SIAM J. CONTROL OPTIM.
Vol. 61, No. 6, pp. 3526–3558

LINEAR CONVERGENCE OF A POLICY GRADIENT METHOD FOR SOME FINITE HORIZON CONTINUOUS TIME CONTROL PROBLEMS*

CHRISTOPH REISINGER†, WOLFGANG STOCKINGER†, AND YUFEI ZHANG‡

Abstract. Despite its popularity in the reinforcement learning community, a provably convergent policy gradient method for continuous space-time control problems with nonlinear state dynamics has been elusive. This paper proposes proximal gradient algorithms for feedback controls of finite-time horizon stochastic control problems. The state dynamics are nonlinear diffusions with control-affine drift, and the cost functions are nonconvex in the state and nonsmooth in the control. The system noise can degenerate, which allows for deterministic control problems as special cases. We prove under suitable conditions that the algorithm converges linearly to a stationary point of the control problem and is stable with respect to policy updates by approximate gradient steps. The convergence result justifies the recent reinforcement learning heuristics that adding entropy regularization or a fictitious discount factor to the optimization objective accelerates the convergence of policy gradient methods. The proof exploits careful regularity estimates of backward stochastic differential equations.

Key words. reinforcement learning, policy gradient method, stochastic control, linear convergence, stationary point, backward stochastic differential equation

MSC codes. 68Q25, 93E20, 49M05

DOI. 10.1137/22M1492180

1. Introduction.

Stochastic control problems seek optimal strategies to control continuous time stochastic systems and are ubiquitous in modern science, engineering, and economics [26, 35]. In most applications, the agent aims to construct a feedback control mapping states of the system to optimal actions. A feedback control has the advantage that it allows for implementing an optimal control in real time through evaluating the feedback map at observed system states. An effective approach to generate (nearly) optimal feedback controls for high-dimensional control problems is via gradient-based algorithms (see, e.g., [32, 16, 40, 22]). These algorithms, often referred to as *policy gradient methods* (PGMs) in the reinforcement learning community, approximate a policy (i.e., a feedback control) in a parametric form and update the policy parameterization iteratively based on gradients of the control objective.

Despite the notable success of PGMs, a mathematical theory that guarantees the convergence of these algorithms for general (continuous time) stochastic control problems has been elusive. It is known that the objective of a control problem is typically nonconvex with respect to *feedback controls,* even if all cost functions are convex in state and control variables; see [10, Proposition 2.4] for a concrete example with deterministic linear state dynamics and strongly convex quadratic costs.

*Received by the editors April 22, 2022; accepted for publication (in revised form) September 7, 2023; published electronically November 29, 2023.
https://doi.org/10.1137/22M1492180
Funding: The second author's research was supported by a special Upper Austrian Government grant.
†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK (christoph.reisinger@maths.ox.ac.uk, wolfgang.stockinger@stcatz.ox.ac.uk).
‡Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (yufei.zhang@imperial.ac.uk).

This lack of convexity creates an essential challenge in analyzing the convergence behavior of PGMs. Most existing theoretical results of PGMs, especially those establishing (optimal) linear convergence, focus on discrete time problems and restrict policies within specific parametric families. This includes Markov decision problems (MDPs) with softmax parameterized policies [30] or overparameterized one-hidden-layer neural-network policies [43, 12, 23] and discrete time linear-quadratic (LQ) control problems with linear parameterized policies [9, 15]. The analysis therein exploits heavily the specific structure of the considered (discrete time) control problems and policy parameterization and hence is difficult to extend to general continuous time control problems or general policy parameterizations. This leads to the following natural question:

Can one design provably convergent gradient-based algorithms for feedback controls of continuous time nonlinear control problems without requiring specific policy parameterization?

Analyzing PGMs in the continuous space-time setting avoids discretization artifacts and yields algorithms whose convergence behavior is robust with respect to time and space mesh sizes [42]. Similarly, analyzing gradient descent algorithms without specific policy parametrization avoids searching for controls in a suboptimal class. This approach also highlights the essential structures of the control problem that affect the algorithmic performance, which subsequently provides a basis for developing improved algorithms with more effective policy parameterizations (see Remark 2.2).

This work takes an initial step towards answering the above challenging question and designs a convergent PGM for certain control problems with uncontrolled diffusion coefficients and affine control of the drift. Let $T \in (0,\infty)$ be a given terminal time, $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space on which a d-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ is defined, $\mathbb{F}$ be the natural filtration of W augmented with an independent σ-algebra $\mathcal{F}_0$, and $\mathcal{H}^2(\mathbb{R}^k)$ be the set of $\mathbb{R}^k$-valued square integrable $\mathbb{F}$-progressively measurable processes $\alpha = (\alpha_t)_{t\in[0,T]}$. For any initial state $\xi_0 \in L^2(\mathcal{F}_0;\mathbb{R}^n)$ and any $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$, consider the following controlled dynamics:

$$\mathrm{d}X_t = b_t(X_t,\alpha_t)\,\mathrm{d}t + \sigma_t(X_t)\,\mathrm{d}W_t,\ t\in[0,T], \quad X_0 = \xi_0, \tag{1.1}$$

where $b:[0,T]\times\mathbb{R}^n\times\mathbb{R}^k \to \mathbb{R}^n$ and $\sigma:[0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times d}$ are differentiable functions such that (1.1) admits a unique strong solution $X^{\xi_0,\alpha}$. The agent's objective is to minimize the cost functional

$$J(\alpha;\xi_0) = \mathbb{E}\left[\int_0^T e^{-\rho t}\left(f_t(X_t^{\xi_0,\alpha},\alpha_t) + \ell(\alpha_t)\right)\mathrm{d}t + e^{-\rho T} g(X_T^{\xi_0,\alpha})\right] \tag{1.2}$$

over all admissible controls $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$, where $\rho \geq 0$ is a given discount factor,[1] $f:[0,T]\times\mathbb{R}^n\times\mathbb{R}^k \to \mathbb{R}$ and $g:\mathbb{R}^n\to\mathbb{R}$ are differentiable functions, and $\ell:\mathbb{R}^k \to \mathbb{R}\cup\{\infty\}$ is a (possibly nondifferentiable) convex function.

The precise conditions on the coefficients in (1.1)–(1.2) will be given in section 2.1. In particular, we require the drift coefficient to be affine in the control but allow both drift and diffusion coefficients to be nonlinear in the state. The diffusion coefficient can degenerate, and hence (1.1) includes as a special case the deterministic *control-affine* system in nonlinear control theory [20]. We allow the cost functions f and g to be nonconvex in the state but require the running cost $f+\ell$ to be strongly convex

[1] We specify the explicit dependence on ρ, as it impacts the convergence rate of PGMs.

in the control. The function ℓ can be discontinuous and can take the value infinity, which are important characteristics of control problems with control constraints and entropy regularizations; see Examples 2.1, 2.2, and 2.3 for details. Note that these structural conditions in general do not imply convexity of the control objective J in either the open-loop or feedback controls.

Proximal PGMs for the control problem (1.1)–(1.2). By interpreting (1.1)–(1.2) as a minimization problem over $\mathcal{H}^2(\mathbb{R}^k)$, one can design a gradient descent algorithm for *open-loop* controls of the problem. Let $H^{\mathrm{re}} : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$ be such that for all $(t,x,a,y) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n$,

$$H_t^{\mathrm{re}}(x,a,y) := \langle b_t(x,a), y\rangle + f_t(x,a) - \rho\,\langle x,y\rangle, \tag{1.3}$$

and let $H : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}$ be the Hamiltonian such that for all $(t,x,a,y,z) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^{n\times d}$,

$$H_t(x,a,y,z) := H_t^{\mathrm{re}}(x,a,y) + \langle \sigma_t(x), z\rangle\,. \tag{1.4}$$

Note that H^{re} and H only involve the differentiable component of the running cost, while the nonsmooth component ℓ will be handled separately by the proximal map defined in (1.8). Then for an initial guess $\alpha^0 \in \mathcal{H}^2(\mathbb{R}^k)$ and a stepsize $\tau > 0$, consider the sequence $(\alpha^m)_{m\in\mathbb{N}} \subset \mathcal{H}^2(\mathbb{R}^k)$ such that for all $m \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$,

$$\alpha_t^{m+1} = \mathrm{prox}_{\tau\ell}\big(\alpha_t^m - \tau\partial_a H_t^{\mathrm{re}}(X_t^{\xi_0,\alpha^m}, \alpha_t^m, Y_t^{\xi_0,\alpha^m})\big) \quad \text{for } \mathrm{d}t \otimes \mathrm{d}\mathbb{P} \text{ a.e.}, \tag{1.5}$$

where $(X^{\xi_0,\alpha^m}, Y^{\xi_0,\alpha^m}, Z^{\xi_0,\alpha^m})$ are adapted processes satisfying the following forward-backward stochastic differential equation (FBSDE): for all $t \in [0,T]$,

$$\mathrm{d}X_t^{\xi_0,\alpha^m} = b_t(X_t^{\xi_0,\alpha^m}, \alpha_t^m)\,\mathrm{d}t + \sigma_t(X_t^{\xi_0,\alpha^m})\,\mathrm{d}W_t, \qquad X_0^{\xi_0,\alpha^m} = \xi_0, \tag{1.6}$$

$$\begin{aligned} \mathrm{d}Y_t^{\xi_0,\alpha^m} &= -\partial_x H_t(X_t^{\xi_0,\alpha^m}, \alpha_t^m, Y_t^{\xi_0,\alpha^m}, Z_t^{\xi_0,\alpha^m})\,\mathrm{d}t + Z_t^{\xi_0,\alpha^m}\,\mathrm{d}W_t, \\ Y_T^{\xi_0,\alpha^m} &= \partial_x g(X_T^{\xi_0,\alpha^m}), \end{aligned} \tag{1.7}$$

and $\mathrm{prox}_{\tau\ell} : \mathbb{R}^k \to \mathbb{R}^k$ is the proximal map of $\tau\ell$ defined by

$$\mathrm{prox}_{\tau\ell}(a) = \arg\min_{p\in\mathbb{R}^k} \left(\frac{1}{2}|p-a|^2 + \tau\ell(p)\right) \quad \forall a \in \mathbb{R}^k. \tag{1.8}$$

Note that (1.7) involves undiscounted costs and the term $\rho Y_t^{\xi_0,\alpha^m}$ and that it arises from the stochastic maximum principle for the discounted problem (see [29]).

The iteration (1.5) is a proximal gradient method for (1.2). In particular, the term $\partial_a H_t^{\mathrm{re}}(X_t^{\xi_0,\alpha^m}, \alpha_t^m, Y_t^{\xi_0,\alpha^m})$ is related to (up to an exponential time scaling) the Fréchet derivative of the differentiable component of $J(\cdot;\xi_0)$ at the iterate α^m, while the function $\mathrm{prox}_{\tau\ell}$ can be identified as the proximal map of the nonsmooth component of $J(\cdot;\xi_0)$ (see the proof of Theorem 3.13). We refer the reader to [36] for a detailed derivation of the algorithm and to [25, 39] for similar gradient-based algorithms without the nonsmooth term ℓ.

The main drawback of the proximal gradient algorithm (1.5) (as well as the algorithms in [25, 39]) is that it iterates over open-loop controls. As for each $m \in \mathbb{N}$ the iterate $\alpha^m \in \mathcal{H}^2(\mathbb{R}^k)$ is a stochastic process depending on the initial information and the driving Brownian noise terms from previous iterates, the iteration (1.5) is difficult to implement in practice. In what follows, we overcome the shortcoming of (1.5) and

introduce an analogue proximal gradient method for feedback controls of (1.1)–(1.2), which is referred to as the *proximal policy gradient method* (PPGM).

To this end, we consider a class $\mathcal{V}_{\mathbf{A}}$ of Lipschitz continuous policies $\phi : [0,T] \times \mathbb{R}^n \to \mathbb{R}^k$, whose precise definition is given in Definition 2.1. For a given initial guess $\phi^0 \in \mathcal{V}_{\mathbf{A}}$ and a stepsize $\tau > 0$, the PPGM generates the sequence $(\phi^m)_{m\in\mathbb{N}} \subset \mathcal{V}_{\mathbf{A}}$ such that for all $m \in \mathbb{N}_0$,

$$\phi_t^{m+1}(x) = \mathrm{prox}_{\tau\ell}\big(\phi_t^m(x) - \tau\partial_a H_t^{\mathrm{re}}(x, \phi_t^m(x), Y_t^{t,x,\phi^m})\big) \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n, \tag{1.9}$$

where $\mathrm{prox}_{\tau\ell}$ is defined in (1.8), and for each $\phi \in \mathcal{V}_{\mathbf{A}}$ and $(t,x) \in [0,T] \times \mathbb{R}^n$, $(X^{t,x,\phi}, Y^{t,x,\phi}, Z^{t,x,\phi})$ are adapted processes satisfying the following FBSDE: for all $s \in [t,T]$,

$$\mathrm{d}X_s^{t,x,\phi} = b_s(X_s^{t,x,\phi}, \phi_s(X_s^{t,x,\phi}))\,\mathrm{d}s + \sigma_s(X_s^{t,x,\phi})\,\mathrm{d}W_s, \quad X_t^{t,x,\phi} = x, \tag{1.10}$$

$$\begin{aligned} \mathrm{d}Y_s^{t,x,\phi} &= -\partial_x H_s(X_s^{t,x,\phi}, \phi_s(X_s^{t,x,\phi}), Y_s^{t,x,\phi}, Z_s^{t,x,\phi})\,\mathrm{d}s + Z_s^{t,x,\phi}\,\mathrm{d}W_s, \\ Y_T^{t,x,\phi} &= \partial_x g(X_T^{t,x,\phi}). \end{aligned} \tag{1.11}$$

The iteration (1.9) is motivated by the observation that if $\alpha_t^{\phi^m} = \phi_t^m(X_t^{\xi_0,\phi^m})$ with X^{ξ_0,ϕ^m} being the state process controlled by the policy ϕ^m, then for $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$ a.e., $Y_t^{\xi_0,\alpha^m} = Y_t^{t,x,\phi^m}\big|_{x=X_t^{\xi_0,\phi^m}}$ and

$$\partial_a H_t^{\mathrm{re}}(x, \phi_t^m(x), Y_t^{t,x,\phi^m})\big|_{x=X_t^{\xi_0,\phi^m}} = e^{\rho t}\left(\nabla_\alpha J_{\mathrm{diff}}(\alpha;\xi_0)\big|_{\alpha=\alpha^{\phi^m}}\right)_t,$$

where $J_{\mathrm{diff}}(\cdot;\xi_0)$ is the differentiable component of $J(\cdot;\xi_0)$ in (1.2): for all $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$,

$$J_{\mathrm{diff}}(\alpha;\xi_0) := \mathbb{E}\left[\int_0^T e^{-\rho t} f_s(X_s^{\xi_0,\alpha}, \alpha_s)\,\mathrm{d}s + e^{-\rho T} g(X_T^{\xi_0,\alpha})\right],$$

and $\nabla_\alpha J_{\mathrm{diff}}(\alpha;\xi_0) \in \mathcal{H}^2(\mathbb{R}^k)$ is the Fréchet derivative of J_{diff} at α. In other words, at the mth iteration, (1.9) evaluates the functional derivative of $J_{\mathrm{diff}}(\cdot;\xi_0)$ at the open-loop control α^{ϕ^m} induced by the current policy ϕ^m and obtains the update direction based on a Markovian representation of the gradient. This choice of gradient directions is crucial for the well-posedness and convergence of the policy iterates $(\phi^m)_{m\in\mathbb{N}_0}$ in (1.9); see the end of section 2.2 for a detailed comparison between the proposed gradient and the vanilla gradient direction of J_{diff} over feedback controls.

The PPGM (1.9) improves the efficiency of the policy iteration (see [24, 21]) and the method of successive approximation (see [27, 25]) by avoiding a pointwise minimization of the Hamiltonian over the action space, which may be expensive, especially in a high-dimensional setting. It has been successfully applied to high-dimensional control problems in [36] by solving the linear BSDE (1.11) numerically; see, e.g., [19, 36] and references therein for various numerical schemes.

Our contributions. This paper identifies conditions under which the PPGM (1.9) converges linearly to a stationary point of (1.1)–(1.2). These conditions allow for nonlinear state dynamics with degenerate noise, unbounded action space, and unbounded cost functions that are nonconvex in state and involve a nonsmooth regularizer in control. To the best of our knowledge, this is the first work which proposes a linearly convergent PGM for a continuous time finite horizon control problem. The convergence result theoretically underpins experimental observations where recent reinforcement learning heuristics, including entropy regularization or a fictitious discount factor, accelerate the convergence of PGMs.

We further prove that the PPGM (1.9) remains linearly convergent even if the FBSDEs are solved only approximately and the policies are updated based on these approximate gradients. This stability result allows for computationally efficient algorithms, as it shows that it is sufficient to solve the linear BSDEs with low accuracy at the initial iterations, while an accurate BSDE solver is only required for the last few iterations; a similar strategy has been used to design approximate policy iteration algorithms in [21].

Our approach and related works. There are various reasons for the relatively slow theoretical progress in PGMs for continuous time stochastic control problems. Due to the nonconvexity of most objective functions of control problems with respect to the policies, establishing linear convergence of PGMs can be linked to analyzing nonasymptotic performance of gradient search for nonconvex objectives, which has always been one of the formidable challenges in optimization theory. Allowing nonparametric policies in the algorithm further compounds the complexity, as the analysis has to be carried out in a suitable function space, instead of in a finite-dimensional parameter space.

Due to these technical difficulties, most existing works on linear performance guarantees of PGMs concentrate on discrete time control problems with specific policy parameterization. The arguments therein often require specific problem structure, in order to derive a suitable Polyak–Łojasiewicz inequality (also known as the gradient dominance property) for the loss landscape. For instance, in the tabular MDP setting, the policies must be uniformly lower bounded away from zero over the entire state space [30], while in the LQ setting, eigenvalues of state covariance matrices must be lower bounded away from zero over the entire time horizon [9, 15]. Consequently, these analyses are difficult to extend to general control problems (such as those with deterministic initial condition and degenerate noise) or to more sophisticated policy parameterizations (such as deep neural networks).

Here, we introduce a new analytical technique to analyze the PPGM (1.9) without relying on the Polyak–Łojasiewicz condition or convexity. By carrying out a precise regularity estimate of associated FBSDEs, we establish uniform Lipschitz continuity and uniform linear growth of the iterates $(\phi^m)_{m\in\mathbb{N}}$. These estimates further allow us to prove that $(\phi^m)_{m\in\mathbb{N}}$ forms a contraction in a weighted sup-norm, whose limit can be identified as a stationary point of (1.2). To the best of our knowledge, this is the first time BSDEs have been used to study convergence of PGMs.

Notation. For each Euclidean space $(E, |\cdot|)$, $t \in [0,T]$ and $p \geq 2$, we introduce the following spaces:

- $\mathcal{S}^p(t,T;E)$ is the space of E-valued $\mathbb{F}$-progressively measurable processes $Y : [t,T]\times\Omega \to E$ satisfying $\|Y\|_{\mathcal{S}^p} = \mathbb{E}[\sup_{s\in[t,T]} |Y_s|^p]^{1/p} < \infty$;[2]
- $\mathcal{H}^p(t,T;E)$ is the space of E-valued $\mathbb{F}$-progressively measurable processes $Z : [t,T]\times\Omega \to E$ satisfying $\|Z\|_{\mathcal{H}^p} = \mathbb{E}[(\int_t^T |Z_s|^2\,\mathrm{d}s)^{p/2}]^{1/p} < \infty$.

For notational simplicity, we denote $\mathcal{S}^p(E) = \mathcal{S}^p(0,T;E)$ and $\mathcal{H}^p(E) = \mathcal{H}^p(0,T;E)$.

2. Main results.

This section summarizes the model assumptions and presents the main results on the linear convergence of the PPGM (1.9).

2.1. Standing assumptions.

The following assumptions on the coefficients of (1.1)–(1.2) are imposed throughout the paper.

[2]With a slight abuse of notation, we denote by sup the essential supremum of a real-valued (Borel) measurable function.

H. 1. *Let $T > 0$, $\xi_0 \in L^2(\mathcal{F}_0;\mathbb{R}^n)$, $\ell : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\}$, $f : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$, $b : [0,T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ be measurable functions such that the following hold:*

(1) *ℓ is lower semicontinuous, and its effective domain $\boldsymbol{A} := \{z \in \mathbb{R}^k \,|\, \ell(z) < \infty\}$ is nonempty;*[3]

(2) *for all $t \in [0,T]$, $\mathbb{R}^n \times \mathbb{R}^k \ni (x,a) \mapsto f_t(x,a) \in \mathbb{R}$ is continuously differentiable, $|f_t(0,0)| < \infty$, and there exist constants $C_{fx}, C_{fa}, L_{fx}, L_{fa} \geq 0$ such that for all $t \in [0,T]$, $(x,a),(x',a') \in \mathbb{R}^n \times \boldsymbol{A}$,*

$$|\partial_x f_t(x,a)| \leq C_{fx}, \quad |\partial_x f_t(x,a) - \partial_x f_t(x',a')| \leq L_{fx}(|x-x'| + |a-a'|), \tag{2.1}$$

$$|\partial_a f_t(0,0)| \leq C_{fa}, \quad |\partial_a f_t(x,a) - \partial_a f_t(x',a')| \leq L_{fa}(|x-x'| + |a-a'|); \tag{2.2}$$

(3) *there exist constants $\mu, \nu \geq 0$ such that $\mu + \nu > 0$ and for all $(t,x) \in [0,T] \times \mathbb{R}^n$, $a, a' \in \boldsymbol{A}$, and $\eta \in [0,1]$,*

$$\eta f_t(x,a) + (1-\eta) f_t(x,a') \geq f_t(x, \eta a + (1-\eta)a') + \eta(1-\eta)\tfrac{\mu}{2}|a-a'|^2, \tag{2.3}$$

$$\eta \ell(a) + (1-\eta)\ell(a') \geq \ell(\eta a + (1-\eta)a') + \eta(1-\eta)\tfrac{\nu}{2}|a-a'|^2; \tag{2.4}$$

(4) *g is differentiable, and there exist constants $C_g, L_g \geq 0$ such that for all x, $x' \in \mathbb{R}^n$,*

$$|\partial_x g(x)| \leq C_g, \quad |\partial_x g(x) - \partial_x g(x')| \leq L_g |x-x'|; \tag{2.5}$$

(5) *there exist $\hat{b} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\bar{b} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times k}$ such that*

$$b_t(x,a) = \hat{b}_t(x) + \bar{b}_t(x)a \quad \forall (t,x,a) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^k, \tag{2.6}$$

with $\mathbb{R}^n \ni x \mapsto (\hat{b}_t(x), \bar{b}_t(x)) \in \mathbb{R}^n \times \mathbb{R}^{n\times k}$ differentiable for all $t \in [0,T]$, and there exist constants $C_{\hat{b}}, C_{\bar{b}}, L_{\hat{b}}, L_{\bar{b}} \geq 0$ and $\kappa_{\hat{b}} \in \mathbb{R}$ such that for all $t \in [0,T]$, $(x,a),(x',a') \in \mathbb{R}^n \times \boldsymbol{A}$,

$$|\hat{b}_t(0)| + |\partial_x \hat{b}_t(0)| \leq C_{\hat{b}}, \quad |\bar{b}_t(x)| \leq C_{\bar{b}}, \tag{2.7}$$

$$\langle x - x', \hat{b}_t(x) - \hat{b}_t(x') \rangle \leq \kappa_{\hat{b}} |x-x'|^2, \quad |\partial_x \hat{b}_t(x) - \partial_x \hat{b}_t(x')| \leq L_{\hat{b}}|x-x'|, \tag{2.8}$$

$$\begin{aligned} &|\bar{b}_t(x) - \bar{b}_t(x')| + |\bar{b}_t(x)a - \bar{b}_t(x')a'| + |\partial_x \bar{b}_t(x)a - \partial_x \bar{b}_t(x')a'| \\ &\qquad \leq L_{\bar{b}}(|x-x'| + |a-a'|); \end{aligned} \tag{2.9}$$

(6) *there exist constants $C_\sigma, L_\sigma \geq 0$ such that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^n$,*

$$|\sigma_t(x)| \leq C_\sigma, \quad |\sigma_t(x) - \sigma_t(x')| + |\partial_x \sigma_t(x) - \partial_x \sigma_t(x')| \leq L_\sigma |x-x'|. \tag{2.10}$$

Remark 2.1. The action set $\mathbf{A}$ may be unbounded, and hence (2.9) cannot be further simplified. If one assumes further that $\mathbf{A}$ is bounded, then, by the boundedness of $\bar{b}$ in (2.7), (2.9) is equivalent to the Lipschitz continuity of $\bar{b}$ and $\partial_x \bar{b}$. Alternatively, if $\mathbf{A}$ is unbounded, then (2.9) is equivalent to the condition that $\bar{b}$ is independent of x.

[3]We say a function $f : X \to \mathbb{R} \cup \{\infty\}$ is proper if it has a nonempty effective domain $\operatorname{dom} f := \{x \in X \mid f(x) < \infty\}$.

To consider nonlinear state-dependent drift and diffusion coefficients, we impose in (2.1) and (2.5) the boundedness conditions on the spatial partial derivatives of cost functions. Observe from (1.4) and (2.6) that $\partial_x H$ (resp., $\partial_a H$) involve the term $(\partial_x \hat{b}_t(x) + \partial_x \bar{b}_t(x)a)^\top y + \partial_x \sigma(x)^\top z$ (resp., $\bar{b}_t(x)^\top y$), whose modulus of continuity in x depends on the magnitude of y and z. By exploiting the boundedness of $\partial_x f$ and $\partial_x g$, we establish an a priori bound of the adjoint processes and subsequently prove that the iterative scheme (1.9) generates Lipschitz continuous policies $(\phi^m)_{m\in\mathbb{N}_0}$ (see Proposition 3.7). If the drift and diffusion coefficients are affine in x, then (2.1) and (2.5) can be relaxed to quadratically growing functions, which include as special cases the linear-convex control problems studied in [14, 41].

For clarity, (2.3) and (2.4) assume convexity of $a \mapsto f_t(x,a)$ and $a \mapsto \ell(a)$ and strong convexity of $a \mapsto f_t(x,a) + \ell(a)$. This allows for characterizing the rate of convergence of $(\phi^m)_{m\in\mathbb{N}_0}$ in terms of μ and ν. Similar analysis can be performed if (2.3) is relaxed into the following semiconvexity condition, i.e., there exists $\mu \in [-L_{fa}, L_{fa}]$ such that for all $(t,x) \in [0,T] \times \mathbb{R}^n$, $a, a' \in \mathbf{A}$, and $\eta \in [0,1]$,

$$\eta f_t(x,a) + (1-\eta) f_t(x,a') \geq f_t(x, \eta a + (1-\eta)a') + \eta(1-\eta)\tfrac{\mu}{2}|a-a'|^2,$$

and ℓ is ν-strongly convex with a sufficiently large ν (cf. condition (iii) below). Such an assumption allows f to be concave in a and can be satisfied if the objective function (1.2) involves entropy regularization (see Example 2.3).

Here we present several important nonsmooth costs used in engineering and machine learning.

Example 2.1 (control constraint). Let $\mathbf{A} \subset \mathbb{R}^k$ be a nonempty closed convex set and $\ell : \mathbb{R}^k \to [0,\infty]$ be the indicator of $\mathbf{A}$ satisfying $\ell(a) = 0$ for $a \in \mathbf{A}$ and $\ell(a) = \infty$ for $a \in \mathbb{R}^k \setminus \mathbf{A}$. Then (2.4) holds with $\nu = 0$, and for all $\tau > 0$, $\mathrm{prox}_{\tau\ell}$ is the orthogonal projection on $\mathbf{A}$. In this case, (1.9) extends the projected PGM in [15] to general stochastic control problems.

Example 2.2 (sparse control). Let $(\gamma_i)_{i=1}^k \subset [0,\infty)$, and let $\ell : \mathbb{R}^k \to [0,\infty)$ be such that $\ell(a) = \sum_{i=1}^k \gamma_i |a_i|$ for $a = (a_i)_{i=1}^k \in \mathbb{R}^k$. Then (2.4) holds with $\nu = 0$, and for all $\tau > 0$, $\mathrm{prox}_{\tau\ell}(a) = (\max\{|a_i| - \tau\gamma_i, 0\}\,\mathrm{sgn}(a_i))_{i=1}^k$ for each $a = (a_i)_{i=1}^k \in \mathbb{R}^k$. In this case, (1.9) can be viewed as an infinite-dimensional extension of the iterative shrinkage-thresholding algorithm (see [4, 36]).

Example 2.3 ($\mathfrak{f}$-divergence regularized control). Let $\Delta_k := \{a \in [0,1]^k \mid \sum_{i=1}^k a_i = 1\}$, let $\mathfrak{u} = (\mathfrak{u}_i)_{i=1}^k \in \Delta_k \cap (0,1)^k$, and let $\ell : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\}$ be the $\mathfrak{f}$-divergence defined by

$$\ell(a) := \sum_{i=1}^k \mathfrak{u}_i \mathfrak{f}\Big(\frac{a_i}{\mathfrak{u}_i}\Big), \quad a \in \Delta_k; \quad \ell(a) = \infty, \quad a \notin \Delta_k,$$

with a given lower semicontinuous function $\mathfrak{f} : [0,\infty) \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathfrak{f}(0) = \lim_{x\to 0} \mathfrak{f}(x)$, $\mathfrak{f}(1) = 0$ and being $\kappa_{\mathfrak{u}}$-strongly convex on $[0, \frac{1}{\min_i \mathfrak{u}_i}]$ with some $\kappa_{\mathfrak{u}} > 0$. As shown in [14, Example 2.2], ℓ satisfies (H.1) with $\nu = \frac{\kappa_{\mathfrak{u}}}{\max_i \mathfrak{u}_i} > 0$.

Note that an $\mathfrak{f}$-divergence ℓ is typically nondifferentiable and may have nonclosed effective domain $\mathbf{A}$ (see [14] for concrete examples). For commonly used forms of $\mathfrak{f}$-divergence, the proximal map prox_ℓ can be computed by solving (1.8) with Lagrange multipliers. For instance, let ℓ be the relative entropy corresponding to $\mathfrak{f}(s) = s\log s$, $s \in \mathbb{R}$. Then, for each $\tau > 0$ and $a = (a_i)_{i=1}^k \in \mathbb{R}^k$, $\mathrm{prox}_{\tau\ell}(a)_i = \tau W\left(\frac{\mathfrak{u}_i}{\tau}\exp\left(\frac{\lambda + a_i}{\tau} - 1\right)\right)$ for all $i = 1, \ldots, k$, where $W : [0,\infty) \to [0,\infty)$ is the Lambert W-function, and $\lambda \in \mathbb{R}$ is the unique solution to $\sum_{i=1}^k \tau W\left(\frac{\mathfrak{u}_i}{\tau}\exp\left(\frac{\lambda + a_i}{\tau} - 1\right)\right) = 1$.

2.2. Well-posedness of the iterates. In what follows, we focus on Lipschitz continuous feedback controls such that the corresponding controlled state dynamics (1.1) admits a strong solution. Due to the (possible) unboundedness of the action set $\mathbf{A}$, these controls in general grow linearly with respect to the state variable.

DEFINITION 2.1. *Let $\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k)$ be the space of measurable functions $\phi:[0,T]\times\mathbb{R}^n\to\mathbb{R}^k$, and let $|\cdot|_0,[\cdot]_1:\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k)\to[0,\infty]$ be such that for all $\phi\in\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k)$,*

$$|\phi|_0=\sup_{(t,x)\in[0,T]\times\mathbb{R}^n}\frac{|\phi_t(x)|}{1+|x|},\qquad [\phi]_1=\sup_{t\in[0,T],x,y\in\mathbb{R}^n,x\neq y}\frac{|\phi_t(x)-\phi_t(y)|}{|x-y|}.$$

We define the following space of feedback controls:

$$\mathcal{V}_{\boldsymbol{A}}:=\left\{\phi\in\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k)\,|\,|\phi|_0+[\phi]_1<\infty,\phi_t(x)\in\boldsymbol{A}\ \textit{for a.e.}(t,x)\in[0,T]\times\mathbb{R}^n\right\},\tag{2.11}$$

and for each $\phi\in\mathcal{V}_{\boldsymbol{A}}$, we define the associated control process $\alpha^\phi\in\mathcal{H}^2(\mathbb{R}^k)$ by $\alpha_t^\phi=\phi_t(X_t^{\xi_0,\phi})\,\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$-a.e., where $X^{\xi_0,\phi}\in\mathcal{S}^2(\mathbb{R}^n)$ is the solution to the following SDE (cf. (1.1)*):*

$$\mathrm{d}X_t=b_t(X_t,\phi_t(X_t))\,\mathrm{d}t+\sigma_t(X_t)\,\mathrm{d}W_t,\quad t\in[0,T];\quad X_0=\xi_0.\tag{2.12}$$

The Lipschitz regularity of $\phi\in\mathcal{V}_{\mathbf{A}}$ ensures that the system (1.10)–(1.11) admits a unique strong solution. We refer the reader to [24, 14] for sufficient conditions under which the control problem admits an optimal feedback control in the class $\mathcal{V}_{\mathbf{A}}$. However, we emphasise that in this work we do not require the control problem (1.1)–(1.2) to have an optimal feedback control. Instead, we focus on constructing a policy $\phi\in\mathcal{V}_{\mathbf{A}}$ whose associated (open-loop) control process is a stationary point of $J(\cdot;\xi_0)$; see section 2.3 for details.

Under (H.1), the iterative scheme (1.9) is well-defined for any given guess $\phi^0\in\mathcal{V}_{\mathbf{A}}$ and stepsize $\tau>0$. The proof of this relies on the well-posedness and stability of the FBSDEs (1.10)–(1.11), with extra difficulties arising from possibly non-Lipschitz and unbounded coefficients; i.e., $\hat{b}_t$ may be non-Lipschitz in x, and $\partial_x H$ non-Lipschitz in (x,y), and unbounded in x. The detailed arguments can be found in Appendix A of the arXiv version [37].

PROPOSITION 2.1. *Suppose* (H.1) *holds. Then, for all $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$ and $\tau>0$, the iterates $(\phi^m)_{m\in\mathbb{N}_0}$ are well-defined functions in $\mathcal{V}_{\boldsymbol{A}}$.*

Regularity of the gradient direction in (1.9). Here we emphasize the importance of the gradient direction in (1.9) on the well-posedness of the iterates $(\phi^m)_{m\in\mathbb{N}_0}$. Observe that if $\phi^m\in\mathcal{V}_{\mathbf{A}}$, then classical stability results of (1.10)–(1.11) imply that the map $x\mapsto Y_t^{t,x,\phi^m}$ in (1.9) is Lipschitz continuous uniformly in t, which subsequently ensures that $\phi^{m+1}\in\mathcal{V}_{\mathbf{A}}$. Such a Lipschitz regularity holds even if the diffusion coefficient of (1.1) degenerates, which includes deterministic control problems as special cases.

The above regularity estimate in general does not hold if one updates a feedback control using the gradient of J at the feedback map itself, especially when the diffusion coefficient of (1.1) degenerates. To see this, we assume for simplicity that all variables are one-dimensional and consider minimizing the following cost (with $\xi_0=x_0$ and $\sigma=0$ in (1.1) and $\ell=0$ and $\rho=0$ in (1.2)) over all $\phi\in\mathcal{V}_{\mathbf{A}}$:

$$J(\phi;x_0)=\int_0^T f_t(X_t^{x_0,\phi},\phi_t(X_t^{x_0,\phi}))\,\mathrm{d}t+g(X_T^{x_0,\phi}),\tag{2.13}$$

where for each $\phi \in \mathcal{V}_{\mathbf{A}}$, $X_t^{x_0,\phi} = x_0 + \int_0^t b_s(X_s^{x_0,\phi}, \phi_s(X_s^{x_0,\phi}))\,\mathrm{d}s$ for all $t \in [0,T]$. By [6, section 4.1], for any given $\psi \in \mathcal{V}_{\mathbf{A}}$, the derivative of J at $\phi \in \mathcal{V}_{\mathbf{A}}$ in the direction ψ is

$$\frac{\mathrm{d}J(\phi+\varepsilon\psi)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} = \int_0^T \partial_a H_t^{\mathrm{re}}(X_t^{x_0,\phi}, \phi_t(X_t^{x_0,\phi}), \partial_x \mathfrak{u}_t^{\phi}(X_t^{x_0,\phi}))\psi_t(X_t^{x_0,\phi})\,\mathrm{d}t, \tag{2.14}$$

where H^{re} is defined in (1.3), and $\mathfrak{u}^{\phi} : [0,T] \times \mathbb{R} \to \mathbb{R}$ satisfies that for all $(t,x) \in [0,T] \times \mathbb{R}$,

$$\partial_t \mathfrak{u}_t(x) + H_t^{\mathrm{re}}\big(x, \phi_t(x), \partial_x \mathfrak{u}_t(x)\big) = 0; \quad \mathfrak{u}_T(x) = g(x). \tag{2.15}$$

Observe that (2.14) is an L^2-inner product between $(t,x) \mapsto \partial_a H_t^{\mathrm{re}}(x, \phi_t(x), \partial_x \mathfrak{u}_t^{\phi}(x))$ and ψ with respect to the law of $X^{x_0,\phi}$. This leads to the following iterative scheme, which is a direct application of the gradient descent algorithm for feedback controls:

$$\phi_t^{m+1}(x) = \phi_t^m(x) - \tau \partial_a H_t^{\mathrm{re}}\big(x, \phi_t^m(x), \partial_x \mathfrak{u}_t^{\phi^m}(x)\big), \quad (t,x) \in [0,T] \times \mathbb{R}. \tag{2.16}$$

However, the iteration (2.16) in general does not preserve the regularity of the iterates $(\phi^m)_{m\in\mathbb{N}}$ and hence may not be well-defined. To see this, assume that $\phi^m \in \mathcal{V}_{\mathbf{A}}$ for some $m \in \mathbb{N}$. Then, by (2.16), the regularity of ϕ^{m+1} depends on the regularity of $\partial_x \mathfrak{u}^{\phi^m}$. Formally taking derivatives of (2.15) with respect to x implies that $w^m := \partial_x \mathfrak{u}^{\phi^m}$ satisfies the following PDE: for all $(t,x) \in [0,T] \times \mathbb{R}$,

$$\begin{aligned} &\partial_t w_t(x) + b_t(x, \phi_t^m(x))\partial_x w_t(x) + \partial_x H_t^{\mathrm{re}}\big(x, \phi_t^m(x), w_t(x)\big) \\ &\quad = -\partial_a H_t^{\mathrm{re}}\big(x, \phi_t^m(x), w_t(x)\big)\partial_x \phi_t^m(x), \end{aligned} \tag{2.17}$$

with $w_T(x) = \partial_x g(x)$. The term $\partial_a H^{\mathrm{re}}\,\partial_x \phi^m$ on the right-hand side of (2.17) appears due to the application of the chain rule. The Lipschitz continuity of ϕ^m only implies the boundedness of $\partial_x \phi^m$, and consequently both $\partial_x \mathfrak{u}^{\phi^m}$ and ϕ^{m+1} are in general not Lipschitz continuous. More crucially, it indicates that estimating the derivatives of ϕ^{m+1} requires bounds on higher order derivatives of ϕ^m, and it is unclear how to close this norm gap.

In contrast, such a loss of regularity does not occur in (1.9). Indeed, in the setting of (2.13), $Z^{t,x,\phi} \equiv 0$ in (1.11), and hence by the Feynman–Kac formula, (1.9) is equivalent to

$$\phi_t^{m+1}(x) = \phi_t^m(x) - \tau \partial_a H_t^{\mathrm{re}}(x, \phi_t^m(x), u_t^m(x)) \quad \forall (t,x) \in [0,T] \times \mathbb{R},$$

where u^m is the unique continuous viscosity solution to the following PDE: for all $(t,x) \in [0,T] \times \mathbb{R}$,

$$\partial_t u_t(x) + b_t(x, \phi_t^m(x))\partial_x u_t(x) + \partial_x H_t^{\mathrm{re}}(x, \phi_t^m(x), u_t(x)) = 0, \quad u_T(x) = \partial_x g(x). \tag{2.18}$$

Note that (2.18) does not involve the term $\partial_x H_t^{\mathrm{re}} \partial_x \phi^m$, and under (H.1), all coefficients of (2.18) are sufficiently regular such that u^m is indeed Lipschitz continuous in x (uniformly in t), according to classical Lipschitz estimates of viscosity solution (see, e.g., [2]).

2.3. Linear convergence of the iterates. The main contribution of this article is to identify conditions under which $(\phi^m)_{m\in\mathbb{N}_0} \subset \mathcal{V}_{\mathbf{A}}$ converge linearly to a

stationary point of the control problem (1.1)–(1.2). As the functional $J(\cdot;\xi_0) : \mathcal{H}^2(\mathbb{R}^k) \to \mathbb{R} \cup \{\infty\}$ is typically nonsmooth and nonconvex, we first recall a notion of stationary points for nonsmooth nonconvex functionals on Hilbert spaces, defined as in [31]. By [31, Proposition 1.114], every local minimizer $\alpha^\star \in \operatorname{dom} J(\cdot;\xi_0)$ is a stationary point in the sense of Definition 2.2. In practice, a stationary point found in this way often gives a good solution candidate [27].

DEFINITION 2.2. *Let X be a Hilbert space equipped with the norm $\|\cdot\|_X$ and the inner product $\langle\cdot,\cdot\rangle_X$, $F : X \to \mathbb{R} \cup \{\infty\}$, and let $x^\star \in \operatorname{dom} F = \{x \in X \mid F(x) < \infty\}$. The Fréchet subdifferential of F at $x^\star$ is defined by*

$$\partial F(x^\star) = \left\{ \bar{x} \in X \,\middle|\, \liminf_{x \to x^\star} \frac{F(x) - F(x^\star) - \langle \bar{x}, x - x^\star\rangle_X}{\|x - x^\star\|_X} \geq 0 \right\}.$$

We say $x^\star \in \operatorname{dom} F$ is a stationary point of F if $0 \in \partial F(x^\star)$.

As alluded to earlier, the map $\mathcal{V}_{\mathbf{A}} \ni \phi \mapsto J(\alpha^\phi;\xi_0) \in \mathbb{R}\cup\{\infty\}$ is typically nonconvex and may not satisfy the Polyak–Łojasiewicz condition as in the setting with parametric policies (see [9, 43, 30, 12, 15, 23]). Hence to ensure the linear convergence of the PPGM (1.9), we impose further conditions on the coefficients which guarantee that we are in *one of the following six cases:*

(i) Time horizon T is small.
(ii) Discount factor ρ is large.
(iii) Running cost is sufficiently convex in control, i.e., $\mu + \nu$ is sufficiently large.
(iv) Costs depend weakly on state, i.e., C_{fx}, L_{fx}, C_g, and L_g are small.
(v) Control affects state dynamics weakly, i.e., $C_{\bar{b}}$ is small.
(vi) State dynamics is strongly dissipative; i.e., $\kappa_{\hat{b}}$ is sufficiently negative.

The above conditions will be made precise in (3.25) and (3.43). Here we give some practical implications of these conditions.

Remark 2.2. Conditions (i) and (ii) are commonly used conditions to ensure the convergence of iterative algorithms for nonconvex problems (see, e.g., [5, 3, 18, 21]). Condition (ii) also justifies the use of a fictitious discount factor to accelerate the convergence of PGMs for continuous time control problems (see [13] and references therein).

Conditions (iii)–(v) help to ease the nonconvexity of $\phi \mapsto J(\alpha^\phi;\xi_0)$ and to reduce the oscillation of the loss function's curvature, which subsequently promotes the convergence of gradient-based algorithms (see [33]). Condition (iii), along with Example 2.3, also justifies recent reinforcement learning heuristics that adding $\mathfrak{f}$-divergences, such as the relative entropy, to the optimization objective can accelerate the convergence of PGMs (see, e.g., [39, 23]).

Condition (vi) indicates that a strong dissipativity of the state dynamics enhances the efficiency of learning algorithms. Such a phenomenon has already been observed in the LQ setting with $\hat{b}_t(x) = A_t x$ in (2.6), where the desired dissipativity can be ensured if eigenvalues of A_t are sufficiently negative (see [18, 15]). Condition (vi) also motivates a residual correction method for solving nonlinear control problems. Consider a control problem (1.1)–(1.2) whose drift involves nondissipative coefficient $\hat{b}$. Then one can search feedback controls of the form $\phi = \overline{\phi} + \widetilde{\phi}$, where $\overline{\phi}$ is a precomputed candidate policy, and $\tilde{\phi}$ is an unknown residual correction. Observe that the state dynamics now has the drift coefficient $b = (\hat{b} + \bar{b}\overline{\phi}) + \bar{b}\widetilde{\phi}$, and the function $\hat{b} + \bar{b}\overline{\phi}$ may be dissipative for suitably chosen policy $\overline{\phi}$; see [36] and references therein for computing

$\overline{\phi}$ via linearization and the efficiency improvement of the residual correction method over plain PGMs.

Now we present the main theorem on the linear convergence of the PPGM (1.9) as m tends to infinity. The precise statement and proof will be given in section 3.5 (see Theorem 3.13).

THEOREM 2.2. *Suppose* (H.1) *holds. For all* $\phi^0 \in \mathcal{V}_{\boldsymbol{A}}$ *and* $\tau \in (0, \frac{2}{\mu + L_{fa}} \wedge \frac{1}{\nu}]$, *if one of conditions* (i)–(vi) *holds, then there exist* $\phi^\star \in \mathcal{V}_{\boldsymbol{A}}$ *and constants* $c \in [0,1)$ *and* $\widetilde{C} \geq 0$ *such that the following hold:*

(1) $\alpha^{\phi^\star}$ *is a stationary point of* $J(\cdot;\xi_0) : \mathcal{H}^2(\mathbb{R}^k) \to \mathbb{R} \cup \{\infty\}$ *defined as in* (1.2)*;*

(2) *for all* $m \in \mathbb{N}_0$, $|\phi^{m+1} - \phi^\star|_0 \leq c|\phi^m - \phi^\star|_0$ *and* $\|\alpha^{\phi^m} - \alpha^{\phi^\star}\|_{\mathcal{H}^2} \leq \widetilde{C} c^m$.

The precise constant c, which determines the rate of convergence, will be given in the proof based on conditions (i)–(vi). Roughly speaking, the stronger the cost convexity (resp., the stronger the state dissipativity, the weaker the state/control coupling, the smaller the time horizon, and the larger the discount factor), the smaller one can choose c, and hence the faster the iteration converges.

Note that Theorem 2.2 does not require nondegeneracy of ξ_0 and σ and can be extended to quadratically growing cost functions (see Remark 2.1). As (1.9) concerns iterations of unbounded and nonlinear feedback controls, the proof of convergence is rather technical. Here we outline the key steps for the reader's convenience.

***Sketched proof of Theorem* 2.2**. Observe that a necessary condition on the convergence of $(\alpha^{\phi^m})_{m\in\mathbb{N}_0}$ is that $(\|X^{\xi_0,\phi^m}\|_{\mathcal{H}^2})_{m\in\mathbb{N}_0}$ are uniformly bounded in m. By standard moment estimates of SDEs, it seems unavoidable to control the Lipschitz constant of $(\phi^m)_{m\in\mathbb{N}}$ in order to obtain the desired convergence result. This uniform regularity estimate is the main technical difficulty in analyzing (1.9), compared with the analyses of iterative algorithms for open-loop controls in [27, 39, 25].

To this end, suppose that $\phi^m \in \mathcal{V}_{\mathbf{A}}$ for a given $m \in \mathbb{N}_0$. By exploiting (1.9) and the convexity of f and ℓ, for all sufficiently small $\tau > 0$,

$$[\phi^{m+1}]_1 \leq (1-\tau C)[\phi^m]_1 + \tau C \left(\sup_{t,x,x'} \frac{|Y_t^{t,x,\phi^m} - Y_t^{t,x',\phi^m}|}{|x-x'|} + \sup_{t,x} |Y_t^{t,x,\phi^m}| + 1 \right),$$

where the constant $C > 0$ depends only on coefficients (see Lemma 3.4). An a priori estimate of (1.11) and the boundedness of $\partial_x f$ and $\partial_x g$ imply that $\sup_{m,t,x} |Y_t^{t,x,\phi^m}| < \infty$, while Lipschitz estimates of (1.10) and (1.11) imply that $x \mapsto Y_t^{t,x,\phi^m}$ is Lipschitz continuous uniformly in t, where the Lipschitz constant $L_Y([\phi^m]_1)$ depends *exponentially* on $[\phi^m]_1$ due to the feedback controlled dynamics (1.10) (see Proposition 3.7). Combining these estimates gives $[\phi^{m+1}]_1 \leq (1-\tau C)[\phi^m]_1 + \tau C(L_Y([\phi^m]_1) + 1)$. We then show in Theorem 3.8 that under suitable conditions on the coefficients, such an exponential dependence can be controlled, and we can further deduce that $\sup_m [\phi^m]_1 < \infty$.

We then proceed to prove the linear convergence of $(\phi^m)_{m\in\mathbb{N}}$. Using the strong convexity of costs, for sufficiently small $\tau > 0$,

(2.19)
$$|\phi^{m+1} - \phi^m|_0 \leq (1-\tau C)|\phi^m - \phi^{m-1}|_0 + \tau C \sup_{t,x} \frac{|Y_t^{t,x,\phi^m} - Y_t^{t,x,\phi^{m-1}}|}{1+|x|} \quad \forall m \in \mathbb{N}.$$

Based on $\sup_m [\phi^m]_1 < \infty$, we prove by Malliavin calculus that $\sup_{m,t,x,s} |Z_s^{t,x,\phi^m}| < \infty$ (see Lemma 3.10) and further by stability estimates of (1.10) and (1.11) that

$|Y_t^{t,x,\phi^m} - Y_t^{t,x,\phi^{m-1}}| \le \tilde{C}(1+|x|)|\phi^m - \phi^{m-1}|_0$ for some constant $\tilde{C}$ independent of t, x, m (see Proposition 3.9). By quantifying C in (2.19) and $\tilde{C}$ precisely, we prove under each of the conditions (i)–(vi) that there exists $c \in [0,1)$ such that $|\phi^{m+1} - \phi^m|_0 \le c|\phi^m - \phi^{m-1}|_0$ for all m, which subsequently implies the convergence of $(\phi^m)_{m\in\mathbb{N}}$ due to Banach's fixed point theorem. Finally, we show that the limit of $(\phi^m)_{m\in\mathbb{N}}$ induces a stationary point of $J(\cdot;\xi_0)$, based on an equivalent characterization of stationary points of $J(\cdot;\xi_0)$ in terms of adjoint processes and the proximal map of ℓ (see Theorem 3.13). ∎

In practice, (1.11) can only be solved approximately and the update step (1.9) for the feedback controls can only be performed with this approximate solution, which feeds the errors into subsequent iterations. Hence we further quantify this effect by establishing a stability property of (1.9) under perturbations of solutions to (1.11). For clarity, we only carry out perturbation analysis for the computation of $Y_t^{t,x,\phi}$, but similar analysis can be performed for (1.9) with inexact computation of the proximal map $\mathrm{prox}_{\tau\ell}$. Our analysis allows for stochastic approximations of $Y_t^{t,x,\phi}$ resulting from applying probabilistic numerical methods to solve (1.11) (see, e.g. [11, 17]).

More precisely, let $\phi^0 \in \mathcal{V}_{\mathbf{A}}$ be an initial guess and $\tau > 0$ be a stepsize. At the mth iteration with $m \in \mathbb{N}_0$, let $\widetilde{\phi}^m$ be the (random) feedback control obtained at the previous iteration (with $\widetilde{\phi}^0 = \phi^0$). That is, $\widetilde{\phi}^m : [0,T] \times \mathbb{R}^n \times \Omega \to \mathbf{A}$ is a measurable function such that $\widetilde{\phi}^m_\cdot(\cdot,\omega) \in \mathcal{V}_{\mathbf{A}}$ for a.s. $\omega \in \Omega$. Consider a measurable function $\widetilde{\mathcal{Y}}^{\widetilde{\phi}^m} : [0,T] \times \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ such that for a.s. $\omega \in \Omega$, $(t,x) \mapsto \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x,\omega)$ approximates $(t,x) \mapsto \mathcal{Y}_t^{\widetilde{\phi}^m}(x,\omega) := Y_t^{t,x,\widetilde{\phi}^m_\cdot(\cdot,\omega)}$, where $Y_t^{t,x,\widetilde{\phi}^m_\cdot(\cdot,\omega)} \in \mathbb{R}^n$ satisfies (1.11) with the realized control $\widetilde{\phi}^m_\cdot(\cdot,\omega) \in \mathcal{V}_{\mathbf{A}}$. The (random) feedback control for the next iteration is then obtained via a proximal gradient update (1.9) based on $\widetilde{\mathcal{Y}}^{\widetilde{\phi}^m}$:

$$\widetilde{\phi}_t^{m+1}(x) = \mathrm{prox}_{\tau\ell}\big(\widetilde{\phi}_t^m(x) - \tau\partial_a H_t^{\mathrm{re}}(x, \widetilde{\phi}_t^m(x), \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x))\big) \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n, \tag{2.20}$$

where the identity is understood in an almost sure sense.

The following theorem shows the accuracy of (2.20), whose precise statement and proof will be given in section 3.5 (see Theorem 3.14). Here we assume that $\widetilde{\mathcal{Y}}^{\widetilde{\phi}^m}$ approximates the function $\mathcal{Y}^{\widetilde{\phi}^m}$ well enough such that the resulting controls $\widetilde{\phi}^m$ are uniformly bounded in time and uniformly Lipschitz in space.

THEOREM 2.3. *Suppose* (H.1) *holds. For all* $\phi^0 \in \mathcal{V}_{\boldsymbol{A}}$ *and* $\tau \in (0, \frac{2}{\mu + L_{fa}} \wedge \frac{1}{\nu}]$, *if* $\sup_{m\in\mathbb{N},\omega\in\Omega}(|\widetilde{\phi}^m_\cdot(\cdot,\omega)|_0 + [\widetilde{\phi}^m_\cdot(\cdot,\omega)]_1) < \infty$, *and one of the conditions* (i)–(vi) *holds, then there exist constants* $c \in [0,1)$ *and* $C \ge 0$ *such that for a.s.* $\omega \in \Omega$ *and for all* $m \in \mathbb{N}_0$,

$$|\widetilde{\phi}^m_\cdot(\cdot,\omega) - \phi^\star|_0 \le c^m|\phi^0 - \phi^\star|_0 + C\sum_{j=0}^{m-1} c^{m-1-j}|\mathcal{Y}^{\widetilde{\phi}^j}_\cdot(\cdot,\omega) - \widetilde{\mathcal{Y}}^{\widetilde{\phi}^j}_\cdot(\cdot,\omega)|_0,$$

where $\phi^\star \in \mathcal{V}_{\boldsymbol{A}}$ *is the limit function in Theorem* 2.2. *Consequently, for all* $p \ge 1$ *and* $m \in \mathbb{N}_0$,

$$\mathbb{E}[|\widetilde{\phi}^m - \phi^\star|_0^p]^{\frac{1}{p}} \le c^m|\phi^0 - \phi^\star|_0 + C\sum_{j=0}^{m-1} c^{m-1-j}\mathbb{E}[|\mathcal{Y}^{\widetilde{\phi}^j} - \widetilde{\mathcal{Y}}^{\widetilde{\phi}^j}|_0^p]^{\frac{1}{p}}.$$

By Lemma 3.4, the condition $\sup_{m\in\mathbb{N}}(|\widetilde{\phi}^m|_0 + [\widetilde{\phi}^m]_1) < \infty$ holds if there exists $C > 0$ such that for all $m \in \mathbb{N}$, $\omega \in \Omega$, and $(t,x,x') \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n$, $|\widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x,\omega)| \le C$

and $|\widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x,\omega) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x',\omega)| \le C|x-x'|$. This highlights the fact that the numerical approximations of the gradient directions must be sufficiently regular in space, to prevent a spatial oscillation of the iterates and to ensure the convergence of the iterates. This is a reasonable assumption, as the exact gradient directions $(\mathcal{Y}^{\phi^m})_{m\in\mathbb{N}_0}$ enjoy these properties (see Propositions 3.5 and 3.7 and Theorem 3.8), and any reasonable approximation $\widetilde{\mathcal{Y}}^{\widetilde{\phi}^m}$ of $\mathcal{Y}^{\widetilde{\phi}^m}$ should retain these properties; see, e.g., [2] for approximation schemes that preserve boundedness and Lipschitz continuity of exact solutions. It would be interesting to derive explicit conditions on model coefficients to ensure the required regularity of $(\widetilde{\phi}^m)_{m\in\mathbb{N}_0}$. This would entail imposing precise dependencies of the Lipschitz regularity of $\widetilde{\mathcal{Y}}^{\widetilde{\phi}^m}$ on the seminorm $[\widetilde{\phi}^m]_1$, and is left for future research.

3. Proofs.

Throughout the rest of this work, we establish estimates with explicit dependence on the constants $T, \rho, C_{fx}, L_{fx}, L_{fa}, \mu, \nu, C_g, L_g, \kappa_{\hat{b}}, C_{\bar{b}}$, which are important for the convergence of (1.9). For notational simplicity, we write $(x)_+ = \max(0,x)$ for all $x\in\mathbb{R}$ and denote by $C>0$ a generic constant which depends on the remaining constants appearing in (H.1) and may take a different value at each occurrence. We shall refer to $C>0$ as an absolute constant if its value is independent of the constants in (H.1). Dependence of C on important quantities will be indicated explicitly by $C_{(\cdot)}$, e.g., $C_{(\phi)}$ for $\phi\in\mathcal{V}_{\mathbf{A}}$.

3.1. Auxiliary lemmas.

In this section, we present some technical lemmas used in the subsequent analysis. The following lemma establishes stability of SDEs with non-Lipschitz drift coefficients. The upper bounds involve explicit dependence on relevant constants, whose proof is given in Appendix A of the arXiv version [37].

LEMMA 3.1. *Let $T>0$, and for each $i=1,2$, let $\mu_i\in\mathbb{R}$, $\nu_i\ge 0$, let $b^i:[0,T]\times\mathbb{R}^n\to\mathbb{R}^n$ and $\sigma^i:[0,T]\times\mathbb{R}^n\to\mathbb{R}^{n\times d}$ be measurable functions such that for all $t\in[0,T]$ and $x,x'\in\mathbb{R}^n$, $\sup_{(t,x)\in[0,T]\times\mathbb{R}^n}\frac{|b_t^i(x)|+|\sigma_t^i(x)|}{1+|x|}<\infty$, $\langle x-x', b_t^i(x)-b_t^i(x')\rangle\le\mu_i|x-x'|^2$, and $|\sigma_t^i(x)-\sigma_t^i(x')|\le\nu_i|x-x'|$, and for each $(t,x)\in[0,T]\times\mathbb{R}^n$, let $X^{t,x,i}\in\mathcal{S}^2(t,T;\mathbb{R}^n)$ satisfy*

$$\mathrm{d}X_s = b_s^i(X_s)\,\mathrm{d}s + \sigma_s^i(X_s)\,\mathrm{d}W_s,\quad s\in[t,T];\quad X_t=x. \tag{3.1}$$

Then for all $p\ge 2$ there exists an absolute constant $C_{(p)}$ such that for all $t\in[0,T]$, $x_1,x_2\in\mathbb{R}^n$,

$$\begin{aligned}&\|X^{t,x_1,1}-X^{t,x_2,2}\|_{\mathcal{S}^p}\\ &\quad\le C_{(p)}e^{T(2\mu_1+C_{(p)}\nu_1^2)_+}\Big(|x_1-x_2|+\sqrt{T}\|b^1(X^{t,x_2,2})-b^2(X^{t,x_2,2})\|_{\mathcal{H}^p}\\ &\qquad\qquad+\|\sigma^1(X^{t,x_2,2})-\sigma^2(X^{t,x_2,2})\|_{\mathcal{H}^p}\Big).\end{aligned}$$

If we further assume that $\sigma^1\equiv\sigma^2$, then for all $(t,x)\in[0,T]\times\mathbb{R}^n$,

$$\mathbb{E}\left[|X_T^{t,x,1}-X_T^{t,x,2}|^2\right]\le\mathbb{E}\left[\int_t^T|b_s^1(X_s^{t,x,2})-b_s^2(X_s^{t,x,2})|^2e^{(T-s)(2\mu_1+\nu_1^2+1)}\,\mathrm{d}s\right]. \tag{3.2}$$

The following lemma establishes stability of BSDEs with monotone nonlinearity. It has been proved in [34] for $p=1$ and in [7, Proposition 3.2] for $p>1$.

LEMMA 3.2. *For each $i=1,2$ and $t\in[0,T]$, let $\xi^i\in L^2(\mathcal{F}_T;\mathbb{R}^n)$, $\gamma_i\ge 0$, and $\mu_i\in\mathbb{R}$, and let $f^i:[t,T]\times\Omega\times\mathbb{R}^n\times\mathbb{R}^{n\times d}\to\mathbb{R}^n$ be such that for all $(y,z)\in\mathbb{R}^n\times\mathbb{R}^{n\times d}$,*

$(f_s^i(\cdot,y,z))_{s\in[t,T]}$ is progressively measurable, and for all $(s,\omega)\in[t,T]\times\Omega$, $y,y'\in\mathbb{R}^n$ and $z,z'\in\mathbb{R}^{n\times d}$, $|f_s^i(\omega,y,z)-f_s^i(\omega,y,z')|\le\gamma_i|z-z'|$ and $\langle y-y',f_s^i(\omega,y,z)-f_s^i(\omega,y',z)\rangle\le\mu_i|y-y'|^2$, and let $(Y^i,Z^i)\in\mathcal{S}^2(t,T;\mathbb{R}^n)\times\mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ satisfy

$$\mathrm{d}Y_s=-f_s^i(\cdot,Y_s,Z_s)\,\mathrm{d}s+Z_s\,\mathrm{d}W_s,\quad s\in[t,T];\quad Y_T=\xi^i.$$

Then, for all $p\ge1$ and $\varepsilon\in(0,1)$, there exists an absolute constant $C_{(p,\varepsilon)}>0$ such that for all $t\in[0,T]$ and $\alpha\ge\varepsilon^{-1}\gamma_1^2+2\mu_1$,

$$\begin{aligned}&\mathbb{E}\bigg[\sup_{s\in[t,T]}e^{p\alpha s}|Y_s^1-Y_s^2|^{2p}+\bigg(\int_t^T e^{\alpha s}|Z_s^1-Z_s^2|^2\,\mathrm{d}s\bigg)^p\bigg]\\&\quad\le C_{(p,\varepsilon)}\mathbb{E}\bigg[e^{p\alpha T}|\xi^1-\xi^2|^{2p}+\bigg(\int_t^T e^{\frac{\alpha}{2}s}|f_s^1(\cdot,Y_s^2,Z_s^2)-f_s^2(\cdot,Y_s^2,Z_s^2)|\,\mathrm{d}s\bigg)^{2p}\bigg].\end{aligned}$$

The next lemma estimates the monotonicity and Lipschitz continuity of $\partial_x H$. The proof follows directly from (1.4) and (H.1) and is given in Appendix A of the arXiv version [37]. Due to the presence of $(\partial_x\hat{b}_t(x))^\top y$ and the unboundedness of $\partial_x\hat{b}$, $\partial_x H$ is not globally Lipschitz continuous in y.

LEMMA 3.3. *Suppose* (H.1) *holds, and let H be defined by* (1.4). *Then, for all $t\in[0,T]$, $x,x'\in\mathbb{R}^n$, $a,a'\in\boldsymbol{A}$, $y,y'\in\mathbb{R}^n$ and $z,z'\in\mathbb{R}^{n\times d}$,*

$$\begin{aligned}&\langle y-y',\partial_x H_t(x,a,y,z)-\partial_x H_t(x,a,y',z)\rangle\\&\quad\le(\kappa_{\hat{b}}-\rho+L_{\bar{b}})|y-y'|^2,|\partial_x H_t(x,a,y,z)-\partial_x H_t(x',a',y,z')|\\&\quad\le\big(L_{\hat{b}}|x-x'|+L_{\bar{b}}(|x-x'|+|a-a'|)\big)|y|+L_\sigma|x-x'||z|\\&\qquad+L_\sigma|z-z'|+L_{fx}(|x-x'|+|a-a'|).\end{aligned}\tag{3.3, 3.4}$$

We then present a Lipschitz estimate for the proximal gradient mapping (1.9).

LEMMA 3.4. *Suppose* (H.1) *holds, and let H^{re} be defined as in* (1.3). *Then, for all $t\in[0,T]$, $x,x'\in\mathbb{R}^n$, $a,a'\in\boldsymbol{A}$, $y,y'\in\mathbb{R}^n$ and $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$,*

$$\begin{aligned}&|\mathrm{prox}_{\tau\ell}(a-\tau\partial_a H_t^{re}(x,a,y))-\mathrm{prox}_{\tau\ell}(a'-\tau\partial_a H_t^{re}(x',a',y'))|\\&\quad\le\bigg(1-\tau\frac{1}{2}\bigg(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\bigg)\bigg)|a-a'|+\tau C_{\bar{b}}|y-y'|+\tau(L_{\bar{b}}|y'|+L_{fa})|x-x'|.\end{aligned}$$

Proof. For each $\tau>0$, since $\tau\ell$ is proper, lower-semicontinuous, and $\tau\nu$-strongly convex (cf. (2.4)), by Theorem 12.56 and Exercise 12.59 in [38],

$$|\mathrm{prox}_{\tau\ell}(x)-\mathrm{prox}_{\tau\ell}(y)|\le\frac{1}{1+\tau\nu}|x-y|\quad\forall x,y\in\mathbb{R}^k.$$

Hence, for any $t\in[0,T]$, $x,x'\in\mathbb{R}^n$, $a,a'\in\mathbf{A}$, and $y,y'\in\mathbb{R}^n$,

$$\begin{aligned}&|\mathrm{prox}_{\tau\ell}(a-\tau\partial_a H_t^{\mathrm{re}}(x,a,y))-\mathrm{prox}_{\tau\ell}(a'-\tau\partial_a H_t^{\mathrm{re}}(x',a',y'))|\\&\quad\le\tfrac{1}{1+\tau\nu}|(a-\tau\partial_a H_t^{\mathrm{re}}(x,a,y))-(a'-\tau\partial_a H_t^{\mathrm{re}}(x',a',y'))|\\&\quad\le\tfrac{1}{1+\tau\nu}|(a-\tau\partial_a H_t^{\mathrm{re}}(x,a,y))-(a'-\tau\partial_a H_t^{\mathrm{re}}(x,a',y))|\\&\qquad+\tfrac{\tau}{1+\tau\nu}|\partial_a H_t^{\mathrm{re}}(x,a',y)-\partial_a H_t^{\mathrm{re}}(x',a',y')|.\end{aligned}\tag{3.5}$$

We now estimate the two terms in (3.5) separately. Observe that $\partial_a H_t^{\mathrm{re}}(x,a,y) = \bar{b}_t(x)^\top y + \partial_a f_t(x,a)$ for all $(t,x,a,y) \in [0,T] \times \mathbb{R}^n \times \mathbf{A} \times \mathbb{R}^n$. Then, by (2.7), (2.1), and (2.2), the second term in (3.5) can be bounded by

$$\begin{aligned} &|\partial_a H_t^{\mathrm{re}}(x,a',y) - \partial_a H_t^{\mathrm{re}}(x',a',y')| \\ &\quad \le |\bar{b}_t(x)^\top y - \bar{b}_t(x')^\top y'| + |\partial_a f_t(x,a') - \partial_a f_t(x',a')| \\ &\quad \le C_{\bar{b}}|y-y'| + (L_{\bar{b}}|y'| + L_{fa})|x-x'|. \end{aligned} \tag{3.6}$$

To estimate the first term in (3.5), observe that for all $(t,x,y) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n$, by (2.6), (2.7), (2.2), and (2.3), $\mathbf{A} \ni a \mapsto H_t^{\mathrm{re}}(x,a,y) \in \mathbb{R}$ is μ-strongly convex, and $\mathbf{A} \ni a \mapsto \partial_a H_t^{\mathrm{re}}(x,a,y) \in \mathbb{R}^k$ is L_{fa}-Lipschitz continuous, which along with [33, Theorem 2.1.12], implies that for all $a, a' \in \mathbf{A}$,

$$\begin{aligned} &\langle \partial_a H_t^{\mathrm{re}}(x,a,y) - \partial_a H_t^{\mathrm{re}}(x,a',y), a-a' \rangle \\ &\quad \ge \frac{\mu L_{fa}}{\mu + L_{fa}}|a-a'|^2 + \frac{1}{\mu + L_{fa}}|\partial_a H_t^{\mathrm{re}}(x,a,y) - \partial_a H_t^{\mathrm{re}}(x,a',y)|^2. \end{aligned}$$

Hence, for all $a, a' \in \mathbf{A}$ and $(t,x,y) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n$, and $\tau \in (0, \frac{2}{\mu + L_{fa}}]$,

$$\begin{aligned} &|(a - \tau\partial_a H_t^{\mathrm{re}}(x,a,y)) - (a' - \tau\partial_a H_t^{\mathrm{re}}(x,a',y))|^2 \\ &\quad = |a-a'|^2 - 2\tau\langle a-a', \partial_a H_t^{\mathrm{re}}(x,a,y) - \partial_a H_t^{\mathrm{re}}(x,a',y)\rangle \\ &\qquad + \tau^2|\partial_a H_t^{\mathrm{re}}(x,a,y) - \partial_a H_t^{\mathrm{re}}(x,a',y)|^2 \\ &\quad \le \left(1 - 2\tau\frac{\mu L_{fa}}{\mu + L_{fa}}\right)|a-a'|^2 + \tau\left(\tau - \frac{2}{\mu + L_{fa}}\right)|\partial_a H_t^{\mathrm{re}}(x,a,y) - \partial_a H_t^{\mathrm{re}}(x,a',y)|^2 \\ &\quad \le \left(1 - 2\tau\frac{\mu L_{fa}}{\mu + L_{fa}}\right)|a-a'|^2. \end{aligned}$$

Taking the square root of both sides of the above estimate and using the inequality $\sqrt{1-\gamma\tau} \le 1 - \gamma\tau/2$ for all $\gamma, \tau \ge 0$ and $\gamma\tau \le 1$ give that

$$|(a - \tau\partial_a H_t^{\mathrm{re}}(x,a,y)) - (a' - \tau\partial_a H_t^{\mathrm{re}}(x,a',y))| \le \left(1 - \tau\frac{\mu L_{fa}}{\mu + L_{fa}}\right)|a-a'|.$$

This, along with (3.5), (3.6), and $\frac{\tau}{1+\tau\nu} \le \tau$ shows that for all $\tau \in (0, \frac{2}{\mu + L_{fa}}]$,

$$\begin{aligned} &|\mathrm{prox}_{\tau\ell}(a - \tau\partial_a H_t^{\mathrm{re}}(x,a,y)) - \mathrm{prox}_{\tau\ell}(a' - \tau\partial_a H_t^{\mathrm{re}}(x',a',y'))| \\ &\quad \le \frac{1}{1+\tau\nu}\left(1 - \tau\frac{\mu L_{fa}}{\mu + L_{fa}}\right)|a-a'| + \tau C_{\bar{b}}|y-y'| + \tau(L_{\bar{b}}|y'| + L_{fa})|x-x'|. \end{aligned}$$

Observe that for all $a, b, \tau \ge 0$ with $0 \le \tau b \le 1$, $1 - \tau a \le (1+\tau b)(1 - \tau\frac{a+b}{2})$. Then setting $a = \frac{\mu L_{fa}}{\mu + L_{fa}}$ and $b = \nu$ in the inequality shows that the desired estimate holds with $\tau \in (0, \frac{2}{\mu + L_{fa}} \wedge \frac{1}{\nu}]$. ∎

3.2. Uniform boundedness in time. To establish the boundedness of $\phi_t^m(0)$, we first prove that the adjoint processes $(Y^{t,x,\phi}, Z^{t,x,\phi})$ defined in (1.11) have bounded pth moments.

PROPOSITION 3.5. *Suppose* (H.1) *holds. For each* $\phi \in \mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x) \in [0,T] \times \mathbb{R}^n$, *let* $(Y^{t,x,\phi}, Z^{t,x,\phi}) \in \mathcal{S}^2(t,T;\mathbb{R}^n) \times \mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ *be defined by* (1.11). *Then for all* $p \geq 1$ *there exists* $C_{(p)} \geq 0$, *such that for all* $\phi \in \mathcal{V}_{\boldsymbol{A}}$, $(t,x) \in [0,T] \times \mathbb{R}^n$,

$$\begin{aligned} &\mathbb{E}\bigg[\sup_{s\in[t,T]} e^{p\widetilde{\alpha}s}|Y_s^{t,x,\phi}|^{2p} + \bigg(\int_t^T e^{\widetilde{\alpha}s}|Z_s^{t,x,\phi}|^2\,\mathrm{d}s\bigg)^p\bigg] \\ &\qquad \leq C_{(p)}\bigg(e^{p\widetilde{\alpha}T}C_g^{2p} + C_{fx}^{2p}\bigg(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\,\mathrm{d}s\bigg)^{2p}\bigg), \end{aligned} \tag{3.7}$$

with $\widetilde{\alpha} = 2(\kappa_{\hat{b}} - \rho + L_{\bar{b}} + L_\sigma^2)$. *Consequently, there exists an absolute constant* $C \geq 0$ *such that for all* $\phi \in \mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x) \in [0,T] \times \mathbb{R}^n$,

$$|Y_t^{t,x,\phi}| \leq C_Y := C(C_g + C_{fx}T)e^{(\kappa_{\hat{b}}-\rho+C)_+T}. \tag{3.8}$$

Proof. Let $\bar{f}^1 : [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$ be such that for all $(s,\omega,y,z) \in [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d}$, $\bar{f}_s^1(\omega,y,z) = \partial_x H_s(X_s^{t,x,\phi}(\omega), \phi_s(X_s^{t,x,\phi}(\omega)), y, z)$, where H is defined in (1.4), and $X^{t,x,\phi}$ is defined by (1.10). Then, by Lemma 3.3, $|\bar{f}_t^1(\omega,y,z) - \bar{f}_t^1(\omega,y,z')| \leq L_\sigma|z-z'|$, and

$$\langle y-y', \bar{f}_t^1(\omega,y,z) - \bar{f}_t^1(\omega,y',z)\rangle \leq (\kappa_{\hat{b}} - \rho + L_{\bar{b}})|y-y'|^2. \tag{3.9}$$

By applying Lemma 3.2 with $f^1 = \bar{f}^1$, $\xi^1 = \partial_x g(X_T^{t,x,\phi})$, $f^2 = 0$, $\xi^2 = 0$, $Y^2 = Z^2 = 0$, $\varepsilon = 1/2$, and $\alpha = 2(\kappa_{\hat{b}} - \rho + L_{\bar{b}} + L_\sigma^2)$, it holds with some constant $C \geq 0$ that, for all $p \geq 1$,

$$\begin{aligned} &\mathbb{E}\bigg[\sup_{s\in[t,T]} e^{p\alpha s}|Y_s^{t,x,\phi}|^{2p} + \bigg(\int_t^T e^{\alpha s}|Z_s^{t,x,\phi}|^2\,\mathrm{d}s\bigg)^p\bigg] \\ &\quad \leq C_{(p)}\mathbb{E}\bigg[e^{p\alpha T}|\partial_x g(X_T^{t,x,\phi})|^{2p} + \bigg(\int_t^T e^{\frac{\alpha}{2}s}|\partial_x f_s(X_s^{t,x,\phi}, \partial_x\phi(X_s^{t,x,\phi}))|\,\mathrm{d}s\bigg)^{2p}\bigg] \\ &\quad \leq C_{(p)}\bigg(e^{p\alpha T}C_g^{2p} + C_{fx}^{2p}\bigg(\int_t^T e^{\frac{\alpha}{2}s}\,\mathrm{d}s\bigg)^{2p}\bigg), \end{aligned}$$

where the last inequality follows from (2.1) and (2.5).

Consequently, by setting $p = 1$ in the above estimate and taking the square root of both sides, there exists an absolute constant $C \geq 0$ such that for all $(t,x) \in [0,T] \times \mathbb{R}^n$,

$$|Y_t^{t,x,\phi}| \leq Ce^{-\frac{\alpha}{2}t}\bigg(e^{\frac{\alpha}{2}T}C_g + C_{fx}\int_t^T e^{\frac{\alpha}{2}s}\,\mathrm{d}s\bigg) \leq C(C_g + C_{fx}T)e^{(\frac{\alpha}{2}T)_+}. \tag{3.10}$$

This finishes the proof of the proposition. ∎

Based on Lemma 3.4 and Proposition 3.5, we now establish the uniform boundedness of $\phi_t^m(0)$.

THEOREM 3.6. *Suppose* (H.1) *holds. Let* $a_0 \in \boldsymbol{A}$ *and* $z^{a_0} \in \mathbb{R}^k$ *such that* $z^{a_0} \in \partial^s\ell(a_0)$.[4] *For each* $\phi^0 \in \mathcal{V}_{\boldsymbol{A}}$, $\tau > 0$, *and* $m \in \mathbb{N}$, *let* ϕ^m *be defined by* (1.9). *Then, for all* $\phi^0 \in \mathcal{V}_{\boldsymbol{A}}$ *and* $\tau \in (0, \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}]$,

$$\sup_{m\in\mathbb{N}_0, t\in[0,T]} |\phi_t^m(0)| \leq C_{(\phi^0)}, \tag{3.11}$$

[4]For any $a \in \mathbf{A} = \operatorname{dom}\ell$, the convex subdifferential of ℓ at a is defined as $\partial^s\ell(a) := \{z \in \mathbb{R}^k \mid \ell(a') - \ell(a) \geq \langle z, a'-a\rangle \quad \forall a' \in \mathbb{R}^k\}$. As ℓ is proper, lower semicontinuous, and convex, $\partial^s\ell(\cdot)$ is nonempty on a dense subset of $\mathbf{A}$ by [1, Corollary 2.44].

where

$$C_{(\phi^0)} := \sup_{t\in[0,T]} |\phi_t^0(0)| + 2\Big(\tfrac{1}{\mu+\nu}C_{\bar b}C_Y + \tfrac{2}{\mu+\nu}(C_{fa} + L_{fa}|a_0| + |z^{a_0}|) + |a_0|\Big) + 4C_Y C_{\bar b}\frac{\mu + L_{fa}}{(\mu+\nu)L_{fa} + \mu\nu},$$

and the constant $C_Y \geq 0$ *is defined by* (3.8).

Proof. For each $(t,x,a) \in [0,T]\times\mathbb{R}^n\times\mathbb{R}^k$ and $u\in\mathbb{R}^n$, let $h_t(x,a) = f_t(x,a)+\ell(a)$ and $\phi_t^\star[u] = \arg\min_{a\in\mathbb{R}^k}(H_t^{\mathrm{re}}(0,a,u)+\ell(a))$, with H^{re} defined as in (1.3). By (1.3) and (2.6),

$$\phi_t^\star[u] = \arg\min_{a\in\mathbb{R}^k}\big(\langle \bar b_t(0)a, u\rangle + h_t(0,a)\big) = \partial_z h_t^*(0, -\bar b_t^\top(0)u), \tag{3.12}$$

where $h^*:[0,T]\times\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^k$ is the convex conjugate function of h defined by

$$h_t^*(x,z) := \sup\{\langle a,z\rangle - h_t(x,a) \mid a\in\mathbb{R}^k\}. \tag{3.13}$$

Note that by (H.1), for each $(t,x)\in[0,T]\times\mathbb{R}^n$, the function $a\mapsto h_t(x,a)$ is proper, lower semicontinuous, and $(\mu+\nu)$-strongly convex, which implies that $z\mapsto h_t^*(x,z)$ is finite and differentiable on $\mathbb{R}^k$, $z\mapsto\partial_z h_t^*(x,z)$ is $\frac{1}{\mu+\nu}$-Lipschitz continuous, and $\partial_z h_t^*(x,z) = \arg\max_{a\in\mathbb{R}^k}\big(\langle a,z\rangle - h_t(x,a)\big)$.

Let $a_0\in\operatorname{dom}\ell$ and $z^{a_0}\in\mathbb{R}^k$ such that $z^{a_0}\in\partial^s\ell(a_0)\neq\emptyset$. Then, for all $t\in[0,T]$, by the differentiability and convexity of $a\mapsto f_t(0,a)$, $\partial_a f_t(0,a_0)+z^{a_0}\in\partial^s h_t(0,a_0)$ (see [38, Corollary 10.9]). Hence, by the fact that $\partial_z h_t^*(0,0) = \arg\min_{a\in\mathbb{R}^k} h_t(0,a)$ and the $(\mu+\nu)$-strong convexity of $a\mapsto h_t(0,a)$,

$$\begin{aligned} h_t(0,a_0) &\geq h_t(0,\partial_z h_t^*(0,0)) \\ &\geq h_t(0,a_0) + \langle \partial_a f_t(0,a_0) + z^{a_0}, \partial_z h_t^*(0,0) - a_0\rangle + \tfrac{\mu+\nu}{2}|\partial_z h_t^*(0,0) - a_0|^2, \end{aligned}$$

which implies that

$$|\partial_z h_t^*(0,0) - a_0| \leq \tfrac{2}{\mu+\nu}|\partial_a f_t(0,a_0) + z^{a_0}| \leq \tfrac{2}{\mu+\nu}(C_{fa} + L_{fa}|a_0| + |z^{a_0}|),$$

where the last inequality follows from (2.2). Hence, by using (2.6) and the $\frac{1}{\mu+\nu}$-Lipschitz continuity of $z\mapsto\partial_z h_t^*(0,z)$, for all $t\in[0,T]$ and $u\in\mathbb{R}^n$,

$$\begin{aligned} |\phi_t^\star[u]| &\leq |\partial_z h_t^*(0,-\bar b_t^\top(0)u) - \partial_z h_t^*(0,0)| + |\partial_z h_t^*(0,0)| \\ &\leq \tfrac{1}{\mu+\nu}C_{\bar b}|u| + \tfrac{2}{\mu+\nu}(C_{fa} + L_{fa}|a_0| + |z^{a_0}|) + |a_0|. \end{aligned} \tag{3.14}$$

Let $\phi^0\in\mathcal{V}_{\mathbf{A}}$ and $\tau\in(0, \frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$ be fixed in the subsequent analysis. For each $t\in[0,T]$, let $c_t^0 := \phi_t^\star[Y_t^{t,0,\phi^0}]$. Observe that for all $t\in[0,T]$, $c\in\mathbb{R}^k$, and $u\in\mathbb{R}^n$, by the definition of $\mathrm{prox}_{\tau\ell}$ in (1.8),

$$\begin{aligned} c = \arg\min_{a\in\mathbb{R}^k}(H_t^{\mathrm{re}}(0,a,u)+\ell(a)) &\iff 0\in\partial_a H_t^{\mathrm{re}}(0,c,u) + \partial^s\ell(c) \\ &\iff 0\in\big(c - (c - \tau\partial_a H_t^{\mathrm{re}}(0,c,u))\big) + \partial^s(\tau\ell)(c) \\ &\iff c = \mathrm{prox}_{\tau\ell}(c - \tau\partial_a H_t^{\mathrm{re}}(0,c,u)). \end{aligned}$$

Then, for all $t \in [0,T]$, the fact that $c_t^0 = \arg\min_{a\in\mathbb{R}^k}(H_t^{\mathrm{re}}(0,a,Y_t^{t,0,\phi^0}) + \ell(a))$ implies that $c_t^0 = \mathrm{prox}_{\tau\ell}(c_t^0 - \tau\partial_a H_t^{\mathrm{re}}(0,c_t^0,Y_t^{t,0,\phi^0}))$. Hence for all $m \in \mathbb{N}_0$ and $t \in [0,T]$, (1.9) and Lemma 3.4 imply that

$$
\begin{aligned}
&|\phi_t^{m+1}(0) - c_t^0| \\
&\quad = \left|\mathrm{prox}_{\tau\ell}\big(\phi_t^m(0) - \tau\partial_a H_t^{\mathrm{re}}(0,\phi_t^m(0),Y_t^{t,0,\phi^m})\big) - \mathrm{prox}_{\tau\ell}(c_t^0 - \tau\partial_a H_t^{\mathrm{re}}(0,c_t^0,Y_t^{t,0,\phi^0}))\right| \\
&\quad \le \left(1 - \tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)|\phi_t^m(0) - c_t^0| + \tau C_{\bar{b}}|Y_t^{t,0,\phi^m} - Y_t^{t,0,\phi^0}|.
\end{aligned}
$$

By Proposition 3.5, there exists an absolute constant $C \ge 0$ such that for all $t \in [0,T]$ and $\phi \in \mathcal{V}_{\mathbf{A}}$, $|Y_t^{t,0,\phi}| \le C_Y := C(C_g + C_{fx}T)(e^{\alpha_+ T})$, with $\alpha = \kappa_{\hat{b}} - \rho + L_{\bar{b}} + L_\sigma^2$, which implies that for all $m \in \mathbb{N}_0$,

$$
\begin{aligned}
|\phi_t^m(0) - c_t^0| &\le |\phi_t^0(0) - c_t^0| + 2C_{\bar{b}}\frac{\mu + L_{fa}}{(\mu+\nu)L_{fa} + \mu\nu}\sup_{m\in\mathbb{N}_0}|Y_t^{t,0,\phi^m} - Y_t^{t,0,\phi^0}| \\
&\le |\phi_t^0(0)| + |c_t^0| + 4C_Y C_{\bar{b}}\frac{\mu + L_{fa}}{(\mu+\nu)L_{fa} + \mu\nu}.
\end{aligned}
$$

By (3.14), for all $t \in [0,T]$, $|c_t^0| \le \frac{1}{\mu+\nu}C_{\bar{b}}C_Y + \frac{2}{\mu+\nu}(C_{fa} + L_{fa}|a_0| + |z^{a_0}|) + |a_0|$, from which for all $m \in \mathbb{N}_0$ and $t \in [0,T]$,

$$
\begin{aligned}
|\phi_t^m(0)| &\le |\phi_t^0(0)| + 2|c_t^0| + 4C_Y C_{\bar{b}}\frac{\mu + L_{fa}}{(\mu+\nu)L_{fa} + \mu\nu} \\
&\le \sup_{t\in[0,T]}|\phi_t^0(0)| + 2\Big(\tfrac{1}{\mu+\nu}C_{\bar{b}}C_Y + \tfrac{2}{\mu+\nu}(C_{fa} + L_{fa}|a_0| + |z^{a_0}|) + |a_0|\Big) \\
&\quad + 4C_Y C_{\bar{b}}\tfrac{\mu+L_{fa}}{(\mu+\nu)L_{fa}+\mu\nu}.
\end{aligned}
$$

This finishes the proof of the uniform boundedness of $\phi_t^m(0)$. ∎

3.3. Uniform Lipschitz continuity in space. This section proves that the iterates $(\phi^m)_{m\in\mathbb{N}_0}$ satisfy $\sup_{m\in\mathbb{N}_0}[\phi^m]_1 < \infty$ if one of the conditions (i)–(vi) holds. The following proposition estimates the Lipschitz continuity of the function $x \mapsto Y_t^{t,x,\phi}$ in terms of the Lipschitz continuity of a given feedback control $\phi \in \mathcal{V}_{\mathbf{A}}$.

PROPOSITION 3.7. *Suppose* (H.1) *holds. For each* $\phi \in \mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x) \in [0,T]\times\mathbb{R}^n$, *let* $(Y^{t,x,\phi}, Z^{t,x,\phi}) \in \mathcal{S}^2(t,T;\mathbb{R}^n) \times \mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ *be defined by* (1.11). *Then there exists a constant* $C \ge 0$ *such that for all* $\phi \in \mathcal{V}_{\boldsymbol{A}}$, $t \in [0,T], x, x' \in \mathbb{R}^n$,

$$
|Y_t^{t,x,\phi} - Y_t^{t,x',\phi}| \le L_Y([\phi]_1)|x - x'|, \tag{3.15}
$$

where for each $M \ge 0$, *the constant* $L_Y(M) \ge 0$ *is defined by*

$$
\begin{aligned}
L_Y(M) := C\Bigg[&L_g e^{\left(2L_{\bar{b}}M + 2\kappa_{\hat{b}} + C\right)_+ T} e^{\alpha_+ T} + \Bigg(\Big((1 + L_{\bar{b}}M)(C_g + C_{fx}T)e^{(\kappa_{\hat{b}} - \rho + C)_+ T} \\
&+ L_{fx}(1+M)\Big)\tfrac{e^{\alpha T}-1}{\alpha} + \sqrt{T}\Big(e^{\alpha_+ T}C_g + C_{fx}\tfrac{e^{\alpha T}-1}{\alpha}\Big)\Bigg)e^{\left(2L_{\bar{b}}M + 2\kappa_{\hat{b}} + C\right)_+ T}\Bigg], \\
\alpha := \kappa_{\hat{b}} - \rho + L_{\bar{b}} + L_\sigma^2.
\end{aligned} \tag{3.16}
$$

Proof. Let $\bar{f}^1, \bar{f}^2 : [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$ be such that for all $(s,\omega,y,z) \in [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d}$,

$$\bar{f}^1_s(\omega,y,z) = \partial_x H_s(X^{t,x,\phi}_s(\omega), \phi_s(X^{t,x,\phi}_s(\omega)), y, z),$$
$$\bar{f}^2_s(\omega,y,z) = \partial_x H_s(X^{t,x',\phi}_s(\omega), \phi_s(X^{t,x',\phi}_s(\omega)), y, z),$$

where H is defined by (1.4), and for each $\phi \in \mathcal{V}_{\mathbf{A}}$, $X^{t,x,\phi} \in \mathcal{S}^2(t,T;\mathbb{R}^n)$ is defined by (1.10). By using (3.9) and applying Lemma 3.2 with $p=1$, $f^1 = \bar{f}^1$, $\xi^1 = \partial_x g(X^{t,x,\phi}_T)$, $f^2 = \bar{f}^2$, $\xi^2 = \partial_x g(X^{t,x',\phi}_T)$, $(Y^2, Z^2) = (Y^{t,x',\phi}, Z^{t,x',\phi})$, and $\varepsilon = 1/2$, there exists a absolute constant $C \geq 0$ such that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{s\in[t,T]} e^{\widetilde{\alpha} s}|Y^{t,x,\phi}_s - Y^{t,x',\phi}_s|^2 + \int_t^T e^{\widetilde{\alpha} s}|Z^{t,x,\phi}_s - Z^{t,x',\phi}_s|^2 \,\mathrm{d}s\Big] \\
&\quad \leq C\mathbb{E}\Big[e^{\widetilde{\alpha} T}|\partial_x g(X^{t,x,\phi}_T) - \partial_x g(X^{t,x',\phi}_T)|^2 \\
&\qquad + \Big(\int_t^T e^{\frac{\widetilde{\alpha}}{2} s}|\bar{f}^1_s(\cdot, Y^{t,x',\phi}_s, Z^{t,x',\phi}_s) - \bar{f}^2_s(\cdot, Y^{t,x',\phi}_s, Z^{t,x',\phi}_s)|\,\mathrm{d}s\Big)^2\Big],
\end{aligned} \tag{3.17}$$

where we defined $\widetilde{\alpha} := 2(\kappa_{\hat{b}} - \rho + L_{\bar{b}} + L^2_\sigma)$ above and hereafter. Recall that in the subsequent analysis, C denotes a generic constant independent of the constants $T, \rho, \kappa_{\hat{b}}, C_{\bar{b}}, C_{fx}, L_{fx}, \mu, \nu, L_{fa}, C_g, L_g$.

We now estimate the two terms on the right-hand side of (3.17). Let $b^1, b^2 : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ be such that for all $(t,x) \in [0,T] \times \mathbb{R}^n$, $b^1_t(x) = b^2_t(x) = b_t(x, \phi_t(x))$, and let $\Delta X^{t,x,x'} = X^{t,x,\phi} - X^{t,x',\phi}$. Then, by (2.8) and (2.9), for all $t \in [0,T]$ and $x, x' \in \mathbb{R}^n$,

$$\begin{aligned}
\langle x - x', b^1_t(x) - b^1_t(x')\rangle &\leq \langle x - x', \hat{b}_t(x) - \hat{b}_t(x') + \bar{b}_t(x)\phi_t(x) - \bar{b}_t(x')\phi_t(x')\rangle \\
&\leq \kappa_{\hat{b}}|x - x'|^2 + L_{\bar{b}}|x - x'|(|x - x'| + |\phi_t(x) - \phi_t(x')|) \\
&\leq \big(\kappa_{\hat{b}} + L_{\bar{b}}(1 + [\phi]_1)\big)|x - x'|^2.
\end{aligned}$$

By Lemma 3.1 for $p \geq 2$ and (2.10), for all $x, x' \in \mathbb{R}^n$,

$$\|\Delta X^{t,x,x'}\|_{\mathcal{S}^p} \leq C_{(p)} e^{T\left(2(\kappa_{\hat{b}} + L_{\bar{b}}(1+[\phi]_1)) + C_{(p)}L^2_\sigma\right)_+}|x - x'|, \tag{3.18}$$

from which by setting $p = 2$ and using (2.5) we obtain

$$\begin{aligned}
&e^{-\frac{\widetilde{\alpha}}{2}t}\mathbb{E}\Big[e^{\widetilde{\alpha} T}|\partial_x g(X^{t,x,\phi}_T) - \partial_x g(X^{t,x',\phi}_T)|^2\Big]^{1/2} \\
&\quad \leq C L_g e^{(T-t)\frac{\widetilde{\alpha}}{2} + T\left(2(\kappa_{\hat{b}} + L_{\bar{b}}[\phi]_1) + C\right)_+}|x - x'|.
\end{aligned} \tag{3.19}$$

We then proceed to estimate the second term in (3.17). Note that for all $t \in [0,T]$, $x, x' \in \mathbb{R}^n$, $\phi \in \mathcal{V}_{\mathbf{A}}$, $(y,z) \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$, by Lemma 3.3,

$$\begin{aligned}
&|\partial_x H_t(x, \phi_t(x), y, z) - \partial_x H_t(x', \phi_t(x'), y, z)| \\
&\quad \leq \big(L_{\hat{b}}|x - x'| + L_{\bar{b}}(|x - x'| + |\phi_t(x) - \phi_t(x')|)\big)|y| + L_\sigma|x - x'||z| \\
&\qquad + L_{fx}(|x - x'| + |\phi_t(x) - \phi_t(x')|) \\
&\quad \leq \Big(\big(L_{\hat{b}} + L_{\bar{b}}(1 + [\phi]_1)\big)|y| + L_{fx}(1 + [\phi]_1) + L_\sigma|z|\Big)|x - x'|,
\end{aligned}$$

which along with the Cauchy–Schwarz inequality implies that

(3.20)
$$\begin{aligned}
&\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}|\bar{f}_s^1(\cdot,Y_s^{t,x',\phi},Z_s^{t,x',\phi})-\bar{f}_s^2(\cdot,Y_s^{t,x',\phi},Z_s^{t,x',\phi})|\,\mathrm{d}s\bigg)^2\bigg]^{\frac{1}{2}}\\
&\le \mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\Big((C+L_{\bar{b}}[\phi]_1)|Y_s^{t,x',\phi}|+L_{fx}(1+[\phi]_1)+L_\sigma|Z_s^{t,x',\phi}|\Big)|\Delta X_s^{t,x,x'}|\,\mathrm{d}s\bigg)^2\bigg]^{\frac{1}{2}}\\
&\le \mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\Big((C+L_{\bar{b}}[\phi]_1)|Y_s^{t,x',\phi}|+L_{fx}(1+[\phi]_1)+C|Z_s^{t,x',\phi}|\Big)\mathrm{d}s\bigg)^4\bigg]^{\frac{1}{4}}\|\Delta X^{t,x,x'}\|_{\mathcal{S}^4}\\
&\le \bigg(\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\Big((C+L_{\bar{b}}[\phi]_1)|Y_s^{t,x',\phi}|+L_{fx}(1+[\phi]_1)\Big)\mathrm{d}s\bigg)^4\bigg]^{\frac{1}{4}}\\
&\quad +C\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}|Z_s^{t,x',\phi}|\,\mathrm{d}s\bigg)^4\bigg]^{\frac{1}{4}}\bigg)\|\Delta X^{t,x,x'}\|_{\mathcal{S}^4}.
\end{aligned}$$

By Proposition 3.5, there exists an absolute constant $C\ge 0$ such that for all $(t,x')\in[0,T]\times\mathbb{R}^n$,

$$|Y_t^{t,x',\phi}|\le C_Y:=C(C_g+C_{fx}T)e^{(\kappa_{\hat{b}}-\rho+C)_+T}. \tag{3.21}$$

The Markovian property of $Y^{t,x',\phi}$ implies that $Y_s^{t,x',\phi}=Y_s^{s,X_s^{t,x',\phi},\phi}$ (see, e.g., [44, Theorem 5.1.3]), which subsequently shows that

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\Big((C+L_{\bar{b}}[\phi]_1)|Y_s^{t,x',\phi}|+L_{fx}(1+[\phi]_1)\Big)\mathrm{d}s\bigg)^4\bigg]^{1/4}\\
&\quad\le\Big((C+L_{\bar{b}}[\phi]_1)C_Y+L_{fx}(1+[\phi]_1)\Big)\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\,\mathrm{d}s.
\end{aligned}\tag{3.22}$$

On the other hand, by the Cauchy–Schwarz inequality and Proposition 3.5 with $p=2$, there exists an absolute constant $C\ge 0$ such that

$$\begin{aligned}
\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}|Z_s^{t,x',\phi}|\,\mathrm{d}s\bigg)^4\bigg]^{1/4}&\le\sqrt{T-t}\,\mathbb{E}\bigg[\bigg(\int_t^T e^{\tilde{\alpha}s}|Z_s^{t,x',\phi}|^2\,\mathrm{d}s\bigg)^2\bigg]^{1/4}\\
&\le C\sqrt{T-t}\bigg(e^{\frac{\tilde{\alpha}}{2}T}C_g+C_{fx}\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\,\mathrm{d}s\bigg).
\end{aligned}\tag{3.23}$$

Combining (3.18) (with $p=4$), (3.20), (3.22), and (3.23) gives that

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg(\int_t^T e^{\frac{\tilde{\alpha}}{2}s}|\bar{f}_s^1(\cdot,Y_s^{t,x',\phi},Z_s^{t,x',\phi})-\bar{f}_s^2(\cdot,Y_s^{t,x',\phi},Z_s^{t,x',\phi})|\,\mathrm{d}s\bigg)^2\bigg]^{1/2}\\
&\quad\le C\bigg(\Big((1+L_{\bar{b}}[\phi]_1)C_Y+L_{fx}(1+[\phi]_1)\Big)\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\,\mathrm{d}s\\
&\qquad+\sqrt{T}\bigg(e^{\frac{\tilde{\alpha}}{2}T}C_g+C_{fx}\int_t^T e^{\frac{\tilde{\alpha}}{2}s}\,\mathrm{d}s\bigg)\bigg)e^{T\left(2(\kappa_{\hat{b}}+L_{\bar{b}}[\phi]_1)+C\right)_+}|x-x'|,
\end{aligned}$$

with C_Y defined in (3.21). Consequently, by using (3.17) and (3.19) and the identity that $e^{-\frac{\widetilde{\alpha}}{2}t}\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\,\mathrm{d}s = \frac{2}{\widetilde{\alpha}}(e^{\frac{\widetilde{\alpha}}{2}(T-t)}-1)$,

$$
\begin{aligned}
&|Y_t^{t,x,\phi} - Y_t^{t,x',\phi}| \\
&\quad \le C\Bigg[L_g e^{(T-t)\frac{\widetilde{\alpha}}{2}+T\left(2(\kappa_{\hat{b}}+L_{\bar{b}}[\phi]_1)+C\right)_+} \\
&\qquad + \Bigg(\Big((1+L_{\bar{b}}[\phi]_1)(C_g+C_{fx}T)e^{(\kappa_{\hat{b}}-\rho+C)_+T} + L_{fx}(1+[\phi]_1)\Big)\frac{2}{\widetilde{\alpha}}(e^{\frac{\widetilde{\alpha}}{2}(T-t)}-1) \\
&\qquad + \sqrt{T}\Big(e^{\frac{\widetilde{\alpha}}{2}(T-t)}C_g + C_{fx}\frac{2}{\widetilde{\alpha}}(e^{\frac{\widetilde{\alpha}}{2}(T-t)}-1)\Big)\Bigg)e^{T\left(2(\kappa_{\hat{b}}+L_{\bar{b}}[\phi]_1)+C\right)_+}\Bigg]|x-x'|.
\end{aligned}
$$

Then the fact that for all $\widetilde{\alpha}\in\mathbb{R}$, $[0,T]\ni t\mapsto \frac{2}{\widetilde{\alpha}}(e^{\frac{\widetilde{\alpha}}{2}(T-t)}-1)\in(0,\infty)$ is maximized at $t=0$ and that $\widetilde{\alpha}=2(\kappa_{\hat{b}}-\rho+L_{\bar{b}}+L_\sigma^2)$ lead to the desired Lipschitz estimate uniformly in t. ∎

With Proposition 3.7 in hand, we prove that under suitable assumptions, for any initial guess $\phi^0\in\mathcal{V}_{\mathbf{A}}$, the sequence of feedback controls $(\phi^m)_{m\in\mathbb{N}_0}$ generated by (1.9) is uniformly Lipschitz continuous. For notational simplicity, let $C>0$ be a constant such that (3.8) and (3.15) hold, let C_Y and α be defined in (3.8) and (3.16), respectively, and for each $\phi^0\in\mathcal{V}_{\mathbf{A}}$, define

(3.24)
$$
\begin{aligned}
A_1 &:= C_{\bar{b}}C\Big((L_g+\sqrt{T}C_g)e^{\alpha_+T} + \tfrac{e^{\alpha T}-1}{\alpha}\Big((C_g+C_{fx}T)e^{\alpha_+T}+L_{fx}+\sqrt{T}C_{fx}\Big)\Big), \\
A_2 &:= C_{\bar{b}}C\tfrac{e^{\alpha T}-1}{\alpha}\Big((C_g+C_{fx}T)e^{\alpha_+T}L_{\bar{b}}+L_{fx}\Big), \\
\mu_0 &:= \frac{1}{4}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right), \\
K &:= \max\left\{2TL_{\bar{b}}[\phi^0]_1, \frac{2TL_{\bar{b}}A_1+2TL_{\bar{b}}(L_{\bar{b}}C_Y+L_{fa})}{\mu_0}+1\right\}.
\end{aligned}
$$

THEOREM 3.8. *Suppose* (H.1) *holds. For each $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$, $\tau>0$, and $m\in\mathbb{N}$, let ϕ^m be defined by* (1.9) *with the initial guess ϕ^0 and stepsize τ. Let $C\ge 0$ be a constant such that* (3.8) *and* (3.15) *hold, let $C_Y\ge 0$ be defined in* (3.8)*, let $\alpha\in\mathbb{R}$ be defined in* (3.16)*, and let $A_1,A_2,\mu_0,K\ge 0$ be defined in* (3.24)*. Then for all $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$ satisfying*

$$
2TL_{\bar{b}}A_1e^{(2\kappa_{\hat{b}}+C)T+K}\le\mu_0 \quad \textit{and} \quad A_2(e^{(2\kappa_{\hat{b}}+C)T+K}+1)\le\mu_0 \tag{3.25}
$$

and for all $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$ and $m\in\mathbb{N}_0$,

$$
[\phi^m]_1\le L_{(\phi^0)} := [\phi^0]_1 + \frac{1}{\mu_0}\Big(L_{\bar{b}}C_Y+L_{fa}+A_1\big(e^{(2\kappa_{\hat{b}}+C)T+K}+1\big)\Big). \tag{3.26}
$$

Proof. Throughout this proof, let $\phi^0\in\mathcal{V}_{\mathbf{A}}$ and $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$ be fixed. Suppose that $\phi^m\in\mathcal{V}_{\mathbf{A}}$ for some $m\in\mathbb{N}_0$. For all $(t,x,x')\in[0,T]\times\mathbb{R}^n\times\mathbb{R}^n$, by applying Lemma 3.4 with $a=\phi_t^m(x)$, $a'=\phi_t^m(x')$, $y=Y_t^{t,x,\phi^m}$, and $y'=Y_t^{t,x',\phi^m}$,

$$
\begin{aligned}
&|\phi_t^{m+1}(x)-\phi_t^{m+1}(x')| \\
&\quad\le\left(1-\tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right)\right)|\phi_t^m(x)-\phi_t^m(x')| + \tau C_{\bar{b}}|Y_t^{t,x,\phi^m}-Y_t^{t,x',\phi^m}| \\
&\qquad + \tau(L_{\bar{b}}|Y_t^{t,x',\phi^m}|+L_{fa})|x-x'|.
\end{aligned}
$$

By using the Lipschitz continuity of ϕ^m and the Lipschitz continuity and uniform boundedness of the mapping $(t,x)\mapsto Y_t^{t,x,\phi^m}$ (see Propositions 3.5 and 3.7), we further deduce that

$$\begin{aligned}&|\phi_t^{m+1}(x)-\phi_t^{m+1}(x')|\\&\quad\le\left([\phi^m]_1\left(1-\tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right)\right)+\tau C_{\bar b}L_Y([\phi^m]_1)+\tau(L_{\bar b}C_Y+L_{fa})\right)|x-x'|,\end{aligned}$$

where C_Y is defined by (3.8) and $L_Y([\phi^m]_1)$ is defined by (3.16). Consequently, we have

$$[\phi^{m+1}]_1\le[\phi^m]_1\left(1-\tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right)\right)+\tau(L_{\bar b}C_Y+L_{fa})+\tau C_{\bar b}L_Y([\phi^m]_1). \tag{3.27}$$

In what follows, we aim to establish a uniform bound of $([\phi^m]_1)_{m\in\mathbb{N}_0}$ based on (3.27). Observe from the definition of $L_Y([\phi^m]_1)$ in (3.16) that

$$\begin{aligned}&L_Y([\phi^m]_1)\\&:=C\Bigg[L_g e^{\left(2L_{\bar b}[\phi^m]_1+2\kappa_{\hat b}+C\right)_+T}e^{\alpha_+T}+\Bigg(\Big((1+L_{\bar b}[\phi^m]_1)(C_g+C_{fx}T)e^{(\kappa_{\hat b}-\rho+C)_+T}\\&\quad+L_{fx}(1+[\phi^m]_1)\Big)\tfrac{e^{\alpha T}-1}{\alpha}+\sqrt{T}\Big(e^{\alpha_+T}C_g+C_{fx}\tfrac{e^{\alpha T}-1}{\alpha}\Big)\Bigg)e^{\left(2L_{\bar b}[\phi^m]_1+2\kappa_{\hat b}+C\right)_+T}\Bigg]\\&=C\Bigg[\Bigg(\big(L_g+\sqrt{T}C_g\big)e^{\alpha_+T}+\Big((C_g+C_{fx}T)e^{(\kappa_{\hat b}-\rho+C)_+T}+L_{fx}+\sqrt{T}C_{fx}\Big)\tfrac{e^{\alpha T}-1}{\alpha}\Bigg)\\&\quad\times\Big(e^{(2L_{\bar b}[\phi^m]_1+2\kappa_{\hat b}+C)T}+1\Big)\\&\quad+\Big((C_g+C_{fx}T)e^{(\kappa_{\hat b}-\rho+C)_+T}L_{\bar b}+L_{fx}\Big)\tfrac{e^{\alpha T}-1}{\alpha}[\phi]_1\Big(e^{(2L_{\bar b}[\phi^m]_1+2\kappa_{\hat b}+C)T}+1\Big)\Bigg].\end{aligned} \tag{3.28}$$

Let A_1,A_2,μ_0,K be defined as in (3.24). Then, by writing $\widetilde{A}_1:=2TL_{\bar b}A_1$ and $[\widetilde{\phi^m}]_1:=2TL_{\bar b}[\phi^m]_1$ for all $m\in\mathbb{N}_0$, multiplying both sides of (3.27) by $2TL_{\bar b}$, and using (3.28), we have

$$\begin{aligned}[\widetilde{\phi^{m+1}}]_1&\le[\widetilde{\phi^m}]_1(1-2\mu_0\tau)+2\tau TL_{\bar b}(L_{\bar b}C_Y+L_{fa})\\&\quad+\tau\Big(\widetilde{A}_1\big(e^{(2\kappa_{\hat b}+C)T+[\widetilde{\phi^m}]_1}+1\big)+A_2[\widetilde{\phi^m}]_1\big(e^{(2\kappa_{\hat b}+C)T+[\widetilde{\phi^m}]_1}+1\big)\Big).\end{aligned} \tag{3.29}$$

Now we prove by induction that $\sup_{m\in\mathbb{N}_0}[\widetilde{\phi^m}]_1\le K$ under the conditions that

$$\widetilde{A}_1e^{(2\kappa_{\hat b}+C)T+K}\le\mu_0\quad\text{and}\quad A_2(e^{(2\kappa_{\hat b}+C)T+K}+1)\le\mu_0. \tag{3.30}$$

The statement holds for $m=0$ by the definition of K. Suppose that $[\widetilde{\phi^m}]_1\le K$ for some $m\in\mathbb{N}_0$. Then, by the induction hypothesis, (3.29), and (3.30),

$$\begin{aligned}[\widetilde{\phi^{m+1}}]_1&\le[\widetilde{\phi^m}]_1(1-2\mu_0\tau)+2\tau TL_{\bar b}(L_{\bar b}C_Y+L_{fa})+\tau\Big(\widetilde{A}_1+\mu_0+\mu_0[\widetilde{\phi^m}]_1\Big)\\&=[\widetilde{\phi^m}]_1(1-\mu_0\tau)+\tau\Big(2TL_{\bar b}(L_{\bar b}C_Y+L_{fa})+\widetilde{A}_1+\mu_0\Big)\\&\le K(1-\mu_0\tau)+\tau\mu_0K\le K.\end{aligned} \tag{3.31}$$

This finishes the proof of the fact that $\sup_{m\in\mathbb{N}_0}[\widetilde{\phi^m}]_1 \le K$. Substituting this a priori bound into (3.29) and using (3.30) give that

$$\begin{aligned}
&[\widetilde{\phi^{m+1}}]_1 \\
&\quad\le [\widetilde{\phi^m}]_1\,(1-2\mu_0\tau) + 2\tau T L_{\bar b}(L_{\bar b}C_Y + L_{fa}) \\
&\qquad + \tau\Big(\widetilde{A}_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big) + A_2[\widetilde{\phi^m}]_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big)\Big) \\
&\quad\le [\widetilde{\phi^m}]_1\,(1-2\mu_0\tau) + 2\tau T L_{\bar b}(L_{\bar b}C_Y + L_{fa}) + \tau\Big(\widetilde{A}_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big) + \mu_0[\widetilde{\phi^m}]_1\Big) \\
&\quad\le [\widetilde{\phi^m}]_1\,(1-\mu_0\tau) + \tau\Big(2TL_{\bar b}(L_{\bar b}C_Y + L_{fa}) + \widetilde{A}_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big)\Big),
\end{aligned}$$

from which one can deduce that for all $m\in\mathbb{N}_0$,

$$[\widetilde{\phi^m}]_1 \le [\widetilde{\phi^0}]_1 + \frac{1}{\mu_0}\Big(2TL_{\bar b}(L_{\bar b}C_Y + L_{fa}) + \widetilde{A}_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big)\Big).$$

Dividing both sides of the above inequality by $2TL_{\bar b}$ shows that

$$[\phi^m]_1 \le [\phi^0]_1 + \frac{1}{\mu_0}\Big(L_{\bar b}C_Y + L_{fa} + A_1\big(e^{(2\kappa_{\hat b}+C)T+K}+1\big)\Big)$$

with constants A_1, K, μ_0 defined as in (3.24). ∎

3.4. Contraction in a weighted sup-norm. Based on the uniform Lipschitz continuity of $(\phi^m)_{m\in\mathbb{N}_0}$ in Theorem 3.8, we prove the contractivity of the iterates $(\phi^m)_{m\in\mathbb{N}_0}$ with respect to the weighted sup-norm norm $|\cdot|_0$ (see Definition 2.1).

The following proposition estimates the Lipschitz stability of the adjoint process $Y^{t,x,\phi}$ with respect to the feedback control $\phi\in\mathcal{V}_{\mathbf{A}}$.

PROPOSITION 3.9. *Suppose* (H.1) *holds. For each* $\phi\in\mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x)\in[0,T]\times\mathbb{R}^n$, *let* $(Y^{t,x,\phi}, Z^{t,x,\phi})\in\mathcal{S}^2(t,T;\mathbb{R}^n)\times\mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ *be defined by* (1.11). *Suppose that for all* $\phi'\in\mathcal{V}_{\boldsymbol{A}}$ *and* $(t,s,x)\in[0,T]\times[t,T]\times\mathbb{R}^n$, *it holds with some constant* $C_Z^{\phi'}\ge 0$ *that* $|Z_s^{t,x,\phi'}|\le C_Z^{\phi'}$ *for* $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$*-a.e. Then there exists a constant* $C\ge 0$ *such that for all* $\phi,\phi'\in\mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x)\in[0,T]\times\mathbb{R}^n$,

$$|Y_t^{t,x,\phi} - Y_t^{t,x,\phi'}| \le B[\phi,\phi',C_Z^{\phi'}](1+|x|)|\phi-\phi'|_0, \tag{3.32}$$

where the constant $B[\phi,\phi',C_Z^{\phi'}]$ *is defined by*

$$\begin{aligned}
&B[\phi,\phi',C_Z^{\phi'}] \\
&\qquad := CC_{\bar b}\Big(1+T+TC_{\bar b}\sup_{t\in[0,T]}|\phi_t'(0)|\Big)e^{T\beta_+} \\
&\qquad\quad\times\Big(L_g\mathfrak{m}_{(\alpha,\beta)}^{1/2} + \tfrac{e^{T\alpha}-1}{\alpha}\Big[\Big((C_Y+L_{fx})(1+[\phi]_1)+C_Z^{\phi'}\Big)Te^{T\beta_+} + L_{fx} + C_Y\Big]\Big), \\
&\qquad\beta := 2\kappa_{\hat b} + 2L_{\bar b}\max\{[\phi]_1,[\phi']_1\} + C, \\
&\mathfrak{m}_{(\alpha,\beta)} := \sup_{t\in[0,T]} e^{2\alpha(T-t)}\int_t^T e^{(T-s)\beta}\,\mathrm{d}s,
\end{aligned} \tag{3.33}$$

with C_Y *and* α *defined as in* (3.8) *and* (3.16), *respectively.*

Proof. Let $\bar{f}^1, \bar{f}^2 : [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$ be such that for all $(s,\omega,y,z) \in [t,T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d}$,

$$\bar{f}^1_s(\omega,y,z) = \partial_x H_s(X^{t,x,\phi}_s(\omega), \phi_s(X^{t,x,\phi}_s(\omega)), y, z),$$
$$\bar{f}^2_s(\omega,y,z) = \partial_x H_s(X^{t,x,\phi'}_s(\omega), \phi'_s(X^{t,x,\phi'}_s(\omega)), y, z),$$

where H is defined in (1.4), and for each $\phi \in \mathcal{V}_{\mathbf{A}}$ and $(t,x) \in [0,T] \times \mathbb{R}^n$, $X^{t,x,\phi} \in \mathcal{S}^2(t,T;\mathbb{R}^n)$ is defined by (1.10). By using (3.9) and applying Lemma 3.2 with $p=1$, $f^1 = \bar{f}^1$, $\xi^1 = \partial_x g(X^{t,x,\phi}_T)$, $f^2 = \bar{f}^2$, $\xi^2 = \partial_x g(X^{t,x,\phi'}_T)$, $(Y^2, Z^2) = (Y^{t,x,\phi'}, Z^{t,x,\phi'})$, and $\varepsilon = 1/2$, it holds with an absolute constant $C \geq 0$ that

$$\begin{aligned}
&\mathbb{E}\bigg[\sup_{s\in[t,T]} e^{\widetilde{\alpha}s}|Y^{t,x,\phi}_s - Y^{t,x,\phi'}_s|^2 + \int_t^T e^{\widetilde{\alpha}s}|Z^{t,x,\phi}_s - Z^{t,x,\phi'}_s|^2\,\mathrm{d}s\bigg] \\
&\quad\leq C\mathbb{E}\bigg[e^{\widetilde{\alpha}T}|\partial_x g(X^{t,x,\phi}_T) - \partial_x g(X^{t,x,\phi'}_T)|^2 \\
&\qquad + \bigg(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}|\bar{f}^1_s(\cdot, Y^{t,x,\phi'}_s, Z^{t,x,\phi'}_s) - \bar{f}^2_s(\cdot, Y^{t,x,\phi'}_s, Z^{t,x,\phi'}_s)|\,\mathrm{d}s\bigg)^2\bigg],
\end{aligned} \tag{3.34}$$

where we defined $\widetilde{\alpha} := 2(\kappa_{\hat{b}} - \rho + L_{\bar{b}} + L^2_\sigma)$ above and hereafter. In the subsequent analysis, we denote by $C \geq 1$ a generic constant independent of the constants $T, \rho, \kappa_{\hat{b}}, C_{\bar{b}}, C_{fx}, L_{fx}, \mu, \nu, L_{fa}, C_g, L_g$.

To estimate the right-hand side of (3.34), we first quantify the dependence of $X^{t,x,\phi}$ on ϕ. For all $(t,x) \in [0,T] \times \mathbb{R}^n$, let $\Delta X^{t,x} = X^{t,x,\phi} - X^{t,x,\phi'}$. Similar to (3.18), by Lemma 3.1 with $p=2$, $b^1_t(x) = b_t(x, \phi_t(x))$, and $b^2_t(x) = b_t(x, \phi'_t(x))$, for all $(t,x) \in [0,T] \times \mathbb{R}^n$, $\phi, \phi' \in \mathcal{V}_{\mathbf{A}}$,

$$\begin{aligned}
\|\Delta X^{t,x}\|_{\mathcal{S}^2} &\leq C_{\bar{b}}\sqrt{T} M_{([\phi]_1)} \|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{H}^2} \\
&\leq C_{\bar{b}} T M_{([\phi]_1)} \|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2},
\end{aligned} \tag{3.35}$$

where the constant $M_{([\phi]_1)}$ is defined by

$$M_{([\phi]_1)} := Ce^{T\left(2(\kappa_{\hat{b}} + L_{\bar{b}}(1+[\phi]_1)) + CL^2_\sigma\right)_+} \leq Ce^{2T\left(\kappa_{\hat{b}} + L_{\bar{b}}[\phi]_1 + C\right)_+}. \tag{3.36}$$

Moreover, by using the fact that $|\phi(x) - \phi'(x)| \leq |\phi - \phi'|_0(1+|x|)$,

$$\|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2} \leq (1 + \|X^{t,x,\phi'}\|_{\mathcal{S}^2})|\phi - \phi'|_0. \tag{3.37}$$

We estimate $\|X^{t,x,\phi'}\|_{\mathcal{S}^2}$ by setting $\|\phi'(0)\|_\infty = \sup_{t\in[0,T]}|\phi'_t(0)|$ and applying Lemma 3.1 with $p=2$, $x_1 = x$, $b^1_t(x) = b_t(x, \phi'_t(x))$, $\sigma^1_t(x) = \sigma_t(x)$, $x_2 = 0$, $b^2_t(x) = 0$, and $\sigma^2_t(x) = 0$:

$$\begin{aligned}
\|X^{t,x,\phi'}\|_{\mathcal{S}^2} &\leq M_{([\phi']_1)}\big(|x| + \sqrt{T}\|b(0,\phi'(0))\|_{\mathcal{H}^2} + \|\sigma(0)\|_{\mathcal{H}^2}\big) \\
&\leq CM_{([\phi']_1)}\big(|x| + T + TC_{\bar{b}}\|\phi'(0)\|_\infty + \sqrt{T}\big) \\
&\leq CM_{([\phi']_1)}\big(1 + T + TC_{\bar{b}}\|\phi'(0)\|_\infty\big)(1+|x|),
\end{aligned} \tag{3.38}$$

which along with (3.35), (3.37), and $M_{([\phi']_1)} \geq 1$ shows that

$$\begin{aligned}
&\|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2} \leq CM_{([\phi']_1)}\big(1 + T + TC_{\bar{b}}(1+|x|)\|\phi'(0)\|_\infty\big)|\phi - \phi'|_0, \\
&\|\Delta X^{t,x}\|_{\mathcal{S}^2} \leq CC_{\bar{b}} T M_{([\phi]_1)} M_{([\phi']_1)}\big(1 + T + TC_{\bar{b}}\|\phi'(0)\|_\infty\big)(1+|x|)|\phi - \phi'|_0.
\end{aligned} \tag{3.39}$$

Similarly, by setting $\widetilde{\beta} = 2(\kappa_{\hat{b}} + L_{\bar{b}}(1+[\phi]_1)) + L_\sigma^2 + 1$ and using (3.2) with $b_t^1(x) = b_t(x,\phi_t(x))$ and $b_t^2(x) = b_t(x,\phi_t'(x))$, we have

$$\begin{aligned}
&\mathbb{E}\left[|\Delta X_T^{t,x}|^2\right]^{\frac{1}{2}} \\
&\quad\le CC_{\bar{b}}\left(\int_t^T e^{(T-s)\widetilde{\beta}}\,\mathrm{d}s\right)^{\frac{1}{2}} \|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2} \\
&\quad\le CC_{\bar{b}}\left(\int_t^T e^{(T-s)\beta}\,\mathrm{d}s\right)^{\frac{1}{2}} M_{([\phi']_1)}\big(1+T+TC_{\bar{b}}\|\phi'(0)\|_\infty\big)(1+|x|)|\phi-\phi'|_0,
\end{aligned} \tag{3.40}$$

with $\beta := 2(\kappa_{\hat{b}} + L_{\bar{b}}[\phi]_1) + C$, where the last inequality used (3.39).

Now we are ready to estimate the right-hand side of (3.34). By using (3.40),

$$\begin{aligned}
e^{-\frac{\widetilde{\alpha}}{2}t}\mathbb{E}\Big[e^{\widetilde{\alpha}T}|\partial_x g(X_T^{t,x,\phi}) - \partial_x g(X_T^{t,x,\phi'})|^2\Big]^{\frac{1}{2}} &\le L_g e^{\frac{\widetilde{\alpha}}{2}(T-t)}\mathbb{E}\left[|\Delta X_T^{t,x}|^2\right]^{\frac{1}{2}} \\
&\le CC_{\bar{b}}L_g\mathfrak{m}_{(\alpha,\beta)}^{1/2} M_{([\phi']_1)}\big(1+T+TC_{\bar{b}}\|\phi'(0)\|_\infty\big)(1+|x|)|\phi-\phi'|_0,
\end{aligned} \tag{3.41}$$

where we recall that $\mathfrak{m}_{(\alpha,\beta)} = \sup_{t\in[0,T]} e^{\widetilde{\alpha}(T-t)}\int_t^T e^{(T-s)\beta}\,\mathrm{d}s$. On the other hand, by Lemma 3.3, for all $t\in[0,T]$, $x,x'\in\mathbb{R}^n$, $\phi,\phi'\in\mathcal{V}_{\mathbf{A}}$, $(y,z)\in\mathbb{R}^n\times\mathbb{R}^{n\times d}$,

$$\begin{aligned}
&|\partial_x H_t(x,\phi_t(x),y,z) - \partial_x H_t(x',\phi_t'(x'),y,z)| \\
&\quad\le \big(L_{\hat{b}}|x-x'| + L_{\bar{b}}(|x-x'| + |\phi_t(x)-\phi_t'(x')|)\big)|y| + L_\sigma|x-x'||z| \\
&\qquad + L_{fx}(|x-x'| + |\phi_t(x)-\phi_t'(x')|) \\
&\quad\le \Big(\big(L_{\hat{b}} + L_{\bar{b}}(1+[\phi]_1)\big)|y| + L_\sigma|z| + L_{fx}(1+[\phi]_1)\Big)|x-x'| \\
&\qquad + (L_{fx} + L_{\bar{b}}|y|)|\phi_t(x') - \phi_t'(x')|.
\end{aligned}$$

This along with (3.8) and the assumption that $|Z_s^{t,x,\phi}| \le C_Z^\phi$ implies that

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}|\bar{f}_s^1(\cdot,Y_s^{t,x,\phi'},Z_s^{t,x,\phi'}) - \bar{f}_s^2(\cdot,Y_s^{t,x,\phi'},Z_s^{t,x,\phi'})|\,\mathrm{d}s\right)^2\right]^{\frac{1}{2}} \\
&\quad\le \mathbb{E}\left[\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\Big((C+L_{\bar{b}}[\phi]_1)|Y_s^{t,x,\phi'}| + L_{fx}(1+[\phi]_1) + C|Z_s^{t,x,\phi'}|\Big)|\Delta X_s^{t,x}|\,\mathrm{d}s\right)^2\right]^{\frac{1}{2}} \\
&\qquad + \mathbb{E}\left[\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\Big((L_{fx} + L_{\bar{b}}|Y_s^{t,x,\phi'}|)|\phi_s(X_s^{t,x,\phi'}) - \phi_s'(X_s^{t,x,\phi'})|\Big)\,\mathrm{d}s\right)^2\right]^{\frac{1}{2}} \\
&\quad\le C\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\,\mathrm{d}s\right)\Big[\Big((1+[\phi]_1)C_Y + L_{fx}(1+[\phi]_1) + C_Z^{\phi'}\Big)\|\Delta X^{t,x}\|_{\mathcal{S}^2} \\
&\qquad + (L_{fx}+C_Y)\|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2}\Big].
\end{aligned}$$

Substituting (3.39) into the above estimate yields

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}|\bar{f}_s^1(\cdot,Y_s^{t,x,\phi'},Z_s^{t,x,\phi'}) - \bar{f}_s^2(\cdot,Y_s^{t,x,\phi'},Z_s^{t,x,\phi'})|\,\mathrm{d}s\right)^2\right]^{\frac{1}{2}} \\
&\quad\le CC_{\bar{b}}\left(\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\,\mathrm{d}s\right)\Big[\Big((C_Y+L_{fx})(1+[\phi]_1) + C_Z^{\phi'}\Big)TM_{([\phi]_1)}M_{([\phi']_1)}\big(1+T \\
&\qquad + TC_{\bar{b}}\|\phi'(0)\|_\infty\big) + (L_{fx}+C_Y)M_{([\phi']_1)}(1+T+TC_{\bar{b}}\|\phi'(0)\|_\infty)\Big](1+|x|)|\phi-\phi'|_0.
\end{aligned}$$

Combining the above estimate with (3.34) and (3.41), and using $e^{-\frac{\widetilde{\alpha}}{2}t}\int_t^T e^{\frac{\widetilde{\alpha}}{2}s}\,\mathrm{d}s \le \frac{2}{\widetilde{\alpha}}(e^{\frac{\widetilde{\alpha}}{2}T}-1)$ with $\widetilde{\alpha}=2(\kappa_{\hat{b}}-\rho+L_{\bar{b}}+L_\sigma^2)$, we conclude the desired estimate

$$
\begin{aligned}
&|Y_t^{t,x,\phi}-Y_t^{t,x,\phi'}| \le CC_{\bar{b}}(1+|x|)|\phi-\phi'|_0\big(1+T+TC_{\bar{b}}\|\phi'(0)\|_\infty\big)M_{([\phi']_1)}\\
&\quad\times\left(L_g\mathfrak{m}_{(\alpha,\beta)}^{1/2}+\frac{e^{\alpha T}-1}{\alpha}\left[\Big((C_Y+L_{fx})(1+[\phi]_1)+C_Z^{\phi'}\Big)TM_{([\phi]_1)}+(L_{fx}+C_Y)\right]\right),
\end{aligned}
$$

with $\alpha=\kappa_{\hat{b}}-\rho+L_{\bar{b}}+L_\sigma^2$, $\beta=2(\kappa_{\hat{b}}+L_{\bar{b}}[\phi]_1)+C$, and $M_{([\phi]_1)}$ defined in (3.36). ∎

The next lemma establishes an upper bound of the adjoint process $Z^{t,x,\phi}$ in terms of the Lipschitz constant of $x\mapsto Y^{t,x,\phi}$. The proof is given in Appendix A of the arXiv version [37] and extends the arguments of [28, Proposition 3.7] to the present setting, where (1.10) has non-Lipschitz drift coefficients and multiplicative noises, and (1.11) has unbounded coefficients in front of Y.

LEMMA 3.10. *Suppose* (H.1) *holds. For each* $\phi\in\mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x)\in[0,T]\times\mathbb{R}^n$, *let* $(Y^{t,x,\phi},Z^{t,x,\phi})\in\mathcal{S}^2(t,T;\mathbb{R}^n)\times\mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ *be defined by* (1.11). *Then, for all* $\phi\in\mathcal{V}_{\boldsymbol{A}}$ *and* $(t,x)\in[0,T]\times\mathbb{R}^n$,

$$
|Z_s^{t,x,\phi}|\le C_Z^\phi:=C_\sigma L_Y([\phi]_1)\quad \textit{for a.e. } (s,\omega)\in[t,T]\times\Omega, \tag{3.42}
$$

where the constant $L_Y([\phi]_1)\ge 0$ *is defined by* (3.16).

Armed with Theorem 3.6, Theorem 3.8, and Proposition 3.9, we prove that under suitable assumptions, for any initial guess $\phi^0\in\mathcal{V}_{\mathbf{A}}$, the sequence of feedback controls $(\phi^m)_{m\in\mathbb{N}_0}$ generated by (1.9) is a contraction with respect to the norm $|\cdot|_0$.

THEOREM 3.11. *Suppose* (H.1) *holds. For each* $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$, $\tau>0$, *and* $m\in\mathbb{N}$, *let* ϕ^m *be defined by* (1.9) *with the initial guess* ϕ^0 *and stepsize* τ. *Let* $C\ge 0$ *be a constant such that* (3.8), (3.15), *and* (3.32) *hold, let* $C_Y\ge 0$ *be defined in* (3.8), *and let* $\alpha\in\mathbb{R}$ *be defined in* (3.16). *For each* $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$ *and* $M\ge 0$, *let* $C_{(\phi^0)}\ge 0$ *be defined in* (3.11), *let* $L_{(\phi^0)}\ge 0$ *be defined in* (3.26), *and let* $L_Y(M)\ge 0$ *be defined in* (3.16). *Then, for all* $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$, *if we assume further that* (3.25) *holds and*

$$
C(1+T+TC_{\bar{b}}C_{(\phi^0)})e^{T\beta_+}(Te^{T\beta_+}+1)B_{(\phi^0)}<\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right), \tag{3.43}
$$

with the constants $\beta\in\mathbb{R}$, $\mathfrak{m}_{(\alpha,\beta)}>0$, *and* $B_{(\phi^0)}\ge 0$ *defined by*

$$
\begin{aligned}
&\beta:=2\kappa_{\hat{b}}+2L_{\bar{b}}L_{(\phi^0)}+C,\quad \mathfrak{m}_{(\alpha,\beta)}:=\sup_{t\in[0,T]}e^{2\alpha(T-t)}\int_t^T e^{(T-s)\beta}\,\mathrm{d}s,\\
&B_{(\phi^0)}:=C_{\bar{b}}^2\Big[L_g\mathfrak{m}_{(\alpha,\beta)}^{1/2}+\tfrac{e^{T\alpha}-1}{\alpha}\Big((C_Y+L_{fx})(1+L_{(\phi^0)})+C_\sigma L_Y\big(L_{(\phi^0)}\big)\Big)\Big],
\end{aligned} \tag{3.44}
$$

then for all $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$, *there exists a constant* $c\in[0,1)$ *such that*

$$
|\phi^{m+1}-\phi^m|_0\le c|\phi^m-\phi^{m-1}|_0\quad\forall m\in\mathbb{N}. \tag{3.45}
$$

Remark 3.1. Theorem 3.11 shows that if (3.25) and (3.43) hold, then the iterates $(\phi^m)_{m\in\mathbb{N}_0}$ form a Cauchy sequence, whose limit will be characterized in Theorem 3.13. We now observe that the inequalities (3.25) and (3.43) can be ensured if one of the conditions (i)–(vi) holds. To this end, we focus on (3.43), as (3.25) can be analyzed similarly. Suppose all remaining parameters are fixed. Then one can clearly see that (3.43) holds if (a) $\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu$ is sufficiently large or (b) $B_{(\phi^0)}$ is sufficiently small.

The former case holds if either μ or ν is sufficiently large (note that (2.2) and (2.3) imply that $\mu \le L_{fa}$). The latter case holds for (a) small $C_{\bar{b}}$, or (b) small $\frac{e^{T\alpha}-1}{\alpha}$ and $\mathfrak{m}_{(\alpha,\beta)}$, or (c) small L_g, C_Y, L_{fx}, and $L_Y(L_{(\phi^0)})$. By the definitions of α and β, $\frac{e^{T\alpha}-1}{\alpha}$ and $\mathfrak{m}_{(\alpha,\beta)}$ tend to 0 as $T \to 0$, or $\kappa_{\hat{b}} \to -\infty$, or $\rho \to \infty$ (see Lemma A.2 of the arXiv version [37]), while by (3.8) and (3.15), C_Y and $L_Y(L_{(\phi^0)})$ scale linearly in C_g, L_g, C_{fx}, L_{fx}, and hence $B_{(\phi^0)}$ is close to zero if C_g, L_g, C_{fx}, L_{fx} are sufficiently small.

Proof. For any $(t,x) \in [0,T] \times \mathbb{R}^n$, Lemma 3.4 with $x = x'$, $y = Y_t^{t,x,\phi^m}$, $y' = Y_t^{t,x,\phi^{m-1}}$, $a = \phi_t^m(x)$, and $a' = \phi_t^{m-1}(x)$ immediately yields that for all $\tau \in (0, \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}]$,

$$
\begin{aligned}
|\phi_t^{m+1}(x) - \phi_t^m(x)| \le \left(1 - \tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)|\phi_t^m(x) - \phi_t^{m-1}(x)| \\
+ \tau C_{\bar{b}}|Y_t^{t,x,\phi^m} - Y_t^{t,x,\phi^{m-1}}|.
\end{aligned} \tag{3.46}
$$

Applying Proposition 3.9 further gives

$$
\begin{aligned}
\frac{|\phi_t^{m+1}(x) - \phi_t^m(x)|}{1+|x|} \le \left(1 - \tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)\frac{|\phi_t^m(x) - \phi_t^{m-1}(x)|}{1+|x|} \\
+ \tau C_{\bar{b}} B[\phi^m, \phi^{m-1}, C_Z^{\phi^{m-1}}]|\phi^m - \phi^{m-1}|_0.
\end{aligned}
$$

Hence taking supremum over $(t,x) \in [0,T] \times \mathbb{R}^n$ results in

$$
|\phi^{m+1} - \phi^m|_0 \le \left(1 + \tau\left(C_{\bar{b}} B[\phi^m, \phi^{m-1}, C_Z^{\phi^{m-1}}] - \frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)\right)|\phi^m - \phi^{m-1}|_0,
$$

with the constant $B[\phi^m, \phi^{m-1}, C_Z^{\phi^{m-1}}]$ defined in (3.33). Observe that under (3.25), Theorem 3.8 shows that for all $\tau \in (0, \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}]$ and $m \in \mathbb{N}_0$, $[\phi^m]_1 \le L_{(\phi^0)}$, which along with Proposition 3.9 implies that $C_Z^{\phi^m} \le C_\sigma L_Y(L_{(\phi^0)})$. Hence, by Theorem 3.6 and (3.33),

$$
\begin{aligned}
&C_{\bar{b}} B[\phi^m, \phi^{m-1}, C_Z^{\phi^{m-1}}] \\
&\quad \le C C_{\bar{b}}^2 \Big(1 + T + T C_{\bar{b}} \sup_{t\in[0,T]} |\phi_t^{m-1}(0)|\Big) e^{T\beta_+} \\
&\qquad \times \left(L_g \mathfrak{m}_{(\alpha,\beta)}^{1/2} + \tfrac{e^{T\alpha}-1}{\alpha}\Big((C_Y + L_{fx})(1 + [\phi^m]_1) + C_Z^{\phi^{m-1}}\Big)(Te^{T\beta_+} + 1)\right) \\
&\quad \le C(1 + T + T C_{\bar{b}} C_{(\phi^0)}) e^{T\beta_+}(Te^{T\beta_+} + 1) B_{(\phi^0)},
\end{aligned}
$$

with β, $\mathfrak{m}_{(\alpha,\beta)}$ and $B_{(\phi^0)}$ defined as in (3.44). Then, under (3.43), the desired estimate holds with

$$
c = 1 + \tau\left(C(1 + T + T C_{\bar{b}} C_{(\phi^0)}) e^{T\beta_+}(Te^{T\beta_+} + 1) B_{(\phi^0)} - \frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right) \in [0,1).
$$

Note that (3.43) implies that $c < 1$, and $\tau \le \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}$ implies that $c \ge 1 - \frac{\tau}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}} + \nu\right) \ge 0$. This finishes the proof. ∎

3.5. Linear convergence to stationary points. Based on Theorem 3.11, we prove the linear convergence of the iterates $(\phi^m)_{m\in\mathbb{N}_0}$ in the weighted sup-norm $|\cdot|_0$ (see Definition 2.1) and the associated control processes $(\alpha^{\phi^m})_{m\in\mathbb{N}_0}$ to stationary points of $J(\cdot;\xi_0)$.

The following proposition characterizes stationary points of the summation of a nonconvex differentiable function and a convex nonsmooth function.

PROPOSITION 3.12. *Let X be a Hilbert space equipped with the norm $\|\cdot\|_X$, let $F : X \to \mathbb{R}$ be a Fréchet differentiable function, let $G : X \to \mathbb{R}\cup\{\infty\}$ be a proper, lower semicontinuous, convex function, and let $x^\star \in \operatorname{dom} G$. Then $x^\star$ is a stationary point of $F+G$ if and only if for some $\tau > 0$,*

$$x^\star = \operatorname{prox}_{\tau G}(x^\star - \tau\nabla F(x^\star)),$$

where for all $x \in X$, $\operatorname{prox}_{\tau G}(x) = \arg\min_{z\in X}\big(\frac{1}{2}\|z-x\|_X^2 + \tau G(z)\big)$.

Proof. By [31, Proposition 1.107], the Fréchet differentiability of F implies that $\partial(F+G)(x^\star) = \nabla F(x^\star) + \partial G(x^\star)$. Hence $x^\star$ is a stationary point of $F+G$ if and only if $-\nabla F(x^\star) \in \partial G(x^\star)$. By the properties of G, ∂G agrees with the convex subdifferential of G (see [31, equation 1.51 and Theorem 1.93]), which along with the definition of prox shows that for all $x, u \in X$ and $\tau > 0$, $u = \operatorname{prox}_{\tau G}(x)$ if and only if $x - u \in \partial(\tau G)(u)$. Hence, by $-\nabla F(x^\star) \in \partial G(x^\star)$, for all $\tau > 0$, $\big(x^\star - \tau\nabla F(x^\star)\big) - x^\star \in \partial(\tau G)(x^\star)$, which leads to the desired result. ∎

The following theorem presents a precise statement of Theorem 2.2, which establishes the linear convergence of the iterates $(\phi^m)_{m\in\mathbb{N}_0}$, and characterizes the limit of the associated control processes $(\alpha^{\phi^m})_{m\in\mathbb{N}_0}$ based on Proposition 3.12.

THEOREM 3.13. *Assume the same notation as in Theorem 3.11. For each $\phi \in \mathcal{V}_{\boldsymbol{A}}$, let $\alpha^\phi \in \mathcal{H}^2(\mathbb{R}^k)$ be the associated control process. Then, for all $\phi^0 \in \mathcal{V}_{\boldsymbol{A}}$ satisfying (3.25) and (3.43), and for all $\tau \in (0, \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}]$, there exist $c \in [0,1)$, $\widetilde{C} \geq 0$, and $\phi^\star \in \mathcal{V}_{\boldsymbol{A}}$ such that the following hold:*

(1) *for all $m \in \mathbb{N}_0$, $|\phi^{m+1} - \phi^\star|_0 \leq c|\phi^m - \phi^\star|_0$;*
(2) *for all $m \in \mathbb{N}_0$, $\|\alpha^{\phi^m} - \alpha^{\phi^\star}\|_{\mathcal{H}^2} \leq \widetilde{C}c^m|\phi^0 - \phi^\star|_0$;*
(3) *$\alpha^{\phi^\star}$ is a stationary point of $J(\cdot;\xi_0) : \mathcal{H}^2(\mathbb{R}^k) \to \mathbb{R}\cup\{\infty\}$ defined as in (1.2).*

Proof. Throughout the proof, let $\phi^0 \in \mathcal{V}_{\mathbf{A}}$ satisfy (3.25) and (3.43), and let $\tau \in (0, \frac{2}{\mu+L_{fa}} \wedge \frac{1}{\nu}]$. In the present setting, Theorem 3.8 implies that $\sup_{m\in\mathbb{N}_0}[\phi^m]_1 \leq L_{(\phi^0)}$, and Theorem 3.11 shows that $(\phi^m)_{m\in\mathbb{N}_0}$ is a Cauchy sequence in $(\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k), |\cdot|_0)$. As $(\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k), |\cdot|_0)$ is a Banach space, the Banach fixed point theorem shows that there exists $\phi^\star \in \mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k)$ such that $\lim_{m\to\infty}|\phi^m - \phi^\star|_0 = 0$. The convergence of $(\phi^m)_{m\in\mathbb{N}_0}$ in the $|\cdot|_0$-norm and $\sup_{m\in\mathbb{N}_0}[\phi^m]_1 \leq L_{(\phi^0)}$ imply that $[\phi^\star]_1 \leq L_{(\phi^0)}$. Hence, to show that $\phi^\star \in \mathcal{V}_{\mathbf{A}}$, it remains to prove that $\phi^\star$ takes values in $\mathbf{A}$ a.e.

By Proposition 3.9 and $\sup_{m\in\mathbb{N}_0, t\in[0,T]}(|\phi_t^m(0)| + [\phi^m]_1) < \infty$, there exists $C \geq 0$ such that for all $(t,x) \in [0,T]\times\mathbb{R}^n$ and $m, m' \in \mathbb{N}_0$, $|Y_t^{t,x,\phi^m} - Y_t^{t,x,\phi^{m'}}| \leq C(1+|x|)|\phi^m - \phi^{m'}|_0$. This along with the fact that $(\phi^m)_{m\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{B}([0,T]\times\mathbb{R}^n;\mathbb{R}^k), |\cdot|_0)$ shows that for all $(t,x) \in [0,T]\times\mathbb{R}^n$, $(Y_t^{t,x,\phi^m})_{m\in\mathbb{N}_0}$ is a Cauchy sequence in $\mathbb{R}^n$. Hence there exists a function $\mathcal{Y} : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ such that $\lim_{m\to\infty} Y_t^{t,x,\phi^m} = \mathcal{Y}_t(x)$ for all $(t,x) \in [0,T]\times\mathbb{R}^n$. Then, for any $(t,x) \in [0,T]\times\mathbb{R}^n$, by the continuity of $\operatorname{prox}_{\tau\ell}$ and $\partial_a H_t^{\mathrm{re}}$ and the pointwise convergence of $(\phi^m)_{m\in\mathbb{N}_0}$ and $(Y_t^{t,x,\phi^m})_{m\in\mathbb{N}_0}$, one can pass m to infinity in (1.9) and show for a.e. $(t,x) \in [0,T]\times\mathbb{R}^n$:

$$\phi_t^\star(x) = \lim_{m\to\infty} \phi_t^{m+1}(x) = \lim_{m\to\infty} \operatorname{prox}_{\tau\ell}\big(\phi_t^m(x) - \tau\partial_a H_t^{\mathrm{re}}(x,\phi_t^m(x),Y_t^{t,x,\phi^m})\big) = \operatorname{prox}_{\tau\ell}\big(\phi_t^\star(x) - \tau\partial_a H_t^{\mathrm{re}}(x,\phi_t^\star(x),\mathcal{Y}_t(x))\big). \tag{3.47}$$

As $\operatorname{prox}_{\tau\ell}(z) \in \operatorname{dom}\ell = \mathbf{A}$ for all $z\in\mathbb{R}^k$, $\phi_t^\star(x)\in\mathbf{A}$ for a.e. $(t,x)\in[0,T]\times\mathbb{R}^n$, and hence $\phi^\star\in\mathcal{V}_{\mathbf{A}}$. Furthermore, by $\phi^\star\in\mathcal{V}_{\mathbf{A}}$, $\lim_{m\to\infty}|\phi^m-\phi^\star|_0=0$, and Proposition 3.9, $\lim_{m\to\infty} Y_t^{t,x,\phi^m} = Y_t^{t,x,\phi^\star} = \mathcal{Y}_t(x)$ for all $(t,x)\in[0,T]\times\mathbb{R}^n$, which along with (3.47) shows that

$$\phi_t^\star(x) = \operatorname{prox}_{\tau\ell}\big(\phi_t^\star(x) - \tau\partial_a H_t^{\mathrm{re}}(x,\phi_t^\star(x),Y_t^{t,x,\phi^\star})\big). \tag{3.48}$$

We are now ready to establish the desired statements. To prove item (1), for any $(t,x)\in[0,T]\times\mathbb{R}^n$, Lemma 3.4 with $x'=x$, $y=Y_t^{t,x,\phi^m}$, $y'=Y_t^{t,x,\phi^\star}$, $a=\phi_t^m(x)$, and $a'=\phi_t^\star(x)$ and (3.48) immediately yield that for all $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$,

$$\begin{aligned}&|\phi_t^{m+1}(x)-\phi^\star(x)|\\ &\quad\le\left(1-\tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right)\right)|\phi_t^m(x)-\phi_t^\star(x)| + \tau C_{\bar b}|Y_t^{t,x,\phi^m}-Y_t^{t,x,\phi^\star}|.\end{aligned}$$

Now, following the exact same lines as the proof of Theorem 3.11 (cf. (3.46)) and using the above fact that $\phi^\star\in\mathcal{V}_{\mathbf{A}}$, $\sup_{t\in[0,T]}|\phi_t^\star(0)|\le C_{(\phi^0)}$ and $[\phi^\star]_1\le L_{(\phi^0)}$, we deduce that$|\phi^{m+1}-\phi^\star|_0\le c|\phi^m-\phi^\star|_0$, with the same constant $c\in[0,1)$ as in Theorem 3.11.

To prove item (2), observe that for each $m\in\mathbb{N}_0$, $\alpha^{\phi^m}=\phi^m(X^{\xi_0,\phi^m})$ and $\alpha^{\phi^\star}=\phi^\star(X^{\xi_0,\phi^\star})$, which implies that

$$\begin{aligned}&\|\alpha^{\phi^{m+1}}-\alpha^{\phi^\star}\|_{\mathcal{H}^2}\\ &\quad\le\|\phi^{m+1}(X^{\xi_0,\phi^{m+1}})-\phi^{m+1}(X^{\xi_0,\phi^\star})\|_{\mathcal{H}^2}+\|\phi^{m+1}(X^{\xi_0,\phi^\star})-\phi^\star(X^{\xi_0,\phi^\star})\|_{\mathcal{H}^2}\\ &\quad\le[\phi^{m+1}]_1\|X^{\xi_0,\phi^{m+1}}-X^{\xi_0,\phi^\star}\|_{\mathcal{H}^2}+|\phi^{m+1}-\phi^\star|_0(1+\|X^{\xi_0,\phi^\star}\|_{\mathcal{H}^2}).\end{aligned} \tag{3.49}$$

By using $\phi^\star\in\mathcal{V}_{\mathbf{A}}$ and $\sup_{m\in\mathbb{N}}(|\phi_t^m(0)|+[\phi^m]_1)<\infty$ and Lemma 3.1, one can easily show that there exists $C\ge0$ such that for all $m\in\mathbb{N}_0$, $\|X^{\xi_0,\phi^\star}\|_{\mathcal{H}^2}\le C$ and $\|X^{\xi_0,\phi^m}-X^{\xi_0,\phi^\star}\|_{\mathcal{H}^2}\le C|\phi^m-\phi^\star|_0$, which along with (3.49) leads to the desired estimate $\|\alpha^{\phi^{m+1}}-\alpha^{\phi^\star}\|_{\mathcal{H}^2}\le\widetilde{C}c^m|\phi^0-\phi^\star|_0$ for all $m\in\mathbb{N}_0$, with some constant $\widetilde{C}\ge0$ independent of m.

It remains to prove item (3). Let $\tilde H^{\mathrm{re}}:[0,T]\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}^n\to\mathbb{R}$ and $\tilde H:[0,T]\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}^n\times\mathbb{R}^{n\times d}\to\mathbb{R}$ be such that for all $(t,x,a,y,z)\in[0,T]\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}^n\times\mathbb{R}^{n\times d}$, $\tilde H_t^{\mathrm{re}}(x,a,y):=\langle b_t(x,a),y\rangle+e^{-\rho t}f_t(x,a)$ and $\tilde H_t(x,a,y,z):=\tilde H_t^{\mathrm{re}}(x,a,y)+\langle\sigma_t(x),z\rangle$. For each $(t,x)\in[0,T]\times\mathbb{R}^n$, let $X^{t,x,\phi^\star}\in\mathcal{S}^2(t,T;\mathbb{R}^n)$ satisfy (1.10) with $\phi=\phi^\star$, let $(\tilde Y^{t,x,\phi^\star},\tilde Z^{t,x,\phi^\star})\in\mathcal{S}^2(t,T;\mathbb{R}^n)\times\mathcal{H}^2(t,T;\mathbb{R}^{n\times d})$ satisfy $\tilde Y_T^{t,x,\phi^\star}=e^{-\rho T}\partial_x g(X_T^{t,x,\phi^\star})$, and let

$$\mathrm{d}\tilde Y_s^{t,x,\phi^\star}=-\partial_x\tilde H_s(X_s^{t,x,\phi^\star},\phi_s^\star(X_s^{t,x,\phi^\star}),\tilde Y_s^{t,x,\phi^\star},\tilde Z_s^{t,x,\phi^\star})\,\mathrm{d}s+\tilde Z_s^{t,x,\phi^\star}\,\mathrm{d}W_s\quad\forall s\in[t,T).$$

The affineness of H and $\tilde H$ in y and z implies that $\tilde Y_s^{t,x,\phi^\star}=e^{-\rho s}Y_s^{t,x,\phi^\star}$ for all $(t,x)\in[0,T]\times\mathbb{R}^n$ and $s\in[t,T]$. Moreover, by (H.1) and (1.8), for all $a,u\in\mathbb{R}^k$ and $\eta>0$,

$$\begin{aligned}a=\operatorname{prox}_\ell(a-\eta u)&\iff0\in(a-(a-\eta u))+\partial\ell(a)\iff0\in u+\partial(\eta^{-1}\ell)(a)\\ &\iff0\in(a-(a-u))+\partial(\eta^{-1}\ell)(a)\iff a=\operatorname{prox}_{\eta^{-1}\ell}(a-u).\end{aligned}$$

Hence, by (3.48) and the affineness of H^{re} and $\tilde{H}^{\mathrm{re}}$ in y, for all $(t,x)\in[0,T]\times\mathbb{R}^n$,

$$\begin{aligned}\phi_t^\star(x) &= \mathrm{prox}_{\tau\ell}\big(\phi_t^\star(x)-\tau e^{\rho t}\partial_a\tilde{H}_t^{\mathrm{re}}(x,\phi_t^\star(x),e^{-\rho t}Y_t^{t,x,\phi^\star})\big)\\ &= \mathrm{prox}_{\tau e^{-\rho t}\ell}\big(\phi_t^\star(x)-\tau\partial_a\tilde{H}_t^{\mathrm{re}}(x,\phi_t^\star(x),\tilde{Y}_t^{t,x,\phi^\star})\big).\end{aligned} \tag{3.50}$$

Now consider the solution $(X^{\xi_0,\phi^\star},\tilde{Y}^{\xi_0,\phi^\star},\tilde{Z}^{\xi_0,\phi^\star})\in\mathcal{S}^2(\mathbb{R}^n)\times\mathcal{S}^2(\mathbb{R}^n)\times\mathcal{H}^2(\mathbb{R}^{n\times d})$ to the following FBSDE: for all $t\in[0,T]$,

$$\begin{aligned}\mathrm{d}X_t^{\xi_0,\phi^\star} &= b_t(X_t^{\xi_0,\phi^\star},\phi_t^\star(X_t^{\xi_0,\phi^\star}))\,\mathrm{d}t+\sigma_t(X_t^{\xi_0,\phi^\star})\,\mathrm{d}W_t,\\ \mathrm{d}\tilde{Y}_t^{\xi_0,\phi^\star} &= -\partial_x\tilde{H}_t(X_t^{\xi_0,\phi^\star},\phi_t^\star(X_t^{\xi_0,\phi^\star}),\tilde{Y}_t^{\xi_0,\phi^\star},\tilde{Z}_t^{\xi_0,\phi^\star})\,\mathrm{d}t+\tilde{Z}_t^{\xi_0,\phi^\star}\,\mathrm{d}W_t,\\ X_0^{\xi_0,\phi^\star} &= \xi_0,\quad \tilde{Y}_T^{\xi_0,\phi^\star}=e^{-\rho T}\partial_x g(X_T^{\xi_0,\phi^\star}).\end{aligned}$$

The Markov property in [44, Theorem 5.1.3] implies that $\tilde{Y}_t^{\xi_0,\phi^\star}=\tilde{Y}_t^{t,X_t^{\xi_0,\phi^\star},\phi^\star}$ $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$ a.e., which along with (3.50) gives that for $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$ a.e.,

$$\phi_t^\star(X_t^{\xi_0,\phi^\star})=\mathrm{prox}_{\tau e^{-\rho t}\ell}\big(\phi_t^\star(X_t^{\xi_0,\phi^\star})-\tau\partial_a\tilde{H}_t^{\mathrm{re}}(X_t^{\xi_0,\phi^\star},\phi_t^\star(X_t^{\xi_0,\phi^\star}),\tilde{Y}_t^{\xi_0,\phi^\star})\big). \tag{3.51}$$

Observe that for all $\alpha\in\mathcal{H}^2(\mathbb{R}^k)$, $J(\alpha;\xi_0)=F(\alpha)+G(\alpha)$, where

$$F(\alpha):=\mathbb{E}\bigg[\int_0^T e^{-\rho t}f_t(X_t^{\xi_0,\alpha},\alpha_t)\,\mathrm{d}t+e^{-\rho T}g(X_T^{\xi_0,\alpha})\bigg],\quad G(\alpha):=\mathbb{E}\bigg[\int_0^T e^{-\rho t}\ell(\alpha_t)\,\mathrm{d}t\bigg],$$

where $X^{\xi_0,\alpha}$ satisfies (1.1). Then the regularity of coefficients and [8, Corollary 4.11] imply that F is Fréchet differentiable, and the derivative ∇F at $\alpha_\cdot^{\phi^\star}=\phi_\cdot^\star(X_\cdot^{\xi_0,\phi^\star})$ is given by

$$\nabla F(\alpha^{\phi^\star})_t=\partial_a\tilde{H}_t^{\mathrm{re}}(X_t^{\xi_0,\phi^\star},\phi_t^\star(X_t^{\xi_0,\phi^\star}),\tilde{Y}_t^{\xi_0,\phi^\star}),\quad \mathrm{d}t\otimes\mathrm{d}\mathbb{P}\text{ a.e.}$$

Moreover, by (H.1(1)), one can easily prove that G is proper, lower semicontinuous, and convex and satisfies for all $\alpha\in\mathcal{H}^2(\mathbb{R}^k)$ and $\tau>0$, $\mathrm{prox}_{\tau G}(\alpha)=\mathrm{prox}_{\tau e^{-\rho t}\ell}(\alpha)$ for $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$ a.e. Hence, Proposition 3.12 shows that $\alpha^{\phi^\star}\in\mathcal{H}^2(\mathbb{R}^k)$ is a stationary point of $J(\cdot;\xi_0)$. ∎

The following theorem presents a precise statement of Theorem 2.3, where the feedback controls $(\phi^m)_{m\in\mathbb{N}_0}$ are updated with approximate gradients.

THEOREM 3.14. *Suppose* (H.1) *holds. Let $C\geq 0$ be a constant such that* (3.8), (3.15), *and* (3.32) *hold, let $C_Y\geq 0$ be defined in* (3.8), *and let $\alpha\in\mathbb{R}$ be defined in* (3.16). *Let $\phi^0\in\mathcal{V}_{\boldsymbol{A}}$, let $C_{(\phi^0)}\geq 0$ be defined in* (3.11), *let $L_{(\phi^0)}\geq 0$ be defined in* (3.26), *and let $L_Y(M)$, $M\geq 0$, be defined in* (3.16). *Suppose that* (3.25) *is satisfied, $\tau\in(0,\frac{2}{\mu+L_{fa}}\wedge\frac{1}{\nu}]$, and there exist constants $\widetilde{C},\widetilde{L}\geq 0$ such that for all $\omega\in\Omega$, $\sup_{t\in[0,T]}|\widetilde{\phi}_t^m(0,\omega)|\leq\widetilde{C}$ and $[\widetilde{\phi}_\cdot^m(\cdot,\omega)]_1\leq\widetilde{L}$ for all $m\in\mathbb{N}_0$, and*

$$\mathfrak{D}:=\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu+L_{fa}}+\nu\right)-C(1+T+TC_{\bar{b}}\widetilde{C})e^{T\beta_+}(Te^{T\beta_+}+1)\widetilde{B}>0, \tag{3.52}$$

with the constants $\beta\in\mathbb{R}$, $\mathfrak{m}_{(\alpha,\beta)}>0$, and $\widetilde{B}\geq 0$ defined by

$$\begin{aligned}&\beta:=2\kappa_{\hat{b}}+2L_{\bar{b}}\max\{L_{(\phi^0)},\widetilde{L}\}+C,\quad \mathfrak{m}_{(\alpha,\beta)}:=\sup_{t\in[0,T]}e^{2\alpha(T-t)}\int_t^T e^{(T-s)\beta}\,\mathrm{d}s,\\ &\widetilde{B}:=C_{\bar{b}}^2\Big[L_g\mathfrak{m}_{(\alpha,\beta)}^{1/2}+\tfrac{e^{T\alpha}-1}{\alpha}\Big((C_Y+L_{fx})(1+L_{(\phi^0)})+C_\sigma L_Y(\widetilde{L})\Big)\Big].\end{aligned} \tag{3.53}$$

Let $c = 1 - \tau\mathfrak{D} \in [0,1)$. *Then, for a.s.* $\omega \in \Omega$ *and for all* $m \in \mathbb{N}_0$,

(3.54)
$$|\phi^\star - \widetilde{\phi}^m_\cdot(\cdot,\omega)|_0 \le c^m|\phi^0 - \phi^\star|_0 + \sum_{j=0}^{m-1} c^{m-1-j}\tau C_{\bar{b}} \sup_{(t,x)\in[0,T]\times\mathbb{R}^n} \frac{|\mathcal{Y}_t^{\widetilde{\phi}^j}(x,\omega) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^j}(x,\omega)|}{1+|x|},$$

where $\phi^\star \in \mathcal{V}_{\boldsymbol{A}}$ *is the limit function in Theorem* 3.13. *Consequently, for all* $p \ge 1$ *and for all* $m \in \mathbb{N}_0$,

$$\mathbb{E}[|\phi^\star - \widetilde{\phi}^m|_0^p]^{\frac{1}{p}} \le c^m|\phi^0 - \phi^\star|_0 + \sum_{j=0}^{m-1} c^{m-1-j}\tau C_{\bar{b}}\mathbb{E}[|\mathcal{Y}^{\widetilde{\phi}^j} - \widetilde{\mathcal{Y}}^{\widetilde{\phi}^j}|_0^p]^{\frac{1}{p}}. \tag{3.55}$$

Proof. Throughout this proof, we fix $\omega \in \Omega$ and omit the explicit dependence on ω if no confusion occurs. First, observe that the conditions $\sup_{t\in[0,T]}|\widetilde{\phi}^m_t(0)| \le \widetilde{C}$ and $[\widetilde{\phi}^m]_1 \le \widetilde{L}$ for all $m \in \mathbb{N}_0$ guarantee $\widetilde{\phi}^m \in \mathcal{V}_{\mathbf{A}}$. We continue by quantifying $|\phi_t^{m+1}(x) - \widetilde{\phi}_t^{m+1}(x)|$ for any $(t,x) \in [0,T]\times\mathbb{R}^n$, where $\widetilde{\phi}_t^{m+1}(x)$ is defined by (2.20). By Lemma 3.4 with $x = x'$, $a = \phi_t^m(x)$, $a' = \widetilde{\phi}_t^m(x)$, $y = \mathcal{Y}_t^{\phi^m}(x)$, and $y' = \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x)$,

$$\begin{aligned} &|\phi_t^{m+1}(x) - \widetilde{\phi}_t^{m+1}(x)| \\ &\quad \le \left(1 - \tau\frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)|\phi_t^m(x) - \widetilde{\phi}_t^m(x)| + \tau C_{\bar{b}}|\mathcal{Y}_t^{\phi^m}(x) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x)|. \end{aligned} \tag{3.56}$$

Now, by applying Proposition 3.9 and using the definition of $\widetilde{B}$ given in (3.53),

$$\begin{aligned} C_{\bar{b}}|\mathcal{Y}_t^{\phi^m}(x) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x)| &\le C_{\bar{b}}|\mathcal{Y}_t^{\phi^m}(x) - \mathcal{Y}_t^{\widetilde{\phi}^m}(x)| + C_{\bar{b}}|\mathcal{Y}_t^{\widetilde{\phi}^m}(x) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x)| \\ &\le C(1 + T + TC_{\bar{b}}\widetilde{C})e^{T\beta_+}(Te^{T\beta_+} + 1)\widetilde{B}(1+|x|)|\phi^m - \widetilde{\phi}^m|_0 + C_{\bar{b}}|\mathcal{Y}_t^{\widetilde{\phi}^m}(x) - \widetilde{\mathcal{Y}}_t^{\widetilde{\phi}^m}(x)|. \end{aligned}$$

Now let $c = 1 - \tau\mathfrak{D}$, with $\mathfrak{D} > 0$ defined in (3.52). The condition (3.52) implies that $c < 1$, and $\tau \le \frac{2}{\mu + L_{fa}} \wedge \frac{1}{\nu}$ implies that $c \ge 1 - \frac{\tau}{2}(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu) \ge 0$. Hence, for a.s. $\omega \in \Omega$,

$$|\phi^{m+1} - \widetilde{\phi}^{m+1}_\cdot(\cdot,\omega)|_0 \le c|\phi^m - \widetilde{\phi}^m_\cdot(\cdot,\omega)|_0 + \tau C_{\bar{b}}|\mathcal{Y}_\cdot^{\widetilde{\phi}^m}(\cdot,\omega) - \widetilde{\mathcal{Y}}_\cdot^{\widetilde{\phi}^m}(\cdot,\omega)|_0, \tag{3.57}$$

which along with $\phi^0 = \widetilde{\phi}^0$ implies that $|\phi^{m+1} - \widetilde{\phi}^{m+1}|_0 \le \sum_{j=0}^m c^{m-j}\tau C_{\bar{b}}|\mathcal{Y}_\cdot^{\widetilde{\phi}^m}(\cdot,\omega) - \widetilde{\mathcal{Y}}_\cdot^{\widetilde{\phi}^m}(\cdot,\omega)|_0$. The estimate (3.54) then follows from Theorem 3.13, item (1), and the estimate (3.55) follows by taking the L^p-norm on both sides of (3.54). ∎

REFERENCES

[1] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces*, Springer, Dordrecht, 2012.

[2] G. Barles and E. R. Jakobsen, *On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations*, ESAIM Math. Model. Numer. Anal., 36 (2002), pp. 33–54.

[3] E. Bayraktar, A. Budhiraja, and A. Cohen, *A numerical scheme for a mean field game in some queueing systems based on Markov chain approximation method*, SIAM J. Control Optim., 56 (2018), pp. 4017–4044, https://doi.org/10.1137/17M1154357.

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, https://doi.org/10.1137/080716542.

[5] C. Bender and J. Zhang, *Time discretization and Markovian iteration for coupled FBSDEs*, Ann. Appl. Probab., 18 (2008), pp. 143–177.

[6] A. Bensoussan, J. Frehse, and P. Yam, *Mean Field Games and Mean Field Type Control Theory*, SpringerBriefs Math. 101, Springer, New York, 2013.

[7] P. Briand and R. Carmona, *BSDEs with polynomial growth generators*, J. Appl. Math. Stoch. Anal., 13 (2000), pp. 207–238.

[8] R. Carmona, *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*, SIAM, Philadelphia, 2016, https://doi.org/10.1137/1.9781611974249.

[9] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, *Global convergence of policy gradient methods for the linear quadratic regulator*, in Proceedings of the International Conference on Machine Learning, PMLR, 2018, pp. 1467–1476.

[10] M. Giegrich, C. Reisinger, and Y. Zhang, *Convergence of Policy Gradient Methods for Finite-Horizon Stochastic Linear-Quadratic Control Problems*, preprint, arXiv:2211.00617, 2022.

[11] E. Gobet, J.-P. Lemor, and X. Warin, *A regression-based Monte Carlo method to solve backward stochastic differential equations*, Ann. Appl. Probab., 15 (2005), pp. 2172–2202.

[12] H. Gu, X. Guo, X. Wei, and R. Xu, *Mean-Field Multi-agent Reinforcement Learning: A Decentralized Network Approach*, preprint, arXiv:2108.02731, 2021.

[13] X. Guo, A. Hu, and J. Zhang, *Theoretical Guarantees of Fictitious Discount Algorithms for Episodic Reinforcement Learning and Global Convergence of Policy Gradient Methods*, preprint, arXiv:2109.06362, 2021.

[14] X. Guo, A. Hu, and Y. Zhang, *Reinforcement Learning for Linear-Convex Models with Jumps via Stability Analysis of Feedback Controls*, preprint, arXiv:2104.09311, 2021.

[15] B. Hambly, R. Xu, and H. Yang, *Policy gradient methods for the noisy linear quadratic regulator over a finite horizon*, SIAM J. Control Optim., 59 (2021), pp. 3359–3391, https://doi.org/10.1137/20M1382386.

[16] J. Han and W. E, *Deep Learning Approximation for Stochastic Control Problems*, preprint, arXiv:1611.07422, 2016.

[17] J. Han, A. Jentzen, and W. E, *Solving high-dimensional partial differential equations using deep learning*, Proc. Natl. Acad. Sci. USA, 115 (2018), pp. 8505–8510.

[18] R. Hu, *Deep Fictitious Play for Stochastic Differential Games*, preprint, arXiv:1903.09376, 2019.

[19] C. Huré, H. Pham, and X. Warin, *Deep backward schemes for high-dimensional nonlinear PDEs*, Math. Comp., 89 (2020), pp. 1547–1579.

[20] A. Isidori, *Nonlinear Control Systems: An Introduction*, Springer-Verlag, Berlin, 1985.

[21] K. Ito, C. Reisinger, and Y. Zhang, *A neural network-based policy iteration algorithm with global H^2-superlinear convergence for stochastic games on domains*, Found. Comput. Math., 21 (2021), pp. 331–374.

[22] Y. Jia and X. Y. Zhou, *Policy Gradient and Actor-Critic Learning in Continuous Time and Space: Theory and Algorithms*, preprint, arXiv:2111.11232, 2021.

[23] B. Kerimkulov, J.-M. Leahy, D. Šiška, and Ł. Szpruch, *Convergence of Policy Gradient for Entropy Regularized MDPs with Neural Network Approximation in the Mean-Field Regime*, preprint, arXiv:2201.07296, 2022.

[24] B. Kerimkulov, D. Šiška, and L. Szpruch, *Exponential convergence and stability of Howard's policy improvement algorithm for controlled diffusions*, SIAM J. Control Optim., 58 (2020), pp. 1314–1340, https://doi.org/10.1137/19M1236758.

[25] B. Kerimkulov, D. Šiška, and Ł. Szpruch, *A modified MSA for stochastic control problems*, Appl. Math. Optim., 84 (2021), pp. 3417–3436.

[26] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Appl. Math. (N. Y.) 24, Springer-Verlag, New York, 2001.

[27] Q. Li, L. Chen, C. Tai, and W. E, *Maximum principle based algorithms for deep learning*, J. Mach. Learn. Res., 18 (2017), 165.

[28] A. Lionnet, G. Dos Reis, and L. Szpruch, *Time discretization of FBSDE with polynomial growth drivers and reaction–diffusion PDEs*, Ann. Appl. Probab., 25 (2015), pp. 2563–2625.

[29] B. Maslowski and P. Veverka, *Sufficient stochastic maximum principle for discounted control problem*, Appl. Math. Optim., 70 (2014), pp. 225–252.

[30] J. Mei, C. Xiao, C. Szepesvari, and D. Schuurmans, *On the global convergence rates of softmax policy gradient methods*, in Proceedings of the International Conference on Machine Learning, PMLR, 2020, pp. 6820–6829.

[31] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I: *Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, Heidelberg, 2006.

[32] R. Munos, *Policy gradient in continuous time*, J. Mach. Learn. Res., 7 (2006), pp. 771–791.

[33] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Appl. Optim. 87, Kluwer Academic Publishers, Boston, MA, 2004.

[34] É. Pardoux, *Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order*, in Stochastic Analysis and Related Topics VI, Birkhäuser Boston, Boston, MA, 1998, pp. 79–127.

[35] H. Pham, *Continuous-Time Stochastic Control and Optimization with Financial Applications*, Stoch. Model. Appl. Probab. 61, Springer-Verlag, Berlin, 2009.

[36] C. Reisinger, W. Stockinger, and Y. Zhang, *A Fast Iterative PDE-based Algorithm for Feedback Controls of Nonsmooth Mean-Field Control Problems*, preprint, arXiv: 2108.06740, 2021.

[37] C. Reisinger, W. Stockinger, and Y. Zhang, *Linear Convergence of a Policy Gradient Method for Finite Horizon Continuous Time Stochastic Control Problems*, preprint, arXiv:2203.11758, 2022.

[38] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, Heidelberg, 2009.

[39] D. Šiška and Ł. Szpruch, *Gradient Flows for Regularized Stochastic Control Problems*, preprint, arXiv:2006.05956, 2020.

[40] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 2018.

[41] Ł. Szpruch, T. Treetanthiploet, and Y. Zhang, *Exploration-Exploitation Trade-Off for Continuous-Time Episodic Reinforcement Learning with Linear-Convex Models*, preprint, arXiv:2112.10264, 2021.

[42] C. Tallec, L. Blier, and Y. Ollivier, *Making deep q-learning methods robust to time discretization*, in Proceedings of the International Conference on Machine Learning, PMLR, 2019, pp. 6096–6104.

[43] L. Wang, Q. Cai, Z. Yang, and Z. Wang, *Neural Policy Gradient Methods: Global Optimality and Rates of Convergence*, preprint, arXiv:1909.01150, 2019.

[44] J. Zhang, *Backward Stochastic Differential Equations*, Springer, New York, 2017.