ON ITERATION IMPROVEMENT FOR AVERAGED EXPECTED COST CONTROL FOR 1D ERGODIC DIFFUSIONS[*][†]

SVETLANA V. ANULOVA[‡], HILMAR MAI[§], AND ALEXANDER YU. VERETENNIKOV[¶]

Abstract. An ergodic Bellman's (HJB) equation is proved for a uniformly ergodic 1D controlled diffusion with variable diffusion and drift coefficients both depending on control; convergence of the values provided by Howard's reward improvement algorithm to the value which is a component of the unique solution of Bellman's equation is established.

Key words. controlled diffusion processes, averaged expected control, Hamilton-Jacobi-Bellman equation, existence and uniqueness, reward improvement algorithm

AMS subject classifications. 93E20; 60H10

1. Introduction. The paper is a complete version of the short presentation without detailed proofs in [1]. Issues of reliability which was in the title of [1] are not addressed here, all proofs are completed and the results are extended in comparison to the cited article. However, an application to reliability seems fruitful and is one of the motivations for the present paper; a corresponding remark about it can be found below. One more motivation is to allow the diffusion coefficient to depend on control. Indirectly, the main result below may be considered as a version of a rigorous realisation of the rather instructive and deliberately non-rigorous example from [15, Ch. 1, §1] where the point was the vanishing at infinity of the expectation of a current cost. Beside a more detailed calculus in step 3 of the proof, here we tackle the issue of the HJB equation(s) satisfied everywhere and/or almost everywhere more precisely than in [1].

We consider a one-dimensional stochastic differential equation (SDE) on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ with a one-dimensional $(\mathcal{F}_t)$ Wiener process $B = (B_t)_{t\geq 0}$ with coefficients b and σ, and with a stationary control function α (called strategy in the sequel)

$$dX_t^\alpha = b(\alpha(X_t^\alpha), X_t^\alpha)\, dt + \sigma(\alpha(X_t^\alpha), X_t^\alpha)\, dW_t, \quad t \geq 0, \tag{1.1}$$
$$X_0^\alpha = x.$$

Let a compact set $U \subset \mathbb{R}$ be a set where any strategy takes its values. The functions b and σ on $U \times \mathbb{R}$ are assumed Borel; later on some further conditions will be imposed, but we note straight

[*]THE PAPER IS A FULL VERSION OF THE SHORT PRESENTATION IN [1].

[†]Submitted to SICON.

Funding: For the first author this research has been supported by the Russian Foundation for Basic Research grant no. 17-01-00633_a. The second author thanks the Institut Louis Bachelier for financial support. The third author is grateful to the financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" at Bielefeld University during his stay there in August 2017; also, for this author this study has been funded by the Russian Academic Excellence Project '5-100' and by the Russian Foundation for Basic Research grant no. 17-01-00633_a. All the authors gratefully acknowledge the support and hospitality of the Oberwolfach Research Institute for Mathematics (MFO) during the RiP programme in June 2014 where this study was initiated.

[‡]Institute for Information Transmission Problems, Moscow, Russia (anulova@mail.ru).

[§]CREST and ENSAE ParisTech, France; email: hilmar.mai@gmail.com.

[¶]University of Leeds, UK, & National Research University Higher School of Economics, & Institute for Information Transmission Problems, Moscow, Russia; email: a.veretennikov@leeds.ac.uk.

away that σ will be assumed non-degenerate and that a weak solution of the equation (1.1) always exists and is Markov and strong Markov, see [16, 17, 14]. Denote the class of all Borel functions α with values in U by $\mathcal{A}$. For $u \in U$ and $\alpha(\cdot) \in \mathcal{A}$ denote

$$L^u(x) = b(u,x)\frac{d}{dx} + \frac{1}{2}\sigma^2(u,x)\frac{d^2}{dx^2}, \quad x \in \mathbb{R},$$

and

$$L^\alpha(x) = b(\alpha(x),x)\frac{d}{dx} + \frac{1}{2}\sigma^2(\alpha(x),x)\frac{d^2}{dx^2}, \quad x \in \mathbb{R}.$$

Denote by $\mathcal{K}$ the class of functions on $U \times \mathbb{R}$ (also just on $\mathbb{R}$) growing no faster than some polynomial. The *running cost* function f will always be chosen from this class. The *averaged cost* function corresponding to the strategy $\alpha \in \mathcal{A}$ is then defined as

$$\rho^\alpha(x) := \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f(\alpha(X_t^\alpha), X_t^\alpha)\,dt. \tag{1.2}$$

For a strategy $\alpha \in \mathcal{A}$ the function $f^\alpha : \mathbb{R} \to \mathbb{R},\ f^\alpha(x) = f(\alpha(x),x),\ x \in \mathbb{R}$, is defined. Then (1.2) has an equivalent form

$$\rho^\alpha(x) = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f^\alpha(X_t^\alpha)\,dt. \tag{1.3}$$

Now, the *cost function* for the model under consideration is defined as

$$\rho(x) := \inf_{\alpha\in\mathcal{A}} \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f^\alpha(X_t^\alpha)\,dt. \tag{1.4}$$

It will be assumed that for every $\alpha \in \mathcal{A}$ the solution of the equation (1.1) X^α is Markov ergodic, i.e., there exists a limiting in total variation distribution μ^α of $X_t^\alpha,\ t \to \infty$, this distribution μ^α does not depend on the initial condition $X_0 = x \in \mathbb{R}$, is unique and is invariant for the generator L^α. The cost function ρ^α then does not depend on x and can be rewritten as

$$\rho^\alpha(x) \equiv \rho^\alpha := \int f^\alpha(x)\,\mu^\alpha(dx) =: \langle f^\alpha, \mu^\alpha\rangle. \tag{1.5}$$

Then what we want to find (compute) is the value

$$\rho := \inf_{\alpha\in\mathcal{A}} \int f^\alpha(x)\,\mu^\alpha(dx) = \inf_{\alpha\in\mathcal{A}} \langle f^\alpha, \mu^\alpha\rangle. \tag{1.6}$$

For any strategy $\alpha \in \mathcal{A}$ let us also define an auxiliary function

$$v^\alpha(x) := \int_0^\infty \mathbb{E}_x(f^\alpha(X_t^\alpha) - \rho^\alpha)\,dt.$$

The convergence of this integral will follow from the assumptions.

The first goal of this paper is to show the *ergodic HJB or Bellman's equation* on the pair (V,ρ)

$$\inf_{u\in U}[L^u V(x) + f^u(x) - \rho] = 0, \quad x \in \mathbb{R}. \tag{1.7}$$

This assumes showing uniqueness of the second component (ρ) along with the property that it coincides with the cost from (1.6). The meaning of the first component V will be explained later. The uniqueness of V will be shown up to an additive constant.

The class where the solution (V, ρ) will be studied is the family of all Borel functions V and constants $\rho \in \mathbb{R}$ such that V has two Sobolev derivatives which are all locally integrable in any power, and V itself should have a moderate grow at infinity not faster than some polynomial. Respectively, the equation (1.7) is to be understood almost everywhere; yet, in the 1D situation and under our assumptions it will follow straightforwardly that this equation is actually satisfied for all $x \in \mathbb{R}$. Note that the first derivative can be considered as continuous (due to the embedding theorems), and the second derivative will be always taken Borel, as one of the Borel representatives of Lebesgue's measurable function.

The second goal of the paper is to show how to approach the solution ρ of the main problem by some successive approximation procedure called the Reward Improvement Algorithm (RIA). It is interesting that under our minimal assumptions on regularity of strategies for the weak SDE solution setting it is yet possible to justify a *monotonic convergence* of the "exact" RIA; compare to [15, ch.1, §4] where it was necessary to work with "approximate" RIA (called Bellman–Howard's iteration procedure there) and with regularized Lipschitz strategies.

Concerning the equation (1.7), it may look like it lacks some boundary conditions: indeed, a 2nd order PDE normally does require certain boundary conditions, which, for example, in the considered 1D case simply means two boundary conditions at two end-points if the equation is on a bounded interval. However, this is the equation "in the whole space" and we are going to solve it in a specific class of functions V – namely, bounded (if f is assumed bounded), or, at most, moderately growing (if f may admit some moderate growth), – which in some sense substitutes the (Dirichlet) boundary conditions at $\pm\infty$. Note that a similar situation can be found in the theory of Poisson equations in the whole space (see, for example, [**?**, 32]).

Concerning a full uniqueness for the solution of (1.7), note that with any solution (V, ρ) and for any constant C, the couple $(V + C, \rho)$ is also a solution. There are two close enough options how to tackle this fact: either accept that uniqueness will be established up to a constant, or choose a certain "natural" constant satisfying some "centering condition" as will be done below.

To guarantee ergodicity, we will assume the "blanket" recurrence conditions (see below), which in some sense provide a uniform recurrence for *any* strategy. Conditions of this type are sometimes considered too restrictive; however, they do allow to include models and cases not covered earlier in this theory and for this reason we regard this restriction as a reasonable price for the time being. It is likely that such restrictions may be relaxed so as to include the "near monotonicity" type conditions (see [5]).

Let us say just a few words about the history of the problem. More can be found in the references provided below. Earlier results on ergodic control in continuous time were obtained in [22], [26], [6], et al. In his book [22] Mandl established apparently first results on ergodic (averaged) control for controlled 1D diffusion on a finite interval with boundary conditions including jumps from the boundary. The author established the HJB equation and proved uniqueness of the couple (up to a constant for the first component). Improvement of control was discussed, too, however, without convergence.

Morton [26] considered the 1D case (a multi-dimensional case too but under stronger assumptions: we do not touch it in this paper) with a price function defined by (1.6) without any relation to (1.4). He proved ([26, Theorem 1]) that the optimal price does satisfy the ergodic Bellman's equation; that the policy determined by Argmax (in our setting Argmin) in the Bellman's equation is optimal within some rather special class of Markov policies which are fixed functions outside some bounded interval; a certain inequality for the optimal price and any solution of Bellman's equation; a remark about RIA; however, neither is the uniqueness for the Bellman's equation solutions established, nor is the convergence of RIA towards a solution proved.

Discrete time controlled models were considered in the monographs [9], [11], [12], [28], and others, and in the papers [2], [24], [29], etc.

Continuous time controlled processes were treated in the 80s in a chapter of the monograph [6] where ergodic control for stable diffusions was considered. Arapostathis and Borkar [4], Arapostathis [3], Arapostathis, Borkar and Ghosh [5] treated diffusions with "relaxed control" and the diffusion coefficient not depending on the control, under weaker recurrence assumptions (i.e., under two types of condition, stable or near-monotone). In this setting, they establish Bellman's equation, existence, uniqueness, and RIA convergence. In this paper we allow the diffusion coefficient to depend on control and we do not use relaxed control.

The latest works include [3], [5], [29], see also the references therein. Although devoted to another type of models – piecewise-linear Markov ones – the monograph [8] may also be mentioned here. In the very first papers and books compact cases with some auxiliary boundary conditions – so as to simplify ergodicity – were studied; convergence of the improvement control algorithms were studied only partially. In later investigations noncompact spaces are allowed; however, apparently, *ergodic* control in the diffusion coefficient σ of the process has not been tackled earlier. The reader may consult [6] and [15] for research on controlled diffusion processes on a finite horizon, or on infinite horizon with discount (technically equivalent to killing).

In most of the works on the topic, measurability of the optimal or improved strategy (see below) is assumed. Yet, it is a subtle issue and in our case we give references – the basic one is [30] – and verify the conditions which provide this measurability.

The paper consists of four sections: 1 – Introduction, 2 – Assumptions and some auxiliaries, 3 – Main result and its proof, and the last one is the Appendix (not numbered). We will use the convention that arbitrary constants C in the calculus may change from line to line.

2. Assumptions and some auxiliaries. To ensure ergodicity of X^α under any stationary control strategy $\alpha \in \mathcal{A}$, we make the following assumptions on the drift and diffusion coefficients.

(A1) (boundedness, non-degeneracy, regularity) The functions b and σ are Borel bounded in their variables; $|b(u,x)| \le C_b$, $|\sigma(u,x)| \le C_\sigma$, σ is uniformly non-degenerate, $|\sigma(u,x)|^{-1} \le C_\sigma$; the functions $\sigma(u,x)$, $b(u,x)$, $f^u(x)$ are continuous in u for every x.

(A2) (recurrence)

$$\lim_{|x|\to\infty} \sup_{u\in U} x\, b(u,x) = -\infty. \tag{2.1}$$

(A3) (running cost) The function f belongs to the class $\mathcal{K}$ of functions which are Borel measurable in x for each u and admit a uniform in u polynomial bound: there exist constants $C_1, m_1 > 0$ such that for any x,

$$\sup_{u\in U} |f^u(x)| \le C_1(1+|x|^{m_1}).$$

(A4) (compactness of U) The set U is compact.

(A5) (additional regularity) The functions b, σ, and f are of the class C^1 in x for each u with uniformly bounded derivatives.

We will need the following three lemmata.

LEMMA 2.1. *Let the assumptions (A1) – (A3) hold true. Then*

- *For any $C_1, m_1 > 0$ there exist $C, m > 0$ such that for any strategy $\alpha \in \mathcal{A}$ and for any function g growing no faster than $C_1(1+|x|^{m_1})$,*

$$\sup_{t\ge 0} |\mathbb{E}_x g(X^\alpha_t)| \le C(1+|x|^m). \tag{2.2}$$

- *For any $\alpha \in \mathcal{A}$, the invariant measure μ^α integrates any polynomial and*

$$\sup_{\alpha\in\mathcal{A}} \int |x|^k \, \mu^\alpha(dx) < \infty, \quad \forall\, k>0.$$

- *For any strategy $\alpha \in \mathcal{A}$ the function ρ^α is a constant, and*

$$\sup_{\alpha\in\mathcal{A}} |\rho^\alpha| \le C < \infty; \tag{2.3}$$

moreover, for any $k>0$ and $f \in \mathcal{K}$, there exist $C, m > 0$ such that

$$\sup_{\alpha\in\mathcal{A}} |\mathbb{E}_x f^\alpha(X^\alpha_t) - \rho^\alpha| \le C\frac{1+|x|^m}{1+t^k}, \tag{2.4}$$

and

$$\sup_{\alpha\in\mathcal{A}} \left| \frac{1}{T}\int_0^T \mathbb{E}_x f^\alpha(X^\alpha_t)\,dt - \rho^\alpha \right| \to 0, \quad T\to\infty. \tag{2.5}$$

Proof. Follows from [31, Theorems 5, 6]. Note that in [31] the solution of the SDE under investigation should be weakly unique, and it also must be a homogeneous Markov and strong Markov process; for the equation (1.1) it is all true by virtue of [16, Theorem 3], [17], and [14, Theorems 2, 3], as no continuity of the diffusion coefficient is required for this in the 1D case. (NB: In [14, Theorem 3] no continuity is needed even for $D \ge 1$, but then weak uniqueness is established in the 1D case only [16, Theorem 3].)

COROLLARY 2.2. *Under the same assumptions,*

$$\sup_{t\ge 0} |\mathbb{E}_x 1(|X^\alpha_t| > N)| \le \sup_{t\ge 0} \mathbb{E}_x \frac{|X^\alpha_t|^m}{N^m} \le \frac{C(1+|x|^m)}{N^m}. \tag{2.6}$$

The proof is straightforward by Bienaymé – Chebyshev –Markov's inequality.

REMARK 2.3. *Note that because $D=1$, under the assumptions (A1)–(A2) for any Borel function $\alpha \in \mathcal{A}$ there is a unique stationary measure μ^α, which is* equivalent to the Lebesgue measure Λ. *The latter follows from the formula for the unique stationary density*

$$p^\alpha(x) := \frac{d\mu^\alpha(x)}{dx} = C_\alpha \frac{1}{\sigma^2(\alpha(x),x)} \exp\left(2\int_0^x \frac{b(\alpha(y),y)}{\sigma^2(\alpha(y),y)}\,dy\right), \tag{2.7}$$

where C_α is a normed constant. The fact that p^α is a stationary density can be seen from a substitution to the equation of stationarity $(L^\alpha)^*p = 0$ (see, for example, [13, Lemma 4.16, equation

(4.70)]); its uniqueness in the class of integrable functions satisfying the normalizing condition $\int p\,dx = 1$ can be justified via the explicit solution of the stationarity equation in the 1D case which we leave to the readers.

In the next Lemma (as well as later in the main Theorem) we use Sobolev spaces $W^2_{p,loc}$ with $p > 1$. (this notation are taken from [19, Chapter 2], although, in some other sources it is denoted by $W^{2,p}_{loc}$.) Although all main statements can be stated without them, this is done in order to mimick the steps in the proof where these spaces show up naturally due to the direct references, even though the dimension equals one, in which case, of course, some calculus can be simlipified.

LEMMA 2.4. *Let the assumptions (A1) – (A3) be satisfied. Then for any strategy $\alpha \in \mathcal{A}$ the cost function v^α has the following properties:*

1. *The function v^α is continuous as well as $(v^\alpha)'$, and there exist $C, m > 0$ both depending only on the constants in (A1)–(A3) such that*

$$\sup_\alpha (|v^\alpha(x)| + |v^\alpha(x)'|) \le C(1 + |x|^m). \tag{2.8}$$

2. *$v^\alpha \in W^2_{p,loc}$ for any $p \ge 1$.*
3. *$v^\alpha \in C^{1,Lip}$ (i.e., $(v^\alpha)'$ is locally Lipschitz).*
4. *v^α satisfies a Poisson equation in the whole space,*

$$L^\alpha v^\alpha + f^\alpha - \langle f^\alpha, \mu^\alpha \rangle = 0, \tag{2.9}$$

in the Sobolev sense; in particular, for almost every $x \in \mathbb{R}$

$$L^\alpha(x) v^\alpha(x) + f^\alpha(x) - \langle f^\alpha, \mu^\alpha \rangle = 0. \tag{2.10}$$

5. *The solution of the equation (2.9) is unique up to an additive constant in the class of Sobolev solutions $W^2_{p,loc}$ with any $p > 1$ with no more than some (any) polynomial growth of the solution v^α.*
6. $\langle v^\alpha, \mu^\alpha \rangle = 0$.

Proof. Firstly, the inequality

$$\sup_\alpha |v^\alpha(x)| \le C(1 + |x|^m)$$

follows immediately from (2.2) and from the assumptions.

Further, let us use a random change of time in the definition of v^α:

$$v^\alpha(x) = \int_0^\infty \mathbb{E}_x (f^\alpha(X^\alpha_t) - \rho^\alpha)\,dt = \int_0^\infty \mathbb{E}_x \bar f^\alpha(\bar X^\alpha_s)\,ds, \tag{2.11}$$

where

$$\bar f^\alpha(x) = \frac{f^\alpha(x) - \rho^\alpha}{a^\alpha(x)},$$

and $\bar X^\alpha_s$ is the process X^α_t with a changed time which makes the diffusion coefficient equal to one:

$$\bar X^\alpha_t := X^\alpha_{t'(t)},$$

where the function $t'(t)$ is the inverse to the mapping

$$t \mapsto \int_0^t \sigma^2(X_s^\alpha)\, ds,$$

see [23, Chapter 2.5], or [10, Theorem 15.5]. The process $\bar X_t^\alpha$ satisfies an SDE

$$d\bar X_t^\alpha = d\bar W_t + \bar b^\alpha(\bar X_t^\alpha)dt, \quad \bar b^\alpha(x) = \frac{b^\alpha(x)}{\sigma^2(\alpha(x), x)}, \tag{2.12}$$

with a new Wiener process $\bar W_t = \displaystyle\int_0^{t'(t)} \sigma(\alpha(X_s^\alpha), X_s^\alpha)\, dW_s$, see the same references [23, Chapter 2.5], or [10, Theorem 15.5].

Further, it follows from (2.11) and (2.12) that the function v^α is a solution of the equation

$$\bar L^\alpha v(x) + \bar f^\alpha(x) = 0, \tag{2.13}$$

where

$$\bar L^\alpha(x) = \bar b(\alpha(x), x)\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2}, \quad x \in \mathbb{R}.$$

Moreover, the last integral in (2.11) can only converge if $\langle \bar f^\alpha, \bar\mu^\alpha\rangle = 0$, where $\bar\mu^\alpha$ is the unique invariant measure of the Markov diffusion $\hat X_t^\alpha$, since otherwise the integral in the right hand side of (2.11) diverges. Existence and uniqueness of such an invariant measure (along with a convergence rate) follows, for example, from [31, Theorem 5] (among many other possible references) due to the assumption (A1). The property $v^\alpha \in W^2_{p,loc}$ for any $p \ge 1$ and the bound

$$\sup_\alpha |(v^\alpha)'(x)| \le C(1 + |x|^m)$$

for some $m > 0$ follow both from [27, Theorem 1] due to the equation (2.13).

Further, given (2.8), the bound $v^\alpha \in C^{1,Lip}$ (which means a local, not global Lipschitz condition for $(v^\alpha)'$) follows from the equation (2.13), as $(v^\alpha)''$ turns out to be locally bounded by virtue of this equation. The same equation(2.13) implies (2.9) and (2.10). Uniqueness of solution for the equation (2.13) and, hence, also for (2.9) up to an additive constant follows from [27]; see also [13, Lemma 4.13 and Remark 4.3]. Finally, the last assertion of the Lemma is due to the Fubini theorem,

$$\int v^\alpha(x)\mu^\alpha(dx) = \int\int_0^\infty \mathbb{E}_x(f^\alpha(X_t^\alpha) - \rho^\alpha)\, dt\mu^\alpha(dx) = \int_0^\infty \int \mathbb{E}_x(f^\alpha(X_t^\alpha) - \rho^\alpha)\mu^\alpha(dx)\, dt = 0,$$

by virtue of the absolute convergence

$$\int\int_0^\infty |E_x(f^\alpha(X_t^\alpha) - \rho^\alpha)|\, dt\mu^\alpha(dx) < \infty.$$

∎

LEMMA 2.5. *Let the assumptions (A1) – (A2) hold true. Then $\exists\ 0 < C_1 < C_2$ such that for any strategy α for the constant C_α from (2.7) we have,*

$$C_1 \le C_\alpha \le C_2.$$

Also, for any k there is a constant C such that for every x uniformly in α

$$p^\alpha(x) \le \frac{C}{1+|x|^k},$$

and there exist constants $c, \kappa > 0$ such that uniformly in α

$$p^\alpha(x) \ge c\exp(-\kappa|x|).$$

Proof. Follows straightforwardly from the recurrence and boundedness assumptions and from the formula (2.7).

3. Main results. We accept in this section that a solution of the SDE with any Markov strategy exists and is a *weak* solution. However, it is important in the proof that it is unique in distribution, strong Markov and Markov ergodic; repeat what was already mentioned in the proof of the Lemma 2.1, that all of these follow from [16] and from the assumptions (A1) and (A2) (see [31] about ergodicity).

For any pair $(v,\rho):\ v\in\bigcap_{p>1}W^2_{p,loc}$, $\rho\in\mathbb{R}$, define

$$F[v,\rho](x) := \inf_{u\in U}\left[L^u v(x) + f^u(x) - \rho\right], \quad G[v](x) := \inf_{u\in U}\left[L^u v(x) + f^u(x)\right],$$

and

$$F_1[v',\rho](x) := \inf_{u\in U}[\hat b^u v' + \hat f^u - \hat\rho](x),$$

where

$$a^u(x) = \frac{1}{2}(\sigma^u(x))^2, \quad \hat b^u(x) = b^u(x)/a^u(x),$$
$$\hat f^u(x) = f^u(x)/a^u(x), \quad \hat\rho^u(x) = \rho/a^u(x).$$

The functions v and v' may be regarded as continuous and absolutely continuous due to the embedding theorems [19]. The function $F[v,\rho](\cdot)$ is defined by the formula above as a function of the class $L_{p,loc}$ for any $p>1$; in particular, it is Lebesgue measurable and as such it is defined only a.e. with respect to x. We may and will use a (any) Borel measurable version of the function $F[v,\rho]$, the existence of which follows, for example, from Luzin's Theorem [21]). It will be shown in the sequel that the function $F_1[v',\rho](x)$ is continuous in x and locally Lipschitz in the two other variables.

Let us recall what a reward improvement algorithm (RIA) is. We start with some (any) stationary strategy $\alpha_0\in\mathcal{A}$. Denote the corresponding cost, the invariant measure, and the auxiliary function $\rho_0=\rho^{\alpha_0}=\langle f^{\alpha_0},\mu^{\alpha_0}\rangle$, and $v_0=v^{\alpha_0}$. If for some $n=0,1,\dots$ the triple (α_n,ρ_n,v_n) is determined, then the strategy α_{n+1} is defined as follows: for a.e. x the function α_{n+1} is chosen so that for each x

$$L^{\alpha_{n+1}}v_n(x) + f^{\alpha_{n+1}}(x) = G[v_n](x), \tag{3.1}$$

or, in other words,

$$\alpha_{n+1}(x)\in \mathrm{Argmin}_{u\in U}\left[L^u v_n(x) + f^u(x)\right].$$

We assume that a Borel measurable version of such strategy may be chosen; see the reference in the Appendix. To this strategy α_{n+1} there correspond the unique invariant measure $\mu^{\alpha_{n+1}}$, the value $\rho_{n+1} := \langle f^{\alpha_{n+1}},\mu^{\alpha_{n+1}}\rangle$, and the function $v_{n+1}=v^{\alpha_{n+1}}$.

THEOREM 3.1. *Let the assumptions (A1) – (A4) be satisfied. Then:*

1. For any n, $\rho_{n+1} \le \rho_n$, *and there exists a limit* $\rho_n \downarrow \tilde\rho$.

2. The sequence (v_n) *is tight in* $C^1[-N,N]$ *for each* $N>0$, *and there exists a bounded sequence of constants* β_n *such that there exists a limit* $\lim_n (v_n(x)+\beta_n) =: \tilde v(x)$.

3. The couple $(\tilde v, \tilde\rho)$ *solves the equation (1.7).*

4. This solution $(\tilde v, \tilde\rho)$ *is unique – up to an additive constant for* $\tilde v$ *– in the class of functions growing no faster than some (any) polynomial and belonging to the class* $W^2_{p,loc}$ *for any* $p>0$ *for the first component and for* $\tilde\rho \in \mathbb{R}$.

5. The component $\tilde\rho$ *in the couple* $(\tilde v,\tilde\rho)$ *coincides with* ρ.

6. Under the additional assumption (A5), $\tilde v'' \in Lip_{loc}$.

In the short presentation [1], beside the restrictive assumption $f\in[0,1]$ and maximisation instead of minimisation, only a sketch of the proof was offered with many details explained too briefly; uniqueness of $\tilde v$ was not addressed. Here the full proof is given. NB: We never compare the trajectories of two SDE solutions in one formula and the processes corresponding to different strategies may be defined on different probability spaces.

Proof. **1**. Due to (3.1) and (2.9), for almost every (a.e.) $x\in\mathbb{R}$,

$$\rho_n = L^{\alpha_n} v_n(x) + f^{\alpha_n}(x) \ge G[v_n](x) = L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x)$$

and also for a.e. $x\in\mathbb{R}$,

$$\rho_{n+1} = L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x)$$

So,

$$\begin{aligned}\rho_n - \rho_{n+1} &\overset{a.e.}{\ge} (L^{\alpha_{n+1}} v_n + f^{\alpha_{n+1}})(x) - (L^{\alpha_{n+1}} v_{n+1} + f^{\alpha_{n+1}})(x) \\ &= (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(x).\end{aligned} \tag{3.2}$$

Let us apply Ito – Krylov's formula (see [15]) with expectations (also known as Dynkin's formula) to $(v_n - v_{n+1})(X_t^{\alpha_{n+1}})$: we have for any $x\in\mathbb{R}$,

$$\begin{aligned}&\mathbb{E}_x\left(v_n(X_t^{\alpha_{n+1}}) - v_{n+1}(X_t^{\alpha_{n+1}})\right) - (v_n - v_{n+1})(x) \\ &= \mathbb{E}_x \int_0^t (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(X_s^{\alpha_{n+1}})\,ds \le \mathbb{E}_x \int_0^t (\rho_n - \rho_{n+1})\,ds = (\rho_n-\rho_{n+1})\,t.\end{aligned} \tag{3.3}$$

The equality in the equation (3.3) holds for all $x\in\mathbb{R}$ and not just a.e. since the functions v_n are Sobolev solutions of Poisson equations locally integrable in any degree with their derivatives up to the second order. Such functions can be regarded as continuous due to the embedding theorems [19]. In addition, the functions $\mathbb{E}_x v_n(X_t^{\alpha_{n+1}})$, $\mathbb{E}_x v_{n+1}(X_t^{\alpha_{n+1}})$, and $\mathbb{E}_x \int_0^t (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(X_s^{\alpha_{n+1}})\,ds$ as functions of x for each $t>0$ are all Hölder continuous, being solutions of non-degenerate parabolic equations [18]. We also used the fact that the distribution of $X_s^{\alpha_{n+1}}$ for almost all $s>0$ is absolutely continuous with respect to the Lebesgue measure due to the non-degeneracy and by virtue of Krylov's estimates [15]; due to this reason and because $v_n, v_{n+1}\in C$, the a.e. inequality (3.2) implies (3.3) for every x. Further, since the left hand side in (3.3) is bounded for a fixed x by virtue of the Lemma 2.4, we divide all terms of the latter inequality by t and let $t \to \infty$ to get,

$$0 \le \rho_n - \rho_{n+1},$$

as required. Thus, $\rho_n \ge \rho_{n+1}$, so that $\rho_n \downarrow \tilde\rho$ (since the sequence ρ_n is bounded for $f \in \mathcal{K}$, see (2.3) in the Lemma 2.1) with some $\tilde\rho$. So, the RIA does converge.

Note that clearly $\tilde\rho \ge \rho$, since ρ is the infimum over all Markov strategies, while $\tilde\rho$ is the infimum over some countable subset of them. Later on we shall show that they do coincide.

Now we want to show that there exists a bounded sequence of real values (non-random!) $\{\beta_n\}$ such that $v_n + \beta_n \to \tilde v$, so that the couple $(\tilde v, \tilde\rho)$ satisfies the equation (1.7), and that $\tilde\rho$ here is unique, as well as $\tilde v$ in some sense. In the first instance we will do it for some subsequence n_j; eventually the convergence of the whole sequence v_n will follow from the uniqueness of the solution of Bellman's equation, although, it is not important for the proof of the Theorem.

2. Let us show local tightness of the family of functions (v_n) in C^1. Note that the equation (1.7) is equivalent to the following:

$$V''(x) + \inf_{u\in U}\left[\frac{b(u,x)}{a(u,x)}V'(x) + \frac{f(u,x)}{a(u,x)} - \frac{\rho}{a(u,x)}\right] = 0, \qquad (3.4)$$

while the equation

$$L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x) - \rho_{n+1} \overset{a.e.}{=} 0 \qquad (3.5)$$

is equivalent to

$$v''_{n+1}(x) + \frac{b(\alpha_{n+1}(x),x)}{a(\alpha_{n+1}(x),x)}v'_{n+1}(x) + \frac{f(\alpha_{n+1}(x),(x))}{a(\alpha_{n+1}(x),x)} - \frac{\rho_{n+1}}{a(\alpha_{n+1}(x),x)} = 0.$$

According to the Lemma 2.4, the functions v'_{n+1} are uniformly locally bounded. Since the sequence ρ_{n+1} is bounded and due to the uniform local boundedness of the functions $f(\alpha_{n+1}(x),x)$ and uniform nondegeneracy of a, it follows that (v''_n) locally are uniformly bounded and satisfy the uniform in n growth bounds similar to (2.8) for the function itself and for its first derivative due to the equation (for example, due to (3.4)). This guarantees compactness of (v_n) in C^1 locally.

3. Due to the (local) compactness property showed in the previous step, by the diagonal procedure from any infinite sub-family of functions v_n it is possible to choose a converging in C^1_{loc} subsequence. We want to show that up to a constant the limit is unique. For this aim, first of all we shall see shortly that if some $v_{n_j}(x)$ has a limit as $n_j \to \infty$, say, $\tilde v(x)$ (locally in C) then $v_{n_j+1}(x) + \beta_{n_j}$ has the same limit, where β_n is some bounded sequence of real values. (In fact, what will be established is a little bit more complicated but still enough for our purposes.) We have,

$$L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x) - \rho_{n+1} \overset{a.e.}{=} 0,$$

and

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n =: -\psi_{n+1}(x) \overset{a.e.}{\le} 0. \qquad (3.6)$$

Let us rewrite it as follows,

$$L^{\alpha_{n+1}}v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n + \psi_{n+1}(x) \overset{a.e.}{=} 0.$$

In other words, the function v_n solves the Poisson equation with the second order operator $L^{\alpha_{n+1}}$ and the "right hand side" $-(f^{\alpha_{n+1}}(x) + \psi_{n+1}(x) - \rho_n)$. This is only possible if the expression $f^{\alpha_{n+1}}(x) + \psi_{n+1}(x) - \rho_n$ is centered with respect to the invariant measure μ^{n+1} because Poisson equations in the whole space have no solutions for non-centered right hand sides (see, for example, [27]). This implies that

$$\langle f^{\alpha_{n+1}}(x) + \psi_{n+1} - \rho_n, \mu^{n+1}\rangle = 0$$

So,

$$\langle \psi_{n+1}, \mu^{n+1}\rangle = \rho_n - \rho_{n+1}. \tag{3.7}$$

Now denote

$$w_n(x) := v_n(x) - v_{n+1}(x).$$

We have,

$$L^{\alpha_{n+1}}w_n(x) + \psi_{n+1}(x) - (\rho_n - \rho_{n+1}) \overset{a.e.}{=} 0.$$

So, there is a constant $\beta_n = \langle w_n, \mu^{n+1}\rangle$ such that

$$w_n(x) - \beta_n = \int_0^\infty \mathbb{E}_x(\psi_{n+1}(X_t^{n+1}) - (\rho_n - \rho_{n+1}))\, dt. \tag{3.8}$$

Let us show that for any $N > 0$,

$$\int_{-N}^N \psi_n^2(x)\, dx \to 0, \quad n \to \infty. \tag{3.9}$$

First of all, note that all functions ψ_n and, hence, ψ_n^2 are uniformly locally bounded and may only grow polynomially fast,

$$(0 \le\ )\ \psi_n(x) \le C(1 + |x|^m), \tag{3.10}$$

with some C, m the same for all values of n. which follows from the definition (3.6), and the properties of derivatives v_n' and v_n'', and from the Lemma 2.5, and due to

$$\langle \psi_{n+1}, \mu^{n+1}\rangle = \rho_n - \rho_{n+1} \to 0, \quad n \to \infty.$$

Now let us rewrite the equation (3.8) via a stationary version of our diffusion, say, $\tilde X_t^{n+1}$:

$$w_n(x) - \beta_n = \int_0^\infty \mathbb{E}_x(\psi_{n+1}(X_t^{n+1}) - E_{\mu^{n+1}}(\psi_{n+1}(\tilde X_t^{n+1}))\, dt.$$

(Note that if we knew that w_n were centered with respect to the invariant measure μ^{n+1} then we would have $\beta_n = 0$; however, the functions v_n and v_{n+1} are both centered with respect to two different measures, and this is the reason why their difference is not just small, but small up to some additive constant; this very constant is denoted by β_n.) Using the coupling idea (see, for example, [31]), let us consider the independent processes X_t^{n+1} and $\tilde X_t^{n+1}$ on the same probability space (just considering the product space) and denote the moment of the first meeting

$$\tau := \inf(t \ge 0:\ X_t^{n+1} = \tilde X_t^{n+1}).$$

It is known (see [31, Theorem 5]) that under our recurrence assumptions for any $k > 0$ there are some constants C_k, m such that uniformly with respect to n,

$$\mathbb{E}_{x,\mu^{n+1}} \tau^k \le C_k (1 + |x|^m).$$

Denote

$$\hat X_t^{n+1} := 1(t < \tau) X_t^{n+1} + 1(t \ge \tau) \tilde X_t^{n+1}.$$

Since τ is a stopping time and because the couple $(X_t^{n+1}, \tilde X_t^{n+1})$ is strong Markov (see [14]), the process $(\hat X_t^{n+1})$ is also strong Markov equivalent to (X_t^{n+1}). Therefore, it is possible to rewrite,

$$w_n(x) - \beta_n = \int_0^\infty \mathbb{E}_{x,\mu} (\psi_{n+1}(\hat X_t^{n+1}) - \psi_{n+1}(\tilde X_t^{n+1}))\, dt.$$

Hence, using the fact that after τ the processes $\hat X_t^{n+1}$ and $\tilde X_t^{n+1}$ coincide, we obtain

$$\begin{aligned}
w_n(x) - \beta_n &= \int_0^\infty \mathbb{E}_{x,\mu} 1(t < \tau)(\psi_{n+1}(\hat X_t^{n+1}) - \psi_{n+1}(\tilde X_t^{n+1}))\, dt \\
&= \int_0^\infty \mathbb{E}_{x,\mu} \sum_{i=0}^\infty 1(i \le \tau < i+1) 1(t < \tau)(\psi_{n+1}(\hat X_t^{n+1}) - \psi_{n+1}(\tilde X_t^{n+1}))\, dt \\
&= \sum_{i=0}^\infty \mathbb{E}_{x,\mu} \int_0^\infty 1(i \le \tau < i+1) 1(t < \tau)(\psi_{n+1}(\hat X_t^{n+1}) - \psi_{n+1}(\tilde X_t^{n+1}))\, dt.
\end{aligned}$$

Thus, using Cauchy-Buniakovsky-Schwarz inequality and Fubini Theorem, we have,

$$|w_n(x) - \beta_n| \le \sum_{i=0}^{\infty} \mathbb{E}_{x,\mu} \int_0^{i+1} 1(\tau > i) |\psi_{n+1}(\hat{X}_t^{n+1}) - \psi_{n+1}(\tilde{X}_t^{n+1})| \, dt$$

$$\le \sum_{i=0}^{\infty} \int_0^{i+1} \mathbb{E}_{x,\mu} 1(\tau > i)(|\psi_{n+1}(\hat{X}_t^{n+1})| + |\psi_{n+1}(\tilde{X}_t^{n+1})|) \, dt$$

$$\le \sum_{i=0}^{\infty} \int_0^{i+1} (\mathbb{E}_{x,\mu} 1(\tau > i))^{1/2} (\mathbb{E}_{x,\mu}(|\psi_{n+1}(\hat{X}_t^{n+1})| + |\psi_{n+1}(\tilde{X}_t^{n+1})|)^2)^{1/2} \, dt$$

$$\le 2 \sum_{i=0}^{\infty} (\mathbb{E}_{x,\mu} 1(\tau > i))^{1/2} \int_0^{i+1} (\mathbb{E}_{x,\mu} |\psi_{n+1}(\hat{X}_t^{n+1})|^2 + \mathbb{E}_{x,\mu} |\psi_{n+1}(\tilde{X}_t^{n+1})|^2)^{1/2} \, dt$$

$$\le 2 \sum_{i=0}^{\infty} (\mathbb{E}_{x,\mu} 1(\tau > i))^{1/2} \int_0^{i+1} [(\mathbb{E}_{x,\mu}(\psi_{n+1}(\hat{X}_t^{n+1}))^2)^{1/2} + (\mathbb{E}_{x,\mu}(\psi_{n+1}(\tilde{X}_t^{n+1}))^2)^{1/2}] \, dt.$$

Now, let us take any $\epsilon > 0$ and use the inequality $\sqrt{a} \le \frac{\epsilon}{2} + \frac{a}{2\epsilon}$. We estimate,

$$\int_0^{i+1} [(\mathbb{E}_{x,\mu}(\psi_{n+1}(\hat{X}_t^{n+1}))^2)^{1/2} + (\mathbb{E}_{x,\mu}(\psi_{n+1}(\tilde{X}_t^{n+1}))^2)^{1/2}] \, dt$$

$$\le \epsilon(i+1) + \frac{1}{2\epsilon} \int_0^{i+1} [\mathbb{E}_{x,\mu} \psi_{n+1}^2(\hat{X}_t^{n+1}) + \mathbb{E}_{x,\mu} \psi_{n+1}^2(\tilde{X}_t^{n+1})] \, dt.$$

Let us first consider the stationary term. We have,

$$\frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2(\tilde{X}_t^{n+1}) \, dt + \frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2(\hat{X}_t^{n+1}) \, dt$$

$$= \frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2 1_{[-N,N]}(\tilde{X}_t^{n+1}) \, dt + \frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2 1_{\mathbb{R}\setminus[-N,N]}(\tilde{X}_t^{n+1}) \, dt$$

$$+ \frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2 1_{[-N,N]}(\hat{X}_t^{n+1}) \, dt + \frac{1}{2\epsilon} \int_0^{i+1} \mathbb{E}_{x,\mu} \psi_{n+1}^2 1_{\mathbb{R}\setminus[-N,N]}(\hat{X}_t^{n+1}) \, dt.$$

Given (3.10) and because any stationary measure integrates uniformly any power function, let us find such N that uniformly with respect to n,

$$\langle C(1 + |x|^{2m}) 1_{\mathbb{R}\setminus[-N,N]}, \mu^{n+1} \rangle < \epsilon^2/2, \tag{3.11}$$

which is possible due to the Lemmata 2.1 and 2.5, and also such that $N > \epsilon^{-2}$. Then choose $n(\epsilon)$ such that

$$\sup_{n\geq n(\epsilon)} \int_{|x|\leq N} \psi_n^2(x)\,dx < \epsilon^2/2.$$

Due to Krylov's estimate

$$\mathbb{E}\int_0^T g(\tilde X_t^{n+1})\,dt \leq K_T \|g\|_{L_1(\mathbb{R})}$$

for any function $g \geq 0$, and also

$$\mathbb{E}\int_s^{s+T} g(\tilde X_t^{n+1})\,dt \leq K_T \|g\|_{L_1(\mathbb{R})}$$

for any $s > 0$ (follows from [15, Theorem 2.2.3]), we evaluate with $n \geq n(\epsilon)$:

$$\frac{1}{2\epsilon}\int_0^{i+1} \mathbb{E}_{x,\mu}\psi_{n+1}^2 1_{[-N,N]}(\tilde X_t^{n+1})\,dt$$

$$= \frac{1}{2\epsilon}\sum_{k=0}^{i} \mathbb{E}_x \int_k^{k+1} \psi_{n+1}^2 1_{[-N,N]}(\tilde X_t^{n+1})\,dt \leq \frac{i+1}{2\epsilon} K \|\psi_{n+1}^2 1_{[-N,N]}\|_{L^1} \leq \frac{(i+1)K\epsilon}{2}.$$

Indeed, for any $k \geq 0$ we have,

$$\mathbb{E}_x \int_k^{k+1} \psi_{n+1}^2 1_{[-N,N]}(\tilde X_t^{n+1})\,dt = \mathbb{E}_x\mathbb{E}_x\left(\int_k^{k+1} \psi_{n+1}^2 1_{[-N,N]}(\tilde X_t^{n+1})\,dt|\mathcal{F}_k\right)$$

$$= \mathbb{E}_x\mathbb{E}_{\tilde X_k^{n+1}}\int_0^1 \psi_{n+1}^2 1_{[-N,N]}(\tilde X_t^{n+1})\,dt$$

$$\leq \frac{1}{2\epsilon} K \|\psi_{n+1}^2 1_{[-N,N]}\|_{L^2} \leq \frac{(i+1)K\epsilon}{2}.$$

This argument works for the non-stationary process as well: due to the same Krylov's estimate,

$$\frac{1}{2\epsilon}\int_0^{i+1} \mathbb{E}_{x,\mu}\psi_{n+1}^2 1_{[-N,N]}(\hat X_t^{n+1})\,dt$$

$$= \frac{1}{2\epsilon}\sum_{k=0}^{i} \mathbb{E} \int_k^{k+1} \psi_{n+1}^2 1_{[-N,N]}(\hat X_t^{n+1})\,dt \leq \frac{i+1}{2\epsilon} K \|\psi_{n+1}^2 1_{[-N,N]}\|)_{L^1} \leq \frac{(i+1)K\epsilon}{2}.$$

Further,

$$\frac{1}{2\epsilon}\int_0^{i+1} \mathbb{E}_{x,\mu}\psi_{n+1}^2 1_{\mathbb{R}\setminus[-N,N]}(\tilde X_t^{n+1})\,dt \leq \frac{i+1}{2\epsilon}\times\frac{\epsilon^2}{2} = \frac{(i+1)\epsilon}{4}.$$

Finally, using (2.6), we obtain with some m,

$$\frac{1}{2\epsilon}\int_0^{i+1} \mathbb{E}_{x,\mu}\psi^2_{n+1}1_{\mathbb{R}\setminus[-N,N]}(\hat X^{n+1}_t)\,dt = \frac{1}{2\epsilon}\int_0^{i+1} \mathbb{E}_{x,\mu}\psi^2_{n+1}1_{\mathbb{R}\setminus[-N,N]}(X^{n+1}_t)\,dt$$

$$\le C\frac{i+1}{2\epsilon}\frac{(1+|x|^m)}{N} \le C(i+1)(1+|x|^m)\epsilon.$$

Overall, this shows that with the appropriately chosen (uniformly bounded) β_n,

$$|w_n(x)-\beta_n| \le C(1+|x|^{2m})\epsilon\sum_{i=0}^{\infty}(i+1)(\mathbb{E}_{x,\mu}1(\tau>i))^{1/2}, \quad n\ge n(\epsilon). \tag{3.12}$$

By virtue of the results in [31], for any $k>0$ there are $C,m>0$ such that

$$\mathbb{P}_{x,\mu}1(\tau>i) \le C\frac{1+|x|^m}{1+i^k}.$$

Therefore, taking any $k>1$, we have that the series in (3.12) converges providing us an estimate

$$|w_n(x)-\beta_n| \le C(1+|x|^{3m})\epsilon, \quad n\ge n(\epsilon). \tag{3.13}$$

In other words, the difference $w_n(x)-\beta_n = v_n - v_{n+1} - \beta_n$ is locally uniformly converging to zero as $n\to\infty$. Naturally, it also implies that for any subsequence n_j such that v_{n_j} converges locally uniformly in C^1 we have that v'_{n_j} and v'_{n_j+1} may only converge to the same limit, i.e., derivatives $v'_{n_j} - v'_{n_j+1}\to 0$ (locally uniformly) as $j\to\infty$. Indeed, otherwise we just integrate to show that the limits of v_{n_j} and $v_{n_j+1}+\beta_{n_j}$ are different, which contradicts to what was established earlier.

4. What we want to do now is to pass to the limit as $j\to\infty$ in the equations

$$L^{\alpha_{n_j+1}}v_{n_j+1}(x) + f^{\alpha_{n_j+1}}(x) - \rho_{n_j+1} \overset{a.e.}{=} 0, \quad \& \quad G[v_{n_j}](x) - \rho_{n_j} \le 0,$$

where $(n_j, j\to\infty)$ is any sequence such that v_{n_j} converges (locally uniformly) in C^1. From

$$G[v_{n_j}](x) - \rho_{n_j} = L^{\alpha_{n_j+1}}v_{n_j}(x) + f^{\alpha_{n_j+1}}(x) - \rho_{n_j}$$

$$(= \inf_{u\in U}[L^u v_{n_j}(x) + f^u(x) - \rho_{n_j}] \overset{a.e.}{\le} 0),$$

by subtracting zero a.e. (3.5), we obtain a.e.,

$$G[v_{n_j}](x) - \rho_{n_j} = L^{\alpha_{n_j+1}}(v_{n_j}(x) - v_{n_j+1}(x)) - (\rho_{n_j} - \rho_{n_j+1}). \tag{3.14}$$

Now we want to show that

$$\tilde v'(x) - \tilde v'(r) + \int_r^x F_1[s,\tilde v'(s),\tilde\rho]\,ds = 0, \tag{3.15}$$

which in turn implies by differentiation the equation equivalent to (1.7),

$$\tilde v''(x) + F_1[x,\tilde v',\tilde\rho](x) = 0, \tag{3.16}$$

for any x, with the note that $\tilde v'$ is absolutely continuous.

Let us show that (3.14), indeed, implies (3.15). Note that $G[v_{n_j}](x) - \rho_{n_j} \le 0$ (a.e.). Let us divide (3.14) by $a_{n_j+1} = a^{\alpha_{n_j+1}}$ and use $\delta := \inf_{u,x} a^u(x) > 0$: we get a.e. with some $K > 0$,

$$
\begin{aligned}
0 \ge \frac{(G[v_{n_j}](x) - \rho_n)}{a_{n_j+1}} &= (v''_{n_j}(x) - v''_{n_j+1}(x)) + (\hat b^{\alpha_{n_j+1}}(v'_{n_j} - v'_{n_j+1})) - \frac{(\rho_{n_j} - \rho_{n_j+1})}{a_{n_j+1}} \\
&\ge (v''_{n_j}(x) - v''_{n_j+1}(x)) - \frac{K}{\delta}|v'_{n_j}(x) - v'_{n_j+1}(x)| - \frac{1}{\delta}(\rho_{n_j} - \rho_{n_j+1}). \qquad (3.17)
\end{aligned}
$$

So, we have just shown that a.e.,

$$0 \ge (v''_{n_j}(x) - v''_{n_j+1}(x)) - \frac{K}{\delta}|v'_{n_j}(x) - v'_{n_j+1}(x)| - \frac{\rho_{n_j} - \rho_{n_j+1}}{\delta}. \tag{3.18}$$

The next trick is to note that again due to (3.17) and $\rho_{n_j} \ge \rho_{n_j+1}$, and since $\delta \le a \le C$,

$$0 \overset{a.e.}{\ge} G[v_{n_j}](x) - \rho_{n_j} \ge a_{n_j+1}(v''_{n_j} - v''_{n_j+1})(x) - C'|v'_{n_j} - v'_{n_j+1}|(x) - (\rho_{n_j} - \rho_{n_j+1}),$$

which implies that with some $C, c > 0$,

$$0 \overset{a.e.}{\ge} v''_{n_j} + F_1[v'_{n_j}, \rho_{n_j}] \ge ((v''_{n_j} - v''_{n_j+1}) - C|v'_{n_j} - v'_{n_j+1}|) - c(\rho_{n_j} - \rho_{n_j+1}). \tag{3.19}$$

Since v'_{n_j} is absolutely continuous, we can integrate (3.19) to get the following: for any (not a.e.!) x and r with $x > r$,

$$
\begin{aligned}
0 &\ge v'_{n_j}(x) - v'_{n_j}(r) + \int_r^x F_1[v'_{n_j}(s), \rho_{n_j}](s)\, ds \\
&= \int_r^x \Big(v''_{n_j}(x) + F_1[v'_{n_j}(s), \rho_{n_j}](s)\Big)\, ds \\
&\ge \int_r^x ((v''_{n_j} - v''_{n_j+1})(s) - C|v'_{n_j} - v'_{n_j+1}|(s) - c(\rho_{n_j} - \rho_{n_j+1}))\, ds \qquad (3.20) \\
&= v'_{n_j}(x) - v'_{n_j}(r) - v'_{n_j+1}(x) + v'_{n_j+1}(r) \\
&\quad - C\int_r^x |v'_{n_j} - v'_{n_j+1}|(s) ds - c(\rho_{n_j} - \rho_{n_j+1})(x - r).
\end{aligned}
$$

As it was explained earlier, due to the compactness in C^1 we may assume that

$$v_{n_j} \to \tilde v, \quad v'_{n_j} \to \tilde v', \quad \& \quad v'_{n_j+1} \to \tilde v', \quad j \to \infty,$$

in C locally for some $\tilde v \in C^1$, as $j \to \infty$. Note that $\tilde v'$ is absolutely continuous, which follows from the uniform local boundedness of v''_n. Therefore, it is possible to get to the limit in the inequality (3.20) as $j \to \infty$: for any $x > r$,

$$0 \ge \tilde v'(x) - \tilde v'(r) + \lim_{j\to\infty} \int_r^x F_1[s, v'_{n_j}(s), \rho_{n_j}]\, ds \ge 0,$$

since the right hand side in (3.20) clearly goes to zero.

Here

$$F_1[v'_{n_j}(s),\rho_{n_j}](s) = \inf_u \left[\frac{b^u}{a^u}v'_{n_j}(s) + \frac{f^u}{a^u}(s) - \frac{\rho_{n_j}}{a^u}(s)\right]$$

$$\to \inf_{u\in U}\left[\frac{b^u}{a^u}\tilde v'(s) + \frac{f^u}{a^u}(s) - \frac{\rho_n}{a^u}(s)\right] = F_1[\tilde v'(s),\tilde\rho](s), \quad j\to\infty.$$

So, from (3.20) we obtain the desired equation (3.15)

$$\tilde v'(x) - \tilde v'(r) + \int_r^x F_1[s,\tilde v'(s),\tilde\rho]\,ds = 0.$$

In turn, since $F_1[\tilde v'(s),\tilde\rho](s)$ is continuous and absolutely continuous in s, it implies $\tilde v\in C^2$, and by (well-defined) differentiation we get the equation (3.16) for every $x\in\mathbb{R}$.

In the sequel it will follow from the uniqueness of solution to the Bellman's equation that actually the whole sequence v_n converges up to an additive constant sequence locally uniformly in C^1 to a single limit. However, it is not needed for our proof.

5. *Uniqueness for ρ in (1.7).* Assume that there are two solutions of the (HJB) equation, (v^1,ρ^1) and (v^2,ρ^2) with $v^i\in\mathcal{K}$, $i=1,2$:

$$\inf_{u\in U}(L^u v^1(x) + f^u(x) - \rho^1) = \inf_{u\in U}(L^u v^2(x) + f^u(x) - \rho^2) = 0.$$

Earlier it was shown that both v^1 and v^2 are classical solutions with locally Lipschitz second derivatives. Let $w(x) := v^1(x) - v^2(x)$ and consider two strategies $\alpha_1,\alpha_2\in\mathcal{A}$ such that $\alpha_1(x)\in \operatorname{Argmax}_{u\in U}(L^u w(x))$ and $\alpha_2(x)\in\operatorname{Argmin}_{u\in U}(L^u w(x))$, and let X_t^1, X_t^2 be solutions of the SDEs corresponding to each strategy α_i, $i=1,2$. Note that due to the measurable choice arguments – see the Appendix – such Borel strategies exist; corresponding weak solutions also exist. Let us denote

$$h_1(x) := \sup_{u\in U}(L^u w(x) - \rho^1 + \rho^2), \quad h_2(x) := \inf_{u\in U}(L^u w(x) - \rho^1 + \rho^2).$$

Then,

$$h_2(x) = \inf_{u\in U}(L^u v^1(x) + f^u(x) - \rho^1 - (L^u v^2(x) + f^u(x) - \rho^2))$$

$$\le \inf_{u\in U}(L^u v^1(x) + f^u(x) - \rho^1) - \inf_{u\in U}(L^u v^2(x) + f^u(x) - \rho^2) = 0.$$

Similarly,

$$h_1(x) = -\inf_u(L^u(-v^2)(x) - \rho^2 + \rho^1)$$

$$= -\inf_u(L^u v^2(x) + f^u(x) + \rho^2 - (L^u v^1(x) + f^u(x) + \rho^1))$$

$$\ge -\left[\inf_u(L^u v^2(x) + f^u(x) - \rho^2) - \inf_u(L^u v^1(x) + f^u(x) - \rho^1)\right] = 0.$$

We have,

$$L^{\alpha_2} w(x) = h_2(x) - \rho^2 + \rho^1,$$

and

$$L^{\alpha_1} w(x) = h_1(x) - \rho^2 + \rho^1.$$

Due to Dynkin's formula we have,

$$\mathbb{E}_x w(X_t^1) - w(x) = \mathbb{E}_x \int_0^t L^{\alpha_1} w(X_s^1)\, ds$$

$$= \mathbb{E}_x \int_0^t h_1(X_s^1)\, ds + (\rho^1 - \rho^2)\, t \stackrel{(h_1 \ge 0)}{\ge} (\rho^1 - \rho^2)\, t.$$

Since the left hand side here is bounded for a fixed x, due to the Lemma 2.1 we get,

$$\rho^1 - \rho^2 \le 0.$$

Similarly, considering α_2 we conclude that

$$E_x w(X_t^2) - w(x) = E_x \int_0^t L^{\alpha_2} w(X_s^2)\, ds$$

$$= E_x \int_0^t h_2(X_s^2)\, ds + (\rho^1 - \rho^2)\, t.$$

From here, due to the boundedness of the left hand side (Lemma 2.1) we get,

$$\rho^2 - \rho^1 = \liminf_{t\to 0} (t^{-1} E_x \int_0^t h_2(X_s^2)\, ds) \stackrel{(h_2 \le 0)}{\le} 0.$$

Thus, $\rho^1 - \rho^2 \ge 0$ and, hence,

$$\rho^1 = \rho^2.$$

6. *Why $\rho = \tilde\rho$?* Recall that for any initial $\alpha_0 \in \mathcal{A}$, the sequence ρ_n converges to the same value $\tilde\rho$, which is a unique component of solution of the equation (1.7). Let us take any $\epsilon > 0$ and consider a strategy α_0 such that

$$\rho_0 = \rho^{\alpha_0} < \rho + \epsilon.$$

Since the sequence (ρ_n) decreases, the limit $\tilde\rho$ must satisfy the same inequality,

$$\tilde\rho = \lim_{n\to\infty} \rho_n < \rho + \epsilon.$$

Due to uniqueness of $\tilde\rho$ as a component of solution of the equation (1.7) and since $\epsilon > 0$ is arbitrary, we find that

$$\tilde\rho \le \rho.$$

But also $\tilde\rho \ge \rho$ since $\tilde\rho$ is the infimum of the cost function values over a smaller – just countable – family of strategies. So, in fact,

$$\tilde\rho = \rho.$$

7. *Uniqueness for V*. Let us have another look at the earlier equations in the step 6, replacing $\rho^2 - \rho^1$ by zero as we already know that the second component in the solution is unique:

$$\mathbb{E}_x w(X_t^1) - w(x) = \mathbb{E}_x \int_0^t h_1(X_s^1)\, ds.$$

Clearly, $h_1 \geq 0$ with $h_1 \neq 0$ – i.e., with $\Lambda(x: \ h_1(x) > 0) > 0$ – would imply that $\langle h_1, \mu_1\rangle > 0$, which contradicts a zero left hand side (after division by t with $t \to \infty$). So, we conclude that

$$h_1 = 0, \quad \mu_1 - \text{a.s.}$$

Since $\mu_1 \sim \Lambda$ due to (2.7), by virtue of Krylov's estimate we have that $0 \leq \mathbb{E}_x \int_0^t h_1(X_s^1)\, ds \leq N\|h_1\|_{L_1} = 0$. So, in fact,

$$\mathbb{E}_x w(X_t^1) - w(x) = 0. \tag{3.21}$$

Further, from (3.21) and due to the last statement of the Lemma 2.1 it follows that

$$w(x) = \lim_{t\to\infty} \mathbb{E}_x w(X_t^1) = \langle w, \mu_1\rangle.$$

Hence, $w(x)$ is a constant. Recall that uniqueness of the first component V is stated up to a constant, and it was just established that

$$v^1(x) - v^2(x) = \text{const}.$$

8. Returning to the second statement of Theorem 3.1, note that due to uniqueness of the solution of the HJB equation, convergence of the whole sequence (v_n) up to additive constants depending only on n is to the unique limit v.

9. *Local Lipschitz for $\tilde v''$*. Recall that a certain additional regularity of the coefficients is assumed. We have from (3.16) and (2.8),

$$|\tilde v''(x)| = |F_1[\tilde v'(x), \tilde\rho](x)| \leq C(1 + |\tilde v'(x)|) \leq C(1 + |x|).$$

Therefore, it follows from the Cauchy Mean Value Theorem that

$$|\tilde v'(x) - \tilde v'(x')| \leq C(1 + |x|^m + |x'|^m)|x - x'|.$$

So, due to Lipschitz condition on b^u, a^u in x and in virtue of the nondegeneracy of a^u,

$$\begin{aligned}
|\tilde v''(x) - \tilde v''(x')| &= |F_1[\tilde v'(x), \tilde\rho](x) - F_1[\tilde v'(x), \tilde\rho](x')| \\
&= |\inf_u [\hat b^u(x)\tilde v'(x) + \hat f^u(x) - \frac{\tilde\rho}{a^u(x)}] - \inf_u [\hat b^u(x')\tilde v'(x') + \hat f^u(x') - \frac{\tilde\rho}{a^u(x')}]| \\
&\leq \sup_u |\hat b^u(x)\tilde v'(x) + \hat f^u(x) - \frac{\tilde\rho}{a^u(x)} - \hat b^u(x')\tilde v'(x') - \hat f^u(x') + \frac{\tilde\rho}{a^u(x')}| \\
&\leq C\left(|\tilde v'(x) - \tilde v'(x')| + |x - x'|\right) \leq C(1 + |x|^m + |x'|^m)|x - x'|.
\end{aligned}$$

The required local Lipschitz property of the function $\tilde v''$ has been verified. ∎

Appendix A. On a measurable choice. For the reader's convenience we repeat the main arguments from [1] concerning the measurable choice a little bit more precisely. Recall that in the presentation of RIA in the beginning of the section 3 existence of a Borel measurable version of such a strategy was assumed, which minimizes some function for any fixed x. In our case existence of such a Borel strategy can be justified by using Stschegolkow's (Shchegolkov's) theorem [30] (see also [20, Satz 39], or [7, Theorem 1]). According to this result, if any *section* of a (nonempty) Borel set E in the direct product of two complete separable metric spaces is sigma-compact (i.e., equals a countable sum of closed bounded sets) then a Borel selection belonging to this set E exists.

In our case we have, $F[v,\rho](x) = \inf_{u\in U}[L^u v(x) + f^u(x) - \rho]$. For a fixed v representing any v_n in the proof, denote $\chi(u,x) := L^u v(x) + f^u(x) - \rho$ and $\bar\chi(x) := F[v,\rho](x)$, and let $E = \{(u,x) : \chi(u,x) = \bar\chi(x)\}$. This set is nonempty because the minima here are attained for each x. Its section for any $x \in \mathbb{R}$ is $E_x := \{u : \chi(u,x) = \bar\chi(x)\}$. Any such section is nonempty and closed and, hence, Borel. Indeed, if $E_x \ni u_n \to u$, $n\to\infty$, then $\chi(u_n,x) \to \chi(u,x)$ due to the continuity of $\chi(\cdot,x)$.

The set E itself is Borel, too. To show this, take any $\epsilon > 0$ and denote

$$E(\epsilon) := \{(u,x) : \chi(u,x) - \bar\chi(x) < \epsilon\}.$$

This set is Borel because the functions $\chi(u,x)$ and $\bar\chi(x\}$ are: the latter one since the minimum in $\min_u \chi(u,x)$ can be taken over some countable dense subset of U. (Recall that the second derivative v'' is Borel measurable by our convention.) It remains to note that

$$E = \bigcap_{k=1}^{\infty} E(1/k),$$

so that E is also Borel.

Thus, Stschegolkow's theorem is applicable and, hence, a Borel measurable improved strategy α_{n+1} in the induction step of the RIA does exist for each step n. By the same reason Borel strategies α_1 and α_2 exist in the steps 6 and (implicitly) 8.

REFERENCES

[1] S.V. Anulova, H. Mai, A.Yu. Veretennikov, On averaged expected cost control as reliability for 1D ergodic diffusions, *Reliability: Theory & Applications (RT&A)*, 12, 4(47), 31-38, 2017.
[2] A. Arapostathis, V.S. Borkar, E. Fernãndes-Gaucherand, M.K. Ghosh, and S.I. Markus, Discrete-time controlled Markov processes with average cost criterion: a survey, *SIAM J. Control and Optimization*, 31(2), 282-344, 1993.
[3] A. Arapostathis, On the policy iteration algorithm for nondegenerate controlled diffusions under the ergodic criterion. In: *Optimization, control, and applications of stochastic systems*, Systems Control Found. Appl., 1–12. Birkhäuser/Springer, New York, 2012.
[4] A. Arapostathis, V.S. Borkar, A relative value iteration algorithm for non-degenerate controlled diffusions, *SIAM Journal on Control and Optimization*, 50(4), 1886-1902, 2012.
[5] A. Arapostathis, V.S. Borkar, M.K. Ghosh, *Ergodic control of diffusion processes,* Encyclopedia of Mathem. and its Appl. 143. Cambridge: CUP, 2012.
[6] V.S. Borkar, *Optimal control of diffusion processes,* Harlow: Longman Scientific & Technical; New York: John Wiley & Sons, 1989.
[7] L.D. Brown, R. Purves, Measurable selections of extrema. *Ann. Stat.*, 1, 902–912, 1973.
[8] O.L. do Valle Costa, F. Dufour, *Continuous Average Control of Piecewise Deterministic Markov Processes*, Springer, New York et al., 2013.
[9] E.B. Dynkin and A.A. Yushkevich, *Upravlyaemye markovskie protsessy i ikh prilozheniya*, Moskva: "Nauka", 1975 (in Russian).
[10] I.I. Gikhman, A.V. Skorokhod, *Stochastic Differential Equations,* Berlin, Heidelberg: Springer, 1972.

[11] R.A. Howard, *Dynamic programming and Markov processes,* New York – London: John Wiley & Sons, Inc. and the Technology Press of the MIT, 1960.
[12] R.A. Howard, *Dynamic probabilistic systems. Vol. II: Semi-Markov and decision processes, reprint of the 1971 original ed.* Mineola, NY: Dover Publ., 577-1108, 2007.
[13] R. Khasminskii, *Stochastic stability of differential equations. With contributions by G.N. Milstein and M.B. Nevelson. 2nd completely revised and enlarged ed.* Berlin: Springer, 2012.
[14] N.V. Krylov, On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. *Math. USSR Izv.* 7: 691–709, 1973.
[15] N.V. Krylov, *Controlled diffusion processes, 2nd ed.* Berlin: Springer, 2009.
[16] N.V. Krylov, On Ito's stochastic integral equations. *Theory Probab. Appl.*, 14: 330–336, 1969.
[17] N.V. Krylov, Addendum: On Ito's Stochastic Integral Equations. *Theory Probab. Appl.*, 17: 373–374, 1973.
[18] N.V. Krylov, M.V. Safonov, A certain property of solutions of parabolic equations with measurable coefficients. *Math. USSR Izv.,* 16:1, 151–164, 1981.
[19] O.A. Ladyzhenskaja, V.A. Solonnikov, N.N. Ural'ceva, *Linear and Quasi-linear Equations of Parabolic Type*, AMS, 1968.
[20] A.A. Ljapunow, E.A. Stschegolkow, and W.J. Arsenin, *Arbeiten zur deskriptiven Mengenlehre.* Mathematische Forschungsberichte. 1. Berlin: VEB Deutscher Verlag der Wissenschaften, 1955.
[21] N. Lusin, Sur les propriétés des fonctions mesurables, *Comptes rendus de l'Académie des Sciences de Paris*, 154, 1688-1690, 1912.
[22] P. Mandl, *Analytical treatment of one-dimensional Markov processes,* Berlin-Heidelberg-New York: Springer, 1968.
[23] H.P. McKean, Stochastic integrals. AMS Chelsea Publishing, 2005.
[24] S. 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.
[25] H. Mine and S. Osaki, *Markovian decision processes.* New York: American Elsevier Publishing Company, Inc., 1970.
[26] R. Morton, On the optimal control of stationary diffusion processes with inaccessible boundaries and no discounting. *J. Appl. Probab.*, 8: 551–560, 1971.
[27] É. Pardoux and A.Yu. Veretennikov, On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3): 1061–1085, 2001.
[28] M.L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, 2005.
[29] V.V. Rykov, Controllable Queueing Systems: From the Very Beginning up to Nowadays, *Reliability: Theory & Applications (RT&A)*, vol. 12, 2(45), 39-61, 2017.
[30] E.A. Shchegol'kov, Über die Uniformisierung gewisser B-Mengen. *Dokl. Akad. Nauk SSSR, n. Ser.*, 59: 1065–1068, 1948.
[31] A.Yu. Veretennikov, On Polynomial Mixing and Convergence Rate for Stochastic Difference and Differential Equations, *Theory Probab. Appl.,* 44(2): 361-374, 2000.
[32] A.Yu. Veretennikov, On Sobolev Solutions of Poisson Equations in $\mathbb{R}^d$ with a Parameter, *Journal of Mathematical Sciences*, 179(1), 48-79, 2011.