A Modified MSA for Stochastic Control Problems

B. Kerimkulov[1,2] · D. Šiška[2,3] · L. Szpruch[2,4]

Abstract
The classical Method of Successive Approximations (MSA) is an iterative method for solving stochastic control problems and is derived from Pontryagin's optimality principle. It is known that the MSA may fail to converge. Using careful estimates for the backward stochastic differential equation (BSDE) this paper suggests a modification to the MSA algorithm. This modified MSA is shown to converge for general stochastic control problems with control in both the drift and diffusion coefficients. Under some additional assumptions the rate of convergence is shown. The results are valid without restrictions on the time horizon of the control problem, in contrast to iterative methods based on the theory of forward-backward stochastic differential equations.

Keywords Stochastic Control · Pontryagin Optimality · Backward Stochastic Differential Equation

✉ D. Šiška
D.Siska@ed.ac.uk
https://www.maths.ed.ac.uk

B. Kerimkulov
B.Kerimkulov@sms.ed.ac.uk
http://www.maxwell.ac.uk/migsaa

L. Szpruch
L.Szpruch@ed.ac.uk
https://www.turing.ac.uk

[1] Maxwell Institute Graduate School in Analysis and Applications, University of Edinburgh, Edinburgh, UK

[2] School of Mathematics, University of Edinburgh, Edinburgh, UK

[3] Vega Protocol, Gibraltar, UK

[4] Alan Turing Institute, London, UK

Mathematics Subject Classification 93E20 · 60H30 · 49M05

1 Introduction

Stochastic control problems appear naturally in a range of applications in engineering, economics and finance. With the exception of very specific cases such as the linear-quadratic control problem in engineering or Merton portfolio optimization task in finance, stochastic control problems typically have no closed form solutions and have to be solved numerically. In this work, we consider a modification to the method of successive approximations (MSA), see Algorithm 1. The MSA is essentially a way of applying the Pontryagin's optimality principle to get numerical solutions of stochastic control problems.

We will consider the continuous space, continuous time problem where the controlled system is modelled by an $\mathbb{R}^d$-valued diffusion process. Let W be a d'-dimensional Wiener martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. We will provide exact assumptions we need in Sect. 2. For now, let us fix a finite time $T \in (0, \infty)$ and consider the controlled stochastic differential equation (SDE) for given measurable functions $b : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R}^{d \times d'}$

$$dX_s = b(s, X_s, \alpha_s) \, ds + \sigma(s, X_s, \alpha_s) \, dW_s, \quad s \in [0, T], \quad X_0 = x. \quad (1)$$

Here $\alpha = (\alpha_s)_{s \in [0,T]}$ is a control process belonging to the space of admissible controls $\mathcal{A}$, valued in a separable metric space A and we will write X^α to denote the unique solution of (1) which starts from x at time 0 whilst being controlled by α. Furthermore let $f : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ be given measurable functions and consider the gain functional

$$J(x, \alpha) := \mathbb{E} \left[\int_0^T f(s, X_s^\alpha, \alpha_s) ds + g(X_T^\alpha) \right] \quad (2)$$

for all $x \in \mathbb{R}^d$ and $\alpha \in \mathcal{A}$. We want to solve the optimisation problem i.e. to find the optimal control α^* which achieves the minimum of (2) (or, if the infimum cannot be reached by $\alpha \in \mathcal{A}$ then an ε-optimal control $\alpha^\varepsilon \in \mathcal{A}$ such that $\inf_{\alpha \in \mathcal{A}} J(x, \alpha) \leq J(x, \alpha^\varepsilon) + \varepsilon$).

In the present paper, we study an approach based on Pontryagin's optimality principle, see e.g. [4,7] or [25]. The main idea is to consider optimality conditions for controls of the problem (2). Given b, σ and f we define the Hamiltonian $\mathcal{H} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d'} \times A \to \mathbb{R}$ as

$$\mathcal{H}(t, x, y, z, a) = b(t, x, a) \cdot y + \mathrm{tr}(\sigma^\top(t, x, a)z) + f(t, x, a). \quad (3)$$

Consider for each $\alpha \in \mathcal{A}$, the BSDE, called the adjoint equation

$$dY_s^\alpha = -D_x \mathcal{H}(s, X_s^\alpha, Y_s^\alpha, Z_s^\alpha, \alpha_s) \, ds + Z_s^\alpha \, dW_s, \quad Y_T^\alpha = D_x g(X_T^\alpha), \quad s \in [0, T]. \quad (4)$$

It is well known from Pontryagin's optimality principle that, if an admissible control $\alpha^* \in \mathcal{A}$ is optimal, X^{α^*} is the corresponding optimally controlled dynamic (1) and $(Y^{\alpha^*}, Z^{\alpha^*})$ is the solution to the associated adjoint equation (4), then $\forall a \in A$ and $\forall s \in [0, T]$ the following holds

$$\mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*) \leq \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a) \quad \text{a.s.} \tag{5}$$

We now define the augmented Hamiltonian $\tilde{\mathcal{H}} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d'} \times A \times A \to \mathbb{R}$ for some $\rho \geq 0$ by

$$\tilde{\mathcal{H}}(t, x, y, z, a', a) := \mathcal{H}(t, x, y, z, a) + \frac{1}{2}\rho|b(t, x, a) - b(t, x, a')|^2$$
$$\frac{1}{2}\rho|\sigma(t, x, a) - \sigma(t, x, a')|^2 + \frac{1}{2}\rho \left|D_x\mathcal{H}(t, x, y, z, a) - D_x\mathcal{H}(t, x, y, z, a')\right|^2. \tag{6}$$

Notice that when $\rho = 0$ we have exactly the definition of Hamiltonian (3). Given the augmented Hamiltonian, let us introduce the modified MSA in Algorithm 1 which consists of successive integrations of the state and adjoint systems and updates to the control. Notice that the backward SDE depends on the Hamiltonian $\mathcal{H}$, while the control update step comes from minimizing the augmented Hamiltonian $\tilde{\mathcal{H}}$.

Algorithm 1 Modified Method of Successive Approximations:

Initialisation: make a guess of the control $\alpha^0 = (\alpha_s^0)_{s \in [0, T]}$.

while difference between $J(x, \alpha^n)$ and $J(x, \alpha^{n-1})$ is large **do**

Given a control $\alpha^{n-1} = (\alpha_s^{n-1})_{s \in [0, T]}$ solve the following forward SDE, then solve backward SDE:

$$dX_s^n = b(s, X_s^n, \alpha_s^{n-1}) \, ds + \sigma(s, X_s^n, \alpha_s^{n-1}) \, dW_s, \quad X_0^n = x,$$
$$dY_s^n = -D_x\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) \, ds + Z_s^n \, dW_s, \quad Y_T^n = D_x g(X_T^n). \tag{7}$$

Update the control

$$\alpha_s^n \in \arg\min_{a \in A} \tilde{\mathcal{H}}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}, a), \quad \forall s \in [0, T]. \tag{8}$$

end while
return α^n.

The method of successive approximations (i.e. case $\rho = 0$) for numerical solution of deterministic control problems was proposed already in [5]. Recent application of the modified MSA to a deep learning problem has been studied in [32], where they formulated the training of deep neural networks as an optimal control problem and introduced the modified method of successive approximations as an alternative training algorithm for deep learning. For us, the main motivation to explore the modified MSA for stochastic control problems is to obtain convergence, ideally with rate, of an iterative algorithm, applicable to problems with the control in the diffusion part of the controlled dynamics. This is in contrast to [36] where convergence rate of an

the Bellman–Howard policy iteration is shown but only for control problems with no control in the diffusion part of the controlled dynamics.

In Lemma 2.3, which can be established using careful BSDE estimates, we can see the estimate on the change of J when we do a minimization step of Hamiltonian as in (8). If the sum of the last three terms of (14) is bigger than the first term, then for classical MSA algorithm (i.e. case $\rho = 0$) we cannot guarantee that we do an update of the control in optimal descent direction of J. That means that the method of successive approximations may diverge. To overcome this, we need to modify the algorithm in such way so that we ensure convergence. With this in mind the desirability of the the augmented Hamiltonian (6) for updating the control becomes clear, as long as it still characterises optimal controls like $\mathcal{H}$ does. Theorem 2.4 answers this question affirmatively which opens the way to the modified MSA. In Theorem 2.5 we show that the modified method of successive approximations, converges for arbitrary T, and in Corollary 2.6, we show logarithmic convergence rate for certain stochastic control problems.

We observe that the forward and backward dynamics in (7) are decoupled, due to the iteration used. Therefore, it can be efficiently approximated, even in high dimension, using deep learning methods, see [30,31]. However, the minimization step (8) might be computationally expensive for some problems. A possible approach circumventing this is to replace the full minimization of (8) by gradient descent. A continuous version of this gradient flow is analyzed in [37].

The main contributions of this paper are the probabilistic proof of convergence of the modified method of successive approximations and establishing convergence rate for a specific type of optimal control problems.

This paper is organised as follow: in Sect. 1.1 we compare our results with existing work. In Sect. 2 we state the assumptions and main results. In Sect. 3 we collect all proofs. Finally, in Appendix 1 we recall an auxiliary lemma which is needed in the proof of Corollary 2.6.

1.1 Related Work

One can solve the stochastic optimal control problem using dynamic programming principle. It is well known, see e.g. Krylov [8], that under reasonable assumptions the value function, defined as infimum of (2) over all admissible controls, satisfies the Bellman partial differential equation (PDE). There are several approaches to solve this nonlinear problem. One may apply a finite difference method to discretise the Bellman PDE and get a high dimensional nonlinear system of equations, see e.g [19] or [22]. Or one may linearize the Bellman PDE and then iterate. The classical approach is the Bellman–Howard policy improvement/iteration algorithm, see e.g. [1,2] or [3]. The algorithm is initialised with a "guess" of Markovian control. Given a Markovian control strategy at step n one solves a linear PDE with the given control fixed and then one uses the solution to the linear PDE to update the Markovian control, see e.g. [27,28] or [29]. In [36], a global rate of convergence and stability for the policy iteration algorithm has been established using backward stochastic differential equations (BSDE) theory.

However, the result only applies to stochastic control problems with no control in the diffusion coefficient of the controlled dynamics.

It is known that the solution of the stochastic optimal control problem can be obtained from a corresponding forward backward stochastic differential equation (FBSDE) via the stochastic optimality principle, see [26, Chap. 8.1]. Indeed, let us consider (1) and (4), and recall from the stochastic optimality principle, see [25, Theorem 4.12], that for the optimal control $\alpha^* = (\alpha_s^*)_{s \in [0,T]}$ we have that (5) holds. Assume that under some conditions on b, σ and f we have that the first order condition stated above uniquely determines α^* for $s \in [0, T]$ by

$$\alpha_s^* = \varphi(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}), \tag{9}$$

for some function φ. Therefore, after plugging (9) into (1) and (4), we obtain the following coupled FBSDE:

$$\begin{aligned} dX_s^{\alpha^*} &= \bar{b}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*})\, ds + \bar{\sigma}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*})\, dW_s, \quad s \in [0, T], \quad X_0^{\alpha^*} = x. \\ dY_s^{\alpha^*} &= -D_x \bar{\mathcal{H}}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*})\, ds + Z_s^{\alpha^*}\, dW_s, \quad Y_T = D_x g(X_T^{\alpha^*}), \quad s \in [0, T], \end{aligned} \tag{10}$$

where $(\bar{b}, \bar{\sigma})(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}) = (b, \sigma)(s, X_s^{\alpha^*}, \varphi(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}))$ and $\bar{\mathcal{H}}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}) = \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \varphi(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}))$. It is worth mentioning that when σ does not depend on the control $\bar{\sigma}$ will depend on forward process and time only. This means that $\bar{\sigma}$ does not have Y and Z components.

The theory of FBSDE has been studied widely and there are several methods to show the existence and uniqueness result, and a number of numerical algorithms have been proposed based on those methods. First is the method of contraction mapping. It was first studied by Antonelli [9] and later by Pardoux and Tang [15]. The main idea there is to show that a certain map is a contraction, and then to apply a fixed point argument. However, it turns out that this method works only for small enough time horizon T. In the case when $\bar{\sigma}$ does not depend on Y and Z, having small T is sufficient to get contraction. Otherwise, one needs to assume additionally that the Lipschitz constants of $\bar{\sigma}$ in z and that of g in x satisfy a certain inequality, see [26, Theorem 8.2.1]. Using the method of contraction mapping one can then implement a Picard-iteration-type numerical algorithm and show exponential convergence for small T. The second method is the Four Step Scheme. It was introduced by Ma et al., see [10], and was later studied by Delarue [17]. The idea is to use a decoupling function and then study an associated quasi-linear PDE. We note that in [10,17] the forward diffusion coefficient $\bar{\sigma}$ does not depend on Z. This corresponds to stochastic control problems with the uncontrolled diffusion coefficient. The numerical algorithms based on this method exploits the numerical solution of the associated quasi-linear PDE and therefore faces some limitations for high dimensional problems, see Douglas et al. [12], Milstein and Tretyakov [20], Ma et al. [21] and Delarue and Menozzi [18]. Guo et al. [24] proposed a numerical scheme for high-dimensional quasi-linear PDE associated with the coupled FBSDE when $\bar{\sigma}$ does not depend on Z, which is based on a monotone scheme and on probabilistic approach. Finally, there is the method of continuation. This method was developed by Hu and Peng [11], Peng and Wu [16] and by Yong [14]. It allows

them to show the existence and uniqueness result for arbitrary T under monotonicity conditions on the coefficients, which one would not expect to apply to FBSDEs arising from a control problem as described by (9), (10). Recently, deep learning methods have been applied to solving FBSDEs. In [35], three algorithms for solving fully coupled FBSDEs which have good accuracy and performance for high-dimensional problems are provided. One of the algorithms is based on the Picard iteration and it converges, but only for small enough T. In [34], an alternative algorithm for solving high-dimensional fully coupled FBSDEs based on deep learning was proposed, and the convergence result was shown assuming small T and other structural conditions (sometimes referred to as weak coupling and monotonicity conditions).

2 Main Results

We fix a finite horizon $T \in (0, \infty)$. Let A be a separable metric space. This is the space where the control processes α take values. We fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$. Let $W = (W_t)_{t \in [0,T]}$ be a d'-dimensional Wiener martingale on this space. By $\mathbb{E}_t$ we denote the conditional expectation with respect to $\mathcal{F}_t$. Let $|\cdot|$ denote any norm in a finite dimensional Euclidean space. By $\|\cdot\|_{L^\infty}$ we denote the norm in $L^\infty(\Omega)$. Let $\|Z\|_{\mathbb{H}^\infty} := \text{ess sup}_{(t,\omega)} |Z_t(\omega)|$ for any predictable process Z. We understand the following as $D_x \sigma = D_{x_l} \sigma^{ij}$, $D_x^2 b = D_{x_l x_n}^2 b^i$ and $D_x^2 \sigma = D_{x_l x_n}^2 \sigma^{ij}$, where $i, l, n = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, d'$. By $Z^\top$ we denote the transpose of Z.

The state of the system is governed by the controlled SDE (1). The corresponding adjoint equation satisfies (4).

Assumption 2.1 The functions b and σ are jointly continuous in t and twice differentiable in x. There exists $K \ge 0$ such that $\forall x \in \mathbb{R}^d, \forall a \in A, \forall t \in [0, T]$,

$$|D_x b(t, x, a)| + |D_x \sigma(t, x, a)| + |D_x^2 b(t, x, a)| \le K. \tag{11}$$

Moreover, assume that $D_x^2 \sigma(t, x, a) = 0 \ \forall x \in \mathbb{R}^d, \forall a \in A, \forall t \in [0, T]$.

Clearly the assumption (11) implies that $\forall x, x' \in \mathbb{R}^d, \forall a \in A, \forall t \in [0, T]$ we have

$$|b(t, x, a) - b(t, x', a)| + |\sigma(t, x, a) - \sigma(t, x', a)| \le K|x - x'|. \tag{12}$$

The assumption that $D_x^2 \sigma(t, x, a) = 0 \ \forall x \in \mathbb{R}^d, \forall a \in A, \forall t \in [0, T]$ is needed so that (21), in the proof of Lemma 3.1, holds. Without this assumption (21) would only hold if we could show that $\|Z^\alpha\|_{\mathbb{H}^\infty} < \infty$. Without additional regularization of the control problem this is impossible. Indeed, with [13, Proposition 5.3] we see that Z_t^α is a version of $D_t Y_t^\alpha$ (the Malliavin derivative of Y_t^α) and $D_t Y_t^\alpha$ itself satisfies a linear BSDE. However, to obtain the estimates using this representation, one term that arises is $D_t \alpha_s$ where $t \in [0, T]$ and $s \in [t, T]$. So we would need $\text{ess sup}_{\omega \in \Omega, t \in (0,T), s \in (t,T)} |D_t \alpha_s(\omega)| < \infty$. This is not necessarily the case here.

Assumption 2.2 The functions f is joinly continuous in t, and f and σ are twice differentiable in x. There is a constant $K \ge 0$ such that $\forall x, \forall a \in A, \forall t \in [0, T]$

$$|D_x g(x)| + |D_x f(t, x, a)| + |D_x^2 g(x)| + |D_x^2 f(t, x, a)| \leq K. \qquad (13)$$

Under these assumptions, we can obtain the following estimate.

Lemma 2.3 *Let Assumption* 2.1 *and* 2.2 *hold. Then for any admissible controls φ and θ there exists a constant $C > 0$ such that*

$$
\begin{aligned}
J(x, \varphi) - J(x, \theta) \leq \mathbb{E} \int_0^T & [\mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)] \, ds \\
& + C\mathbb{E} \int_0^T |b(s, X_s^\theta, \varphi_s) - b(s, X_s^\theta, \theta_s)|^2 \, ds \\
& + C\mathbb{E} \int_0^T |\sigma(s, X_s^\theta, \varphi_s) - \sigma(s, X_s^\theta, \theta_s)|^2 \, ds \\
& + C\mathbb{E} \int_0^T |D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) \\
& - D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)|^2 \, ds.
\end{aligned}
$$

$$(14)$$

The proof will be given in Sect. 3. We now state a necessary condition for optimality for the augmented Hamiltonian.

Theorem 2.4 [Extended Pontryagin's optimality principle] *Let α^* be the (locally) optimal control, X^{α^*} be the associated controlled state solving* (1)*, and $(Y^{\alpha^*}, Z^{\alpha^*})$ be the associated adjoint processes solving* (4)*. Then for any $a \in A$ we have*

$$\tilde{\mathcal{H}}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*, \alpha_s^*) \leq \tilde{\mathcal{H}}(t, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*, a), \quad \forall s \in [0, T]. \quad (15)$$

The proof of Theorem 2.4 will come in Sect. 3. We are now ready to present the main result of the paper.

Theorem 2.5 *Let Assumptions* 2.1 *and* 2.2 *hold. Then Algorithm* 1 *converges to a local minimum of* (2) *for sufficiently large $\rho > 0$.*

Theorem 2.5 will be proved in Sect. 3. It can be seen from the proof that ρ needs to be two times larger than the constant appearing in Lemma 2.3, which itself depends increases with T, d and constants from Assumption 2.1 and 2.2.

We cannot guarantee that the Algorithm 1 converges to the optimal control which minimizes (2), since the extended Pontryagin's optimality principle, see Theorem 2.4, is the necessary condition for optimality. The sufficient condition for optimality tells us that to get the optimal control we need to assume convexity of the Hamiltonian in state and control variables, and need to assume convexity of the terminal cost function. To that end, we need to assume convexity of b, σ, f and g in x and a.

In the following corollary, we show that under a particular setting of the problem we have logarithmic convergence of the modified method of successive approximations to the true solution of the problem.

Corollary 2.6 *Let Assumptions* 2.1 *and* 2.2 *hold. Moreover, if* b, σ *and* f *are in the form of*

$$b(t, x, a) = b_1(t)x + b_2(t, a),$$
$$\sigma(t, x, a) = \sigma_1(t)x + \sigma_2(t, a),$$
$$f(t, x, a) = f_1(t, x) + f_2(t, a)$$

for $\forall t \in [0, T], \ \forall x \in \mathbb{R}^d, \ \forall a \in A.$ *In addition, assume that* f *and* g *are convex in* $x, \ f_2, b_2, \sigma_2$ *are convex in* a. *Then we have the following estimate for the sequence* $(\alpha^n)_{n \in \mathbb{N}}$ *from Algorithm* 1:

$$0 \le J(x, \alpha^n) - J(x, \alpha^*) \le \frac{C}{n},$$

where α^* *is the optimal control for* (2) *and* C *is a positive constant.*

The proof of Corollary 2.6 will be given in Sect. 3. Theorem 2.5 and Corollary 2.6 are extensions of the result in [5] to the stochastic case.

3 Proofs

We start working towards the proof of Theorem 2.5. Recall the adjoint equation for an admissible control α:

$$dY_s^\alpha = -D_x \mathcal{H}(s, X_s^\alpha, Y_s^\alpha, Z_s^\alpha, \alpha_s) \, ds + Z_s^\alpha \, dW_s, \ s \in [0, T], \quad Y_T = D_x g(X_T^\alpha). \tag{16}$$

From now on, we shall use Einstein notation, so that repeated indices in a single term imply summation over all the values of that index.

Lemma 3.1 *Assume that there exists* $K \ge 0$ *such that* $\forall x \in \mathbb{R}^d, \forall a \in A, \forall t \in [0, T]$ *we have*

$$|D_x b(t, x, a)| + |D_x \sigma(t, x, a)| \le K,$$

and

$$|D_x g(x)| + |D_x f(t, x, a)| \le K.$$

Then $\|Y^\alpha\|_{\mathbb{H}^\infty}$ *is bounded.*

Proof From the definition of the Hamiltonian (3) we have

$$D_{x_i} \mathcal{H}(s, X_s^\alpha, Y_s^\alpha, Z_s^\alpha, \alpha_s) = D_{x_i} b^j(s, X_s^\alpha, \alpha_s)(Y_s^\alpha)^j + D_{x_i} \sigma^{jp}(s, X_s^\alpha, \alpha_s)(Z_s^\alpha)^{jp}$$
$$+ D_{x_i} f(s, X_s^\alpha, \alpha_s), \quad \forall s \in [0, T], \quad i = 1, 2, \dots, d.$$

Hence, one can observe that (16) is a linear BSDE. Therefore, from [33, Proposition 3.2] we can write the formula for the solution of (16):

$$Y_t^\alpha = \mathbb{E}_t \left[S_t^{-1} S_T D_x g(X_T^\alpha) + \int_t^T S_t^{-1} S_s D_x f(s, X_s^\alpha, \alpha_s) \, ds \right],$$

where the process S is the unique strong solution of

$$dS_t^{ij} = S_t^{il} D_{x_l} b^j(t, X_t^\alpha, \alpha_t)\, dt + S_t^{il} D_{x_l} \sigma^{jp}(t, X_t^\alpha, \alpha_t)\, dW_t^p, \quad i, j = 1, 2, \ldots, d, \ S_0 = I_d,$$

and S^{-1} is the inverse process of S. Thus, due to [33, Corollary 3.7] and assumptions of lemma we have the following bound:

$$\|Y^\alpha\|_{\mathbb{H}^\infty} \leq C \|D_x g(X_T^\alpha)\|_{L^\infty} + C T \|D_x f(\cdot, X_\cdot^\alpha, \alpha.)\|_{\mathbb{H}^\infty}.$$

Hence, due to assumptions of lemma we conclude that $\|Y^\alpha\|_{\mathbb{H}^\infty}$ is bounded.

Proof of Lemma 2.3 Let φ and θ be some generic admissible controls. We will write $(X_s^\varphi)_{s \in [0,T]}$ for the solution of (1) controlled by φ and $(X_s^\theta)_{s \in [0,T]}$ for the solution of (1) controlled by θ. We denote solutions of corresponding adjoint equations by $(Y_s^\varphi, Z_s^\varphi)_{s \in [0,T]}$ and $(Y_s^\theta, Z_s^\theta)_{s \in [0,T]}$. Due to Taylor's theorem, we note that for some $R^1(\omega) \in [0, 1]$, we have $\forall \omega \in \Omega$ that

$$
\begin{aligned}
g(X_T^\varphi) - g(X_T^\theta) &= (D_x g(X_T^\theta))^\top (X_T^\varphi - X_T^\theta) \\
&\quad + \frac{1}{2}(X_T^\varphi - X_T^\theta)^\top D_x^2 g(X_T^\theta + R^1(X_T^\varphi - X_T^\theta))(X_T^\varphi - X_T^\theta) \\
&\leq (D_x g(X_T^\theta))^\top (X_T^\varphi - X_T^\theta) \\
&\quad + \frac{1}{2}(X_T^\varphi - X_T^\theta)^\top \left| D_x^2 g(X_T^\theta + R^1(X_T^\varphi - X_T^\theta)) \right| (X_T^\varphi - X_T^\theta) \\
&\leq (D_x g(X_T^\theta))^\top (X_T^\varphi - X_T^\theta) + \frac{K}{2} \left| X_T^\varphi - X_T^\theta \right|^2.
\end{aligned}
$$

The last inequality holds due to Assumption 2.2. Recall that $Y_T^\theta = D_x g(X_T^\theta)$. Hence, using Itô's product rule, we get

$$
\begin{aligned}
\mathbb{E}[g(X_T^\varphi) - g(X_T^\theta)] &\leq \mathbb{E}\left[(Y_T^\theta)^\top (X_T^\varphi - X_T^\theta) + \frac{K}{2} \left| X_T^\varphi - X_T^\theta \right|^2 \right] \\
&\leq \mathbb{E}\int_0^T (X_s^\varphi - X_s^\theta)^\top dY_s^\theta + \mathbb{E}\int_0^T (Y_s^\theta)^\top [dX_s^\varphi - dX_s^\theta] \\
&\quad + \mathbb{E}\int_0^T \mathrm{tr}[(\sigma(s, X_s^\varphi, \varphi_s) - \sigma(s, X_s^\theta, \theta_s))^\top Z_s^\theta]\, ds \\
&\quad + \frac{K}{2}\mathbb{E}\left[\left| X_T^\varphi - X_T^\theta \right|^2 \right].
\end{aligned}
$$

From this, the forward SDE (1) and the adjoint equation (4) we thus get

$$
\begin{aligned}
\mathbb{E}[g(X_T^\varphi) - g(X_T^\theta)] &\leq -\mathbb{E}\int_0^T (X_s^\varphi - X_s^\theta)^\top D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)\, ds \\
&\quad + \mathbb{E}\int_0^T (Y_s^\theta)^\top [b(s, X_s^\varphi, \varphi_s) - b(s, X_s^\theta, \theta_s)]\, ds
\end{aligned}
$$

$$+ \mathbb{E} \int_0^T \text{tr}[(\sigma(s, X_s^\varphi, \varphi_s) - \sigma(s, X_s^\theta, \theta_s))^\top Z_s^\theta] \, ds$$

$$+ \frac{K}{2} \mathbb{E} \left[|X_T^\varphi - X_T^\theta|^2 \right]. \tag{17}$$

On the other hand, by definition of the Hamiltonian we have

$$\mathbb{E} \int_0^T [f(s, X_s^\varphi, \varphi_s) - f(s, X_s^\theta, \theta_s)] \, ds$$

$$= \mathbb{E} \int_0^T [\mathcal{H}(s, X_s^\varphi, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)] \, ds \tag{18}$$

$$- \mathbb{E} \int_0^T (Y_s^\theta)^\top [b(s, X_s^\varphi, \varphi_s) - b(s, X_s^\theta, \theta_s)] \, ds$$

$$- \mathbb{E} \int_0^T \text{tr}[(\sigma(s, X_s^\varphi, \varphi_s) - \sigma(s, X_s^\theta, \theta_s))^\top Z_s^\theta] \, ds.$$

Summing up (17) and (18) we get

$$J(x, \varphi) - J(x, \theta) = \mathbb{E}[g(X_T^\varphi) - g(X_T^\theta)] + \mathbb{E} \int_0^T [f(s, X_s^\varphi, \varphi_s) - f(s, X_s^\theta, \theta_s)] \, ds$$

$$\leq \mathbb{E} \int_0^T [\mathcal{H}(s, X_s^\varphi, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)$$

$$- (X_s^\varphi - X_s^\theta)^\top D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)] \, ds$$

$$+ \frac{K}{2} \mathbb{E} \left[|X_T^\varphi - X_T^\theta|^2 \right]. \tag{19}$$

Due to Taylor's theorem, there exists $(R_s^2(\omega))_{s \in [0,T]} \in [0, 1]$ such that $\forall \omega \in \Omega$ we have

$$\mathcal{H}(s, X_s^\varphi, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)$$

$$= \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)$$

$$+ (X_s^\varphi - X_s^\theta)^\top D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) \tag{20}$$

$$+ \frac{1}{2}(X_s^\varphi - X_s^\theta)^\top D_x^2 \mathcal{H}(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), Y_s^\theta, Z_s^\theta, \varphi_s)(X_s^\varphi - X_s^\theta).$$

Since $D_x^2 \sigma(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), \varphi_s) = 0$ by Assumption 2.1, we have that

$$\left| D_{x_i x_j}^2 \mathcal{H}(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), Y_s^\theta, Z_s^\theta, \varphi_s) \right|$$

$$= \left| D_{x_i x_j}^2 b^l(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), \varphi_s)(Y_s^\theta)^l \right.$$

$$\left. + D_{x_i x_j}^2 f(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), \varphi_s) \right|, \quad i, j = 1, 2, \ldots, d.$$

From Lemma 3.1 we know that $|Y_s^\theta|$ is bounded a.s. for all $s \in [0, T]$. Hence by Assumption 2.1 and 2.2 we have

$$|D_x^2 \mathcal{H}(s, X_s^\theta + R_s^2(X_s^\varphi - X_s^\theta), Y_s^\theta, Z_s^\theta, \varphi_s)| < \infty. \tag{21}$$

Therefore, after substituting (20) into (19), and by 21 we get

$$\begin{aligned} J(x, \varphi) - J(x, \theta) = \mathbb{E}\Bigg[\int_0^T & [\mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s) \\ & + (X_s^\varphi - X_s^\theta)^\top (D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) \\ & - D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)) \\ & + \frac{K}{2} \left| X_s^\varphi - X_s^\theta \right|^2 ds \Bigg] + \frac{K}{2} \mathbb{E}\left[\left| X_T^\varphi - X_T^\theta \right|^2 \right]. \end{aligned}$$

Let us now get a standard SDE estimate for the difference of X^φ and X^θ. From $(a + b)^2 \leq 2a^2 + 2b^2$, from taking the expectation, from Hölder's inequality, from Assumption 2.1, from the Burkholder–Davis–Gundy inequality and from Gronwall's inequality we obtain

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} |X_t^\varphi - X_t^\theta|^2 \leq C\mathbb{E} \int_0^T & |b(s, X_s^\theta, \varphi_s) - b(s, X_s^\theta, \theta_s)|^2 ds \\ & + C\mathbb{E} \int_0^T |\sigma(s, X_s^\theta, \varphi_s) - \sigma(s, X_s^\theta, \theta_s)|^2 ds. \end{aligned} \tag{22}$$

Young's inequality allows us to get the estimate

$$\begin{aligned} J(x, \varphi) &- J(x, \theta) \\ &\leq \mathbb{E} \int_0^T [\mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)] ds \\ &+ \frac{1}{2} \mathbb{E} \int_0^T |X_s^\varphi - X_s^\theta|^2 ds \\ &+ \frac{1}{2} \mathbb{E} \Bigg[\int_0^T |D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)|^2 \\ &+ \frac{K}{2} \left| X_s^\varphi - X_s^\theta \right|^2 ds \Bigg] + \frac{K}{2} \mathbb{E}\left[\left| X_T^\varphi - X_T^\theta \right|^2 \right]. \end{aligned}$$

Hence, from (22) we have that

$$\begin{aligned} J(x, \varphi) - J(x, \theta) \leq \mathbb{E} \int_0^T & [\mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s) - \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)] ds \\ & + C\mathbb{E} \int_0^T |b(s, X_s^\theta, \varphi_s) - b(s, X_s^\theta, \theta_s)|^2 ds \end{aligned}$$

$$+ C\mathbb{E} \int_0^T |\sigma(s, X_s^\theta, \varphi_s) - \sigma(s, X_s^\theta, \theta_s)|^2 \, ds$$

$$+ C\mathbb{E} \int_0^T |D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \varphi_s)$$

$$- D_x \mathcal{H}(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \theta_s)|^2 \, ds,$$

for some constant $C > 0$, which depends on K, T, and d. $\square$

Proof of Theorem 2.4 Since α^* is the (locally) optimal control for the problem (2), the Pontryagin's optimality principle holds, see e.g. [23]. Hence for any $a \in A$ we have

$$\mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*) \leq \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a), \quad \forall s \in [0, T]. \tag{23}$$

By definition of the augmented Hamiltonian (6) for all $s \in [0, T]$ we have

$$
\begin{aligned}
\tilde{\mathcal{H}}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*, a) &= \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a) \\
&+ \frac{1}{2}\rho |b(s, X_s^{\alpha^*}, a) - b(s, X_s^{\alpha^*}, \alpha_s^*)|^2 \\
&+ \frac{1}{2}\rho |\sigma(s, X_s^{\alpha^*}, a) - \sigma(s, X_s^{\alpha^*}, \alpha_s^*)|^2 \\
&+ \frac{1}{2}\rho |D_x \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a) - D_x \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*)|^2.
\end{aligned}
\tag{24}
$$

Therefore, due to (23) and (24) we have

$$
\begin{aligned}
\tilde{\mathcal{H}}&(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*, \alpha_s^*) \\
&= \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*) \leq \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a) \\
&+ \frac{1}{2}\rho |b(s, X_s^{\alpha^*}, a) - b(s, X_s^{\alpha^*}, \alpha_s^*)|^2 \\
&+ \frac{1}{2}\rho |\sigma(s, X_s^{\alpha^*}, a) - \sigma(s, X_s^{\alpha^*}, \alpha_s^*)|^2 \\
&+ \frac{1}{2}\rho |D_x \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, a) - D_x \mathcal{H}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*)|^2 \\
&= \tilde{\mathcal{H}}(s, X_s^{\alpha^*}, Y_s^{\alpha^*}, Z_s^{\alpha^*}, \alpha_s^*, a).
\end{aligned}
$$

This concludes the proof. $\square$

Proof of Theorem 2.5 Let us apply Lemma 2.3 for $\varphi = \alpha^n$ and $\theta = \alpha^{n-1}$. Hence, for some $C > 0$ we have

$$
\begin{aligned}
J(x, \alpha^n) &- J(x, \alpha^{n-1}) \\
&\leq \mathbb{E} \int_0^T [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds \\
&\quad + C\mathbb{E} \int_0^T |b(s, X_s^n, \alpha_s^n) - b(s, X_s^n, \alpha_s^{n-1})|^2 \, ds \\
&\quad + C\mathbb{E} \int_0^T |\sigma(s, X_s^n, \alpha_s^n) - \sigma(s, X_s^n, \alpha_s^{n-1})|^2 \, ds \\
&\quad + C\mathbb{E} \int_0^T \left| D_x\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - D_x\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) \right|^2 \, ds.
\end{aligned}
\tag{25}
$$

Let

$$
\mu(\alpha^{n-1}) = \mathbb{E} \int_0^T [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds.
$$

Due to the definition of α^n (8) and (15) we have for all $s \in [0, T]$

$$
\begin{aligned}
\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) &+ \frac{1}{2}\rho|b(s, X_s^n, \alpha_s^n) - b(s, X_s^n, \alpha_s^{n-1})|^2 \\
&+ \frac{1}{2}\rho|\sigma(s, X_s^n, \alpha_s^n) - \sigma(s, X_s^n, \alpha_s^{n-1})|^2 \\
&+ \frac{1}{2}\rho|D_x\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - D_x\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})|^2 \\
&\leq \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}).
\end{aligned}
$$

Therefore, we can observe that $\mu(\alpha^{n-1}) \leq 0$. Hence we can rewrite the inequality (25) as

$$
J(x, \alpha^n) - J(x, \alpha^{n-1}) \leq \mu(\alpha^{n-1}) - \frac{2C}{\rho}\mu(\alpha^{n-1}) = D\mu(\alpha^{n-1}),
\tag{26}
$$

where $D := 1 - \frac{2C}{\rho}$. By choosing $\rho > 2C$ we have that $D > 0$. Notice that for any integer $M > 1$ we have

$$
\begin{aligned}
\sum_{n=1}^M (-\mu(\alpha^{n-1})) &\leq D^{-1} \sum_{n=1}^M (J(x, \alpha^{n-1}) - J(x, \alpha^n)) \\
&= D^{-1}(J(x, \alpha^0) - J(x, \alpha^M)) \\
&\leq D^{-1}(J(x, \alpha^0) - \inf_{\alpha \in \mathcal{A}} J(x, \alpha)) < \infty.
\end{aligned}
$$

Since $(-\mu(\alpha^{n-1})) \geq 0$ and $\sum_{n=1}^{\infty}(-\mu(\alpha^{n-1})) < +\infty$ we have that $\mu(\alpha^{n-1}) \to 0$ as $n \to 0$. This concludes the proof. $\qquad\square$

We need to introduce new notation, which will be used in the proof of Corollary 2.6. Denote the set

$$I_{\tau,h} := [\tau - h, \tau + h] \cap [0, T], \quad \tau \in [0, T], \quad h \in [0, +\infty). \qquad (27)$$

Let us define for all $s \in [0, T]$

$$\Delta_{\alpha^{n-1}}\mathcal{H}(s) := \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}),$$

and

$$\mu(\alpha^{n-1}) := \mathbb{E}\int_0^T \Delta_{\alpha^{n-1}}\mathcal{H}(s)\,ds.$$

By definition of α^n notice that $\Delta_{\alpha^{n-1}}\mathcal{H}(t) \leq 0$ for all $t \in [0, T]$. Let us show an auxiliary lemma.

Lemma 3.2 *For any $h > 0$ there exists τ, which depends on h and α^{n-1}, such that*

$$\mathbb{E}\int_{I_{\tau,h}} \Delta_{\alpha^{n-1}}\mathcal{H}(t)\,dt \leq \frac{h\mu(\alpha^{n-1})}{T}.$$

Proof We will prove by contradiction. Assume that there exists $h^* > 0$ such that $\forall \tau \in [0, T]$ we have

$$\mathbb{E}\int_{I_{\tau,h^*}} \Delta_{\alpha^{n-1}}\mathcal{H}(t)\,dt > \frac{h^*\mu(\alpha^{n-1})}{T}. \qquad (28)$$

Denote $\tau_i = ih^*$, $i = 1, \ldots, N(h^*)$, where $N(h^*) = [T/h^*]$ - integer part. Since $\Delta_{\alpha^{n-1}}\mathcal{H}(t) \leq 0$ for all $t \in [0, T]$ by definition of α^n and $\cup_{i=1}^{N(h^*)} I_{\tau_i,h^*}$ is a superset of $[0, T]$ we have

$$\mu(\alpha^{n-1}) = \mathbb{E}\int_0^T \Delta_{\alpha^{n-1}}\mathcal{H}(t)\,dt \geq \sum_{i=1}^{N(h^*)} \mathbb{E}\int_{I_{\tau_i,h^*}} \Delta_{\alpha^{n-1}}\mathcal{H}(t)\,dt. \qquad (29)$$

Hence, by (28) we get

$$\mu(\alpha^{n-1}) > \sum_{i=1}^{N(h^*)} \frac{h^*\mu(\alpha^{n-1})}{T} = \frac{h^*N(h^*)}{T}\mu(\alpha^{n-1}) \geq \mu(\alpha^{n-1}).$$

Last inequality holds since $\frac{h^*N(h^*)}{T}$ is less or equal to 1. Hence we get the contradiction. $\qquad\square$

Now we are ready to prove Corollary 2.6.

Proof of Corollary 2.6 First, observe that

$$
\begin{aligned}
b(s, X_s^n, \alpha_s^n) - b(s, X_s^n, \alpha_s^{n-1}) &= b_2(s, \alpha_s^n) - b_2(s, \alpha_s^{n-1}), \\
\sigma(s, X_s^n, \alpha_s^n) - \sigma(s, X_s^n, \alpha_s^{n-1}) &= \sigma_2(s, \alpha_s^n) - \sigma_2(s, \alpha_s^{n-1}), \\
D_x \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - D_x \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) &= 0.
\end{aligned}
$$

Let us consider the set $I_{\tau,h}$ given by (27). We will specify the choice of τ and h later. Hence, after applying Lemma 2.3 for α^n and α^{n-1} we have for some $C > 0$

$$
\begin{aligned}
&J(x, \alpha^n) - J(x, \alpha^{n-1}) \\
&\leq \mathbb{E} \int_{I_{\tau,h}} [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds \\
&\quad + C\mathbb{E} \int_{I_{\tau,h}} |b_2(s, \alpha_s^n) - b_2(s, \alpha_s^{n-1})|^2 + |\sigma_2(s, \alpha_s^n) - \sigma_2(s, \alpha_s^{n-1})|^2 \, ds \\
&\quad + \mathbb{E} \int_{[0,T] \setminus I_{\tau,h}} [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds \\
&\quad + C\mathbb{E} \int_{[0,T] \setminus I_{\tau,h}} |b_2(s, \alpha_s^n) - b_2(s, \alpha_s^{n-1})|^2 + |\sigma_2(s, \alpha_s^n) - \sigma_2(s, \alpha_s^{n-1})|^2 \, ds.
\end{aligned}
$$

Since the following holds for all $s \in [0, T]$ and $\rho \geq 0$:

$$
\begin{aligned}
&\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) \\
&+ \frac{1}{2}\rho |b_2(s, \alpha_s^n) - b_2(s, \alpha_s^{n-1})|^2 + \frac{1}{2}\rho |\sigma_2(s, \alpha_s^n) - \sigma_2(s, \alpha_s^{n-1})|^2 \leq 0,
\end{aligned}
$$

we have for $\rho \geq 2C$

$$
\begin{aligned}
&J(x, \alpha^n) - J(x, \alpha^{n-1}) \\
&\leq \mathbb{E} \int_{I_{\tau,h}} [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds \\
&\quad + C\mathbb{E} \int_{I_{\tau,h}} |b_2(s, \alpha_s^n) - b_2(s, \alpha_s^{n-1})|^2 + |\sigma_2(s, \alpha_s^n) - \sigma_2(s, \alpha_s^{n-1})|^2 \, ds.
\end{aligned}
$$

Therefore, from Lemma 3.2 and from similar calculations as in (26), there exists τ such that

$$
\begin{aligned}
&J(x, \alpha^n) - J(x, \alpha^{n-1}) \\
&\leq \left(1 - \frac{2C}{\rho}\right) \mathbb{E} \int_{I_{\tau,h}} [\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})] \, ds \\
&\leq \left(1 - \frac{2C}{\rho}\right) \frac{h\mu(\alpha^{n-1})}{T}.
\end{aligned}
$$

Let us choose $h = -(\rho - 2C)\mu(\alpha^{n-1})/(\rho T)$. Hence

$$J(x, \alpha^n) - J(x, \alpha^{n-1}) \leq -(\rho - 2C)^2 (\mu(\alpha^{n-1}))^2 / (\rho^2 T^2). \tag{30}$$

Let α^* be the optimal control. Indeed, by the sufficient condition for optimality, see e.g. [23], and by assumptions of corollary, we have the existence of the optimal control. Therefore, by convexity of g, and by Itô's product rule we have

$$
\begin{aligned}
0 \leq{} & J(x, \alpha^{n-1}) - J(x, \alpha^*) \\
={} & \mathbb{E}\left[\int_0^T (f(s, X_s^n, \alpha_s^{n-1}) - f(s, X_s, \alpha_s^*))\, ds + g(X_T^n) - g(X_T)\right] \\
\leq{} & \mathbb{E}\left[\int_0^T (f(s, X_s^n, \alpha_s^{n-1}) - f(s, X_s, \alpha_s^*))\, ds\right] + \mathbb{E}[(D_x g(X^n))^\top (X_T^n - X_T)] \\
\leq{} & \mathbb{E}\left[\int_0^T (f(s, X_s^n, \alpha_s^{n-1}) - f(s, X_s, \alpha_s^*))\, ds\right] \\
& + \mathbb{E}\left[\int_0^T (Y_s^n)^\top d(X_s^n - X_s) + \int_0^T (X_s^n - X_s)^\top dY_s^n\right] \\
& + \mathbb{E}\left[\int_0^T \operatorname{tr}((\sigma(s, X_s^n, \alpha_s^{n-1}) - \sigma(s, X_s, \alpha_s^*))^\top Z_s^n)\, ds\right].
\end{aligned}
$$

Hence, we have that

$$
\begin{aligned}
0 \leq{} & J(x, \alpha^{n-1}) - J(x, \alpha^*) \\
\leq{} & \mathbb{E}\left[\int_0^T f(s, X_s^n, \alpha_s^{n-1}) - f(s, X_s, \alpha_s^*)\, ds\right] \\
& + \mathbb{E}\left[\int_0^T (Y_s^n)^\top (b(s, X_s^n, \alpha_s^{n-1}) - b(s, X_s, \alpha_s^*))\, ds\right] \\
& - \mathbb{E}\left[\int_0^T (X_s^n - X_s)^\top D_x \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1})\, ds\right] \\
& + \mathbb{E}\left[\int_0^T \operatorname{tr}((\sigma(s, X_s^n, \alpha_s^{n-1}) - \sigma(s, X_s, \alpha_s^*))^\top Z_s^n)\, ds\right].
\end{aligned}
$$

Recalling the form of b, σ and observing that

$$D_x \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) = b_1(s) Y_s^n + \sigma_1(s) Z_s^n + D_x f(s, X_s^n, \alpha_s^{n-1}),$$

we have

$$
\begin{aligned}
0 \leq\ & J(x, \alpha^{n-1}) - J(x, \alpha^*) \\
\leq\ & \mathbb{E}\left[\int_0^T f(s, X_s^n, \alpha_s^{n-1}) - f(s, X_s, \alpha_s^*)\,ds\right] \\
& + \mathbb{E}\left[\int_0^T \mathrm{tr}((\sigma_2(s, \alpha_s^{n-1}) - \sigma_2(s, \alpha_s^*))^\top Z_s^n)\,ds\right] \\
& + \mathbb{E}\left[\int_0^T (Y_s^n)^\top (b_2(s, \alpha_s^{n-1}) - b_2(s, \alpha_s^*))\,ds - \int_0^T (X_s^n - X_s)^\top D_x f(s, X_s^n, \alpha_s^{n-1})\,ds\right].
\end{aligned}
$$

Since f is convex in x we have for all $s \in [0, T]$ that

$$
f(s, X_s, \alpha_s^{n-1}) \geq f(s, X_s^n, \alpha_s^{n-1}) + (X_s - X_s^n)^\top D_x f(s, X_s^n, \alpha_s^{n-1}).
$$

Therefore, we obtain

$$
\begin{aligned}
& J(x, \alpha^{n-1}) - J(x, \alpha^*) \\
& \qquad \leq \mathbb{E} \int_0^T \left[\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^{n-1}) - \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^*)\right] ds \qquad (31) \\
& \qquad \leq -\mu(\alpha^{n-1}),
\end{aligned}
$$

where the second inequality holds due to

$$
\mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^n) \leq \mathcal{H}(s, X_s^n, Y_s^n, Z_s^n, \alpha_s^*).
$$

Let $b^n := J(x, \alpha^n) - J(x, \alpha)$, then due to (30) and (31) we have that

$$
b^n - b^{n-1} \leq \frac{-(\rho - 2C)^2 \mu(\alpha^{n-1})^2}{(\rho^2 T^2)} \leq \frac{-(\rho - 2C)^2 (b^{n-1})^2}{\rho^2 T^2}.
$$

Therefore, due to Lemma A.1 we have

$$
J(x, \alpha^n) - J(x, \alpha^*) \leq \frac{C_1}{n}.
$$

for some constant $C_1 > 0$. This concludes the proof. $\qquad\square$

Appendix A: Auxiliary Lemma

Lemma A.1 *Let $\{b_k\}_{k\in\mathbb{N}}$ be the sequence of nonnegative numbers such that*

$$b_{k+1} \le b_k - qb_k^2,$$

where q is a positive constant. Then $b_k = O(1/k)$.

One can find the proof in [6, Lemma 1.4, p. 93]. However, the proof is written in Russian. For convenience of the reader we provide it here.

Proof Let $b_k = \frac{c_k}{k}$ for some nonnegative sequence $(c_k)_{k\in\mathbb{N}}$. Then it is enough to show that c_k is bounded for all $k \in \mathbb{N}$. By assumption we have

$$b_k - b_{k+1} = \frac{c_k}{k} - \frac{c_{k+1}}{k+1} = \frac{c_k}{k}\left(1 - \frac{c_{k+1}}{c_k}\frac{k}{k+1}\right) \ge q\frac{c_k^2}{k^2}.$$

Therefore,

$$1 - \frac{c_{k+1}}{c_k}\frac{k}{k+1} \ge q\frac{c_k}{k}.$$

After some transformation, we can rewrite the equation above as

$$\left(1+\frac{1}{k}\right)\left(1-q\frac{c_k}{k}\right) \ge \frac{c_{k+1}}{c_k}.$$

Thus

$$1 + \frac{1}{k}(1-qc_k) - q\frac{c_k}{k^2} \ge \frac{c_{k+1}}{c_k}.$$

If $1 - qc_k < 0$ we have

$$1 > 1 + \frac{1}{k}(1-qc_k) - q\frac{c_k}{k^2} \ge \frac{c_{k+1}}{c_k}.$$

Hence $c_{k+1} < c_k$. On the other hand, if $1 - qc_k \ge 0$, we have $c_k \le \frac{1}{q}$. Therefore, we conclude that for all k we have

$$c_k \le \max\left\{c_1, \frac{1}{q}\right\}.$$

$\square$

References

1. Antonelli, F.: Backward-forward stochastic differential equations. Ann. Appl. Probab. **3**, 777–793 (1993)
2. Bellman, R.: Functional equations in the theory of dynamic programming. V. Positivity and quasi-linearity. Proc. Natl Acad. Sci. U.S.A. **41**(10), 743–746 (1955)
3. Bellman, R.: Dynamic Programming. Princeton University Press, Princeton (1957)
4. Boltyanskii, V.G., Gamkrelidze, R.V., Pontryagin, L.S.: Theory of optimal processes. I. The maximum principle. Izv. Akad. Nauk SSSR Ser. Mat. **24**(1), 3–42 (1960)
5. Carmona, R.: Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications. SIAM, Philadelphia (2016)
6. Delarue, F.: On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. Stoch. Process. Appl. **99**, 209–289 (2002)
7. Delarue, F., Menozzi, S.: A forward-backward stochastic algorithm for quasi-linear PDEs. Ann. Appl. Probab. **16**(1), 140–184 (2006)
8. Demyanov, V.D., Rubinov, A.M.: Approximate Methods in Extremal Problems. Leningrad State University, Leningrad (1968). (in Russian)
9. Dong, H., Krylov, N.V.: The rate of convergence of finite-difference approximations for parabolic Bellman equations with Lipschitz coefficients in cylindrical domains. Appl. Math. Optim. **56**(1), 37–66 (2007)
10. Douglas, J., Ma, J., Protter, P.: Numerical methods for forward-backward stochastic differential equations. Ann. Appl. Probab. **6**(3), 940–968 (1996)
11. El Karoui, N., Peng, S., Quenez, M.C.: Backward stochastic differential equations in finance. Math. Finance **7**(1), 1–71 (1997)
12. Guo, W., Zhang, J., Zhuo, J.: A monotone scheme for high-dimensional fully nonlinear PDEs. Ann. Appl. Probab. **25**(3), 1540–1580 (2015)
13. Gyöngy, I., Šiska, D.: On finite-difference approximations for normalized Bellman equations. Appl. Math. Optim. **60**, 297–339 (2009)
14. Han, J., Long, J.: Convergence of the deep BSDE method for coupled FBSDEs. Probab. Uncertain. Quant. Risk. **5**(1), 1–33 (2020)
15. Han, J., Jentzen, A., Weinan, E.: Solving high-dimensional partial differential equations using deep learning. Proc. Natl. Acad. Sci. U.S.A. **115**(34), 8505–8510 (2018)
16. Harter, J., Richou, A.: A stability approach for solving multidimensional quadratic BSDEs. Electron. J. Probab. **24**(4), 1–51 (2019)
17. Howard, R.A.: Dynamic Programming and Markov Processes. MIT Press, Cambridge (1960)
18. Hu, Y., Peng, S.: Solution of forward-backward stochastic differential equations. Probab. Theory Relat. Fields **103**, 273–283 (1995)
19. Jacka, S.D., Mijatović, A.: On the policy improvement algorithm in continuous time. Stochastics **89**(1), 348–359 (2017)
20. Jacka, S.D., Mijatović, A., Siraj, A.: Coupling and a generalised Policy Iteration Algorithm in continuous time. arXiv:1707.07834 (2017)
21. Ji, S., Peng, S., Peng, Y., Zhang, X.: Three algorithms for solving high-dimensional fully-coupled FBSDEs through deep learning. IEEE Intell. Syst. **35**(3), 71–84 (2020)
22. Kerimkulov, B., Šiška, D., Szpruch, Ł.: Exponential convergence and stability of Howard's policy improvement algorithm for controlled diffusions. SIAM J. Control Optim. **58**(3), 1314–1340 (2020)
23. Krylov, N.V.: Controlled Diffusion Processes. Springer, Berlin (1980)
24. Krylov, I.A., Chernousko, F.L.: On the method of successive approximations for solution of optimal control problems (in Russian). U.S.S.R. Comput. Math. Math. Phys. **2**(6), 1371–1382 (1963)
25. Li, Q., Chen, L., Tai, C., Weinan, E.: Maximum principle based algorithms for deep learning. J. Mach. Learn. Res. **18**(165), 1–29 (2018)
26. Ma, J., Protter, P., Yong, J.M.: Solving forward-backward stochastic differential equations explicitly—a four step scheme. Probab. Theory Relat. Fields **98**, 339–359 (1994)
27. Ma, J., Shen, J., Zhao, Y.: On numerical approximations of forward-backward stochastic differential equations. SIAM J. Numer. Anal. **46**(5), 2636–2661 (2008)
28. Maeda, A., Jacka, S.D.: Evaluation of the rate of convergence in the PIA. arXiv:1709.06466 (2017)
29. Milstein, G., Tretyakov, M.: Discretization of forward-backward stochastic differential equations and related quasi-linear parabolic equations. IMA J. Numer. Anal. **27**(1), 24–44 (2007)

30. Pardoux, E., Tang, S.: Forward-backward stochastic differential equations and quasilinear parabolic PDEs. Probab. Theory Relat. Fields **114**(2), 123–150 (1999)
31. Peng, S., Wu, Z.: Fully coupled forward-backward stochastic differential equations and applications to optimal control. SIAM J. Control Optim. **37**, 825–843 (1999)
32. Pham, H.: Continuous-Time Stochastic Control and Optimization with Financial Applications. Springer, Berlin (2009)
33. Pontryagin, L.S.: Mathematical Theory of Optimal Processes. CRC Press, Boca Raton (1987)
34. Sabate Vidales, M., Šiška, D., Szpruch, Ł.: Unbiased deep solvers for parametric PDEs. arXiv:1810.05094v2 (2018)
35. Šiška, D., Szpruch, Ł.: Gradient flows for regularized stochastic control problems. arXiv:2006.05956 (2020)
36. Yong, J.: Finding adapted solutions of forward-backward stochastic differential equations: method of continuation. Probab. Theory Relat. Fields **107**, 537–572 (1997)
37. Zhang, J.: Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory. Springer, New York (2017)