

ON THE RELATIVE VALUE ITERATION WITH A RISK-SENSITIVE CRITERION

ARI ARAPOSTATHIS

*Department of Electrical and Computer Engineering, The University of Texas at Austin
2501 Speedway, EER 7.824, Austin, TX 78712, USA
ORCID: 0000-0003-2207-357X E-mail: ari@utexas.edu*

VIVEK S. BORKAR

*Department of Electrical Engineering, Indian Institute of Technology
Powai, Mumbai 400076, India
ORCID: 0000-0003-0756-5402 E-mail: borkar@ee.iitb.ac.in*

Abstract. A multiplicative relative value iteration algorithm for solving the dynamic programming equation for the risk-sensitive control problem is studied for discrete time controlled Markov chains with a compact Polish state space, and controlled diffusions in the whole Euclidean space. The main result is a proof of convergence to the desired limit in each case.

1. Introduction. Risk-sensitive control problems on an infinite horizon seek to minimize or maximize a functional defined as the exponential growth rate of a multiplicative cost, resp. reward. Thus unlike the more classical and commonplace criteria, they lead to a multiplicative dynamic programming equation, in fact a nonlinear eigenvalue problem for a positive, positively 1-homogeneous continuous nonlinear operator. This has been extensively studied for the discrete time discrete state (both finite and countable) and continuous time and state problems, but the important case of discrete time and general state space has received relatively less attention in comparison, with only a small number of contributions such as [1, 20, 24]. The same also holds for the corresponding development of the value iteration algorithm, which ends up being a multiplicative analog of

2010 *Mathematics Subject Classification*: Primary: 90C40, 93E20, 49K40; Secondary: 60J25, 60J60.

Key words and phrases: risk-sensitive control, relative value iteration, controlled Markov process. The paper is in final form and no version of it will be published elsewhere.

the algorithm encountered in average cost problems, alternatively, in its simplest scenario, a nonlinear counterpart of the power iteration method for computing the principal eigenvector and eigenvalue of an irreducible non-negative matrix. This again has been studied in the discrete time and state case [16–18], but not for the general state space. In this work we take a first step towards filling in this gap by proposing and analyzing a multiplicative relative value iteration algorithm for two instances of risk-sensitive control on a general state space: the discrete time compact Polish state space problem, and the continuous time controlled diffusion in a Euclidean space. In the case of controlled diffusions, we would like to cite here the work in [19, 22, 23, 25] which is very much related to this problem.

2. Results in discrete time. We consider a controlled Markov chain on a compact Polish space \mathcal{S} with a compact metric action space U and controlled transition kernel

$$(x, u) \in \mathcal{S} \times U \mapsto p(dy | x, u) = \varphi(y | x, u) \gamma(dy) \in \mathcal{P}(\mathcal{S}),$$

where γ is some positive measure on \mathcal{S} with full support and $\varphi(\cdot | \cdot, \cdot) > 0$ is continuous. Also given is a ‘per stage’ continuous cost function

$$(x, u) \in \mathcal{S} \times U \mapsto k(x, u).$$

We shall denote by $X_n, n \geq 0$, and $Z_n, n \geq 0$, resp., the \mathcal{S} -valued state process and U -valued control process. Thus

$$P(X_{n+1} \in A | X_m, Z_m, m \leq n) = p(A | X_n, Z_n) \quad \forall n \in \mathbb{N}, \forall A \text{ Borel in } \mathcal{S}.$$

When $Z_n = v(X_n)$ for all n for some measurable $v: \mathcal{S} \mapsto U$, we call it a stationary Markov control policy and denote it simply by v . When

$$P(Z_n \in B | X_m, Z_m, m < n; X_n) = \phi(B | X_n) \quad \forall n \in \mathbb{N},$$

for some $\phi: \mathcal{S} \mapsto \mathcal{P}(U)$, we call it a randomized Markov control policy and denote it simply by ϕ .

The objective is to minimize the asymptotic risk-sensitive cost

$$\limsup_{n \uparrow \infty} \frac{1}{n} \log \mathbb{E} \left[\exp \left(\sum_{m=0}^{n-1} k(X_m, Z_m) \right) \right].$$

The ‘dynamic programming equation’ for this problem ends up being the nonlinear eigenvalue problem

$$\Lambda V(x) = \min_{u \in U} \left(e^{k(x, u)} \int_{\mathcal{S}} p(dy | x, u) V(y) \right), \quad x \in \mathcal{S}. \quad (2.1)$$

By Theorem 2.2 of [1], this has a solution $V(\cdot) \in C(\mathcal{S}; [0, \infty))$, $\Lambda \in (0, \infty)$, where Λ is unique, and V is unique up to a multiplicative positive scalar. Our objective is to propose a recursive scheme to compute these. Specifically, we consider the ‘Value Iteration’ (VI) algorithm given by

$$\begin{aligned} J_{n+1}(x) &= \frac{\min_{u \in U} \int_{\mathcal{S}} p(dy | x, u) e^{k(x, u)} J_n(y)}{\Lambda} \\ &= \frac{\int_{\mathcal{S}} p(dy | x, u_n(x)) e^{k(x, u)} J_n(y)}{\Lambda} \end{aligned}$$

for suitable $u_n(\cdot)$ guaranteed by a standard measurable selection theorem [29]. This is not a practicable algorithm since Λ is unknown. But it will serve a useful purpose in the analysis of the more realistic scheme, the ‘Relative Value Iteration’ (RVI). Choose some $x_0 \in \mathcal{S}$, which is kept fixed. The RVI is given by

$$\begin{aligned} V_{n+1}(x) &= \frac{\min_{u \in U} \int_{\mathcal{S}} p(dy | x, u) e^{k(x,u)} V_n(y)}{V_n(x_0)} \\ &= \frac{\int_{\mathcal{S}} p(dy | x, u''_n(x)) e^{k(x,u)} V_n(y)}{V_n(x_0)}, \end{aligned}$$

for suitable $u''_n(\cdot)$, initiated at $J_0 = V_0 > 0$ so that $V_n, J_n > 0$ for all n .

We have

$$\begin{aligned} \max_{x \in \mathcal{S}} \left(\frac{V_{n+1}(x)}{J_{n+1}(x)} \right) &= \max_{x \in \mathcal{S}} \left(\frac{\min_{u \in U} \int_{\mathcal{S}} p(dy | x, u) e^{k(x,u)} J_n(y) (V_n(y)/J_n(y))}{\min_{u \in U} \int_{\mathcal{S}} p(dy | x, u) e^{k(x,u)} J_n(y)} \right) \frac{\Lambda}{V_n(x_0)} \\ &\leq \max_{x \in \mathcal{S}} \left(\frac{V_n(x)}{J_n(x)} \right) \frac{\Lambda}{V_n(x_0)}. \end{aligned}$$

Similarly,

$$\min_{x \in \mathcal{S}} \left(\frac{V_{n+1}(x)}{S_{n+1}(x)} \right) \geq \min_{x \in \mathcal{S}} \left(\frac{V_n(x)}{J_n(x)} \right) \frac{\Lambda}{V_n(x_0)}.$$

Therefore

$$1 \leq \frac{\max_{x \in \mathcal{S}} ((V_{n+1}(x)/J_{n+1}(x)))}{\min_{x \in \mathcal{S}} ((V_{n+1}(x)/J_{n+1}(x)))} \leq \frac{\max_{x \in \mathcal{S}} ((V_n(x)/J_n(x)))}{\min_{x \in \mathcal{S}} ((V_n(x)/J_n(x)))} \leq \dots \leq 1,$$

implying that equality must hold throughout, that is, $V_n(x) = C_n J_n(x)$ for some constant C_n independent of x . We can then show inductively that

$$C_n := \frac{V_n(x)}{J_n(x)} = \prod_{m=0}^{n-1} \frac{\Lambda}{V_m(x_0)}.$$

Furthermore,

$$\frac{V_{n+1}(x_0)}{J_{n+1}(x_0)} = \frac{V_n(x_0)}{J_n(x_0)} \frac{\Lambda}{V_n(x_0)} = \frac{\Lambda}{J_n(x_0)}.$$

We say that the VI (RVI) converges if the sequence of functions $\{J_n\}_{n \in \mathbb{N}}$ ($\{V_n\}_{n \in \mathbb{N}}$) converges pointwise. If the VI converges, in particular $J_n(x_0)$ does, and by the above equations, the RVI will also converge. Thus we only need to establish the convergence of the VI.

Let $V(\cdot)$ and Λ be as in (2.1). Let $v^*(\cdot)$ denote a measurable minimizer of the right-hand side of (2.1). This is always possible by a measurable selection theorem [29]. Define

$$p^*(dy | x) := (\Lambda V(x))^{-1} p(dy | x, v^*(x)) e^{k(x, v^*(x))} V(y).$$

Then

$$\frac{J_{n+1}(x)}{V(x)} \leq \int_{\mathcal{S}} p^*(dy | x) \left(\frac{J_n(y)}{V(y)} \right).$$

Let $\{X_n^*\}$ denote the stationary chain governed by $p^*(\cdot|\cdot)$. Let $Y_n := X_{-n}^*$, for $n \in \mathbb{N}$. It then follows that

$$\frac{J_n(Y_n)}{V(Y_n)}, \quad n < 0,$$

is a reverse submartingale that converges a.s. and in $L_1(\nu)$ [28] to a random variable ζ (say). For any open $O \subset \mathcal{S}$, the martingale law of large numbers [28] yields

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \left(\mathbb{I}\{X_{m+1}^* \in O\} - p^*(O | X_m^*, v^*(X_m^*)) \right) = 0 \quad \text{a.s.}$$

Under our assumptions,

$$\min_{x,u} p^*(O | x, u) > \delta \gamma(O) > 0 \quad (2.2)$$

for some $\delta > 0$. Thus

$$\liminf_{n \uparrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{1}\{X_n^* \in O\} \geq \delta \gamma(O) \quad \text{a.s.},$$

implying $X_n^* \in O$ i.o., a.s. Fix $\eta > 0$ and let O be an open ϵ -ball centered at x for a prescribed $\epsilon > 0$, chosen such that

$$y \in O \implies |V(y) - V(x)| < \eta.$$

Pick a zero probability set \mathcal{N} outside which all ‘a.s.’ results above hold for $\epsilon = \frac{1}{m}$, $\eta = \frac{1}{k}$, and $m, k \geq 1$. Fix $x \in \mathcal{S}$. Fix a sample point $\omega \notin \mathcal{N}$. Take (possibly random) $n_0 \geq 1$ such that (say)

$$n \geq n_0 \implies |J_n(X_n^*) - \zeta V(X_n^*)| < \eta = \frac{1}{k}.$$

Then on $\{X_n^* \in O\}$ with $\epsilon = \frac{1}{m}$ (say), we have

$$\begin{aligned} |J_n(X_n^*) - \zeta V(x)| &\leq |J_n(X_n^*) - \zeta V(X_n^*)| + |\zeta(V(X_n^*) - V(x))| \\ &\leq (\zeta + 1) \frac{1}{k}. \end{aligned}$$

If we consider $k, m \uparrow \infty$, it follows that if $X_n^* \rightarrow x$ along a subsequence, then $J_n(X_n^*) \rightarrow \zeta V(x)$ along that subsequence. By (2.2), it also follows that $J_n(x) \rightarrow \zeta V(x)$ for γ -a.s. x . It then follows that $V_n(x) \rightarrow$ some $\bar{V}(x)$ γ -a.s. But then, passing to the limit in the defining equation for RVI, \bar{V} satisfies (2.1) with $\bar{V}(x_0) = \Lambda$, which uniquely specifies it.

3. Results in continuous time. In this section we consider the risk-sensitive control problem for a controlled diffusion on \mathbb{R}^d taking the form

$$dX_t = b(X_t, U_t) dt + \sigma(X_t) dW_t. \quad (3.1)$$

All random processes in (3.1) 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 , and the control process $\{U_t\}_{t \geq 0}$ lives in a compact metrizable space \mathbb{U} . The sets of admissible controls \mathfrak{U} , and stationary Markov controls \mathfrak{U}_{sm} are defined in the standard manner.

We let $a := \sigma \sigma^T$, and denote by B_R the open ball of radius R in \mathbb{R}^d centered at 0. We impose the following set assumptions on the coefficients, and the running cost $c: \mathbb{R}^d \times \mathbb{U} \rightarrow \mathbb{R}$.

ASSUMPTION 3.1. The following hold.

- (i) The drift $b: \mathbb{R}^d \times \mathbb{U} \rightarrow \mathbb{R}^d$ and running cost c are continuous, and for some positive constants C_R depending on $R > 0$, and C_0 , we have

$$|c(x, u) - c(y, u)| + |b(x, u) - b(y, u)| + \|\sigma(x) - \sigma(y)\| \leq C_R |x - y|$$

for all $x, y \in B_R$ and $u \in \mathbb{U}$, and

$$\sum_{i,j=1}^d a^{ij}(x) \zeta_i \zeta_j \geq C_0^{-1} |\zeta|^2 \quad \forall (x, \zeta) \in \mathbb{R}^d \times \mathbb{R}^d,$$

where $\|\sigma\| := (\text{Tr } \sigma \sigma^\top)^{1/2}$ denotes the Hilbert–Schmidt norm of the matrix σ .

- (ii) The function $a: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ is bounded, and for some $\theta \in [0, 1)$ and a constant κ_0 , we have

$$|b(x, u)| \leq \kappa_0(1 + |x|^\theta), \quad \text{and} \quad |c(x, u)| \leq \kappa_0(1 + |x|^{2\theta}) \quad (3.2)$$

for all $(x, u) \in \mathbb{R}^d \times \mathbb{U}$. In addition,

$$\min_{x \in B_R} \min_{u \in \mathbb{U}} c(x, u) \xrightarrow{R \rightarrow \infty} \infty, \quad (3.3)$$

and

$$\max_{x \in B_R} \frac{1}{|x|^{1-\theta}} \max_{u \in \mathbb{U}} \langle b(x, u), x \rangle^+ \xrightarrow{R \rightarrow \infty} 0. \quad (3.4)$$

DEFINITION 3.1. For $U \in \mathfrak{U}$ we define *the risk-sensitive value* under a control $U \in \mathfrak{U}_{\text{sm}}$, by

$$\Lambda_U^x = \Lambda_U^x(c) := \limsup_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_U^x \left[\exp \left(\int_0^T c(X_t, U_t) dt \right) \right], \quad (3.5)$$

and the *risk-sensitive optimal values* by

$$\Lambda_*^x := \inf_{U \in \mathfrak{U}} \Lambda_U^x, \quad \text{and} \quad \Lambda_* := \inf_{x \in \mathbb{R}^d} \Lambda_*^x. \quad (3.6)$$

Also let

$$\mathcal{G}f(x) := \frac{1}{2} \text{Tr}(a(x) \nabla^2 f(x)) + \min_{u \in \mathbb{U}} [\langle b(x, u), \nabla f(x) \rangle + c(x, u) f(x)], \quad f \in C^2(\mathbb{R}^d),$$

and

$$\lambda_* = \lambda_*(c) := \inf \{ \lambda \in \mathbb{R} : \exists \phi \in \mathcal{W}_{\text{loc}}^{2,d}(\mathbb{R}^d), \phi > 0, \mathcal{G}\phi - \lambda\phi \leq 0 \text{ a.e. in } \mathbb{R}^d \}. \quad (3.7)$$

Some discussion is in order here. The quantity λ_* is the generalized principal eigenvalue of the semilinear operator \mathcal{G} in \mathbb{R}^d . We assume that $\lambda_* < \infty$. Note that in specific problems, this is verified via a Foster–Lyapunov equation of the form

$$\frac{1}{2} \text{Tr}(a(x) \nabla^2 \mathcal{V}(x)) + \langle b_v(x), \nabla \mathcal{V}(x) \rangle + c_v(x) \mathcal{V}(x) \leq \kappa_0 - \kappa_1 \mathcal{V}(x)$$

for some positive function $\mathcal{V} \in C^2(\mathbb{R}^d)$ which is bounded away from 0, and for some $v \in \mathfrak{U}_{\text{sm}}$ and constants κ_0 and κ_1 . In this equation we used the convenient notation

$$b_v(x) := b(x, v(x)), \quad \text{and} \quad c_v(x) := c(x, v(x)) \quad \text{for } v \in \mathfrak{U}_{\text{sm}},$$

which we adopt for the rest of the paper.

3.1. The risk-sensitive HJB. As shown in [3, Lemmas 2.2 and 2.3], there exists a positive eigenfunction $\Psi \in C^2(\mathbb{R}^d)$ which solves

$$\mathcal{G}\Psi(x) = \lambda_* \Psi(x), \quad x \in \mathbb{R}^d, \quad (3.8)$$

and $\lambda_* \leq \Lambda_*^x$ for all $x \in \mathbb{R}^d$. We let $\mathfrak{U}_{\text{sm}}^*$ denote the controls $v \in \mathfrak{U}_{\text{sm}}$ which satisfy

$$\langle b_v(x), \nabla \Psi(x) \rangle + c_v(x) \Psi(x) = \min_{u \in \mathbb{U}} [\langle b(x, u), \nabla \Psi(x) \rangle + c(x, u) \Psi(x)] \quad \text{a.e. } x \in \mathbb{R}^d.$$

In other words, $\mathfrak{U}_{\text{sm}}^*$ is the set of measurable selectors from the minimizer of (3.8).

A variation of [3, Lemma 3.2], by (3.4), shows that

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}_U^x [|X_t|^{1+\theta}] = 0 \quad \forall U \in \mathfrak{U}. \quad (3.9)$$

Indeed, using the function $|x|^{2(1+\theta)}$ in equation (3.1) of [3] following the rest of the proof of [3, Lemma 3.2], we obtain (3.9). On the other hand, [4, Lemma 4.1] shows that (3.2) and (3.4) imply that there exists a constant $\tilde{C}_0 > 0$ such that any positive solution $\phi \in \mathcal{W}_{\text{loc}}^{2,d}(\mathbb{R}^d)$ of

$$\frac{1}{2} \text{Tr}(a(x) \nabla^2 \phi(x)) + \langle b_v(x), \nabla \phi(x) \rangle + c_v(x) \phi(x) = \lambda \phi(x)$$

for $v \in \mathfrak{U}_{\text{sm}}$, satisfies

$$\frac{|\nabla \phi(x)|}{\phi(x)} \leq \tilde{C}_0 (1 + |x|^\theta). \quad (3.10)$$

Therefore, by (3.10), the eigenfunction Ψ in (3.8) satisfies

$$e^{-C(1+|x|^{1+\theta})} \leq \Psi(x) \leq e^{C(1+|x|^{1+\theta})} \quad \forall x \in \mathbb{R}^d, \quad (3.11)$$

for some constant $C > 0$. An application of Fatou's lemma to the stochastic representation of the solution Ψ of (3.8) shows that

$$\Psi(x) \geq \mathbb{E}_{v^*}^x \left[\exp \left(\int_0^T [c_{v^*}(X_t) - \lambda_*] dt \right) \Psi(X_T) \right] \quad \forall T > 0, \quad (3.12)$$

with $v^* \in \mathfrak{U}_{\text{sm}}^*$. Taking logarithms on both sides of (3.12), applying Jensen's inequality, and dividing by T , we obtain

$$\frac{1}{T} \mathbb{E}_{v^*}^x \left[\int_0^T c_{v^*}(X_t) dt \right] + \frac{1}{T} \mathbb{E}_{v^*}^x [\log \Psi(X_T)] \leq \lambda_* + \frac{1}{T} \log \Psi(x). \quad (3.13)$$

Using (3.9) and (3.11) and taking limits as $T \rightarrow \infty$ in (3.13), we obtain

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{v^*}^x \left[\int_0^T c_{v^*}(X_t) dt \right] \leq \lambda_*.$$

This together with (3.3) implies that the diffusion in (3.1) controlled by $v^* \in \mathfrak{U}_{\text{sm}}^*$ has an invariant probability measure, and, therefore, it is positive recurrent [21, Theorem 3.3] (see also [14]). An application of [3, Lemma 2.1] then shows that Ψ is inf-compact, which in turn implies that $\Lambda_{v^*}^x \leq \lambda_*$ for all $x \in \mathbb{R}^d$, by [3, Lemma 2.1 (d) and (f)]. Since we have already asserted the converse inequality, this shows that

$$\Lambda_*^x = \Lambda_* = \lambda_* \quad \forall x \in \mathbb{R}^d,$$

or in other words, the optimal risk-sensitive value is equal to the generalized principal eigenvalue defined in (3.7). Note also that the inf-compactness of Ψ implies by (3.8) that

the diffusion in (3.1) controlled under $v^* \in \mathfrak{U}_{\text{sm}}^*$ is exponentially ergodic, or in other words, the transition probability of the process $\{X_t\}_{t \geq 0}$ in (3.1) under the control v^* , converges to its invariant probability measure in total variation at an exponential rate [27].

Uniqueness of the eigenfunction Ψ , which we refer to as the *ground state*, is related to the ergodic properties of the *ground state diffusion*, which takes the form

$$dX_t^* = (b(X_t^*, U_t) + a(X_t^*)\nabla\psi(X_t^*)) dt + \sigma(X_t^*) dW_t^*, \quad (3.14)$$

with $\psi := \log \Psi$. First, we have equality in (3.12) if and only if (3.14) controlled under $U_t = v^*(X_t^*)$ is regular. This is shown in [5, Lemma 2.3 and Corollary 2.2]. Note that (3.8) can be written in the form

$$\begin{aligned} \mathcal{G}^*\psi(x) &:= \frac{1}{2} \text{Tr}(a(x)\nabla^2\psi(x)) \\ &+ \min_{u \in \mathbb{U}} \left[\langle b(x, u) + \frac{1}{2}a(x)\nabla\psi(x), \nabla\psi(x) \rangle + c(x, u)\psi(x) \right] = \lambda\psi(x). \end{aligned} \quad (3.15)$$

Naturally, the sets of measurable selectors from the minimizers of (3.8) and (3.15) are equal. By (3.10), the hypothesis that a is bounded, and the growth assumptions of the drift in (3.2), it follows that (3.14) is regular for any $U \in \mathfrak{U}$. Thus, mimicking the proof of [5, Lemma 2.3] we obtain

$$\Psi(x) \leq \mathbb{E}_U^x \left[\exp \left(\int_0^T [c(X_t, U_t) - \lambda_*] dt \right) \Psi(X_T) \right] \quad \forall T > 0, \quad (3.16)$$

with equality when $U_t = v^*(X_t)$ for any $v^* \in \mathfrak{U}_{\text{sm}}^*$.

We review one important property of the generalized principal eigenvalue which concerns its dependence on the running cost c . Let

$$\mathcal{L}_u f(x) := \frac{1}{2} \text{Tr}(a(x)\nabla^2 f(x)) + \langle b(x, u), \nabla f(x) \rangle, \quad u \in \mathbb{U}, \quad (3.17)$$

and \mathcal{L}_v for $v \in \mathfrak{U}_{\text{sm}}$, denote the operator defined as above, but with $b(x, u)$ replaced by $b_v(x)$. For $v \in \mathfrak{U}_{\text{sm}}$ let

$$\lambda_v(c) := \inf \left\{ \lambda \in \mathbb{R} : \exists \phi \in \mathcal{W}_{\text{loc}}^{2,d}(\mathbb{R}^d), \phi > 0, \mathcal{L}_v \phi + c_v \phi - \lambda \phi \leq 0 \text{ a.e. in } \mathbb{R}^d \right\}. \quad (3.18)$$

Naturally, we have $\lambda_{v^*}(c) = \lambda_*$ for all $v^* \in \mathfrak{U}_{\text{sm}}^*$. Let $C_o^+(\mathbb{R}^d)$ denote the collection of all non-trivial, non-negative, continuous functions which vanish at infinity. We say that λ_v is *strictly monotone at c on the right* if $\lambda_v(c+h) > \lambda_v(c)$ for all $h \in C_o^+(\mathbb{R}^d)$. We can of course define the analogous property for λ_* , independently of the control $v^* \in \mathfrak{U}_{\text{sm}}^*$, using the definition in (3.7). Since $\mathfrak{U}_{\text{sm}}^*$ is the set of measurable selectors from the minimizer, it is clear that monotonicity at c on the right for λ_* and λ_{v^*} are equivalent.

Let $\hat{\tau}(A)$ denote the first hitting time of the set A . By [5, Lemma 2.7, Corollary 2.3, and Theorem 2.3], together with the equivalence of strict monotonicity on the right of λ_* and λ_{v^*} for $v^* \in \mathfrak{U}_{\text{sm}}^*$, we can assert that the following statements are equivalent.

1. The eigenvalue λ_* is simple.
2. For any open ball \mathcal{B} and $v^* \in \mathfrak{U}_{\text{sm}}^*$,

$$\Psi(x) = \mathbb{E}_{v^*}^x \left[\exp \left(\int_0^{\hat{\tau}(\mathcal{B})} [c_{v^*}(X_s) - \lambda_*] ds \right) \Psi(X_{\hat{\tau}(\mathcal{B})}) \mathbb{1}_{\{\hat{\tau}(\mathcal{B}) < \infty\}} \right] \quad \forall x \in \bar{\mathcal{B}}^c. \quad (3.19)$$

3. The ground state process in (3.14) controlled under any $v^* \in \mathfrak{U}_{\text{sm}}^*$ is recurrent.

We summarize the above discussion in the following theorem which is a slight variation of [4, Proposition 5.1].

THEOREM 3.1. *Grant Assumption 3.1, and suppose that λ_* is finite. Then the HJB equation*

$$\min_{u \in \mathbb{U}} [\mathcal{L}_u \Psi(x) + c(x, u) \Psi(x)] = \lambda_* \Psi(x) \quad \forall x \in \mathbb{R}^d \quad (3.20)$$

has a solution $\Psi \in C^2(\mathbb{R}^d)$, satisfying $\inf_{\mathbb{R}^d} \Psi > 0$, and the following hold:

- (a) $\Lambda_*^x = \Lambda_* = \lambda_*$ for all $x \in \mathbb{R}^d$.
- (b) Any $v^* \in \mathfrak{U}_{\text{sm}}^*$ renders the SDE in (3.1) exponentially ergodic and is optimal, that is, $\Lambda_{v^*}^x = \Lambda_*$ for all $x \in \mathbb{R}^d$.
- (c) For any $v \in \mathfrak{U}_{\text{sm}}^*$,

$$\Psi(x) = \mathbb{E}_{v^*}^x \left[\exp \left(\int_0^T [c(X_t, v(X_t)) - \lambda_*] dt \right) \Psi(X_T) \right] \quad \forall (T, x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

and, in addition, (3.16) holds.

- (d) The function $\psi = \log \Psi$ satisfies $|\nabla \psi| \leq \tilde{C}_0(1 + |x|)$ for some constant \tilde{C}_0 .
- (e) If λ_* is strictly monotone at c on the right, then there exists a unique (up to a positive multiplicative constant) positive solution to (3.20) (ground state), and any optimal $v \in \mathfrak{U}_{\text{sm}}^*$ lies in $\mathfrak{U}_{\text{sm}}^*$. In addition, the ground state Ψ satisfies (3.19), and (3.14) controlled under $U_t = v^*(X_t^*)$ with $v^* \in \mathfrak{U}_{\text{sm}}^*$ is recurrent.

There is another important property that we need in the study of convergence of the value iteration, which we explain next. Let $v \in \mathfrak{U}_{\text{sm}}$. We say that $\lambda_v(c)$, defined in (3.18), is *strictly monotone at c* if $\lambda_v(c - h) < \lambda_v(c)$ for some $h \in C_0^+(\mathbb{R}^d)$. Of course, strict monotonicity implies strict monotonicity on the right as can be seen from the fact that $c \mapsto \lambda_v(c)$ is convex. By [5, Theorem 2.1] strict monotonicity of $\lambda_{v^*}(c)$ at c is equivalent to the statement that the ground state diffusion in (3.14) controlled under v^* is positive recurrent.

3.2. The value iteration. Let

$$C_{\Psi, +}^2(\mathbb{R}^d) := \{g \in C^2(\mathbb{R}^d) : g > 0, \|g\|_{\Psi} < \infty\}.$$

We introduce the equation

$$\partial_t \bar{\Phi}(t, x) = \min_{u \in \mathbb{U}} [\mathcal{L}_u \bar{\Phi}(t, x) + c(x, u) \bar{\Phi}(t, x)] - \lambda_* \bar{\Phi}(t, x), \quad t > 0, \quad (3.21)$$

with $\bar{\Phi}(0, x) = \Phi_0(x)$, $\Phi_0 \in C_{\Psi, +}^2(\mathbb{R}^d)$.

DEFINITION 3.2. Let $\{\hat{v}_t\}_{t \geq 0}$ be an a.e. measurable selector from the minimizer of (3.21). We define the corresponding (non-stationary) *Markov control*

$$\hat{v}^t := \{\hat{v}_s^t = \hat{v}_{t-s}(x), \quad s \in [0, t]\}.$$

and denote the set of these controls by $\hat{\mathcal{U}}(\Phi_0)$, including explicitly the dependence on the initial condition Φ_0 in the notation.

We do not care so much about uniqueness of solutions to (3.21); however, see [11, Theorems 3.12–3.13]. We work with the solution $\bar{\Phi}(t, x)$ which satisfies

$$\begin{aligned}\bar{\Phi}(t, x) &= \inf_{U \in \mathfrak{U}} \mathbb{E}_U^x \left[\exp \left(\int_0^t [c(X_s, U_s) - \lambda_*] ds \right) \Phi_0(X_t) \right] \\ &= \mathbb{E}_{\hat{v}^t}^x \left[\exp \left(\int_0^t [c(X_s, \hat{v}_s^t(X_s)) - \lambda_*] ds \right) \Phi_0(X_t) \right] \quad \forall \{\hat{v}^t\}_{t \geq 0} \in \widehat{\mathcal{U}}(\Phi_0).\end{aligned}$$

Note that for any element of $\widehat{\mathcal{U}}(\Phi_0)$ we have $\hat{v}_{s+\tau}^t = \hat{v}_s^t$ for all $t \geq s \geq 0$ and $\tau \geq 0$. Also, by (3.16), we obtain

$$\Psi(x) \leq \mathbb{E}_{\hat{v}^t}^x \left[\exp \left(\int_0^t [c(X_s, \hat{v}_s^t(X_s)) - \lambda_*] ds \right) \Psi(X_t) \right] \quad \forall \{\hat{v}^t\}_{t \geq 0} \in \widehat{\mathcal{U}}(\Phi_0).$$

Incorporating explicitly the dependence on the initial condition Φ_0 in the notation, we let $\mathcal{S}_t[\Phi_0](x)$, $t \geq 0$, denote the solution of (3.21). It is clear that $\mathcal{S}_t[\Psi] = \Psi$ for all $t \geq 0$ by Theorem 3.1 (c), and that the uniqueness of the ground state in Theorem 3.1 (e) implies that any positive initial condition Φ_0 satisfying $\mathcal{S}_t[\Phi_0] = \Phi_0$ for all $t \geq 0$ must equal the ground state Ψ up to a positive multiplicative constant.

Let \mathcal{E} denote the set of equilibria of the semiflow \mathcal{S}_t , or equivalently, the set of solutions of the HJB in (3.20), that is,

$$\mathcal{E} := \{r\Psi : r > 0\}.$$

By $C_\Psi(\mathbb{R}^d)$ we denote the class of continuous functions ϕ satisfying

$$\|\phi\|_\Psi := \sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{\Psi(x)} < \infty.$$

For $\kappa > 0$ we define the set $\mathcal{H}_\kappa \subset C^2(\mathbb{R}^d)$ by

$$\mathcal{H}_\kappa := \{h \in C^2(\mathbb{R}^d) : h \geq \kappa^{-1}\Psi, \|h\|_\Psi < \kappa\}. \quad (3.22)$$

We have

$$\begin{aligned}\kappa^{-1}\Psi(x) &= \mathcal{S}_t[\kappa^{-1}\Psi](x) \leq \mathcal{S}_t[\Phi_0](x) \\ &\leq \mathcal{S}_t[\|\Phi_0\|_\Psi \Psi](x) \leq \mathcal{S}_t[\kappa\Psi](x) = \kappa\Psi(x) \quad \forall \Phi_0 \in \mathcal{H}_\kappa,\end{aligned} \quad (3.23)$$

where the first and the last equalities follow by Theorem 3.1 (c), and the inequalities by the monotonicity of $f \mapsto \mathcal{S}_t[f]$ and the definition of \mathcal{H}_κ . It follows from (3.23) that if $\Phi_0 \in \mathcal{H}_\kappa$ then $\mathcal{S}_t[\Phi_0] \in \mathcal{H}_\kappa$ for all $t \geq 0$. So the set \mathcal{H}_κ is positively invariant under the semiflow \mathcal{S}_t .

Recall the definition of \mathcal{L} in (3.17), and let

$$\tilde{\mathcal{L}}_u := \mathcal{L}_u + \langle \nabla \psi(x), a(x) \nabla \rangle, \quad u \in \mathbb{U}. \quad (3.24)$$

This definition can be extended to $\tilde{\mathcal{L}}_v$ for any Markov control v (not necessarily stationary) by replacing $u \in \mathbb{U}$ with v in (3.24). Clearly then \mathcal{L}_v , with $v \in \mathfrak{U}_{\text{sm}}$, is the extended generator of (3.14) controlled by v . The operator $\tilde{\mathcal{L}}_u$ satisfies a very important identity. If $\Phi \in C^2(\mathbb{R}^d)$ is a positive function then

$$\tilde{\mathcal{L}}_u \left(\frac{\Phi}{\Psi} \right) = \left(\frac{\mathcal{L}_u \Phi}{\Phi} - \frac{\mathcal{L}_u \Psi}{\Psi} \right) \frac{\Phi}{\Psi} \quad \forall u \in \mathbb{U}. \quad (3.25)$$

In the sequel we work under the following hypothesis.

(H1) *The ground state diffusion in (3.14) is positive recurrent under some $v^* \in \mathfrak{U}_{\text{sm}}^*$. We let $\tilde{\mu}_*$ denote its invariant probability measure, and $\tilde{\mathbb{E}}_*^x$ expectation operator on the canonical space of the process controlled under v^* .*

As explained in Subsection 3.1, under (H1), $\lambda_{v^*}(c)$ is strictly monotone at c . Therefore, by Theorem 3.1, we have unicity of the ground state Ψ , and complete verification of optimality results. In what follows v^* is the control in (H1).

We present the following important convergence result.

THEOREM 3.2. *Grant (H1). For each $\Phi_0 \in \mathcal{H}_\kappa$, $\kappa > 0$, the semiflow $\mathcal{S}_t[\Phi_0]$ converges to $\kappa_0\Psi \in \mathcal{E}$ for some $\kappa_0 \in [\kappa^{-1}, \kappa]$ as $t \rightarrow \infty$. Moreover, if A is a bounded subset of $C_\Psi(\mathbb{R}^d)$, then the only subsets of $\mathcal{H}_\kappa \cap A$, with $\kappa > 0$, which are invariant under the semiflow are the points (singletons) of $\mathcal{E} \cap \mathcal{H}_\kappa \cap A$.*

Proof. Define $\Phi_\Psi(t, x) := \frac{\mathcal{S}_t[\Phi_0](x)}{\Psi(x)}$. By (3.25) applied to (3.8) and (3.21), we have

$$\partial_t \Phi_\Psi(t, x) - \tilde{\mathcal{L}}_{v^*} \Phi_\Psi(t, x) \leq 0. \quad (3.26)$$

Since $\Phi_\Psi(t, x)$ is bounded by (3.23), we obtain from (3.26)

$$\Phi_\Psi(t, x) \leq \tilde{\mathbb{E}}_*^x[\Phi_\Psi(\tau, X_\tau)], \quad 0 \leq \tau \leq t. \quad (3.27)$$

Integrating (3.27) with respect to $\tilde{\mu}_*$, and using the abbreviated notation $\tilde{\mu}_*(f) = \int_{\mathbb{R}^d} f(x) \tilde{\mu}_*(dx)$, we obtain

$$\tilde{\mu}_*(\Phi_\Psi(t, x)) \leq \tilde{\mu}_*(\Phi_\Psi(s, x)) \quad \text{for all } t > s.$$

Thus, since $t \mapsto \tilde{\mu}_*(\Phi_\Psi(t, x))$ is non-increasing, and $\Phi_\Psi(t, x) \in \mathcal{H}_\kappa$ by (3.23), it converges to some constant $\kappa_0 \in [\kappa^{-1}, \kappa]$ as $t \rightarrow \infty$. It is clear that $\sup_{t>0} \|\mathcal{S}_t[\Phi_0]\|_\Psi < \|\Phi_0\|_\Psi$ by (3.23). Therefore by the interior estimates of solutions of (3.21) (see [26, Theorem 6.2, p. 457]), $\{\mathcal{S}_t[\Phi_0] : t > 0\}$ is locally precompact in $C^2(\mathbb{R}^d)$. Hence the ω -limit set of Φ_0 under the semiflow \mathcal{S}_t , denoted by $\omega(\Phi_0)$, is non-empty, and is a subset of $C^2(\mathbb{R}^d)$. Note that the convergence of $\tilde{\mu}_*(\Phi_\Psi(t, x))$ to κ_0 as $t \rightarrow \infty$ implies that

$$\tilde{\mu}_*\left(\frac{h}{\Psi}\right) = \kappa_0 \quad \forall h \in \omega(\Phi_0). \quad (3.28)$$

Fix some $h \in \omega(\Phi_0)$, and define

$$g(t, x) := \mathcal{L}_{v^*} \mathcal{S}_t[h](x) + c_{v^*}(x) \mathcal{S}_t[h](x) - \min_{u \in \mathbb{U}} [\mathcal{L}_u \mathcal{S}_t[h](x) + c(x, u) \mathcal{S}_t[h](x)]. \quad (3.29)$$

Therefore, by (3.21) and (3.29), we have

$$\partial_t \bar{\Phi}(t, x) = \mathcal{L}_{v^*} \mathcal{S}_t[h](x) + c_{v^*}(x) \mathcal{S}_t[h](x) - g(t, x) - \lambda_* \bar{\Phi}(t, x), \quad t > 0, \quad (3.30)$$

which we write as

$$\partial_t \bar{\Phi}(t, x) = \mathcal{L}_{v^*} \mathcal{S}_t[h](x) + \left(c_{v^*}(x) - \frac{g(t, x)}{\mathcal{S}_t[h](x)} \right) \mathcal{S}_t[h](x) - \lambda_* \bar{\Phi}(t, x).$$

Using (3.21), (3.25), and (3.30), we obtain

$$\partial_t \frac{\mathcal{S}_t[h](x)}{\Psi(x)} - \tilde{\mathcal{L}}_{v^*} \frac{\mathcal{S}_t[h](x)}{\Psi(x)} = - \frac{g(t, x)}{\Psi(x)}. \quad (3.31)$$

Since $\|\mathcal{S}_t[h]\|_\Psi \leq \kappa$ by the positive invariance of \mathcal{H}_κ , we can apply Itô's formula to (3.31) to obtain

$$\frac{\mathcal{S}_t[h](x)}{\Psi(x)} = -\tilde{\mathbb{E}}_x^* \left[\int_0^t \frac{g(t-s, X_s)}{\Psi(X_s)} ds \right] + \tilde{\mathbb{E}}_x^* \left[\frac{h(X_t)}{\Psi(X_t)} \right] \quad \forall t > 0. \quad (3.32)$$

As argued earlier $t \mapsto \tilde{\mu}_* \left(\frac{\mathcal{S}_t[h](x)}{\Psi(x)} \right)$ is constant. Hence, integrating (3.32) with respect to $\tilde{\mu}_*$, we obtain

$$\int_0^t \int_{\mathbb{R}^d} g(t-s, x) \frac{1}{\Psi(x)} \tilde{\mu}_*(dx) ds = 0 \implies g(t, x) = 0 \quad (t, x)\text{-a.e.}$$

where we used the fact that $\Psi(x) > 0$. Therefore, the first term on the right-hand side of (3.29) is identically equal to 0. Since $\frac{h}{\Psi}$ is bounded and the diffusion governed by $\tilde{\mathcal{L}}^*$ is ergodic, the second term on the right-hand side of (3.32) converges as $t \rightarrow \infty$ to some constant κ_0 by (3.28). Thus, again by (3.32), $\mathcal{S}_t[h]$ converges to $\kappa_0\Psi$ along any subsequence as $t \rightarrow \infty$, and the invariance of the ω -limit set of $\mathcal{S}_t[\Phi_0]$ implies that $h = \kappa_0\Psi$. This completes the proof. ■

3.3. The relative value iteration. We modify (3.21) as follows:

$$\partial_t \Phi(t, x) = \min_{u \in \mathbb{U}} [\mathcal{L}_u \Phi(t, x) + f(x, u) \Phi(t, x)] - \Phi(t, 0) \Phi(t, x), \quad t > 0, \quad (3.33)$$

with $\Phi(0, x) = \Phi_0(x)$. Existence of solutions to (3.33) is evident from the following observation: If Φ solves (3.33) then

$$\bar{\Phi}(t, x) = \Phi(t, x) \exp \left(\int_0^t (\Phi(s, 0) - \lambda_*) ds \right) \quad (3.34)$$

solves (3.21). Therefore,

$$\frac{\bar{\Phi}(t, x)}{\Phi(t, x)} = \frac{\bar{\Phi}(t, 0)}{\Phi(t, 0)} \quad \forall (t, x) \in (0, \infty) \times \mathbb{R}^d, \quad (3.35)$$

so that $\frac{\bar{\Phi}(t, x)}{\Phi(t, x)}$ does not depend on x . By (3.34)–(3.35) we have

$$\frac{d}{dt} \frac{\bar{\Phi}(t, x)}{\Phi(t, x)} = -\bar{\Phi}(t, 0) + \lambda_* = -\bar{\Phi}(t, 0) \frac{\bar{\Phi}(t, x)}{\bar{\Phi}(t, x)} + \lambda_*.$$

Thus

$$\frac{\bar{\Phi}(t, x)}{\bar{\Phi}(t, x)} = \exp \left(- \int_0^t \bar{\Phi}(s, 0) ds \right) + \lambda_* \int_0^t \exp \left(- \int_\tau^t \bar{\Phi}(s, 0) ds \right) d\tau. \quad (3.36)$$

It follows by (3.36) that if $\bar{\Phi}(t, 0) \rightarrow C > 0$ as $t \rightarrow \infty$ for some positive constant C , then $\frac{\bar{\Phi}(t, x)}{\bar{\Phi}(t, x)}$ converges to a positive constant as $t \rightarrow \infty$, and thus by (3.34) we have

$$\int_0^t (\Phi(s, 0) - \lambda_*) ds \xrightarrow{t \rightarrow \infty} \text{constant.}$$

In particular $\Phi(t, 0) \rightarrow \lambda_*$ as $t \rightarrow \infty$.

3.4. Results under blanket exponential ergodicity. Under blanket exponential ergodicity, we can remove the hypotheses in Assumption 3.1 (ii). We keep Assumption 3.1 (i), and add an affine growth condition of the form

$$\sup_{u \in \mathbb{U}} (b(x, u), x)^+ + \|\sigma(x)\|^2 \leq C_0(1 + |x|^2) \quad \forall x \in \mathbb{R}^d. \quad (3.37)$$

Concerning the running cost, we assume that it is bounded below, and, without loss of generality, we normalize it so that $\inf_{\mathbb{R}^d \times \mathbb{U}} c = 0$.

The essential hypothesis in this subsection is the following.

ASSUMPTION 3.2. We distinguish two cases.

- (i) If c is bounded, we assume that there exist a function $\mathcal{V} \in C^2(\mathbb{R}^d)$ taking values in $[1, \infty)$, a compact set $\mathcal{K} \subset \mathbb{R}^d$, and constants \widehat{C} and $\gamma > \|c\|_\infty$ which satisfy

$$\mathcal{L}_u \mathcal{V}(x) \leq \widehat{C} \mathbf{1}_{\mathcal{K}}(x) - \gamma \mathcal{V}(x) \quad \forall u \in \mathbb{U}. \quad (3.38)$$

- (ii) If c is not bounded, we assume that there exist an inf-compact function F and a constant $\beta \in (0, 1)$ such that $\beta F - c$ is also inf-compact, and \mathcal{V} , \mathcal{K} , and \widehat{C} as in part (i), such that

$$\mathcal{L}_u \mathcal{V}(x) \leq \widehat{C} \mathbf{1}_{\mathcal{K}}(x) - F(x) \mathcal{V}(x). \quad (3.39)$$

The reason for differentiating cases (i) and (ii) in Assumption 3.2 is because if the coefficients a and b are bounded, it is not, in general, possible to find an inf-compact function F which satisfies (3.39).

Under Assumption 3.2 we obtain a much stronger version of Theorem 3.1. Recall the definitions in (3.5)–(3.7), and $\mathfrak{U}_{\text{sm}}^*$ in the beginning of Subsection 3.1. The following theorem is a combination of [5, Theorems 4.1 and 4.2], and the results in [5, Section 3].

THEOREM 3.3. *Grant Assumption 3.1 (i), (3.37), and Assumption 3.2. Then λ_* is finite, and the equation*

$$\min_{u \in \mathbb{U}} [\mathcal{L}_u \Psi(x) + c(x, u) \Psi(x)] = \lambda_* \Psi(x) \quad \forall x \in \mathbb{R}^d \quad (3.40)$$

has a unique positive solution $\Psi \in C^2(\mathbb{R}^d)$, $\Psi(0) = 1$, and the following hold:

- (a) $\Lambda_*^x = \Lambda_* = \lambda_*$ for all $x \in \mathbb{R}^d$.
- (b) A stationary Markov control is optimal, if and only if it belongs to $\mathfrak{U}_{\text{sm}}^*$.
- (c) Part (c) of Theorem 3.1 holds, and also (3.19).
- (d) The ground state diffusion (3.14) is exponentially ergodic under any stationary Markov control.

We review part (d) of Theorem 3.3 which is not discussed in [5]. First, it is straightforward to show, by using (3.38) and (3.39) as a barrier in the construction of the solution Ψ , that $\frac{\mathcal{V}}{\Psi}$ is bounded away from 0 on \mathbb{R}^d . Second, note that the nonnegativity of c implies that $\lambda_* \geq 0$. Thus, from (3.25), (3.38), and (3.40) we obtain

$$\tilde{\mathcal{L}}_u \left(\frac{\mathcal{V}}{\Psi} \right) (x) \leq (\mathcal{V}^{-1}(x) \widehat{C} \mathbf{1}_{\mathcal{K}}(x) - \lambda_* + c(x, u) - \gamma) \frac{\mathcal{V}(x)}{\Psi(x)} \quad \forall (x, u) \in \mathbb{R}^d \times \mathbb{U}. \quad (3.41)$$

Under (3.39), γ gets replaced by F in (3.41). It is well known (see [9, Lemma 2.5.5]) that (3.41) implies that there exist positive constants $\tilde{\kappa}_0$ and $\tilde{\kappa}_1$ such that

$$\tilde{\mathbb{E}}_U^x \left[\frac{\mathcal{V}}{\Psi}(X_t) \right] \leq \tilde{\kappa}_0 + \frac{\mathcal{V}}{\Psi}(x) e^{-\tilde{\kappa}_1 t} \quad \forall x \in \mathbb{R}^d, \forall U \in \mathfrak{U}. \quad (3.42)$$

Let $\tilde{P}_t^v(x, dy)$ denote the transition probability of the process $\{X_t^*\}_{t \geq 0}$ in (3.14) under the control $v \in \mathfrak{U}_{\text{sm}}$, and $\tilde{\mu}_v$ its invariant probability measure. Then, using the argument as in the proof of [13, Theorem 2.1 (b)], one can show that (3.41) implies that there exist positive constants γ_\circ and C_{γ_\circ} , which do not depend on $v \in \mathfrak{U}_{\text{sm}}$, such that

$$\|\tilde{P}_t^v(x, \cdot) - \tilde{\mu}_v(\cdot)\|_{\text{TV}} \leq C_{\gamma_\circ} \frac{\mathcal{V}(x)}{\Psi(x)} e^{-\gamma_\circ t} \quad \forall (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm.

REMARK 3.1. We want to point out that the proof of [5, Theorems 4.1 and 4.2], shows that under the hypotheses of Theorem 3.3, the generalized principal eigenvalue λ_v defined in (3.18) is finite for any $v \in \mathfrak{U}_{\text{sm}}$, and there exists a positive $\Psi_v \in \mathcal{W}_{\text{loc}}^{2,p}(\mathbb{R}^d)$, for any $p \geq d$, which solves

$$\mathcal{L}_v \Psi_v(x) + c_v(x) \Psi_v(x) = \lambda_v \Psi_v(x) \quad \text{a.e. } x \in \mathbb{R}^d. \quad (3.43)$$

In addition, Ψ_v is the unique positive solution of (3.43) in $\mathcal{W}_{\text{loc}}^{2,d}(\mathbb{R}^d)$ up to a positive multiplicative constant, and $\lambda_v = \Lambda_v^x$ for all $x \in \mathbb{R}^d$, or in other words, the risk-sensitive value equals the generalized principal eigenvalue of the operator $\mathcal{L}_v + c_v$. Another important result is given in [5, Theorem 4.3] which shows that, under Assumption 3.2, $v \mapsto \lambda_v$ is continuous in the topology of Markov controls (see [15] for a definition of this topology).

Moving on to the VI algorithm under the assumptions of Theorem 3.3, note that by (3.27) we have

$$\Phi_\Psi(t, x) = \frac{\mathcal{S}_t[\Phi_0](x)}{\Psi(x)} \leq \tilde{\mathbb{E}}_*^x \left[\frac{\Phi_0}{\Psi}(X_t) \right] \quad \forall t \geq 0. \quad (3.44)$$

This gives us an upper bound. To obtain a lower bound, we use the measurable selector $\{\hat{v}_t\}$ in Definition 3.2 and combine (3.21), (3.25), and (3.40), to write

$$\partial_t \Phi_\Psi(t, x) - \tilde{\mathcal{L}}_{\hat{v}_t} \Phi_\Psi(t, x) \geq 0. \quad (3.45)$$

With $\bar{\varphi}(t, x) := \log \bar{\Phi}(t, x)$ and $\varphi_0 := \log \Phi_0$, we deduce from (3.45) that

$$\bar{\varphi}(t, x) \geq \psi(x) + \tilde{\mathbb{E}}_{\hat{v}_t}^x [\varphi_0(X_t) - \psi(X_t)], \quad (3.46)$$

where the expectation is under the non-stationary control $\{\hat{v}_t\}_{t \geq 0} \in \widehat{\mathcal{U}}(\Phi_0)$ in Definition 3.2.

We borrow the following result. As shown in the proof of [5, Theorem 4.3], under Assumption 3.2, there exist positive constants $\hat{\kappa}_0$, and $\delta_\circ > 1$ such that $\mathcal{V} \geq \hat{\kappa}_0 \Psi^{\delta_\circ}$. This together with (3.42) and Jensen's inequality shows that there exists a constant $\hat{\kappa}_1$ such that

$$\tilde{\mathbb{E}}_{\hat{v}_t}^x [\psi(X_t)] \geq \frac{1}{\delta_\circ - 1} \log \left(\frac{\tilde{\kappa}_0}{\hat{\kappa}_0} + \frac{\mathcal{V}(x)}{\hat{\kappa}_0 \Psi(x)} e^{-\tilde{\kappa}_1 t} \right). \quad (3.47)$$

Combining (3.46) and (3.47), we obtain

$$\liminf_{t \rightarrow \infty} \bar{\varphi}(t, x) \geq \psi(x) + \left(\inf_{\mathbb{R}^d} \varphi_0 \right) + \frac{1}{\delta_o - 1} \log \left(\frac{\tilde{\kappa}_0}{\kappa_0} \right). \quad (3.48)$$

Equations (3.44) and (3.48) shows that as long as the initial condition Φ_0 is bounded from below away from 0 in \mathbb{R}^d , and $\|\Phi_0\|_V < \infty$, then any limit point in $C^2(\mathbb{R}^d)$ of the semiflow $\mathcal{S}_t[\Phi_0]$ lies in the set \mathcal{H}_κ for some $\kappa > 0$ (recall the definition in (3.22)). Using the interior estimates of solutions and the bounds in (3.44), (3.46), and (3.47), as in the proof of Theorem 3.2, it is straightforward to show that the ω -limit set of Φ_0 is a non-empty subset of $C^2(\mathbb{R}^d)$, therefore also of \mathcal{H}_κ . Hence, following the arguments in [6, Section 4.2] which is based on convergence of reverse supermartingales, or the method in [7] that has a dynamical systems flavor (see also [10, Theorem 3.1]), one can establish the following result.

THEOREM 3.4. *Grant Assumption 3.1 (i), (3.37), and Assumption 3.2, and suppose that the initial condition $\Phi_0 \in C^2(\mathbb{R}^d)$ is bounded from below away from 0, and satisfies $\|\Phi_0\|_V < \infty$. Then there exists a positive constant $\kappa_0 = \kappa_0(\Phi_0)$ such that the value iteration $\bar{\Phi}(t, x)$ in (3.21) converges to $\kappa_0 \Psi(x)$ as $t \rightarrow \infty$ uniformly on compact sets.*

REMARK 3.2. When the state space is compact, stronger results can be obtained. Such a scenario is investigated in [12], and Theorem 4.3 in that paper shows in fact that under mild assumptions, and for a large class of abstract problems, the convergence is exponential.

REMARK 3.3. It is worth investigating if the global convergence result in Theorem 3.4 holds under additional assumptions in the near-monotone case. Suppose that $\theta = 1$ in (3.2) and (3.4) and that c has strictly quadratic growth. Then, by (3.10), c satisfies

$$\min_{u \in \mathbb{U}} c(x, u) \geq \theta_1 \psi(x) - \theta_2 \quad \forall x \in \mathbb{R}^d \quad (3.49)$$

for some positive constants θ_1 and θ_2 . In the case of the ergodic control problem, under the structural condition in (3.49), with ψ replaced by the solution of the HJB equation, global convergence can be established for the value iteration in continuous [11, Theorem 3.2], as well as in discrete time [8, Theorems 6.1–6.2] (see also [2]). For the risk-sensitive problem, this inequality has to be modified to account for the *relative entropy rate* term arising from the logarithmic transformation. We strengthen (3.49) to

$$\min_{u \in \mathbb{U}} c(x, u) - \frac{1}{2} \left| \sigma^\top(x) \nabla \psi(x) \right|^2 \geq \theta_1 \psi(x) - \theta_2 \quad \forall x \in \mathbb{R}^d. \quad (3.50)$$

Note that (3.50) implies (H1). We conjecture that under the structural assumption in (3.50) the value iteration $\bar{\Phi}(t, x)$ in (3.21), starting from any initial condition $\Phi_0 \in C_{\Psi, +}^2(\mathbb{R}^d)$, converges to an equilibrium in \mathcal{E} .

Acknowledgments. The work of Ari Arapostathis was supported in part by the Army Research Office through grant W911NF-17-1-001, in part by the National Science Foundation through grant DMS-1715210, and in part by the Office of Naval Research through grant N00014-16-1-2956 and was approved for public release under DCN #43-6054-19. The work of Vivek Borkar was supported by a J. C. Bose Fellowship.

References

- [1] V. Anantharam, V. S. Borkar, *A variational formula for risk-sensitive reward*, SIAM J. Control Optim. 55 (2017), 961–988.
- [2] A. Arapostathis, *Open problem—Convergence and asymptotic optimality of the relative value iteration in ergodic control*, Stoch. Syst. 9 (2019), 292–294.
- [3] A. Arapostathis, A. Biswas, *Infinite horizon risk-sensitive control of diffusions without any blanket stability assumptions*, Stochastic Process. Appl. 128 (2018), 1485–1524.
- [4] A. Arapostathis, A. Biswas, V. S. Borkar, K. Suresh Kumar, *A variational characterization of the risk-sensitive average reward for controlled diffusions in \mathbb{R}^d* , SIAM J. Control Optim., in press; ArXiv: 1903.08346.
- [5] A. Arapostathis, A. Biswas, S. Saha, *Strict monotonicity of principal eigenvalues of elliptic operators in \mathbb{R}^d and risk-sensitive control*, J. Math. Pures Appl. (9) 124 (2019), 169–219.
- [6] A. Arapostathis, V. S. Borkar, *A relative value iteration algorithm for nondegenerate controlled diffusions*, SIAM J. Control Optim. 50 (2012), 1886–1902.
- [7] A. Arapostathis, V. S. Borkar, *A correction to “A relative value iteration algorithm for nondegenerate controlled diffusions”*, SIAM J. Control Optim. 55 (2017), 1711–1715.
- [8] A. Arapostathis, V. S. Borkar, *Average cost optimal control under weak ergodicity hypotheses: Relative value iterations*, ArXiv:1902.01048.
- [9] A. Arapostathis, V. S. Borkar, M. K. Ghosh, *Ergodic Control of Diffusion Processes*, Encyclopedia Math. Appl. 143, Cambridge Univ. Press, Cambridge, 2012.
- [10] A. Arapostathis, V. S. Borkar, K. Suresh Kumar, *Relative value iteration for stochastic differential games*, in: Advances in Dynamic Games, Ann. Internat. Soc. Dynam. Games 13, Birkhäuser/Springer, Cham, 2013, 3–27.
- [11] A. Arapostathis, V. S. Borkar, K. Suresh Kumar, *Convergence of the relative value iteration for the ergodic control problem of nondegenerate diffusions under near-monotone costs*, SIAM J. Control Optim. 52 (2014), 1–31.
- [12] A. Arapostathis, V. S. Borkar, K. Suresh Kumar, *Risk-sensitive control and an abstract Collatz–Wielandt formula*, J. Theoret. Probab. 29 (2016), 1458–1484.
- [13] A. Arapostathis, H. Hmedi, G. Pang, *On uniform exponential ergodicity of Markovian multiclass many-server queues in the Halfin–Whitt regime*, Math. Oper. Res., in press, ArXiv: 1812.03528.
- [14] R. N. Bhattacharya, *Criteria for recurrence and existence of invariant measures for multi-dimensional diffusions*, Ann. Probab. 6 (1978), 541–553.
- [15] V. S. Borkar, *A topology for Markov controls*, Appl. Math. Optim. 20 (1989), 55–62.
- [16] V. S. Borkar, S. P. Meyn, *Risk-sensitive optimal control for Markov decision processes with monotone cost*, Math. Oper. Res. 27 (2002), 192–209.
- [17] R. Cavazos-Cadena, R. Montes-de Oca, *The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space*, Math. Oper. Res. 28 (2003), 752–776.
- [18] R. Cavazos-Cadena, R. Montes-de Oca, *Nonstationary value iteration in controlled Markov chains with risk-sensitive average criterion*, J. Appl. Probab. 42 (2005), 905–918.
- [19] E. Chasseigne, N. Ichihara, *Ergodic problems for viscous Hamilton–Jacobi equations with inward drift*, SIAM J. Control Optim. 57 (2019), 23–52.
- [20] G. B. Di Masi, L. Stettner, *Risk-sensitive control of discrete-time Markov processes with infinite horizon*, SIAM J. Control Optim. 38 (1999), 61–78.

- [21] R. Z. Has'minskiĭ, *Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations*, Teor. Veroyatnost. i Primenen. 5 (1960), 196–214; English transl.: Theory Probab. Appl. 5 (1960), 179–196.
- [22] N. Ichihara, *The generalized principal eigenvalue for Hamilton–Jacobi–Bellman equations of ergodic type*, Ann. Inst. H. Poincaré Anal. Non Linéaire 32 (2015), 623–650.
- [23] N. Ichihara, S.-J. Sheu, *Large time behavior of solutions of Hamilton–Jacobi–Bellman equations with quadratic nonlinearity in gradients*, SIAM J. Math. Anal. 45 (2013), 279–306.
- [24] A. Jaśkiewicz, *Average optimality for risk-sensitive control with general state space*, Ann. Appl. Probab. 17 (2007), 654–675.
- [25] H. Kaise, S.-J. Sheu, *On the structure of solutions of ergodic type Bellman equation related to risk-sensitive control*, Ann. Probab. 34 (2006), 284–320.
- [26] O. A. Ladyženskaja, V. A. Solonnikov, N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, Transl. Math. Monogr. 23, Amer. Math. Soc., Providence, RI, 1968.
- [27] S. P. Meyn, R. L. Tweedie, *Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes*, Adv. in Appl. Probab. 25 (1993), 518–548.
- [28] J. Neveu, *Discrete-Parameter Martingales*, revised edition, North-Holland Math. Library 10, American Elsevier Publishing Co., New York, 1975.
- [29] T. Parthasarathy, *Selection Theorems and Their Applications*, Lecture Notes in Math. 263, Springer, Berlin, 1972.