On the policy iteration algorithm for non-degenerate controlled diffusions under the ergodic criterion

Ari Arapostathis

Dedicated to Onésimo Hernández-Lerma on the occasion of his 65th birthday.

Abstract The subject of this paper is the policy iteration algorithm for non-degenerate controlled diffusions. The results parallel the ones in Meyn [11] for discrete-time controlled Markov chains. The model in [11] uses norm-like running costs, while we opt for the milder assumption of near-monotone costs. Also, instead of employing a blanket Lyapunov stability hypothesis, we provide a characterization of the region of attraction of the optimal control.

Key words: policy iteration, controlled diffusions, Markov processes, ergodic control

1 Introduction

The policy iteration algorithm (PIA) for controlled Markov chains has been known since the fundamental work of Howard [2]. For controlled Markov chains on Borel state spaces most studies of the PIA rely on blanket Lyapunov conditions [9]. A study of the PIA that treats the model of near-monotone costs can be found in [11], some ideas of which we follow closely. An analysis of the PIA for piecewise deterministic Markov processes has appeared in [6].

In this paper we study the PIA for controlled diffusion processes $X = \{X_t,\ t \ge 0\}$ taking values in the d-dimensional Euclidean space $\mathbb{R}^d$, and governed by the Itô stochastic differential equation

$$\mathrm{d}X_t = b(X_t, U_t)\,\mathrm{d}t + \boldsymbol{\sigma}(X_t)\,\mathrm{d}W_t\,. \tag{1}$$

Ari Arapostathis
Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, e-mail: ari@mail.utexas.edu.

All random processes in (1) live in a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. The process W is a d-dimensional standard Wiener process independent of the initial condition X_0. The control process U takes values in a compact, metrizable set $\mathbb{U}$, and $U_t(\omega)$ is jointly measurable in $(t,\omega) \in [0,\infty) \times \Omega$. Moreover, it is *non-anticipative*: for $s < t$, $W_t - W_s$ is independent of

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

Such a process U is called an *admissible control*, and we let $\mathfrak{U}$ denote the set of all admissible controls.

We impose the following standard assumptions on the drift b and the diffusion matrix σ to guarantee existence and uniqueness of solutions to (1).

(A1) *Local Lipschitz continuity:* The functions

$$b = \bigl[b^1, \dots, b^d\bigr]^{\mathsf{T}} : \mathbb{R}^d \times \mathbb{U} \mapsto \mathbb{R}^d \quad \text{and} \quad \sigma = \bigl[\sigma^{ij}\bigr] : \mathbb{R}^d \mapsto \mathbb{R}^{d\times d}$$

are locally Lipschitz in x with a Lipschitz constant K_R depending on $R > 0$. In other words, if B_R denotes the open ball of radius R centered at the origin in $\mathbb{R}^d$, then for all $x, y \in B_R$ and $u \in \mathbb{U}$,

$$|b(x,u) - b(y,u)| + \|\sigma(x) - \sigma(y)\| \le K_R |x-y|\,,$$

where $\|\sigma\|^2 := \operatorname{trace}\bigl(\sigma\sigma^{\mathsf{T}}\bigr)$.

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

$$|b(x,u)|^2 + \|\sigma(x)\|^2 \le K_1\bigl(1 + |x|^2\bigr)\,, \quad \forall (x,u) \in \mathbb{R}^d \times \mathbb{U}\,.$$

(A3) *Local non-degeneracy:* For each $R > 0$, there exists a positive constant κ_R such that

$$\sum_{i,j=1}^{d} a^{ij}(x)\xi_i\xi_j \ge \kappa_R |\xi|^2\,, \quad \forall x \in B_R\,,$$

for all $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$, where $a := \frac{1}{2}\sigma\,\sigma^{\mathsf{T}}$.

We also assume that b is continuous in (x,u).

In integral form, (1) is written as

$$X_t = X_0 + \int_0^t b(X_s, U_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s\,. \tag{2}$$

The second term on the right hand side of (2) is an Itô stochastic integral. We say that a process $X = \{X_t(\omega)\}$ is a solution of (1), if it is $\mathfrak{F}_t$-adapted, continuous in t, defined for all $\omega \in \Omega$ and $t \in [0,\infty)$, and satisfies (2) for all $t \in [0,\infty)$ at once a.s.

With $u \in \mathbb{U}$ treated as a parameter, we define the family of operators $L^u : \mathcal{C}^2(\mathbb{R}^d) \mapsto \mathcal{C}(\mathbb{R}^d)$ by

$$L^u f(x) = \sum_{i,j} a^{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_i b^i(x,u) \frac{\partial f}{\partial x_i}(x)\,, \quad u \in \mathbb{U}\,. \tag{3}$$

We refer to L^u as the *controlled extended generator* of the diffusion.

Of fundamental importance in the study of functionals of X is Itô's formula. For $f \in \mathcal{C}^2(\mathbb{R}^d)$ and with L^u as defined in (3),

$$f(X_t) = f(X_0) + \int_0^t L^{U_s} f(X_s)\,\mathrm{d}s + M_t\,, \quad \text{a.s.}, \tag{4}$$

where

$$M_t := \int_0^t \big\langle \nabla f(X_s), \boldsymbol{\sigma}(X_s)\,\mathrm{d}W_s \big\rangle$$

is a local martingale. Krylov's extension of the Itô formula [10, p. 122] extends (4) to functions f in the Sobolev space $\mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$.

Recall that a control is called *stationary Markov* if $U_t = v(X_t)$ for a measurable map $v : \mathbb{R}^d \mapsto \mathbb{U}$. Correspondingly, the equation

$$X_t = x_0 + \int_0^t b\big(X_s, v(X_s)\big)\,\mathrm{d}s + \int_0^t \boldsymbol{\sigma}(X_s)\,\mathrm{d}W_s \tag{5}$$

is said to have a *strong solution* if given a Wiener process $(W_t, \mathfrak{F}_t)$ on a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, there exists a process X on $(\Omega, \mathfrak{F}, \mathbb{P})$, with $X_0 = x_0 \in \mathbb{R}^d$, which is continuous, $\mathfrak{F}_t$-adapted, and satisfies (5) for all t at once, a.s. A strong solution is called *unique*, if any two such solutions X and X' agree $\mathbb{P}$-a.s., when viewed as elements of $\mathcal{C}\big([0,\infty), \mathbb{R}^d\big)$. It is well known that under Assumptions A1–A3, for any stationary Markov control v, (5) has a unique strong solution [8].

Let $\mathfrak{U}_{\mathrm{SM}}$ denote the set of stationary Markov controls. Under $v \in \mathfrak{U}_{\mathrm{SM}}$, the process X is strong Markov, and we denote its transition function by $P^v(t,x,\cdot)$. It also follows from the work of [5, 12] that under $v \in \mathfrak{U}_{\mathrm{SM}}$, the transition probabilities of X have densities which are locally Hölder continuous. Thus L^v defined by

$$L^v f(x) = \sum_{i,j} a^{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_i b^i(x, v(x)) \frac{\partial f}{\partial x_i}(x)\,, \quad v \in \mathfrak{U}_{\mathrm{SM}}\,,$$

for $f \in \mathcal{C}^2(\mathbb{R}^d)$, is the generator of a strongly-continuous semigroup on $\mathcal{C}_b(\mathbb{R}^d)$, which is strong Feller. We let $\mathbb{P}^v_x$ denote the probability measure and $\mathbb{E}^v_x$ the expectation operator on the canonical space of the process under the control $v \in \mathfrak{U}_{\mathrm{SM}}$, conditioned on the process X starting from $x \in \mathbb{R}^d$ at $t = 0$.

In Section 2 we define our notation. Section 3 reviews the ergodic control problem for near-monotone costs and the basic properties of the PIA. Section 4 is dedicated to the convergence of the algorithm.

2 Notation

The standard Euclidean norm in $\mathbb{R}^d$ is denoted by $|\cdot|$, and $\langle\cdot,\,\cdot\rangle$ stands for the inner product. The set of non-negative real numbers is denoted by $\mathbb{R}_+$, $\mathbb{N}$ stands for the set of natural numbers, and $\mathbb{I}$ denotes the indicator function. We denote by $\uptau(A)$ the *first exit time* of the process $\{X_t\}$ from the set $A \subset \mathbb{R}^d$, defined by

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

The open ball of radius R in $\mathbb{R}^d$, centered at the origin, is denoted by B_R, and we let $\uptau_R := \uptau(B_R)$, and $\check{\uptau}_R := \uptau(B_R^c)$.

The term *domain* in $\mathbb{R}^d$ refers to a nonempty, connected open subset of the Euclidean space $\mathbb{R}^d$. We introduce the following notation for spaces of real-valued functions on a domain $D \subset \mathbb{R}^d$. The space $\mathcal{L}^p(D)$, $p \in [1,\infty)$, stands for the Banach space of (equivalence classes) of measurable functions f satisfying $\int_D |f(x)|^p\,\mathrm{d}x < \infty$, and $\mathcal{L}^\infty(D)$ is the Banach space of functions that are essentially bounded in D. The space $\mathcal{C}^k(D)$ ($\mathcal{C}^\infty(D)$) refers to the class of all functions whose partial derivatives up to order k (of any order) exist and are continuous, and $\mathcal{C}^k_c(D)$ is the space of functions in $\mathcal{C}^k(D)$ with compact support. The standard Sobolev space of functions on D whose generalized derivatives up to order k are in $\mathcal{L}^p(D)$, equipped with its natural norm, is denoted by $\mathcal{W}^{k,p}(D)$, $k \ge 0$, $p \ge 1$.

In general if $\mathcal{X}$ is a space of real-valued functions on D, $\mathcal{X}_{\mathrm{loc}}$ consists of all functions f such that $f\varphi \in \mathcal{X}$ for every $\varphi \in \mathcal{C}^\infty_c(D)$. In this manner we obtain the spaces $\mathcal{L}^p_{\mathrm{loc}}(D)$ and $\mathcal{W}^{2,p}_{\mathrm{loc}}(D)$.

Let $h \in \mathcal{C}(\mathbb{R}^d)$ be a positive function. We denote by $\mathcal{O}(h)$ the set of functions $f \in \mathcal{C}(\mathbb{R}^d)$ having the property

$$\limsup_{|x|\to\infty} \frac{|f(x)|}{h(x)} < \infty\,, \tag{6}$$

and by $\mathfrak{o}(h)$ the subset of $\mathcal{O}(h)$ over which the limit in (6) is zero.

We adopt the notation $\partial_i := \frac{\partial}{\partial x_i}$ and $\partial_{ij} := \frac{\partial^2}{\partial x_i \partial x_j}$. We often use the standard summation rule that repeated subscripts and superscripts are summed from 1 through d. For example,

$$a^{ij}\partial_{ij}\varphi + b^i\partial_i\varphi := \sum_{i,j=1}^d a^{ij}\frac{\partial^2\varphi}{\partial x_i \partial x_j} + \sum_{i=1}^d b^i \frac{\partial\varphi}{\partial x_i}\,.$$

3 Ergodic Control and the PIA

Let $c\colon \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ be a continuous function bounded from below. As well known, the ergodic control problem, in its *almost sure* (or *pathwise*) formulation, seeks to a.s. minimize over all admissible $U \in \mathfrak{U}$

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t c(X_s, U_s)\, \mathrm{d}s\,. \tag{7}$$

A weaker, *average* formulation seeks to minimize

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{E}^U \big[c(X_s, U_s)\big]\, \mathrm{d}s\,. \tag{8}$$

We let ϱ^* denote the infimum of (8) over all admissible controls. We assume that $\varrho^* < \infty$.

We assume that the cost function $c\colon \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}_+$ is continuous and locally Lipschitz in its first argument uniformly in $u \in \mathbb{U}$. More specifically, for some function $K_c\colon \mathbb{R}_+ \to \mathbb{R}_+$,

$$\big|c(x,u) - c(y,u)\big| \le K_c(R)|x-y| \qquad \forall x, y \in B_R\,,\ \forall u \in \mathbb{U}\,,$$

and all $R > 0$.

An important class of running cost functions arising in practice for which the ergodic control problem is well behaved are the *near-monotone* cost functions. Let $M^* \in \mathbb{R}_+ \cup \{\infty\}$ be defined by

$$M^* := \liminf_{|x|\to\infty} \min_{u\in\mathbb{U}} c(x,u)\,.$$

The running cost function c is called near-monotone if $\varrho^* < M^*$. Note that inf-compact functions c are always near-monotone.

We adopt the following abbreviated notation. For a function $g : \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ and $v \in \mathfrak{U}_{\mathrm{SSM}}$ we let

$$g_v(x) := g\big(x, v(x)\big)\,, \quad x \in \mathbb{R}^d\,.$$

The ergodic control problem for near-monotone cost functions is characterized as follows:

Theorem 1. *There exists a unique function $V \in \mathcal{C}^2(\mathbb{R}^d)$ which is bounded below in $\mathbb{R}^d$ and satisfies $V(0) = 0$ and the Hamilton–Jacobi–Bellman (HJB) equation*

$$\min_{u\in\mathbb{U}} \big[L^u V(x) + c(x,u)\big] = \varrho^*\,, \quad x \in \mathbb{R}^d\,.$$

The control $v^* \in \mathfrak{U}_{\mathrm{SM}}$ is optimal with respect to the criteria (7) *and* (8) *if and only if it satisfies*

$$\min_{u\in\mathbb{U}}\left[\sum_{i=1}^{d} b^i(x,u)\frac{\partial V}{\partial x_i}(x)+c(x,u)\right]=\sum_{i=1}^{d} b^i_{v^*}(x)\frac{\partial V}{\partial x_i}(x)+c_{v^*}(x)$$

a.e. in $\mathbb{R}^d$. *Moreover, with* $\breve{\tau}_r=\tau(B_r^c)$, $r>0$, *we have*

$$V(x)=\limsup_{r\downarrow 0}\inf_{v\in\mathfrak{U}_{\rm SSM}}\mathbb{E}^v_x\left[\int_0^{\breve{\tau}_r}\bigl(c_v(X_t)-\varrho^*\bigr)\,\mathrm{d}t\right],\quad x\in\mathbb{R}^d.$$

A control $v\in\mathfrak{U}_{\rm SM}$ is called *stable*, if the associated diffusion is positive recurrent. We denote the set of such controls by $\mathfrak{U}_{\rm SSM}$. Also we let μ_v denote the unique invariant probability measure on $\mathbb{R}^d$ for the diffusion under the control $v\in\mathfrak{U}_{\rm SSM}$. Recall that $v\in\mathfrak{U}_{\rm SSM}$ if and only if there exists an inf-compact function $\mathcal{V}\in\mathcal{C}^2(\mathbb{R}^d)$ a bounded domain $D\subset\mathbb{R}^d$ and a constant $\varepsilon>0$ satisfying

$$L^v\mathcal{V}(x)\le-\varepsilon\qquad\forall x\in D^c. \tag{9}$$

It follows that the optimal control v in Theorem 1 is stable. For $v\in\mathfrak{U}_{\rm SSM}$ we define

$$\varrho_v:=\limsup_{t\to\infty}\frac{1}{t}\int_0^t\mathbb{E}^v\bigl[c_v(X_s)\bigr]\,\mathrm{d}s.$$

A difficulty in synthesizing an optimal control $v\in\mathfrak{U}_{\rm SM}$ via the HJB equation lies in the fact that the optimal cost ϱ^* is not known. The PIA provides an iterative procedure for obtaining the HJB equation at the limit. In order to describe the algorithm we first need to review some properties of the Poisson equation

$$L^vV(x)+c_v(x)=\varrho,\quad x\in\mathbb{R}^d. \tag{10}$$

We need the following definition.

Definition 1. For $v\in\mathfrak{U}_{\rm SSM}$, and provided $\varrho_v<\infty$, define

$$\Psi^v(x):=\lim_{r\downarrow 0}\mathbb{E}^v_x\left[\int_0^{\breve{\tau}_r}\bigl(c_v(X_t)-\varrho_v\bigr)\,\mathrm{d}t\right],\quad x\neq 0.$$

For $v\in\mathfrak{U}_{\rm SM}$ and $\alpha>0$, let J^v_α denote the α-discounted cost

$$J^v_\alpha(x):=\mathbb{E}^v_x\left[\int_0^\infty \mathrm{e}^{-\alpha t}c_v(X_t)\,\mathrm{d}t\right],\quad x\in\mathbb{R}^d.$$

We borrow the following result from [1, Lemma 7.4]. If $v\in\mathfrak{U}_{\rm SSM}$ and $\varrho_v<\infty$, then there exist a function $V\in\mathcal{W}^{2,p}_{\rm loc}(\mathbb{R}^d)$, for any $p>1$, and a constant $\varrho\in\mathbb{R}$ which satisfy (10) a.e. in $\mathbb{R}^d$ and such that, as $\alpha\downarrow 0$, $\alpha J^v_\alpha(0)\to\varrho$ and $J^v_\alpha-J^v_\alpha(0)\to V$ uniformly on compact subsets of $\mathbb{R}^d$. Moreover,

$$\varrho=\varrho_v\quad\text{and}\quad V(x)=\Psi^v(x).$$

We refer to the function $V(x) = \Psi^v(x) \in \mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$ as the *canonical solution* of the Poisson equation $L^v V + c_v = \varrho_v$ in $\mathbb{R}^d$.

It can be shown that the canonical solution V to the Poisson equation is the unique solution in $\mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$ which is bounded below and satisfies $V(0) = 0$. Note also that (9) implies that any control v satisfying $\varrho_v < M^*$ is stable.

The PIA takes the following familiar form:

Algorithm (PIA).

1. *Initialization. Set $k = 0$ and select any $v_0 \in \mathfrak{U}_{\mathrm{SM}}$ such that $\varrho_{v_0} < M^*$.*
2. *Value determination. Obtain the canonical solution $V_k = \Psi^{v_k} \in \mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$, $p > 1$, to the Poisson equation*

$$L^{v_k} V_k + c_{v_k} = \varrho_{v_k}$$

   in $\mathbb{R}^d$.
3. *If $v_k(x) \in \operatorname{Arg\,min}_{u \in \mathbb{U}} \left[b^i(x,u) \partial_i V_k(x) + c(x,u) \right]$ x-a.e., return v_k.*
4. *Policy improvement. Select an arbitrary $v_{k+1} \in \mathfrak{U}_{\mathrm{SM}}$ which satisfies*

$$v_{k+1}(x) \in \operatorname*{Arg\,min}_{u \in \mathbb{U}} \left[\sum_{i=1}^{d} b^i(x,u) \frac{\partial V_k}{\partial x_i}(x) + c(x,u) \right], \qquad x \in \mathbb{R}^d .$$

Since $\varrho_{v_0} < M^*$ it follows that $v_0 \in \mathfrak{U}_{\mathrm{SSM}}$. The algorithm is well defined, provided $v_k \in \mathfrak{U}_{\mathrm{SSM}}$ for all $k \in \mathbb{N}$. This follows from the next lemma which shows that $\varrho_{v_{k+1}} \le \varrho_{v_k}$, and in particular that $\varrho_{v_k} < M^*$, for all $k \in \mathbb{N}$.

Lemma 1. *Suppose $v \in \mathfrak{U}_{\mathrm{SSM}}$ satisfies $\varrho_v < M^*$. Let $V \in \mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$, $p > 1$, be the canonical solution to the Poisson equation*

$$L^v V + c_v = \varrho_v , \quad \text{in } \mathbb{R}^d .$$

Then any measurable selector $\hat{v}$ from the minimizer

$$\operatorname*{Arg\,min}_{u \in \mathbb{U}} \left[b^i(x,u) \partial_i V(x) + c(x,u) \right]$$

satisfies $\varrho_{\hat{v}} \le \varrho_v$. Moreover, the inequality is strict unless v satisfies

$$L^v V(x) + c_v(x) = \min_{u \in \mathbb{U}} \left[L^u V(x) + c(x,u) \right] = \varrho_v , \quad \text{for almost all } x . \tag{11}$$

Proof. Let $\mathcal{V}$ be a Lyapunov function satisfying $L^v \mathcal{V}(x) \le k_0 - g(x)$, for some inf compact g such that $c_v \in \mathfrak{o}(g)$ (see [1, Lemma 7.1]). For $n \in \mathbb{N}$, define

$$\hat{v}_n(x) = \begin{cases} \hat{v}(x) & \text{if } x \in B_n \\ v(x) & \text{if } x \in B_n^c . \end{cases}$$

Clearly $\hat v_n \to \hat v$ as $n \to \infty$ in the topology of Markov controls (see [1, Section 3.3]). It is evident that $\mathcal{V}$ is a stochastic Lyapunov function relative to $\hat v_n$, i.e., there exist constants k_n such that $L^{\hat v_n}\mathcal{V}(x) \le k_n - g(x)$, for all $n \in \mathbb{N}$. Since $V \in \mathfrak{o}(\mathcal{V})$ it follows that (see [1, Lemma 7.1])

$$\frac{1}{t}\,\mathbb{E}_x^{\hat v_n}[V(X_t)] \xrightarrow[t\to\infty]{} 0 \tag{12}$$

Let

$$h(x) := \varrho_v - \min_{u\in\mathbb{U}} \left[L^u V(x) + c(x,u)\right], \quad x \in \mathbb{R}^d.$$

Also, by definition of $\hat v_n$, for all $m \le n$, we have

$$L^{\hat v_n} V(x) + c_{\hat v_n}(x) \le \varrho_v - h(x)\,\mathbb{I}_{B_m}(x). \tag{13}$$

By Itô's formula we obtain from (13) that

$$\begin{aligned}\frac{1}{t}\big(\mathbb{E}_x^{\hat v_n}[V(X_t)] - V(x)\big) + \frac{1}{t}\,\mathbb{E}_x^{\hat v_n}\left[\int_0^t c_{\hat v_n}(X_s)\,\mathrm{d}s\right] \\ \le \varrho_v - \frac{1}{t}\,\mathbb{E}_x^{\hat v_n}\left[\int_0^t h(X_s)\,\mathbb{I}_{B_m}(X_s)\,\mathrm{d}s\right],\end{aligned} \tag{14}$$

for all $m \le n$. Taking limits in (14) as $t \to \infty$ and using (12), we obtain

$$\varrho_{\hat v_n} \le \varrho_v - \int_{\mathbb{R}^d} h(x)\,\mathbb{I}_{B_m}(x)\,\mu_{\hat v_n}(\mathrm{d}x). \tag{15}$$

Note that $v \mapsto \varrho_v$ is lower semicontinuous. Therefore, taking limits in (15) as $n \to \infty$, we have

$$\varrho_{\hat v} \le \varrho_v - \limsup_{n\to\infty} \int_{\mathbb{R}^d} h(x)\,\mathbb{I}_{B_m}(x)\,\mu_{\hat v_n}(\mathrm{d}x). \tag{16}$$

Since c is near monotone and $\varrho_{\hat v_n} \le \varrho_v < M^*$, there exists $\hat R > 0$ and $\delta > 0$, such that $\mu_{\hat v_n}(B_{\hat R}) \ge \delta$ for all $n \in \mathbb{N}$. Then with $\psi_{\hat v_n}$ denoting the density of $\mu_{\hat v_n}$ Harnack's inequality [7, Theorem 8.20, p. 199] implies that there exists a constant $C_H = C_H(R)$ such that for every $R > \hat R$, with $|B_R|$ denoting the volume of $B_R \subset \mathbb{R}^d$, it holds that

$$\inf_{B_R} \psi_{\hat v_n} \ge \frac{\delta}{C_H |B_R|}, \quad \forall n \in \mathbb{N}.$$

This in turn implies by (16) that $\varrho_{\hat v} < \varrho_v$ unless $h = 0$ a.e. $\square$

4 Convergence of the PIA

We start with the following lemma.

Lemma 2. *The sequence $\{V_k\}$ of the PIA has the following properties:*

(i) *For some constant $C_0 = C_0(\varrho_{v_0})$ we have $\inf_{\mathbb{R}^d} V_k > C_0$ for all $k \geq 0$.*
(ii) *Each V_k attains it's minimum on the compact set*

$$\mathcal{K}(\varrho_{v_0}) := \left\{x \in \mathbb{R}^d : \min_{u \in \mathbb{U}} c(x,u) \leq \varrho_{v_0}\right\}.$$

(iii) *For any $p > 1$, there exists a constant $\tilde{C}_0 = \tilde{C}_0(R, \varrho_{v_0}, p)$ such that*

$$\left\|V_k\right\|_{\mathcal{W}^{2,p}(B_R)} \leq \tilde{C}_0 \qquad \forall R > 0\,.$$

(iv) *There exist positive numbers α_k and β_k, $k \geq 0$, such that $\alpha_k \downarrow 1$ and $\beta_k \downarrow 0$ as $k \to \infty$ and*

$$\alpha_{k+1} V_{k+1}(x) + \beta_{k+1} \leq \alpha_k V_k + \beta_k \quad \forall k \geq 0\,.$$

Proof. Parts (i) and (ii) follow directly from [3, Lemmas 3.6.1 and 3.6.4].

For part (iii) note first that the near monotone assumption implies that

$$\mu_{v_k}\left(\mathcal{K}\left(\tfrac{M^* + \varrho_{v_k}}{2}\right)\right) \geq \frac{M^* - \varrho_{v_k}}{M^* + \varrho_{v_k}} \qquad \forall k \geq 0\,.$$

Consequently

$$\mu_{v_k}\left(\mathcal{K}\left(\tfrac{M^* + \varrho_{v_0}}{2}\right)\right) \geq \frac{M^* - \varrho_{v_0}}{M^* + \varrho_{v_0}} \qquad \forall k \geq 0\,.$$

uniformly on compact subsets of $\mathbb{R}^d$. Hence since $J_\alpha^{v_k} - J_\alpha^{v_k}(0) \to V_k$ weakly in $\mathcal{W}^{2,p}(B_R)$ for any $R > 0$, (iii) follows from [3, Theorem 3.7.4].

Part (iv) follows as in [11, Theorem 4.4]. [1] □

As the corollary below shows, the PIA always converges.

Corollary 1. *There exists a constant $\hat{\varrho}$ and a function $\hat{V} \in \mathcal{C}^2(\mathbb{R}^d)$ with $\hat{V}(0) = 0$, such that, as $k \to \infty$, $\varrho_{v_k} \downarrow \hat{\varrho}$ and $V_k \to \hat{V}$ weakly in $\mathcal{W}^{2,p}(B_R)$, $p > 1$, for any $R > 0$. Moreover, $(\hat{V}, \hat{\varrho})$ satisfy the HJB equation*

$$\min_{u \in \mathbb{U}} \left[L^u \hat{V}(x) + c(x,u)\right] = \hat{\varrho}\,, \quad x \in \mathbb{R}^d\,. \tag{17}$$

[1] Theorem 4.4 in [11] applies to Markov chains on Borel state spaces. Also the model in [11] involves only inf-compact running costs. Nevertheless, the essential arguments can be followed to adapt the proof to controlled diffusions. We skip the details.

Proof. By Lemma 1, ϱ_{v_k} is decreasing monotonically in k, and hence converges to some $\hat\varrho \ge \varrho^*$. By Lemma 2 (iii) the sequence V_k is weakly compact in $\mathcal{W}^{2,p}(B_R)$, $p>1$, for any $R>0$, while by Lemma 2 (iv) any weakly convergent subsequence has the same limit $\hat V$. Also repeating the argument in the proof of Lemma 1, with

$$h_k(x) := \varrho_{v_{k-1}} - \min_{u\in\mathbb{U}} \left[L^u V_{k-1}(x) + c(x,u)\right], \quad x\in\mathbb{R}^d,$$

we deduce that for any $R>0$ there exists some constant $K(R)$ such that

$$\int_{B_R} h_k(x)\,\mathrm{d}x \le K(R)\big(\varrho_{v_{k-1}} - \varrho_{v_k}\big) \qquad \forall k\in\mathbb{N}.$$

Therefore $h_k\to 0$ weakly in $\mathcal{L}^1(D)$ as $k\to\infty$ for any bounded domain D. Taking limits in the equation

$$\min_{u\in\mathbb{U}} \left[L^u V_{k-1}(x) + c(x,u)\right] = \varrho_{v_{k-1}} - h_k(x)$$

and using [3, Lemma 3.5.4] yields (17). □

It is evident that $v\in\mathfrak{U}_{\rm SM}$ is an equilibrium of the PIA if it satisfies $\varrho_v < M^*$ and

$$\min_{u\in\mathbb{U}} \left[L^u \Psi^v(x) + c(x,u)\right] = \varrho_v, \quad x\in\mathbb{R}^d. \tag{18}$$

For one-dimensional diffusions one can show that (18) has a unique solution, and hence this is the optimal solution with $\varrho_v=\varrho^*$. For higher dimensions, to the best of our knowledge there is no such result. There is also the possibility that the PIA converges to $\hat v\in\mathfrak{U}_{\rm SSM}$ which is not an equilibrium. This happens if (17) satisfies

$$L^{\hat v}\hat V(x) + c_{\hat v}(x) = \min_{u\in\mathbb{U}} \left[L^u \hat V(x) + c(x,u)\right] = \hat\varrho > \varrho_{\hat v}, \quad x\in\mathbb{R}^d. \tag{19}$$

This is in fact the case with the example in [4]. In this example the controlled diffusion takes the form $\mathrm{d}X_t = U_t\,\mathrm{d}t + \mathrm{d}W_t$, with $\mathbb{U}=[-1,1]$ and running cost $c(x) = 1-\mathrm{e}^{-|x|}$. If we define

$$\xi_\varrho := \log\frac{3}{2} + \log(1-\varrho), \quad \varrho\in[1/3,1)$$

and

$$V_\varrho(x) := 2\int_{-\infty}^x \mathrm{e}^{2|y-\xi_\varrho|}\mathrm{d}y \int_{-\infty}^y \mathrm{e}^{-2|z-\xi_\varrho|}\big(\varrho - c(z)\big)\,\mathrm{d}z, \quad x\in\mathbb{R},$$

then direct computation shows that

$$\tfrac{1}{2}V_\varrho''(x) - |V_\varrho'(x)| + c(x) = \varrho \qquad \forall \varrho \in [1/3, 1)\ ,$$

and so the pair (V_ϱ, ϱ) satisfies the HJB. The stationary Markov control corresponding the this solution of the HJB is $w_\varrho(x) = -\operatorname{sign}(x - \xi_\varrho)$. The controlled process under w_ϱ has invariant probability density $\varphi_\varrho(x) = \mathrm{e}^{-2|x-\xi_\varrho|}$. A simple computation shows that

$$\int_{-\infty}^{\infty} c(x)\varphi_\varrho(x)\,\mathrm{d}x = \varrho - \tfrac{9}{8}(1-\varrho)(3\varrho - 1) < \varrho\,, \quad \forall \varrho \in (1/3, 1)\ .$$

Thus if $\varrho > 1/3$, then V_ϱ is not a canonical solution of the Poisson equation corresponding to the stable control w_ϱ. Therefore, this example satisfies (19) and shows that in general we cannot preclude the possibility that the limiting value of the PIA is not an equilibrium of the algorithm.

In [11, Theorem 5.2] a blanket Lyapunov condition is imposed to guarantee convergence of the PIA to an optimal control. Instead, we use Lyapunov analysis to characterize the domain of attraction of the optimal value.

We need the following definition.

Definition 2. Let v^* be an optimal control as characterized in Theorem 1. Let $\mathfrak{V}$ denote the class of all non-negative functions $\mathcal{V} \in \mathcal{C}^2(\mathbb{R}^d)$ satisfying $L^{v^*}\mathcal{V} \le k_0 - h(x)$ for some non-negative, inf-compact $h \in \mathcal{C}(\mathbb{R}^d)$ and a constant k_0. We denote by $\mathfrak{o}(\mathfrak{V})$ the class of inf-compact functions g satisfying $g \in \mathfrak{o}(\mathcal{V})$ for some $\mathcal{V} \in \mathfrak{V}$.

The theorem below asserts that if the PIA is initialized at a $v_0 \in \mathfrak{U}_{\mathrm{SSM}}$ whose associated canonical solution to the Poisson equation lies in $\mathfrak{o}(\mathfrak{V})$ then it converges to an optimal $v^* \in \mathfrak{U}_{\mathrm{SSM}}$.

Theorem 2. *If $v_0 \in \mathfrak{U}_{\mathrm{SSM}}$ satisfies $\Psi^{v_0} \in \mathfrak{o}(\mathfrak{V})$ then $\varrho_{v_k} \to \varrho^*$ as $k \to \infty$.*

Proof. The proof is straightforward. By Lemma 2 (iv), $\hat{V} \in \mathfrak{o}(\mathfrak{V})$. Also by (17), we have

$$L^{v^*}\hat{V}(x) + c_{v^*}(x) \ge \hat{\varrho}\,, \quad x \in \mathbb{R}^d\,,$$

and applying Dynkin's formula we obtain

$$\frac{1}{t}\big(\mathbb{E}_x^{v^*}\big[\hat{V}(X_t)\big] - V(x)\big) + \frac{1}{t}\,\mathbb{E}_x^{v^*}\left[\int_0^t c_{v^*}(X_s)\,\mathrm{d}s\right] \ge \hat{\varrho}\,, \tag{20}$$

Since $\hat{V} \in \mathfrak{o}(\mathfrak{V})$, by [1, Lemma 7.1] we have

$$\frac{1}{t}\,\mathbb{E}_x^{v^*}\big[\hat{V}(X_t)\big] \xrightarrow[t\to\infty]{} 0$$

and thus taking limits as $t \to \infty$ in (20) we obtain $\varrho^* \ge \hat{\varrho}$. Therefore, we must have $\hat{\varrho} = \varrho^*$. □

5 Concluding Remarks

We have concentrated on the model of controlled diffusions with near monotone running costs. The case of stable controls with a blanket Lyapunov condition is much simpler. If for example we impose the assumption that there exist a constant $k_0 > 0$, and a pair of nonnegative, inf-compact functions $(\mathcal{V}, h) \in \mathcal{C}^2(\mathbb{R}^d) \times \mathcal{C}(\mathbb{R}^d)$ satisfying $1 + c \in \mathfrak{o}(h)$ and such that

$$L^u \mathcal{V}(x) \le k_0 - h(x,u) \qquad \forall (x,u) \in \mathbb{R}^d \times \mathbb{U}\,,$$

then the PIA always converges to the optimal solution.

Acknowledgements This work was supported in part by the Office of Naval Research through the Electric Ship Research and Development Consortium.

References

1. Arapostathis, A., Borkar, V.S.: Uniform recurrence properties of controlled diffusions and applications to optimal control. SIAM J. Control Optim. **48**(7), 152–160 (2010)
2. Arapostathis, A., Borkar, V.S., Fernández-Gaucherand, E., Ghosh, M.K., Marcus, S.I.: Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. **31**(2), 282–344 (1993)
3. Arapostathis, A., Borkar, V.S., Ghosh, M.K.: Ergodic control of diffusion processes, *Encyclopedia of Mathematics and its Applications*, vol. 143. Cambridge University Press, Cambridge (2011)
4. Bensoussan, A., Borkar, V.: Ergodic control problem for one-dimensional diffusions with near-monotone cost. Systems Control Lett. **5**(2), 127–133 (1984)
5. Bogachev, V.I., Krylov, N.V., Röckner, M.: On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions. Comm. Partial Differential Equations **26**(11-12), 2037–2080 (2001)
6. Costa, O.L.V., Dufour, F.: The policy iteration algorithm for average continuous control of piecewise deterministic Markov processes. Appl. Math. Optim. **62**(2) (2010)
7. Gilbarg, D., Trudinger, N.S.: Elliptic partial differential equations of second order, *Grundlehren der Mathematischen Wissenschaften*, vol. 224, second edn. Springer-Verlag, Berlin (1983)
8. Gyöngy, I., Krylov, N.: Existence of strong solutions for Itô's stochastic equations via approximations. Probab. Theory Related Fields **105**(2), 143–158 (1996)
9. Hernández-Lerma, O., Lasserre, J.B.: Policy iteration for average cost Markov control processes on Borel spaces. Acta Appl. Math. **47**(2) (1997)
10. Krylov, N.V.: Controlled diffusion processes, *Applications of Mathematics*, vol. 14. Springer-Verlag, New York (1980)
11. Meyn, S.P.: The policy iteration algorithm for average reward Markov decision processes with general state space. IEEE Trans. Automat. Control **42**(12), 1663–1680 (1997)
12. Stannat, W.: (Nonsymmetric) Dirichlet operators on L^1: existence, uniqueness and associated Markov processes. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **28**(1), 99–140 (1999)