^d$ there exists an $\mathbb{R}^d$-valued process $X_{\pi,x} = (X_{\pi,x}^t)_{t \geq 0}$ that satisfies the following SDE:
\[
X_{\pi,x}^t = x + \int_0^t \sigma \left( X_{\pi,x}^s, \pi(X_{\pi,x}^s) \right) dB_s + \int_0^t \mu \left( X_{\pi,x}^s, \pi(X_{\pi,x}^s) \right) ds, \quad t \geq 0.
\]

Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration with respect to which $(X_{\pi,x}, B)$ is $(\mathcal{F}_t)$-adapted and $B$ is an $(\mathcal{F}_t)$-Brownian motion. Such a filtration $(\mathcal{F}_t)_{t \geq 0}$ exists by the definition of a solution of SDE $(4)$, see e.g. [12] Def. 5.3.1, p. 300]. Then $(\mathcal{F}_t)_{t \geq 0}$ can be taken to be the filtration in the definition of the policy $\pi(X_{\pi,x}) \in \mathcal{A}(x)$. Moreover, without loss of generality, we may assume that $(\mathcal{F}_t)_{t \geq 0}$ satisfies the usual conditions.

For any function $\pi : \mathbb{R}^d \to A$ that is Lipschitz on compacts (i.e. locally Lipschitz) in $\mathbb{R}^d$, Assumption 1 implies (see e.g. [5] p. 45) and the references therein) that the SDE in $(4)$ has a unique, strong non-exploding solution, thus making $\pi$ a Markov policy. The payoff function of a locally Lipschitz Markov policy is a classical solution of a linear PDE, a fact key for (gPIA) to work. To state this formally, recall that $h : \mathbb{R}^d \to \mathbb{R}^k$ (for any $k \in \mathbb{N}$) is $(1/2)$-Hölder continuous
on a compact $D \subset \mathbb{R}^d$ if $\|h(x') - h(x'')\| \leq K_D \|x' - x''\|^{1/2}$ holds for some constant $K_D > 0$ and all $x', x'' \in D$. Streamline the notation for Markov policies as follows:

\begin{align}
V_\pi(\cdot) := V_\pi(X_\pi(\cdot)) \quad \text{and} \quad L_\pi h := \frac{1}{2} \text{Tr} \left( \sigma^T h \sigma \right) + \mu^T \nabla h \quad \text{for} \ h \in C^2(\mathbb{R}^d),
\sigma_\pi(\cdot) := \sigma(\cdot, \pi(\cdot)) \quad \mu_\pi(\cdot) := \mu(\cdot, \pi(\cdot)) \quad \alpha_\pi(\cdot) := \alpha(\cdot, \pi(\cdot)),
\end{align}

where $H h$ and $\nabla h$ are the Hessian and gradient of $h$, respectively, and $\text{Tr}(M)$ denotes the trace of any matrix $M \in \mathbb{R}^{m \times m}$.

**Proposition 1.** Let Assumption [1] hold. For a locally Lipschitz Markov policy $\pi$ we have: $V_\pi \in C^2(\mathbb{R}^d)$ is the unique bounded solution of the Poisson equation $L_\pi V_\pi - \alpha_\pi V_\pi + f_\pi = 0$ and $HV_\pi$ is $(1/2)$-Hölder on compacts in $\mathbb{R}^d$.

**Remark 3.** The proof of Proposition [1] given in Section 5.2 depends crucially on the coupling in Lindvall and Rogers [18] of two $d$-dimensional diffusions via a coupling of the corresponding $d$-dimensional driving Brownian motions (see Lemma [7] below). Moreover, it is easy to see that the probability of coupling could in general be zero if the dimension $m$ of the Brownian noise is strictly less than $d$. In this case the controlled diffusions in $\mathbb{R}^d$, started at distinct points, could remain forever on disjoint submanifolds of positive co-dimension in $\mathbb{R}^d$ for any choice of control. In particular, Proposition [1] fails in the case $m < d$, as demonstrated by Example [1] below.

**Example 1.** Let $m = 1$, $d = 2$, $A = [-1, 1]$, $f(x_1, x_2, a) := |x_1 + x_2| + |x_1 - x_2| + a^2/2$, $\alpha(x_1, x_2, a) \equiv 1$, $\sigma(x_1, x_2, a) \equiv (1, 1)^T$ and $\mu(x_1, x_2, a) = a(1, -1)^T$. Then, for a constant policy $\pi_0 \equiv a \in A$, the controlled process started at $x \in \mathbb{R}^2$ is given by $X^{\pi_0}_t = x + (1, 1)^T t B + a(1, -1)^T t$, for $t \geq 0$. In particular, for $a = 0$ and any $x = (x_1, x_2)^T$, we have $V_{\pi_0}(x) = |x_1 - x_2| + g(x_1 + x_2)$, $g(y) := \mathbb{E}|y + 2 B_{c_1}|$, $y \in \mathbb{R}$, and $c_1$ is an exponential random variable with mean $1$, independent of $B$. Since $B_{c_1}$ has a smooth density, it follows that $g$ is also smooth, implying that $V_{\pi_0}$ cannot satisfy the conclusion of Proposition [1].

**Remark 4.** As we are using the standard weak formulation of the control problem (i.e. the filtered probability space is not specified in advance), all that matters for a Markov policy $\pi$ is the law of the controlled process $X^{\pi,\cdot}$, which solves the martingale problem in [4]. Moreover, this law is uniquely determined by the symmetric-matrix valued function $(x, a) \mapsto \sigma(x, a)\sigma(x, a)^T \in \mathbb{R}^{d \times d}$ and of course the drift $(x, a) \mapsto \mu(x, a))$. Since the symmetric square root of $\sigma(x, a)\sigma(x, a)^T$ in $\mathbb{R}^{d \times d}$ satisfies Assumption [1] in the remainder of the paper we assume, without loss of generality, that the noise and the controlled process are of the same dimension, i.e. $d = m$ (cf. Remark 2).

**Remark 5.** For any locally Lipschitz Markov policy $\pi$, the process $X^{\pi,\cdot}$ is strong Markov. Hence [8] Thm. 1.7 implies that $V_\pi$ is in the domain $\mathcal{D}(A_\pi)$ of the generator $A_\pi$ of $X^{\pi,\cdot}$ and that the Poisson equation $A_\pi V_\pi - \alpha_\pi V_\pi + f_\pi = 0$ holds. Recall that for a bounded continuous $g : \mathbb{R}^d \to \mathbb{R}$ in $\mathcal{D}(A_\pi)$ the limit $(A_\pi g)(x) := \lim_{t \to 0} (\mathbb{E} g(X^{\pi,\cdot}_t) - g(x))/t$ exists for all $x \in \mathbb{R}^d$. Furthermore, if $g$ is also in $C^2(\mathbb{R}^d)$, it is known that $A_\pi g = L_\pi g$. However, [8] Thm. 1.7 does not imply that $V_\pi$ is in $C^2(\mathbb{R}^d)$. The PDE in Proposition [1] key for (gPIA) to work, is established via the coupling argument in Section 5.2.

If a policy $\pi : \mathbb{R}^d \to A$ is constant (i.e. $\pi \equiv p \in A$), write $\sigma_p$, $\mu_p$, $\alpha_p$, $f_p$, $L_p$ and $V_p$ instead of $\sigma_\pi$, $\mu_\pi$, $\alpha_\pi$, $f_\pi$, $L_\pi$ and $V_\pi$, respectively. Let $S : \mathbb{R}^d \times A \to (0, \infty)$ be a continuous function and, for any $p \in A$, denote $S_p(x) := S(x, p)$. Under Assumption [1] the function
\[ p \mapsto S_p(x)(L_p h(x) - \alpha_p(x)h(x) + f_p(x)), \quad p \in A, \]
continues for any \( x \in \mathbb{R}^d \) and \( h \in C^2(\mathbb{R}^d) \).
Since \( A \) is compact, there exists \( I_h(x) \in A \), which minimises this function.

**Assumption 2.** For any \( h \in C^2(\mathbb{R}^d) \), the function \( I_h : \mathbb{R}^d \to A \) can be chosen to be locally Lipschitz on \( \mathbb{R}^d \). The continuous scaling function \( S \) satisfies \( \epsilon_S < S < M_S \) for some \( \epsilon_S, M_S \in (0, \infty) \).

**Generalised Policy Iteration Algorithm**

**Input:** \( \sigma, \mu, \alpha, f, S \) satisfying Assumptions 1, 2 constant policy \( \pi_0 \) and \( N \in \mathbb{N} \).

**for** \( n = 0 \) to \( N - 1 \) **do**

1. Compute \( V_{\pi_n} \) from the PDE in Proposition 1
2. Choose the policy \( \pi_{n+1} \) as follows:

\[
\text{(gPIA)} \quad \pi_{n+1}(x) \in \arg\min_{p \in A} \{ S_p(x)(L_p V_{\pi_n}(x) - \alpha_p(x)V_{\pi_n}(x) + f_p(x)) \}, \quad x \in \mathbb{R}^d.
\]

**end**

**Output:** Approximation \( V_{\pi_N} \) of the value function \( V \).

**Remark 6.** By Assumption 2 the policy \( \pi_{n+1} \) defined in (gPIA) is locally Lipschitz for all \( 0 \leq n \leq N - 1 \). Hence, by Proposition 1 the algorithm is well defined.

**Remark 7.** In the classical case of (gPIA) we take \( S \equiv 1 \). A non-trivial scaling function \( S \), which plays an important role in the one-dimensional context (see Section 3 below), makes the algorithm into a generalised Policy Iteration Algorithm. Thm 2 requires only the positivity and continuity of \( S \). The uniform bounds on \( S \) in Assumption 2 are used only in the proof of Thm 5 (gPIA) always leads to an improved payoff (Theorem 2 is proved in Section 5.2 below):

**Theorem 2.** Under Assumptions 1, 2 the inequality \( V_{\pi_{n+1}} \leq V_{\pi_n} \) holds on \( \mathbb{R}^d \) for all \( n \in \{0, \ldots, N - 1\} \).

The sequence \( \{V_{\pi_N}\}_{N \in \mathbb{N}} \), obtained by running (gPIA) from a given policy \( \pi_0 \) for each \( N \in \mathbb{N} \), is non-increasing and bounded. Hence we may define

\[
V_{\text{lim}}(x) := \lim_{N \to \infty} V_{\pi_N}(x), \quad x \in \mathbb{R}^d.
\]

However, the sequence of the corresponding Markov policies \( \{\pi_N\}_{N \in \mathbb{N}} \) need not converge and, even if it did, the limit may be discontinuous and hence not necessarily a Markov policy.

**Remark 8.** If the algorithm stops before \( N \), i.e. \( \pi_{n+1} = \pi_n \) for some \( n < N \), then clearly \( V_{\pi_N} = V_{\pi_n} \) and \( \pi_N = \pi_n \). As this holds for any \( N > n \), with (gPIA) started at a given \( \pi_0 \), we may proceed directly to the verification lemma (Theorem 5 below) to conclude that \( V_{\pi_n} \) is the value function with an optimal policy \( \pi_n \).

Controlling the convergence of the policies requires the following additional assumptions.

Introduce the set \( S_{B,K} := \{ h \in C^2(\mathbb{R}^d) : \|\nabla h(x)\| < B^1, \|Hh(x)\| < B^2 \text{ for } x \in D_K \} \), where \( D_K := \{ x \in \mathbb{R}^d : \|x\| \leq K \} \) is a ball of radius \( K > 0 \) and \( B := (B^1, B^2) \in (0, \infty)^2 \) are constants.

**Assumption 3.** For any \( K > 0 \), there exist constants \( B_K \) and \( C_K \) satisfying the following: if \( h \in S_{B_K,K} \), then \( I_h \) (defined in Assumption 2) satisfies \( d(I_h(x), I_h(y)) \leq C_K \|x - y\| \) for all \( x, y \in D_K \).

**Assumption 4.** For any \( K > 0 \), let \( B_K, C_K \) be as in Assumption 3. Then for a locally Lipschitz Markov policy \( \pi : \mathbb{R}^d \to A \), such that \( d(\pi(x), \pi(y)) \leq C_K \|x - y\| \), \( x, y \in D_K \), the following holds:

\( V_{\pi} \in S_{B_K,K} \).
Theorem 4. Under Assumptions 1–4, the equality $V_{n} = \lim_{\pi} V_{\pi}$ holds on $\mathbb{R}^d$ for a policy $\pi_{\lim}$.

Let $\pi_{\lim} : \mathbb{R}^d \to A$ denote a locally Lipschitz Markov policy that is a locally uniform limit of a subsequence of $\{\pi_N\}_{N \in \mathbb{N}}$. By Propositions 1, $V_{\pi_{\lim}}$ is a well-defined function in $C^2(\mathbb{R}^d)$ that solves the corresponding Poisson equation. However, since $\pi_{\lim}$ clearly depends on its defining subsequence, so may $V_{\pi_{\lim}}$. Furthermore, $V_{\lim}$ may depend on the choice of $\pi_0$ in (gPIA). But this is not so, since $V_{\lim}$ equals both the value function $V$ and the payoff for the policy $\pi_{\lim}$.

Remark 9. Non-trivial problem data that satisfy Assumptions 1–4 are described in Section 3 below. It is precisely these types of examples that motivated the form Assumptions 2–4 take. Assumption 1 is standard and Assumptions 2–3 concern only the deterministic data specifying the problem. Assumption 4 essentially states that $\|HV_n\|$ has a prescribed bound on the ball $D_K$ if the coefficients of the PDE in Proposition 1 have a prescribed Lipschitz constant. Schauder’s boundary estimates for elliptic PDEs [10, p. 86] suggest that this requirement is both natural and feasible. In fact, Assumption 4 may follow from assumptions of the type 1–3 on the problem data. This is left for future research.

Proposition 3 and Theorems 4 and 5, proved in Section 5.2 below, show that (gPIA) converges.

Proposition 3. Let Assumptions 1–4 hold. Then there exists a subsequence of $\{\pi_N\}_{N \in \mathbb{N}}$ that converges uniformly on every compact subset of $\mathbb{R}^d$ to a locally Lipschitz Markov policy.

Theorem 5. Under Assumptions 1–4 the equality $V_{\lim} = V_{\pi_{\lim}}$ holds on $\mathbb{R}^d$ for a policy $\pi_{\lim}$.

Remark 10. The key technical issue in the proof of Theorem 5 is that the policies in the convergent subsequence constructed in the proof of Proposition 3 are not improvements of their predecessors (cf. (gPIA)). The idea of the proof is to work with a convergent subsequence of the pairs of policies $\{(\pi_N, \pi_{N+1})\}_{N \in \mathbb{N}}$, where $\pi_N$ is produced by the (gPIA) (see Section 5.2 for details).

3. THE ONE-DIMENSIONAL CASE

There are two reason for considering the one-dimensional control problem in its own right. (A) The canonical choice for the scaling function $S := 1/\sigma^2$ simplifies the (gPIA) to

$$\pi_{n+1}(x) \in \arg\min_{\pi \in A} \left\{ (\mu(x,p)V_{\pi_n}(x) - H(x,p))/\sigma^2 \right\},$$

by removing the second derivative of the payoff function $V_{\pi_n}$ from the minimisation procedure. This reduction appears to make the numerical implementation of the (gPIA) converge extremely fast: in the example in Section 4.2 below the optimal payoff and policy are obtained in fewer than half a dozen iterations.

(B) It is natural to control the process $X^{\Pi,x}$ only up to its first exit from an interval $(a,b)$, where $a, b \in [-\infty, \infty]$, and generalise the payoff as follows:

$$V_{\Pi}(x) := \mathbb{E} \left( \int_0^{\tau_{b}(X^{\Pi,x})} e^{-\int_0^s \alpha_n(x^{\Pi,x}) ds} f_{\Pi}(X^{\Pi,x}_s) dt + e^{-\int_0^{b}(X^{\Pi,x}) \alpha_n(x^{\Pi,x}) dt} g(X^{\Pi,x}_{\tau_{b}(X^{\Pi,x})}) \right).$$

Here $\mu, \alpha, f : (a,b) \times A \to \mathbb{R}$ are measurable with the same notational convention as in Section 2 ($f_p(x) = f(x,p)$ etc.). Furthermore, $\Pi \in A(x)$ if $X^{\Pi,x}$ follows SDE 1 on the stochastic interval...
Data satisfying Assumptions 1–4. The main aim in the present section is not to be exhaustive, but merely to demonstrate that the form (particularly) of Assumptions 3–4 is natural in the context of control problems considered here. The example we give is in dimension one. But it is clear that the proofs, in Sections 5.3 and 5.4 below, rely on the theory of ODEs and scalar SDEs. In particular, we need to prove that the payoff has a continuous extension to a finite boundary point of the state space. In the interest of brevity, we omit their statement. We stress that the main difference lies in the fact that the proofs, in Sections 5.3 and 5.4 below, rely on the theory of ODEs and scalar SDEs. In particular, we need to prove that the payoff has a continuous extension to a finite boundary point of the state space.

4. Examples

4.1. Data satisfying Assumptions 1–4. We now describe a class of models that provably satisfies Assumptions 1–4. The main aim in the present section is not to be exhaustive, but merely to demonstrate that the form (particularly) of Assumptions 3–4 is natural in the context of control problems considered here. The example we give is in dimension one. But it is clear from the construction below that it can easily be generalised.

Let $A := [-a, a]$, for some constant $a > 0$, and $\sigma, \mu, f, \alpha : \mathbb{R} \times [-a, a] \to \mathbb{R}$ be given by

\begin{equation}
\sigma(x, p) := \sigma_1(x), \quad \mu(x, p) := \mu_1(x) + p\mu_2, \quad f(x, p) := f_1(x) + f_2(p), \quad \alpha(x, p) \equiv \alpha_0,
\end{equation}

where $\sigma_1, \mu_1, f_1 \in C^1(\mathbb{R})$, $f_2 \in C^2((-a, a))$ is convex and symmetric (i.e. $f_2(p) = f_2(-p)$ for all $p \in A$) and $\mu_2$ and $\alpha_0$ are constants. For any $h \in \{\sigma_1, \mu_1, f_1, f_2\}$ (resp. $h' \in \{\sigma_1', \mu_1', f_1', f_2'\}$) let the positive constant $C_h$ (resp. $C_h'$) satisfy $|h| \leq C_h$ (resp. $|h'| \leq C_h'$). In particular, we assume that the derivatives of $\sigma_1, \mu_1, f_1, f_2$ are bounded. Moreover we may (and do) take $C_{f_2} := f_2'(a)$. Assume also that $\alpha_0 > 0$ and $\sigma_2^2 > \lambda > 0$ (so that As 1 is satisfied) and the scaling function $S \equiv 1$. Then the following proposition holds.
Proposition 6. Assume \( \alpha_0 > C_{\mu_1} + |\mu_2|(2 + C_{f_1}/C_{f_2}) \) and \( |\mu_2|B^2_\infty < L_{f_2} := \inf_{p \in A} f''_2(p) \), where

\[
B^2_\infty := \left[ 2(C_{f_1} + C_{f_2}) + (C_{\mu_1} + a|\mu_2|)B^1_\infty \right] / \lambda \quad \text{and} \quad B^1_\infty := (C_{f_1} + C_{f_2})/(\alpha_0 - C_{\mu_1} - |\mu_2|),
\]

then Assumptions 1–4 hold. Moreover, in Assumptions 3–4 we have \( B_K = (B^1_\infty, B^2_\infty) \) and \( C_K = 1 \) for any \( K > 0 \).

Remark 13. It is clear that the assumptions in Proposition 6 define a non-empty subclass of models [8]. Moreover, these assumptions are much stronger than what is required by our general Assumptions 3–4 since the proposition yields global (rather than local) bounds on the derivatives of the payoff functions and the Lipschitz coefficients of the policies arising in (gPIA).

Proof. Pick \( h \in C^2(\mathbb{R}) \), such that \( |h'(x)| < B^1_\infty \) and \( |h''(x)| < B^2_\infty \) for all \( x \in \mathbb{R} \). Then the function \( I_h \) in Ass 3 satisfies \( I_h(x) = \arg \min_{p \in A} \{ \mu_2h'(x) + f_2(p) \} \). By assumption we have

\[
|\mu_2h'(x)| \leq |\mu_2|B^1_\infty < C_{f_2}' < f_2'(a), \quad \text{implying} \quad I_h(x) = (f_2')^{-1}(-\mu_2h'(x)) \quad \forall x \in \mathbb{R}.
\]

Differentiate \( I_h \) to obtain \( |I_h'(x)| \leq |\mu_2|B^2_\infty / L_{f_2} < 1 \), \( x \in \mathbb{R} \), and note that Assumptions 2–3 follow.

We now establish As 4. The idea is to start with any policy \( \pi : \mathbb{R} \to A \) in \( C^4(\mathbb{R}) \), such that its derivative satisfies \( |\pi'_{\omega}| \leq 1 \) on all of \( \mathbb{R} \) (e.g. a constant policy), and apply stochastic flow of diffeomorphisms [21 Sec. V.10] to deduce the necessary regularity of the payoff function \( V_{\pi} \). In the notation from [8], we have \( |\mu_{\pi}'| \leq C_{\mu_1} + |\mu_2| \) and \( |\sigma_{\pi}'| = |\sigma_1'| \leq C_{\sigma_1}' \). Hence, for each \( x \in \mathbb{R} \), the stochastic exponential \( Y = (Y_t)_{t \in \mathbb{R}_+} \), given by

\[
Y_t = 1 + \int_0^t \mu_{\pi}'(X_{s,x}^\pi)Y_s ds + \int_0^t \sigma_{\pi}'(X_{s,x}^\pi)Y_s dW_s,
\]

exists (we suppress the dependence on \( x \) and \( \pi \) from the notation). Since the coefficients SDE [1] are in \( C^1(\mathbb{R}) \) with bounded and locally Lipschitz first derivative, [21 Sec. V.10, Thm 49] implies that the flow of controlled processes \( \{X_{s,x}^\pi\}_{x \in \mathbb{R}} \) may be constructed on the single probability space so that it is smooth in the initial condition \( x \) with \( \frac{\partial}{\partial x}X_{s,x}^\pi = Y \). The unshot here is that, by the argument in the proof of [9 Prop. 3.2], we obtain a stochastic representation for the derivative \( \frac{\partial}{\partial x} \mathbb{E}f_{\pi}(X_{t,x}^\pi) = \mathbb{E}Y_t f'_{\pi}(X_{t,x}^\pi) \) for every \( t \in \mathbb{R}_+ \). Since \( Y_t = M_t \exp \int_0^t \mu_{\pi}'(X_{s,x}^\pi)ds \), where the stochastic exponential \( M = \mathcal{E}(\int_0^\cdot \sigma_{\pi}'(X_{s,x}^\pi)dW_s) \) is a true martingale by Novikov’s condition, the following inequality holds: \( \mathbb{E}Y_t \leq \exp(t(C_{\mu_1} + |\mu_2|)) \). Since \( \alpha_0 > C_{\mu_1} + |\mu_2| \) by assumption and the inequality \( |f''_2| \leq C_{f_1}' + C_{f_2}' \) holds, we have \( \mathbb{E} \int_0^\infty e^{-\alpha_0 s}Y_t f''_{\pi}(X_{s,x}^\pi)ds < B^1_\infty \) for all \( x \in \mathbb{R} \).

Recall that \( V_{\pi}(x) = \mathbb{E} \int_0^\infty e^{-\alpha_0 s}f_{\pi}(X_{s,x}^\pi)ds \). By [21 Sec. V.8, Thm 43], the family of random variables indexed by \( \delta \in (0, 1) \),

\[
\frac{1}{\delta} \int_0^\infty e^{-\alpha_0 s}e_{\pi}(X_{s,x+\delta}) - e_{\pi}(X_{s,x})|ds \leq \frac{1}{\delta} (C_{f_1} + C_{f_2}')(\int_0^\infty e^{-\alpha_0 s}|X_{s,x+\delta} - X_{s,x}|ds,
\]

is uniformly integrable. Hence \( \lim_{\delta \to 0} (V_{\pi}(x + \delta) - V_{\pi}(x)) / \delta \) takes the form

\[
V''_{\pi}(x) = \mathbb{E}(\int_0^\infty e^{-\alpha_0 s}Y_s f''_{\pi}(X_\pi,x)ds) \quad \text{implying} \quad |V''_{\pi}(x)| < B^2_\infty \quad \text{for all} \quad x \in \mathbb{R}.
\]

This inequality, the fact \( \sigma_{\pi}' > \lambda \) and Proposition 1 imply \( |V''_{\pi}| < B^2_\infty \), concluding the proof. \( \square \)
Remark 14. The process $Y$ in the proof of Proposition 6 exists in the multidimensional setting, see [21, Sec. V.10, Thm 49]. Hence the same argument works in higher dimensions if we can deduce a bound on the Hessian of the payoff function from the PDE in Proposition 1.

4.2. Numerical examples. Consider the one-dimensional control problem: $A = [-1, 1]$, $a = -10$, $b = 10$, $g(a) := a^2$, $g(b) := b^2$, $\sigma(x, p) := 1$, $\mu(x, p) := 1$ and $f(x, p) := x^2 + p^2$, which is in the class discussed in Section 4.1. Explicitly, we seek to compute $\inf_{\Pi \in A(x)} V_{\Pi}(x)$ for every $x \in (-10, 10)$, where the payoff $V_{\Pi}(x)$ of a policy $\Pi$ is defined in Section 3.

We implemented (gPIA), with the main step given by (7), in Matlab. The payoff function at each step is obtained as the solution to the differential equation from Proposition 1 with the boundary conditions given by the function $g$. The new policy at each step can be calculated explicitly (cf. the proof of Proposition 6 above). Figures 1 and 2 graph the payoff functions and the policies (colour coded). The initial policy $\pi_0 \equiv 1$ and its payoff correspond to the blue graphs.

The graphs suggest that convergence effectively occurs in just a few steps. Figures 3 and 4, containing the graphs of the differences of the consecutive payoffs and policies on the logarithmic scale, confirm this. In Figures 1 and 2 it seems that fewer graphs are presented than is stated in the caption. The reason for this is that the final few graphs coincide. Moreover, the policies only differ on a subinterval $(-2, 2)$, because outside of it they coincide as it is optimal to chose one of the boundary points of $A = [-1, 1]$. Finally, there is no numerical indication that the sequence of policies have more than one accumulation point as they appear to converge very fast indeed.

5. Proofs

5.1. Auxiliary results - the multidimensional case.

5.1.1. The reflection coupling of Lindvall and Rogers [18] and the continuity of the payoff $V_{\pi}$. We now establish the continuity of the payoff function for a locally Lipschitz Markov policy $\pi$ under Assumption 1. The reflection coupling of Lindvall and Rogers [18] plays a crucial role in this. In fact, the continuity of $V_{\pi}$ hinges on the following property of the coupling in [18]: copies of $X^{\pi, x}$ and $X^{\pi, x'}$, started very close to each other, will meet with high probability before moving apart by a certain distance greater than $\|x - x'\|$ (see Lemma 7 below).
We first show that the coupling from [18] can be applied to the diffusion $X^{\pi, c}$. As explained in Remark 4, we may (and do) assume that the dimension of the noise and the controlled process are equal, i.e. $d = m$. By Assumption (i) above $\sigma_0$ and $\mu_0$ are bounded and hence [18] As. (12)(ii) holds. Inequality (i) in Assumption (i) implies that $\lambda_{\text{max}}(\sigma_0^{-1} \sigma_0^{-T}) \leq 1/\lambda$. Hence, by Remark 4, we have $\|\sigma_0^{-1}\| \leq 1/\sqrt{\lambda}$ and [18] As. (12)(ii) also holds. The assumptions in [18] (12)(i) requires that $\sigma_0$ and $\mu_0$ are globally Lipschitz. But this assumption is only used in [18] as a guarantee that the corresponding SDE has a unique strong solution, which is the case in our setting under the locally Lipschitz condition in Assumption (i). Hence, for any $x, x' \in \mathbb{R}^d$, the coupling from [18] can be applied to construct the process $(X^{\pi, x}, X^{\pi, x'})$ so that $X^{\pi, x}$ follows SDE (4) and $X^{\pi, x'}$ satisfies

$$X_t^{\pi, x'} = x' + \int_0^t \mu_x \left( X_s^{\pi, x'} \right) ds + \int_0^t \sigma_x \left( X_s^{\pi, x'} \right) H_s dB_s, \quad \text{for } t \in [0, \rho_0(Y)),$$

where $\rho_0(Y) := \inf\{t \geq 0 : \|Y_t\| = 0\}$ (inf $\emptyset := \infty$) is the coupling time, $Y := X^{\pi, x} - X^{\pi, x'}$, and

$$H_t := I - 2u_t u_t^T, \quad \text{defined via } \ u_t := \frac{\sigma_0^{-1} \left( X_t^{\pi, x'} \right) Y_t}{\|\sigma_0^{-1} \left( X_t^{\pi, x'} \right) Y_t\|} \quad \text{for } t \in [0, \rho_0(Y)),$$

is the reflection on $\mathbb{R}^d$ about the hyperplane orthogonal to the unit vector $u_t$. Moreover, we have $X_t^{\pi, x'} = X_t^{\pi, x}$ for all $t \in [\rho_0(Y), \infty)$. Note also that $H_t \in O(d)$ is an orthogonal matrix for $t \in [0, \rho_0(Y))$ and the process $B' = (B'_t)_{t \in \mathbb{R}_+}$, given by $B'_t := \int_0^t (\mathbb{1}_{\{s < \rho_0(Y)\}} H_t + \mathbb{1}_{\{s \geq \rho_0(Y)\}} I) dB_s$, is a Brownian motion by the Lévy characterisation theorem. Hence $X^{\pi, x'}$ satisfies the SDE $dX_t^{\pi, x'} = \sigma_x (X_t^{\pi, x'}) dB'_t + \mu_x (X_t^{\pi, x'}) dt$ with $X_0^{\pi, x'} = x'$ (see [18] Sec. 3 for more details).

**Lemma 7.** Fix a locally Lipschitz Markov policy $\pi : \mathbb{R}^d \to A$ and $x \in \mathbb{R}^d$. Then for every $\epsilon \in (0, 1)$ there exist $\tilde{\varphi} \in (0, 1]$ with the property: $\forall \varphi \in (0, \tilde{\varphi}) \exists \varphi' \in (0, \varphi)$ such that $\mathbb{P}(\rho_c(Y) \leq \rho_0(Y)) < \epsilon$ if $\|x - x'\| < \varphi'$, where $\rho_c(Y) := \inf\{t \geq 0 : \|Y_t\| = c\}$ (inf $\emptyset = \infty$) for any $c > 0$.

**Remark 15.** Note that the main assumption in [18] Thm. 1 is not satisfied in Lemma 7 as we have no assumption on the global variability of $\sigma_0$. Hence the coupling $(X^{\pi, x}, X^{\pi, x'})$ is not necessarily successful even if the starting points $x$ and $x'$ are very close to each other, i.e. possibly $\mathbb{P}(\rho_0(Y) < \infty) < 1$ even if $\|Y_0\| = \|x - x'\|$ is very close to zero. However, by Lemma 7 the
coupling will occur with probability at least $1 - \epsilon$ before the diffusions are more than $\bar{\varphi} = \bar{\varphi}(\epsilon)$ away from each other, implying the continuity of $V_\pi$ (cf. Lemma 3 and Remark 10 below). 

**Proof.** Let $\bar{S} := \|Y\|^2, \delta := \sigma_\pi(X^{x,x}) - \sigma_\pi(X^{x,x'})$ and $\beta := \mu_\pi(X^{x,x}) - \mu_\pi(X^{x,x'})$. Define
\begin{equation}
\alpha_t := \sigma_\pi(X^{x,x}_t) - \sigma_\pi(X^{x,x'}_t)H_t \quad \text{and} \quad v_t := Y_t/\|Y_t\| \quad \text{for} \ t \in [0, \rho_0(Y)).
\end{equation}
In this proof $x \in \mathbb{R}^d$ is fixed and $x' \in \mathbb{R}^d$ is arbitrary in the ball of radius one centred at $x$. Recall that $\nabla h(z) = 2z$ and $\mathcal{H}a(z) = 2I$ for $h(z) := \|z\|^2$, $z \in \mathbb{R}^d$, and apply Itô’s lemma to $\bar{S}$:
\begin{equation}
\bar{S}_t = \|x - x'\|^2 + \int_0^t 2\sqrt{\bar{S}_s}v_s^T \alpha_s dB_s + \int_0^t \left(2\sqrt{\bar{S}_s}v_s^T \beta_s + \text{Tr}(\alpha_s \alpha_s^T)\right) ds, \quad t \in [0, \rho_0(Y)).
\end{equation}

Our task is to study the behaviour of $\bar{S}$ when started very close to zero. To do this, we first establish the facts in (12) and (14) below, which in turn allow us to apply time-change and coupling techniques to prove the lemma. We start by proving the following:
\begin{equation}
0 \leq \text{Tr}(\alpha_t \alpha_t^T) - \|v_t^T \alpha_t\|^2 = \text{Tr}(\delta_t \delta_t^T) - \|v_t^T \delta_t\|^2 \leq M_d^2 \|Y_t\|^2 \quad \text{for} \ t \in [0, \rho_0(Y) \wedge \rho_1(Y)),
\end{equation}
where $M_d > 1$ is a Lipschitz constant for $\sigma_\pi$ and $\mu_\pi$ in the ball around $x$ of radius one. The first inequality in (12) follows since the trace is the sum of the eigenvalues of $\alpha_t \alpha_t^T$, which are all non-negative, while $\|v_t^T \alpha_t\|^2$ is at most the largest eigenvalue. The second inequality follows since $\sigma_\pi$ is Lipschitz on any ball around $x$ and $\|Y_t\| < 1$ for $t < \rho_1(Y)$. To establish the equality in (12) note that, as $\|v_t^T A\|^2 = \text{Tr}(A^T v_t v_t^T) = \text{Tr}(v_t v_t^T A^T A)$ for any $A \in \mathbb{R}^{d \times d}$, we have
\[\text{Tr}(\alpha_t \alpha_t^T) - \|v_t^T \alpha_t\|^2 - (\text{Tr}(\delta_t \delta_t^T) - \|v_t^T \delta_t\|^2) = \text{Tr}((I - v_t v_t^T)(\alpha_t \alpha_t^T - \delta_t \delta_t^T)).\]
Recall that $H_t^{-1} = H_t$. We therefore find
\begin{equation}
\alpha_t \alpha_t^T - \delta_t \delta_t^T = \sigma_\pi(X^{x,x}_t)(I - H_t)\sigma_\pi(X^{x,x}_t)^T + \sigma_\pi(X^{x,x}_t)(I - H_t)\sigma_\pi(X^{x,x'}_t)^T
\end{equation}
\begin{equation}
= 2(v_t v_t^T \sigma_\pi(X^{x,x}_t)^T + \sigma_\pi(X^{x,x}_t)u_t u_t^T) \|Y_t\|/\|\sigma_\pi(X^{x,x}_t)\| \|Y_t\|/\|\sigma_\pi(X^{x,x'}_t)\| Y_t^2,
\end{equation}
where the second equality follows by definition (9) and identity $v_t = Y_t/\|Y_t\|$. Hence (12) follows.

Since $\text{Tr}(v_t v_t^T \sigma_\pi(X^{x,x}_t)T) = \langle \sigma_\pi(X^{x,x}_t) \sigma_\pi(X^{x,x'}_t)^{-1} v_t, v_t \rangle \|Y_t\|/\|\sigma_\pi(X^{x,x'}_t)^{-1} Y_t\|^2$ holds for times $t \in [0, \rho_0(Y))$, equalities (12) and (13) yield:
\[\|v_t^T \alpha_t\|^2 \geq \|v_t^T \delta_t\|^2 = 4\langle \sigma_\pi(X^{x,x}_t) \sigma_\pi(X^{x,x'}_t)^{-1} v_t, v_t \rangle \|Y_t\|^2/\|\sigma_\pi(X^{x,x'}_t)^{-1} Y_t\|^2.\]
Inequality (3) in Assumption 1 implies $\|\sigma^{-1}\| \leq 1/\sqrt{\lambda}$ (cf. the second paragraph of this section). Hence $\|Y_t\|/\|\sigma_\pi(X^{x,x'}_t)^{-1} Y_t\| \geq \sqrt{\lambda}$. By the definition of $\delta_t$ above we get
\[\|v_t^T \alpha_t\|^2 \geq 4\lambda(1 + \langle \delta_t \sigma_\pi(X^{x,x'}_t)^{-1} v_t, v_t \rangle) \geq 4\lambda(1 - \|\delta_t\|/\|\sigma_\pi(X^{x,x'}_t)^{-1}\|) \geq 4\lambda(1 - \|\delta_t\|/\sqrt{\lambda}).\]
For any $\epsilon \in (0, 1)$, define
\[\bar{\varphi} := \min\{1, \epsilon \sqrt{\lambda}/M_x, (1 - \epsilon) \lambda/M_d^2\},\]
where $M_x$ is as in (12) above. Then, if $\|x - x'\| < \bar{\varphi}$ and $t \in [0, \rho_\varphi(Y))$, we have $\|Y_t\| < \bar{\varphi}$ and hence $\|\delta_t\| \leq M_x \|Y_t\| \leq \epsilon \sqrt{\lambda}$. In particular, we get
\begin{equation}
\|v_t^T \alpha_t\|^2 \geq 4\lambda(1 - \epsilon) > 0 \quad \text{for any} \ t \in [0, T), \quad \text{where} \ T := \rho_0(Y) \wedge \rho_\varphi(Y).
\end{equation}
Let $M > 0$ denote a global upper bound on $\sigma_\pi$ and $\mu_\pi$, which exists by Assumption 1. Since the inequalities $\|v_t^T \alpha_t\| \leq \|\alpha_t\| \leq \|\sigma_\pi(X^{x,x}_t)\| + \|\sigma_\pi(X^{x,x'}_t)\| \leq 2M$ hold for all $t \in [0, \rho_0(Y))$, the increasing process $[N] = ([N]_t)_{t \in \mathbb{R}^+}$, given by $[N]_t := \int_0^t 1_{(s < \rho_0(Y))} v_s^T \alpha_s^2 ds$, is well-defined and $[N]_t < \infty$ for every $t \in \mathbb{R}^+$. Hence $N = (N_t)_{t \in \mathbb{R}^+}$, given by $N_t := \int_0^t 1_{(s < \rho_0(Y))} v_s^T \alpha_s^2 ds$.}
is a well-defined local martingale with a quadratic variation process given by $[N]$. Let $\tau = (\tau_s)_{s \in \mathbb{R}^+}$ and $W = (W_s)_{s \in \mathbb{R}^+}$ be the Dambis Dubins-Schwarz (DDS) time-change and Brownian motion, respectively, for the local martingale $N$ (see [14] Thm 3.4.6, p. 174). More precisely, let $s \mapsto \tau_s := \inf\{t \in \mathbb{R}^+ : [N]_t > s\}$ (with $\inf \emptyset = \infty$) be the inverse of $t \mapsto [N]_t$. Then $W$ satisfies $W[N]_t = N_t$ for all $t \in \mathbb{R}^+$. Moreover it holds that $\tau_s < \infty$ for $s < [N]_\infty := \lim_{t \to \infty} [N]_t$. If $[N]_\infty < \infty$ with positive probability, we have to extend the probability space to support $W$ (see e.g. [14] Prob. 3.4.7, p. 175). This extension however has no bearing on the coupling $(X^{\pi,x}, X^{\pi,x'})$.

Let $\tilde{\alpha}_s := \alpha_{\tau_s}$, $\tilde{\delta}_s := \delta_{\tau_s}$, $\tilde{\beta}_s := \beta_{\tau_s}$, $\tilde{v}_s := v_{\tau_s}$, and $\hat{S}_s := S_{\tau_s}$ for $s \in [0, [N]_{\rho_0(Y)})$, cf. (10) above. Assume $\|x - x'\| < \varphi$ and time-change the integrals in (11) (see [14] Prop. 3.4.8, p. 176) to get

$$\hat{S}_u = \|x - x'\|^2 + \int_0^\infty 2\sqrt{\hat{S}_s} dW_s + \int_0^u (1 + \nu_s) ds,$$

for any $u \in [0, [N]_T)$, where $\nu_s := (2\sqrt{\hat{S}_s} \tilde{\beta}_s + \text{Tr}(\tilde{\delta}_s \tilde{\delta}_s^T) / \|\tilde{\delta}_s \tilde{\delta}_s^T\|^2$ and $T$ is defined in (14). By (14) it holds that $\|\tilde{\delta}_s \tilde{\delta}_s^T\|^2 \geq 4A_1(1 - \epsilon)$ for all $s \in [0, [N]_T)$. Then (12) and the definitions of $\tilde{\beta}$, $\tilde{\delta}$ and $\nu_s$ imply the inequalities $0 \leq \nu_s \leq M_2^2 [Y_{\tau_s}]^2 / (\lambda(1 - \epsilon))$ for all $s \in [0, [N]_T)$. Any $\varphi \in (0, \bar{\varphi})$ satisfies $\varphi < \epsilon(1 - \epsilon)\lambda/M^2_2$ and $R := \rho_0(Y) / \rho_\varphi(Y) \leq T$. Hence the Lipschitz property of $\sigma_\varphi$ and $\nu_s$ on the ball of radius $\varphi$ around $x$ implies

$$\nu_s < \epsilon \quad \text{for all } s \in [0, [N]_R).$$

The SDE $S_s = \|x - x'\| + \int_0^s 2\sqrt{S_r} dW_r + (1 + \epsilon)s$, $s \in \mathbb{R}^+$, for the squared Bessel process of dimension $1 + \epsilon$ has a pathwise unique (and hence strong) solution $S_s := (S_s)_{s \in \mathbb{R}^+}$, see [6] App. A.3, p. 108. Note that $S$ is driven by the DDS Brownian motion $W$ introduced above. Hence the coupling $(S, \hat{S})$ on the stochastic interval $[0, \rho_\varphi(Y)]$ allows us to compare the two processes pathwise. Assume $\|x - x'\| < \varphi$. Then the following equality holds:

$$\sqrt{S_s} - \sqrt{\hat{S}_s} = \frac{1}{2} \int_0^s \left( \epsilon / \sqrt{S_r} - \nu_r / \sqrt{\hat{S}_r} \right) dr \quad \text{for any } s \in [0, \rho_\varphi(Y)].$$

Almost surely, the path of the process $(\sqrt{S_s} - \sqrt{\hat{S}_s})_{s \in [0, \rho_\varphi(Y)]}$ is continuously differentiable and, by (15), its derivative is strictly positive at every zero of the path. Since the derivative is continuous, it must be strictly positive on a neighbourhood of each zero. This implies that the only zero is at $s = 0$ (i.e. $S_0 = \hat{S}_0$), and it holds that

$$S_s \geq \hat{S}_s \quad \text{for all } s \in [0, \rho_\varphi(Y)].$$

We now conclude the proof of the lemma. Assume as before that $\|x - x'\| < \varphi$ and define $Y_\varphi(\hat{S}) := \inf\{s \in [0, \rho_\varphi(Y)] : \hat{S}_s = \varphi\}$ (with $\inf \emptyset = \infty$). Note that the events $\{Y_\varphi(\hat{S}) < \infty\}$ and $\{[N]_{\rho_\varphi(Y)} < [N]_{\rho_0(Y)}\}$ coincide, since on either event we have $Y_\varphi(\hat{S}) = \rho_\varphi(Y) = \rho_0(Y)$. Hence,

$$\{\rho_\varphi(Y) < \rho_0(Y)\} = \{Y_\varphi(\hat{S}) < \infty\} \subseteq \{S \text{ exits interval } (0, \varphi) \text{ at } \varphi\},$$

where the inclusion follows by (16). Recall that $s(z) = z^{(1 - \epsilon)/2}$, $z \in \mathbb{R}^+$, is a scale function of the diffusion $S$. Hence $\mathbb{P}(S \text{ exits interval } (0, \varphi) \text{ at } \varphi) = s(\|x - x'\| / s(\varphi))$. Define $\varphi' := \epsilon \varphi$ and note that by (17) we have: $\mathbb{P}(\rho_\varphi(Y) < \rho_0(Y)) < \epsilon$ for any $x' \in \mathbb{R}^d$ satisfying $\|x - x'\| < \varphi'$. ~\(\square\)

**Lemma 8.** Pick a locally Lipschitz Markov policy $\pi : \mathbb{R}^d \to A$ and let Assumption [1] hold. Then the corresponding payoff function in (5), $V_\pi : \mathbb{R}^d \to \mathbb{R}^+$, is continuous.
Proof. Fix $x \in \mathbb{R}^d$ and pick arbitrary $\varepsilon \in (0, 1)$. By Assumption [1], there exists $\epsilon_0 > 0$, such that $\alpha_\pi \geq \epsilon_0$, and a constant $M > 1$ that simultaneously bounds $\alpha_\pi, |f_\pi| < M$ and is a Lipschitz constant on the ball of radius one around $x$ for $\alpha_\pi$ and $f_\pi$. Apply Lemma [7] to $x, \varepsilon := \varepsilon \epsilon_0/(6M)$ and $\pi$ to obtain $\varphi (x, \varepsilon) \geq C(1, \varphi)$, and $\pi_0 \leq M^2/(3\epsilon_0^2)$. Specifically, define

$$\varphi : \min \{ \varphi / 2, \varepsilon / (3 + 1 + M/\epsilon_0) e/\epsilon_0^2, (\epsilon e \epsilon_0)/(3M^2) \}$$

and fix $\varphi' \in (0, \varphi)$ such that the conclusion of Lemma [7] holds. Throughout this proof we use the notation and notions from Lemma [7]. In particular, $(X^{\pi, x}, X^{\pi, x'})$ denotes the coupling of two controlled processes started at $(x, x')$ and we assume that $\|x - x'\| < \varphi'$.

Recall that $V_\pi(x') = EF_\pi(x')$ for any $x' \in \mathbb{R}^d$, where $F_\pi(x')$ is given in [19]. By decomposing the probability space into complementary events $\{\rho_\varphi > \rho_0\}$ and $\{\rho_\varphi < \rho_0\}$, we obtain the following inequality $|V_\pi(x') - V_\pi(x')| \leq A + A' + A''$, where

$$A := E \left( \mathbb{I}_{\{\rho_\varphi > \rho_0\}} \int_0^{\rho_0} e^{-f_\pi \alpha_\pi ds} f_\pi \left( X^{\pi, x} \right) \right),$$

$$A' := E \left( \mathbb{I}_{\{\rho_\varphi > \rho_0\}} \int_0^{\rho_0} e^{-f_\pi \alpha_\pi ds} f_\pi \left( X^{\pi, x} \right) \right),$$

$$A' := E \left( \mathbb{I}_{\{\rho_\varphi < \rho_0\}} \int_0^{\rho_0} e^{-f_\pi \alpha_\pi ds} f_\pi \left( X^{\pi, x} \right) \right).$$

Hence, by Lemma [7] we have $A'' \leq E \left( \mathbb{I}_{\{\rho_\varphi < \rho_0\}} 2M/\epsilon_0 \leq 2M/\epsilon_0 = \varepsilon/3$. Since in the summands $A$ and $A'$ the coupling succeeds before the components of $(X^{\pi, x}, X^{\pi, x'})$ grow at least $\varphi$ apart, we can control these terms using the local regularity of $\alpha_\pi$ and $f_\pi$. Consider $A$. Add and subtract $e^{-f_\pi \alpha_\pi ds} f_\pi \left( X^{\pi, x} \right)$ to obtain the bound:

$$A \leq E \left( \mathbb{I}_{\{\rho_\varphi > \rho_0\}} \int_0^{\rho_0} \left( e^{-\epsilon_0 t} f_\pi \left( X^{\pi, x} \right) \right) + M e^{-f_\pi \alpha_\pi ds} f_\pi \left( X^{\pi, x} \right) \right) \right) dt.$$
successful, since the global Lipschitz condition controls globally the local variability of the coefficients. The coupling may fail because the assumptions in [18, Thm. 1] constrain the global variability of $\sigma_t$. In fact, the idea of the proof of Lemma 7 can be used to construct an example where $\mathbb{P}(\rho_0(Y) < \infty) < 1$ by bounding the norm of $\|Y\|^2$ from below by a squared Bessel process of dimension greater than two on an event of positive probability.

5.1.2. A version of the Ascoli-Arzelà Theorem. The following fact is key for proving the existence of the optimal strategy and showing that a subsequence of $\{\pi_N\}_{N \in \mathbb{N}}$ in [gPIA] converges to it.

**Lemma 9.** Let $(M_1, d_1)$ and $(M_2, d_2)$ be compact metric spaces, and for every $n \in \mathbb{N}$ let $f_n : M_1 \to M_2$. If the sequence $\{f_n\}_{n \in \mathbb{N}}$ is equicontinuous, i.e.

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall x, y \in M_1 \; \forall n \in \mathbb{N} : \; d_1(x, y) < \delta \implies d_2(f_n(x), f_n(y)) < \epsilon,$$

then there exists a uniformly convergent subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$, i.e. $\exists f : M_1 \to M_2$ such that for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\sup_{x \in M_1} d_2(f_{n_k}(x), f(x)) < \epsilon$ for all $k \geq N$.

**Proof.** Let $B(x, 1/m) := \{y \in M_1 : d_1(x, y) < 1/m\}$ be a ball of radius $1/m$, $m \in \mathbb{N}$, centred at $x \in M_1$. Since $M_1$ is compact and metric, it is totally bounded: $\exists S_m \subseteq M_1$ finite satisfying $M_1 = \bigcup_{x \in S_m} B(x, 1/m)$. Then $S := \bigcup_{m \in \mathbb{N}} S_m = \{x_n \in M_1 : n \in \mathbb{N}\}$ is countable and dense in $M_1$. We now apply the standard diagonalisation argument to find the subsequence in the lemma.

Let $\iota_1 : \mathbb{N} \to \mathbb{N}$ be an increasing function defining a subsequence $\{f_{\iota_1(n)}\}_{n \in \mathbb{N}}$ that converges at $x_1$, i.e. $\lim_{n \to \infty} f_{\iota_1(n)}(x_1)$ exists in $M_2$. Such a function $\iota_1$ exists since $M_2$ is compact. Assume now that we have constructed an increasing $\iota_k : \mathbb{N} \to \mathbb{N}$ such that $\{f_{\iota_k(n)}\}_{n \in \mathbb{N}}$ converges on the set $\{x_1, \ldots, x_k\}$ for some $k \in \mathbb{N}$. Then there exists an increasing $\iota : \mathbb{N} \to \mathbb{N}$ such that the sequence of functions $\{f_{\iota_k(n)}\}_{n \in \mathbb{N}}$, where $\iota_{k+1} := \iota_k \circ \iota$, converges at $x_{k+1}$ as well as on the set $\{x_1, \ldots, x_k\}$, as it is a subsequence of $\{f_{\iota_k(n)}\}_{n \in \mathbb{N}}$. Since $k \in \mathbb{N}$ was arbitrary, we have defined a sequence of subsequences of $\{f_n\}_{n \in \mathbb{N}}$, such that the $k$-th subsequence converges on $\{x_1, \ldots, x_k\}$.

Consider the “diagonal” subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$, $n_k := \iota_k(k)$ for any $k \in \mathbb{N}$. By construction it converges on $S$. We now prove that it is uniformly Cauchy, which implies uniform convergence since $M_2$ is complete. Pick any $\epsilon > 0$. By equicontinuity $\exists m \in \mathbb{N}$ such that for any $k \in \mathbb{N}$ and $x, y \in M_1$ satisfying $d_1(x, y) < 1/m$, it holds that $d_2(f_{n_k}(x), f_{n_k}(y)) < \epsilon/3$. Furthermore, since $S_m$ is finite, $\exists N \in \mathbb{N}$ such that for all natural numbers $k_1, k_2 \geq N$ we have $d_2(f_{n_{k_1}}(y), f_{n_{k_2}}(y)) < \epsilon/3$ for all $y \in S_m$. Finally, for any $x \in M_1$ there exists $y \in S_m$ such that $d_1(x, y) < 1/m$. Hence, for any $k_1, k_2 \geq N$ it holds that

$$d_2(f_{n_{k_1}}(x), f_{n_{k_2}}(x)) \leq d_2(f_{n_{k_1}}(x), f_{n_{k_1}}(y)) + d_2(f_{n_{k_1}}(y), f_{n_{k_2}}(y)) + d_2(f_{n_{k_2}}(y), f_{n_{k_2}}(x)) < \epsilon.$$ 

Since $x \in M_1$ was arbitrary, the lemma follows. □

5.1.3. A uniformly integrable martingale. If the process $X^{\pi,x}$ in [4], controlled by a Markov policy $\pi$, exists for all $x \in \mathbb{R}^d$, then $X^{\pi,:}$ is a strong Markov process [14] Thm 4.30, p. 322, since $\sigma$ and $\mu$ are bounded by Assumption 1. Define the additive functional $F(X^{\pi,x}) = (F_t(X^{\pi,x}))_{t \in [0, \infty]}$,

$$F_t(X^{\pi,x}) := \int_0^t e^{-\int_0^t \alpha_s(X^{\pi,x}) ds} f_\pi(X^{\pi,x}) \, du \quad \text{for } t \in [0, \infty].$$

**Remark 17.** Note that $V_\pi(x) = \mathbb{E}F_\infty(X^{\pi,x})$ and, by Assumption 1, the process $|F(X^{\pi,x})|$ is bounded by some constant $C_0 > 0$. Hence $|F_\infty(X^{\pi,x})| < C_0$ and $|V_\pi(x)| < C_0$. 


Lemma 10. The following holds for every Markov policy \( \pi, x \in \mathbb{R}^d \) and \((\mathcal{F}_t)\)-stopping time \( T \):
\[
\mathbb{E}(F_\infty(X^{\pi,x}) | \mathcal{F}_T) = F_T(X^{\pi,x}) + \mathbb{E} \left[ \mathbb{1}_{\{T < \infty\}} e^{-\int_0^T \alpha_\pi(X^{\pi,x}_s) ds} V_\pi(X^{\pi,x}_T) \right].
\]
In particular, the process \( M = (M_t)_{t \in [0, \infty]} \) is a uniformly integrable martingale, where
\[
M_t := F_t(X^{\pi,x}) + \mathbb{1}_{\{t < \infty\}} e^{-\int_0^t \alpha_\pi(X^{\pi,x}_s) ds} V_\pi(X^{\pi,x}_t).
\]

Proof. The following calculations imply the lemma:
\[
\begin{align*}
\mathbb{E}(F_\infty(X^{\pi,x}) | \mathcal{F}_T) & = F_T(X^{\pi,x}) + \mathbb{E} \left[ \mathbb{1}_{\{T < \infty\}} \int_0^\infty e^{-\int_0^t \alpha_\pi(X^{\pi,x}_s) ds} f_\pi(X^{\pi,x}_t) \ dt \bigg| \mathcal{F}_T \right] \\
& = F_T(X^{\pi,x}) + \mathbb{E} \left[ \mathbb{1}_{\{T < \infty\}} \int_0^\infty e^{-\int_0^t \alpha_\pi(X^{\pi,x}_s) ds} f_\pi(X^{\pi,x}_t) \ dt \bigg| \mathcal{F}_T \right] \\
& = F_T(X^{\pi,x}) + \mathbb{E} \left[ \int_0^\infty e^{-\int_0^t \alpha_\pi(X^{\pi,x}_s) ds} f_\pi(X^{\pi,x}_t) \ dt \bigg| \mathcal{F}_T \right]
\end{align*}
\]
where we applied the strong Markov property of \( X^{\pi,x} \) in the last step. \( \square \)

5.2. Proofs of results in Section 2

Proof of Proposition 4 Assume \( m = d \), cf. Remark 4 It suffices to prove that the PDE holds on the ball \( D := \{ y \in \mathbb{R}^d : ||y - x'|| < 1 \} \) for any \( x' \in \mathbb{R}^d \). Fix \( x \in D \) and define \( \tau := \inf\{t \in \mathbb{R}_+ : X^{\pi,x}_t \in \partial D \} \) (with \( \inf \emptyset = \infty \)) to be the first time the process \( X^{\pi,x} \) hits the boundary \( \partial D := \{ y \in \mathbb{R}^d : ||y - x'|| = 1 \} \) of \( D \). Note that, by Assumption 1, we have \( \tau < \infty \).

Let \( v \in C^2(D) \cap C(\bar{D}) \), where \( \bar{D} := D \cup \partial D \), denote a solution of the boundary value problem
\[
L_\pi v - \alpha_\pi v + f_\pi = 0 \quad \text{in} \quad D, \quad \text{where} \quad v = V_\pi \quad \text{on} \ \partial D.
\]
Since \( \pi \) is locally Lipschitz and \( \alpha_\pi \) in Assumption 1 holds, the coefficients \( \sigma_\pi, \mu_\pi, f_\pi, \alpha_\pi \) are \((1/2)\)-Hölder (in fact Lipschitz) on \( \bar{D} \). The boundary data \( V_\pi|_{\partial D} \) is continuous by Lemma 8 \( \alpha_\pi \geq 0 \) and \( \sigma_\pi \) satisfies (3). Hence, by [10] (Thm 19, p. 87), the function \( v \) exists, is unique and \( \partial_\nu \) is \((1/2)\)-Hölder.

Note that, for all \( t \in [0, \infty] \), we have \( X^{\pi,x}_{t \wedge \tau} \in D \). Hence we can define
\[
Y_t := F_{t \wedge \tau}(X^{\pi,x}) + e^{-\int_0^{t \wedge \tau} \alpha_\pi(X^{\pi,x}_s) ds} v(X^{\pi,x}_{t \wedge \tau}), \quad \text{for} \ t \in [0, \infty],
\]
where the process \( F(X^{\pi,x}) \) is given in (19) above. The process \( Y := (Y_t)_{t \in [0, \infty]} \) is bounded by a constant and by definition converges almost surely \( \lim_{t \to \infty} Y_t = Y_\infty \). Since \( v \) solves the boundary value problem above and \( X^{\pi,x} \) satisfies SDE (4), Itô’s formula on the stochastic interval \( [0, \tau] \subset \mathbb{R}_+ \) yields
\[
Y_t = v(x) + \int_0^{t \wedge \tau} e^{-\int_0^s \alpha_\pi(X^{\pi,x}_r) dr} \nabla v(X^{\pi,x}_s)^T \sigma_\pi(X^{\pi,x}_s) dB_s, \quad t \in [0, \infty],
\]
making \( Y \) into a local martingale. Since \( Y \) is bounded, it is a uniformly integrable martingale satisfying \( v(x) = Y_0 = \mathbb{E}[Y_\infty] \). Since \( v = V_\pi \) on \( \partial D \) and \( X^{\pi,x}_{t \wedge \tau} \in \partial D \), the definition of \( Y \) in (20) and Lemma 10 (applied to the stopping time \( T := \tau \)) yield
\[
Y_\infty = F_\tau(X^{\pi,x}) + e^{-\int_0^\tau \alpha_\pi(X^{\pi,x}_r) dr} V_\pi(X^{\pi,x}_\tau) = \mathbb{E}(F_\infty(X^{\pi,x}) | \mathcal{F}_\tau),
\]
implying \( v(x) = \mathbb{E}(F_\infty(X^{\pi,x})) = V_\pi(x) \).

The uniqueness follows similarly: let \( \bar{v} \) be another bounded solution of the Poisson equation on \( \mathbb{R}^d \). Define the process \( \bar{Y} \) as in (20) with \( \tau \equiv \infty \) and \( t < \infty \). As above we have \( v(x) = \mathbb{E} \bar{Y}_t \) for all \( t \in \mathbb{R}_+ \). Then the DCT, applicable since \( v \) is bounded, yields \( v(x) = \lim_{t \uparrow \infty} \mathbb{E} \bar{Y}_t = V_\pi(x) \). \( \square \)
Proof of Theorem 2. Let $\pi_n$ and $\pi_{n+1}$ be as in \([\text{gPIA}], n \in \mathbb{N} \cup \{0\}$. Define $Y = (Y_t)_{t \in \mathbb{R}_+}$ by

\[Y_t := F_t(X_{t+1}^n) + e^{-\int_0^t \alpha_{n+1}(x) \, dx} V_{\pi_n}(X_t^n), \quad t \in \mathbb{R}_+.
\]

where $F_t(x)$ is defined in \([19]\) above. Define $\tau_m := \inf\{t \geq 0 : \|X_{t+1}^n\| = m\}$ for any fixed $m > \|x\|$ and note that $\tau_m < \infty$ by Assumption \([1]\). Itô’s formula, applicable by Proposition \([1]\), yields

\[Y_{t+\tau_m} = V_{\pi_n}(x) + M + \int_0^{t+\tau_m} e^{-\int_0^s \alpha_{n+1}(x) \, ds} \left( f_{\pi_n+1} + L_{\pi_n+1} V_{\pi_n} - \alpha_{n+1} V_{\pi_n} \right) (X_s^n) \, ds,
\]

where $M = (M_t)_{t \in \mathbb{R}_+}$, $M_t := \int_0^{t+\tau_m} e^{-\int_0^s \alpha_{n+1}(x) \, ds} \left( \nabla V_{\pi_n} \sigma_{\pi_{n+1}} \right) (X_s^n) \, dB_s$, is a local martingale. Since the functions $\sigma_{\pi_{n+1}}$ and $\nabla V_{\pi_n}$ are bounded on the ball $\{y \in \mathbb{R}^d : \|y\| \leq m\}$ by Assumption \([1]\) and Proposition \([1]\) respectively, and $\alpha_{n+1} > \epsilon_0 > 0$, the quadratic variation of $M$ is bounded above by a constant. Hence $M$ is a uniformly integrable martingale. In particular, $\mathbb{E}M_t = 0$ for all $t \in \mathbb{R}_+$. By \([\text{gPIA}]\) and Proposition \([1]\), we have

\[(f_{\pi_n+1} + L_{\pi_n+1} V_{\pi_n} - \alpha_{n+1} V_{\pi_n}) S_{\pi_{n+1}} \leq (f_{\pi_n} + L_{\pi_n} V_{\pi_n} - \alpha_{n+1} V_{\pi_n}) S_{\pi_n} = 0 \quad \text{on } \mathbb{R}^d.
\]

Since $S_{\pi_{n+1}} > 0$, we have $E(Y_{t+\tau_m}) \leq V_{\pi_n}(x)$. Hence \([21]\), Assumption \([1]\) and the DCT, as $t \uparrow \infty$, yield

\[V_{\pi_n}(x) \geq EF_{\pi_n}(X_{t+1}^n) + EV_{\pi_n}(X_{t+1}^n) - \int_0^t \alpha_{n+1}(x) \, dx.
\]

Hence, by Remark \([17]\), we have $V_{\pi_n}(x) \geq EF_{\pi_n}(X_{t+1}^n) - C_0 \mathbb{E}e^{-\tau_m}$. Since $X_{t+1}^n$ satisfies SDE \([4]\) for all $t \in \mathbb{R}_+$, we have $\lim_{m \uparrow \infty} \tau_m = \infty$. The DCT and Remark \([17]\) yield $V_{\pi_{n+1}}(x) = EF_{\pi_{n+1}}(X_{t+1}^n) = \lim_{m \uparrow \infty} EF_{\pi_{n+1}}(X_{t+1}^n) - C_0 \mathbb{E}e^{-\tau_m} \leq V_{\pi_n}(x)$, which concludes the proof. \(\square\)

Proof of Proposition 3. Run \([\text{gPIA}]\) to produce a sequence of policies $\{\pi_N\}_{N \in \mathbb{N}}$, starting from a constant policy $\pi_0$. Fix an arbitrary $K_0 > 0$ and consider the restriction of this sequence onto the closed ball $D_{K_0}$. Since the Lipschitz constant of $\pi_0$ is equal to zero and hence smaller than $C_{K_0}$, Assumption \([4]\) implies $V_{\pi_0} \in S_{D_{K_0} \setminus K_0}$. Assumption \([3]\) implies that the Lipschitz constant of $\pi_1$ is also at most $C_{K_0}$. Iterating this argument implies that all the policies in the sequence $\{\pi_N\}_{N \in \mathbb{N}}$ have the same Lipschitz constant on $D_{K_0}$, making it equicontinuous on $D_{K_0}$. By Lemma \([9]\) above, there exists a subsequence that converges uniformly on $D_{K_0}$ to a function $\pi^0_{\infty} : D_{K_0} \rightarrow A$. Moreover, $\pi^0_{\infty}$ is also Lipschitz with a constant bounded above by $C_{K_0}$.

Let $K_1 := 2K_0$ and repeat the argument above for $K_1$ and the subsequence of $\{\pi_N\}_{N \in \mathbb{N}}$ constructed in the previous paragraph. This yields a further subsequence of the policies that converges uniformly to a Lipschitz function $\pi^1_{\infty} : D_{K_1} \rightarrow A$ with the Lipschitz constant bounded above by $C_{K_1}$. Since the sequence we started with converges pointwise to $\pi^0_{\infty}$ on $D_{K_0} \subset D_{K_1}$, so must its every subsequence. Hence it holds that $\pi^1_{\infty}(x) = \pi^0_{\infty}(x)$ for all $x \in D_{K_0}$.

For $k \in \mathbb{N}$, let $K_k := 2K_{k-1}$ and construct inductively $\pi^k_{\infty} : D_{K_k} \rightarrow A$ as above. Then the function $\pi^k_{\lim} : \mathbb{R}^d \rightarrow A$, given by $\pi^k_{\lim}(x) := \pi^k_{\infty}(x)$ for any $n \in \mathbb{N}$ such that $x \in D_{K_n}$, is well-defined and locally Lipschitz. Let the policy $\pi_{n_k} : \mathbb{R}^d \rightarrow A$ be the $k$-th element of the convergent subsequence used to define $\pi^k_{\infty} : D_{K_k} \rightarrow A$. Then, by construction, the “diagonal” subsequence $\{\pi_{n_k}\}_{k \in \mathbb{N}}$ of $\{\pi_N\}_{N \in \mathbb{N}}$ converges uniformly to $\pi^k_{\lim}$ on $D_K$ for any $K > 0$. \(\square\)

Proof of Theorem 4. Let $\{\pi_{n_k}\}_{k \in \mathbb{N}}$ be a subsequence of the output of \([\text{gPIA}]\), $\{\pi_N\}_{N \in \mathbb{N}}$, that converges locally uniformly to a policy $\pi^k_{\lim} \equiv \lim_{k \uparrow \infty} \pi_{n_k}$. By \([0]\) and Theorem \([2]\) $V_{\pi_{n_k}} \searrow V_{\pi^k_{\lim}}$
as \( k \to \infty \). Fix \( K > 0 \) and let \( \tau_K := \inf \{ t \in \mathbb{R}_+ : X_t^{\pi_{\text{lim},x}} - x \in \partial D_K \} \) be the first time \( X_t^{\pi_{\text{lim},x}} \) hits the boundary of the closed ball \( x + D_K \) with radius \( K \), centred at an arbitrary \( x \in \mathbb{R}^d \).

Pick \( k \in \mathbb{N} \), \( t \in \mathbb{R}_+ \) and define

\[
S_t^k := \int_0^t e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} f_{\pi_{\text{lim}}} (X_s^{\pi_{\text{lim},x}}) \, ds + e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} V_{\pi_{\text{lim}}} (X_t^{\pi_{\text{lim},x}}) .
\]

Apply Itô’s formula to the process \( S^k = (S_t^k)_{t \geq 0} \) on the stochastic interval \([0, \tau_K]\) to get

\[
S_{t \wedge \tau_K}^k = V_{\pi_{\text{lim}}} (x) + \int_0^{t \wedge \tau_K} e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} \left( \nabla V_{\pi_{\text{lim}}} (X_s^{\pi_{\text{lim},x}}) \right)^T \sigma_{\pi_{\text{lim}}} (X_s^{\pi_{\text{lim},x}}) \, dB_s
\]

\[+ \int_0^{t \wedge \tau_K} e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} (f_{\pi_{\text{lim}}} + L_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} - \alpha_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}}) (X_s^{\pi_{\text{lim},x}}) \, ds .
\]

Note that \( \sigma_{\pi_{\text{lim}}} \) and \( \nabla V_{\pi_{\text{lim}}} \) are bounded on \( D_K \) by Assumption \( \square \) and Proposition \( \square \) respectively, and \( \alpha_{\pi_{\text{lim}}} > \epsilon_0 > 0 \). Hence the quadratic variation of the stochastic integral is bounded, making it into a true martingale. This fact and the equality \( \alpha_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} - f_{\pi_{\text{lim}}} = L_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} \) (Prop. \( \square \)) yield

\[
\mathbb{E}_{t \wedge \tau_K} S_{t \wedge \tau_K}^k = V_{\pi_{\text{lim}}} (x) + \mathbb{E} \int_0^{t \wedge \tau_K} e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} \left( f_{\pi_{\text{lim}}} + L_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} - \alpha_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} \right) (X_s^{\pi_{\text{lim},x}}) \, ds
\]

\[+ \mathbb{E} \int_0^{t \wedge \tau_K} e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} L_{\pi_{\text{lim}}} V_{\pi_{\text{lim}}} (X_s^{\pi_{\text{lim},x}}) \, ds .
\]

Note \( |L_{\pi_{\text{lim}}} - L_{\pi_{\text{lim}}}| V_{\pi_{\text{lim}}} = (\mu_{\pi_{\text{lim}}} - \mu_{\pi_{\text{lim}}})^T \nabla V_{\pi_{\text{lim}}} + \frac{1}{2} \text{Tr}((\sigma_{\pi_{\text{lim}}} + \sigma_{\pi_{\text{lim}}})^T \nabla^2 V_{\pi_{\text{lim}}} (\sigma_{\pi_{\text{lim}}} - \sigma_{\pi_{\text{lim}}})) \).

Since, for every \( k \), \( V_{\pi_{\text{lim}}} \) solves the corresponding Poisson equation in Proposition \( \square \) and, by Assumption \( \square \) and Remark \( \square \) the family of functions \( \{ \sigma_{\pi_{\text{lim}}} \} \) is uniformly bounded on the ball \( x + D_K \), Schauder’s boundary estimate for elliptic PDEs \( \square \) p. 86] implies that the sequences \( \{ \nabla V_{\pi_{\text{lim}}} \}_{k \in \mathbb{N}} \) and \( \{ \nabla^2 V_{\pi_{\text{lim}}} \}_{k \in \mathbb{N}} \) are also uniformly bounded on \( x + D_K \). Since \( \alpha_{\pi_{\text{lim}}} > \epsilon_0 > 0 \) for all \( k \in \mathbb{N} \) and the limits \( \lim_{k \to \infty} \mu_{\pi_{\text{lim}}} = \mu_{\pi_{\text{lim}}} \) and \( \lim_{k \to \infty} \sigma_{\pi_{\text{lim}}} = \sigma_{\pi_{\text{lim}}} \) are uniform on \( x + D_K \), the DCT and the equality in \( \square \) imply \( \lim_{k \to \infty} \mathbb{E} S_{t \wedge \tau_K}^k = V_{\pi_{\text{lim}}} (x) \). Hence, the definition of \( S^k \) above, Assumption \( \square \) and a further application of the DCT yield

\[
V_{\pi_{\text{lim}}} (x) = \mathbb{E} \int_0^{t \wedge \tau_K} e^{-\int_0^t \alpha(x^{\pi_{\text{lim},x}}) \, dt} f_{\pi_{\text{lim}}} (X_s^{\pi_{\text{lim},x}}) \, ds + E_{t \wedge \tau_K} ,
\]

where \( E_{t \wedge \tau_K} := \mathbb{E} e^{-\int_0^{t \wedge \tau_K} \alpha(x^{\pi_{\text{lim},x}}) \, dt} V_{\text{lim}} (X_{t \wedge \tau_K}) .
\]

By \( \square \) and Remark \( \square \) the inequality \( |V_{\text{lim}} (y)| \leq C_0 \) holds for all \( y \in \mathbb{R}^d \). By Assumption \( \square \) we hence get

\[
0 \leq \lim_{t \wedge \tau_K \to \infty} \sup_{t \wedge \tau_K} |E_{t \wedge \tau_K}| \leq C_0 \lim_{t \wedge \tau_K \to \infty} \mathbb{E} e^{-\epsilon_0 (t \wedge \tau_K)} = 0 ,
\]

since \( \tau_K \uparrow \infty \) as \( K \uparrow \infty \). The DCT applied to the first summand in \( \square \), as \( t \wedge \tau_K \to \infty \), yields the equality \( V_{\text{lim}} (x) = V_{\pi_{\text{lim}}} (x) \). Since \( x \in \mathbb{R}^d \) was arbitrary, the theorem follows.

\[ \square \]

**Proof of Theorem \( \square \)** The second assertion in the theorem follows from the first one and Theorem \( \square \). We now establish the first assertion of Theorem \( \square \). Equip \( A \times A \) with a product metric, e.g. \( d_{\infty}((p_1, p_2), (a_1, a_2)) := \max \{ d_A (a_1, p_1), d_A (a_2, p_2) \} \); and let \( \{ \pi_N \}_{N \in \mathbb{N}} \) be constructed by the (gPTA). As in the proof of Proposition \( \square \) \( \{ \pi_N \}_{N \in \mathbb{N}} : \mathbb{R}^d \to A \times A \). Each \( \pi_N \) is Lipschitz on a closed ball \( D_K \) of radius \( K > 0 \) with the Lipschitz constant \( C_K \), independent of \( N \). Hence as in the proof of Proposition \( \square \) there exists a subsequence \( \{ \pi_{N_k} \} \) that converges uniformly on every compact subset of \( \mathbb{R}^d \) to a locally Lipschitz function \( \tilde{\pi}_{\text{lim}, \pi_{\text{lim}}} : \mathbb{R}^d \to A \times A \).
Pick any \( x \in \mathbb{R}^d \), a policy \( \Pi \in \mathcal{A}(x) \), \( K > 0 \) and let \( \tau_K := \inf\{ t \in \mathbb{R}_+ : X_t^\Pi - x \in \partial D_K \} \) be the first time the controlled process \( X_t^\Pi \) hits the boundary of the closed ball \( x + D_K \) with radius \( K \) (centred at \( x \)). Since \( \Pi_t \in \mathcal{A} \) for all \( s \in \mathbb{R}_+ \), the (EPTA) implies the inequality
\[
S_{\Pi_t}(f_{\Pi_t} + L_{\Pi_t} V_{\pi_{\nu_k}} - \alpha_{\Pi_t} V_{\pi_{\nu_k}}) \geq S_{\pi_{\nu_k}}(f_{\pi_{\nu_k}} + L_{\pi_{\nu_k}} V_{\pi_{\nu_k}} - \alpha_{\pi_{\nu_k}} V_{\pi_{\nu_k}}) \quad \text{on} \quad \mathbb{R}^d.
\]
Denote \( L_{\pi} h := L_{\pi} h - \alpha_{\pi} h + f_{\pi} \) for any policy \( \pi \) and \( h \in C^2(\mathbb{R}^d) \). Then, for \( k \in \mathbb{N} \), we find that
\[
\mathbb{E} \left( \int_0^{\tau_{\Pi_K}} e^{- \int_0^t \alpha(x(t)) dt} f_{\Pi_t}(X_t^\Pi) \, ds + e^{- \int_0^{\tau_{\Pi_K}} \alpha(x(t)) dt} V_{\pi_{k+1}} \right) \\
\leq \mathbb{E} \left( \int_0^{\tau_{\Pi_K}} e^{- \int_0^t \alpha(x(t)) dt} \left( f_{\Pi_t} + L_{\Pi_t} V_{\pi_{k+1}} - \alpha_{\Pi_t} V_{\pi_{k+1}} \right) \left( X_t^\Pi \right) \, ds \right)
\]
where the last inequality follows from Assumption 2 and inequality (24).

The next task is to take the limit as \( K \to \infty \) on both sides of inequality (25). Since the sequence \( \{\pi_{k+1}\}_{k \in \mathbb{N}} \) converges locally uniformly to the locally Lipschitz policy \( \pi_{\lim} \) (resp. \( \pi_{\lim} \)), Theorem 4 implies \( V_{\pi_{\lim}} = V_{\lim} \) (resp. \( V_{\pi_{\lim}} = V_{\lim} \)). Proposition 1 implies \( \mathcal{L}_{\pi_{\lim}} V_{\lim} = 0 = \mathcal{L}_{\pi_{\lim}} V_{\lim} \). Hence we can express \( \mathcal{L}_{\pi_{k+1}} V_{\pi_{k+1}} - \mathcal{L}_{\pi_{\lim}} V_{\pi_{k+1}} \) as \( \mathcal{L}_{\pi_{\lim}} V_{\pi_{k+1}} - \mathcal{L}_{\pi_{\lim}} V_{\pi_{k+1}} \). By Schauder’s boundary estimate for elliptic PDEs [10] p. 86, the sequences \( \{\nabla V_{\pi_{k+1}}\}_{k \in \mathbb{N}} \) and \( \{H V_{\pi_{k}}\}_{k \in \mathbb{N}} \) are uniformly bounded on \( x + D_K \). By Assumption 1 and Remark 17, the bounded sequence \( \{\sigma_{\pi_{k+1}}, \mu_{\pi_{k+1}}, \alpha_{\pi_{k+1}}, f_{\pi_{k+1}}, V_{\pi_{k+1}}\}_{k \in \mathbb{N}} \) tends to the limit \( \sigma_{\pi_{\lim}}, \mu_{\pi_{\lim}}, \alpha_{\pi_{\lim}}, f_{\pi_{\lim}}, V_{\pi_{\lim}} \) uniformly on \( x + D_K \) as \( k \to \infty \). Hence, so does
\[
\mathcal{L}_{\pi_{k+1}} V_{\pi_{k+1}} - \mathcal{L}_{\pi_{\lim}} V_{\pi_{k+1}} = [L_{\pi_{k+1}} - L_{\pi_{\lim}}] V_{\pi_{k+1}} - (\alpha_{\pi_{k+1}} - \alpha_{\pi_{\lim}}) V_{\pi_{k+1}} + (f_{\pi_{k+1}} - f_{\pi_{\lim}}) \to 0.
\]

By the elliptic version of Theorem 15 in [10] p. 80 applied to the family of PDEs \( \pi_{\pi_{k+1}} V_{\pi_{k+1}} = 0 \), \( k \in \mathbb{N} \), there exists a subsequence of \( \{V_{\pi_{k+1}}\}_{k \in \mathbb{N}} \) (again denoted by \( \{V_{\pi_{k+1}}\}_{k \in \mathbb{N}} \)), such that the corresponding subsequence \( \{V_{\pi_{k+1}}, \nabla V_{\pi_{k+1}}, H V_{\pi_{k+1}}\}_{k \in \mathbb{N}} \) converges uniformly on the closed ball \( x + D_K \) to \( V_{\pi_{\lim}}, V_{\pi_{\lim}}, H V_{\pi_{\lim}} \). Hence it follows that
\[
\mathcal{L}_{\pi_{\lim}} V_{\pi_{\lim}} - \mathcal{L}_{\pi_{\lim}} V_{\lim} = \mathcal{L}_{\pi_{\lim}} (V_{\pi_{\lim}} - V_{\lim}) - \alpha_{\pi_{\lim}} (V_{\pi_{\lim}} - V_{\lim}) \to 0, \quad \text{as} \quad k \to \infty.
\]
Equations (26) and (27) imply that \( \mathcal{L}_{\pi_{k+1}} V_{\pi_{k+1}} \to 0 \) as \( k \to \infty \) uniformly on the ball \( x + D_K \).

Apply the DCT to the right-hand side of (25) and Assumption 1 and Remark 17 to its left-hand side:
\[
V_{\lim}(x) \leq \mathbb{E} \int_0^{\tau_{\Pi_K}} e^{- \int_0^t \alpha(x(t)) dt} f_{\Pi_t}(X_t^\Pi) \, ds + C_0 \mathbb{E} e^{- \epsilon_0 (t + \tau_{\Pi_K})}.
\]
Since this inequality holds for all \( K, t > 0 \) and \( \tau_K \uparrow \infty \) as \( K \uparrow \infty \), the inequality \( V_{\lim}(x) \leq V_{\Pi}(x) \) follows by the DCT as \( t \wedge K \to \infty \) (cf. the last paragraph of the proof of Theorem 4).

5.3. Auxiliary results - the one-dimensional case. Throughout Sections 5.3 and 5.4, define \( \tau^d(Z) := \inf\{ t \geq 0 : Z_t \in [c, d] \} \) (inf 0 = \( \infty \)) for any continuous stochastic process \( (Z_t)_{t \in \mathbb{R}_+} \) in \( \mathbb{R} \) and \( -\infty \leq c < d \leq \infty \).

**Lemma 11.** For any Markov policy \( \pi : (a, b) \to A \), the payoff function \( V_{\pi} : (a, b) \to \mathbb{R} \) can be continuously extended by defining \( V_{\pi}(a) := g(a) \) if \( a > -\infty \) and \( V_{\pi}(b) := g(b) \) if \( b < \infty \).
Proof. Let \( \{x_n\}_{n \in \mathbb{N}} \) be a decreasing sequence in \((a, b)\) that converges to \(a > -\infty\). We now prove that \( \lim_{n \to \infty} V_\pi(x_n) = g(a) \) (the argument for \(b\) is analogous).

Pick arbitrary \(\epsilon > 0\). Since \(\mu\) is bounded and \(\sigma^2\) bounded and bounded away from 0, a simple coupling argument yields that the process \(X^{\pi, x_n}\) can be bounded by a Brownian motion with drift so that \(\mathbb{P}(\tau_{a}^\infty(X^{\pi, x_n}) > \epsilon) < \epsilon\) and \(\mathbb{P}(\tau_{b}^\infty(X^{\pi, x_n}) > \tau_{a}^\infty(X^{\pi, x_n})) < \epsilon\) hold for large \(n \in \mathbb{N}\). Hence, there exists \(n_0 \in \mathbb{N}\) such that

\[
P(\tau_{a}^\infty(X^{\pi, x_n}) > \epsilon \land \tau_{b}^\infty(X^{\pi, x_n})) \leq \mathbb{P}(\tau_{a}^\infty(X^{\pi, x_n}) > \epsilon) + \mathbb{P}(\tau_{a}^\infty(X^{\pi, x_n}) > \tau_{b}^\infty(X^{\pi, x_n})) < 2\epsilon
\]

for all \(n \geq n_0\). Define the quantities \(B_a^b := |e^{-\int_a^b \alpha_s(x^{\pi, x_n}) \, ds} g(x^{\pi, x_n})|\) and \(B_b^a := \int_a^b \alpha_s(x^{\pi, x_n}) \, ds f(x^{\pi, x_n})|\) for all \(n \geq n_0\). The expectation on the event \(\Omega \setminus C\), which has probability less than \(2\epsilon\), is smaller than \(2M\epsilon\) since \(f, g\) are bounded and \(\alpha \geq \epsilon_0 > 0\). On the event \(C\) we have \(\tau_{a}^b(X^{\pi, x_n}) \leq \epsilon\), which implies \(E(A_a^b)_{\Omega \setminus C} < M\epsilon\). By Theorem 19.1 in [10], the elementary inequality \(1 - e^{-x} \leq x\) for \(x \geq 0\), yields an upper bound on \(E_B^a_{\Omega \setminus C}\) of the form \(|g(a)| E(\int_0^\epsilon a\, \pi(x^{\pi, x_n}) \, dt)\). This concludes the proof.

Lemma 12 is the analogue of Lemma 10 with an analogous proof, which we omit for brevity.

**Lemma 12.** The following holds for every Markov policy \(\pi, x \in (a, b)\) and stopping time \(\rho:\)

\[
E \left( F_{\tau_{a}^b(X^{\pi, x})}(X^{\pi, x}) + e^{-\int_a^b \alpha_s(x^{\pi, x}) \, ds} g(X^{\pi, x}_{\tau_{a}^b(X^{\pi, x})}) I_{\{\tau_{a}^b(X^{\pi, x}) < \infty\}} \right) = M_{\rho},
\]

where \(M_{\rho} := F_{\tau_{a}^b(X^{\pi, x})}(X^{\pi, x}) + I_{[r, \infty[} e^{-\int_a^b \alpha_s(x^{\pi, x}) \, ds} V_\pi(X^{\pi, x}_{\tau_{a}^b(X^{\pi, x})})\) for \(r \in [0, \infty[\). In particular, the process \(M = (M_{\rho})_{\tau \in [0, \infty[}\) is a uniformly integrable martingale.

5.4. **Proofs of results in Section 3.**

**Proof of Proposition 2 in dimension one.** Recall that Assumption 1 holds. We need to show that for any locally Lipschitz Markov policy \(\pi : (a, b) \to A\) we have \(V_\pi \in C^2((a, b))\) and \(L_\pi V_\pi - \alpha V_\pi + f_\pi = 0\).

Let \(a < a' < a'' < x < b' < b'' < b\), and for any \(c < d\) denote \(\tau_{c}^{d} := \tau_{c}^{d}(X^{\pi, x})\). Let \(v \in C^2((a', b')) \cap C([a', b'])\) be the unique solution of the boundary value problem \(L_\pi v - \alpha_v v + f_\pi = 0\), \(v(a') = V_\pi(a')\), \(v(b') = V_\pi(b')\), guaranteed to exist by Theorem 19 in [10] p. 87, which is applicable by Assumption 1. Let \(S_{t}^{a''}\) be \(F_{\tau_{a''}^{b''}(X^{\pi, x})} + e^{-\int_{a''}^{b''} \alpha_s(x^{\pi, x}) \, ds} v(X^{\pi, x}_{\tau_{a''}^{b''}})\). Then, by Itô’s formula on \([0, \tau_{b''}^a]\) and the definition of \(v\), the process \(S_{t}^{a''}\) satisfies

\[
S_{t}^{a''} = v(x) + \int_0^{\tau_{a''}^{b''}} e^{-\int_s^{b''} \alpha_s(x^{\pi, x}) \, ds} \sigma_s v'(X^{\pi, x}_s) \, dB_s
\]

Hence \(S_{t}^{a''}\) is clearly a uniformly integrable martingale and the following equalities hold:

\[
\lim_{t \to \infty} E[S_{t}^{a''} - S_{\infty}^{a''}] = 0 \text{ and } v(x) = E[S_{\infty}^{a''}].
\]

Define \(S_{t}^{a''}\) by substituting \(\tau_{a''}^{b''}\) in the
definition of $S_{a',b'}^{y',z'}$ with $\tau_{a'}^{b'}$. Since $X^y_{x'}$ is continuous, we have $\lim_{a' \downarrow a, b' \uparrow b} E S_{a',b'}^{y',z'} = \tau_{a'}^{b'}$ a.s. Hence, by the DCT, $v(x) = \lim_{a' \downarrow a, b' \uparrow b} E S_{a',b'}^{y',z'} = E S_{a,b}^{y,z}$.

Note that the boundary conditions for $v$, the fact $X_{\tau_{a'}^{b'}}\in \{a', b'\}$ and Lemma 12 (with $\rho = \tau_{a'}^{b'}$) imply $E S_{\tau_{a'}^{b'}}^{y',z'} = E(\int_{\tau_{a'}^{b'}}^{\infty} \alpha(s) X_{\tau_{a'}^{b'}}^{y',z'}(s) ds g(X_{\tau_{a'}^{b'}}^{y',z'}(s)) I_{\{\tau_{a'}^{b'} < \infty\}} | F_{\tau_{a'}^{b'}})$. Taking expectations on both sides of this equality yields $v(x) = V_x(x)$. \hfill \Box

\textbf{Proof of Theorem 2 in dimension one.} We claim that under Assumptions 1–2, the inequality $V_{x_n+1}(x) \leq V_{x_n}(x)$ holds for all $x \in (a, b)$ and $n \in \mathbb{N}$, where $\pi_{n+1}$ is defined in (7).

Define the process $Y$ as in (21) in Section 5.2 and consider the stopped process $Y_{\cdot \wedge \tau_{a'}^{b'}}$, where $a < a' < x < b' < b$ and $\tau_{a'}^{b'} := \tau_{a'}^{b'}(X^{y,x}_{\cdot})$ for any $c < d$. Then the proof follows the same steps as the proof of Theorem 2 in Section 5.2. The only difference is that in the penultimate line of the proof of Section 5.2 we apply the DCT and Lemma 11 (instead of the DCT only) to obtain $V_{x_n}(x) \geq V_{x_{n+1}}(x)$. \hfill \Box

The proof of the one-dimensional case of Proposition 3 is completely analogous to the multi-dimensional one and is hence omitted.

\textbf{Proof of Theorem 4 in one dimension.} We need to show that $V_{\lim}(x) = V_{\pi_{\lim}}(x)$ holds for all $x \in (a, b)$. The proof follows along the same lines as in the multi-dimensional case of Section 5.2. The only difference lies in the fact that we stop the process $X_{\pi_{\lim}}^{x}$ at $\tau_{a'}^{b'}(X_{\pi_{\lim}}^{x})$, where $a < a' < x < b' < b$, and take the limit as $(a', b', t) \to (a, b, \infty)$. \hfill \Box

The verification lemma in the one-dimensional case is established exactly as in the proof of Theorem 5. The details are omitted.

\textbf{References}

[1] A. Arapostathis, On the Policy Iteration Algorithm for Nondegenerate Controlled Diffusions Under the Ergodic Criterion. Optimization, Control, and Applications of Stochastic Systems, 1–12, 2012.

[2] A. Araposthathis and V.S. Borkar, Uniform recurrence properties of controlled diffusions and applications to optimal control. SIAM J. Control Optim., 48(7), 152–160, 2010.

[3] P. Azimzadeh, E. Bayraktar and G. Labahn, Convergence of approximation schemes for weakly nonlocal second order equations. arXiv:1705.02922v2, 2017.

[4] P. Bertsekas, Approximate Policy Iteration: a Survey and Some New Methods. Journal of Control Theory and Applications, 9(3):310–335, 2011.

[5] A. N. Borodin, P. Salminen, Handbook of Brownian Motion – Facts and Formulae. Birkhauser, Basel, second edition, 2002.

[6] A. S. Cherny and H.-J. Engelbert, Singular Stochastic Differential Equations. Lecture Notes in Mathematics, Springer-Verlag, Berlin Heidelberg, 2005.

[7] B. T. Doshi, Continuous Time Control of Markov Processes on an Arbitrary State Space: Discounted Rewards. The Annals of Statistics, 4(6):1219–1235, 1976.

[8] E.B. Dynkin, Markov Processes, vol I. Springer-Verlag, Berlin, 1965.

[9] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, N. Touzi, Applications of Malliavin calculus to Monte Carlo methods in finance. Finance and Stochastics, 3:391–412, 1999.

[10] A. Friedman, Partial Differential Equations of Parabolic Type. Prentice-Hall, Englewood Cliffs, N.J., 1964.

[11] O. Hernández-Lerma and J. B. Lasserre, Policy iteration for average cost Markov control processes on Borel spaces. Acta Applicandae Mathematicae, 47(2):125–154, 1997.

[12] A. Hordijk and M. L. Puterman, On the convergence of policy iteration in finite state undiscounted Markov decision processes: the unichain case. Mathematics of Operations Research, 12(1):163–176, 1987.

[13] R. A. Howard, Dynamic Programming and Markov Processes. MIT Press, Cambridge, 1960.
[14] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, second edition, 1991.

[15] N. V. Krylov, *Controlled Diffusion Processes*. Springer-Verlag, New York, 1980.

[16] M. G. Lagoudakis and R. Parr, Least-Squares Policy Iteration. *Journal of Machine Learning Research*, 4:1107-1149, 2003.

[17] J. B. Lasserre, A new policy iteration scheme for Markov decision processes using Schweitzer’s formula. *Journal of Applied Probability*, 31(1):268–273, 1994.

[18] T. Lindvall and L. C. G. Rogers, Coupling of Multidimensional Diffusions by Reflection. *The Annals of Probability*, 14(3):860–872, 1986.

[19] S. Mahadevan, Representation Policy Iteration. arXiv:1207.1408 [cs.AI], 2012.

[20] S. P. Meyn, The policy iteration algorithm for average reward Markov decision processes with general state space. *IEEE Transactions on Automatic Control*, 42(12):1663–1680, 1997.

[21] P. E. Protter, *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin Heidelberg, 2004.

[22] J. P. Rust, A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-Random, and Deterministic Discretizations. Available at http://ssrn.com/abstract=37768, 1997.

[23] M. S. Santos and J. Rust, Convergence properties of policy iteration. *SIAM Journal on Control and Optimization*, 42(6):2094–2115, 2004.

[24] Q. X. Zhu, X. S. Yang and C. X. Huang, Policy iteration for continuous-time average reward Markov decision processes in Polish spaces. *Abstract and Applied Analysis*, 2009, Article ID103723, 17 pages.

DEPARTMENT OF STATISTICS, UNIVERSITY OF WARWICK, AND, THE ALAN TURING INSTITUTE, UK
E-mail address: s.d.jacka@warwick.ac.uk

DEPARTMENT OF MATHEMATICS, KING’S COLLEGE LONDON, AND, THE ALAN TURING INSTITUTE, UK
E-mail address: aleksandar.mijatovic@kcl.ac.uk

DEPARTMENT OF STATISTICS, UNIVERSITY OF WARWICK, UK
E-mail address: d.siraj@warwick.ac.uk