THE UNIVERSITY OF WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

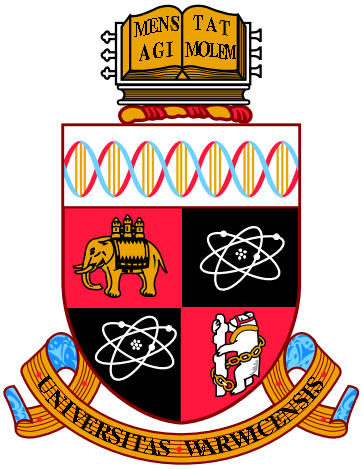
A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/73446

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.



Coupling and the policy improvement algorithm for controlled diffusion processes

by

Dejan Širaj

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

March 2015

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments

First I would like to thank my parents, not only for their unconditional support during my studies, but also for everything they did for me before that, in particular for encouraging my curiosity. A big thank you also goes to my brother Mitja – among other things, it has been really convenient to have free accommodation in central London at any time.

I am most grateful to my supervisors, Professor Saul Jacka and Dr Aleksandar Mijatović, for their guidance, support and advice, always coupled with a good deal of patience. Under their mentorship, it has been really rewarding to learn how Mathematics is being made and presented.

Furthermore, I would like to thank my friends and colleagues, both in England and Slovenia, for mutual support, fruitful discussions, all the fun moments, those little pieces of advice about LaTeX or Matlab that save you a huge amount of time, and much more.

Special thanks go to all my teachers – I would not have been able to start the PhD without the previously acquired skills and knowledge. In particular, I would like to thank Professor Matjaž Omladič for delivering my first course on probability, being my undergraduate theses supervisor, and supporting my decision to pursue my doctorate abroad.

Finally, I would like to thank the Slovene Human Resources Development and Scholarship Fund for their scholarship. I am also grateful to the Department of Statistics, University of Warwick, not only for the partial bursary award, but also for providing stimulating and friendly research environment, for giving me an opportunity to teach, and for enabling me to go to conferences.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Chapters 2, 3 and 4 are joint work with my supervisors Prof. S. D. Jacka and Dr A. Mijatović. The content of Chapters 3 and 4 is in preparation for publication, whereas the content of Chapter 2 has been published as follows:

- S. D. Jacka, A. Mijatović and D. Širaj, Mirror and Synchronous Couplings of Geometric Brownian Motions. *Stochastic Processes and their Applications*, 123(2):1055–1069, 2014.

Abstract

The thesis deals with the mirror and synchronous couplings of geometric Brownian motions, the policy improvement (or iteration) algorithm in completely continuous settings, and an application where the latter is applied to the former.

First we investigate whether the mirror and synchronous couplings of Brownian motions minimise and maximise, respectively, the coupling time of the corresponding geometric Brownian motions. We prove (via Bellman's principle) that this is indeed the case in the infinite horizon and ergodic average problems, but not necessarily in the finite horizon and exponential efficiency problems, for which we characterise when the two couplings are suboptimal.

Then we describe the policy improvement algorithm for controlled diffusion processes in the framework of the discounted infinite horizon problem, both in one and several dimensions. Under some assumptions on the data of the problem, we prove that the algorithm yields a sequence of Markov policies such that its accumulation point is an optimal policy, and that the corresponding payoff functions converge monotonically to the value function. We use no discretisation procedures at any stage. We show that a large class of data satisfies the assumptions, and an example implemented in Matlab demonstrates that the convergence is numerically fast.

Next we study the policy improvement algorithm for continuous finite horizon problem. We obtain analogous results as for the infinite horizon problem. Finally we apply the algorithm to a certain sequence of data to approximate the value function of the (partially unsolved) finite horizon problem for geometric Brownian motions.

Chapter 1

Introduction

1.1 Structure of the thesis

1.1.1 Structure of this chapter

In Section 1.2 we present the basic coupling terminology and prove the coupling inequality. Then we investigate the mirror coupling of Brownian motions. This particular example was chosen not only because it illustrates well the newly introduced concepts, but also because it prepares the ground for the next chapter, in which we deal with the same type of problems but for more complex processes. The section closes with a short literature review on coupling.

Section 1.3 introduces the policy improvement algorithm in a simple setting. We treat the discrete discounted infinite-horizon minimisation problem with countable state space and finite action space. We prove that the algorithm indeed improves the policy at each step. Although the treatment of the policy improvement algorithm in Chapters 3 and 4 will be quite different and more involved, this simple exposition manages to convey the main idea behind the algorithm concisely, which is why it is included. The section again ends with a brief literature review.

1.1.2 Structure of the rest of the thesis

Chapter 2 investigates whether the mirror and synchronous couplings of geometric Brownian motions are optimal in four different (although related) problems. Chapter 3 presents the policy improvement algorithm for the discounted infinite-horizon problem in a continuous setting. In the final chapter we make a connection between the two main topics. We develop the policy improvement algorithm for the continuous finite-horizon problem, and then apply it to approximate the value function for one of the partially unsolved coupling problems.

The three chapters have a similar structure. First comes the introduction, where the problems that will be treated in the chapter are motivated. Then most of the subsequent sections deal with one of them. Statement of the problem and the result(s) are always in an independent subsection for greater transparency. They are followed by the proof(s) in either one or two subsections, depending on whether extensive auxiliary results are required. Each chapter ends with the conclusion, which includes a very brief summary, certain observations and comparisons, reasons why the assumptions had or had not been made, comments about possible generalisations, and other remarks.

The numbering of theorems, lemmas, assumptions, etc., is unified. It includes the chapter number, section number, and the consecutive number of the theorem, lemma, assumption, etc., in that section.

1.2 Coupling

1.2.1 Terminology and the coupling inequality

This subsection is very standard, see e.g. [22] or [30].

Let $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ and $(\hat{\Omega}', \hat{\mathcal{F}}', \hat{\mathbb{P}}')$ be probability spaces, $(E, \mathcal{E})$ a measurable space, $\hat{X} : \hat{\Omega} \to E$ an $(\hat{\mathcal{F}}, \mathcal{E})$-measurable mapping, and $\hat{X}' : \hat{\Omega}' \to E$ an $(\hat{\mathcal{F}}', \mathcal{E})$-measurable mapping. A coupling of random elements $\hat{X}$ and $\hat{X}'$ is an $(\mathcal{F}, \mathcal{E} \otimes \mathcal{E})$-measurable mapping $(X, X') : \Omega \to E \times E$ such that

$$X \stackrel{\mathcal{L}}{=} \hat{X} \quad \text{and} \quad X' \stackrel{\mathcal{L}}{=} \hat{X}',$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\stackrel{\mathcal{L}}{=}$ denotes the equality in law (i.e. distribution).

The coupling where we make the random elements independent always exists due to the product space construction. However, usually we want some dependence because such couplings can be more informative.

The total variation distance between the probability measures $\mathbb{P}$ and $\mathbb{Q}$ on a

measurable space $(E, \mathcal{E})$ is defined as[1]

$$||\mathbb{P} - \mathbb{Q}|| := \sup_{A \in \mathcal{E}} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

Let for any random element X the symbol P_X denote the law of X.

The following lemma, called the coupling inequality, provides an upper bound for the total variation distance between the laws of the coupled random elements. Note that the left-hand side does not depend on the coupling (i.e. the joint law) whereas the right-hand side does.

Lemma 1.2.1. *Let $(E, \mathcal{E})$ be a Polish[2] space, $\hat{X}$ and $\hat{X}'$ random elements on it, and (X, X') their coupling, which is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$||P_{\hat{X}} - P_{\hat{X}'}|| \leq \mathbb{P}(X \neq X').$$

Proof. For any $A \in \mathcal{E}$ we obtain

$$\begin{aligned} P_{\hat{X}}(A) - P_{\hat{X}'}(A) &= \mathbb{P}(X \in A) - \mathbb{P}(X' \in A) \\ &= \mathbb{P}(X \in A, X \neq X') + \mathbb{P}(X \in A, X = X') \\ &\quad - \mathbb{P}(X' \in A, X \neq X') - \mathbb{P}(X' \in A, X = X') \\ &= \mathbb{P}(X \in A, X \neq X') - \mathbb{P}(X' \in A, X \neq X') \\ &\leq \mathbb{P}(X \in A, X \neq X') \\ &\leq \mathbb{P}(X \neq X'). \end{aligned}$$

By symmetry we get $|P_X(A) - P_{X'}(A)| \leq \mathbb{P}(X \neq X')$, and by taking the supremum over $A \in \mathcal{E}$ the desired inequality follows. □

If X and X' are stochastic processes with the index set $I \subseteq \mathbb{R}$, their coupling

[1] Some authors define the total variation distance between the probability measures $\mathbb{P}$ and $\mathbb{Q}$ as $2 \sup_{A \in \mathcal{E}} |\mathbb{P}(A) - \mathbb{Q}(A)|$ since this expression is equal to

$$\sup \left\{ \left| \int_E X \, \mathrm{d}\mathbb{P} - \int_E X \, \mathrm{d}\mathbb{Q} \right| ; \; X : E \to [-1, 1] \text{ is } (\mathcal{E}, \mathcal{B}([-1, 1]))\text{-measurable} \right\}.$$

[2] In fact the space $(E, \mathcal{E})$ does not have to be Polish (i.e. separable and completely metrizable), only the diagonal has to be measurable, i.e.

$$\{(x, x); \; x \in E\} \in \mathcal{E} \otimes \mathcal{E}.$$

time τ is defined as

$$\tau = \inf\{t \in I;\ X_s = X'_s \text{ for all } s \geq t\}, \quad (\inf \emptyset := \infty).$$

The random time τ is neither necessarily a stopping time (with respect to either of the natural filtrations) nor an almost surely finite random variable. We say that coupling is successful if τ is almost surely finite.

Since the inclusion $\{X_t \neq X'_t\} \subseteq \{\tau > t\}$ holds for every $t \in I$, the coupling inequality implies

$$||P_{X_t} - P_{X'_t}|| \leq P(\tau > t), \quad t \in I.$$

Coupling of stochastic processes X and X' is called maximal if equality is achieved in the previous inequality for every $t \in I$. See [30, Ch. 3] for a comprehensive treatment of maximal couplings.

1.2.2 The mirror coupling of Brownian motions

The content of this subsection is well-known, see e.g. [13] or [22].

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space that supports an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B_t)_{t\geq 0}$. (Brownian motion in this thesis will mean one-dimensional Brownian motion started at 0, unless stated otherwise.) Let $\mathcal{V}$ be the set of all $(\mathcal{F}_t)_{t\geq 0}$-Brownian motions. Let $x_1, x_2 \in \mathbb{R}$, and for any $V \in \mathcal{V}$ define

$$\tau(V) := \inf\{t \geq 0;\ x_1 + B_t = x_2 + V_t\}, \quad (\inf \emptyset := \infty).$$

For any $T > 0$, we would like to solve the following problem:

$$\text{find } B^- \in \mathcal{V} \text{ such that} \quad \inf_{V\in\mathcal{V}} \mathbb{P}(\tau(V) > T) = \mathbb{P}(\tau(B^-) > T). \tag{P}$$

Remark 1.2.2. The analogous maximisation problem, i.e.

$$\text{find } B^+ \in \mathcal{V} \text{ such that} \quad \sup_{V\in\mathcal{V}} \mathbb{P}(\tau(V) > T) = \mathbb{P}(\tau(B^+) > T),$$

has an obvious solution $B^+ = B$ since $\tau(B) = \infty$ if $x_1 \neq x_2$. We call (B, B) the synchronous coupling of Brownian motions.

Note that $\tau(V)$ is the first meeting time of the processes $x_1 + B$ and $x_2 + V$,

and not necessarily their coupling time. However, if we define the process $\overline{V}$ by

$$\overline{V}_t := \begin{cases} V_t & \text{if} \quad t \in [0, \tau(V)], \\ B_t & \text{if} \quad t \in (\tau(V), \infty), \end{cases}$$

then the following become apparent: $\overline{V} \in \mathcal{V}$ since both B and V are strong Markov processes with respect to the same filtration, $\tau(V) = \tau(\overline{V})$, and $\tau(\overline{V})$ is the coupling time of B and $\overline{V}$. Therefore by Lemma 1.2.1, Problem (P) is equivalent to finding a maximal coupling of two $(\mathcal{F}_t)_{t\geq 0}$-Brownian motions started at x_1 and x_2. The following theorem provides a solution.

Theorem 1.2.3. *For any $x_1, x_2 \in \mathbb{R}$ and $T > 0$, a solution to Problem* (P) *is given by*

$$B^- = -B.$$

Remark 1.2.4. It follows from the proof that the coupling $(x_1 + B, x_2 + \overline{-B})$ is maximal and successful. It is usually called the mirror (or reflection) coupling of (one-dimensional) Brownian motions started at x_1 and x_2, however we will refer to $(B, -B)$ as the mirror coupling of Brownian motions.

Proof. It is enough to prove that the coupling inequality becomes equality for the mirror coupling, i.e.

$$||P_{x_1+B_t} - P_{x_2-B_t}|| = \mathbb{P}(\tau(-B) > t), \quad t \geq 0.$$

We can assume $x_1 < x_2$ without loss of generality. Note

$$\tau(-B) = \inf\left\{t > 0;\ B_t = \frac{x_2 - x_1}{2}\right\},$$

and therefore

$$\mathbb{P}(\tau(-B) > t) = \mathbb{P}\left(\sup_{s\leq t} B_s < \frac{x_2 - x_1}{2}\right) = \mathbb{P}\left(|B_t| \leq \frac{x_2 - x_1}{2}\right), \quad t \geq 0,$$

by the Reflection Principle.

On the other hand we have the following:

$$\tilde{A} := \left\{y \in \mathbb{R};\ \mathrm{e}^{-\frac{(x_1-y)^2}{2t}} \geq \mathrm{e}^{-\frac{(x_2-y)^2}{2t}}\right\} = \left\{y \in \mathbb{R};\ y \leq \frac{x_1 + x_2}{2}\right\}.$$

Hence we obtain

$$
\begin{aligned}
||P_{x_1+B_t} - P_{x_2-B_t}|| &= \sup_{A\in\mathcal{B}(\mathbb{R})} |\mathbb{P}(x_1 + B_t \in A) - \mathbb{P}(x_2 - B_t \in A)| \\
&= \sup_{A\in\mathcal{B}(\mathbb{R})} (\mathbb{P}(x_1 + B_t \in A) - \mathbb{P}(x_2 - B_t \in A)) \\
&= \frac{1}{\sqrt{2\pi t}} \sup_{A\in\mathcal{B}(\mathbb{R})} \int_A \left(\mathrm{e}^{-\frac{(x_1-y)^2}{2t}} - \mathrm{e}^{-\frac{(x_2-y)^2}{2t}} \right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi t}} \int_{\tilde{A}} \left(\mathrm{e}^{-\frac{(x_1-y)^2}{2t}} - \mathrm{e}^{-\frac{(x_2-y)^2}{2t}} \right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\frac{x_1+x_2}{2}} \mathrm{e}^{-\frac{(x_1-y)^2}{2t}} \mathrm{d}y - \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\frac{x_1+x_2}{2}} \mathrm{e}^{-\frac{(x_2-y)^2}{2t}} \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\frac{x_2-x_1}{2}} \mathrm{e}^{-\frac{y^2}{2t}} \mathrm{d}y - \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\frac{x_1-x_2}{2}} \mathrm{e}^{-\frac{y^2}{2t}} \mathrm{d}y \\
&= \mathbb{P}\left(B_t \le \frac{x_2 - x_1}{2} \right) - \mathbb{P}\left(B_t \le \frac{x_1 - x_2}{2} \right) \\
&= \mathbb{P}\left(|B_t| \le \frac{x_2 - x_1}{2} \right), \quad t \ge 0.
\end{aligned}
$$

□

1.2.3 Literature review

Coupling is a very useful technique, and a popular topic in probability. See the classical books [22] and [30] for the general theory and numerous applications.

The mirror and synchronous couplings of Brownian motions and related processes have attracted much attention in the literature. Paper [23] introduces the mirror coupling of Brownian motions and diffusion processes. In [13] it is established that the mirror coupling of Brownian motions is not the only maximal coupling, although it is the unique maximal coupling in the family of Markovian (or immersion) couplings. More about the Markovian maximal couplings for diffusion processes can be found in [2] and [19].

In [3] it is proved that the tracking error of two driftless diffusions is minimised by the synchronous coupling of the driving Brownian motions. In [15] generalised mirror coupling and generalised synchronous coupling of Brownian motions are introduced; the former minimises the coupling time and maximises the tracking error of two regime-switching martingales, whereas the latter does the opposite.

Articles [1], [6], and [26] discuss various applications of the mirror coupling of reflected Brownian motions and other processes. In particular in [6], the notion of

efficiency of a Markovian coupling, also used in this thesis, is studied in the context of the spectral gap of the generator of a Markov process.

1.3 Policy improvement algorithm

1.3.1 Discrete discounted infinite-horizon problem

The algorithm has become an established method. See [14] and [27] for reference.

We are given the following data:

- the state space S, which is a countable (i.e. finite or denumerable) set;
- the action space A, which is a finite set;
- the discount factor $\alpha \in (0, 1)$;
- the cost function $R : S \times A \to \mathbb{R}$, which satisfies

$$\sup_{i \in S} \max_{a \in A} |R(i, a)| \leq M$$

  for some $M \in \mathbb{R}$ (if the state space is finite, this condition holds automatically);

- the transition probabilities $\{P_{i,j}(a);\ i, j \in S, a \in A\}$, which satisfy

$$\forall i, j \in S\ \forall a \in A : P_{i,j}(a) \in [0, 1] \quad \text{and} \quad \forall i \in S\ \forall a \in A : \sum_{j \in S} P_{i,j}(a) = 1.$$

Now we define the following objects (for any sets C and D, C^D is the set of all mappings from D to C):

- the set of Markov policies:

$$\mathcal{A}_M := \left\{\pi = \{\pi_k\}_{k \in \mathbb{N}_0};\ \forall k \in \mathbb{N}_0 : \pi_k \in A^S\right\};$$

- the set of stationary policies: $\mathcal{A}_S := A^S$; note that even though $\mathcal{A}_S$ is not a subset of $\mathcal{A}_M$, we will treat it as such due to the following natural embedding: $\pi \mapsto (\pi, \pi, \cdots)$;
- (for every $\pi \in \mathcal{A}_M$) the controlled process $\{X_k^\pi\}_{k \in \mathbb{N}_0}$, which is a time-inhomogeneous Markov chain with the transition matrix at step $k \in \mathbb{N}_0$ equal to $\{P_{i,j}(\pi_k(i));\ i, j \in S\}$; the underlying probability space is not important

(we know that it exists) since we will actually only need the law of the process $\{X_k^\pi\}_{k\in\mathbb{N}_0}$, which is unique given X_0^π; note that if $\pi \in \mathcal{A}_S$, the controlled process is a time-homogeneous Markov chain with the transition matrix equal to $\{P_{i,j}(\pi(i));\ i,j \in S\}$;

- (for every $\pi \in \mathcal{A}_M$) the payoff function $V_\pi : S \to \mathbb{R}$ given by

$$V_\pi(i) := \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_k^\pi, \pi_k\left(X_k^\pi\right)\right) \,\middle|\, X_0^\pi = i\right), \quad i \in S;$$

  note that it is well-defined since

$$\sum_{k=0}^{\infty} \alpha^k \left|R\left(X_k^\pi, \pi_k\left(X_k^\pi\right)\right)\right| \le \sum_{k=0}^{\infty} \alpha^k M = \frac{M}{1-\alpha};$$

- the value function $V : S \to \mathbb{R}$, defined by

$$V(\cdot) := \inf_{\pi\in\mathcal{A}_M} V_\pi(\cdot);$$

  note that V is a bounded function due to the estimate above;

- the optimality equation, which is the following function equation (for v):

$$v(i) = \min_{a\in A}\left(R(i,a) + \alpha\sum_{j\in S} P_{i,j}(a)v(j)\right), \quad i \in S;$$

- the shift operator θ: for any sequence $x = \{x_k\}_{k\in\mathbb{N}_0}$ let the sequence $x\circ\theta$ be defined as

$$(x\circ\theta)_k := x_{k+1}, \quad k \in \mathbb{N}_0.$$

The problem is to find the value function and an optimal policy, if it exists. Policy $\pi \in \mathcal{A}_M$ is optimal if $V_\pi(\cdot) = V(\cdot)$.

The following theorem characterises the value function and optimal policy. We will not prove it (it can be found in [27, Ch. 2]), but we will also not use it in the proof of Theorem 1.3.3, which is the main result of this section.

Theorem 1.3.1. *The value function V is the unique bounded solution of the optimality equation. Moreover, if we define the stationary policy π as*

$$\pi(i) := \operatorname*{argmin}_{a\in A}\left(R(i,a) + \alpha\sum_{j\in S} P_{i,j}(a)V(j)\right), \quad i \in S,$$

then it is an optimal policy.

We will follow the convention that if the minimum can be achieved by several arguments, then argmin is any of them.

Mimicking the above formula, for any stationary policy π define

$$\pi'(i) := \operatorname*{argmin}_{a \in A} \left(R(i,a) + \alpha \sum_{j \in S} P_{i,j}(a) V_\pi(j) \right), \quad i \in S. \tag{1.1}$$

Is π' better than π, i.e. is $V_{\pi'}$ smaller than V_π? Before we reveal the answer, we will prove the following lemma.

Lemma 1.3.2. *For any Markov policy π the following holds:*

$$V_\pi(i) = R(i, \pi_0(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi_0(i)) V_{\pi \circ \theta}(j), \quad i \in S.$$

In particular, for any stationary policy π we have

$$V_\pi(i) = R(i, \pi(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi(i)) V_\pi(j), \quad i \in S.$$

Proof. We obtain

$$\begin{aligned}
V_\pi(i) &= \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_k^\pi, \pi_k(X_k^\pi)\right) \,\middle|\, X_0^\pi = i \right) \\
&= R(i, \pi_0(i)) + \alpha\, \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_{k+1}^\pi, \pi_{k+1}(X_{k+1}^\pi)\right) \,\middle|\, X_0^\pi = i \right) \\
&= R(i, \pi_0(i)) + \alpha \sum_{j \in S} \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_{k+1}^\pi, \pi_{k+1}(X_{k+1}^\pi)\right) \,\middle|\, X_0^\pi = i, X_1^\pi = j \right) \\
&\qquad\qquad \cdot \mathbb{P}\left(X_1^\pi = j \,\middle|\, X_0^\pi = i\right) \\
&= R(i, \pi_0(i)) + \alpha \sum_{j \in S} \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_{k+1}^\pi, \pi_{k+1}(X_{k+1}^\pi)\right) \,\middle|\, X_1^\pi = j \right) P_{i,j}(\pi_0(i)) \\
&= R(i, \pi_0(i)) + \alpha \sum_{j \in S} \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_k^{\pi \circ \theta}, (\pi \circ \theta)_k(X_k^{\pi \circ \theta})\right) \,\middle|\, X_0^{\pi \circ \theta} = j \right) P_{i,j}(\pi_0(i)) \\
&= R(i, \pi_0(i)) + \alpha \sum_{j \in S} V_{\pi \circ \theta}(j) P_{i,j}(\pi_0(i)).
\end{aligned}$$

□

The following theorem establishes that π' is indeed an improvement of π.

Theorem 1.3.3. *For every stationary policy π the following holds:*

$$V_{\pi'}(\cdot) \leq V_{\pi}(\cdot).$$

Proof. For every $n \in \mathbb{N}_0$ define the Markov policy $\pi^{(n)}$ by

$$\pi_k^{(n)}(i) = \begin{cases} \pi'(i) & \text{if} \quad k \leq n, \\ \pi(i) & \text{if} \quad k > n, \end{cases} \qquad i \in S, \quad k \in \mathbb{N}_0.$$

We will now show by induction that $V_{\pi^{(n)}}(\cdot) \leq V_{\pi}(\cdot)$ holds for every $n \in \mathbb{N}_0$. First we notice that $\pi^{(0)} \circ \theta = \pi$. Using Lemma 1.3.2 and the definition of π' in (1.1), we obtain

$$\begin{aligned} V_{\pi^{(0)}}(i) &= R(i, \pi'(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi'(i)) V_{\pi}(j) \\ &= \min_{a \in A} \left(R(i,a) + \alpha \sum_{j \in S} P_{i,j}(a) V_{\pi}(j) \right) \\ &\leq R(i, \pi(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi(i)) V_{\pi}(j) \\ &= V_{\pi}(i), \quad i \in S. \end{aligned}$$

Now suppose that $V_{\pi^{(m)}}(\cdot) \leq V_{\pi}(\cdot)$ holds for some $m \in \mathbb{N}_0$. Using the observation $\pi^{(m)} \circ \theta = \pi^{(m-1)}$ (if $m \geq 1$), Lemma 1.3.2, the induction hypothesis and the previous inequality, we obtain

$$\begin{aligned} V_{\pi^{(m+1)}}(i) &= R(i, \pi'(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi'(i)) V_{\pi^{(m)}}(j) \\ &\leq R(i, \pi'(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi'(i)) V_{\pi}(j) \\ &= V_{\pi^{(0)}}(i) \\ &\leq V_{\pi}(i), \quad i \in S, \end{aligned}$$

which concludes the induction.

We will need the following estimate, which follows from the definition of $\pi^{(n)}$

and boundedness of R:

$$\begin{aligned}
V_{\pi'}(i) - V_{\pi^{(n)}}(i) &= \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_k^{\pi'}, \pi'(X_k^{\pi'})\right) \,\middle|\, X_0^{\pi'} = i\right) \\
&\quad - \mathbb{E}\left(\sum_{k=0}^{\infty} \alpha^k R\left(X_k^{\pi^{(n)}}, \pi_k^{(n)}(X_k^{\pi^{(n)}})\right) \,\middle|\, X_0^{\pi^{(n)}} = i\right) \\
&= \mathbb{E}\left(\sum_{k=n+1}^{\infty} \alpha^k R\left(X_k^{\pi'}, \pi'(X_k^{\pi'})\right) \,\middle|\, X_0^{\pi'} = i\right) \\
&\quad - \mathbb{E}\left(\sum_{k=n+1}^{\infty} \alpha^k R\left(X_k^{\pi^n}, \pi_k^n(X_k^{\pi^n})\right) \,\middle|\, X_0^{\pi^n} = i\right) \\
&\leq 2 \sum_{k=n+1}^{\infty} \alpha^k M \\
&= \frac{2M\alpha^{n+1}}{1-\alpha}, \quad i \in S, \quad n \in \mathbb{N}_0.
\end{aligned}$$

To finish the proof, we note

$$V_{\pi'}(i) - V_{\pi}(i) = V_{\pi'}(i) - V_{\pi^{(n)}}(i) + V_{\pi^{(n)}}(i) - V_{\pi}(i) \leq \frac{2M\alpha^{n+1}}{1-\alpha}, \quad i \in S, \quad n \in \mathbb{N}_0,$$

and send n to ∞. □

The policy improvement algorithm is now defined as follows: take a stationary policy π^0 and then

$$\pi^{n+1}(\cdot) := (\pi^n)'(\cdot), \quad n \in \mathbb{N}_0.$$

If it happens for some $n \in \mathbb{N}_0$ that $V_{\pi^{n+1}}(\cdot) = V_{\pi^n}(\cdot)$, then V_{π^n} clearly satisfies the optimality equation and is therefore the value function (and π_n an optimal policy) by Theorem 1.3.1. In the case of the finite state space S this means that the algorithm always achieves an optimal policy (and usually this happens very quickly). In the general case the following theorem states that the sequence $\{V_{\pi^n}\}_{n\in\mathbb{N}_0}$ converges to the value function.

Theorem 1.3.4. *For any initial stationary policy π_0, the sequence $\{V_{\pi^n}\}_{n\in\mathbb{N}_0}$ converges uniformly to V.*

Proof. Since the cost function R is bounded, there exists $\tilde{M} \in \mathbb{R}$ such that $\sup_{i\in S} |V(i) - V_{\pi^0}(i)| \leq \tilde{M}$, i.e.

$$V_{\pi^0}(i) \leq V(i) + \tilde{M}, \quad i \in S.$$

We will prove that the statement

$$V_{\pi^n}(i) \leq V(i) + \tilde{M}\alpha^n, \quad i \in S$$

holds for every $n \in \mathbb{N}_0$, which implies the theorem. Assume that the statement holds for some $m \in \mathbb{N}_0$. Then we obtain

$$\begin{aligned}
V_{\pi^{m+1}}(i) &\overset{\text{Lemma 1.3.2}}{=} R(i, \pi^{m+1}(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi^{m+1}(i)) V_{\pi^{m+1}}(j) \\
&\overset{\text{Thm. 1.3.3}}{\leq} R(i, \pi^{m+1}(i)) + \alpha \sum_{j \in S} P_{i,j}(\pi^{m+1}(i)) V_{\pi^m}(j) \\
&\overset{\text{PIA}}{=} \min_{a \in A} \left(R(i,a) + \alpha \sum_{j \in S} P_{i,j}(a) V_{\pi^m}(j) \right) \\
&\overset{\text{I.H.}}{\leq} \min_{a \in A} \left(R(i,a) + \alpha \sum_{j \in S} P_{i,j}(a) (V(j) + \tilde{M}\alpha^m) \right) \\
&\overset{\text{Thm. 1.3.1}}{=} V(i) + \tilde{M}\alpha^{m+1}, \quad i \in S,
\end{aligned}$$

which conludes the induction and hence the proof. □

1.3.2 Literature review

Since Howard's book [12] containing the policy improvement (or policy iteration) algorithm was published in 1960, a lot of work has been done on this subject. A survey of approximate policy iteration methods for finite state, discrete time, stochastic dynamic programming problems is given in [4]. The algorithm has proved to be useful in deterministic control theory, too (see e.g. [8] and [31]). Nevertheless, it has probably been applied most often to Markov decision processes in various settings (see [7], [10], [11], [21], [25], [20], [24], [28], [29] and [32]).

Most of the settings of the above papers are discrete, but some are continuous (or general) to a certain extent. For example, article [7] deals with continuous time Markov decision processes on a fairly general state space, but the policy improvement algorithm is only proved to work in the special case of finite action space. Paper [32] removes this restriction, but even there the controlled processes are not continuous. We have not been able to find any mention of the policy improvement algorithm for controlled diffusion processes or any processes with continuous paths.

Chapter 2

Mirror and synchronous couplings of geometric Brownian motions

2.1 Introduction

Recall that in Subsection 1.2.2 we solved the following finite horizon problem for any $T > 0$:

$$\text{minimise/maximise} \quad \mathbb{P}(\tau(V) > T) \quad \text{over all Brownian motions } V,$$

where $\tau(V)$ is the first meeting time of the processes $x + B$ and $y + V$, $x, y \in \mathbb{R}$, and B and V are Brownian motions with respect to the same filtration (and B is considered to be fixed). We proved that a solution is given by the mirror coupling (i.e. an optimal Brownian motion is $V = -B$) in the case of minimisation, and by the synchronous coupling (i.e. an optimal Brownian motion is $V = B$) in the case of maximisation. Since the solution is the same for every $T > 0$, the two couplings must also solve the problems obtained by replacing the expression $\mathbb{P}(\tau(V) > T)$ above by

$$\int_0^\infty \mathrm{e}^{-qt}\, \mathbb{P}\big(\tau(V) > t\big)\, \mathrm{d}t, \quad (q > 0),$$

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{P}(\tau(V) > t)\, \mathrm{d}t,$$

and

$$\liminf_{t\to\infty} \frac{1}{t} \log\left(\mathbb{P}\left(\tau(V) > t\right)\right),$$

which are called the infinite horizon problem, ergodic average problem and exponential efficiency problem, respectively.

It is natural to investigate the analogous problems for other processes, especially geometric Brownian motions. In this case $\tau(V)$ is the first meeting time of the processes X and $Y(V)$, where X is a geometric Brownian motion started at x and driven by B, and $Y(V)$ is a geometric Brownian motion started at y and driven by V. Since X_t and $Y(V)_t$ are at any time t given by explicit deterministic functions of B_t and V_t respectively, we might expect that the mirror and synchronous couplings of B and V will again be optimal in the finite horizon problem. However, as we shall see, this turns out to be false in general. Consequently, the other problems are not trivial, and we will look into them, too.

An application in mathematical finance of the finite horizon problem considered in the present chapter can be described as follows. Assume that the performance of a portfolio manager is assessed at some fixed future time (e.g. one year from now) with respect to a benchmark security (e.g. some equity index), which evolves as a geometric Brownian motion X. Put differently, the remuneration of the manager depends on whether their portfolio, which evolves as $Y(V)$, exceeds the benchmark X in normalised terms. Assume also that the manager's mandate stipulates that, over the same time horizon, their portfolio may not exceed a prespecified amount of realised variance. Both of these assumptions are realistic and are used extensively in practice, since the investor wants to beat the index but cannot tolerate arbitrary amounts of volatility in the meantime (e.g. investors like pension funds routinely stipulate such realised variance conditions). Imagine now a situation where the manager has a given amount of time, say T, before the evaluation of their performance, but is behind the benchmark by a certain amount. The question of how to trade in such a way as to minimise the probability of not catching up with the benchmark before T, and to achieve this without taking unnecessary bets that would increase the realised volatility of the portfolio, is precisely the question of the stochastic minimisation of the first meeting time between X and $Y(V)$ (recall that the expected quadratic variation of $Y(V)$, i.e. the realised variance of the manager's portfolio, does not depend on the choice of Brownian motion V).

In the next section we describe the setting and basic notation, which remain throughout the chapter. We also state a lemma from stochastic analysis, which enables us to apply Bellman's principle, on which some of the proofs are based. Then each of the four sections deals with one of the above problems. In Sections 2.3 and 2.5 we prove that the mirror and synchronous couplings always solve the infinite horizon and ergodic average problem, respectively. In Sections 2.4 and 2.6 we prove

that two couplings are not always optimal for the finite horizon and exponential efficiency problem, respectively, and provide a characterisation of when exactly this happens.

2.2 Setting and notation

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions that is rich enough to support an $(\mathcal{F}_t)$-Brownian motion $B = (B_t)_{t\geq 0}$, and set

$$\mathcal{V} := \{V = (V_t)_{t\geq 0};\ V \text{ is an } (\mathcal{F}_t)\text{-Brownian motion}\}.$$

The following well-known lemma will come in useful. For its proof, see e.g. [15, Lemma 2.1].

Lemma 2.2.1. *For any Brownian motion $V \in \mathcal{V}$, there exists an $(\mathcal{F}_t)$-Brownian motion $W \in \mathcal{V}$ and a process $C = (C_t)_{t\geq 0}$ such that B and W are independent, C is progressively measurable with $-1 \leq C_t \leq 1$ for all $t \geq 0$ $\mathbb{P}$-a.s., and the following representation holds:*

$$V_t = \int_0^t C_s \, \mathrm{d}B_s + \int_0^t \sqrt{1 - C_s^2} \, \mathrm{d}W_s, \quad t \geq 0.$$

Remark 2.2.2. The proof of this lemma requires the existence of a Brownian motion $B^\perp \in \mathcal{V}$ that is independent of B. If our probability space did not support such a Brownian motion, we could enlarge it, which would only increase the set $\mathcal{V}$. This means that if B and $-B$ are optimal in the new problem, they also have to be optimal in the original problem. Therefore we can assume that $B^\perp$ exists.

For any $V \in \mathcal{V}$, let $X = (X_t)_{t\geq 0}$ and $Y(V) = (Y_t(V))_{t\geq 0}$ be geometric Brownian motions given by the following stochastic differential equations:

$$X_t = x + \int_0^t X_s \left(\sigma_1 \, \mathrm{d}B_s + a_1 \, \mathrm{d}s\right), \quad Y_t(V) = y + \int_0^t Y_s(V) \left(\sigma_2 \, \mathrm{d}V_s + a_2 \, \mathrm{d}s\right), \tag{2.1}$$

where

$$x, y > 0, \quad a_1, a_2 \in \mathbb{R}, \quad \text{and} \quad \sigma_1, \sigma_2 \in \mathbb{R} \quad \text{such that} \quad \sigma_1\sigma_2 > 0. \tag{2.2}$$

Define the following constants:

$$\mu := a_2 - a_1 + \sigma_1^2/2 - \sigma_2^2/2 \quad \text{and} \quad \sigma_\pm := \sigma_2 \pm \sigma_1. \tag{2.3}$$

Note that (2.2) implies $|\sigma_+| > |\sigma_-|$. The symbol $\pm$ denotes either $+$ or $-$. If $\pm$ and $\mp$ appear in the same expression, then they simultaneously denote either $+$ and $-$, or $-$ and $+$.

Define $\tau(V)$ as the first meeting time of the two processes in (2.1), i.e.

$$\tau(V) := \inf\{t \geq 0;\ X_t = Y_t(V)\} \qquad (\inf \emptyset := \infty).$$

The random variable $\tau(V)$ is zero when the two processes start at the same point and positive $\mathbb{P}$-a.s. otherwise. Since the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions, $\tau(V)$ is an $(\mathcal{F}_t)$-stopping time. Since X and $Y(V)$ are geometric Brownian motions, this stopping time has the following useful representation:

$$\tau(V) = \inf\left\{t \geq 0;\ \log\left(\frac{x}{y}\right) = \sigma_2 V_t - \sigma_1 B_t + \mu t\right\}. \tag{2.4}$$

2.3 Infinite horizon problems

2.3.1 The problems and main theorem

For any $q > 0$, we consider the following two problems: find $V^{\inf} \in \mathcal{V}$ and $V^{\sup} \in \mathcal{V}$ (if they exist) such that

$$\inf_{V\in\mathcal{V}} \int_0^\infty \mathrm{e}^{-qt}\,\mathbb{P}\big(\tau(V) > t\big)\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-qt}\,\mathbb{P}\big(\tau\big(V^{\inf}\big) > t\big)\,\mathrm{d}t \tag{qInf}$$

and

$$\sup_{V\in\mathcal{V}} \int_0^\infty \mathrm{e}^{-qt}\,\mathbb{P}\big(\tau(V) > t\big)\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-qt}\,\mathbb{P}\big(\tau(V^{\sup}) > t\big)\,\mathrm{d}t. \tag{qSup}$$

An application of Fubini's theorem (or integration by parts for Riemann-Stieltjes integral) yields

$$\int_0^\infty \mathrm{e}^{-rt}\,\mathbb{P}(\tau > t)\,\mathrm{d}t = \frac{1 - \mathbb{E}(\mathrm{e}^{-r\tau})}{r}$$

for any nonnegative random variable τ and $r > 0$. Therefore Problems (qInf) and (qSup) are equivalent to finding $V^{(+)} \in \mathcal{V}$ and $V^{(-)} \in \mathcal{V}$ respectively, such that

$$\sup_{V\in\mathcal{V}} \pm\mathbb{E}\left(\mathrm{e}^{-q\tau(V)}\right) = \pm\mathbb{E}\left(\mathrm{e}^{-q\tau\left(V^{(\pm)}\right)}\right). \tag{q$\pm$}$$

Note also that if e_q is an exponential random variable with $\mathbb{E}(e_q) = \frac{1}{q}$, independent of the filtration $(\mathcal{F}_t)_{t\geq 0}$, then Problems (qInf) and (qSup) are equivalent to minimising and maximising $\mathbb{P}(\tau(V) > e_q)$ over $V \in \mathcal{V}$, respectively.

The following theorem holds.

Theorem 2.3.1. *A solution to Problem* (q±) *is (for any $q > 0$) given by*

$$V^{(\pm)} = \mp B.$$

Remark 2.3.2. Observe that by Theorem 2.3.1, the mirror coupling ($V^{(+)} = -B$) solves Problem (qInf) and the synchronous coupling ($V^{(-)} = +B$) is the solution to Problem (qSup). Note that the solution depends neither on the parameters in (2.2) nor on the discount rate q.

2.3.2 Proof

Note that, due to the symmetry in Problem (q±), we may assume without loss of generality that the starting points x, y in (2.1)–(2.2) satisfy $(x, y) \in D$, where the set D is given by

$$D := \{(a, b) \in \mathbb{R}^2;\ a \geq b > 0\}. \tag{2.5}$$

Fix $q > 0$ and define the following function, closely related to the right-hand side in Problem (q±):

$$\Psi^{(\pm)}(x, y) := \mathbb{E}_{x,y}\left(\mathrm{e}^{-q\tau(\mp B)}\right), \qquad (x, y) \in D. \tag{2.6}$$

The proof of Theorem 2.3.1 is in two steps: we first establish sufficient conditions for a function $\Psi : D \to \mathbb{R}_+$ implying that $\pm\Psi$ is equal to the right-hand side in Problem (q±) (Lemmas 2.3.3 and 2.3.4), and then prove that $\Psi^{(\pm)}$ in (2.6) satisfies these conditions (Lemma 2.3.5). Throughout the thesis we denote $\mathbb{R}_+ := [0, \infty)$.

For any measurable function $\Psi : D \to \mathbb{R}_+$ and Brownian motion $V \in \mathcal{V}$, consider the process $U(V, \Psi) = (U_t(V, \Psi))_{t\in[0,\infty)}$ defined by

$$U_t(V, \Psi) := \mathrm{e}^{-q(t\wedge\tau(V))}\Psi(X_{t\wedge\tau(V)}, Y_{t\wedge\tau(V)}(V)) \tag{2.7}$$

(here and in the rest of the thesis we denote $s \wedge t := \min(\{s, t\})$). Then the following lemma (a suitable version of Bellman's principle) holds.

Lemma 2.3.3. *Let $\Psi : D \to \mathbb{R}_+$ be a bounded continuous function satisfying $\Psi(x, x) = 1$ for all $x > 0$. If, for every $(x, y) \in D$, the process $\pm U(V, \Psi)$ is a $\mathbb{P}_{x,y}$-supermartingale for all $V \in \mathcal{V}$ and $U(\mp B, \Psi)$ is a $\mathbb{P}_{x,y}$-martingale, then $V^{(\pm)} = \mp B$ solves Problem* (q±).

Proof. Since $X_{\tau(V)} = Y_{\tau(V)}(V)$ $\mathbb{P}_{x,y}$-a.s. on the event $\{\tau(V) < \infty\}$ for any $V \in \mathcal{V}$, Ψ is continuous and bounded, $\Psi(x, x) = 1$ holds for any $x > 0$ and $q > 0$, the

supermartingale property and the Dominated Convergence Theorem imply

$$\pm\mathbb{E}_{x,y}\left(\mathrm{e}^{-q\tau(V)}\right) = \mathbb{E}_{x,y}\left(\pm U_{\tau(V)}(V,\Psi)\mathbb{I}_{\{\tau(V)<\infty\}}\right) \leq \mathbb{E}_{x,y}\left(\pm U_0(V,\Psi)\right) = \pm\Psi(x,y),$$

for all $(x,y) \in D$ and $V \in \mathcal{V}$ ($\mathbb{I}_{\{\cdot\}}$ denotes the indicator of the event $\{\cdot\}$). Since $U(\mp B,\Psi)$ is a martingale, for $V^{(\pm)} = \mp B$ this inequality becomes an equality and the lemma follows. □

Our next task is to establish a verification lemma for Problem (q±). Let D° be the interior (in $\mathbb{R}^2$) of the set D defined in (2.5). For any function $f \in \mathcal{C}^{2,2}(D^\circ)$ we define the function $\mathcal{L}^{(\pm)}f$ by the formula

$$\begin{aligned}&\left(\mathcal{L}^{(\pm)}f\right)(x,y) := \\ &\left(a_1 x f_x + a_2 y f_y + \frac{1}{2}\sigma_1^2 x^2 f_{xx} + \frac{1}{2}\sigma_2^2 y^2 f_{yy} \mp \sigma_1\sigma_2 xy f_{xy} - qf\right)(x,y),\end{aligned} \tag{2.8}$$

where $(x,y) \in D^\circ$ and f_x, f_y, f_{xx}, f_{yy} and f_{xy} denote the partial derivatives of f. For any function $\Psi : D \to \mathbb{R}_+$ such that $\Psi \in \mathcal{C}^{2,2}(D^\circ)$, and Brownian motion $V \in \mathcal{V}$, the local martingale $M(V,\Psi) = (M_t(V,\Psi))_{t\in[0,\infty)}$, given by

$$\begin{aligned}M_t(V,\Psi) := &\int_0^{t\wedge\tau(V)} \mathrm{e}^{-qs}\sigma_1 X_s \Psi_x(X_s, Y_s(V))\,\mathrm{d}B_s \\ &+ \int_0^{t\wedge\tau(V)} \mathrm{e}^{-qs}\sigma_2 Y_s(V)\Psi_y(X_s, Y_s(V))\,\mathrm{d}V_s,\end{aligned} \tag{2.9}$$

is well-defined.

Lemma 2.3.4. *Assume the following hold:*

(I) $\Psi : D \to \mathbb{R}_+$ *is a bounded continuous function with* $\Psi(x,x) = 1$ *for all* $x > 0$*;*

(II) $\Psi \in \mathcal{C}^{2,2}(D^\circ)$ *and, in the interior* D°, $\Psi_{xy} \leq 0$ *and* $\mathcal{L}^{(\pm)}\Psi = 0$*;*

(III) $M(V,\Psi)$ *is a* $\mathbb{P}_{x,y}$*-martingale for all* $(x,y) \in D$ *and* $V \in \mathcal{V}$*.*

Then for any $(x,y) \in D$, $V \in \mathcal{V}$, *the process* $\pm U(V,\Psi)$, *defined in* (2.7), *is a* $\mathbb{P}_{x,y}$*-supermartingale and* $U(\mp B,\Psi)$ *is a* $\mathbb{P}_{x,y}$*-martingale.*

Proof. The definition of X and $Y(V)$ in (2.1) and Lemma 2.2.1 imply $\mathrm{d}[X,Y(V)]_t = C_t\sigma_1 X_t \sigma_2 Y_t(V)\,\mathrm{d}t$, where $C = (C_t)_{t\in[0,\infty)}$ is $(\mathcal{F}_t)$-adapted and $\mathbb{P}(C_t \in [-1,1]) = 1$ for all $t \in [0,\infty)$. Itô's lemma, the assumptions in Lemma 2.3.4 and definition (2.7)

of $U(V,\Psi)$ yield

$$\begin{aligned}&\pm U_t(V,\Psi)\\&= \pm\Psi(x,y) \pm M_t(V,\Psi) + \int_0^{t\wedge\tau(V)} \mathrm{e}^{-qs}\sigma_1\sigma_2(1\pm C_s)X_sY_s(V)\Psi_{xy}(X_s,Y_s(V))\,\mathrm{d}s\end{aligned}$$

for all $(x,y)\in D$ and $V\in\mathcal{V}$. Since X, $Y(V)$ and $1\pm C$ are non-negative processes and, by assumption (2.2), we have $\sigma_1\sigma_2>0$, the integrand in the representation of $\pm U(V,\Psi)$ is non-positive, making $\pm U(V,\Psi)$ a $\mathbb{P}_{x,y}$-supermartingale. For $\mp B$ we have $C_s=\mp 1$ for every $s\geq 0$, which implies that $U(\mp B,\Psi)$ is a $\mathbb{P}_{x,y}$-martingale. □

Note the following equivalence:

$$\mathbb{P}_{x,y}(\tau(\mp B)=\infty)=1 \quad \text{for all } (x,y)\in D^\circ \iff \mp=+,\ \sigma_2=\sigma_1,\ a_2\leq a_1. \tag{2.10}$$

It is clear that under condition (2.10) Theorem 2.3.1 holds. Lemmas 2.3.3 and 2.3.4 imply that in order to establish Theorem 2.3.1 in general, it is sufficient to prove that, when (2.10) fails, the function $\Psi^{(\pm)}: D\to\mathbb{R}_+$ in (2.6) satisfies the assumptions of Lemma 2.3.4. More precisely, the following lemma holds.

Lemma 2.3.5. *If for some $(x,y)\in D^\circ$ we have $\mathbb{P}_{x,y}(\tau(\mp B)=\infty)<1$, Assumptions (I)–(III) of Lemma 2.3.4 hold for the function $\Psi^{(\pm)}: D\to\mathbb{R}_+$ in* (2.6).

Proof. Under the assumption of the lemma, the following representation holds:

$$\Psi^{(\pm)}(x,y)=\left(\frac{y}{x}\right)^{k_\pm},\quad (x,y)\in D, \tag{2.11}$$

where

$$k_\pm := \begin{cases} -\mu/\sigma_\pm^2+\sqrt{(\mu/\sigma_\pm^2)^2+2q/\sigma_\pm^2} & \text{if}\quad \sigma_\pm\neq 0,\\ q/\mu & \text{if}\quad \sigma_\pm=0,\end{cases}$$

and $\sigma_\pm$ and μ are defined in (2.3). Since, by assumption, the condition on the right-hand side in (2.10) is not satisfied, the equality $\sigma_\pm=0$ implies $\mu>0$, making $k_\pm$ a well-defined real number. Formula (2.11) follows from the fact that $\tau(\mp B)$ equals the first passage time of the Brownian motion with drift, $(\mp\sigma_\pm B_t+\mu t)_{t\in[0,\infty)}$, over the level $\log\left(\frac{x}{y}\right)$. The Laplace transform of this random time is given in [5, p. 295] and amounts to the right-hand side of (2.11).

Assumption (I) in Lemma 2.3.4 follows from (2.11). Furthermore it is clear that $\Psi^{(\pm)}\in\mathcal{C}^{2,2}(D^\circ)$. The formula in (2.11) and some simple calculations imply

that for $(x,y) \in D^\circ$ the following holds:

$$\Psi_x^{(\pm)}(x,y) = -\frac{k_\pm}{x}\,\Psi^{(\pm)}(x,y), \quad \Psi_y^{(\pm)}(x,y) = \frac{k_\pm}{y}\,\Psi^{(\pm)}(x,y), \tag{2.12}$$

and

$$\Psi_{xy}^{(\pm)}(x,y) = -\frac{k_\pm^2}{xy}\,\Psi^{(\pm)}(x,y) \le 0, \quad \left(\mathcal{L}^{(\pm)}\Psi^{(\pm)}\right)(x,y) = 0.$$

Hence assumption (II) of Lemma 2.3.4 is also satisfied. The equalities in (2.12) and the definition in (2.9) of the local martingale $M(V,\Psi^{(\pm)})$ imply that the integrands in the stochastic integrals are bounded processes and therefore square integrable. Hence $M(V,\Psi^{(\pm)})$ is a $\mathbb{P}_{x,y}$-martingale for all $(x,y) \in D$ and $V \in \mathcal{V}$ and assumption (III) of Lemma 2.3.4 also holds. □

2.4 Finite horizon problems

2.4.1 The problems and main theorem

For any $T > 0$, consider the following problem(s):

$$\text{find } V^{(\pm)} \in \mathcal{V} \text{ such that } \inf_{V\in\mathcal{V}} \pm\mathbb{P}_{x,y}\big(\tau(V) > T\big) = \pm\mathbb{P}_{x,y}\big(\tau\big(V^{(\pm)}\big) > T\big). \tag{T$\pm$}$$

Unlike in the infinite horizon problem, the mirror and synchronous couplings are not always optimal. The following theorem characterises precisely when this is the case. Recall that μ and $\sigma_\pm$ are given in (2.3), and D in (2.5).

Theorem 2.4.1. *The following holds for any $T > 0$ and $(x,y) \in D^\circ$:*

(a) if $\mu > 0$ and $\sigma_\pm \neq 0$, then $V^{(\pm)} = \mp B$ does NOT solve Problem (T$\pm$)*;*

(b) if $\mu \le 0$, then $V^{(\pm)} = \mp B$ solves Problem (T$\pm$)*.*

Remark 2.4.2. In the case $\mu > 0$ and $\sigma_\pm = 0$ we have $\pm = -$, $\sigma_1 = \sigma_2$ and $\Phi^{(-)}(x,y,t) = \mathbb{I}_{\{t\mu<\log(x/y)\}}$ for all $(x,y) \in D^\circ$, $t \in [0,T]$ (recall (2.4)), which implies that the synchronous coupling is suboptimal if and only if $T \ge \frac{1}{\mu}\log\left(\frac{x}{y}\right)$.

Remark 2.4.3. Intuition behind this theorem follows from the representation in (2.4): starting from 0, we want to hit a positive level in a given time with as high probability as possible (for the minimisation problem); when the drift is against us (i.e. non-positive), we are desperate and therefore choose the maximal variance at every moment, which corresponds to the mirror coupling; when the drift is helping us, neither of the extreme solutions is optimal. The same representation implies that

in the case $0 < x < y$, the theorem still holds if μ gets the opposite sign in the statement.

2.4.2 Proof

Define the set $E := D \times [0,T]$ and recall that the value function for Problem (T$\pm$) is defined by

$$F(x,y,t) := \inf_{V \in \mathcal{V}} \pm \mathbb{P}_{x,y}(\tau(V) > t), \qquad (x,y,t) \in E. \tag{2.13}$$

Define also

$$\Phi^{(\pm)}(x,y,t) := \mathbb{P}_{x,y}\left(\tau(\mp B) > t\right), \qquad (x,y,t) \in E, \tag{2.14}$$

and $\mathcal{A}^{(\pm)} f$ for any $f \in \mathcal{C}^{2,2,1}(E^\circ)$ (E° is the interior of E in $\mathbb{R}^3$) by the formula

$$\begin{aligned}&\left(\mathcal{A}^{(\pm)} f\right)(x,y,t) := \\ &\left(a_1 x f_x + a_2 y f_y + \frac{1}{2}\sigma_1^2 x^2 f_{xx} + \frac{1}{2}\sigma_2^2 y^2 f_{yy} \mp \sigma_1\sigma_2 xy f_{xy} - f_t\right)(x,y,t),\end{aligned}$$

where $(x,y,t) \in E^\circ$ and f_x, f_y, f_t, etc., denote the partial derivatives of f. For any sufficiently smooth function $\Phi : E \to \mathbb{R}_+$ and any Brownian motion $V \in \mathcal{V}$, we define the local martingale $N(V,\Phi) = (N_t(V,\Phi))_{t\in[0,T]}$ by

$$\begin{aligned}N_t(V,\Phi) := &\int_0^{t\wedge\tau(V)} \sigma_1 X_s \Phi_x(X_s, Y_s(V), T-s)\,\mathrm{d}B_s \\ &+ \int_0^{t\wedge\tau(V)} \sigma_2 Y_s(V) \Phi_y(X_s, Y_s(V), T-s)\,\mathrm{d}V_s.\end{aligned} \tag{2.15}$$

The following proposition provides the key ingredient in the proof of Theorem 2.4.1.

Proposition 2.4.4. *Let a bounded function $\Phi : E \to \mathbb{R}_+$ satisfy:*

(i) $\Phi(x,x,t) = 0$ for all $x > 0$ and $t \in [0,T]$, and $\Phi(x,y,0) = 1$ for all $(x,y) \in D^\circ$;

(ii) $\Phi \in \mathcal{C}^{2,2,1}(E^\circ)$ and, in the interior E°, the equality $\mathcal{A}^{(\pm)}\Phi = 0$ holds;

(iii) $N(V,\Phi)$ is a $\mathbb{P}_{x,y}$-martingale for all $(x,y) \in D$ and $V \in \mathcal{V}$.

Then the following equivalence holds:

$$\Phi_{xy} \geq 0 \text{ on } E^\circ \iff V^{(\pm)} = \mp B \text{ solves Problem (T}\pm\text{) and } \pm\Phi \text{ is its value funct.}$$

Proof. ($\Rightarrow$): The proof of this implication is analogous to that of Lemmas 2.3.3 (Bellman's principle) and 2.3.4 (submartingale property) in Section 2.3. The process $\pm U(V,\Phi) = (\pm U_t(V,\Phi))_{t\in[0,T]}$, given by

$$U_t(V,\Phi) := \Phi(X_{t\wedge\tau(V)}, Y_{t\wedge\tau(V)}(V), T-t), \tag{2.16}$$

is a $\mathbb{P}_{x,y}$-submartingale for any $V \in \mathcal{V}$ and $(x,y) \in D$ (proof as in Lemma 2.3.4). For any $t \in [0,T]$, the boundary conditions in assumption (i) imply

$$U_t(V,\Phi) = U_{\tau(V)}(V,\Phi) = 0 \quad \mathbb{P}_{x,y}\text{-a.s. on } \{t \geq \tau(V)\}.$$

Hence, for any $(x,y) \in D$ and $V \in \mathcal{V}$, the submartingale property yields the inequality

$$\begin{aligned}
\pm\mathbb{P}_{x,y}\left(\tau(V) > T\right) &= \mathbb{E}_{x,y}\left(\pm U_T(V,\Phi)\mathbb{I}_{\{\tau(V)>T\}}\right) \\
&= \mathbb{E}_{x,y}\left(\pm U_T(V,\Phi)\right) \\
&\geq \pm\mathbb{E}_{x,y}U_0(V,\Phi) \\
&= \pm\Phi(x,y,T).
\end{aligned}$$

As in Lemma 2.3.3, this establishes the implication (note that, unlike in Lemma 2.3.3, in this case we do not need, and in fact do not have, the continuity of Φ on E).
($\Leftarrow$): Assume that there exists $(x_0,y_0,T_0) \in E^\circ$ such that $\Phi_{xy}(x_0,y_0,T_0) < 0$, and that $\pm\Phi$ is the value function of Problem (T$\pm$). Bellman's principle implies that the process $\pm U(V,\Phi)$, defined in (2.16), is a $\mathbb{P}_{x,y}$-submartingale for any $V \in \mathcal{V}$ and $(x,y) \in D$. Using our assumption, we will construct a Brownian motion $\tilde{V}^{(\pm)} \in \mathcal{V}$ such that $\pm U(\tilde{V}^{(\pm)},\Phi)$ fails to be a $\mathbb{P}_{x,y}$-submartingale (for any pair $(x,y) \in D^\circ$), which will imply the proposition.

The continuity of Φ_{xy} implies that there exists $r > 0$, such that Φ_{xy} is strictly negative on the set $K_2 := H_2 \times [T_0 - 2r, T_0 + 2r] \subset E^\circ$, where

$$H_2 := [x_0 - 2r, x_0 + 2r] \times [y_0 - 2r, y_0 + 2r].$$

Let

$$H_1 := [x_0 - r, x_0 + r] \times [y_0 - r, y_0 + r]$$

and define the stopping times $\tau_1^{(\pm)}$ and $\tau_2^{(\pm)}$ by:

$$\tau_1^{(\pm)} := \inf\{t \in [0,T];\ (X_t, Y_t(\mp B)) \in H_1\}, \quad (\inf\emptyset := T),$$

$$\tau_2^{(\pm)} := \inf\{t \in [\tau_1, T];\ (X_t, Y_t(\pm B)) \notin H_2\}, \quad (\inf\emptyset := T).$$

Note that $\tau_1^{(\pm)} \leq \tau_2^{(\pm)} \leq T$ $\mathbb{P}_{x,y}$-a.s. and $\mathbb{P}_{x,y}(\tau_1^{(\pm)} < \tau_2^{(\pm)}) > 0$ (there is a slight abuse of notation in the definition of $\tau_2^{(\pm)}$ as it is assumed that the process $Y(\pm B)$, defined in (2.1), is driven by the Brownian motion $\pm B$ as indicated, but started at the random time $\tau_1^{(\pm)}$ and point $Y_{\tau_1^{(\pm)}}(\mp B)$; ditto for X).

Define the process $\tilde{V}^{(\pm)} = (\tilde{V}_t^{(\pm)})_{t\in[0,\infty)}$ by the following formula:

$$\tilde{V}_t^{(\pm)} := \int_0^t \left(\mp\mathbb{I}_{\{s<\tau_1^{(\pm)}\}} \pm \mathbb{I}_{\{\tau_1^{(\pm)}\leq s<\tau_2^{(\pm)}\}} \mp \mathbb{I}_{\{s\geq\tau_2^{(\pm)}\}}\right) \mathrm{d}B_s.$$

Note that $\tilde{V}^{(\pm)}$ is an $(\mathcal{F}_t)$-Brownian motion by Lévy's characterisation theorem. Itô's formula on the stochastic interval $[\tau_1^{(\pm)}, \tau_2^{(\pm)}]$ and assumptions (i)–(iii) in the proposition imply the following representation:

$$\begin{aligned}\mathbb{E}_{x,y}\left(\pm U_{\tau_2^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi)\,\middle|\,\mathcal{F}_{\tau_1^{(\pm)}}\right) &= \pm U_{\tau_1^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi) \\ &\quad + \mathbb{E}_{x,y}\left(\int_{\tau_1^{(\pm)}}^{\tau_2^{(\pm)}} 2\sigma_1\sigma_2 X_s Y_s(\tilde{V}^{(\pm)})\Phi_{xy}(X_s, Y_s(\tilde{V}^{(\pm)}), T-s)\,\mathrm{d}s\,\middle|\,\mathcal{F}_{\tau_1^{(\pm)}}\right).\end{aligned}$$

The event

$$A := \{\tau_1^{(\pm)} \in (T_0 - r, T_0 + r), \tau(\tilde{V}^{(\pm)}) > T_0 + 2r\}$$

has a strictly positive probability and the integrand under the conditional expectation is strictly negative on this event. We therefore find

$$\mathbb{E}_{x,y}\left(\pm U_{\tau_2^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi)\,\middle|\,\mathcal{F}_{\tau_1^{(\pm)}}\right) < \pm U_{\tau_1^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi) \quad \text{on} \quad A$$

$\mathbb{P}_{x,y}$-a.s. This inequality contradicts the $\mathbb{P}_{x,y}$-a.s. inequality

$$\mathbb{E}_{x,y}\left(\pm U_{\tau_2^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi)\,\middle|\,\mathcal{F}_{\tau_1^{(\pm)}}\right) \geq \pm U_{\tau_1^{(\pm)}}(\tilde{V}^{(\pm)}, \Phi),$$

which follows from the optional sampling theorem applied to the bounded $\mathbb{P}_{x,y}$-submartingale $U(\tilde{V}^{(\pm)}, \Phi)$. This concludes the proof. □

We will now apply Proposition 2.4.4 to study the question of whether $\pm\Phi^{(\pm)}$, defined in (2.14), is the value function for Problem (T±).

Lemma 2.4.5. *Assume $\sigma_\pm \neq 0$. Then assumptions (i)–(iii) of Proposition 2.4.4*

hold for the function $\Phi^{(\pm)}$ *defined in* (2.14). *Furthermore, we have*

$$\Phi^{(\pm)}_{xy}(x,y,t) = \frac{2\log\left(\frac{x}{y}\right) - 4\mu t}{xy(|\sigma_\pm|\sqrt{t})^3}\, n\left(\frac{\log\left(\frac{x}{y}\right) - \mu t}{|\sigma_\pm|\sqrt{t}}\right) + \frac{4\mu^2}{xy\sigma_\pm^4}\left(\frac{x}{y}\right)^{\frac{2\mu}{\sigma_\pm^2}} N\left(\frac{-\log\left(\frac{x}{y}\right) - \mu t}{|\sigma_\pm|\sqrt{t}}\right)$$

for all $(x,y) \in D^\circ$ *and* $t > 0$, *where* $N(\cdot)$ *is the standard normal distribution function and* $n(\cdot)$ *is its density.*

Proof. The explicit formula for the distribution of the running maximum of a Brownian motion with drift (see e.g. [5, p. 250]) yields the following representation of the function in (2.14):

$$\Phi^{(\pm)}(x,y,t) = h^{(\pm)}\left(\log\left(\frac{x}{y}\right), t\right) \quad \text{for} \quad (x,y) \in D, \tag{2.17}$$

where, for any $z \geq 0$ and $s > 0$, we define

$$h^{(\pm)}(z,s) := N\left(\frac{z-\mu s}{|\sigma_\pm|\sqrt{s}}\right) - \exp\left(\frac{2\mu z}{\sigma_\pm^2}\right) N\left(\frac{-z-\mu s}{|\sigma_\pm|\sqrt{s}}\right). \tag{2.18}$$

Simple (but tedious) calculations using this representation yield the properties required in assumptions (i)–(iii) of Proposition 2.4.4. Indeed, note that the partial derivatives $h^{(\pm)}_z$, $h^{(\pm)}_{zz}$ and $h^{(\pm)}_s$ take the following form (recall $n'(x) = -xn(x)$):

$$\begin{aligned}
h^{(\pm)}_z(z,s) &= \frac{2}{|\sigma_\pm|\sqrt{s}}\, n\left(\frac{z-\mu s}{|\sigma_\pm|\sqrt{s}}\right) - \frac{2\mu}{\sigma_\pm^2}\exp\left(\frac{2\mu z}{\sigma_\pm^2}\right) N\left(\frac{-z-\mu s}{|\sigma_\pm|\sqrt{s}}\right),\\
h^{(\pm)}_{zz}(z,s) &= \frac{4s\mu - 2z}{(|\sigma_\pm|\sqrt{s})^3}\, n\left(\frac{z-\mu s}{|\sigma_\pm|\sqrt{s}}\right) - \frac{4\mu^2}{\sigma_\pm^4}\exp\left(\frac{2\mu z}{\sigma_\pm^2}\right) N\left(\frac{-z-\mu s}{|\sigma_\pm|\sqrt{s}}\right),\\
h^{(\pm)}_s(z,s) &= -\frac{z}{|\sigma_\pm| s^{3/2}}\, n\left(\frac{z-\mu s}{|\sigma_\pm|\sqrt{s}}\right).
\end{aligned}$$

These formulae and the representation in (2.17) imply the formula for $\Phi^{(\pm)}_{xy}(x,y,t)$, as well as assumptions (i) and (ii) of Proposition 2.4.4. The martingale property of the process in (2.15) (i.e. assumption (iii) in Proposition 2.4.4) follows by Itô's isometry from the fact that both functions

$$x\Phi^{(\pm)}_x(x,y,t) = h^{(\pm)}_z\left(\log\left(\frac{x}{y}\right), t\right) \quad \text{and} \quad y\Phi^{(\pm)}_y(x,y,t) = -h^{(\pm)}_z\left(\log\left(\frac{x}{y}\right), t\right)$$

are bounded on E. This completes the proof of the lemma. □

We are now ready to prove Theorem 2.4.1.

Proof of Theorem 2.4.1. (a) By Proposition 2.4.4 it suffices to show that for any $t > 0$ there exists $(x, y) \in D^\circ$ (see (2.5) for the definition of D) such that $\Phi^{(\pm)}_{xy}(x\,y, t) < 0$.

Define $z := \frac{1}{|\sigma_\pm|\sqrt{t}} \log\left(\frac{x}{y}\right) > 0$ and $\alpha := \frac{\mu\sqrt{t}}{|\sigma_\pm|} > 0$. Note that, since we are allowed to choose the point $(x, y) \in D^\circ$ arbitrarily close to the diagonal half-line in the boundary of D, a Taylor expansion of order one of $z \mapsto n(z - \alpha)$ and $z \mapsto N(-z - \alpha)$ around $z = 0$, the representation of $\Phi^{(\pm)}_{xy}$ in Lemma 2.4.5 and the inequality

$$\alpha N(-\alpha) < n(-\alpha) \tag{2.19}$$

imply that $\Phi^{(\pm)}_{xy}(x, y, t) < 0$ for some $(x, y) \in D^\circ$. To check (2.19), note that $un(u) = -n'(u)$ and

$$\alpha N(-\alpha) = \int_\alpha^\infty \alpha n(u)\,\mathrm{d}u < \int_\alpha^\infty un(u)\,\mathrm{d}u = n(-\alpha).$$

(b) Assume first $\sigma_\pm \neq 0$. Then the representation of $\Phi^{(\pm)}_{xy}$ in Lemma 2.4.5 and the assumption $\mu \leq 0$ imply $\Phi^{(\pm)}_{xy} \geq 0$ on E°. Hence Proposition 2.4.4 yields the theorem. If $\sigma_\pm = 0$, we have $\pm = -$, $\sigma_1 = \sigma_2$ and, by (2.4), it follows that $\Phi^{(-)}(x, y, t) = 1$ holds for all $(x, y) \in D^\circ$, $t \in [0, T]$. Hence $-\Phi^{(-)}$ is the value function for Problem (T−), and the theorem is proved. □

2.5 Ergodic average problems

2.5.1 The problems and main theorem

We would like to solve the following problems: find $V^{\inf} \in \mathcal{V}$ and $V^{\sup} \in \mathcal{V}$ such that

$$\inf_{V\in\mathcal{V}} \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}(\tau(V) > t)\,\mathrm{d}t = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}(\tau(V^{\inf}) > t)\,\mathrm{d}t \tag{EAInf}$$

and

$$\sup_{V\in\mathcal{V}} \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}(\tau(V) > t)\,\mathrm{d}t = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}(\tau(V^{\sup}) > t)\,\mathrm{d}t. \tag{EASup}$$

Note first that Fubini's theorem and the Dominated Convergence Theorem imply that the limit exists and has the following representation:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{P}(\tau(V) > t)\, \mathrm{d}t = \lim_{T\to\infty} \mathbb{E}\left(\frac{\tau(V)}{T} \wedge 1\right) = \mathbb{P}(\tau(V) = \infty). \tag{2.20}$$

A solution to these problems, independent of the values of the parameters of the geometric Brownian motions in (2.1), is given in the following theorem. It is completely analogous to the infinite time horizon case.

Theorem 2.5.1. *The Brownian motions $V^{inf} = -B$ and $V^{sup} = B$ solve Problems* (EAInf) *and* (EASup) *respectively.*

2.5.2 Proof

The proof is rather short due to (2.20) and because we can apply the result obtained for the finite horizon problem.

Proof of Theorem 2.5.1. As in Section 2.3 we may assume that, due to symmetry, the starting points of X and $Y(V)$ satisfy $(x, y) \in D$. If $x = y$ we have $\tau(V) = 0$ for all $V \in \mathcal{V}$ and Proposition 2.5.1 follows. So we can assume $(x, y) \in D^\circ$ in the rest of the proof.

We first analyse the case $\mu > 0$. By (2.20), Problems (EAInf) and (EASup) are equivalent to finding $V^{(\pm)} \in \mathcal{V}$ such that

$$\inf_{V\in\mathcal{V}} \pm\mathbb{P}(\tau(V) = \infty) = \pm\mathbb{P}(\tau(V^{(\pm)}) = \infty). \tag{S$\pm$}$$

The strong law of large numbers for Brownian motion (see e.g. [5, p. 53]), representation (2.4) and $\log\left(\frac{x}{y}\right) > 0$ imply the equality $\mathbb{P}_{x,y}(\tau(V) = \infty) = 0$ for every $V \in \mathcal{V}$ and Theorem 2.5.1 follows.

In the case $\mu \leq 0$, we return to the formulation of Problems (EAInf) and (EASup) above. Observe that Theorem 2.4.1(b) yields the optimal couplings that minimise and maximise the probability $\mathbb{P}(\tau(V) > t)$ for every $t \geq 0$. Since the couplings are independent of t, they also minimise and maximise the ergodic average criteria in Problems (EAInf) and (EASup), which concludes the proof. □

2.6 Exponential efficiency problems

2.6.1 The problems and main theorem

We would like to find $V^{(\pm)} \in \mathcal{V}$ such that:

$$\inf_{V\in\mathcal{V}} \pm \liminf_{t\to\infty} \frac{\log\left(\mathbb{P}_{x,y}\left(\tau(V) > t\right)\right)}{t} = \pm \liminf_{t\to\infty} \frac{\log\left(\mathbb{P}_{x,y}\left(\tau\left(V^{(\pm)}\right) > t\right)\right)}{t}. \quad \text{(EE}\pm\text{)}$$

It turns out that the answer is a dichotomy, as in the finite horizon problem. Recall that μ and $\sigma_\pm$ are given in (2.3), and D in (2.5).

Theorem 2.6.1. *The following holds for any $(x, y) \in D^\circ$:*

(a) If $\mu > 0$, then $V^{(\pm)} = \mp B$ does NOT solve Problem (EE$\pm$).

(b) If $\mu \le 0$, then $V^{(\pm)} = \mp B$ solves Problem (EE$\pm$).

2.6.2 Proof

Proof of Theorem 2.6.1. The second part of the theorem again follows from Theorem 2.4.1. We will prove the first part in the following way: when we claim that B is not optimal we will show that $-B$ is better, and vice-versa.

The following bounds hold for the standard normal distribution function $N(\cdot)$ and its derivative $n(\cdot)$:

$$-\frac{z}{1+z^2}\, n(z) \le N(z) \le -\frac{n(z)}{z} \qquad \text{for any } z < 0. \tag{2.21}$$

The first inequality follows from the identity

$$\int_r^\infty \left(1 + \frac{1}{y^2}\right) \mathrm{e}^{-\frac{y^2}{2}}\, \mathrm{d}y = \frac{1}{r}\, \mathrm{e}^{-\frac{r^2}{2}}, \quad r > 0,$$

and the second is given in (2.19).

Assume first that $\sigma_\pm \neq 0$. Let

$$Z(t) := \frac{\log\left(\frac{x}{y}\right) - \mu t}{|\sigma_\pm|\sqrt{t}} \qquad \text{and} \qquad \widehat{Z}(t) := \frac{-\log\left(\frac{x}{y}\right) - \mu t}{|\sigma_\pm|\sqrt{t}},$$

and note that for all large $t > 0$ we have $\widehat{Z}(t) < Z(t) < 0$, and the equality

$$n\left(Z(t)\right) = \left(\frac{x}{y}\right)^{\frac{2\mu}{\sigma_\pm^2}} n\left(\widehat{Z}(t)\right) \tag{2.22}$$

holds. The representations in (2.17) and (2.18) imply

$$\Phi^{(\pm)}(x,y,t) = N(Z(t))\left(1-\left(\frac{x}{y}\right)^{\frac{2\mu}{\sigma_\pm^2}}\frac{N(\widehat{Z}(t))}{N(Z(t))}\right). \tag{2.23}$$

The inequalities in (2.21) yield

$$\lim_{t\to\infty}\frac{1}{t}\log\big(N(Z(t))\big) = -\frac{\mu^2}{2\sigma_\pm^2}. \tag{2.24}$$

In order to deal with the second factor on the right-hand side of (2.23), we note the following inequalities:

$$1-\left(\frac{x}{y}\right)^{\frac{2\mu}{\sigma_\pm^2}}\frac{N(\widehat{Z}(t))}{N(Z(t))} \geq 1+(1+Z(t)^2)\frac{N(\widehat{Z}(t))}{n(\widehat{Z}(t))Z(t)} \geq 1-\frac{1+Z(t)^2}{\widehat{Z}(t)Z(t)};$$

they are a consequence of two applications of the second inequality in (2.21) and identity (2.22). Let the assumption

$$\log\left(\frac{x}{y}\right) > \frac{\sigma_+^2}{2\mu} \tag{2.25}$$

hold. Then we obtain

$$1-\frac{1+Z(t)^2}{\widehat{Z}(t)Z(t)} = \frac{\frac{1}{t}\left(2\mu\log\left(\frac{x}{y}\right)-\sigma_\pm^2\right)-\frac{2}{t^2}\log\left(\frac{x}{y}\right)^2}{\mu^2-\frac{1}{t^2}\log\left(\frac{x}{y}\right)^2} > 0 \qquad \text{for all large } t>0,$$

and

$$\lim_{t\to\infty}\frac{1}{t}\log\left(1-\frac{1+Z(t)^2}{\widehat{Z}(t)Z(t)}\right) = 0. \tag{2.26}$$

By (2.23) we have

$$N(Z(t))\left(1-\frac{1+Z(t)^2}{\widehat{Z}(t)Z(t)}\right) \leq \Phi^{(\pm)}(x,y,t) \leq N(Z(t)).$$

If the starting points x, y satisfy (2.25), then (2.24), (2.26), the inequalities in the line above and the fact that the function log is increasing imply the following equality:

$$\lim_{t\to\infty}\frac{1}{t}\log\Big(\Phi^{(\pm)}(x,y,t)\Big) = -\frac{\mu^2}{2\sigma_\pm^2}.$$

In order to see that this remains true without assumption (2.25), i.e. for

$(x,y) \in D$ such that $\log\left(\frac{x}{y}\right) \in \left(0, \frac{\sigma_\pm^2}{2\mu}\right]$, define a Brownian motion with drift $W^\pm$ and its first-passage time $T_\pm(z)$:

$$W_t^\pm := \mp\sigma_\pm B_t + \mu t, \quad t \geq 0, \qquad \text{and} \qquad T_\pm(z) := \inf\{t \geq 0;\, W_t^\pm = z\}, \quad z \in \mathbb{R},$$

and note that $\mathbb{P}_{x,y}(\tau(\mp B) > t) = \mathbb{P}\left(T_\pm\left(\log\left(\frac{x}{y}\right)\right) > t\right)$ holds for any $(x,y) \in D$ (cf. (2.4)). Fix $(x,y) \in D$ that violates assumption (2.25) and pick $\alpha_0 < 0$ and $(x_0, y_0) \in D^\circ$ such that the following holds:

$$\log\left(\frac{x_0}{y_0}\right) = \log\left(\frac{x}{y}\right) - \alpha_0 > \frac{\sigma_+^2}{2\mu}.$$

Denote the constant $q_\pm := \mathbb{P}\left(W_1^\pm < \alpha_0, T_\pm\left(\log\left(\frac{x}{y}\right)\right) > 1\right)$, which clearly satisfies $q_\pm \in (0,1)$. The Markov property of $W^\pm$ at time 1 yields the following inequalities for all $t > 1$:

$$\begin{aligned}
\mathbb{P}_{x,y}(\tau(\mp B) > t) &= \mathbb{P}\left(T_\pm\left(\log\left(\frac{x}{y}\right)\right) > t\right) \\
&\geq q_\pm\, \mathbb{P}\left(T_\pm\left(\log\left(\frac{x}{y}\right) - \alpha_0\right) > t - 1\right) \\
&> q_\pm\, \mathbb{P}\left(T_\pm\left(\log\left(\frac{x}{y}\right) - \alpha_0\right) > t\right) \\
&= q_\pm\, \mathbb{P}_{x_0,y_0}(\tau(\mp B) > t).
\end{aligned}$$

Since (2.23) implies the bound $\mathbb{P}_{x,y}(\tau(\mp B) > t) \leq N(Z(t))$ for any $(x,y) \in D^\circ$, we obtain

$$\lim_{t\to\infty} \frac{1}{t} \log\left(\mathbb{P}_{x,y}(\tau(\mp B) > t)\right) = \lim_{t\to\infty} \frac{1}{t} \log\left(\Phi^{(\pm)}(x,y,t)\right) = -\frac{\mu^2}{2\sigma_\pm^2}, \tag{2.27}$$

by the inequality above, our analysis under assumption (2.25) and the limit in (2.24).

Definition (2.3) and assumption $\sigma_\pm \neq 0$ imply $|\sigma_+| > |\sigma_-| > 0$ and hence $-\frac{\mu^2}{2\sigma_-^2} < -\frac{\mu^2}{2\sigma_+^2}$. The mirror coupling is therefore not optimal for Problem (EE+) since it has a strictly thicker exponential tail than the synchronous coupling. Likewise, the synchronous coupling is not optimal for Problem (EE−), which requires the thickest possible exponential tail among all couplings, since it has a thinner tail than the mirror coupling.

In the case $\sigma_\pm = 0$ we have $\sigma_1 = \sigma_2$ and, by (2.4), $\tau(B) = \frac{1}{\mu}\log\left(\frac{x}{y}\right)$. Hence $\mathbb{P}_{x,y}\left(\tau(B) > t\right) = 0$ for all $t \geq \frac{1}{\mu}\log\left(\frac{x}{y}\right)$. Since the equality in (2.27) still holds for

$\pm = +$ (note that $|\sigma_+| > 0$), we obtain

$$\lim_{t\to\infty} \frac{1}{t} \log\left(\mathbb{P}_{x,y}\left(\tau(B) > t\right)\right) = -\infty < -\frac{\mu^2}{2\sigma_+^2} = \lim_{t\to\infty} \frac{1}{t} \log\left(\mathbb{P}_{x,y}\left(\tau(-B) > t\right)\right).$$

This inequality implies that the mirror (resp. synchronous) coupling is not optimal for Problem (EE+) (resp. (EE−)). □

Remark 2.6.2. It is the presence of the positive drift $\mu > 0$ that makes the mirror coupling suboptimal in Problem (T+) (see Theorem 2.4.1). The proof of Theorem 2.6.1 suggests that if the drift is positive, it is in fact better (according to the exponential efficiency criterion) to use the synchronous coupling. This naturally leads to the following conjecture.

Conjecture 2.6.3. If $\mu > 0$, the synchronous (resp. mirror) coupling is optimal in Problem (EE+) (resp. (EE−)).

2.7 Conclusion

We have seen that, unlike in the case of Brownian motions, the mirror and synchronous couplings are not always a solution to the finite horizon problem for geometric Brownian motions. Nevertheless, this does not prevent them from solving the ergodic average and infinite horizon problems (for all discount rates). Interestingly, when it comes to the exponential efficiency, this problem is again not always solved by the two couplings.

For the exponential efficiency problem, we at least have a more or less natural conjecture (although we do not see any natural way of proving or disproving it). In the case of the finite horizon problem, there seems to be no clear candidate. It may even happen that there is no optimal coupling since the supremum and infimum need not be attained. In any case, we will at least "obtain" the value function at the very end of the thesis via the policy improvement algorithm. In fact, this is where the motivation to start looking at the policy improvement algorithm in a continuous setting came from.

Why did we only deal with geometric Brownian motions and not general diffusion processes (with the diffusion coefficients of the same sign)? The reason lies in Lemma 2.3.5, where certain analytical properties of the candidate value functions had to be verified. In the case of geometric Brownian motions, this was easy since we had obtained an explicit formula for the functions. If we had been dealing with other processes, this could have easily become an impossible task.

Chapter 3

The policy improvement algorithm for the general continuous discounted infinite-horizon problem

3.1 Introduction

To the best of the author's knowledge, nothing has been published about the policy improvement algorithm for controlled processes in continuous time with continuous state space, continuous paths and general action space. In this and the next chapter we deal with such processes, which become diffusion processes if controlled by Markov policies.

In the present chapter we investigate the discounted infinite-horizon minimisation problem. In Section 3.2 we treat the one-dimensional case. The first idea, based on the discrete case formula in (1.1), was to define the policy at each step (given an initial policy) by

$$\pi_{n+1}(x) := \operatorname*{argmin}_{p \in A} \left(L_p V_{\pi_n}(x) - \alpha(x,p) V_{\pi_n}(x) + f(x,p) \right), \quad x \in (a,b), \quad n \in \mathbb{N}_0,$$

where A is the compact action space, L_p the infinitesimal generator corresponding to action p, V_{π_n} the payoff function generated by the Markov policy π_n, α the discounting function, f the cost function, and (a,b) the (possibly infinite) state space. (Note that we have an additional term; this is because α in the discrete case, unlike here, was constant.) It turned out that it is better to look at the "normed"

HJB equation (cf. [18, p. 12]), with a normalising multiplier such that the policies no longer depend on the second derivative of the payoff function: for each $n \in \mathbb{N}_0$,

$$\pi_{n+1}(x) := \operatorname*{argmin}_{p \in A} \left(\frac{\mu(x,p)}{\sigma(x,p)^2} V'_{\pi_n}(x) - \frac{\alpha(x,p)}{\sigma(x,p)^2} V_{\pi_n}(x) + \frac{f(x,p)}{\sigma(x,p)^2} \right), \quad x \in (a,b),$$

where μ is the drift and σ the diffusion coefficient.

We solve the problem by finding a convergent subsequence of the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ whose limit is an optimal policy, and by proving that the sequence of payoff functions $\{V_{\pi_n}\}_{n\in\mathbb{N}}$ converges to the value function of the problem; the convergence is monotonic, so the policy on each step of the algorithm indeed improves the previous one (as in the discrete case). No discretisation of time, action space or state space is involved at any point.

In Section 3.3 we treat the multidimensional case. Although the results can be applied to one dimension, they do not imply the results of the previous section. The results of Section 3.2 are more general because they deal with domains other than $\mathbb{R}$ and because we can use the normed HJB equation. Despite this the proofs in the multidimensional case are not easier, in fact there is an additional property that has to be proved. For the purpose of elliptic differential equations theory, we need to establish the continuity of the payoff functions in advance. We do that by invoking the mirror coupling of multidimensional diffusions (Lemma 3.3.13).

In order to carry out our proof that the algorithm works, the data of the problem have to satisfy certain assumptions; in Section 3.4 we show that there is a large family of suitable data. We also present a concrete example that we implemented in Matlab, for which the convergence towards both the optimal policy and the value function is numerically very fast, which again sounds familiar from the discrete case.

3.2 One-dimensional case

3.2.1 Setting and the algorithm

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space (satisfying the usual conditions) that supports an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B_t)_{t\geq 0}$. Let $a, b \in [-\infty, \infty]$, $a < b$, and for any $\mathbb{R}$-valued process $Y = (Y_t)_{t\geq 0}$ define

$$\tau_a^b(Y) := \inf\{t \geq 0;\ Y_t \leq a \text{ or } Y_t \geq b\} \quad (\inf \emptyset := \infty).$$

Let (A, d) be a compact metric space, and for any $x \in (a, b)$ define the set of admissible controls at x as

$$\begin{aligned}\mathcal{A}(x) := \{&\Pi = (\Pi_t)_{t\geq 0};\ \Pi \text{ is an } A\text{-valued process adapted to } (\mathcal{F}_t)_{t\geq 0}, \text{ and} \\ &\text{there exists a pathwise unique process } X^{\Pi,x} = \left(X_t^{\Pi,x}\right)_{t\geq 0} \text{ that satisfies (3.1)}\},\end{aligned}$$

where

$$\begin{aligned} X_t^{\Pi,x} &= x + \int_0^t \sigma\left(X_s^{\Pi,x}, \Pi_s\right) \mathrm{d}B_s + \int_0^t \mu\left(X_s^{\Pi,x}, \Pi_s\right) \mathrm{d}s, \quad 0 \leq t < \tau_a^b\left(X^{\Pi,x}\right), \\ X_t^{\Pi,x} &= X^{\Pi,x}_{\tau_a^b(X^{\Pi,x})}, \quad \tau_a^b\left(X^{\Pi,x}\right) \leq t < \infty, \end{aligned} \tag{3.1}$$

and $\sigma : (a, b) \times A \to \mathbb{R}$ and $\mu : (a, b) \times A \to \mathbb{R}$ are measurable functions. In fact it will not matter what the process $X^{\Pi,x}$ looks like after it reaches a or b, if that occurs at all.

Let $\alpha : (a, b) \times A \to \mathbb{R}$ and $f : (a, b) \times A \to \mathbb{R}$ be measurable functions and $g : \{a, b\} \cap \mathbb{R} \to \mathbb{R}$ an arbitrary function. For any $x \in (a, b)$ and $\Pi \in \mathcal{A}(x)$ define the payoff as

$$\begin{aligned} V_\Pi(x) := \mathbb{E}\Bigg(&\int_0^{\tau_a^b\left(X^{\Pi,x}\right)} \mathrm{e}^{-\int_0^t \alpha\left(X_s^{\Pi,x},\Pi_s\right)\mathrm{d}s} f\left(X_t^{\Pi,x}, \Pi_t\right) \mathrm{d}t \\ &+\mathrm{e}^{-\int_0^{\tau_a^b\left(X^{\Pi,x}\right)} \alpha\left(X_t^{\Pi,x},\Pi_t\right)\mathrm{d}t} g\left(X^{\Pi,x}_{\tau_a^b(X^{\Pi,x})}\right) \mathbb{I}_{\{\tau_a^b(X^{\Pi,x})<\infty\}}\Bigg). \end{aligned}$$

The problem is to find the *value function* V, defined by

$$V(x) := \inf_{\Pi\in\mathcal{A}(x)} V_\Pi(x), \quad x \in (a, b),$$

and an optimal control (which will in general depend on x), if it exists.

In order to solve the problem, we make the following assumptions about the functions σ, μ, α and f.

Assumption 3.2.1. The functions σ, μ, α and f are bounded, and Lipschitz on compacts in $(a, b) \times A$, i.e. for every compact set $K \subseteq (a, b)$ there exists a constant $C > 0$ such that

$$|h(x, p) - h(y, r)| \leq C((x - y)^2 + d(p, r)^2)^{\frac{1}{2}}$$

holds for every $x, y \in K$, $p, r \in A$ and $h \in \{\sigma, \mu, \alpha, f\}$. In addition, σ^2 is bounded away from 0, and α is positive and bounded away from 0.

Assumption 3.2.2. For every $h \in \mathcal{C}^2((a,b))$ and $x \in (a,b)$, let $I_h(x)$ denote a point where the minimum of the function

$$p \mapsto \frac{\mu(x,p)}{\sigma(x,p)^2}\, h'(x) - \frac{\alpha(x,p)}{\sigma(x,p)^2}\, h(x) + \frac{f(x,p)}{\sigma(x,p)^2}, \quad p \in A,$$

is attained. If the sequence $\{h_n'\}_{n\in\mathbb{N}}$ is uniformly Lipschitz (i.e. there exists a constant that is a Lipschitz constant for all the functions in the sequence) on compacts in (a,b), then the points $\{I_{h_n}(x);\ x \in (a,b), n \in \mathbb{N}\}$ can be chosen in such a way that the sequence of functions $\{I_{h_n}\}_{n\in\mathbb{N}}$ ($I_{h_n} : (a,b) \to A$) is also uniformly Lipschitz on compacts in (a,b).

Remark 3.2.3. It is important to note that there are non-trivial data that satisfy the above assumptions. Some are presented in Proposition 3.4.1.

We will need a special class of controls. A measurable function $\pi : (a,b) \to A$ is a *Markov policy* if for every $x \in (a,b)$ there exists a pathwise unique process $X^{\pi,x} = (X_t^{\pi,x})_{t\geq 0}$ that satisfies the following:

$$\begin{aligned} X_t^{\pi,x} &= x + \int_0^t \sigma\left(X_s^{\pi,x}, \pi\left(X_s^{\pi,x}\right)\right) \mathrm{d}B_s + \int_0^t \mu\left(X_s^{\pi,x}, \pi\left(X_s^{\pi,x}\right)\right) \mathrm{d}s \\ &\quad \text{if} \quad 0 \leq t < \tau_a^b\left(X^{\pi,x}\right), \\ X_t^{\pi,x} &= X^{\pi,x}_{\tau_a^b(X^{\pi,x})} \quad \text{if} \quad \tau_a^b\left(X^{\pi,x}\right) \leq t < \infty. \end{aligned} \tag{3.2}$$

If π is a Markov policy, then $\pi(X^{\pi,x}) := (\pi(X_t^{\pi,x}))_{t\geq 0} \in \mathcal{A}(x)$ for every $x \in (a,b)$ (where $\pi(a)$, if $a > -\infty$, and $\pi(b)$, if $b < \infty$, are arbitrary elements of A). For easier notation we define

$$\sigma_\pi(\cdot) := \sigma(\cdot, \pi(\cdot)), \quad \mu_\pi(\cdot) := \mu(\cdot, \pi(\cdot)), \quad \alpha_\pi(\cdot) := \alpha(\cdot, \pi(\cdot)), \quad f_\pi(\cdot) := f(\cdot, \pi(\cdot)),$$

$$V_\pi(\cdot) := V_{\pi(X^{\pi,\cdot})}(\cdot), \quad \text{and} \quad L_\pi h := \frac{1}{2}\sigma_\pi^2 h'' + \mu_\pi h' \ \text{ for } \ h \in \mathcal{C}^2((a,b)).$$

If π is a constant Markov policy with the value $p \in A$, we will write σ_p, μ_p, α_p, f_p and L_p instead of σ_π, μ_π, α_π, f_π and L_π, respectively.

The first proposition establishes that there is a large class of Markov policies. It is the members of this class for which our algorithm will be defined. As explained in Subsection 1.1.2, the proofs do not follow immediately, but are presented in the next two subsections.

Proposition 3.2.4. *If $\pi : (a,b) \to A$ is Lipschitz on compacts in (a,b), then π is a Markov policy.*

It turns out that the corresponding payoff functions satisfy the following differential equation (cf. Lemma 1.3.2).

Proposition 3.2.5. *For any Markov policy π that is Lipschitz on compacts in (a,b), the following holds: $V_\pi \in \mathcal{C}^2((a,b))$ and*

$$L_\pi V_\pi - \alpha_\pi V_\pi + f_\pi = 0.$$

Now we can finally present the algorithm. Let π_0 be a Markov policy that is Lipschitz on compacts in (a,b). The *policy improvement algorithm* is defined inductively in the following way:

$$\pi_{n+1}(x) := \operatorname*{argmin}_{p\in A}\left(\frac{L_p V_{\pi_n}}{\sigma_p^2}(x) - \frac{\alpha_p V_{\pi_n}}{\sigma_p^2}(x) + \frac{f_p}{\sigma_p^2}(x)\right), \quad x \in (a,b), \quad n \in \mathbb{N}_0. \tag{3.3}$$

Note the equality $\pi_{n+1}(x) = \operatorname{argmin}_{p\in A}\left(\frac{\mu(x,p)}{\sigma(x,p)^2}\, V'_{\pi_n}(x) - \frac{\alpha(x,p)}{\sigma(x,p)^2}\, V_{\pi_n}(x) + \frac{f(x,p)}{\sigma(x,p)^2}\right)$. For every $n \in \mathbb{N}_0$, the function V''_{π_n} is continuous by Proposition 3.2.5, hence V'_{π_n} is Lipschitz on compacts in (a,b). Therefore Assumption 3.2.2 (applied separately for every $n \in \mathbb{N}_0$) ensures that the points $\{\pi_{n+1}(x);\ x \in (a,b)\}$ can be chosen in such a way that $\pi_{n+1} : (a,b) \to A$ is Lipschitz on compacts in (a,b).

Remark 3.2.6. If the algorithm stops, i.e. $\pi_{n+1} = \pi_n$ for some $n \in \mathbb{N}_0$, then clearly $V_{\pi_m} = V_{\pi_n}$ and $\pi_m = \pi_n$ hold for every $m \geq n$. We can then proceed directly to the verification lemma (Theorem 3.2.10) to prove that V_{π_n} is the value function and π_n is an optimal policy. In general we first need to establish the existence of the limit payoff function and strategy.

The next theorem justifies the name of the algorithm (cf. Theorem 1.3.3).

Theorem 3.2.7. *For every $n \in \mathbb{N}_0$, $x \in (a,b)$ and Markov policy π_0 that is Lipschitz on compacts in (a,b), the following holds:*

$$V_{\pi_{n+1}}(x) \leq V_{\pi_n}(x).$$

Since $\{V_{\pi_n}\}_{n\in\mathbb{N}}$ is a decreasing bounded sequence, we can define

$$V_{\lim}(x) := \lim_{n\to\infty} V_{\pi_n}(x), \quad x \in (a,b).$$

The sequence of policies might not converge, but the next proposition says that there exists a convergent subsequence.

Proposition 3.2.8. *There exists a subsequence of $\{\pi_n\}_{n\in\mathbb{N}}$ that converges uniformly on every compact subset of (a,b).*

Therefore this subsequence converges (pointwise) on (a, b); denote the limit by π_{lim}. Note that V_{lim} and π_{lim} can in principle depend on π_0. Additionally, π_{lim} could depend on the choice of the subsequence from Proposition 3.2.8. However, this turns out to be irrelevant in the following theorem, which brings together the limit payoff function and the limit policy.

Theorem 3.2.9. *For every $x \in (a, b)$ and Markov policy π_0 that is Lipschitz on compacts in (a, b), the following holds:*

$$V_{\text{lim}}(x) = V_{\pi_{\text{lim}}}(x).$$

The last step establishes $V_{\text{lim}} = V$ via the so called verification lemma.

Theorem 3.2.10. *For every $x \in (a, b)$, $\Pi \in \mathcal{A}(x)$ and Markov policy π_0 that is Lipschitz on compacts in (a, b), the following holds:*

$$V_{\text{lim}}(x) \leq V_{\Pi}(x).$$

Hence V_{lim} is the value function (and does not depend on π_0) and π_{lim} is an optimal policy.

3.2.2 Auxiliary results

Lemma 3.2.11. *For any Markov policy π, 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 $a > -\infty$. If we prove

$$\lim_{n\to\infty} V_\pi(x_n) = g(a),$$

the lemma follows (since the proof for b is analogous).

Let $\epsilon > 0$. Since μ is bounded and σ^2 is bounded and bounded away from 0, the process X^{π,x_n} locally looks like a Brownian motion with drift, therefore

$$\mathbb{P}\left(\tau_a^\infty\left(X^{\pi,x_n}\right) > \epsilon\right) < \epsilon \quad \text{and} \quad \mathbb{P}\left(\tau_a^\infty\left(X^{\pi,x_n}\right) > \tau_{-\infty}^b\left(X^{\pi,x_n}\right)\right) < \epsilon$$

hold for all large enough $n \in \mathbb{N}$. Hence there exists $n_0 \in \mathbb{N}$ such that

$$\begin{aligned}
&\mathbb{P}\left(\tau_a^\infty\left(X^{\pi,x_n}\right) > \epsilon \wedge \tau_{-\infty}^b\left(X^{\pi,x_n}\right)\right) \\
&\leq \mathbb{P}\left(\tau_a^\infty\left(X^{\pi,x_n}\right) > \epsilon\right) + \mathbb{P}\left(\tau_a^\infty\left(X^{\pi,x_n}\right) > \tau_{-\infty}^b\left(X^{\pi,x_n}\right)\right) \\
&< 2\epsilon
\end{aligned}$$

holds for all $n \geq n_0$. We obtain

$$\begin{aligned}
&|V_\pi(x_n) - g(a)| \\
&\leq \mathbb{E}\left(\int_0^{\tau_a^b(X^{\pi,x_n})} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x_n}\right)\mathrm{d}s} \left|f_\pi\left(X_t^{\pi,x_n}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) > \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&\quad + \mathbb{E}\left(\left|\mathrm{e}^{-\int_0^{\tau_a^b(X^{\pi,x_n})} \alpha_\pi\left(X_s^{\pi,x_n}\right)\mathrm{d}t} g\left(X^{\pi,x_n}_{\tau_a^b(X^{\pi,x_n})}\right) - g(a)\right| \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) > \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&\quad + \mathbb{E}\left(\int_0^{\tau_a^b(X^{\pi,x_n})} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x_n}\right)\mathrm{d}s} \left|f_\pi\left(X_t^{\pi,x_n}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) \leq \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&\quad + \mathbb{E}\left(\left|\mathrm{e}^{-\int_0^{\tau_a^b(X^{\pi,x_n})} \alpha_\pi\left(X_s^{\pi,x_n}\right)\mathrm{d}t} g\left(X^{\pi,x_n}_{\tau_a^b(X^{\pi,x_n})}\right) - g(a)\right| \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) \leq \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right).
\end{aligned}$$

We will now see that there exists $M > 0$ such that each of the four terms is bounded by $M\epsilon$ if $n \geq n_0$, which concludes the proof. For the first term we obtain the bound since f is bounded, α is positive and bounded away from 0, and the probability of the event is under 2ϵ. To estimate the second term, we only need to note that g is bounded. In the third term we can change the upper bound in the integral to ϵ, which then yields the desired estimate. Applying the elementary inequality $1 - \mathrm{e}^{-x} \leq x$ for $x \geq 0$, we obtain for the final term

$$\begin{aligned}
&\mathbb{E}\left(\left|\mathrm{e}^{-\int_0^{\tau_a^b(X^{\pi,x_n})} \alpha_\pi\left(X_t^{\pi,x_n}\right)\mathrm{d}t} g\left(X^{\pi,x_n}_{\tau_a^b(X^{\pi,x_n})}\right) - g(a)\right| \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) \leq \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&= \mathbb{E}\left(\left|1 - \mathrm{e}^{-\int_0^{\tau_a^b(X^{\pi,x_n})} \alpha_\pi\left(X_t^{\pi,x_n}\right)\mathrm{d}t}\right| |g(a)|\, \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) \leq \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&\leq \mathbb{E}\left(\int_0^{\tau_a^b(X^{\pi,x_n})} \alpha_\pi\left(X_t^{\pi,x_n}\right) \mathrm{d}t\, |g(a)|\, \mathbb{I}_{\{\tau_a^\infty(X^{\pi,x_n}) \leq \epsilon \wedge \tau_{-\infty}^b(X^{\pi,x_n})\}}\right) \\
&\leq \mathbb{E}\left(\int_0^{\epsilon} \alpha_\pi\left(X_t^{\pi,x_n}\right) \mathrm{d}t\, |g(a)|\right),
\end{aligned}$$

which yields the required bound since α is bounded. □

The processes controlled by Markov policies are strong Markov processes (Theorem 4.20 in [17, p. 322]). This enables us to prove the following lemma.

Lemma 3.2.12. *The following holds for every Markov policy π, $x \in (a,b)$ and stopping time S:*

$$\begin{aligned}
&\mathbb{E}\Bigg(\int_0^{\tau_a^b(X^{\pi,x})} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t \\
&\quad + \mathrm{e}^{-\int_0^{\tau_a^b(X^{\pi,x})} \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} g\left(X^{\pi,x}_{\tau_a^b(X^{\pi,x})}\right) \mathbb{I}_{\{\tau_a^b(X^{\pi,x})<\infty\}} \Bigg| \mathcal{F}_S \Bigg) \\
&= \int_0^{S\wedge\tau_a^b(X^{\pi,x})} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t \\
&\quad + \mathrm{e}^{-\int_0^{S\wedge\tau_a^b(X^{\pi,x})} \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} V_\pi\left(X^{\pi,x}_{S\wedge\tau_a^b(X^{\pi,x})}\right) \mathbb{I}_{\{S\wedge\tau_a^b(X^{\pi,x})<\infty\}}.
\end{aligned}$$

In particular, the process M is a uniformly integrable martingale, where

$$\begin{aligned}
M_r := &\int_0^{r\wedge\tau_a^b(X^{\pi,x})} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t \\
&+ \mathrm{e}^{-\int_0^{r\wedge\tau_a^b(X^{\pi,x})} \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} V_\pi\left(X^{\pi,x}_{r\wedge\tau_a^b(X^{\pi,x})}\right), \quad r \geq 0.
\end{aligned}$$

Proof. Let $\tau := \tau_a^b\left(X^{\pi,x}\right)$ and $\tau_S := \tau \circ \theta_{S\wedge\tau} = \tau_a^b\left(X^{\pi,x}_{\cdot+S\wedge\tau}\right)$. Then $\tau_S = \tau - S\wedge\tau$

holds almost surely. We obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathrm{e}^{-\int_0^\tau \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} g\left(X_\tau^{\pi,x}\right)\mathbb{I}_{\{\tau<\infty\}}\middle|\mathcal{F}_S\right)\\
&= \mathbb{E}\left(\int_0^{\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathrm{e}^{-\int_0^\tau \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} g\left(X_\tau^{\pi,x}\right)\mathbb{I}_{\{\tau<\infty\}}\middle|\mathcal{F}_{S\wedge\tau}\right)\\
&= \int_0^{S\wedge\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t\\
&\quad + \mathbb{E}\Bigg(\mathbb{I}_{\{S\wedge\tau<\infty\}}\int_0^{\tau-S\wedge\tau} \mathrm{e}^{-\int_0^{t+S\wedge\tau} \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_{t+\tau\wedge S}^{\pi,x}\right)\mathrm{d}t\\
&\qquad + \mathrm{e}^{-\int_0^\tau \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} g\left(X_\tau^{\pi,x}\right)\mathbb{I}_{\{\tau_S<\infty\}}\mathbb{I}_{\{S\wedge\tau<\infty\}}\Bigg|\mathcal{F}_{S\wedge\tau}\Bigg)\\
&= \int_0^{S\wedge\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathbb{I}_{\{S\wedge\tau<\infty\}}\mathrm{e}^{-\int_0^{S\wedge\tau} \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\\
&\cdot \mathbb{E}\Bigg(\int_0^{\tau_S} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_{s+S\wedge\tau}^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_{t+S\wedge\tau}^{\pi,x}\right)\mathrm{d}t\\
&\qquad + \mathrm{e}^{-\int_0^{\tau_S} \alpha_\pi\left(X_{t+S\wedge\tau}^{\pi,x}\right)\mathrm{d}t} g\left(X_{\tau_S+S\wedge\tau}^{\pi,x}\right)\mathbb{I}_{\{\tau_S<\infty\}}\Bigg|\mathcal{F}_{S\wedge\tau}\Bigg)\\
&= \int_0^{S\wedge\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathbb{I}_{\{S\wedge\tau<\infty\}}\mathrm{e}^{-\int_0^{S\wedge\tau} \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\\
&\cdot \mathbb{E}_x\left(\int_0^{\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi}\right)\mathrm{d}t \circ \theta_{S\wedge\tau} + \mathrm{e}^{-\int_0^\tau \alpha_\pi\left(X_t^{\pi}\right)\mathrm{d}t} g\left(X_\tau^{\pi}\right)\mathbb{I}_{\{\tau<\infty\}} \circ \theta_{S\wedge\tau}\middle|\mathcal{F}_{S\wedge\tau}\right)\\
&= \int_0^{S\wedge\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathbb{I}_{\{S\wedge\tau<\infty\}}\mathrm{e}^{-\int_0^{S\wedge\tau} \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\\
&\cdot \mathbb{E}_{X_{S\wedge\tau}^{\pi,x}}\left(\int_0^{\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi}\right)\mathrm{d}t + \mathrm{e}^{-\int_0^\tau \alpha_\pi\left(X_t^{\pi}\right)\mathrm{d}t} g\left(X_\tau^{\pi}\right)\mathbb{I}_{\{\tau<\infty\}}\right)\\
&= \int_0^{S\wedge\tau} \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathbb{I}_{\{S\wedge\tau<\infty\}}\mathrm{e}^{-\int_0^{S\wedge\tau} \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} V_\pi\left(X_{S\wedge\tau}^{\pi,x}\right),
\end{aligned}
$$

where we used the strong Markov property in the penultimate step. □

We will need the following version of the Ascoli-Arzela Theorem.

Lemma 3.2.13. *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.*

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

then there exists a uniformly convergent subsequence, i.e.

$$\exists \{f_{n_k}\}_{k\in\mathbb{N}} \subseteq \{f_n\}_{n\in\mathbb{N}} \quad \exists f \in M_2^{M_1} \quad \forall \epsilon > 0 \quad \exists N \in \mathbb{N}$$
$$\forall x \in M_1 \quad \forall k \geq N : \quad d_2(f_{n_k}(x), f(x)) < \epsilon.$$

Proof. First we will show that (M_1, d_1) (and thus every compact metric space) is separable, i.e. it contains a countable dense subset. For every $n \in \mathbb{N}$ and $x \in M_1$ define

$$B\left(x, \frac{1}{n}\right) := \left\{ y \in M_1;\ d_1(x,y) < \frac{1}{n} \right\},$$

which is the open ball of radius $\frac{1}{n}$ around x. The collection $\left\{ B\left(x, \frac{1}{n}\right);\ x \in M_1 \right\}$ is an open cover for M_1, hence by compactness there exists a finite subcover. Let S_n denote the set of all centres of the balls from this finite subcover. Then the distance from any point of M_1 to the set S_n is less than $\frac{1}{n}$, therefore the set

$$S := \bigcup_{n\in\mathbb{N}} S_n$$

is clearly dense and countable.

Now we will find a subsequence of $\{f_n\}_{n\in\mathbb{N}}$ that converges pointwise on S by the standard diagonalisation argument. Since S is countable, it can be written in the following form:

$$S = \{x_n \in M_1;\ n \in \mathbb{N}\}.$$

Since the sequence $\{f_n(x_1)\}_{n\in\mathbb{N}}$ is contained in a compact metric space, it has a convergent subsequence (compactness and sequential compactness coincide for metric spaces), which we will denote by $\{f_n^1(x_1)\}_{n\in\mathbb{N}}$. For every $k \in \mathbb{N}$, we similarly obtain a sequence

$$\{f_n^{k+1}\}_{n\in\mathbb{N}} \subseteq \{f_n^k\}_{n\in\mathbb{N}}$$

that converges at x_{k+1}. Now we look at the diagonal sequence $\{f_n^n\}_{n\in\mathbb{N}}$. It is clearly still a subsequence of the original sequence, and by construction it converges at every point of S.

In the last step we will prove that the sequence $\{f_n^n\}_{n\in\mathbb{N}}$ is uniformly Cauchy. Since every compact metric space is complete, it will follow that the sequence $\{f_n^n\}_{n\in\mathbb{N}}$ is uniformly convergent. Let $\epsilon > 0$. By equicontinuity there exists $M \in \mathbb{N}$ such that the following holds:

$$\forall x, y \in M_1 \quad \forall n \in \mathbb{N} : \quad d_1(x,y) < \frac{1}{M} \implies d_2(f_n^n(x) - f_n^n(y)) < \frac{\epsilon}{3}.$$

Recall that the set S_M is finite and that the sequence $\{f_n^n\}_{n\in\mathbb{N}}$ converges at each of its points, hence there exists $N \in \mathbb{N}$ such that

$$\forall s \in S_M \quad \forall n, m \geq N : \quad d_2(f_n^n(s), f_m^m(s)) < \frac{\epsilon}{3}.$$

Fix now $x \in M_1$ and $n, m \geq N$. By the construction of S_M, there exists $s \in S_M$ such that $d_1(x, s) < \frac{1}{M}$. Then the last two statements yield

$$\begin{aligned} &d_2(f_n^n(x), f_m^m(x)) \leq d_2(f_n^n(x), f_n^n(s)) + d_2(f_n^n(s), f_m^m(s)) + d_2(f_m^m(s), f_m^m(x)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \end{aligned}$$

which concludes the proof. □

3.2.3 Proofs

Proof of Proposition 3.2.4. If $(a, b) = \mathbb{R}$, then the Lipschitz property on compacts and boundedness of σ and μ guarantee the existence of a unique strong non-exploding solution (see for example [5, p. 45]). When (a, b) is not equal to $\mathbb{R}$, we apply the following state-space transformation. Let $\phi : (a, b) \to \mathbb{R}$ be a $\mathcal{C}^2$-diffeomorphism and consider the following stochastic differential equation (this is the equation that the process $\phi(X^{\pi,x})$ would satisfy due to Itô's lemma if we knew that the process $X^{\pi,x}$ existed):

$$\begin{aligned} Y_t &= \phi(x) + \int_0^t \left(\mu_\pi(\phi^{-1}(Y_s)) \cdot \phi'(\phi^{-1}(Y_s)) + \frac{1}{2}\sigma_\pi^2(\phi^{-1}(Y_s)) \cdot \phi''(\phi^{-1}(Y_s)) \right) \mathrm{d}s \\ &\quad + \int_0^t \sigma_\pi(\phi^{-1}(Y_s)) \cdot \phi'(\phi^{-1}(Y_s)) \, \mathrm{d}B_s, \quad 0 \leq t < \tau_{-\infty}^{\infty}(Y), \\ Y_t &= Y_{\tau_{-\infty}^{\infty}(Y)}, \quad \tau_{-\infty}^{\infty}(Y) \leq t < \infty. \end{aligned}$$

The new drift and diffusion functions are clearly Lipschitz on compacts in $\mathbb{R}$, hence there exists a unique strong solution, see for example [5, p. 45]. Then the process $\phi^{-1}(Y)$ takes values in $[a, b]$ and is the unique strong solution of (3.2). (Note that the explosion of Y corresponds to $X^{\pi,x}$ reaching a or b.) □

Proof of Proposition 3.2.5. Let $a < a' < a'' < x < b'' < b' < b$, and for any $c < d$ denote $\tau_c^d := \tau_c^d(X^{\pi,x})$. Let $v \in \mathcal{C}^2((a', b')) \cap \mathcal{C}([a', b'])$ be the unique solution of the boundary value problem

$$L_\pi v - \alpha_\pi v + f_\pi = 0, \quad v(a') = V_\pi(a'), \quad v(b') = V_\pi(b'),$$

which is guaranteed to exist by Theorem 19 in [9, p. 87] (the main assumptions in the theorem are that $\sigma_\pi^2 > 0$, $\alpha_\pi \geq 0$, and that all the coefficients are Hölder continuous, which is satisfied because we imposed the Lipschitz condition on σ, μ, α, f and π). Define the process $S^{a'',b''} = (S_t^{a'',b''})_{t\geq 0}$ and analogously $S^{a',b'}$ by

$$S_t^{a'',b''} := \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f^\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_{a''}^{b''}} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} v\left(X_{t\wedge\tau_{a''}^{b''}}^{\pi,x}\right).$$

Itô's formula on $[0,\tau_{a''}^{b''}]$ and the differential equation for v yield

$$\begin{aligned}
S_t^{a'',b''} &= v(x) + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r}\left(f_\pi + L_\pi v - \alpha_\pi v\right)\left(X_s^{\pi,x}\right)\mathrm{d}s \\
&\quad + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r}\sigma_\pi v'\left(X_s^{\pi,x}\right)\mathrm{d}B_s \\
&= v(x) + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r}\sigma_\pi v'\left(X_s^{\pi,x}\right)\mathrm{d}B_s.
\end{aligned}$$

Hence $S^{a'',b''}$ is a local martingale, and since it is clearly a bounded process, it is a uniformly integrable martingale. Thus the Dominated Convergence Theorem yields

$$v(x) = \lim_{a''\to a'}\lim_{b''\to b'} \mathbb{E}\left(S_0^{a'',b''}\right) = \lim_{a''\to a'}\lim_{b''\to b'} \mathbb{E}\left(S_\infty^{a'',b''}\right) = \mathbb{E}\left(S_\infty^{a',b'}\right).$$

Due to the boundary conditions for v and Lemma 3.2.12 we obtain

$$\begin{aligned}
S_\infty^{a',b'} &= \int_0^{\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau_{a'}^{b'}} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} v\left(X_{\tau_{a'}^{b'}}^{\pi,x}\right)\mathbb{I}_{\{\tau_{a'}^{b'}<\infty\}} \\
&= \int_0^{\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau_{a'}^{b'}} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} V_\pi\left(X_{\tau_{a'}^{b'}}^{\pi,x}\right)\mathbb{I}_{\{\tau_{a'}^{b'}<\infty\}} \\
&= \mathbb{E}\left(\int_0^{\tau_a^b} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau_a^b} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} g\left(X_{\tau_a^b}^{\pi,x}\right)\mathbb{I}_{\{\tau_a^b<\infty\}}\middle|\mathcal{F}_{\tau_{a'}^{b'}}\right),
\end{aligned}$$

and therefore

$$\begin{aligned}
v(x) &= \mathbb{E}\left(\int_0^{\tau_a^b} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau_a^b} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} g\left(X_{\tau_a^b}^{\pi,x}\right)\mathbb{I}_{\{\tau_a^b<\infty\}}\right) \\
&= V_\pi(x).
\end{aligned}$$

□

Proof of Theorem 3.2.7. Let $a < a' < x < b' < b$ and $\tau_c^d := \tau_c^d(X^{\pi_{n+1},x})$ for any $c < d$. Define the process S by

$$S_t := \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} f_{\pi_{n+1}}\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^t \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} V_{\pi_n}\left(X_t^{\pi_{n+1},x}\right), \quad t \geq 0.$$

Itô's formula, which is applicable thanks to Proposition 3.2.5, yields

$$\begin{aligned} S_{t\wedge\tau_{a'}^{b'}} = V_{\pi_n}(x) &+ \int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \sigma_{\pi_{n+1}} V'_{\pi_n}\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}B_s \\ &+ \int_0^{t\wedge\tau_{a'}^{b'}} e^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \left(f_{\pi_{n+1}} + L_{\pi_{n+1}} V_{\pi_n} - \alpha_{\pi_{n+1}} V_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s. \end{aligned}$$

The stochastic integral is a martingale since the functions $\sigma_{\pi_{n+1}}$ and V'_{π_n} are bounded on $[a', b']$ by Assumption 3.2.1 and Proposition 3.2.5, respectively. Hence we obtain

$$\begin{aligned} E\left(S_{t\wedge\tau_{a'}^{b'}}\right) = V_{\pi_n}(x) + E\Bigg(&\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \\ &\cdot\sigma_{\pi_{n+1}}^2 \left(\frac{f_{\pi_{n+1}}}{\sigma_{\pi_{n+1}}^2} + \frac{L_{\pi_{n+1}} V_{\pi_n}}{\sigma_{\pi_{n+1}}^2} - \frac{\alpha_{\pi_{n+1}} V_{\pi_n}}{\sigma_{\pi_{n+1}}^2}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\Bigg). \end{aligned}$$

By the definition of the policy improvement algorithm (3.3) we get

$$\begin{aligned} &E\left(S_{t\wedge\tau_{a'}^{b'}}\right) \\ &= V_{\pi_n}(x) + E\Bigg(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \\ &\qquad\cdot\sigma_{\pi_{n+1}}^2 \min_{p\in A}\left(\frac{f_p}{\sigma_p^2} + \frac{L_p V_{\pi_n}}{\sigma_p^2} - \frac{\alpha_p V_{\pi_n}}{\sigma_p^2}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\Bigg) \\ &\leq V_{\pi_n}(x) + E\Bigg(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \\ &\qquad\cdot\frac{\sigma_{\pi_{n+1}}^2}{\sigma_{\pi_n}^2}\left(f_{\pi_n} + L_{\pi_n} V_{\pi_n} - \alpha_{\pi_n} V_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\Bigg) \\ &= V_{\pi_n}(x), \end{aligned}$$

where we used Proposition 3.2.5 in the last step. By recalling the definition of S_t

and applying the Dominated Convergence Theorem as t tends to ∞, we obtain

$$V_{\pi_n}(x) \geq \mathbb{E}\left(\int_0^{\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}(X_r^{\pi_{n+1},x})\mathrm{d}r} f_{\pi_{n+1}}(X_s^{\pi_{n+1},x})\,\mathrm{d}s \right.$$
$$\left. + \mathrm{e}^{-\int_0^{\tau_{a'}^{b'}} \alpha_{\pi_{n+1}}(X_r^{\pi_{n+1},x})\mathrm{d}r} V_{\pi_n}\left(X_{\tau_{a'}^{b'}}^{\pi_{n+1},x}\right) \mathbb{I}_{\{\tau_{a'}^{b'}<\infty\}}\right).$$

Now we send a' to a and b' to b, and applying the Dominated Convergence Theorem, Lemma 3.2.11 and the definition of $V^{\pi_{n+1}}$ we obtain

$$\begin{aligned} V_{\pi_n}(x) &\geq \mathbb{E}\left(\int_0^{\tau_a^b} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}(X_r^{\pi_{n+1},x})\mathrm{d}r} f_{\pi_{n+1}}(X_s^{\pi_{n+1},x})\,\mathrm{d}s \right. \\ &\qquad \left. + \mathrm{e}^{-\int_0^{\tau_a^b} \alpha_{\pi_{n+1}}(X_r^{\pi_{n+1},x})\mathrm{d}r} g\left(X_{\tau_a^b}^{\pi_{n+1},x}\right) \mathbb{I}_{\{\tau_a^b<\infty\}}\right) \\ &= V_{\pi_{n+1}}(x), \end{aligned}$$

which is what we had to prove. □

Proof of Proposition 3.2.8. Let the sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ be such that the following holds for every $k \in \mathbb{N}$:

$$a < a_{k+1} < a_k < b_k < b_{k+1} < b \quad \text{and} \quad \bigcup_{n\in\mathbb{N}} [a_n, b_n] = (a, b).$$

Applying Interior Estimates[1] from [9, p. 86], Assumption 3.2.1 and Proposition 3.2.5 to the sequences $\{\sigma_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\mu_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\alpha_{\pi_n}\}_{n\in\mathbb{N}}$, $\{f_{\pi_n}\}_{n\in\mathbb{N}}$ and $\{V_{\pi_n}\}_{n\in\mathbb{N}}$, we obtain that the sequence $\{V''_{\pi_n}\}_{n\in\mathbb{N}}$ is uniformly bounded on $[a_k, b_k]$ for every $k \in \mathbb{N}$, which implies that the sequence $\{V'_{\pi_n}\}_{n\in\mathbb{N}}$ is uniformly Lipschitz on $[a_k, b_k]$ for every $k \in \mathbb{N}$. Recalling the policy improvement algorithm (3.3) and applying Assumption 3.2.2, we obtain that the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ is uniformly Lipschitz and hence equicontinuous on $[a_k, b_k]$ for every $k \in \mathbb{N}$. Define $\pi_n^0 := \pi_n$ for every $n \in \mathbb{N}$. Thanks to the version of the Arzela-Ascoli Theorem in Lemma 3.2.13, for every $k \in \mathbb{N}$ there exists a subsequence $\{\pi_n^k\}_{n\in\mathbb{N}} \subseteq \{\pi_n^{k-1}\}_{n\in\mathbb{N}}$ such that $\{\pi_n^k\}_{n\in\mathbb{N}}$ converges uniformly on $[a_k, b_k]$. The diagonal sequence, i.e. $\{\pi_n^n\}_{n\in\mathbb{N}}$, then converges uniformly on $[a_k, b_k]$ for every $k \in \mathbb{N}$, and hence on every compact subset of (a, b). □

[1] The theorem considers a family of uniformly elliptic differential equations. Roughly speaking, it says that if the coefficients have certain uniform boundedness properties away from the boundary, then the solutions (if they are bounded) have similar properties.

Proof of Theorem 3.2.9. Let $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ be a sequence from Proposition 3.2.8 that converges to $\pi_{\lim}$ uniformly on compacts in (a,b). Let $k\in\mathbb{N}$, $a<a'<x<b'<b$ and $\tau_{a'}^{b'}:=\tau_{a'}^{b'}(X^{\pi_{\lim},x})$. Define the process $S=(S_t)_{t\geq 0}$ by

$$S_t:=\int_0^t \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r} f_{\pi_{n_k}}\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s+\mathrm{e}^{-\int_0^t \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r} V_{\pi_{n_k}}\left(X_t^{\pi_{\lim},x}\right).$$

Itô's formula yields

$$\begin{aligned}S_{t\wedge\tau_{a'}^{b'}}&=V_{\pi_{n_k}}(x)+\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\,\sigma_{\pi_{\lim}} V'_{\pi_{n_k}}\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}B_s\\&\quad+\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(f_{\pi_{n_k}}+L_{\pi_{\lim}}V_{\pi_{n_k}}-\alpha_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s.\end{aligned}$$

The stochastic integral is a martingale since the functions $\sigma_{\pi_{\lim}}$ and $V'_{\pi_{n_k}}$ are bounded on $[a',b']$ (by Assumption 3.2.1 and Proposition 3.2.5). Hence we obtain

$$\begin{aligned}E\left(S_{t\wedge\tau_{a'}^{b'}}\right)&=V_{\pi_{n_k}}(x)\\&\quad+E\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(f_{\pi_{n_k}}+L_{\pi_{\lim}}V_{\pi_{n_k}}-\alpha_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right)\\&=V_{\pi_{n_k}}(x)+E\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(L_{\pi_{\lim}}V_{\pi_{n_k}}-L_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right)\\&=V_{\pi_{n_k}}(x)+E\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\right.\\&\qquad\left.\cdot\left(\frac{1}{2}\left(\sigma_{\pi_{\lim}}^2-\sigma_{\pi_{n_k}}^2\right)V''_{\pi_{n_k}}+\left(\mu_{\pi_{\lim}}-\mu_{\pi_{n_k}}\right)V'_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right),\end{aligned}$$

where we used Proposition 3.2.5 and the definition of the operator L. Applying Interior Estimates from [9, p. 86], Assumption 3.2.1 and Proposition 3.2.5 to the sequences $\{\sigma_{\pi_{nm}}\}_{m\in\mathbb{N}}$, $\{\mu_{\pi_{nm}}\}_{m\in\mathbb{N}}$, $\{\alpha_{\pi_{nm}}\}_{m\in\mathbb{N}}$, $\{f_{\pi_{nm}}\}_{m\in\mathbb{N}}$ and $\{V_{\pi_{nm}}\}_{m\in\mathbb{N}}$, we obtain that the sequences $\{V'_{\pi_{nm}}\}_{m\in\mathbb{N}}$ and $\{V''_{\pi_{nm}}\}_{m\in\mathbb{N}}$ are uniformly bounded on $[a',b']$. Now the Dominated Convergence Theorem yields that the last term disap-

pears when k tends to ∞, hence we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\pi_{\lim}}(X_r^{\pi_{\lim},x})\mathrm{d}r} f_{\pi_{\lim}}(X_s^{\pi_{\lim},x})\,\mathrm{d}s\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{t\wedge\tau_{a'}^{b'}} \alpha_{\pi_{\lim}}(X_r^{\pi_{\lim},x})\mathrm{d}r} V_{\lim}\left(X_{t\wedge\tau_{a'}^{b'}}^{\pi_{\lim},x}\right)\right)\\
&= V_{\lim}(x).
\end{aligned}$$

By first sending t to ∞ and then a' to a and b' to b, we obtain the desired equality as in the previous proof. □

When it comes to the policies, we only know that a subsequence converges. The problem with the subsequence is that the policies are no longer improvements of their predecessors. However, this problem can be overcome by considering the original sequence and the shifted sequences as a two-dimensional sequence of policies, as we shall see.

Proof of Theorem 3.2.10. The second assertion follows from Theorem 3.2.9.

Consider the sequence $\{(\pi_{n+1},\pi_n) : (a,b) \to A\times A\}_{n\in\mathbb{N}}$, where $A\times A$ is equipped with any p-product metric[2], $p\in[1,\infty]$. In the same way as in the proof of Proposition 3.2.8 we can find a subsequence $\{(\pi_{1+n_k},\pi_{n_k})\}_{k\in\mathbb{N}}$ that is uniformly convergent on every compact set in (a,b).

For every $k\in\mathbb{N}$, let

$$\hat{\sigma}_k(\cdot) := \sigma(\cdot,\pi_{n_k}(\cdot)),\quad \hat{\pi}_\infty(\cdot) := \lim_{m\to\infty}\pi_{n_m}(\cdot)\quad\text{and}\quad \hat{\sigma}_\infty(\cdot) := \sigma(\cdot,\hat{\pi}_\infty(\cdot)).$$

Define $\hat{\mu}_k$, $\hat{\alpha}_k$, $\hat{f}_k$ and $\hat{\mu}_\infty$, $\hat{\alpha}_\infty$, $\hat{f}_\infty$ in a corresponding fashion. Similarly, let

$$\tilde{\sigma}_k(\cdot) := \sigma(\cdot,\pi_{n_k+1}(\cdot)),\quad \tilde{\pi}_\infty(\cdot) := \lim_{m\to\infty}\pi_{n_m+1}(\cdot)\quad\text{and}\quad \tilde{\sigma}_\infty(\cdot) := \sigma(\cdot,\tilde{\pi}_\infty(\cdot)),$$

and define $\tilde{\mu}_k$, $\tilde{\alpha}_k$, $\tilde{f}_k$ and $\tilde{\mu}_\infty$, $\tilde{\alpha}_\infty$, $\tilde{f}_\infty$ in a corresponding fashion. Set

$$\hat{v}_k(\cdot) := V_{\pi_{n_k}}(\cdot),\quad \tilde{v}_k(\cdot) := V_{\pi_{n_k+1}}(\cdot),\quad\text{and}\quad v(\cdot) := V_{\lim}(\cdot),$$

[2] Let (X,d_X) and (Y,d_Y) be metric spaces, and for any $(x_1,y_1),(x_2,y_2)\in X\times Y$ define

$$d_p((x_1,y_1),(x_2,y_2)) := \left(d_X(x_1,x_2)^p + d_Y(y_1,y_2)^p\right)^{\frac{1}{p}}\quad\text{for}\quad p\in[1,\infty),$$

$$d_\infty((x_1,y_1),(x_2,y_2)) := \max\left(\{d_X(x_1,x_2),d_Y(y_1,y_2)\}\right).$$

It is well-known that $\{d_p;\ p\in[0,\infty]\}$ is a family of equivalent metrics on $X\times Y$.

and define the operators $\hat{\mathcal{L}}_k$ and $\hat{\mathcal{L}}_\infty$ by

$$\hat{\mathcal{L}}_k u := \frac{1}{2}\hat{\sigma}_k^2 u'' + \hat{\mu}_k u' - \hat{\alpha}_k u + \hat{f}_k, \quad \text{and} \quad \hat{\mathcal{L}}_\infty u := \frac{1}{2}\hat{\sigma}_\infty^2 u'' + \hat{\mu}_\infty u' - \hat{\alpha}_\infty u + \hat{f}_\infty,$$

with corresponding definitions for $\tilde{\mathcal{L}}_k$ and $\tilde{\mathcal{L}}_\infty$. Applying the last part of Theorem 15[3] from [9, p. 80], Proposition 3.2.5 and Assumption 3.2.1, we obtain that both $\hat{\mathcal{L}}_\infty v = 0$ and $\tilde{\mathcal{L}}_\infty v = 0$ hold on every compact subset of (a, b), and that the sequence of functions $\{\frac{1}{2}(\tilde{\sigma}_k^2 - \hat{\sigma}_k^2)\hat{v}_k'' + (\tilde{\mu}_k - \hat{\mu}_k)\hat{v}_k' - (\tilde{\alpha}_k - \hat{\alpha}_k)\hat{v}_k + \tilde{f}_k - \hat{f}_k\}_{k\in\mathbb{N}}$ converges uniformly to $\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)v'' + (\tilde{\mu}_\infty - \hat{\mu}_\infty)v' - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)v + \tilde{f}_\infty - \hat{f}_\infty$ on every compact subset of (a, b). However, we know that

$$\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)v'' + (\tilde{\mu}_\infty - \hat{\mu}_\infty)v' - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)v + \tilde{f}_\infty - \hat{f}_\infty = \tilde{\mathcal{L}}_\infty v - \hat{\mathcal{L}}_\infty v = 0. \quad (3.4)$$

[3] The theorem is stated for parabolic equations, but is also valid for the elliptic ones; in the same way as the parabolic version follows from Theorem 5, p. 64, the elliptic version follows from Interior Estimates, p. 86. Given the setting from the Interior Estimates, the theorem establishes the convergence of solutions of uniformly elliptic differential equations provided that the coefficients converge.

Now let $k \in \mathbb{N}$, $a < a' < x < b' < b$ and $\tau_{a'}^{b'} := \tau_{a'}^{b'}(X^{\Pi,x})$. We obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} f_{\Pi_s}\left(X_s^{\Pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_{a'}^{b'}} \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} V_{\pi_{n_k}}\left(X_{t\wedge\tau_{a'}^{b'}}^{\Pi,x}\right)\right)\\
&\overset{\text{Itô}}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\left(f_{\Pi_s} + L_{\Pi_s}V_{\pi_{n_k}} - \alpha_{\Pi_s}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right)\\
&\geq V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\,\sigma_{\Pi_s}^2\right.\\
&\quad\left.\cdot \min_{p\in A}\left(\frac{f_p}{\sigma_p^2} + \frac{L_pV_{\pi_{n_k}}}{\sigma_p^2} - \frac{\alpha_pV_{\pi_{n_k}}}{\sigma_p^2}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right)\\
&\overset{\text{PIA}(3.3)}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\frac{\sigma_{\Pi_s}^2}{\sigma_{\pi_{n_k+1}}^2}\right.\\
&\quad\left.\cdot\left(f_{\pi_{n_k+1}} + L_{\pi_{n_k+1}}V_{\pi_{n_k}} - \alpha_{\pi_{n_k+1}}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right)\\
&\overset{\text{Prop. 3.2.5}}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\frac{\sigma_{\Pi_s}^2}{\sigma_{\pi_{n_k+1}}^2}\right.\\
&\quad\left.\cdot\left(f_{\pi_{n_k+1}} - f_{\pi_{n_k}} + L_{\pi_{n_k+1}}V_{\pi_{n_k}} - L_{\pi_{n_k}}V_{\pi_{n_k}} - \alpha_{\pi_{n_k+1}}V_{\pi_{n_k}} + \alpha_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right)\\
&= V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\frac{\sigma_{\Pi_s}^2}{\sigma_{\pi_{n_k+1}}^2}\right.\\
&\quad\left.\cdot\left(\frac{1}{2}(\tilde{\sigma}_k^2 - \hat{\sigma}_k^2)\hat{v}_k'' + (\tilde{\mu}_k - \hat{\mu}_k)\hat{v}_k' - (\tilde{\alpha}_k - \hat{\alpha}_k)\hat{v}_k + \tilde{f}_k - \hat{f}_k\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right),
\end{aligned}
$$

Sending k to ∞, applying the Dominated Convergence Theorem and using (3.4), we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} f_{\Pi_s}\left(X_s^{\Pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_{a'}^{b'}} \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} V_{\lim}\left(X_{t\wedge\tau_{a'}^{b'}}^{\Pi,x}\right)\right)\\
&\geq V_{\lim}(x).
\end{aligned}
$$

By first sending t to ∞ and then a' to a and b' to b, we obtain the desired inequality in the usual way. □

3.3 Multidimensional case

3.3.1 Setting and the algorithm

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space (satisfying the usual conditions) that supports a d-dimensional $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B_t)_{t\geq 0}$ ($d \in \mathbb{N}$). Let $(A, \tilde{d})$ be a compact metric space, and for any $x \in \mathbb{R}^d$ define the set of admissible controls at x as

$$\begin{aligned}\mathcal{A}(x) := \{&\Pi = (\Pi_t)_{t\geq 0};\ \Pi \text{ is an } A\text{-valued process adapted to } (\mathcal{F}_t)_{t\geq 0}, \text{ and} \\ &\text{there exists a pathwise unique process } X^{\Pi,x} = \left(X_t^{\Pi,x}\right)_{t\geq 0} \text{ that satisfies (3.5)}\},\end{aligned}$$

where

$$X_t^{\Pi,x} = x + \int_0^t \sigma\left(X_s^{\Pi,x}, \Pi_s\right) \mathrm{d}B_s + \int_0^t \mu\left(X_s^{\Pi,x}, \Pi_s\right) \mathrm{d}s, \quad t \geq 0, \tag{3.5}$$

and $\sigma : \mathbb{R}^d \times A \to \mathbb{R}^{d\times d}$ and $\mu : \mathbb{R}^d \times A \to \mathbb{R}^d$ are measurable mappings.

Let $\alpha : \mathbb{R}^d \times A \to \mathbb{R}$ and $f : \mathbb{R}^d \times A \to \mathbb{R}$ be measurable functions. For any $x \in \mathbb{R}^d$ and $\Pi \in \mathcal{A}(x)$ define the payoff as

$$V_\Pi(x) := \mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha\left(X_s^{\Pi,x}, \Pi_s\right)\mathrm{d}s} f\left(X_t^{\Pi,x}, \Pi_t\right) \mathrm{d}t\right).$$

The problem is to find the *value function* V, defined by

$$V(x) := \inf_{\Pi\in\mathcal{A}(x)} V_\Pi(x), \quad x \in \mathbb{R}^d,$$

and an optimal control (which will in general depend on x), if it exists.

Before we continue, we have to mentions the norms. For any vector $v \in \mathbb{R}^d$, $||v||$ will denote its Euclidean norm. For a matrix $M \in \mathbb{R}^{d\times d}$, we will have

$$||M|| := \sup\left\{\frac{||Mv||}{||v||};\ v \in \mathbb{R}^d\backslash\{0\}\right\} = \sqrt{\lambda_{\max}(M^T M)}.$$

This is the spectral norm, which is defined as the induced Euclidean norm, and is equal to the largest singular value of M, i.e. the square root of the largest eigenvalue of the positive-semidefinite matrix $M^T M$. When we say that a function that maps to $\mathbb{R}^d$ or $\mathbb{R}^{d\times d}$ is bounded, we mean that its norm is a bounded function. It is worth noting that all the norms on finite-dimensional spaces are equivalent. The reason why we chose the spectral (and not for example Frobenius) norm is the

representation with the singular value, which will come in useful.

In order to solve the problem, we make the following assumptions about the functions σ, μ, α and f.

Assumption 3.3.1. The functions σ, μ, α and f are bounded, and Lipschitz on compacts in $\mathbb{R}^d \times A$, i.e. for every compact set $K \subseteq \mathbb{R}^d$ there exists a constant $C > 0$ such that

$$||h(x,p) - h(y,r)|| \leq C \left(||x-y||^2 + \tilde{d}(p,r)^2 \right)^{\frac{1}{2}}$$

holds for every $x, y \in K$, $p, r \in A$ and $h \in \{\sigma, \mu, \alpha, f\}$. In addition, α is positive and bounded away from 0, and there exists $\lambda > 0$ such that

$$\sum_{i,j=1}^{d} (\sigma(x,p)^T \sigma(x,p))_{i,j} v_i v_j \geq \lambda ||v||^2 \qquad \text{for all} \quad x \in \mathbb{R}^d, p \in A, v \in \mathbb{R}^d. \tag{3.6}$$

Remark 3.3.2. (3.6) is the uniform ellipticity condition, and it is the multidimensional analogue of being bounded away from 0. It implies (for all x and p) that the smallest eigenvalue of $\sigma(x,p)^T\sigma(x,p)$ is at least as big as λ, and that $\sigma(x,p)$ is invertible.

Assumption 3.3.3. For every $h \in \mathcal{C}^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, let $I_h(x)$ denote a point where the minimum of the function

$$p \mapsto \frac{1}{2} \operatorname{Tr}\left(\sigma(x,p)^T \mathrm{H}h(x) \sigma(x,p)\right) + \mu(x,p)^T \nabla h(x) - \alpha(x,p)h(x) + f(x,p), \quad p \in A,$$

is attained ($\mathrm{H}h$ is the Hessian matrix of the second derivatives of h, and we used the standard notation for the trace, transpose and gradient). If the sequence $\{\mathrm{H}h_n\}_{n\in\mathbb{N}}$ is uniformly bounded on compacts in $\mathbb{R}^d$, then the points $\{I_{h_n}(x);\ x \in \mathbb{R}^d, n \in \mathbb{N}\}$ can be chosen in such a way that the sequence of functions $\{I_{h_n}\}_{n\in\mathbb{N}}$ ($I_{h_n} : \mathbb{R}^d \to A$) is uniformly Lipschitz on compacts in $\mathbb{R}^d$.

Remark 3.3.4. It is important to note that there are non-trivial data that satisfy the above assumptions. Some are presented in Proposition 3.4.2.

It is time to introduce Markov policies. A measurable function $\pi : \mathbb{R}^d \to A$ is a *Markov policy* if for every $x \in \mathbb{R}^d$ there exists a pathwise unique $\mathbb{R}^d$-valued process $X^{\pi,x} = (X_t^{\pi,x})_{t\geq 0}$ that satisfies the following:

$$X_t^{\pi,x} = x + \int_0^t \sigma\left(X_s^{\pi,x}, \pi\left(X_s^{\pi,x}\right)\right) \mathrm{d}B_s + \int_0^t \mu\left(X_s^{\pi,x}, \pi\left(X_s^{\pi,x}\right)\right) \mathrm{d}s, \quad t \geq 0. \tag{3.7}$$

If π is a Markov policy, then $\pi(X^{\pi,x}) := (\pi(X_t^{\pi,x}))_{t\geq 0} \in \mathcal{A}(x)$ for every $x \in \mathbb{R}^d$. For

easier notation we define

$$\sigma_\pi(\cdot) := \sigma(\cdot, \pi(\cdot)), \quad \mu_\pi(\cdot) := \mu(\cdot, \pi(\cdot)), \quad \alpha_\pi(\cdot) := \alpha(\cdot, \pi(\cdot)), \quad f_\pi(\cdot) := f(\cdot, \pi(\cdot)),$$

$$V_\pi(\cdot) := V_{\pi(X^{\pi,\cdot})}(\cdot), \quad \text{and} \quad L_\pi h := \frac{1}{2} \operatorname{Tr}\left(\sigma_\pi^T \mathrm{H} h \, \sigma_\pi\right) + \mu_\pi^T \nabla h \quad \text{for} \quad h \in \mathcal{C}^2(\mathbb{R}^d).$$

If π is a constant Markov policy with the value $p \in A$, we will write σ_p, μ_p, α_p, f_p and L_p instead of σ_π, μ_π, α_π, f_π and L_π, respectively.

According to [5, p. 45] and thanks to Assumption 3.3.1, for every $x \in \mathbb{R}^d$ and function $\pi : \mathbb{R}^d \to A$ that is Lipschitz on compacts in $\mathbb{R}^d$, the stochastic differential equation in (3.7) has a unique strong non-exploding solution, so every such π is a Markov policy.

It turns out that the payoff functions of Lipschitz Markov policies satisfy the following differential equation.

Proposition 3.3.5. *For any Markov policy π that is Lipschitz on compacts in $\mathbb{R}^d$, the following holds: $V_\pi \in \mathcal{C}^2(\mathbb{R}^d)$ and*

$$L_\pi V_\pi - \alpha_\pi V_\pi + f_\pi = 0.$$

Now we can present the algorithm. Let π_0 be a Markov policy that is Lipschitz on compacts in $\mathbb{R}^d$. The *policy improvement algorithm* is defined in the following way:

$$\pi_{n+1}(x) := \operatorname*{argmin}_{p \in A} \left(L_p V_{\pi_n}(x) - \alpha_p(x) V_{\pi_n}(x) + f_p(x)\right), \quad x \in \mathbb{R}^d, \quad n \in \mathbb{N}_0. \tag{3.8}$$

Note that the following equality holds:

$$\pi_{n+1}(x)$$
$$= \operatorname*{argmin}_{p \in A} \left(\frac{1}{2} \operatorname{Tr}\left(\sigma_p(x)^T \, \mathrm{H} V_{\pi_n}(x) \, \sigma_p(x)\right) + \mu_p(x)^T \nabla V_{\pi_n}(x) - \alpha_p(x) V_{\pi_n}(x) + f_p(x) \right).$$

Hence Assumption 3.3.3 (applied separately for each $n \in \mathbb{N}_0$) ensures that the function $\pi_{n+1} : \mathbb{R}^d \to A$ is Lipschitz on compacts in $\mathbb{R}^d$ (since $\mathrm{H} V_{\pi_n}$ is bounded on compacts in $\mathbb{R}^d$ by Proposition 3.3.5).

Remark 3.3.6. If the algorithm stops, i.e. $\pi_{n+1} = \pi_n$ for some $n \in \mathbb{N}_0$, then clearly $V_{\pi_m} = V_{\pi_n}$ and $\pi_m = \pi_n$ hold for every $m \geq n$. We can then proceed directly to the verification lemma (Theorem 3.3.10) to prove that V_{π_n} is the value function and π_n is an optimal policy. In general we first need to establish the existence of the limit payoff function and policy.

The next theorem confirms that the policy is indeed improved at each step.

Theorem 3.3.7. *For every $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ and Markov policy π_0 that is Lipschitz on compacts in $\mathbb{R}^d$, the following holds:*

$$V_{\pi_{n+1}}(x) \leq V_{\pi_n}(x).$$

Since $\{V_{\pi_n}\}_{n\in\mathbb{N}}$ is a decreasing bounded sequence, we can define

$$V_{\lim}(x) := \lim_{n\to\infty} V_{\pi_n}(x), \quad x \in \mathbb{R}^d.$$

The sequence of policies might not converge, but the next proposition says that there exists a convergent subsequence.

Proposition 3.3.8. *There exists a subsequence of $\{\pi_n\}_{n\in\mathbb{N}}$ that converges uniformly on every compact subset of $\mathbb{R}^d$.*

Therefore this subsequence converges (pointwise) on $\mathbb{R}^d$; denote the limit by $\pi_{\lim}$. Note that $V_{\lim}$ and $\pi_{\lim}$ can in principle depend on π_0. Additionally, $\pi_{\lim}$ could depend on the choice of the subsequence from Proposition 3.3.8. However, this turns out to be irrelevant in the following theorem, which brings together the limit payoff function and the limit policy.

Theorem 3.3.9. *For every $x \in \mathbb{R}^d$ and Markov policy π_0 that is Lipschitz on compacts in $\mathbb{R}^d$, the following holds:*

$$V_{\lim}(x) = V_{\pi_{\lim}}(x).$$

The last step establishes $V_{\lim} = V$ via the so called verification lemma.

Theorem 3.3.10. *For every $x \in \mathbb{R}^d$, $\Pi \in \mathcal{A}(x)$ and Markov policy π_0 that is Lipschitz on compacts in $\mathbb{R}^d$, the following holds:*

$$V_{\lim}(x) \leq V_{\Pi}(x).$$

Hence $V_{\lim}$ is the value function (and does not depend on π_0) and $\pi_{\lim}$ is an optimal policy.

3.3.2 Auxiliary results

The processes controlled by Markov policies are strong Markov processes (Theorem 4.20 in [17, p. 322]). This enables us to prove the following lemma.

Lemma 3.3.11. *The following holds for every Markov policy π, $x \in \mathbb{R}^d$ and stopping time S:*

$$\begin{aligned}&\mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t \middle| \mathcal{F}_S\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathrm{e}^{-\int_0^S \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} V_\pi\left(X_S^{\pi,x}\right).\end{aligned}$$

In particular, the process M is a uniformly integrable martingale, where

$$M_r := \int_0^r \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathrm{e}^{-\int_0^r \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} V_\pi\left(X_r^{\pi,x}\right), \quad r \geq 0.$$

Remark 3.3.12. Note that the expression $\mathrm{e}^{-\int_0^S \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} V_\pi\left(X_S^{\pi,x}\right)$ is well-defined on the event $\{S = \infty\}$ since $\mathrm{e}^{-\int_0^\infty \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} = 0$ (because α is positive and bounded away from 0) and V_π is a bounded function (recall Assumption 3.3.3).

Proof. Let $\tau := \tau_a^b\left(X^{\pi,x}\right)$ and $\tau_S := \tau \circ \theta_{S\wedge\tau} = \tau_a^b\left(X^{\pi,x}_{\cdot+S\wedge\tau}\right)$. Then $\tau_S = \tau - S \wedge \tau$ holds almost surely. We obtain

$$\begin{aligned}&\mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t \middle| \mathcal{F}_S\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t\\&\quad + \mathbb{E}\left(\mathbb{I}_{\{S<\infty\}} \int_0^\infty \mathrm{e}^{-\int_0^{t+S} \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_{t+S}^{\pi,x}\right)\mathrm{d}t \middle| \mathcal{F}_S\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t\\&\quad + \mathbb{I}_{\{S<\infty\}} \mathrm{e}^{-\int_0^S \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\, \mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_{s+S}^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_{t+S}^{\pi,x}\right)\mathrm{d}t \middle| \mathcal{F}_S\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t\\&\quad + \mathbb{I}_{\{S<\infty\}} \mathrm{e}^{-\int_0^S \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\, \mathbb{E}_x\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi}\right)\mathrm{d}t \circ \theta_S \middle| \mathcal{F}_S\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t\\&\quad + \mathbb{I}_{\{S<\infty\}} \mathrm{e}^{-\int_0^S \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t}\, \mathbb{E}_{X_S^{\pi,x}}\left(\int_0^\infty \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi}\right)\mathrm{d}t\right)\\&= \int_0^S \mathrm{e}^{-\int_0^t \alpha_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s} f_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t + \mathbb{I}_{\{S<\infty\}} \mathrm{e}^{-\int_0^S \alpha_\pi\left(X_t^{\pi,x}\right)\mathrm{d}t} V_\pi(X_S^{\pi,x}),\end{aligned}$$

where we used the strong Markov property in the penultimate step. □

In the next two rather lengthy lemmas we will prove that the payoff functions are continuous. The main part is to show that we can couple the controlled processes started close together in such a way that the probability of not coupling soon is small. As shown in the following lemma, this can be achieved by the mirror coupling, which in the case of multidimensional diffusions turns out to be much more complicated than for one-dimensional Brownian motions.

Lemma 3.3.13. *For every Lipschitz Markov control and small enough $\epsilon > 0$ there exists $\delta > 0$ such that the following holds for every $x, y \in \mathbb{R}^d$: if $||x - y|| < \delta$ then there exist processes $\tilde{X}^{x,\pi}$ and $\tilde{X}^{y,\pi}$ that have the same laws as $X^{x,\pi}$ and $X^{y,\pi}$, respectively, such that*

$$\left|\left|\tilde{X}_t^{x,\pi} - \tilde{X}_t^{y,\pi}\right|\right|^2 \leq S_{\tau_t} \quad on \quad \{t < \rho_\delta\}$$

and

$$\tilde{X}_t^{x,\pi} = \tilde{X}_t^{y,\pi} \quad on \quad \{t \geq \rho_0\}$$

for every $t \geq 0$, where

$$\rho_c := \inf\left\{t \geq 0;\ \left|\left|\tilde{X}^{x,\pi} - \tilde{X}^{y,\pi}\right|\right| = c\right\}, \quad (\inf \emptyset = \infty),$$

for any $c \geq 0$, S is the squared Bessel process of dimension $1 + \epsilon$ started at $||x - y||$, and $(\tau_t)_{t \geq 0}$ is a stochastic time change with the property

$$\tau_t \leq \frac{t}{\lambda}, \quad t \geq 0.$$

Proof. Let $\tilde{X}^{\pi,x}$ be equal to $X^{\pi,x}$, which is defined as the solution of (3.7). Let $\hat{X}^{\pi,y}$ be the solution of the following stochastic differential equation (it exists and is unique, see [23]):

$$\hat{X}_t^{\pi,y} = y + \int_0^t \mu_\pi\left(\hat{X}_s^{\pi,y}\right) \mathrm{d}s + \int_0^t \sigma_\pi\left(\hat{X}_s^{\pi,y}\right) H_s \,\mathrm{d}B_s, \quad t \geq 0,$$

where

$$H_t := I - 2u_t u_t^T, \quad t \geq 0,$$

(note that H_t is a reflection in $\mathbb{R}^d$ about the hyperplane that is orthogonal to the

unit vector u_t and contains the origin), and

$$u_t := \frac{\sigma_\pi^{-1}\left(\hat{X}_t^{\pi,y}\right) X_t}{\left|\left|\sigma_\pi^{-1}\left(\hat{X}_t^{\pi,y}\right) X_t\right|\right|}, \quad X_t := X_t^{\pi,x} - \hat{X}_t^{\pi,y}, \quad t \geq 0.$$

By Lévy's characterisation theorem the process

$$\int_0^t H_s \, \mathrm{d}B_s, \quad t \geq 0,$$

is a d-dimensional Brownian motion, hence $\hat{X}^{\pi,y}$ has the same law as $X^{\pi,y}$. Defining

$$\tilde{X}_t^{\pi,y} := \hat{X}_t^{\pi,y} \mathbb{I}_{\{t \leq \rho_0\}} + X_t^{\pi,x} \mathbb{I}_{\{t > \rho_0\}}, \quad t \geq 0,$$

we see (by the strong Markov property) that $\tilde{X}^{\pi,y}$ has the same law as $X^{\pi,y}$, and it is obviously equal to $X^{\pi,x}$ after ρ_0.

For easier notation, set (for every $t \geq 0$)

$$\bar{S}_t := ||X_t||^2, \quad A_t := \sigma_\pi\left(X_t^{\pi,x}\right), \quad C_t := \sigma_\pi\left(\hat{X}_t^{\pi,y}\right),$$

$$v_t := \frac{X_t}{||X_t||}, \quad b_t := \mu_\pi\left(X_t^{\pi,x}\right) - \mu_\pi\left(\hat{X}_t^{\pi,y}\right).$$

Itô's lemma, applied to the function

$$h(x_1, \cdots, x_d) := x_1^2 + \cdots + x_d^2,$$

yields for every $t \geq 0$

$$\begin{aligned}
\bar{S}_t &= ||x-y|| + \int_0^t \nabla h\left(X_s\right)^T \left(A_s - C_s H_s\right) \mathrm{d}B_s \\
&\quad + \int_0^t \left(\nabla h\left(X_s\right)^T b_t + \frac{1}{2} \operatorname{Tr}\left(\left(A_s - C_s H_s\right)^T \mathrm{H} h\left(X_s\right) \left(A_s - C_s H_s\right)\right)\right) \mathrm{d}s \\
&= ||x-y|| + \int_0^t 2\sqrt{\bar{S}_s} v_s^T \left(A_s - C_s H_s\right) \mathrm{d}B_s \\
&\quad + \int_0^t \left(2\sqrt{\bar{S}_s} v_s^T b_t + \operatorname{Tr}\left(\left(A_s - C_s H_s\right)\left(A_s - C_s H_s\right)^T\right)\right) \mathrm{d}s.
\end{aligned} \tag{3.9}$$

Before we can time-change the process $\bar{S}$ and compare it to the Bessel process, we need to make quite a long and technical detour. First we will prove

$$\operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \left|\left|v_t^T \alpha_t\right|\right|^2 = \operatorname{Tr}\left(\beta_t \beta_t^T\right) - \left|\left|v_t^T \beta_t\right|\right|^2, \quad t \geq 0, \tag{3.10}$$

where

$$\alpha_t := A_t - C_t H_t \quad \text{and} \quad \beta_t := A_t - C_t, \quad t \geq 0.$$

We obtain

$$\begin{aligned}
&\operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \left|\left|v_t^T \alpha_t\right|\right|^2 - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \left|\left|v_t^T \beta_t\right|\right|^2 \\
&= \operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \operatorname{Tr}\left(v_t^T \alpha_t \alpha_t^T v_t\right) - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \operatorname{Tr}\left(v_t^T \beta_t \beta_t^T v_t\right) \\
&= \operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \operatorname{Tr}\left(\alpha_t \alpha_t^T v_t v_t^T\right) - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \operatorname{Tr}\left(\beta_t \beta_t^T v_t v_t^T\right),
\end{aligned}$$

where we applied the invariance property for trace under cyclic permutations. Simplifying further we get

$$\begin{aligned}
&\operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \left|\left|v_t^T \alpha_t\right|\right|^2 - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \left|\left|v_t^T \beta_t\right|\right|^2 \\
&= \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left(\alpha_t \alpha_t^T - \beta_t \beta_t^T\right)\right) \\
&= \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left((A_t - C_t H_t)(A_t^T - H_t^T C_t^T) - (A_t - C_t)(A_t^T - C_t^T)\right)\right) \\
&= \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left(C_t(I - H_t)A_t^T + A_t(I - H_t)C_t^T\right)\right),
\end{aligned}$$

where we applied $H_t^T H_t = H_t H_t^T = I$ and $H_t^T = H_t$. Since $\operatorname{Tr}(M_1^T) = \operatorname{Tr}(M_1)$ and $\operatorname{Tr}(M_1 M_2) = \operatorname{Tr}(M_2 M_1)$ hold for every square matrices M_1 and M_2, we see that

$$\operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \left|\left|v_t^T \alpha_t\right|\right|^2 - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \left|\left|v_t^T \beta_t\right|\right|^2 = 2 \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left(C_t(I - H_t)A_t^T\right)\right).$$

Recalling the definitions of H_t and u_t we obtain

$$\begin{aligned}
&\operatorname{Tr}\left(\alpha_t \alpha_t^T\right) - \left|\left|v_t^T \alpha_t\right|\right|^2 - \operatorname{Tr}\left(\beta_t \beta_t^T\right) + \left|\left|v_t^T \beta_t\right|\right|^2 \\
&= 4 \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left(C_t u_t u_t^T A_t^T\right)\right) \\
&= \frac{4 \left|\left|X_t\right|\right|^2}{\left|\left|C_t^{-1} X_t\right|\right|^2} \operatorname{Tr}\left(\left(I - v_t v_t^T\right)\left(C_t C_t^{-1} v_t v_t^T C_t^{-T} A_t^T\right)\right) \\
&= \frac{4 \left|\left|X_t\right|\right|^2}{\left|\left|C_t^{-1} X_t\right|\right|^2} \operatorname{Tr}\left(v_t v_t^T C_t^{-T} A_t^T - v_t v_t^T v_t v_t^T C_t^{-T} A_t^T\right) \\
&= 0,
\end{aligned}$$

since $v_t^T v_t = ||v_t||^2 = 1$.

From the equality in (3.10) we obtain

$$\begin{aligned}
&\left|\left|v_t^T\alpha_t\right|\right|^2 - \left|\left|v_t^T\beta_t\right|\right|^2 = \operatorname{Tr}\left(\alpha_t\alpha_t^T\right) - \operatorname{Tr}\left(\beta_t\beta_t^T\right) \\
&= \operatorname{Tr}\left((A_t - C_tH_t)(A_t^T - H_t^TC_t^T) - (A_t - C_t)(A_t^T - C_t^T)\right) \\
&= \operatorname{Tr}\left(A_tC_t^T + C_tA_t^T - A_tH_tC_t^T - C_tH_tA_t^T\right) \\
&= 2\operatorname{Tr}\left(A_t(I - H_t)C_t^T\right) \\
&= 4\operatorname{Tr}\left(A_tu_tu_t^TC_t^T\right) \\
&= \frac{4\left|\left|X_t\right|\right|^2}{\left|\left|C_t^{-1}X_t\right|\right|^2}\operatorname{Tr}\left(A_tC_t^{-1}v_tv_t^TC_t^{-T}C_t^T\right) \\
&= \frac{4\left|\left|X_t\right|\right|^2}{\left|\left|C_t^{-1}X_t\right|\right|^2}v_t^TA_tC_t^{-1}v_t,
\end{aligned}$$

where we used that $\operatorname{Tr}(w_1w_2^T) = w_2^Tw_1$ holds for any vectors w_1 and w_2 of the same dimension. Recall the definition of λ in (3.6) and that our matrix norm is the spectral norm, from where we obtain

$$\left|\left|C_t^{-1}X_t\right|\right| \leq \left|\left|C_t^{-1}\right|\right| \left|\left|X_t\right|\right|$$

and

$$\left|\left|C_t^{-1}\right|\right| = \sqrt{\lambda_{\max}(C_t^{-T}C_t^{-1})} = \frac{1}{\sqrt{\lambda_{\min}(C_t^TC_t)}} \leq \frac{1}{\sqrt{\lambda}},$$

and thus

$$\left|\left|v_t^T\alpha_t\right|\right|^2 \geq \frac{4\left|\left|X_t\right|\right|^2}{\left|\left|C_t^{-1}X_t\right|\right|^2}v_t^TA_tC_t^{-1}v_t \geq 4\lambda v_t^TA_tC_t^{-1}v_t.$$

We have

$$\left|\left|A_tC_t^{-1} - I\right|\right| \leq \left|\left|A_t - C_t\right|\right| \left|\left|C_t^{-1}\right|\right| \leq \left|\left|A_t - C_t\right|\right| \frac{1}{\sqrt{\lambda}},$$

and thus the Lipschitz the continuity of σ implies that there exists $\delta' > 0$ such that

$$\left|\left|A_tC_t^{-1} - I\right|\right| < \epsilon \quad \text{on} \quad \left\{t < \rho_{\delta'}\right\},$$

and hence

$$\left|\left|v_t^T\alpha_t\right|\right|^2 \geq 4\lambda(1-\epsilon) \geq \lambda \quad \text{on} \quad \left\{t < \rho_{\delta'}\right\}.$$

Define now a time-change (see [17, p. 174]) in the following way:

$$\tau_t := \inf\{s \geq 0;\ [N]_s > t\}, \quad t \geq 0,$$

where

$$N_t := \int_0^t v_s^T \alpha_s \, \mathrm{d}B_s, \quad t \geq 0.$$

Then[4]

$$W_t := N_{\tau_t}, \quad t \geq 0,$$

is a (one-dimensional) Brownian motion, and

$$[N]_t \geq \lambda t \quad \text{and} \quad \tau_t \leq \inf\{s \geq 0;\ \lambda s > t\} = \frac{t}{\lambda} \quad \text{on} \quad \{t < \rho_{\delta'}\}.$$

By the time-change integration formula we obtain from (3.9), with the hat denoting the time-changed processes,

$$\begin{aligned}\hat{S}_t &= ||x-y|| + \int_0^t 2\sqrt{\hat{S}_s}\, \mathrm{d}W_s + \int_0^t \frac{2\sqrt{\hat{S}_s}\hat{v}_s^T\hat{b}_s + \mathrm{Tr}\left(\hat{\alpha}_s\hat{\alpha}_s^T\right)}{||\hat{v}_s^T\hat{\alpha}_s||^2}\, \mathrm{d}s \\ &= ||x-y|| + \int_0^t 2\sqrt{\hat{S}_s}\, \mathrm{d}W_s + \int_0^t \left(1 + \frac{2\sqrt{\hat{S}_s}\hat{v}_s^T\hat{b}_s + \mathrm{Tr}\left(\hat{\beta}_s\hat{\beta}_s^T\right) - \left|\left|\hat{v}_s^T\hat{\beta}_s\right|\right|^2}{||\hat{v}_s^T\hat{\alpha}_s||^2}\right) \mathrm{d}s.\end{aligned}$$

Now we know that we can choose $\delta \in (0, \delta')$ such that

$$\nu_s := \frac{2\sqrt{\hat{S}_s}\hat{v}_s^T\hat{b}_s + \mathrm{Tr}\left(\hat{\beta}_s\hat{\beta}_s^T\right) - \left|\left|\hat{v}_s^T\hat{\beta}_s\right|\right|^2}{||\hat{v}_s^T\hat{\alpha}_s||^2} \leq \epsilon \quad \text{on} \quad \{s < \rho_\delta\}.$$

Hence we can rewrite the stochastic differential equation for $\hat{S}$ as

$$\hat{S}_t = ||x-y|| + \int_0^t 2\sqrt{\hat{S}_s}\, \mathrm{d}W_s + \int_0^t (1+\nu_s)\, \mathrm{d}s, \quad t \geq 0,$$

with $1 + \nu_s \leq 1 + \epsilon$ on $\{s < \rho_{\delta'}\}$. We know that the squared Bessel process S with dimension $1+\epsilon$ started at $||x-y||$ solves

$$S_t = ||x-y|| + \int_0^t 2\sqrt{S_s}\, \mathrm{d}W_s + (1+\epsilon)t, \quad t \geq 0.$$

Itô's lemma yields

$$\sqrt{S_t} - \sqrt{\hat{S}_t} = \int_0^t \left(\frac{\epsilon}{2\sqrt{S_s}} - \frac{\nu_s}{2\sqrt{\hat{S}_s}}\right) \mathrm{d}s, \quad t \geq 0.$$

[4] In order not to need to enlarge the probability space, $[N]_\infty = \infty$ must hold. We can achieve this by substituting $v^T\alpha$ in the definition of N by the process that is stopped when $||v^T\alpha||^2$ reaches, say, $\frac{\lambda}{2}$. Clearly this does not change anything up to the time $\rho_{\delta'}$.

Suppose there exists $t_{\text{neg}} > 0$ and $\omega \in \{t_{\text{neg}} < \rho_\delta\}$ such that $\sqrt{S_{t_{\text{neg}}}(\omega)} < \sqrt{\hat{S}_{t_{\text{neg}}}(\omega)}$. Let $t_0(\omega)$ be the largest zero of the function

$$t \mapsto \sqrt{S_t(\omega)} - \sqrt{\hat{S}_t(\omega)}, \quad t \geq 0,$$

that is not bigger than t_{neg} (it exists because the function is of bounded variation). Then the function is increasing on $(t_0(\omega), t_{\text{neg}})$ since the integrand is non-negative. At the same time the function goes from 0 to a negative value. This is a contradiction, which concludes the proof. □

Lemma 3.3.14. *For every Lipschitz Markov policy π, the function V_π is continuous.*

Proof. Let $\epsilon > 0$ and recall the notation from the previous lemma. For any $\hat{\delta}, \tilde{\delta} \leq \delta$, which will be determined later, we obtain: if $||x - y|| \leq \tilde{\delta}$, then

$$\begin{aligned}
&|V_\pi(x) - V_\pi(y)| \\
&\leq \mathbb{E}\left(\int_0^\infty \left|\mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{x,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{x,\pi}\right) - \mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{y,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right| \mathrm{d}t\right) \\
&= \mathbb{E}\left(\int_0^{\rho_0} \left|\mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{x,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{x,\pi}\right) - \mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{y,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\rho_{\hat{\delta}} \geq \rho_0\}}\right) \\
&\quad + \mathbb{E}\left(\int_{\rho_0}^\infty \left|\mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{x,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{x,\pi}\right) - \mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{y,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\rho_{\hat{\delta}} \geq \rho_0\}}\right) \\
&\quad + \mathbb{E}\left(\int_0^\infty \left|\mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{x,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{x,\pi}\right) - \mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{y,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\rho_{\hat{\delta}} < \rho_0\}}\right).
\end{aligned}$$

In the first two terms, the two processes couple before they are more than $\hat{\delta}$ apart, and with the help of continuity this will make the terms small. In the last term, we will apply the previous lemma to show that the probability of the event is small.

Let M be a constant that bounds the functions α_π, f_π and V_π, and is also a Lipschitz constant for α_π and f_π. Let α_m be small enough so that $\alpha_\pi - \alpha_m$ is still bounded away from 0. To estimate the first term, we will use the Triangle Inequality, the elementary inequality

$$\left|\mathrm{e}^{-y} - \mathrm{e}^{-z}\right| \leq |y - z|, \quad y, z \geq 0,$$

and the Lipschitz property of f_π and α_π. We obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{\rho_0}\left|\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_t^{x,\pi}\right)-\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\leq\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s}\left|f_\pi\left(\tilde{X}_t^{x,\pi}\right)-f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\quad+\mathbb{E}\left(\int_0^{\rho_0}\left|f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right|\left|\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s}-\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)\mathrm{d}s}\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\leq\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\alpha_m t}\left|f_\pi\left(\tilde{X}_t^{x,\pi}\right)-f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\quad+M\,\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\alpha_m t}\left|\int_0^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s-\int_0^t\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)\mathrm{d}s\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\leq M\,\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\alpha_m t}\left|\left|\tilde{X}_t^{x,\pi}-\tilde{X}_t^{y,\pi}\right|\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\quad+M^2\,\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\alpha_m t}\int_0^t\left|\left|\tilde{X}_s^{x,\pi}-\tilde{X}_s^{y,\pi}\right|\right|\mathrm{d}s\,\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right).
\end{aligned}
$$

Let $\delta_1\in(0,\delta)$ be small enough that the following holds:

$$
\delta_1 M\,\mathbb{E}\left(\int_0^{\rho_0}\mathrm{e}^{-\alpha_m t}\,\mathrm{d}t\right)+\delta_1 M^2\,\mathbb{E}\left(\int_0^{\rho_0}t\mathrm{e}^{-\alpha_m t}\,\mathrm{d}t\right)<\frac{\epsilon}{3}.
$$

To bound the second term, we recall that the processes $\tilde{X}^{x,\pi}$ and $\tilde{X}^{y,\pi}$ coincide after time ρ_0. Thus we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_{\rho_0}^{\infty}\left|\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_t^{x,\pi}\right)-\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&=\mathbb{E}\left(\int_{\rho_0}^{\infty}\mathrm{e}^{-\alpha_m\rho_0}\left|\mathrm{e}^{-\int_0^{\rho_0}\left(\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)-\alpha_m\right)\mathrm{d}s}-\mathrm{e}^{-\int_0^{\rho_0}\left(\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)-\alpha_m\right)\mathrm{d}s}\right|\right.\\
&\qquad\left.\cdot\left|\mathrm{e}^{-\int_{\rho_0}^t\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_t^{x,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\leq\mathbb{E}\left(\mathrm{e}^{-\alpha_m\rho_0}\left|\int_0^{\rho_0}\alpha_\pi\left(\tilde{X}_s^{x,\pi}\right)\mathrm{d}s-\int_0^{\rho_0}\alpha_\pi\left(\tilde{X}_s^{y,\pi}\right)\mathrm{d}s\right|\right.\\
&\qquad\left.\cdot\int_0^{\infty}\left|\mathrm{e}^{-\int_0^t\alpha_\pi\left(\tilde{X}_{s+\rho_0}^{x,\pi}\right)\mathrm{d}s}f_\pi\left(\tilde{X}_{t+\rho_0}^{x,\pi}\right)\right|\mathrm{d}t\,\mathbb{I}_{\{\rho_{\hat{\delta}}\geq\rho_0\}}\right)\\
&\leq M^2\,\mathbb{E}\left(\mathrm{e}^{-\alpha_m\rho_0}\int_0^{\rho_0}\left|\left|\tilde{X}_s^{x,\pi}-\tilde{X}_s^{y,\pi}\right|\right|\mathrm{d}s\right).
\end{aligned}
$$

Set $\delta_2\in(0,\delta)$ so small that

$$
\delta_2 M^2\,\mathbb{E}\left(\mathrm{e}^{-\alpha_m\rho_0}\rho_0\right)<\frac{\epsilon}{3}.
$$

The final term can be bounded in the following way:

$$\mathbb{E}\left(\int_0^\infty \left|\mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{x,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{x,\pi}\right) - \mathrm{e}^{-\int_0^t \alpha_\pi(\tilde{X}_s^{y,\pi})\mathrm{d}s} f_\pi\left(\tilde{X}_t^{y,\pi}\right)\right| \mathrm{d}t\, \mathbb{I}_{\{\rho_{\hat{\delta}}<\rho_0\}}\right)$$
$$\leq 2M\mathbb{P}\left(\rho_{\hat{\delta}} < \rho_0\right)$$

Recall that $\rho_{\hat{\delta}}$ and ρ_0 are the first hitting times of $\hat{\delta}$ and 0, respectively, by the process $\left|\left|\tilde{X}^{x,\pi} - \tilde{X}^{y,\pi}\right|\right|$. If we denote by $\rho_{\hat{\delta}}(Y)$ and $\rho_0(Y)$ the equivalent times for any process Y, then we obtain by the previous lemma

$$\mathbb{P}\left(\rho_{\hat{\delta}} < \rho_0\right) \leq \mathbb{P}\left(\rho_{\hat{\delta}}\left(S_{\tau_\cdot}\right) < \rho_0\left(S_{\tau_\cdot}\right)\right) \leq \mathbb{P}\left(\rho_{\hat{\delta}}\left(S_{\frac{1}{\lambda}\cdot}\right) < \rho_0\left(S_{\frac{1}{\lambda}\cdot}\right)\right).$$

Using the scale property of the squared Bessel process we get

$$\mathbb{P}\left(\rho_{\hat{\delta}}\left(S_{\frac{1}{\lambda}\cdot}\right) < \rho_0\left(S_{\frac{1}{\lambda}\cdot}\right)\right) = \mathbb{P}\left(\rho_{\hat{\delta}}\left(\frac{1}{\lambda}S\right) < \rho_0\left(\frac{1}{\lambda}S\right)\right) = \mathbb{P}\left(\rho_{\lambda\hat{\delta}}\left(S\right) < \rho_0\left(S\right)\right).$$

Recall that the scale function of the Bessel process with dimension $1+\epsilon$ is given by $s(z) := z^{\frac{1-\epsilon}{2}}$, and that the process S starts at $||x-y|| < \tilde{\delta}$. Hence we obtain

$$\mathbb{P}\left(\rho_{\lambda\hat{\delta}}\left(S\right) < \rho_0\left(S\right)\right) = \frac{s(||x-y||) - s(0)}{s(\lambda\hat{\delta})} \leq \left(\frac{\tilde{\delta}}{\lambda\hat{\delta}}\right)^{\frac{1-\epsilon}{2}}$$

Now we can finally finish the proof. Set $\hat{\delta} := \delta_1 \wedge \delta_2$, and let $\tilde{\delta} \in (0, \hat{\delta})$ be so small that

$$2M\left(\frac{\tilde{\delta}}{\lambda\hat{\delta}}\right)^{\frac{1-\epsilon}{2}} < \frac{\epsilon}{3}.$$

Then we have proved that $||x-y|| \leq \tilde{\delta}$ implies $|V_\pi(x) - V_\pi(y)| < \epsilon$, so we have even derived the uniform continuity of V_π. □

3.3.3 Proofs

Proof of Proposition 3.3.5. It suffices to prove that the differential equation holds in every open ball in $\mathbb{R}^d$. Let D be an open ball with centre $\tilde{x}$ and radius r. Let $x \in D$ and define τ as the first time the process $X^{x,\pi}$ hits the boundary of D. For every $n \in \mathbb{N}$, define D^n as the closed ball with centre $\tilde{x}$ and radius $r - \frac{1}{n}$, and let τ_n be the first time the process $X^{x,\pi}$ hits the boundary of D^n.

Let $v \in \mathcal{C}^2(D) \cap \mathcal{C}(\bar{D})$ be the unique solution of the boundary value problem

$$L_\pi v - \alpha_\pi v + f_\pi = 0 \quad \text{in} \quad D, \qquad v = V_\pi \quad \text{on} \quad \partial D,$$

which is guaranteed to exist by Lemma 3.3.14 and Theorem 19 in [9, p. 87] (the main assumptions in the theorem are that $\sigma_\pi > 0$, $\alpha_\pi \geq 0$, the boundary condition is continuous, and that all the coefficients are Hölder continuous, which is satisfied because we imposed Lipschitz conditions on σ, μ, α, f and π). Let n_0 be large enough such that $x \in D^n$, and for every $n \geq n_0$, define the process $S^n = (S^n)_{t\geq 0}$ by

$$S_t^n := \int_0^{t\wedge\tau_n} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f^\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_n} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} v\left(X_{t\wedge\tau_n}^{\pi,x}\right),$$

and

$$S_t := \int_0^{t\wedge\tau} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f^\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} v\left(X_{t\wedge\tau}^{\pi,x}\right).$$

Itô's formula on $[0, \tau_n]$ and the differential equation for v yield

$$\begin{aligned}
S_t^n &= v(x) + \int_0^{t\wedge\tau_n} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} \left(f_\pi + L_\pi v - \alpha_\pi v\right)\left(X_s^{\pi,x}\right)\mathrm{d}s \\
&\quad + \int_0^{t\wedge\tau_n} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} \left(\nabla v\right)^T \sigma_\pi\left(X_s^{\pi,x}\right)\mathrm{d}B_s \\
&= v(x) + \int_0^{t\wedge\tau_n} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} \left(\nabla v\right)^T \sigma_\pi\left(X_s^{\pi,x}\right)\mathrm{d}B_s, \quad t \geq 0, \quad n \geq n_0.
\end{aligned}$$

Hence S^n is a local martingale, and since it is a bounded process (recall that f is bounded, and α positive and bounded away from 0), it is a uniformly integrable martingale. Thus the Dominated Convergence Theorem yields

$$v(x) = \lim_{n\to\infty} \mathbb{E}\left(S_0^n\right) = \lim_{n\to\infty} \mathbb{E}\left(S_\infty^n\right) = \mathbb{E}\left(S_\infty\right).$$

Due to the boundary conditions for v and Lemma 3.3.11 we obtain

$$\begin{aligned}
S_\infty &= \int_0^{\tau} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} v\left(X_\tau^{\pi,x}\right) \\
&= \int_0^{\tau} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{\tau} \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} V_\pi\left(X_\tau^{\pi,x}\right) \\
&= \mathbb{E}\left(\int_0^{\infty} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s \,\middle|\, \mathcal{F}_\tau\right),
\end{aligned}$$

and therefore

$$v(x) = \mathbb{E}\left(\int_0^{\infty} \mathrm{e}^{-\int_0^s \alpha_\pi\left(X_r^{\pi,x}\right)\mathrm{d}r} f_\pi\left(X_s^{\pi,x}\right)\mathrm{d}s\right) = V_\pi(x).$$

□

Proof of Theorem 3.2.7. For every $m \in \mathbb{N}$, let D_m be the closed ball with centre x and radius m, and τ_m the first time the process $X^{\pi_{n+1},x}$ hits the boundary of D_m. Define the process S by

$$S_t := \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} f_{\pi_{n+1}}\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^t \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} V_{\pi_n}\left(X_t^{\pi_{n+1},x}\right), \quad t \geq 0.$$

Itô's formula, which is applicable thanks to Proposition 3.3.5, yields

$$\begin{aligned} S_{t\wedge\tau_m} = V_{\pi_n}(x) &+ \int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \left(\nabla V_{\pi_n}\right)^T \sigma_{\pi_{n+1}}\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}B_s \\ &+ \int_0^{t\wedge\tau_m} e^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \left(f_{\pi_{n+1}} + L_{\pi_{n+1}}V_{\pi_n} - \alpha_{\pi_{n+1}}V_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s. \end{aligned}$$

The stochastic integral is a martingale since the mappings $\sigma_{\pi_{n+1}}$ and ∇V_{π_n} are bounded on D_m by Assumption 3.3.1 and Proposition 3.3.5, respectively. Hence we obtain

$$\begin{aligned} E\left(S_{t\wedge\tau_m}\right) = V_{\pi_n}(x) + E\Bigg(&\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \\ &\cdot \left(f_{\pi_{n+1}} + L_{\pi_{n+1}}V_{\pi_n} - \alpha_{\pi_{n+1}}V_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\Bigg). \end{aligned}$$

By the definition of the policy improvement algorithm (3.8) we get

$$\begin{aligned} &E\left(S_{t\wedge\tau_m}\right) \\ &= V_{\pi_n}(x) + E\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \min_{p\in A}\left(f_p + L_pV_{\pi_n} - \alpha_pV_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\right) \\ &\leq V_{\pi_n}(x) + E\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} \left(f_{\pi_n} + L_{\pi_n}V_{\pi_n} - \alpha_{\pi_n}V_{\pi_n}\right)\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s\right) \\ &= V_{\pi_n}(x), \end{aligned}$$

where we used Proposition 3.3.5 in the last step. By recalling the definition of S_t and applying the Dominated Convergence Theorem, we obtain the following (for every $m \in \mathbb{N}$) by sending t to ∞:

$$\begin{aligned} V_{\pi_n}(x) \geq \mathbb{E}\Bigg(&\int_0^{\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} f_{\pi_{n+1}}\left(X_s^{\pi_{n+1},x}\right)\mathrm{d}s \\ &+\mathrm{e}^{-\int_0^{\tau_m} \alpha_{\pi_{n+1}}\left(X_r^{\pi_{n+1},x}\right)\mathrm{d}r} V_{\pi_n}\left(X_{\tau_m}^{\pi_{n+1},x}\right)\Bigg). \end{aligned}$$

Now we send m to ∞, and applying the Dominated Convergence Theorem, Remark 3.3.12 and the definition of $V^{\pi_{n+1}}$ we obtain

$$V_{\pi_n}(x) \geq \mathbb{E}\left(\int_0^\infty \mathrm{e}^{-\int_0^s \alpha_{\pi_{n+1}}(X_r^{\pi_{n+1},x})\mathrm{d}r} f_{\pi_{n+1}}(X_s^{\pi_{n+1},x})\,\mathrm{d}s\right) = V_{\pi_{n+1}}(x),$$

which is what we had to prove. □

Proof of Proposition 3.3.8. For every $k \in \mathbb{N}$, let D_k be the closed ball with centre 0 and radius k. Applying Interior Estimates from [9, p. 86], Assumption 3.3.1 and Proposition 3.3.5 to the sequences $\{\sigma_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\mu_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\alpha_{\pi_n}\}_{n\in\mathbb{N}}$, $\{f_{\pi_n}\}_{n\in\mathbb{N}}$ and $\{V_{\pi_n}\}_{n\in\mathbb{N}}$, we obtain that the sequence $\{\mathrm{H}V_{\pi_n}\}_{n\in\mathbb{N}}$ is uniformly bounded on D_k for every $k \in \mathbb{N}$. Recalling the policy improvement algorithm 3.8 and applying Assumption 3.3.3, we obtain that the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ is uniformly Lipschitz and hence equicontinuous on D_k for every $k \in \mathbb{N}$. Define $\pi_n^0 := \pi_n$ for every $n \in \mathbb{N}$. Thanks to the version of the Arzela-Ascoli Theorem in Lemma 3.2.13, for every $k \in \mathbb{N}$ there exists a subsequence $\{\pi_n^k\}_{n\in\mathbb{N}} \subseteq \{\pi_n^{k-1}\}_{n\in\mathbb{N}}$ such that $\{\pi_n^k\}_{n\in\mathbb{N}}$ converges uniformly on D_k. The diagonal sequence, i.e. $\{\pi_n^n\}_{n\in\mathbb{N}}$, then converges uniformly on D_k for every $k \in \mathbb{N}$, and hence on every compact subset of $\mathbb{R}^d$. □

Proof of Theorem 3.3.9. Let $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ be a sequence from Proposition 3.3.8 that converges to $\pi_{\lim}$ uniformly on compacts in $\mathbb{R}^d$. For every $m \in \mathbb{N}$, let D_m be the closed ball with centre x and radius m, and τ_m the first time the process $X^{\pi_{\lim},x}$ hits the boundary of D_m. Let $k \in \mathbb{N}$ and define the process $S = (S_t)_{t\geq 0}$ by

$$S_t := \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}(X_r^{\pi_{\lim},x})\mathrm{d}r} f^{\pi_{n_k}}(X_s^{\pi_{\lim},x})\,\mathrm{d}s + \mathrm{e}^{-\int_0^t \alpha_{\pi_{n_k}}(X_r^{\pi_{\lim},x})\mathrm{d}r} V_{\pi_{n_k}}(X_t^{\pi_{\lim},x}).$$

Itô's formula yields

$$\begin{aligned} S_{t\wedge\tau_m} = V_{\pi_{n_k}}(x) &+ \int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}(X_r^{\pi_{\lim},x})\mathrm{d}r} \left(\nabla V_{\pi_{n_k}}\right)^T \sigma_{\pi_{\lim}}(X_s^{\pi_{\lim},x})\,\mathrm{d}B_s \\ &+ \int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}(X_r^{\pi_{\lim},x})\mathrm{d}r} \left(f_{\pi_{n_k}} + L_{\pi_{\lim}} V_{\pi_{n_k}} - \alpha_{\pi_{n_k}} V_{\pi_{n_k}}\right)(X_s^{\pi_{\lim},x})\,\mathrm{d}s. \end{aligned}$$

The stochastic integral is a martingale since the mappings $\sigma_{\pi_{\lim}}$ and $\nabla V_{\pi_{n_k}}$ are

bounded on D_m (by Assumption 3.3.1 and Proposition 3.3.5). Hence we obtain

$$\begin{aligned}
&E\left(S_{t\wedge\tau_m}\right) = V_{\pi_{n_k}}(x) \\
&\quad + E\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(f_{\pi_{n_k}} + L_{\pi_{\lim}}V_{\pi_{n_k}} - \alpha_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right) \\
&= V_{\pi_{n_k}}(x) + E\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(L_{\pi_{\lim}}V_{\pi_{n_k}} - L_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right) \\
&= V_{\pi_{n_k}}(x) + E\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{n_k}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r}\left(\left(\mu_{\pi_{\lim}} - \mu_{\pi_{n_k}}\right)^T \nabla V_{\pi_{n_k}}\right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{2}\operatorname{Tr}\left(\left(\sigma_{\pi_{\lim}} - \sigma_{\pi_{n_k}}\right)^T \mathrm{H}V_{\pi_{n_k}}\left(\sigma_{\pi_{\lim}} - \sigma_{\pi_{n_k}}\right)\right)\right)\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right),
\end{aligned}$$

where we used Proposition 3.3.5 and the definition of the operator L. Applying Interior Estimates from [9, p. 86], Assumption 3.3.1 and Proposition 3.3.5 to the sequences $\{\sigma_{\pi_{n_m}}\}_{m\in\mathbb{N}}$, $\{\mu_{\pi_{n_m}}\}_{m\in\mathbb{N}}$, $\{\alpha_{\pi_{n_m}}\}_{m\in\mathbb{N}}$, $\{f_{\pi_{n_m}}\}_{m\in\mathbb{N}}$ and $\{V_{\pi_{n_m}}\}_{m\in\mathbb{N}}$, we obtain that the sequences $\{\nabla V_{\pi_{n_m}}\}_{m\in\mathbb{N}}$ and $\{\mathrm{H}V_{\pi_{n_m}}\}_{m\in\mathbb{N}}$ are uniformly bounded on D_m. Now the Dominated Convergence Theorem yields that the last term disappears when k tends to ∞, hence we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\pi_{\lim}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r} f_{\pi_{\lim}}\left(X_s^{\pi_{\lim},x}\right)\mathrm{d}s\right. \\
&\qquad \left. + \mathrm{e}^{-\int_0^{t\wedge\tau_m} \alpha_{\pi_{\lim}}\left(X_r^{\pi_{\lim},x}\right)\mathrm{d}r} V_{\lim}\left(X_{t\wedge\tau_m}^{\pi_{\lim},x}\right)\right) \\
&= V_{\lim}(x).
\end{aligned}$$

By first sending t to ∞ and then m to ∞, we obtain the desired equality as in the previous proof. □

Proof of Theorem 3.3.10. The second assertion follows from Theorem 3.3.9.

Applying Interior Estimates from [9, p. 86], Assumption 3.3.1 and Proposition 3.3.5 to the sequences $\{\sigma_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\mu_{\pi_n}\}_{n\in\mathbb{N}}$, $\{\alpha_{\pi_n}\}_{n\in\mathbb{N}}$, $\{f_{\pi_n}\}_{n\in\mathbb{N}}$ and $\{V_{\pi_n}\}_{n\in\mathbb{N}}$, we obtain that the sequence $\{\mathrm{H}V_{\pi_n}\}_{n\in\mathbb{N}}$ is uniformly bounded on compacts in $\mathbb{R}^d$. According to Assumption 3.3.3, the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ is uniformly Lipschitz on compacts in $\mathbb{R}^d$, and the same holds for the sequence $\{(\pi_{n+1},\pi_n) : \mathbb{R}^d \to A\times A\}_{n\in\mathbb{N}}$ where $A\times A$ is equipped with any p-product metric, $p\in[1,\infty]$. The same diagonalisation argument as in the proof of Proposition 3.3.8 yields a subsequence $\{(\pi_{1+n_k},\pi_{n_k})\}_{k\in\mathbb{N}}$ that is uniformly convergent on every compact set in $\mathbb{R}^d$.

For every $k\in\mathbb{N}$, let

$$\hat{\sigma}_k(\cdot) := \sigma(\cdot,\pi_{n_k}(\cdot)), \quad \hat{\pi}_\infty(\cdot) := \lim_{m\to\infty}\pi_{n_m}(\cdot) \quad \text{and} \quad \hat{\sigma}_\infty(\cdot) := \sigma(\cdot,\hat{\pi}_\infty(\cdot)).$$

Define $\hat{\mu}_k$, $\hat{\alpha}_k$, $\hat{f}_k$ and $\hat{\mu}_\infty$, $\hat{\alpha}_\infty$, $\hat{f}_\infty$ in a corresponding fashion. Similarly, let

$$\tilde{\sigma}_k(\cdot) := \sigma(\cdot, \pi_{n_k+1}(\cdot)), \quad \tilde{\pi}_\infty(\cdot) := \lim_{m\to\infty} \pi_{n_m+1}(\cdot) \quad \text{and} \quad \tilde{\sigma}_\infty(\cdot) := \sigma(\cdot, \tilde{\pi}_\infty(\cdot)),$$

and define $\tilde{\mu}_k$, $\tilde{\alpha}_k$, $\tilde{f}_k$ and $\tilde{\mu}_\infty$, $\tilde{\alpha}_\infty$, $\tilde{f}_\infty$ in a corresponding fashion. Set

$$\hat{v}_k(\cdot) := V_{\pi_{n_k}}(\cdot), \quad \tilde{v}_k(\cdot) := V_{\pi_{n_k+1}}(\cdot), \quad \text{and} \quad v(\cdot) := V_{\text{lim}}(\cdot),$$

and define the operator $\hat{\mathcal{L}}_k$ by

$$\hat{\mathcal{L}}_k u := \frac{1}{2} \operatorname{Tr}\left(\hat{\sigma}_k^T \mathrm{H} u \hat{\sigma}_k\right) + \hat{\mu}_k^T \nabla u - \hat{\alpha}_k u + \hat{f}_k, \quad u \in \mathcal{C}^2(\mathbb{R}^d),$$

with the corresponding definitions for $\hat{\mathcal{L}}_\infty$, $\tilde{\mathcal{L}}_k$ and $\tilde{\mathcal{L}}_\infty$.

For every $m \in \mathbb{N}$, let D_m be the closed ball with centre x and radius m, and τ_m the first time the process $X^{\Pi,x}$ hits the boundary of D_m. Applying the last part of Theorem 15[5] from [9, p. 80], Proposition 3.3.5 and Assumption 3.3.1, we obtain that both $\hat{\mathcal{L}}_\infty v = 0$ and $\tilde{\mathcal{L}}_\infty v = 0$ hold on D_m, and that the sequence $\{\frac{1}{2}(\tilde{\sigma}_k^2 - \hat{\sigma}_k^2)\hat{v}_k'' + (\tilde{\mu}_k - \hat{\mu}_k)\hat{v}_k' - (\tilde{\alpha}_k - \hat{\alpha}_k)\hat{v}_k + \tilde{f}_k - \hat{f}_k\}_{k\in\mathbb{N}}$ converges uniformly to $\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)v'' + (\tilde{\mu}_\infty - \hat{\mu}_\infty)v' - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)v + \tilde{f}_\infty - \hat{f}_\infty$ on D_m for every $m \in \mathbb{N}$. However, we know that

$$\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)v'' + (\tilde{\mu}_\infty - \hat{\mu}_\infty)v' - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)v + \tilde{f}_\infty - \hat{f}_\infty = \tilde{\mathcal{L}}_\infty v - \hat{\mathcal{L}}_\infty v = 0. \quad (3.11)$$

[5] As already mentioned, the theorem is stated for parabolic equations, but is also valid for elliptic ones; in the same way as the parabolic version follows from Theorem 5, p. 64, the elliptic version follows from Interior Estimates, p. 86.

Now let $k \in \mathbb{N}$ and $m \in \mathbb{N}$. We obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} f_{\Pi_s}\left(X_s^{\Pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_m} \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} V_{\pi_{n_k}}\left(X_{t\wedge\tau_m}^{\Pi,x}\right)\right) \\
&\overset{\text{Itô}}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\left(f_{\Pi_s} + L_{\Pi_s}V_{\pi_{n_k}} - \alpha_{\Pi_s}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right) \\
&\geq V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} \min_{p\in A}\left(f_p + L_pV_{\pi_{n_k}} - \alpha_pV_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right) \\
&\overset{\text{PIA(3.8)}}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\right. \\
&\quad \left.\cdot\left(f_{\pi_{n_k+1}} + L_{\pi_{n_k+1}}V_{\pi_{n_k}} - \alpha_{\pi_{n_k+1}}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right) \\
&\overset{\text{Prop. 3.3.5}}{=} V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\right. \\
&\quad \left.\cdot\left(f_{\pi_{n_k+1}} - f_{\pi_{n_k}} + L_{\pi_{n_k+1}}V_{\pi_{n_k}} - L_{\pi_{n_k}}V_{\pi_{n_k}} - \alpha_{\pi_{n_k+1}}V_{\pi_{n_k}} + \alpha_{\pi_{n_k}}V_{\pi_{n_k}}\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right) \\
&= V_{\pi_{n_k}}(x) + \mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r}\right. \\
&\quad \left.\cdot\left(\frac{1}{2}(\tilde{\sigma}_k^2 - \hat{\sigma}_k^2)\hat{v}_k'' + (\tilde{\mu}_k - \hat{\mu}_k)\hat{v}_k' - (\tilde{\alpha}_k - \hat{\alpha}_k)\hat{v}_k + \tilde{f}_k - \hat{f}_k\right)\left(X_s^{\Pi,x}\right)\mathrm{d}s\right),
\end{aligned}
$$

Sending k to ∞, applying the Dominated Convergence Theorem and using (3.11), we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{t\wedge\tau_m} \mathrm{e}^{-\int_0^s \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} f_{\Pi_s}\left(X_s^{\Pi,x}\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{t\wedge\tau_m} \alpha_{\Pi_r}\left(X_r^{\Pi,x}\right)\mathrm{d}r} V_{\lim}\left(X_{t\wedge\tau_m}^{\Pi,x}\right)\right) \\
&\geq V_{\lim}(x).
\end{aligned}
$$

By first sending t to ∞ and then m to ∞, we obtain the desired inequality in the usual way. □

3.4 Examples

3.4.1 Data satisfying the assumptions

Proposition 3.4.1. *Let A be a compact interval. For every $x \in (a, b)$ and $p \in A$ define*

$$\sigma(x,p) := \sigma_1(x), \quad \mu(x,p) := \mu_1(x) + \mu_2 p,$$

$$\alpha(x,p) := \alpha_1(x) + \alpha_2 p, \quad f(x,p) := f_1(x) + f_2(p),$$

where σ_1 is bounded, Lipschitz on compacts in (a,b), and σ_1^2 is bounded away from 0; μ_1 is bounded and Lipschitz on compacts in (a,b), and $\mu_2 \in \mathbb{R}$; α_1 is Lipschitz on compacts in (a,b), bounded below away from $-\alpha_2 \max(A) \vee -\alpha_2 \min(A)$ and bounded above, and $\alpha_2 \in \mathbb{R}$; f_1 is bounded and Lipschitz on compacts in (a,b), and f_2 is continuous on A, differentiable on $\mathring{A}$ (i.e. the interior of A), and f_2' is strictly increasing and its inverse function is Lipschitz on compacts in $f_2'(\mathring{A})$. An example of such a function is $f_2(p) := \sum_{k=1}^n c_k |p|^{m_k}$ where $n \in \mathbb{N}$, $c_k > 0$ and $m_k \in (1,2]$ for every $k \leq n$. Then these data satisfy Assumptions 3.2.1 and 3.2.2.

Proof. The data obviously satisfy Assumptions 3.2.1, so we only need to check Assumptions 3.2.2. We are looking for $I_h(x)$, i.e. the minimiser of the function

$$p \mapsto \frac{\mu_1(x) + \mu_2 p}{\sigma_1(x)^2} h'(x) - \frac{\alpha_1(x) + \alpha_2 p}{\sigma_1(x)^2} h(x) + \frac{f_1(x) + f_2(p)}{\sigma_1(x)^2}, \quad p \in A.$$

The derivative of this function is

$$p \mapsto \frac{\mu_2 h'(x)}{\sigma_1(x)^2} - \frac{\alpha_2 h(x)}{\sigma_1(x)^2} + \frac{f_2'(p)}{\sigma_1(x)^2}, \quad p \in \mathring{A}.$$

Since f_2' is increasing, the function is convex, and hence

$$I_h(x) = \begin{cases} \min(A) & \text{if} \quad \alpha_2 h(x) - \mu_2 h'(x) \leq \inf f_2'(\mathring{A}), \\ (f_2')^{-1}(\alpha_2 h(x) - \mu_2 h'(x)) & \text{if} \quad \alpha_2 h(x) - \mu_2 h'(x) \in f_2'(\mathring{A}), \\ \max(A) & \text{if} \quad \alpha_2 h(x) - \mu_2 h'(x) \geq \sup f_2'(\mathring{A}). \end{cases}$$

If the sequence of functions $\{h_n'\}_{n\in\mathbb{N}}$ is uniformly Lipschitz on compacts in (a,b), then the same holds for $\{I_{h_n}\}_{n\in\mathbb{N}}$ because of the Lipschitz assumption on $(f_2')^{-1}$. □

Essentially the same functions work in the multidimensional case.

Proposition 3.4.2. *Let A be a compact interval. For every $x \in \mathbb{R}^d$ and $p \in A$ define*

$$\sigma(x,p) := \sigma_1(x), \quad \mu(x,p) := \mu_1(x) + p\mu_2,$$

$$\alpha(x,p) := \alpha_1(x) + \alpha_2 p, \quad f(x,p) := f_1(x) + f_2(p),$$

where σ_1 is bounded, Lipschitz on compacts in $\mathbb{R}^d$, and $\sigma_1^T \sigma_1$ is bounded away from 0 in the sense of uniform ellipticity (refer to condition (3.6)); μ_1 is bounded and Lipschitz on compacts in $\mathbb{R}^d$, and $\mu_2 \in \mathbb{R}^d$; α_1 is Lipschitz on compacts in $\mathbb{R}^d$, bounded below away from $-\alpha_2 \max(A) \vee -\alpha_2 \min(A)$ and bounded above, and $\alpha_2 \in \mathbb{R}$; f_1 is bounded and Lipschitz on compacts in $\mathbb{R}^d$, and f_2 is continuous on A,

differentiable on $\mathring{A}$ (i.e. the interior of A), and f_2' is strictly increasing and its inverse function is Lipschitz on compacts in $f_2'(\mathring{A})$. An example of such a function is $f_2(p) := \sum_{k=1}^n c_k|p|^{m_k}$ where $n \in \mathbb{N}$, $c_k > 0$ and $m_k \in (1,2]$ for every $k \leq n$. Then these data satisfy Assumptions 3.3.1 and 3.3.3.

Proof. The function that we want to minimise this time is

$$p \mapsto \frac{1}{2}\operatorname{Tr}\left(\sigma(x)^T \mathrm{H}h(x)\sigma(x)\right) + (\mu_1(x) + p\mu_2)^T \nabla h(x) \\ - (\alpha_1(x) + \alpha_2 p)h(x) + f_1(x) + f_2(p), \quad p \in A.$$

The derivative of this function is

$$p \mapsto \mu_2^T \nabla h(x) - \alpha_2 h(x) + f_2'(p), \quad p \in \mathring{A},$$

thus the proof is completed in the same way the previous one was. □

3.4.2 Numerical examples

We want to solve the following problem: for every $x \in (-10, 10)$ find

$$\inf_{\Pi \in \mathcal{A}} \mathbb{E}\left(\int_0^{\tau\left(X^{\Pi,x}\right)} \mathrm{e}^{-t}\left(\left(X_t^{\Pi,x}\right)^2 + \Pi_t^2\right)\mathrm{d}t + \mathrm{e}^{-\tau\left(X^{\Pi,x}\right)}\left(X_{\tau\left(X^{\Pi,x}\right)}^{\Pi,x}\right)^2 \mathbb{I}_{\{\tau\left(X^{\Pi,x}\right)<\infty\}}\right)$$

together with an optimal policy (if it exists), where

$$\tau\left(X^{\Pi,x}\right) := \inf\left\{t \geq 0;\ X_t^{\Pi,x} \leq -10 \text{ or } X_t^{\Pi,x} \geq 10\right\},$$

$$\mathcal{A} := \{\Pi = (\Pi_t)_{t\geq 0};\ \Pi \text{ is adapted to } (\mathcal{F}_t)_{t\geq 0} \text{ and takes values in } [-1,1]\},$$

and

$$X_t^{\Pi,x} := x + B_t + \int_0^t \Pi_s\,\mathrm{d}s, \quad t \geq 0.$$

Assumptions 3.2.1 and 3.2.2 are satisfied since the data $A := [-1,1]$, $a = -10$, $b := 10$, $g(a) := a^2$, $g(b) := b^2$, $\sigma(x,p) := 1$, $\mu(x,p) := p$, $\alpha(x,p) := 1$, and $f(x,p) := x^2 + p^2$ fit into the model presented in Proposition 3.4.1.

We implemented the algorithm in Matlab. The payoff function on each step is obtained as the solution to the differential equation from Proposition 3.2.5 with the boundary conditions given by the function g. The new policy on each step can be calculated explicitly (cf. the proof of Proposition 3.4.1).

Figures 3.1 and 3.2 show the graphs of the payoff functions and of the policies (a different colour is used for each curve) when the initial policy is constant with

the value 1.

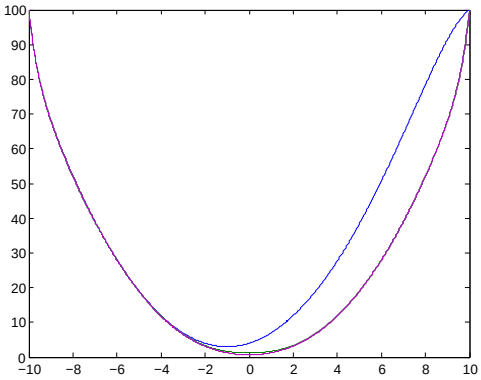


Figure 3.1: The graphs of V_{π_n} for $n \in \{0, 1, 2, 3, 4\}$.

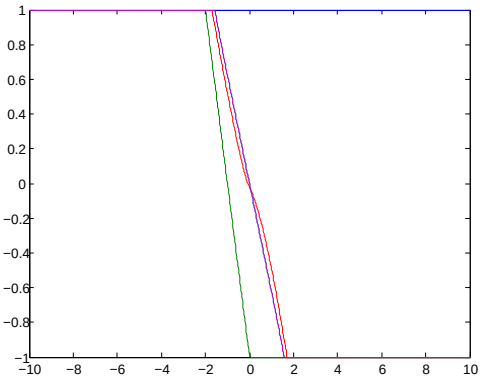


Figure 3.2: The graphs of π_n for $n \in \{0, 1, 2, 3, 4\}$.

The graphs suggest that convergence effectively occurs in just a few steps. Figures 3.3 and 3.4, which contain the graphs of the differences on the logarithmic scale, confirm this. Where it seems that fewer graphs are presented in a figure than stated below the figure, the last few graphs coincide. The policies only differ on a subinterval because outside of it they are all equal to the same constant (cf. the proof of Proposition 3.4.1).

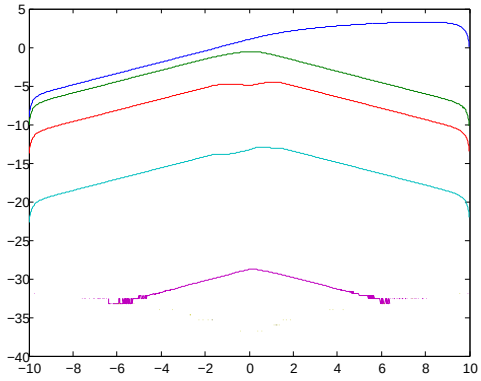


Figure 3.3: The graphs of $\log(|V_{\pi_{n+1}} - V_{\pi_n}|)$ for $n \in \{0, 1, 2, 3, 4, 5, 6, 7, 99\}$.

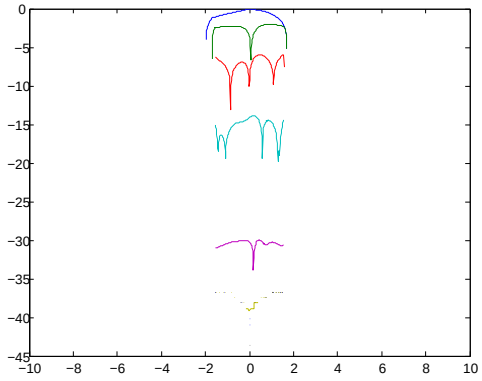


Figure 3.4: The graphs of $\log(|\pi_{n+1} - \pi_n|)$ for $n \in \{1, 2, 3, 4, 5, 6, 7, 99\}$.

The situation is very similar if we start with a different policy. Figures 3.5 and 3.6 show the graphs of the differences on the logarithmic scale for the initial policy $\pi_0(x) := \sin(10x)$.

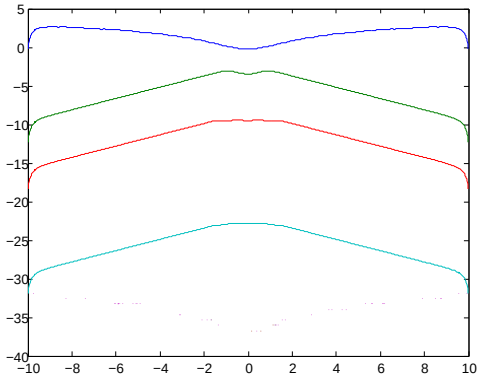


Figure 3.5: The graphs of $\log(|V_{\pi_{n+1}} - V_{\pi_n}|)$ for $n \in \{0, 1, 2, 3, 4, 5, 6, 7, 99\}$.

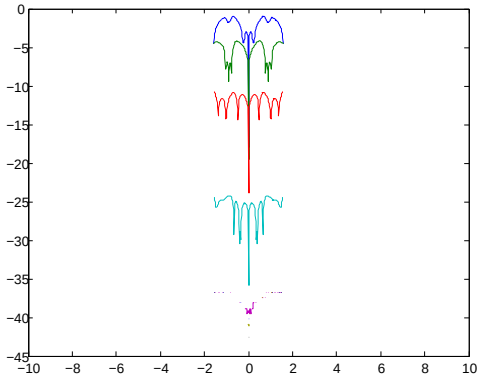


Figure 3.6: The graphs of $\log(|\pi_{n+1} - \pi_n|)$ for $n \in \{1, 2, 3, 4, 5, 6, 7, 99\}$.

3.5 Conclusion

In this chapter we presented the policy improvement algorithm for controlled diffusion processes in the framework of the discounted infinite horizon problem, separately for the one-dimensional and multidimensional setting. In both cases the chain of the main steps is the same. First we show that every function Lipschitz on compacts is a Markov policy, and that the corresponding payoff function satisfies a certain differential equation. Then we are able to define the algorithm, and next we prove that the policy indeed gets improved at each step. The payoff functions therefore converge, and we prove that this is also the case for a subsequence of the policies. We establish that the limit of the payoff functions and the payoff function corresponding to the limit policy are one and the same, which enables us to prove that the function so obtained is the value function of the problem (the limit policy is then an optimal policy).

The reason for treating separately the one-dimensional and multidimensional cases is that in the former we can be more general. Firstly, this is because we can use the normed version of the algorithm, which enables us to form a slightly weaker second assumption. (Note that the normed and non-normed algorithms in general produce different policies, yet all the theorems remain valid in both cases.) Secondly, in $\mathbb{R}$ every (simply) connected domain (i.e. an open interval) is $\mathcal{C}^2$-diffeomorphic to $\mathbb{R}$, whereas in $\mathbb{R}^n$, $n \geq 2$, this is not true. We used this to construct (possibly exiting) solutions on those domains.

Despite the fact that the framework in the multidimensional setting was slightly less general, an additional difficulty appeared. We had to prove a priori the continuity of the payoff functions in order to have a continuous boundary condition,

which enabled us to apply the elliptic differential equations theory (in one dimension we did not have this since the boundary consisted of only two points). We did this with the help of the mirror coupling of multidimensional diffusion processes. If we considered coupling to be a motive or goal in the previous chapter, we applied it here as a method or technique.

Even though we only treated the minimisation problem, it is easy to notice the following. Suppose we have a maximisation problem with the data σ, μ, α, f and g. If the functions σ, μ, α, $-f$ and $-g$ satisfy the assumptions, the value function obtained via the algorithm will be the value function of the maximisation problem multiplied by -1, and the optimal policies will coincide.

It would be interesting to know if the assumptions can be relaxed. Especially the second assumption looks unusual. We needed it to establish the existence of π_{lim}, but could we do without it? However, it was reassuring to find out that there were nontrivial examples of data that satisfied the assumptions. When implemented in Matlab, the convergence seemed very fast. Also, there was no indication that the sequence of policies had more than one accumulation point.

Comparing the continuous and discrete policy improvement algorithm, there are many similarities, such as analogies for policies, operators, the algorithm itself and its outcome. Boundedness of costs and a strictly positive discounting factor also played a role in both cases. However, the order of the steps is different since in the discrete case it is possible to prove (although we did not do it) a priori that the value function satisfies the optimality equation. In the continuous setting it is not clear in advance that the value function satisfies the HJB equation, in fact it might not even be smooth enough. Maybe the second assumption is the price that we need to pay for this.

There is another difference. In the discrete case we had no given probability space, therefore we could not talk about the control processes, just about the policies, and also when we considered the controlled processes, they were not uniquely specified, only their law was. In the continuous framework this would correspond to weak solutions, however we were working with strong solutions, which is not uncommon (see [18]). In principle we could try to work within the weak setting, however there is one possible danger. As we have seen, to obtain the desired estimates it is often useful (maybe even necessary) to use (specific) couplings. In [17, p. 308] it is proved that any two weak solutions of any stochastic differential equation can be coupled in such a way that they are driven by the same Brownian motion. But it is not clear that this is enough.

Chapter 4

The PIA for the continuous finite-horizon problem, and its application to coupling of GBMs

4.1 Introduction

This chapter connects the two main motivations of this thesis, which are: coupling of geometric Brownian motions and the policy improvement algorithm in a continuous setting.

In Section 4.2 we develop the policy improvement algorithm for the finite horizon problem involving time-inhomogeneous controlled processes in a continuous setting (continuous time, state space and paths, and general action space). We obtain analogous results to those in the previous chapter. Although the proofs are similar, too, we still carry them out because time dependence brings in different operators – parabolic differential equations instead of elliptic – and some other differences.

In Section 4.3 we apply the algorithm to characterise the value function of Problem (T+), which was the original motivation to start looking at the policy improvement algorithm in a continuous setting. The algorithm cannot be applied directly, though, due to insufficient smoothness of the data. To overcome this, we construct a family of approximating problems to which the algorithm can be applied, such that their value functions converge to the value function of Problem (T+).

4.2 The policy improvement algorithm for the continuous finite-horizon problem

4.2.1 Setting and the algorithm

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space (satisfying the usual conditions) that supports an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B_t)_{t\geq 0}$. Let $a, b \in [-\infty, \infty]$, $a < b$, and for any $\mathbb{R}$-valued process $Y = (Y_t)_{t\leq T}$ $(T > 0)$ define

$$\tau_a^b(Y) := \inf\{t \in [0,T];\ Y_t \leq a \text{ or } Y_t \geq b\} \quad (\inf \emptyset := \infty).$$

Let (A, d) be a compact metric space, and for any $x \in (a, b)$ and $T > 0$ define the set of admissible controls at (x, T) as

$$\begin{aligned}&\mathcal{A}(x,T) := \{\Pi = (\Pi_t)_{t<T};\ \Pi \text{ is an } A\text{-valued process adapted to } (\mathcal{F}_t)_{t<T}, \text{ and}\\ &\text{there exists a pathwise unique process } X^{x,T,\Pi} = (X_t^{x,T,\Pi})_{t\leq T} \text{ that satisfies (4.1)}\},\end{aligned}$$

where

$$\begin{aligned}&X_t^{x,T,\Pi} = x + \int_0^t \sigma\left(X_s^{x,T,\Pi}, T-s, \Pi_s\right) \mathrm{d}B_s + \int_0^t \mu\left(X_s^{x,T,\Pi}, T-s, \Pi_s\right) \mathrm{d}s\\ &\quad \text{if} \quad t \in \left[0, T \wedge \tau_a^b\left(X^{x,T,\Pi}\right)\right),\\ &X_t^{x,T,\Pi} = X^{x,T,\Pi}_{\tau_a^b(X^{x,T,\Pi})} \quad \text{if} \quad t \in \left[\tau_a^b\left(X^{x,T,\Pi}\right), T\right),\\ &\text{and} \quad X_T^{x,T,\Pi} = \lim_{t\uparrow T} X_t^{x,T,\Pi},\end{aligned} \tag{4.1}$$

and both $\sigma : (a,b)\times(0,\infty)\times A \to \mathbb{R}$ and $\mu : (a,b)\times(0,\infty)\times A \to \mathbb{R}$ are measurable functions.

Let $\alpha : (a,b)\times(0,\infty)\times A \to \mathbb{R}$, $f : (a,b)\times(0,\infty)\times A \to \mathbb{R}$ and $g : [a,b]\cap\mathbb{R} \to \mathbb{R}$ be measurable functions. For any $x \in (a,b)$, $T > 0$ and $\Pi \in \mathcal{A}(x,T)$ define the payoff by

$$\begin{aligned}V^{\Pi}(x,T) := \mathbb{E}\Bigg(&\int_0^{T\wedge\tau_a^b\left(X^{x,T,\Pi}\right)} \mathrm{e}^{-\int_0^t \alpha\left(X_s^{x,T,\Pi}, T-s, \Pi_s\right)\mathrm{d}s} f\left(X_t^{x,T,\Pi}, T-t, \Pi_t\right) \mathrm{d}t\\ &+\mathrm{e}^{-\int_0^{T\wedge\tau_a^b\left(X^{x,T,\Pi}\right)} \alpha\left(X_t^{x,T,\Pi}, T-t, \Pi_t\right)\mathrm{d}t} g\left(X^{x,T,\Pi}_{T\wedge\tau_a^b(X^{x,T,\Pi})}\right)\Bigg).\end{aligned}$$

The problem is to find the *value function* V, defined by

$$V(x,T) := \inf_{\Pi \in \mathcal{A}(x,T)} V^{\Pi}(x,T), \qquad x \in (a,b), \quad T > 0, \tag{4.2}$$

and an optimal policy (which will in general depend on x and T), if it exists.

In order to solve the problem, we will make some assumptions about the functions σ, μ, α, f and g. Before we state them, we will explain what Lipschitz continuity will mean in this chapter. Let $h : (a,b) \times (0,\infty) \to A$ and $K \subseteq (a,b) \times (0,\infty)$. Then h is Lipschitz on K if there exists $C > 0$ such that

$$d(h(x,t), h(y,s)) \leq C \left((x-y)^2 + |t-s|\right)^{\frac{1}{2}}$$

holds for every $(x,t),(y,s) \in K$. This is the standard Lipschitz continuity with respect to a slightly modified metric on $(a,b) \times (0,\infty)$. We will use it because it appears in Chapter 3 of [9], from where we will cite some results.

Assumption 4.2.1. Let σ, μ, α, f and g be bounded functions. Let g be continuous, σ^2 bounded away from 0, and α nonnegative. Additionally, let the following hold: for every $h \in \{\sigma, \mu, \alpha, f\}$ and every compact set $K \subseteq (a,b) \times (0,\infty)$ there exists $C > 0$ such that

$$|h(x,t,p) - h(y,s,r)| \leq C \left(\left((x-y)^2 + |t-s|\right)^{\frac{1}{2}} + d(p,r) \right)$$

holds for every $(x,t),(y,s) \in K$ and $p,r \in A$.

Assumption 4.2.2. For every $h \in \mathcal{C}^{2,1}((a,b) \times (0,\infty))$, $x \in (a,b)$ and $t > 0$, let $I^h(x,t)$ denote a point where the minimum of the function

$$p \mapsto \frac{1}{2}\sigma(x,t,p)^2 h_{xx}(x,t) + \mu(x,t,p) h_x(x,t) - \alpha(x,t,p) h(x,t) + f(x,t,p), \quad p \in A,$$

is attained. If the sequence $\{(h_n)_{xx}\}_{n \in \mathbb{N}}$ is uniformly bounded (i.e. there exists a constant that is a bound for all the functions) on compacts in $(a,b) \times (0,\infty)$, then the points $\{I^{h_n}(x,t);\ x \in (a,b), t > 0, n \in \mathbb{N}\}$ can be chosen in such a way that the sequence of functions $\{I^{h_n} : (a,b) \times (0,\infty) \to A\}_{n \in \mathbb{N}}$ is uniformly Lipschitz on compacts in $(a,b) \times (0,\infty)$.

A measurable function $\pi : (a,b) \times (0,\infty) \to A$ is a *Markov policy* if for every $x \in (a,b)$ and $T > 0$ there exists a pathwise unique process $X^{x,T,\pi} = \left(X_t^{x,T,\pi}\right)_{t \leq T}$

that satisfies the following:

$$\begin{aligned} X_t^{x,T,\pi} &= x + \int_0^t \sigma\left(X_s^{x,T,\pi}, T-s, \pi\left(X_s^{x,T,\pi}, T-s\right)\right) \mathrm{d}B_s \\ &\quad + \int_0^t \mu\left(X_s^{x,T,\pi}, T-s, \pi\left(X_s^{x,T,\pi}, T-s\right)\right) \mathrm{d}s \quad \text{if} \quad t \in \left[0, T \wedge \tau_a^b\left(X^{x,T,\pi}\right)\right), \\ X_t^{x,T,\pi} &= X^{x,T,\pi}_{\tau_a^b(X^{x,T,\pi})} \quad \text{if} \quad t \in \left[\tau_a^b\left(X^{x,T,\pi}\right), T\right), \\ \text{and} \quad X_T^{x,T,\pi} &= \lim_{t \uparrow T} X_t^{x,T,\pi}. \end{aligned} \tag{4.3}$$

If π is a Markov policy, then $\pi\left(X_\cdot^{x,T,\pi}, T-\cdot\right) := \left(\pi\left(X_t^{x,T,\pi}, T-t\right)\right)_{t<T} \in \mathcal{A}(x,T)$ for every $x \in (a,b)$ and $T > 0$ (if $a > -\infty$, $\pi(a,\cdot)$ is defined as any function with values in A; ditto for b). For easier notation we define

$$\sigma_\pi(x,t) := \sigma(x,t,\pi(x,t)), \quad \mu_\pi(x,t) := \mu(x,t,\pi(x,t)),$$

$$\alpha^\pi(x,t) := \alpha(x,t,\pi(x,t)), \quad f^\pi(x,t) := f(x,t,\pi(x,t)),$$

$$V^\pi(x,T) := V^{\pi\left(X_\cdot^{x,T,\pi}, T-\cdot\right)}(x,T) \quad \text{for} \quad x \in (a,b), T > 0,$$

$$L^\pi h := \frac{1}{2}\sigma_\pi^2 h_{xx} + \mu_\pi h_x - h_t \quad \text{for} \quad h \in C^{2,1}((a,b)\times(0,\infty)).$$

If π is a constant Markov policy with the value $p \in A$, we will write σ_p, μ_p, α^p, f^p and L^p instead of σ_π, μ_π, α^π, f^π and L^π, respectively.

Proposition 4.2.3. *If function $\pi : (a,b)\times(0,\infty) \to A$ is Lipschitz on compacts in $(a,b)\times(0,\infty)$, then π is a Markov policy.*

Proposition 4.2.4. *For any Markov policy π that is Lipschitz on compacts in $(a,b)\times(0,\infty)$, the following holds: $V^\pi \in \mathcal{C}^{2,1}((a,b)\times(0,\infty))$ and*

$$L^\pi V^\pi - \alpha^\pi V^\pi + f^\pi = 0.$$

Let π_0 be a Markov policy that is Lipschitz on compacts in $(a,b)\times(0,\infty)$. The *policy improvement algorithm* is defined in the following way:

$$\begin{aligned} &\pi_{n+1}(x,T) := \operatorname*{argmin}_{p\in A}\left(L^p V^{\pi_n}(x,T) - \alpha^p(x,T)V^{\pi_n}(x,T) + f^p(x,T)\right), \\ &x \in (a,b), \quad T > 0, \qquad n \in \mathbb{N}_0. \end{aligned} \tag{4.4}$$

Note $\pi_{n+1}(x,T) = \operatorname{argmin}_{p\in A}\left(\frac{1}{2}\sigma_p^2 V_{xx}^{\pi_n} + \mu_p V_x^{\pi_n} - \alpha^p V^{\pi_n} + f^p\right)(x,T)$. Since $V_{xx}^{\pi_n}$ is bounded on compacts in $(a,b)\times(0,\infty)$), Assumption 4.2.2 (applied separately for every $n \in \mathbb{N}_0$) ensures that the points $\{\pi_{n+1}(x,T);\ x \in (a,b), T > 0\}$ can be

chosen in such a way that $\pi_{n+1} : (a,b) \times (0,\infty) \to A$ is Lipschitz on compacts in $(a,b) \times (0,\infty)$.

Remark 4.2.5. If the algorithm stops, i.e. $\pi_{n+1} = \pi_n$ for some $n \in \mathbb{N}_0$, then clearly $V^{\pi_m} = V^{\pi_n}$ and $\pi_m = \pi_n$ hold for every $m \geq n$. We can then proceed directly to the verification lemma (Theorem 4.2.9) to prove that V^{π_n} is the value function and π_n is an optimal policy.

As usual, we will now prove that every policy generated by the algorithm is at least as good as its predecessor.

Theorem 4.2.6. *For every $n \in \mathbb{N}_0$, $x \in (a,b)$, $T > 0$ and Markov policy π_0 that is Lipschitz on compacts in $(a,b) \times (0,\infty)$, the following holds:*

$$V^{\pi_{n+1}}(x,T) \leq V^{\pi_n}(x,T).$$

Since $\{V^{\pi_n}\}_{n\in\mathbb{N}}$ is a bounded decreasing sequence, we can define

$$V^{\lim}(x,T) := \lim_{n\to\infty} V^{\pi_n}(x,T), \quad x \in (a,b), \quad T > 0.$$

The sequence of policies might not converge, but the next proposition says that there exists a convergent subsequence.

Proposition 4.2.7. *There exists a subsequence of $\{\pi_n\}_{n\in\mathbb{N}}$ that converges uniformly on every compact subset of $(a,b) \times (0,\infty)$.*

Therefore this subsequence converges (pointwise) on $(a,b) \times (0,\infty)$; denote the limit by $\pi_{\lim}$.

Note that $V^{\lim}$ and $\pi_{\lim}$ can in principle depend on π_0. Additionally, $\pi_{\lim}$ could depend on the choice of the subsequence from Proposition 4.2.7. However, this turns out to be irrelevant in the following theorem, which brings together the limit payoff function and the limit policy.

Theorem 4.2.8. *For every $x \in (a,b)$, $T > 0$ and Markov policy π_0 that is Lipschitz on compacts in $(a,b) \times (0,\infty)$, the following holds:*

$$V^{\lim}(x,T) = V^{\pi_{\lim}}(x,T).$$

The last step establishes $V^{\lim} = V$ via the so called verification lemma.

Theorem 4.2.9. *For every $x \in (a,b)$, $T > 0$, $\Pi \in \mathcal{A}(x,T)$ and Markov policy π_0 that is Lipschitz on compacts in $(a,b) \times (0,\infty)$, the following holds:*

$$V^{\lim}(x,T) \leq V^{\Pi}(x,T).$$

Hence $V^{\lim}$ is the value function (and does not depend on π_0) and $\pi_{\lim}$ is an optimal policy.

4.2.2 Auxiliary results

Lemma 4.2.10. *For every Markov policy π, the payoff function V^π, which is defined on $(a,b)\times(0,\infty)$, can be continuously extended by defining*

$$V^\pi(x,0) := g(x) \quad \textit{for} \quad x \in (a,b),$$

$$V^\pi(a,T) := g(a) \quad \textit{for} \quad T \geq 0 \quad \textit{if} \quad a > -\infty,$$

and

$$V^\pi(b,T) := g(b) \quad \textit{for} \quad T \geq 0 \quad \textit{if} \quad b < \infty.$$

Proof. The proof of

$$\lim_{x\downarrow a} V^\pi(x,T) = g(a), \quad \lim_{x\uparrow b} V^\pi(x,T) = g(b), \quad T \geq 0,$$

is analogous to the proof of Lemma 3.2.11. To see

$$\lim_{T\downarrow 0} V^\pi(x,T) = g(x), \quad x \in (a,b),$$

it suffices to apply the Dominated Convergence Theorem (since f and g are bounded by Assumption 4.2.1). □

The processes controlled by Markov policies are strong Markov processes (Theorem 4.20 in [17, p. 322]; according to the introduction of that section the theorem still holds in the time-inhomogeneous case). This enables us to prove the following lemma.

Lemma 4.2.11. *The following holds for every Markov policy π, $x \in (a,b)$, $T > 0$*

and stopping time S that is almost surely less than or equal to $T \wedge \tau_a^b\left(X^{x,T,\pi}\right)$:

$$\begin{aligned}
&\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t \\
&\qquad + \mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} V^\pi\left(X_S^{x,T,\pi},T-S\right) \\
&= \mathbb{E}\left(\int_0^{T\wedge\tau_a^b\left(X^{x,T,\pi}\right)} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right. \\
&\qquad \left. + \mathrm{e}^{-\int_0^{T\wedge\tau_a^b\left(X^{x,T,\pi}\right)} \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} g\left(X^{x,T,\pi}_{T\wedge\tau_a^b\left(X^{x,T,\pi}\right)}\right) \,\middle|\, \mathcal{F}_S\right).
\end{aligned}$$

In particular, the process M defined below is a uniformly integrable martingale:

$$\begin{aligned}
M_t := &\int_0^t \mathrm{e}^{-\int_0^r \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r \\
&+ \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} V^\pi\left(X_t^{x,T,\pi},T-t\right), \quad t \le T.
\end{aligned}$$

Proof. Let $\tau := \tau_a^b\left(X^{x,T,\pi}\right)$ and $\tau_S := \tau \circ \theta_S = \tau_a^b\left(X^{x,T+S,\pi}_{\cdot+S}\right)$. Then $\tau_S = \tau - S$

holds almost surely, and we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{T\wedge\tau} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right.\\
&\quad\left.+\mathrm{e}^{-\int_0^{T\wedge\tau} \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} g\left(X_{T\wedge\tau}^{x,T,\pi}\right)\,\middle|\,\mathcal{F}_S\right)\\
&=\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\\
&\quad+\mathbb{E}\left(\int_0^{(T\wedge\tau)-S} \mathrm{e}^{-\int_0^{t+S} \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_{t+s}^{x,T,\pi},T-S-t\right)\mathrm{d}t\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{T\wedge\tau} \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} g\left(X_{T\wedge\tau}^{x,T,\pi}\right)\,\middle|\,\mathcal{F}_S\right)\\
&=\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t+\mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s}\\
&\quad\cdot\mathbb{E}\left(\int_0^{(T-S)\wedge\tau_S} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_{s+S}^{x,T,\pi},T-S-s\right)\mathrm{d}s} f^\pi\left(X_{t+S}^{x,T,\pi},T-S-t\right)\mathrm{d}t\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{(T-S)\wedge\tau_S} \alpha^\pi\left(X_{s+S}^{x,T,\pi},T-S-s\right)\mathrm{d}s} g\left(X_{(T-S)\wedge\tau_S+S}^{x,T,\pi}\right)\,\middle|\,\mathcal{F}_S\right)\\
&=\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t+\mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s}\\
&\quad\cdot\mathbb{E}\left(\int_0^{(T-S)\wedge\tau} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T-S,\pi},T-S-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T-S,\pi},T-S-t\right)\mathrm{d}t\circ\theta_S\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{(T-S)\wedge\tau} \alpha^\pi\left(X_s^{x,T-S,\pi},T-S-s\right)\mathrm{d}s} g\left(X_{(T-S)\wedge\tau}^{x,T-S,\pi}\right)\circ\theta_S\,\middle|\,\mathcal{F}_S\right).
\end{aligned}
$$

Applying the strong Markov property, we arrive to

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^{T\wedge\tau} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right.\\
&\quad\left.+\mathrm{e}^{-\int_0^{T\wedge\tau} \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} g\left(X_{T\wedge\tau}^{x,T,\pi}\right)\,\middle|\,\mathcal{F}_S\right)\\
&=\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t+\mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s}\\
&\quad\cdot\mathbb{E}_{X_S^{x,T,\pi}}\left(\int_0^{(T-S)\wedge\tau} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{T-S,\pi},T-S-s\right)\mathrm{d}s} f^\pi\left(X_t^{T-S,\pi},T-S-t\right)\mathrm{d}t\right.\\
&\qquad\left.\left.+\mathrm{e}^{-\int_0^{(T-S)\wedge\tau} \alpha^\pi\left(X_s^{T-S,\pi},T-S-s\right)\mathrm{d}s} g\left(X_{(T-S)\wedge\tau}^{T-S,\pi}\right)\right)\right.\\
&=\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\\
&\quad+\mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} V^\pi\left(X_S^{x,T,\pi},T-S\right).
\end{aligned}
$$

□

By applying the expectation, the lemma implies the so called Bellman's principle.

Corollary 4.2.12. *The following holds for every Markov policy π, $x \in (a,b)$, $T > 0$ and stopping time S that is almost surely less than or equal to $T \wedge \tau_a^b\left(X^{x,T,\pi}\right)$:*

$$V^\pi(x,T) = \mathbb{E}\left(\int_0^S \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right) + \mathbb{E}\left(\mathrm{e}^{-\int_0^S \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} V^\pi\left(X_S^{x,T,\pi},T-S\right)\right).$$

In the next two lemmas we will establish the (joint) continuity of the payoff functions. The mirror coupling is again one of the techniques used.

Lemma 4.2.13. *For every Markov policy π, $x \in (a,b)$, $T > 0$ and $\epsilon > 0$ there exists $\delta > 0$ such that for every $y \in (a,b)$ and $T' > 0$ the following holds: $|x-y| < \delta$ and $|T-T'| < \delta$ imply*

$$|V^\pi(x,T') - V^\pi(y,T')| \le \epsilon.$$

Proof. Recall that the process $X^{x,T,\pi}$ solves (4.3). Let the process X^y (for an arbitrary $y \in (a,b)$) solve the analogous stochastic differential equations but with (y instead of x and) $-B$ instead of B, i.e.

$$X_t^y = y - \int_0^t \sigma_\pi\left(X_s^y, T-s\right)\mathrm{d}B_s + \int_0^t \mu\left(X_s^y,T-s\right)\mathrm{d}s \quad \text{if} \quad t \in \left[0, T\wedge\tau_a^b\left(X^y\right)\right),$$
$$X_t^y = X^y_{\tau_a^b(X^y)} \quad \text{if} \quad t \in \left[\tau_a^b\left(X^y\right), T\right), \quad \text{and} \quad X_T^y = \lim_{t\uparrow T} X_t^y.$$

Let $\hat{\tau}$ be their first meeting time, i.e.

$$\hat{\tau} := \inf\left\{t \in [0,T];\ X_t^{x,T,\pi} = X_t^y\right\} \quad (\inf\emptyset := \infty),$$

and define

$$X_t^x := X_t^{x,T,\pi}\mathbb{I}_{\{t<\hat{\tau}\}} + X_t^y\mathbb{I}_{\{t\ge\hat{\tau}\}}, \quad t \le T.$$

Due to the strong Markov property, the laws of the processes X^x and $X^{x,T,\pi}$ are the same, and clearly the equality $X_t^x = X_t^y$ holds on the event $\{\hat{\tau} \le t\}$ for every $t \le T$.

Let M be a constant that bounds all the functions σ_π, μ_π, f^π, V^π and g, and assume (without loss of generality) that $x \ge y$. Then the following holds on the

event $\left\{\hat{\tau} \leq \tau_a^b\left(X^x\right) \wedge \tau_a^b\left(X^y\right)\right\}$:

$$\begin{aligned}
\hat{\tau} &\leq \inf\left\{t \in [0,T];\ x-y = \int_0^t \left(\mu_\pi\left(X_s^x, T-s\right) - \mu_\pi\left(X_s^y, T-s\right)\right) \mathrm{d}s. \right. \\
&\qquad\qquad \left. + \int_0^t \left(\sigma_\pi\left(X_s^x, T-s\right) + \sigma_\pi\left(X_s^y, T-s\right)\right) \mathrm{d}B_s \right\} \\
&\leq \inf\left\{t \in [0,T];\ x-y = -2Mt + \int_0^t \left(\sigma_\pi\left(X_s^x, T-s\right) + \sigma_\pi\left(X_s^y, T-s\right)\right) \mathrm{d}B_s \right\} \\
&= \inf\left\{t \in [0,T];\ x-y = W_{[N]_t} - 2Mt\right\},
\end{aligned}$$

where we used Dambis-Dubins-Schwartz theorem for the martingale N defined as

$$N_t := \int_0^t \left(\sigma_\pi\left(X_s^x, T-s\right) + \sigma_\pi\left(X_s^y, T-s\right)\right) \mathrm{d}B_s, \quad t \leq T,$$

and W is a Brownian motion. Since σ^2 is bounded away from 0, we know that there exists a constant $m > 0$ such that $[N]_t \geq mt$ for all $t \in [0,T]$. Thus we obtain

$$\begin{aligned}
\hat{\tau} &\leq \inf\left\{t \in [0,T];\ x-y = W_{mt} - 2Mt\right\} \\
&=: \tau \overset{\mathcal{L}}{=} \inf\left\{t \in [0,T];\ \frac{x-y}{\sqrt{m}} = W_t - \frac{2M}{\sqrt{m}}t\right\}.
\end{aligned}$$

The density of the first hitting time of the level l by the Brownian motion with drift c is (see e.g. [17, p. 197])

$$p(t) := \frac{|l|}{\sqrt{2\pi t^3}} \exp\left(\frac{-(l-ct)^2}{2t}\right), \quad t > 0,$$

from where we can estimate

$$\begin{aligned}
\mathbb{E}\left(\tau \mathbb{I}_{\{\tau \leq T\}}\right) &= \int_0^T t \frac{\left|\frac{x-y}{\sqrt{m}}\right|}{\sqrt{2\pi t^3}} \exp\left(\frac{-\left(\frac{x-y}{\sqrt{m}} - \frac{2M}{\sqrt{m}}t\right)^2}{2t}\right) \mathrm{d}t \\
&\leq \int_0^T t \frac{|x-y|}{\sqrt{2m\pi t^3}} \,\mathrm{d}t = C_1\sqrt{T}|x-y|
\end{aligned}$$

and

$$\mathbb{P}\left(\tau > T\right) = \int_T^\infty \frac{\left|\frac{x-y}{\sqrt{m}}\right|}{\sqrt{2\pi t^3}} \exp\left(\frac{-\left(\frac{x-y}{\sqrt{m}} - \frac{2M}{\sqrt{m}}t\right)^2}{2t}\right) \mathrm{d}t \leq \int_T^\infty \frac{|x-y|}{\sqrt{2m\pi t^3}} \,\mathrm{d}t = \frac{C_2}{\sqrt{T}}|x-y|$$

for some constants C_1 and C_2. With τ_a^b denoting $\tau_a^b(X^x)$ or $\tau_a^b(X^y)$ when they are the same, we obtain

$$
\begin{aligned}
&|V^\pi(x,T) - V^\pi(y,T)| \\
&\mathbb{E}\left(|V^\pi(x,T) - V^\pi(y,T)|\,\mathbb{I}_{\{\tau > T\wedge\tau_a^b(X^x)\wedge\tau_a^b(X^y)\}}\right) \\
&+ \mathbb{E}\left(\left(\int_0^\tau \left|\mathrm{e}^{-\int_0^t \alpha^\pi(X_s^x,T-s)\mathrm{d}s} f^\pi\left(X_t^x, T-t\right) - \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^y,T-s\right)\mathrm{d}s} f^\pi\left(X_t^y, T-t\right)\right| \mathrm{d}t\right.\right. \\
&+ \int_\tau^{T\wedge\tau_a^b} \left|\mathrm{e}^{-\int_0^t \alpha^\pi(X_s^x,T-s)\mathrm{d}s} - \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^y,T-s\right)\mathrm{d}s}\right| \left|f^\pi\left(X_t^x, T-t\right)\right| \mathrm{d}t \\
&\left.\left. + \left|\mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^\pi(X_s^x,T-s)\mathrm{d}s} - \mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^\pi\left(X_s^y,T-s\right)\mathrm{d}s}\right| \left|g\left(X_{T\wedge\tau_a^b}^x\right)\right|\right) \mathbb{I}_{\{\tau \le T\wedge\tau_a^b(X^x)\wedge\tau_a^b(X^y)\}}\right).
\end{aligned}
$$

For $t > \tau$ we have

$$
\begin{aligned}
&\left|\mathrm{e}^{-\int_0^t \alpha^\pi(X_s^x,T-s)\mathrm{d}s} - \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^y,T-s\right)\mathrm{d}s}\right| \\
&= \left|\mathrm{e}^{-\int_0^\tau \alpha^\pi(X_s^x,T-s)\mathrm{d}s} - \mathrm{e}^{-\int_0^\tau \alpha^\pi\left(X_s^y,T-s\right)\mathrm{d}s}\right| \mathrm{e}^{-\int_\tau^t \alpha^\pi(X_s^x,T-s)\mathrm{d}s} \\
&\le \left|\int_0^\tau \alpha^\pi\left(X_s^x, T-s\right)\mathrm{d}s - \int_0^\tau \alpha^\pi\left(X_s^y, T-s\right)\mathrm{d}s\right| \\
&\le 2M\tau,
\end{aligned}
$$

and therefore for every $\delta' \in (0,T)$ we obtain

$$
\begin{aligned}
&|V^\pi(x,T) - V^\pi(y,T)| \\
&\le 2M\,\mathbb{P}\left(\tau > T\wedge\tau_a^b(X^x)\wedge\tau_a^b(X^y)\right) + \mathbb{E}\left(\left(2M\tau + 2M\tau\cdot MT + 2M\tau\cdot M\right)\mathbb{I}_{\{\tau\le T\}}\right) \\
&\le 2M\mathbb{P}\left(\tau > \delta'\right) + 2M\mathbb{P}\left(\tau_a^b(X^x)\wedge\tau_a^b(X^y) < \delta'\right) \\
&\quad + (2M + 2M^2 + 2M^2T)\mathbb{E}\left(\tau\mathbb{I}_{\{\tau\le T\}}\right) \\
&\le 2M\frac{C_2}{\sqrt{\delta'}}|x-y| + 2M\mathbb{P}\left(\tau_a^b(X^x)\wedge\tau_a^b(X^y) < \delta'\right) \\
&\quad + \left(2M + 2M^2 + 2M^2T\right)C_1\sqrt{T}|x-y|.
\end{aligned}
$$

Fix $\epsilon > 0$, $x \in (a,b)$ and $T > 0$. We know that we can choose δ' such that

$$
2M\mathbb{P}\left(\tau_a^b(X^x)\wedge\tau_a^b(X^y) < \delta'\right) < \frac{\epsilon}{3}
$$

holds for all y close to x and T' close to T (dependence on T is not visible in order

to have lighter notation). Let $\delta > 0$ be so small that

$$2M\frac{C_2}{\sqrt{\delta'}}\,\delta < \frac{\epsilon}{3} \quad \text{and} \quad \left(2M + 2M^2 + 2M^2T\right) C_1\sqrt{T}\,\delta < \frac{\epsilon}{3}.$$

It is clear that the same δ will work for all T' close to T, which concludes the proof. □

Now we will apply the previous two statements to prove the continuity of the payoff functions for Markov policies.

Lemma 4.2.14. *For every Markov policy π, the function V^π is continuous.*

Proof. Let $M > 1$ be a constant that bounds all the functions σ_π, μ_π, f^π and V^π. Later in the proof we will need the following inequality ($x \in (a,b)$, $T > 0$, $t < T$):

$$\mathbb{P}\left(\left|X_t^{x,T,\pi} - x\right| > \sqrt[3]{t},\, \tau_a^b\left(X^{x,T,\pi}\right) \geq t\right) \leq \mathbb{P}\left(|B_1| > \frac{\sqrt[3]{t} - Mt}{M\sqrt{t}}\right). \tag{4.5}$$

Note that this probability goes to 0 as t tends to 0 (B_1 is a standard normal random variable).

We will prove this inequality by comparing $X_t^{x,T,\pi} - x$ to the Brownian motion with drift. Note that there exists a Brownian motion W such that

$$N_s := \int_0^s \sigma_\pi\left(X_r^{x,T,\pi}, T-r\right) \mathrm{d}B_r = W_{[N]_s}, \quad [N]_s \leq M^2 s, \quad s \leq T.$$

Then we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left|X_t^{x,T,\pi} - x\right| > \sqrt[3]{t},\, \tau_a^b\left(X^{x,T,\pi}\right) \geq t\right) \\
&\leq \mathbb{P}\left(\left|\int_0^t \mu_\pi\left(X_r^{x,T,\pi}, T-r\right) \mathrm{d}r\right| + \left|\int_0^t \sigma_\pi\left(X_r^{x,T,\pi}, T-r\right) \mathrm{d}B_r\right| > \sqrt[3]{t}\right) \\
&\leq \mathbb{P}\left(\left|W_{[N]_t}\right| > \sqrt[3]{t} - Mt\right) \leq \mathbb{P}\left(\left|\sup_{s\leq t} W_{[N]_s}\right| > \sqrt[3]{t} - Mt\right) \\
&\leq \mathbb{P}\left(\left|\sup_{s\leq t} W_{M^2 s}\right| > \sqrt[3]{t} - Mt\right) = \mathbb{P}\left(|W_1| > \frac{\sqrt[3]{t} - Mt}{M\sqrt{t}}\right)
\end{aligned}$$

by the Reflection Principle.

For any $x, x' \in (a,b)$ and $T > T' > 0$ the following holds by Corollary 4.2.12:

$$
\begin{aligned}
&\left|V^\pi(x,T) - V^\pi(x',T')\right| \\
&\leq \left|V^\pi(x,T) - V^\pi(x,T')\right| + \left|V^\pi(x,T') - V^\pi(x',T')\right| \\
&\leq \mathbb{E}\left(\left|\int_0^{T-T'} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right.\right. \\
&\quad \left.\left. + \mathrm{e}^{-\int_0^{T-T'} \alpha^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t} V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi(x,T')\right| \mathbb{I}_{\left\{\tau_a^b\left(X^{x,T,\pi}\right)\geq T-T'\right\}}\right) \\
&\quad + \left|V^\pi(x,T) - V^\pi(x,T')\right| \mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) < T-T'\right) + \left|V^\pi(x,T') - V^\pi(x',T')\right|.
\end{aligned}
$$

We further obtain

$$
\begin{aligned}
&\left|V^\pi(x,T) - V^\pi(x',T')\right| \\
&\leq M(T-T') + \mathbb{E}\left(\left|\mathrm{e}^{-\int_0^{T-T'} \alpha^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t} V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right)\right|\right) \\
&\quad + \mathbb{E}\left(\left|V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi(x,T')\right| \mathbb{I}_{\left\{\tau_a^b\left(X^{x,T,\pi}\right)\geq T-T'\right\}}\right) \\
&\quad + 2M\mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) < T-T'\right) + \left|V^\pi(x,T') - V^\pi(x',T')\right| \\
&\leq M(T-T') + \mathbb{E}\left(\left|\int_0^{T-T'} \alpha^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right| \left|V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right)\right|\right) \\
&\quad + \mathbb{E}\left(\left|V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi(x,T')\right| \mathbb{I}_{\left\{\tau_a^b\left(X^{x,T,\pi}\right)\geq T-T',\, \left|X_{T-T'}^{x,T,\pi}-x\right| > \sqrt[3]{T-T'}\right\}}\right) \\
&\quad + \mathbb{E}\left(\left|V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi(x,T')\right| \mathbb{I}_{\left\{\tau_a^b\left(X^{x,T,\pi}\right)\geq T-T',\, \left|X_{T-T'}^{x,T,\pi}-x\right| \leq \sqrt[3]{T-T'}\right\}}\right) \\
&\quad + 2M\mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) < T-T'\right) + \left|V^\pi(x,T') - V^\pi(x',T')\right| \\
&\leq M(T-T') + M^2(T-T') \\
&\quad + 2M\mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) \geq T-T',\, \left|X_{T-T'}^{x,T,\pi} - x\right| > \sqrt[3]{T-T'}\right) \\
&\quad + \mathbb{E}\left(\left|V^\pi\left(X_{T-T'}^{x,T,\pi},T'\right) - V^\pi(x,T')\right| \mathbb{I}_{\left\{\left|X_{T-T'}^{x,T,\pi}-x\right| \leq \sqrt[3]{T-T'}\right\}}\right) \\
&\quad + 2M\mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) < T-T'\right) + \left|V^\pi(x,T') - V^\pi(x',T')\right|
\end{aligned}
$$

Fix now $x \in (a,b)$ and $T > 0$, and set $\epsilon > 0$. Let $\delta_1 > 0$ be such that

$$M^2\delta_1 < \frac{\epsilon}{6}.$$

Due to (4.5) there exists $\delta_2 > 0$ such that

$$2M\mathbb{P}\left(|W_1| > \frac{\sqrt[3]{\delta_2} - M\delta_2}{M\sqrt{\delta_2}}\right) < \frac{\epsilon}{6}.$$

From Lemma 4.2.13 we know that there exists $\delta_3 \in (0,1)$ such that

$$|x - x'| < \delta_3, \quad |T - T'| < \delta_3 \quad \Longrightarrow \quad \left|V^\pi(x,T') - V^\pi(x',T')\right| < \frac{\epsilon}{6}.$$

It is also clear (since $\tau_a^b\left(X^{x,T,\pi}\right)$ is positive almost surely) that there exists $\delta_4 > 0$ such that

$$2M\mathbb{P}\left(\tau_a^b\left(X^{x,T,\pi}\right) < \delta_4\right) < \frac{\epsilon}{6}.$$

Set

$$\delta := \delta_1 \wedge \delta_2 \wedge \delta_3^3 \wedge \delta_4.$$

Then we have proved above that $|x - x'| < \delta$ and $|T - T'| < \delta$ (we assumed $T > T'$ without loss of generality) imply $|V^\pi(x,T) - V^\pi(x',T')| < \epsilon$. □

4.2.3 Proofs

Proof of Proposition 4.2.3. We obtain the process $\left(X_t^{x,T,\pi}\right)_{t<T}$ via the same state-space transformation as in the proof of Proposition 3.2.4. To prove that the limit $\lim_{t\uparrow T} X_t^{x,T,\pi}$ exists, it is enough to do this on the event $\left\{\tau_a^b\left(X^{x,T,\pi}\right) \geq T\right\}$. The equality

$$\lim_{t\uparrow T}\int_0^t \mu_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}s = \int_0^T \mu_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}s$$

holds because μ_π is bounded (and it does not not matter how we define $\mu_\pi(\cdot,0)$). The equality

$$\lim_{t\uparrow T}\int_0^t \sigma_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s = \int_0^T \sigma_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s$$

follows from the Martingale Convergence Theorem, which we apply to

$$Y_t := \int_0^{h^{-1}(t)} \sigma_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s, \quad t \geq 0,$$

where $h : [0,T) \to [0,\infty)$ is a homeomorphism. We then use

$$\lim_{t\uparrow T}\int_0^t \sigma_\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s = \lim_{t\uparrow T} Y_{h(t)} = \lim_{t\uparrow\infty} Y_t.$$

□

Proof of Proposition 4.2.4. Let $a < a' < x < b' < b$ and $T > 0$. Let $v \in \mathcal{C}([a', b'] \times [0, T]) \cap \mathcal{C}^{2,1}((a', b') \times (0, T])$ be the unique solution of the initial-boundary value problem

$$\begin{aligned} &L^\pi v + f^\pi = 0, \quad v(y, 0) = g(y) \quad \text{for} \quad y \in [a', b'], \\ &v(a', t) = V^\pi(a', t) \quad \text{and} \quad v(b', t) = V^\pi(b', t) \quad \text{for} \quad t \in [0, T], \end{aligned}$$

which is guaranteed to exist by Corollary 1 in [9, p. 71] since by Lemma 4.2.14, Lemma 4.2.10 and Assumption 4.2.1 the initial-boundary conditions are continuous and the coefficients are Hölder continuous (in fact they are even Lipschitz). Let $a' < a'' < x < b'' < b'$ and $\tau_c^d := \tau_c^d(X^{x,T,\pi})$ for any $c < d$. For every $\epsilon \geq 0$ define the process $S^{\epsilon,a'',b''}$ and analogously $S^{\epsilon,a',b'}$ by

$$\begin{aligned} S_t^{\epsilon,a'',b''} := &\int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} f^\pi\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}s \\ &+ \mathrm{e}^{-\int_0^{t\wedge\tau_{a''}^{b''}} \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} v\left(X_{t\wedge\tau_{a''}^{b''}}^{x,T,\pi}, T-t\wedge\tau_{a''}^{b''}\right), \quad t \leq T-\epsilon. \end{aligned}$$

Itô's formula on $[0, \tau_{a''}^{b''}]$ and the differential equation for v yield, for every small enough $\epsilon > 0$,

$$\begin{aligned} S_t^{\epsilon,a'',b''} &= v(x,T) + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} \sigma_\pi v_x\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s \\ &\quad + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} \left(f^\pi + L^\pi v - \alpha^\pi v\right)\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}s \\ &= v(x,T) + \int_0^{t\wedge\tau_{a''}^{b''}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} \sigma_\pi v_x\left(X_s^{x,T,\pi}, T-s\right)\mathrm{d}B_s. \end{aligned}$$

Hence $S^{\epsilon,a'',b''}$ is a local martingale, and since it is clearly a bounded process, it is a uniformly integrable martingale. Thus the Dominated Convergence Theorem yields

$$v(x,T) = \lim_{a''\to a'} \lim_{b''\to b'} \lim_{\epsilon\to 0} \mathbb{E}\left(S_0^{\epsilon,a'',b''}\right) = \lim_{a''\to a'} \lim_{b''\to b'} \lim_{\epsilon\to 0} \mathbb{E}\left(S_{T-\epsilon}^{\epsilon,a'',b''}\right) = \mathbb{E}\left(S_T^{0,a',b'}\right).$$

Thanks to the initial-boundary conditions for v, Lemma 4.2.10 and Lemma 4.2.11,

we obtain

$$\begin{aligned}
S_T^{0,a',b'} &= \int_0^{T\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} f^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s \\
&\quad + \mathrm{e}^{-\int_0^{T\wedge\tau_{a'}^{b'}} \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} v\left(X_{T\wedge\rho_b}^{x,T,\pi},T-(T\wedge\tau_{a'}^{b'})\right) \\
&= \int_0^{T\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} f^\pi\left(X_s^{x,T,\pi},T-s\right)\mathrm{d}s \\
&\quad + \mathrm{e}^{-\int_0^{T\wedge\tau_{a'}^{b'}} \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} V^\pi\left(X_{T\wedge\tau_{a'}^{b'}}^{x,T,\pi},T-(T\wedge\tau_{a'}^{b'})\right) \\
&= \mathbb{E}\left(\int_0^{T\wedge\tau_a^b} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right. \\
&\qquad \left. + \mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} g\left(X_{T\wedge\tau_a^b}^{x,T,\pi}\right)\middle|\mathcal{F}_{T\wedge\tau_{a'}^{b'}}\right),
\end{aligned}$$

and therefore

$$\begin{aligned}
v(x,T) &= \mathbb{E}\left(\int_0^{T\wedge\tau_a^b} \mathrm{e}^{-\int_0^t \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} f^\pi\left(X_t^{x,T,\pi},T-t\right)\mathrm{d}t\right. \\
&\qquad \left. + \mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^\pi\left(X_r^{x,T,\pi},T-r\right)\mathrm{d}r} g\left(X_{T\wedge\tau_a^b}^{x,T,\pi}\right)\right) \\
&= V^\pi(x,T).
\end{aligned}$$

□

Proof of Theorem 4.2.6. Let $a < a' < x < b' < b$ and $\tau_c^d := \tau_c^d\left(X^{x,T,\pi_{n+1}}\right)$ for any $c < d$. Define the process S by

$$\begin{aligned}
S_t &:= \int_0^t \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} f^{\pi_{n+1}}\left(X_s^{x,T,\pi_{n+1}},T-s\right)\mathrm{d}s \\
&\quad + \mathrm{e}^{-\int_0^t \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} V^{\pi_n}\left(X_t^{x,T,\pi_{n+1}},T-t\right), \quad t \le T.
\end{aligned}$$

Itô's formula, applicable thanks to Proposition 4.2.4, yields, for every small enough

$\epsilon > 0$,

$$S_{(T-\epsilon)\wedge\tau_{a'}^{b'}} = V^{\pi_n}(x,T) + \int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} \cdot \left(f^{\pi_{n+1}} + L^{\pi_{n+1}}V^{\pi_n} - \alpha^{\pi_{n+1}}V^{\pi_n}\right)\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}s + \int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} \sigma_{\pi_{n+1}} V_x^{\pi_n}\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}B_s.$$

The stochastic integral is a true martingale since the functions $\sigma_{\pi_{n+1}}$ and $V_x^{\pi_n}$ are bounded on $[a',b'] \times [\epsilon, T]$ (by Assumption 4.2.1 and Proposition 4.2.4, respectively). Hence we obtain

$$\mathbb{E}\left(S_{(T-\epsilon)\wedge\tau_{a'}^{b'}}\right) = V^{\pi_n}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} \left(f^{\pi_{n+1}} + L^{\pi_{n+1}}V^{\pi_n} - \alpha^{\pi_{n+1}}V^{\pi_n}\right)\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}s\right).$$

By the definition of the policy improvement algorithm (4.4) we get

$$\begin{aligned}
\mathbb{E}\left(S_{(T-\epsilon)\wedge\tau_{a'}^{b'}}\right) &= V^{\pi_n}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} \cdot \min_{p\in A}\left(f^p + L^pV^{\pi_n} - \alpha^pV^{\pi_n}\right)\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}s\right) \\
&\leq V^{\pi_n}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} \cdot \left(f^{\pi_n} + L^{\pi_n}V^{\pi_n} - \alpha^{\pi_n}V^{\pi_n}\right)\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}s\right) \\
&= V^{\pi_n}(x,T),
\end{aligned}$$

where we used Proposition 4.2.4 in the last step. Therefore we obtain

$$V^{\pi_n}(x,T) \geq \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} f^{\pi_{n+1}}\left(X_s^{x,T,\pi_{n+1}}, T-s\right)\mathrm{d}s + \mathrm{e}^{-\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} V^{\pi_n}\left(X_{(T-\epsilon)\wedge\tau_{a'}^{b'}}^{x,T,\pi_{n+1}}, T-((T-\epsilon)\wedge\tau_{a'}^{b'})\right)\right).$$

Since $\tau_{a'}^{b'}$ converges almost surely to τ_a^b when a' tends to a and b' to b, and the functions f and V^{π_n} are bounded, and V^{π_n} is continuous by Lemma 4.2.14, the

Dominated Convergence Theorem yields

$$\lim_{a'\to a}\lim_{b'\to b}\lim_{\epsilon\to 0}\mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} f^{\pi_{n+1}}\left(X_s^{x,T,\pi_{n+1}},T-s\right)\mathrm{d}s\right)$$
$$=\mathbb{E}\left(\int_0^{T\wedge\tau_a^b} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} f^{\pi_{n+1}}\left(X_s^{x,T,\pi_{n+1}},T-s\right)\mathrm{d}s\right)$$

and

$$\begin{aligned}
&\lim_{a'\to a}\lim_{b'\to b}\lim_{\epsilon\to 0}\mathbb{E}\left(\mathrm{e}^{-\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r}\right.\\
&\qquad\left. V^{\pi_n}\left(X_{(T-\epsilon)\wedge\tau_{a'}^{b'}}^{x,T,\pi_{n+1}},T-((T-\epsilon)\wedge\tau_{a'}^{b'})\right)\right)\\
&=\mathbb{E}\left(\mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r}\cdot V^{\pi_n}\left(X_{T\wedge\tau_a^b}^{x,T,\pi_{n+1}},T-(T\wedge\tau_a^b)\right)\right)\\
&=\mathbb{E}\left(\mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} g\left(X_{T\wedge\tau_a^b}^{x,T,\pi_{n+1}}\right)\right),
\end{aligned}$$

where we also used Lemma 4.2.10. Hence we have

$$\begin{aligned}
V^{\pi_n}(x,T)&\geq\mathbb{E}\left(\int_0^{T\wedge\tau_a^b} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} f^{\pi_{n+1}}\left(X_s^{x,T,\pi_{n+1}},T-s\right)\mathrm{d}s\right)\\
&\quad+\mathbb{E}\left(\mathrm{e}^{-\int_0^{T\wedge\tau_a^b} \alpha^{\pi_{n+1}}\left(X_r^{x,T,\pi_{n+1}},T-r\right)\mathrm{d}r} g\left(X_{T\wedge\tau_a^b}^{x,T,\pi_{n+1}}\right)\right)\\
&=V^{\pi_{n+1}}(x,T).
\end{aligned}$$

□

Proof of Proposition 4.2.7. Let the sequences $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$, $\{\epsilon_n\}_{n\in\mathbb{N}}$ and $\{T_n\}_{n\in\mathbb{N}}$ be such that the following hold for every $k\in\mathbb{N}$:

$$a<a_{k+1}<a_k<b_k<b_{k+1}<b,\quad 0<\epsilon_{k+1}<\epsilon_k<T_k<T_{k+1},$$

and

$$\bigcup_{n\in\mathbb{N}}([a_n,b_n]\times[\epsilon_n,T_n])=(a,b)\times(0,\infty).$$

Thanks to Theorem 5 from [9, p. 64] (and Proposition 4.2.4), the sequence $\{V_{xx}^{\pi_n}\}_{n\in\mathbb{N}}$ is uniformly bounded on compacts in $(a,b)\times(0,\infty)$. Recalling the policy improvement algorithm (4.4) and applying Assumption 4.2.2, we obtain that the sequence

$\{\pi_n\}_{n\in\mathbb{N}}$ is uniformly Lipschitz on $[a_k, b_k] \times [\epsilon_k, T_k]$ for every $k \in \mathbb{N}$. Let $\pi_n^0 := \pi_n$ for every $n \in \mathbb{N}$. Due to the Arzela-Ascoli Theorem as stated in Lemma 3.2.13, for every $k \in \mathbb{N}$ there exists a subsequence $\{\pi_n^k\}_{n\in\mathbb{N}} \subseteq \{\pi_n^{k-1}\}_{n\in\mathbb{N}}$ such that $\{\pi_n^k\}_{n\in\mathbb{N}}$ converges uniformly on $[a_k, b_k] \times [\epsilon_k, T_k]$. The diagonal sequence, i.e. $\{\pi_n^n\}_{n\in\mathbb{N}}$, then converges uniformly on $[a_k, b_k]\times[\epsilon_k, T_k]$ for every $k \in \mathbb{N}$, and hence on every compact subset of $(a, b) \times (0, \infty)$. □

Proof of Theorem 4.2.8. Let $\{\pi_{n_k}\}_{k\in\mathbb{N}}$ be a sequence from Proposition 4.2.7 that converges to $\pi_{\lim}$ uniformly on compacts in $(a, b) \times (0, \infty)$. Let $a < a' < x < b' < b$, $k \in \mathbb{N}$ and $\tau_c^d := \tau_c^d\left(X^{x,T,\pi_{\lim}}\right)$ for any $c < d$. Define the process S by

$$\begin{aligned} S_t := &\int_0^t \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} f^{\pi_{n_k}}\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s \\ &+ \mathrm{e}^{-\int_0^t \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} V^{\pi_{n_k}}\left(X_t^{x,T,\pi_{\lim}},T-s\right), \quad t \le T. \end{aligned}$$

Itô's formula yields, for every small enough $\epsilon > 0$,

$$\begin{aligned} S_{(T-\epsilon)\wedge\tau_{a'}^{b'}} = V^{\pi_{n_k}}(x,T) &+ \int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} \\ &\cdot \left(f^{\pi_{n_k}} + L^{\pi_{\lim}}V^{\pi_{n_k}} - \alpha^{\pi_{n_k}}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s \\ &+ \int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} \sigma_{\pi_{\lim}} V_x^{\pi_{n_k}}\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}B_s. \end{aligned}$$

The stochastic integral is a martingale since the functions $\sigma_{\pi_{\lim}}$ and $V_x^{\pi_{n_k}}$ are bounded on the domain of integration. Hence we obtain

$$\begin{aligned} \mathbb{E}\left(S_{(T-\epsilon)\wedge\tau_{a'}^{b'}}\right) &= V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r}\right. \\ &\quad \left.\cdot \left(f^{\pi_{n_k}} + L^{\pi_{\lim}}V^{\pi_{n_k}} - \alpha^{\pi_{n_k}}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s\right) \\ &= V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r}\right. \\ &\quad \left.\cdot \left(L^{\pi_{\lim}}V^{\pi_{n_k}} - L^{\pi_{n_k}}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s\right) \\ &= V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{n_k}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r}\right. \\ &\quad \left.\cdot \left(\frac{1}{2}\left(\sigma_{\pi_{\lim}}^2 - \sigma_{\pi_{n_k}}^2\right)V_{xx}^{\pi_{n_k}} + \left(\mu_{\pi_{\lim}} - \mu_{\pi_{n_k}}\right)V_x^{\pi_{n_k}}\right)\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s\right), \end{aligned}$$

where we used Proposition 4.2.4 and the definition of the operator L. Due to Theo-

rem 5 in [9, p. 64] (and Proposition 4.2.4), the sequences $\{V_{xx}^{\pi_{n_k}}\}_{k\in\mathbb{N}}$ and $\{V_x^{\pi_{n_k}}\}_{k\in\mathbb{N}}$ are uniformly bounded on $[a',b']\times[\epsilon,T]$. Now the Dominated Convergence Theorem, when k tends to ∞, yields that the last term disappears. Hence we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\pi_{\lim}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} f^{\pi_{\lim}}\left(X_s^{x,T,\pi_{\lim}},T-s\right)\mathrm{d}s\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \alpha^{\pi_{\lim}}\left(X_r^{x,T,\pi_{\lim}},T-r\right)\mathrm{d}r} V^{\lim}\left(X_{(T-\epsilon)\wedge\tau_{a'}^{b'}}^{x,T,\pi_{\lim}},T-((T-\epsilon)\wedge\tau_{a'}^{b'})\right)\right)\\
&= V^{\lim}(x,T).
\end{aligned}$$

By sending ϵ to 0, b' to b and a' to a, we obtain the desired equality as in the previous proof. □

Proof of Theorem 4.2.9. The second assertion follows from Theorem 4.2.8.

Consider the sequence $\{(\pi_{n+1},\pi_n):(a,b)\times(0,\infty)\to A\times A\}_{n\in\mathbb{N}}$, where $A\times A$ is equipped with any p-product metric, $p\in[1,\infty]$. In the same way as in the proof of Proposition 4.2.7 we can find a subsequence $\{(\pi_{1+n_k},\pi_{n_k})\}_{k\in\mathbb{N}}$ that is uniformly convergent on every compact set in $(a,b)\times(0,\infty)$.

For every $(y,t)\in(a,b)\times(0,\infty)$ and $k\in\mathbb{N}$, set

$$\hat{\pi}_\infty(y,t):=\lim_{l\to\infty}\pi_{n_l}(y,t),\quad \tilde{\pi}_\infty(y,t):=\lim_{l\to\infty}\pi_{n_l+1}(y,t),$$

$$\hat{\sigma}_k(y,t):=\sigma(y,t,\pi_{n_k}(y,t)),\quad \tilde{\sigma}_k(y,t):=\sigma(y,t,\pi_{n_k+1}(y,t)),$$

$$\hat{\sigma}_\infty(y,t):=\sigma(y,t,\hat{\pi}_\infty(y,t)),\quad \tilde{\sigma}_\infty(y,t):=\sigma(y,t,\tilde{\pi}_\infty(y,t)).$$

Define $\hat{\mu}_k$, $\hat{\alpha}_k$, $\hat{f}_k$, and $\tilde{\mu}_k$, $\tilde{\alpha}_k$, $\tilde{f}_k$, and $\hat{\mu}_\infty$, $\hat{\alpha}_\infty$, $\hat{f}_\infty$, and $\tilde{\mu}_\infty$, $\tilde{\alpha}_\infty$, $\tilde{f}_\infty$, in a corresponding fashion. Let

$$\hat{u}_k(y,t):=V^{\pi_{n_k}}(y,t),\quad \tilde{u}_k(y,t):=V^{\pi_{n_k+1}}(y,t)\quad\text{and}\quad u(y):=V^{\lim}(y),$$

and define the operator $\hat{\mathcal{L}}^k$ by

$$\hat{\mathcal{L}}^k h:=\frac{1}{2}\hat{\sigma}_k^2 h_{xx}+\hat{\mu}_k h_x-h_t-\hat{\alpha}_k h+\hat{f}_k,\quad h\in\mathcal{C}^{2,1}((a,b)\times(0,\infty)),$$

with the corresponding definitions for $\tilde{\mathcal{L}}^k$, $\hat{\mathcal{L}}^\infty$ and $\tilde{\mathcal{L}}^\infty$. Applying Theorem 15 from [9, p. 80] (and Proposition 4.2.4), we obtain that both $\hat{\mathcal{L}}^\infty u=0$ and $\tilde{\mathcal{L}}^\infty u=0$ hold on every compact subset of $(a,b)\times(0,\infty)$, and that the sequence of functions $\{\frac{1}{2}(\tilde{\sigma}_k^2-\hat{\sigma}_k^2)\cdot(\hat{u}_k)_{xx}+(\tilde{\mu}_k-\hat{\mu}_k)\cdot(\hat{u}_k)_x-(\tilde{\alpha}_k-\hat{\alpha}_k)\hat{u}_k+\tilde{f}_k-\hat{f}_k\}_{k\in\mathbb{N}}$ converges to

the function $\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)u_{xx} + (\tilde{\mu}_\infty - \hat{\mu}_\infty)u_x - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)u + \tilde{f}_\infty - \hat{f}_\infty$ on on every compact subset of $(a,b)\times(0,\infty)$. However, we know that

$$\frac{1}{2}(\tilde{\sigma}_\infty^2 - \hat{\sigma}_\infty^2)u_{xx} + (\tilde{\mu}_\infty - \hat{\mu}_\infty)u_x - (\tilde{\alpha}_\infty - \hat{\alpha}_\infty)u + \tilde{f}_\infty - \hat{f}_\infty = \tilde{\mathcal{L}}^\infty u - \hat{\mathcal{L}}^\infty u = 0. \quad (4.6)$$

Let $k \in \mathbb{N}$, $\epsilon > 0$ and $a < a' < x < b' < b$. Define $\tau_c^d := \tau_c^d\left(X^{x,T,\Pi}\right)$ for any $c < d$. We obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r} f^{\Pi_s}\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right.\\
&\qquad \left.+\mathrm{e}^{-\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r} V^{\pi_{n_k}}\left(X^{x,T,\Pi}_{(T-\epsilon)\wedge\tau_{a'}^{b'}}, T-((T-\epsilon)\wedge\tau_{a'}^{b'})\right)\right)\\
&\overset{\text{Itô}}{=} V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r}\right.\\
&\qquad \left.\cdot\left(f^{\Pi_s} + L^{\Pi_s}V^{\pi_{n_k}} - \alpha^{\Pi_s}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right)\\
&\geq V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r}\right.\\
&\qquad \left.\cdot\left(\min_{p\in A}\left(f^p + L^pV^{\pi_{n_k}} - \alpha^pV^{\pi_{n_k}}\right)\right)\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right)\\
&\overset{\text{PIA (4.4)}}{=} V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r}\right.\\
&\qquad \left.\cdot\left(f^{\pi_{n_k+1}} + L^{\pi_{n_k+1}}V^{\pi_{n_k}} - \alpha^{\pi_{n_k+1}}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right)\\
&\overset{\text{Pr. 4.2.4}}{=} V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r}\left(f^{\pi_{n_k+1}} - f^{\pi_{n_k}}\right.\right.\\
&\qquad \left.\left.+L^{\pi_{n_k+1}}V^{\pi_{n_k}} - L^{\pi_{n_k}}V^{\pi_{n_k}} - \alpha^{\pi_{n_k+1}}V^{\pi_{n_k}} + \alpha^{\pi_{n_k}}V^{\pi_{n_k}}\right)\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right)\\
&= V^{\pi_{n_k}}(x,T) + \mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r}\left(\frac{1}{2}(\tilde{\sigma}_k^2 - \hat{\sigma}_k^2)\cdot(\hat{u}_k)_{xx}\right.\right.\\
&\qquad \left.\left.+(\tilde{\mu}_k - \hat{\mu}_k)\cdot(\hat{u}_k)_x - (\tilde{\alpha}_k - \hat{\alpha}_k)\hat{u}_k + \tilde{f}_k - \hat{f}_k\right)\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right).
\end{aligned}$$

Sending k to ∞ and applying the Dominated Convergence Theorem, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \mathrm{e}^{-\int_0^s \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r} f^{\Pi_s}\left(X_s^{x,T,\Pi},T-s\right)\mathrm{d}s\right.\\
&\qquad\left.+\mathrm{e}^{-\int_0^{(T-\epsilon)\wedge\tau_{a'}^{b'}} \alpha^{\Pi_r}\left(X_r^{x,T,\Pi},T-r\right)\mathrm{d}r} V^{\lim}\left(X^{x,T,\Pi}_{(T-\epsilon)\wedge\tau_{a'}^{b'}}, T-((T-\epsilon)\wedge\tau_{a'}^{b'})\right)\right)\\
&\geq V^{\lim}(x,T)
\end{aligned}$$

by using the first part of the proof. After sending ϵ to 0, b' to b and a' to a, the desired inequality follows in the usual way. □

4.3 Application to the finite horizon problem for geometric Brownian motions

4.3.1 Approximation of the value function

Recall the setting of Problem (T+) from Section 2.4. We have

$$\tau(V) := \inf\{t \geq 0;\ X_t = Y(V)_t\} = \inf\left\{t \geq 0;\ \log\left(\frac{X_t}{Y(V)_t}\right) = 0\right\} \quad (\inf\emptyset := \infty).$$

For any $T > 0$, we would like to solve the following problem:

$$\text{find} \quad \inf_{V\in\mathcal{V}} \mathbb{P}_{x,y}\left(\tau(V) > T\right) =: \tilde{U}(x,y,T).$$

First we will try to accommodate the problem within the setting of the present chapter. For every $V \in \mathcal{V}$, Lemma 2.2.1 yields $W \in \mathcal{V}$ and a process $\Pi = (\Pi_t)_{t\geq 0}$ such that the following hold: B and W are independent, Π is $(\mathcal{F}_t)_{t\geq 0}$-adapted and takes values in $[-1,1]$, and

$$V_t = \int_0^t \Pi_s\,\mathrm{d}B_s + \int_0^t \sqrt{1-\Pi_s^2}\,\mathrm{d}W_s, \quad t \geq 0.$$

Recall $\mu := a_2 - a_1 + \frac{\sigma_1^2}{2} - \frac{\sigma_2^2}{2}$. Using the explicit formula for the geometric Brownian motion (or its stochastic differential equation and Itô's lemma) and this representa-

tion, we obtain

$$\begin{aligned}\log\left(\frac{X_t}{Y(V)_t}\right) &= \log\left(\frac{x}{y}\right) + \sigma_1 B_t + a_1 t - (\sigma_2 V_t + a_2 t) - \frac{\sigma_1^2}{2}\,t + \frac{\sigma_2^2}{2}\,t \\ &= \log\left(\frac{x}{y}\right) - \mu t + \int_0^t (\sigma_1 - \sigma_2\Pi_s)\,\mathrm{d}B_s + \sigma_2 \int_0^t \sqrt{1-\Pi_s^2}\,\mathrm{d}W_s \\ &= \log\left(\frac{x}{y}\right) - \mu t + \int_0^t \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\Pi_s}\,\mathrm{d}B_s^{\Pi}, \quad t \geq 0,\end{aligned}$$

where

$$B_t^{\Pi} := \int_0^t \frac{\sigma_1 - \sigma_2\Pi_s}{\left(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\Pi_s\right)^{\frac{1}{2}}}\,\mathrm{d}B_s + \int_0^t \frac{\sigma_2\left(1-\Pi_s^2\right)^{\frac{1}{2}}}{\left(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\Pi_s\right)^{\frac{1}{2}}}\,\mathrm{d}W_s, \quad t \geq 0.$$

Note that B^{Π} is an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion by Lévy's characterisation theorem.

We know that we can assume $x > y$ and $\sigma_1 > \sigma_2 > 0$ without loss of generality. Let

$$\begin{aligned}\mathcal{A} := \{\Pi = (\Pi_t)_{t<T};\ &\Pi \text{ adapted to } (\mathcal{F}_t)_{t<T}, \\ &\text{and } \Pi_t(\omega) \in [-1,1] \text{ for every } t \geq 0 \text{ and } \omega \in \Omega\}\end{aligned}$$

and for any $x > 0$, $T > 0$ and $\Pi \in \mathcal{A}$ define the controlled process by

$$X_t^{x,T,\Pi} := x - \mu t + \int_0^t \left(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\Pi_s\right)^{\frac{1}{2}}\mathrm{d}B_s, \quad t \leq T,$$

and the payoff by

$$U^{\Pi}(x,T) := \mathbb{P}\left(\tau_0^{\infty}\left(X^{x,T,\Pi}\right) > T\right).$$

The problem is the following:

$$\text{find} \quad \inf_{\Pi\in\mathcal{A}} U^{\Pi}(x,T) =: U(x,T). \tag{4.7}$$

From the previous paragraph it is clear that

$$\tilde{U}(x,y,T) = U\left(\log\left(\frac{x}{y}\right), T\right),$$

hence by solving Problem (4.7) the original problem will be solved, too.

We note

$$U^{\Pi}(x,T) = \mathbb{E}\left(\mathbb{I}_{(0,\infty)}\left(X^{x,T,\Pi}_{T\wedge\tau_0^{\infty}(X^{x,T,\Pi})}\right)\right).$$

We would like to apply our policy improvement algorithm to find the function U, but there are two obstacles: the function $\mathbb{I}_{(0,\infty)}$ is not continuous and hence Assumption 4.2.1 is not satisfied, and $I^{h_n}(x,t) = \operatorname{sgn}((h_n)_x(x,t))$, which is also far from satisfying the demands of Assumption 4.2.2. We will rectify this by constructing similar problems in such a way that we will be able to use the policy improvement algorithm and that they will approximate (in the limit) Problem (4.7).

Let $\epsilon > 0$, and define the following functions:

$$\alpha_{(\epsilon)}(x,t,p) := 0, \quad f_{(\epsilon)}(x,t,p) := \epsilon p^2, \quad g_{(\epsilon)}(x) := \begin{cases} 0 & \text{if } x \le 0, \\ \frac{x}{\epsilon} & \text{if } 0 < x < \epsilon, \\ 1 & \text{if } x \ge \epsilon, \end{cases}$$

$$V^{\Pi}_{(\epsilon)}(x,T) := \mathbb{E}\left(\int_0^{T \wedge \tau_0^\infty\left(X^{x,T,\Pi}\right)} f_{(\epsilon)}\left(X_t^{x,T,\Pi}, T-t, \Pi_t\right) \mathrm{d}t + g_{(\epsilon)}\left(X^{x,T,\Pi}_{T\wedge\tau_0^\infty(X^{x,T,\Pi})}\right)\right),$$

$$V_{(\epsilon)}(x,T) := \inf_{\Pi \in \mathcal{A}} V^{\Pi}_{(\epsilon)}(x,T).$$

Now Assumption 4.2.1 is clearly satisfied, and the same holds for Assumption 4.2.2 since the following function possesses the required Hölder property:

$$I^{h_n}(x,t) = \begin{cases} -1 & \text{if } \frac{\sigma_1\sigma_2(h_n)_x(x,t)}{2\epsilon} \le -1, \\ \frac{\sigma_1\sigma_2(h_n)_x(x,t)}{2\epsilon} & \text{if } -1 < \frac{\sigma_1\sigma_2(h_n)_x(x,t)}{2\epsilon} < 1, \\ 1 & \text{if } \frac{\sigma_1\sigma_2(h_n)_x(x,t)}{2\epsilon} \ge 1. \end{cases}$$

Hence the function $V_{(\epsilon)}$ can be obtained via the policy improvement algorithm for every $\epsilon > 0$. The following theorem says that the value function U can be approximated by these functions, which solves Problem (4.7) and thus the motivating problem of optimal coupling of geometric Brownian motions.

Theorem 4.3.1. *For every $x > 0$ and $T > 0$, the following holds:*

$$U(x,T) = \lim_{\epsilon \to 0} V_{(\epsilon)}(x,T).$$

4.3.2 Proof

First we will look at some properties of the functions U^Π and U.

Lemma 4.3.2. *For every $T > 0$ there exists $C > 0$ such that the following holds for every $\Pi \in \mathcal{A}$ and $x, y > 0$:*

$$\left|U^\Pi(x,T) - U^\Pi(y,T)\right| \le C|x-y|.$$

Proof. For easier notation, define $\sigma(p) := (\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 p)^{\frac{1}{2}}$. Without loss of generality we can assume that $x > y$. Define the processes X^x and X^y in the following way:

$$X_t^x := x - \mu t + \int_0^t \sigma(\Pi_s)\,\mathrm{d}B_s, \quad X_t^y := y - \mu t - \int_0^t \sigma(\Pi_s)\,\mathrm{d}B_s, \quad t \le T.$$

Let ρ be their first meeting time, i.e.

$$\rho := \inf\{t \in [0,T];\ X_t^x = X_t^y\} = \inf\left\{t \in [0,T];\ x - y = -2\int_0^t \sigma(\Pi_s)\,\mathrm{d}B_s\right\}.$$

We will construct a process that has the same law as X^x and stays above or is equal to X^y the entire time. Define $X = (X_t)_{t\in[0,T]}$ and $\tilde{X} = (\tilde{X}_t)_{t\in[0,T]}$ by

$$X_t := X_t^x \mathbb{I}_{\{t<\rho\}} + X_t^y \mathbb{I}_{\{t\ge\rho\}} \quad \text{and} \quad \tilde{X}_t := x - \mu t + \int_0^t \sigma(\Pi_s)\left(\mathbb{I}_{\{s<\rho\}} - \mathbb{I}_{\{s\ge\rho\}}\right)\mathrm{d}B_s.$$

It is obvious that $X_t \ge X_t^y$ holds for all $t \le T$. By looking separately at the events $\{t < \rho\}$ and $\{t \ge \rho\}$, it is also easy to see that $X = \tilde{X}$. Finally, $\tilde{X}$ has the same law as X^x since the process $\left(\int_0^t \left(\mathbb{I}_{\{s<\rho\}} - \mathbb{I}_{\{s\ge\rho\}}\right)\mathrm{d}B_s\right)_{t\ge0}$ is a Brownian motion by Lévy's characterisation theorem. We now obtain the following estimate:

$$\begin{aligned}&\left|U^\Pi(x,T) - U^\Pi(y,T)\right| = \mathbb{P}\left(\tau_0^\infty(X) > T\right) - \mathbb{P}\left(\tau_0^\infty(X^y) > T\right)\\ &= \mathbb{P}\left(\tau_0^\infty(X) > T, \tau_0(X^y) \le T\right) \le \mathbb{P}\left(\rho > T\right).\end{aligned}$$

Due to Dambis-Dubins-Schwarz theorem, there exists a Brownian motion W such that

$$N_t := -2\int_0^t \sigma(\Pi_s)\,\mathrm{d}B_s = W_{[N]_t}$$

holds for every $t \le T$. Since σ is bounded away from 0, there exists $m > 0$ such that $[N]_t \ge mt$ holds for every $t \le T$. Using all these and the Reflection Principle, we obtain

$$\begin{aligned}&\mathbb{P}(\rho > T) = \mathbb{P}\left(\sup_{t\le T} N_t < x - y\right) = \mathbb{P}\left(\sup_{t\le T} W_{[N]_t} < x - y\right) \le \mathbb{P}\left(\sup_{t\le T} W_{mt} < x - y\right)\\ &= \mathbb{P}\left(|W_T| < \frac{x-y}{\sqrt{m}}\right) \le \int_{-\frac{x-y}{\sqrt{m}}}^{\frac{x-y}{\sqrt{m}}} \frac{1}{\sqrt{2\pi T}}\,\mathrm{d}y = C(x-y),\end{aligned}$$

where the constant C does not depend on x, y or Π. □

Lemma 4.3.3. *For every $T > 0$, the function $U(\cdot, T)$ is continuous.*

Proof. Thanks to Lemma 4.3.2 we obtain

$$|U(x,T) - U(y,T)| = \left| \inf_{\Pi \in \mathcal{A}} U^{\Pi}(x,T) - \inf_{\Pi \in \mathcal{A}} U^{\Pi}(y,T) \right| \leq \sup_{\Pi \in \mathcal{A}} \left| U^{\Pi}(x,T) - U^{\Pi}(y,T) \right|$$
$$\leq \sup_{\Pi \in \mathcal{A}} C|x-y| = C|x-y|,$$

which proves that $U(\cdot, T)$ is even Lipschitz. □

Proof of Theorem 4.3.1. First we will prove $U(x,T) \geq \limsup_{\epsilon \to 0} V_{(\epsilon)}(x,T)$. To this end, note that $V_{(\epsilon)}(x,T) \leq V^{\Pi}_{(\epsilon)}(x,T)$ holds for every $\Pi \in \mathcal{A}$ and $\epsilon > 0$. By the Dominated Convergence Theorem, $V^{\Pi}_{(\epsilon)}(x,T)$ converges to $U^{\Pi}(x,T)$ as ϵ tends to 0. Hence we obtain

$$\limsup_{\epsilon \to 0} V_{(\epsilon)}(x,T) \leq \limsup_{\epsilon \to 0} V^{\Pi}_{(\epsilon)}(x,T) = U^{\Pi}(x,T)$$

for every $\Pi \in \mathcal{A}$. Thus we can deduce the following:

$$\limsup_{\epsilon \to 0} V_{(\epsilon)}(x,T) \leq \inf_{\Pi \in \mathcal{A}} U^{\Pi}(x,T) = U(x,T).$$

Now we will prove the other inequality, i.e. $U(x,T) \leq \liminf_{\epsilon \to 0} V_{(\epsilon)}(x,T)$. We note that the following holds for every $\Pi \in \mathcal{A}$ and small enough $\epsilon > 0$:

$$V^{\Pi}_{(\epsilon)}(x,T) \geq \mathbb{E}\left(g_{(\epsilon)}\left(X^{x,T,\Pi}_{T \wedge \tau_0^{\infty}(X^{x,T,\Pi})} \right) \right) \geq \mathbb{E}\left(\mathbb{I}_{(\epsilon,\infty)}\left(X^{x,T,\Pi}_{T \wedge \tau_0^{\infty}(X^{x,T,\Pi})} \right) \right)$$
$$\geq \mathbb{P}\left(\tau_{\epsilon}^{\infty}\left(X^{x,T,\Pi} \right) > T \right) = \mathbb{P}\left(\tau_0^{\infty}\left(X^{x-\epsilon,T,\Pi} \right) > T \right) = U^{\Pi}(x-\epsilon, T).$$

By taking the infimum on both sides we arrive to $V_{(\epsilon)}(x,T) \geq U(x-\epsilon, T)$. By Lemma 4.3.3, the function $U(\cdot, T)$ is continuous, which brings us to

$$\liminf_{\epsilon \to 0} V_{(\epsilon)}(x,T) \geq \liminf_{\epsilon \to 0} U(x-\epsilon, T) = U(x,T).$$

□

4.4 Conclusion

In Section 4.2 we developed the policy improvement algorithm for the finite horizon problem. Compared to the (one-dimensional) infinite horizon problem, the algorithms are conceptually very similar. Nevertheless, there are some important

technical differences. In the finite time horizon case we additionally have time dependence, which naturally leads to parabolic differential equations. Because of this we had to prove continuity of the payoff function, which was not required in the one-dimensional infinite horizon case. The role of discounting is also different. Whereas it is essential in the infinite horizon case (α has to be positive, even bounded away from 0), α can be 0 in the finite horizon problem. We would lose some generality if we did this, but not all since constant α would still be covered by the time-dependent cost function f.

Section 4.3 is the culmination of the thesis because we applied there our new method to the problem that we had originally wanted to solve, and because it brings together the two leading topics of the thesis, i.e. the coupling of geometric Brownian motions and the policy improvement algorithm in a continuous setting. The application was not straightforward, though, due to the lack of smoothness. We constructed a sequence of approximating smooth data and proved the convergence.

The reason why we have not dealt with the policy improvement algorithm for the multidimensional finite-horizon problem is two-fold: having developed the algorithm for three related problems, it should be quite clear how to attempt to do it; and it has never been our intention to try to present the complete or final treatise of it. On the contrary, this was only the first step, and it is the author's wish that many more will follow, both by those who would like to reach the theoretical boundaries of the algorithm, and by those with a more or less concrete application in mind.

Bibliography

[1] R. Atar and K. Burdzy, Mirror Couplings and Neumann Eigenfunctions. *Journal of the American Mathematical Society*, 17(2):243–265, 2004.

[2] S. Banerjee and W. S. Kendall, Rigidity for Markovian Maximal Couplings of Elliptic Diffusions. ArXiv:1412.2647 [math.PR], 2014.

[3] M. T. Barlow and S. D. Jacka, Tracking a Diffusion, and an Application to Weak Convergence. *Advances in Applied Probability*, 18:15–25, 1986.

[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] K. Burdzy and W. S. Kendall, Efficient Markovian Couplings: Examples and Counterexamples. *The Annals of Probability*, 10(2):362–409, 2000.

[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] B. Doya, Reinforcement Learning in Continuous Time and Space. *Neural Computation*, 12(1):219–245, 2000.

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

[10] 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.

[11] 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.

[12] R. A. Howard, *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, 1960.

[13] E. P. Hsu and K. T. Sturm, Maximal Coupling of Euclidean Brownian Motions. *Communications in Mathematics and Statistics*, 1(1):93–104, 2013.

[14] S. D. Jacka, *Dynamic Stochastic Control*. University of Warwick (lecture notes), 2012. Available at http://www2.warwick.ac.uk/fac/sci/statistics/modules/st4/st411/resources/dscnotes5.pdf.

[15] S. D. Jacka and A. Mijatović, Coupling and Tracking of Regime-Switching Martingales. *Electronic Journal of Probability*, 20:1–39, 2015.

[16] S. D. Jacka, A. Mijatović and D. Širaj, Mirror and Synchronous Couplings of Geometric Brownian Motions. *Stochastic Processes and their Applications*, 123(2):1055–1069, 2014.

[17] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, second edition, 1991.

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

[19] K. Kuwada, Characterization of Maximal Markovian Couplings for Diffusion Processes. *Electronic Journal of Probability*, 14:633–662, 2009.

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

[21] 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.

[22] T. Lindvall, *Lectures on the Coupling Method*. Dover Publications, New York, 2002.

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

[24] S. Mahadevan, Representation Policy Iteration. ArXiv:1207.1408 [cs.AI], 2012.

[25] 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.

[26] M. N. Pascu, Mirror Coupling of Reflecting Brownian Motion and an Application to Chavel's Conjecture. *Electronic Journal of Probability*, 16:505–530, 2011.

[27] S. M. Ross, *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, 1983.

[28] 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.

[29] M. S. Santos and J. Rust, Convergence Properties of Policy Iteration. *SIAM Journal on Control and Optimization*, 42(6):2094–2115, 2004.

[30] H. Thorisson, *Coupling, Stationarity, and Regeneration*. Springer-Verlag, New York, 2000.

[31] K. G. Vamvoudakis and F. L. Lewis, Online Actor–Critic Algorithm to Solve the Continuous-Time Infinite Horizon Optimal Control Problem. *Automatica*, 46(5):878–888, 2010.

[32] 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.