Linear convergence of a policy gradient method for finite horizon continuous time stochastic control problems

Christoph Reisinger*, Wolfgang Stockinger*, Yufei Zhang †

Abstract. Despite its popularity in the reinforcement learning community, a provably convergent policy gradient method for general continuous space-time stochastic control problems has been elusive. This paper closes the gap by proposing a proximal gradient algorithm for feedback controls of finite-time horizon stochastic control problems. The state dynamics are continuous time nonlinear diffusions with controlled drift and possibly degenerate noise, and the objectives are nonconvex in the state and nonsmooth in the control. 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

AMS subject classifications. 68Q25, 93E20, 49M05

1 Introduction

Stochastic control problems seek optimal strategies to control continuous time stochastic systems and are ubiquitous in modern science, engineering and economics [23, 33]. 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., [30, 14, 37, 19]). 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 parametrization 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. Analysing the convergence behavior of PGMs is technically challenging, as the objective of a control problem is typically nonconvex with respect to the policies, even in the linear-quadratic (LQ) setting [9, 13]. 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 [28] or overparametrized one-hidden-layer neural-network policies [40, 10, 20], and discrete time LQ control problems with linear parameterized policies [9, 13]. The analysis therein exploits heavily the specific structure of the considered (discrete time) control problems and policy parameterizations, 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 general continuous time stochastic 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 [39]. 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).

Proximal PGMs for nonsmooth control problems. This work provides an affirmative answer to the above challenging open question for control problems with uncontrolled diffusion coefficients. 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, $\mathbb{F} = (\mathcal{F}_t)_{t \in [0,T]}$ be the natural filtration of $W$ augmented with an independent $\sigma$-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:

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

(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 objective of the agent is to minimize the following cost functional

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

(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 condition on the coefficients in (1.1)-(1.2) will be given in Section 2. In particular, we allow $\ell$ to be discontinuous and to 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.

By interpreting (1.1)-(1.2) as an optimization problem over $\mathcal{H}^2(\mathbb{R}^k)$, one can design a gradient-descent algorithm for open-loop controls of the problem. Let $H^{re}: [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$ be defined by

$$
H^{re}_t(x, a, y) := \langle b_t(x, a), y \rangle + f_t(x, a) - \rho \langle x, y \rangle, \quad \forall (t, x, a, y) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n,
$$

(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 defined by

$$
H_t(x, a, y, z) := H^{re}_t(x, a, y) + \langle \sigma_t(x), z \rangle, \quad \forall (t, x, a, y, z) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^{n \times d}.
$$

(1.4)

\(^1\)We specify the explicit dependence on $\rho$, as it impacts the convergence rate of PGMs.
Then for an initial guess \( \alpha^0 \in \mathcal{H}^2(\mathbb{R}^k) \) and a stepsize \( \tau > 0 \), we consider the sequence \((\alpha^m)_{m \in \mathbb{N}} \subset \mathcal{H}^2(\mathbb{R}^k)\) such that for all \( m \in \mathbb{N}_0 \),

\[
\alpha^{m+1}_t = \text{prox}_{\tau \ell}(\alpha^m_t - \tau \partial_\alpha H^e_t(\phi^m_t, \alpha^m_t, Y^\xi_t)), \quad \text{for } dt \otimes d\mathbb{P} \text{ a.e.,}
\]

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

\[
\begin{align*}
\mathrm{d}X^{\xi_0}_{t} &= b_t(X^{\xi_0}_{t}, \alpha^m_t) \mathrm{d}t + \sigma_t(X^{\xi_0}_{t}, \alpha^m_t) \mathrm{d}W_t, \\
\mathrm{d}Y^{\xi_0}_{t} &= -\partial_\alpha H^e_t(X^{\xi_0}_{t}, \alpha^m_t, Y^{\xi_0}_{t}, Z^{\xi_0}_{t}) \mathrm{d}t + \partial_\xi \sigma_t(X^{\xi_0}_{t}, \alpha^m_t) \mathrm{d}W_t,
\end{align*}
\]

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

\[
\text{prox}_{\tau \ell}(a) = \arg \min_{z \in \mathbb{R}^k} \left( \frac{1}{2} \| z - a \|^2 + \tau \ell(z) \right), \quad \forall a \in \mathbb{R}^k.
\]

Note that \((\ref{1.7})\) involves undiscounted costs and the term \( \rho Y^{\xi_0}_{t} \), and arises from stochastic maximum principle for the discounted problem (see \([27]\))

The iteration \((\ref{1.5})\) is a proximal gradient method for \((\ref{1.2})\). The term \( \partial_\alpha H^e_t(X^{\xi_0}_{t}, \alpha^m_t, Y^{\xi_0}_{t}) \) is related to (up to an exponential time scaling) the Fréchet derivative of the differentiable component of \( J(\cdot; \xi_0) \) at the iterate \( \alpha^m \), while the function \( \text{prox}_{\tau \ell} \) 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 \([34]\) for a detailed derivation of the algorithm and to \([22, 36]\) for similar gradient-based algorithms without the nonsmooth term \( \ell \).

The main drawback of the proximal gradient algorithm \((\ref{1.5})\) (as well as the algorithms in \([22, 36]\)) 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 \((\ref{1.5})\) is difficult to implement in practice. In the sequel, we overcome the shortcoming of \((\ref{1.5})\) and introduce an analogue proximal gradient method for feedback controls of \((\ref{1.1})-(\ref{1.2})\), which is referred to as the proximal policy gradient method (PPGM).

To this end, we consider a class \( \mathcal{V}_A \) 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 \( \phi^0 \in \mathcal{V}_A \) and a stepsize \( \tau > 0 \), the PPGM generates the sequence \((\phi^m)_{m \in \mathbb{N}} \subset \mathcal{V}_A \) such that for all \( m \in \mathbb{N}_0 \),

\[
\phi^{m+1}_t(x) = \text{prox}_{\tau \ell}(\phi^m_t(x) - \tau \partial_\alpha H^e_t(x, \phi^m_t(x), Y^{t,x,\phi}_t)), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n,
\]

where \( \text{prox}_{\tau \ell} \) is defined in \((\ref{1.8})\), and for each \( \phi \in \mathcal{V}_A \) 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 FBSDE: for all \( s \in [t, T] \),

\[
\begin{align*}
\mathrm{d}X^{t,x,\phi}_s &= b_s(X^{t,x,\phi}_s, \phi_s(X^{t,x,\phi}_s)) \mathrm{d}s + \sigma_s(X^{t,x,\phi}_s) \mathrm{d}W_s, \\
\mathrm{d}Y^{t,x,\phi}_s &= -\partial_\alpha H_s(X^{t,x,\phi}_s, \phi_s(X^{t,x,\phi}_s), Y^{t,x,\phi}_s, Z^{t,x,\phi}_s) \mathrm{d}s + \partial_\xi \sigma_s(X^{t,x,\phi}_s) \mathrm{d}W_s, \\
Y^{t,x,\phi}_t &= x,
\end{align*}
\]

The iteration \((\ref{1.9})\) is motivated by the observation that if \( \alpha^{m}_t = \phi^m_t(X^{\xi_0}_t) \) with \( X^{\xi_0}_t \) being the state process controlled by the policy \( \phi^m \), then \( Y^{t,x,\phi^m}_t = Y^{t,x,\phi^m}_t|_{x=X^{\xi_0}_t} \) for \( dt \otimes d\mathbb{P} \) a.e. In other words, at the \( m \)-th iteration, \((\ref{1.9})\) evaluates the functional derivative of \( J(\cdot; \xi_0) \) at the open-loop control induced by the current policy \( \phi^m \), and obtains an updated policy based on a Markovian representation of the gradient.

The PPGM \((\ref{1.9})\) improves the efficiency of the policy iteration (see \([21, 18]\)) and the Method of Successive Approximation (see \([24, 22]\)) 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 \([34]\) by solving the linear BSDE \((\ref{1.11})\) numerically; see e.g., \([16]\) and \([34]\) 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, nonconvex and nonsmooth cost functions, and unbounded action space. 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 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 [18].

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 parametrization. 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 [28], while in the LQ setting, eigenvalues of state covariance matrices must be lower bounded away from zero over the entire time horizon [9, 13]. 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 analyse 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 \((E, |\cdot|)\), we introduce the following spaces:

- \(\mathcal{S}^p(t, T; E)\), for \(t \in [0, T]\) and \(p \geq 2\), is the space of \(E\)-valued \(\mathcal{F}\)-progressively measurable processes \(Y : [t, T] \times \Omega \rightarrow E\) satisfying \(\|Y\|_{\mathcal{S}^p} = \mathbb{E}[\sup_{s \in [t, T]} |Y_s|^p]^{1/p} < \infty\); \(^2\)

- \(\mathcal{H}^p(t, T; E)\), for \(t \in [0, T]\) and \(p \geq 2\), is the space of \(E\)-valued \(\mathcal{F}\)-progressively measurable processes \(Z : [t, T] \times \Omega \rightarrow E\) satisfying \(\|Z\|_{\mathcal{H}^p} = \mathbb{E}[\int_t^T |Z_s|^2 |d\sigma_s|^{p/2}]^{1/p} < \infty\).

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

\(^2\)With a slight abuse of notation, we denote by sup the essential supremum of a real-valued (Borel) measurable function throughout this paper.
2 Standing assumptions and main results

In this section, we state our main results on the linear convergence of the PPGM (1.9). The following assumptions on the coefficients of (1.1)-(1.2) are imposed throughout the paper.

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:

1. $\ell$ is lower semicontinuous and its effective domain $A := \{ z \in \mathbb{R}^k \mid \ell(z) < \infty \}$ is nonempty; \footnote{We say a function $f : X \to \mathbb{R} \cup \{\infty\}$ is proper if it has a nonempty effective domain $\text{dom } f := \{ x \in X \mid f(x) < \infty \}$.}

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

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

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

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

$$\eta f_t(x, a) + (1 - \eta) f_t(x, a') \geq f_t(x, \eta a + (1 - \eta) a') + \eta(1 - \eta)\frac{\nu}{2}|a - a'|^2, \quad (2.3)$$

$$\eta f_t(x, a) + (1 - \eta) f_t(x, a') \geq f_t(x, \eta a + (1 - \eta) a') + \eta(1 - \eta)\frac{\nu}{2}|a - a'|^2; \quad (2.4)$$

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

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

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

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

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

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

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

$$|\hat{b}_t(x) - \hat{b}_t(x')| + |\hat{b}_t(x)a - \hat{b}_t(x)a'| + |\partial_x \hat{b}_t(x)a - \partial_x \hat{b}_t(x)a'| \leq L_b(|x - x'| + |a - a'|); \quad (2.9)$$

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

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

Remark 2.1. The action set $A$ may be unbounded, and hence (2.9) cannot be further simplified. If one assumes further that $A$ is bounded, then, by the boundedness of $\hat{b}$ in (2.7), (2.9) is equivalent to the Lipschitz continuity of $\hat{b}$ and $\partial_x \hat{b}$. Alternatively, if $A$ is unbounded, then (2.9) is equivalent to the condition that $\hat{b}$ is independent in $x$. 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 \((\partial_x \tilde{b}_t(x) + \partial_a \tilde{b}_t(x)a)^\top y + \partial_x \sigma(x)^\top z \) (resp. \( \tilde{b}_t(x)^\top y \)), 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 the iterative scheme (1.9) generates Lipschitz continuous policies \((\phi^m)_{m \in N_0}\) (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 [12, 38].

For clarity, (2.3) and (2.4) assume convexity of \( a \mapsto f_i(x, a) \) and \( a \mapsto \ell(a) \) and strong convexity of \( a \mapsto f_i(x, a) + \ell(a) \). This allows for characterizing the rate of convergence of \((\phi^m)_{m \in N_0}\) in terms of \( \mu \) and \( \nu \). Similar analysis can be performed if (2.3) is relaxed into the following semi-convexity condition, i.e., there exists \( \mu \in [-6f_0, 6f_0] \) such that for all \((t, x) \in [0, T] \times \mathbb{R}^n\), \( a, a' \in A \) and \( \eta \in [0, 1] \),

\[
\eta f_i(x, a) + (1 - \eta) f_i(x, a') \geq f_i(x, \eta a + (1 - \eta) a') + \eta(1 - \eta) \mu \| a - a' \|^2;
\]

and \( \ell \) is \( \nu \)-strongly convex with a sufficiently large \( \nu \) (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 \( A \subset \mathbb{R}^k \) be a nonempty closed convex set and \( \ell : \mathbb{R}^k \to [0, \infty] \) be the indicator of \( A \) satisfying \( \ell(a) = 0 \) for \( a \in A \) and \( \ell(a) = \infty \) for \( a \in \mathbb{R}^k \setminus A \). Then (2.3) holds with \( \nu = 0 \), and for all \( \tau > 0 \), \( \text{prox}_{\tau \ell} \) is the orthogonal projection on \( A \). In this case, (1.9) extends the projected PGM in [13] to general stochastic control problems.

**Example 2.2** (Sparse control). Let \( (\gamma_i)_{i=1}^k \subset [0, \infty) \) and \( \ell : \mathbb{R}^k \to [0, \infty) \) 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 \( \tau > 0 \), \( \text{prox}_{\tau \ell}(a) = (\max\{a_i - \tau \gamma_i, 0\} \text{sgn}(a_i))_{i=1}^k \) 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 [5, 34]).

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

\[
\ell(a) := \sum_{i=1}^k u_i f \left( \frac{a_i}{u_i} \right), \quad a \in \Delta_k; \quad \ell(a) = \infty, \quad a \notin \Delta_k
\]

with a given lower semicontinuous function \( f : [0, \infty) \to \mathbb{R} \cup \{\infty\} \) satisfying \( f(0) = \lim_{x \to 0} f(x) \), \( f(1) = 0 \), and being \( \kappa_a \)-strongly convex on \([0, \frac{1}{\min_{i \in \mathbb{N}} u_i}]\) with some \( \kappa_a > 0 \). As shown in [12, Example 2.2], \( \ell \) satisfies (H.1) with \( \nu = \frac{\kappa_a}{\max_{i \in \mathbb{N}} u_i} > 0 \).

Note that an \( f \)-divergence \( \ell \) is typically non-differentiable and may have non-closed effective domain \( A \) (see [12] for concrete examples). For commonly used forms of \( f \)-divergence, the proximal map \( \text{prox}_{\ell} \) can be computed by solving (1.8) with Lagrange multipliers. For instance, let \( \ell \) be the relative entropy corresponding to \( f(s) = s \log s, \ s \in \mathbb{R} \). Then for each \( \tau > 0 \) and \( a = (a_i)_{i=1}^k \in \mathbb{R}^k \), \( \text{prox}_{\tau \ell}(a)_i = \tau W \left( \frac{\mu_i}{\tau} \exp \left( \frac{\lambda_i + a_i}{\tau} \right) - 1 \right) \) 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 \( \sum_{i=1}^k \tau W \left( \frac{\mu_i}{\tau} \exp \left( \frac{\lambda + a_i}{\tau} \right) - 1 \right) = 1 \).

In the sequel, 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 \( A \), 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 $\phi \in \mathcal{B}([0, T] \times \mathbb{R}^n; \mathbb{R}^k)$,

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

We define the following space of feedback controls:

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

(2.11)

and for each $\phi \in \mathcal{V}_A$, define the associated control process $\alpha^\phi \in \mathcal{H}^2(\mathbb{R}^k)$ by

$$\alpha^\phi_t = \phi_t(X_t^{\xi_0, \phi}),$$

where $X_t^{\xi_0, \phi} \in \mathcal{S}^2(\mathbb{R}^n)$ is the solution to the following SDE (cf. (1.1)):

$$dX_t = b_t(X_t, \phi_t(X_t)) \, dt + \sigma_t(X_t) \, dW_t, \quad t \in [0, T]; \quad X_0 = \xi_0.$$ 

(2.12)

Under (H.1), the iterative scheme (1.9) is well-defined for any given guess $\phi^0 \in \mathcal{V}_A$ 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., $b_t$ 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.

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

The main contribution of this article is to identify conditions under which $(\phi^m)_{m \in \mathbb{N}_0} \subset \mathcal{V}_A$ 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 [29]. By [29, Proposition 1.114], every local minimizer $\alpha^* \in \text{dom } J(\cdot; \xi_0)$ 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 [24].

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 $x^* \in \text{dom } F = \{x \in X \mid F(x) < \infty\}$. The Fréchet subdifferential of $F$ at $x^*$ is defined by

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

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

As alluded to earlier, the map $\mathcal{V}_A \ni \phi \mapsto J(\alpha^\phi; \xi_0) \in \mathbb{R} \cup \{\infty\}$ is typically nonconvex and may not satisfy the Polyak-Lojasiewicz condition as in the setting with parametric policies ([9, 40, 28, 10, 13, 20]). 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 $\rho$ 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., $C_b$ is small.

(vi) State dynamics is strongly dissipative, i.e., $\kappa_b$ 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., [6, 4, 15, 18]). Condition (ii) also justifies the use of a fictitious discount factor to accelerate the convergence of PGMs for continuous-time control problems (see [11] 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 [31]). Condition (iii), along with Example 2.3, also justifies recent reinforcement learning heuristics that adding f-divergences, such as the relative entropy, to the optimization objective can accelerate the convergence of PGMs (see e.g., [36, 20]).

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 $b_t(x) = A_t x$ in (2.6), where the desired dissipativity can be ensured if eigenvalues of $A_t$ are sufficiently negative (see [15, 13]). Condition (vi) also motivates a residual correction method for solving nonlinear control problems. Consider a control problem (1.1)-(1.2) whose drift involves non-dissipative coefficient $\tilde{b}$. Then one can search feedback controls of the form $\phi = \tilde{\phi} + \hat{\phi}$, where $\tilde{\phi}$ is a precomputed candidate policy, and $\hat{\phi}$ is an unknown residual correction. Observe that the state dynamics now has the drift coefficient $b = (\tilde{b} + b\phi) + b\tilde{\phi}$, and the function $\tilde{b} + b\tilde{\phi}$ may be dissipative for suitably chosen policy $\tilde{\phi}$; see [34] and references therein for computing $\tilde{\phi}$ 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 $\phi^0 \in \mathcal{V}_A$ and $\tau \in (0, \frac{2}{\mu + L_f + \frac{1}{2}}]$, if one of conditions (i)-(vi) holds, then there exist $\phi^* \in \mathcal{V}_A$ and constants $c \in (0, 1)$ and $\overline{C} \geq 0$ such that

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

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

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, 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 $\xi_0$ and $\sigma$, 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 $(\|\tilde{X}^{\phi^m}\|_{\mathcal{H}^2})_{m \in \mathbb{N}_0}$ 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 [24, 36, 22].

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

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

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 \( \sup_{m,t,x} |Y_{t,x,\phi^m}| < \infty \), while Lipschitz estimates of (1.10) and (1.11) imply that \( x \mapsto Y_{t,x,\phi^m} \) is Lipschitz continuous uniformly in \( t \), where the Lipschitz constant \( L_Y([\phi^m]) \) depends exponentially on \( [\phi^m] \) due to the feedback controlled dynamics (1.10) (see Proposition 3.7). Combining these estimates gives \( |\phi^{m+1}|_1 \leq (1 - \tau C)|\phi^m|_1 + \tau C(L_Y([\phi^m])+1) \). We then show in Theorem 3.8 that under suitable conditions on the coefficients, such an exponential dependence can be controlled, and 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 \( \tau > 0 \),

\[
|\phi^{m+1} - \phi^m|_1 \leq (1 - \tau C)|\phi^m - \phi^m|_1 + \tau C \sup_{t,x} \frac{|Y_{t,x,\phi^m} - Y_{t,x,\phi^{m-1}}|}{1 + |x|}, \quad \forall m \in \mathbb{N}. \tag{2.13}
\]

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 \( |Y_{t,x,\phi^m} - Y_{t,x,\phi^{m-1}}| \leq \tilde{C}(1 + |x|)|\phi^m - \phi^{m-1}|_1 \), for some constant \( \tilde{C} \) independent of \( t, x, m \) (see Proposition 3.9). By quantifying \( C \) in (2.13) and \( \tilde{C} \) precisely, we prove under each of the conditions (i)-(vi) that there exists \( c \in [0,1) \) such that \( |\phi^{m+1} - \phi^m|_1 \leq c|\phi^m - \phi^{m-1}|_1 \) for all \( m \), which subsequently implies the convergence of \( (\phi^m)_{m \in \mathbb{N}} \) due to Banach’s fixed point theorem. Finally, we identify the limit of \( (\phi^m)_{m \in \mathbb{N}} \) as 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 proximal map of \( \ell \) (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 \( Y_{t,x,\phi} \), but similar analysis can be performed for (1.9) with inexact computation of the proximal map \( \text{prox}_{\tau \ell} \).

More precisely, let \( \phi^0 \in \mathcal{V}_A \) be an initial guess and \( \tau > 0 \) be a stepsize. For each \( m \in \mathbb{N}_0 \), given the feedback control \( \phi^m \) at the \( m \)-th iteration (with \( \phi^0 = \phi^0 \)), we consider the corresponding \( (Y_{t,x,\phi^m})_{(t,x) \in [0,T] \times \mathbb{R}^n} \) defined by (1.11) and a function \( \tilde{Y}_{\phi^m} : [0,T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n \) approximating the solution map \( [0,T] \times \mathbb{R}^n \ni (t,x) \mapsto Y_{t,x,\phi^m}(x) = Y_{t,x,\phi^m} \in \mathbb{R}^n \). The feedback control for the next iteration is then obtained via a proximal gradient update (1.9) based on \( \tilde{Y}_{\phi^m} \):

\[
\phi_{t+1}^m(x) = \text{prox}_{\tau \ell}(\phi_t^m(x) - \tau \partial \phi_t^m(x, \phi_t^m(x), \tilde{Y}_{\phi^m}(x))), \quad (t, x) \in [0,T] \times \mathbb{R}^n. \tag{2.14}
\]

The following theorem shows the accuracy of (2.14), whose precise statement and proof will be given in Section 3.5 (see Theorem 3.14). Here we assume that \( \tilde{Y}_{\phi^m} \) approximates the function \( Y_{\phi^m} \) well enough such that the resulting feedback controls \( \phi^m \) are uniformly bounded in time and
uniformly Lipschitz in space. This is a reasonable assumption as the exact iteration enjoys these properties as shown in Theorem 3.11, and any reasonable approximation $\tilde{Y}_{\phi}^{m}$ of $Y_{\phi}^{m}$ should retain these properties; see e.g., [3] for approximation schemes that preserve boundedness and Lipschitz continuity of exact solutions.

**Theorem 2.3.** Suppose (H.1) holds. For all $\phi^{0} \in V_{A}$ and $\tau \in (0, \frac{2}{p+L_{fa}} \wedge \frac{1}{p}]$, if $\sup_{m \in \mathbb{N}}(|\tilde{\phi}^{m}|_{0} + [\tilde{\phi}^{m}]_{1}) < \infty$, and one of the conditions (i)-(vi) holds, then there exist constants $c \in [0, 1)$ and $C \geq 0$ such that for all $m \in \mathbb{N}$,

$$|\tilde{\phi}^{m} - \phi^{*}|_{0} \leq c^{m}|\phi^{0} - \phi^{*}|_{0} + C \sum_{j=0}^{m-1} c^{m-1-j}|Y_{\phi}^{j} - \tilde{Y}_{\phi}^{j}|_{0},$$

where $\phi^{*} \in V_{A}$ is the limit function in Theorem 2.2.

## 3 Proofs

Throughout the rest of this work, we establish estimates with explicit dependence on the constants $T, \rho, C_{f}, L_{fa}, \mu, \nu, C_{g}, L_{g}, \kappa_{b}, C_{\bar{b}}$, which are important for the convergence of (1.9). For notational simplicity, we 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 $\phi \in V_{A}$.

### 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.

**Lemma 3.1.** Let $T > 0$, and for each $i = 1, 2$, let $\mu_{i} \in \mathbb{R}, \nu_{i} \geq 0$, let $b^{i} : [0, T] \times \mathbb{R}^{n} \rightarrow \mathbb{R}^{k}$ and $\sigma^{i} : [0, T] \times \mathbb{R}^{n} \rightarrow \mathbb{R}^{n \times d}$ be measurable functions such that for all $t \in [0, T]$ and $x, x' \in \mathbb{R}^{n}$,

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^{n}} \frac{|b^{i}(x)+\sigma^{i}(x)|}{1+|x|} < \infty, \quad (x-x', b^{i}(x)-b^{i}(x')) \leq \mu_{i}|x-x'|^{2}, \quad \text{and} \quad |\sigma^{i}(x)-\sigma^{i}(x')| \leq \nu_{i}|x-x'|,$$

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

$$dX_{s} = b^{i}(X_{s}) \, ds + \sigma^{i}(X_{s}) \, dW_{s}, \quad s \in [t, T]; \quad X_{t} = x. \tag{3.1}$$

Then for all $p \geq 2$ there exists an absolute constant $C(p)$ such that for all $t \in [0, T]$, $x_{1}, x_{2} \in \mathbb{R}^{n}$,

$$\|X^{t, x_{1}, 1} - X^{t, x_{2}, 2}\|_{\mathcal{S}^{p}} \leq C(p) e^{T(2\mu_{1} + C(p)\nu^{2})} \left( |x_{1} - x_{2}|^{2} + \sqrt{T} \|b^{1}(X^{t, x, 2}) - b^{2}(X^{t, x, 2})\|_{\mathcal{H}^{p}} + \|\sigma^{1}(X^{t, x, 2}) - \sigma^{2}(X^{t, x, 2})\|_{\mathcal{H}^{p}} \right).$$

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

$$\mathbb{E}\left[|X^{t,x,1}_{T} - X^{t,x,2}_{T}|^{2}\right] \leq \mathbb{E}\left[\int_{t}^{T} |b^{1}(X^{t, x, 2}) - b^{2}(X^{t, x, 2})|^{2} e^{(T-s)(2\mu_{1} + \nu^{2} + 1)} \, ds \right]. \tag{3.2}$$

The following lemma establishes stability of BSDEs with monotone nonlinearity. It has been proved in [32] for $p = 1$ and in [7, Proposition 3.2] for $p > 1$.  

10
Lemma 3.3. Suppose (H.1) holds, and let $H$ be defined by (1.4). Then for all $t \in [0, T]$, $x, x' \in \mathbb{R}^n$, $a, a' \in A$, $y, y' \in \mathbb{R}^n$ and $z, z' \in \mathbb{R}^{n \times d}$,

$$
\langle y - y', \partial_x H_t(x, a, y, z) - \partial_x H_t(x, a, y', z) \rangle \leq (\kappa_b - \rho + L_b)|y - y'|^2,
$$

(3.3)

$$
|\partial_x H_t(x, a, y, z) - \partial_x H_t(x', a', y', z')| \\
\leq (L_b|x - x'| + L_b(|x - x'| + |a - a'|)|y| + L_a|x - x'||z| \\
+ L_a|z - z'| + L_{fa}(|x - x'| + |a - a'|)).
$$

(3.4)

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

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

$$
|\text{prox}_{\tau\ell}(a - \tau \partial_a H_t^\text{re}(x, a, y)) - \text{prox}_{\tau\ell}(a' - \tau \partial_a H_t^\text{re}(x', a', y'))| \\
\leq \left(1 - \frac{1}{2}\left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right)|a - a'| + \tau C_b|y - y'| + \tau (L_b|y'| + L_{fa})|x - x'|.
$$

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 [35],

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

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

$$
|\text{prox}_{\tau\ell}(a - \tau \partial_a H_t^\text{re}(x, a, y)) - \text{prox}_{\tau\ell}(a' - \tau \partial_a H_t^\text{re}(x', a', y'))| \\
\leq \frac{1}{1 + \nu\tau}|(a - \tau \partial_a H_t^\text{re}(x, a, y)) - (a' - \tau \partial_a H_t^\text{re}(x', a', y'))| \\
\leq \frac{1}{1 + \nu\tau}|(a - \tau \partial_a H_t^\text{re}(x, a, y)) - (a' - \tau \partial_a H_t^\text{re}(x', a', y'))| \\
+ \frac{\tau}{1 + \nu\tau}|\partial_a H_t^\text{re}(x', a', y') - \partial_a H_t^\text{re}(x', a', y')|.
$$

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

\[
|\partial_a H^{re}_t(x, a', y) - \partial_a H^{re}_t(x', a', y')| \leq |\tilde{b}_t(x)^\top y - \tilde{b}_t(x')^\top y'| + |\partial_a f_t(x, a') - \partial_a f_t(x', a')| \\
\leq C_b|y - y'| + (L_b|y'| + L_{fa})|x - x'|.
\]

(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), \( \mathbb{A} \ni a \mapsto H^{re}_t(x, a, y) \in \mathbb{R} \) is \( \mu \)-strongly convex, and \( \mathbb{A} \ni a \mapsto \partial_a H^{re}_t(x, a, y) \in \mathbb{R}^k \) is \( L_{fa} \)-Lipschitz continuous, which along with [31, Theorem 2.1.12], implies for all \( a, a' \in \mathbb{A} \),

\[
\langle \partial_a H^{re}_t(x, a, y) - \partial_a H^{re}_t(x, a', y), a - a' \rangle \\
\geq \frac{\mu_{L_{fa}}}{\mu + L_{fa}} |a - a'|^2 + \frac{1}{\mu + L_{fa}} |\partial_a H^{re}_t(x, a, y) - \partial_a H^{re}_t(x, a', y)|^2.
\]

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

\[
|(a - \tau \partial_a H^{re}_t(x, a, y)) - (a' - \tau \partial_a H^{re}_t(x, a', y))|^2 \\
= |a - a'|^2 - 2\tau(a - a', \partial_a H^{re}_t(x, a, y) - \partial_a H^{re}_t(x, a', y)) + \tau^2 |\partial_a H^{re}_t(x, a, y) - \partial_a H^{re}_t(x, a', y)|^2 \\
\leq \left(1 - 2\tau \frac{\mu_{L_{fa}}}{\mu + L_{fa}}\right) |a - a'|^2 + \tau^2 |\partial_a H^{re}_t(x, a, y) - \partial_a H^{re}_t(x, a', y)|^2 \\
\leq \left(1 - 2\tau \frac{\mu_{L_{fa}}}{\mu + L_{fa}}\right) |a - a'|^2.
\]

Taking the square root of both sides of the above estimate and using the inequality \( \sqrt{1 - a\tau} \leq 1 - a\tau/2 \) for all \( a, \tau \geq 0 \) and \( a\tau \leq 1 \) give that

\[
|(a - \tau \partial_a H^{re}_t(x, a, y)) - (a' - \tau \partial_a H^{re}_t(x, a', y))| \leq \left(1 - \tau \frac{\mu_{L_{fa}}}{\mu + L_{fa}}\right) |a - a'|.
\]

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

\[
|\\text{prox}_{\tau f}(a - \tau \partial_a H^{re}_t(x, a, y)) - \\text{prox}_{\tau f}(a' - \tau \partial_a H^{re}_t(x', a', y'))| \\
\leq \frac{1}{1 + \tau
u \left(1 - \tau \frac{\mu_{L_{fa}}}{\mu + L_{fa}}\right) |a - a'| + \tau C_b|y - y'| + \tau (L_b|y'| + L_{fa})|x - x'|.
\]

Observe that for all \( a, b, \tau \geq 0 \) with \( 0 \leq \tau b \leq 1, 1 - \tau a \leq (1 + \tau b)(1 - \tau \frac{a + b}{2}) \). 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 \( \phi^{en}_t(0) \), we first prove the adjoint processes \((Y^{t,x,\phi}, Z^{t,x,\phi})\) defined in (1.11) have bounded \( p \)-th moments.

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

\[
\mathbb{E}\left[ \sup_{s \in [t, T]} e^{\tilde{a}s} |Y^{t,x,\phi}_s|^2 + \left( \int_t^T e^{\tilde{a}s} |Z^{t,x,\phi}_s|^2 \, ds \right)^p \right] \leq C_{(p)} \left( e^{\tilde{a}T} C_{g}^{2p} + C_{fa}^{2p} \left( \int_t^T e^{\tilde{a}s} \, ds \right)^{2p} \right),
\]

(3.7)
with \( \tilde{\alpha} = 2(\kappa_h - \rho + L_h + L_2^2) \). Consequently, there exists an absolute constant \( C \geq 0 \) such that for all \( \phi \in \mathcal{V}_A \) and \((t, x) \in [0, T] \times \mathbb{R}^n\),

\[
|Y_{t,x,\phi}^{t,x,\phi}| \leq C_Y := C(C_g + C_{fx} T) e^{(\kappa_h - \rho + C) T}.
\] (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 \( X_s^{t,x,\phi} \) is defined by (1.10). Then by Lemma 3.3, \(|\bar{f}_s^1(\omega, y, z) - \bar{f}_s^1(\omega, y, z')| \leq L_{\sigma}|z - z'|\), and

\[
(y - y', \bar{f}_s^1(\omega, y, z) - \bar{f}_s^1(\omega, y', z)) \leq (\kappa_h - \rho + L_h)|y - y'|^2.
\] (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 \( \alpha = 2(\kappa_h - \rho + L_h + L_2^2) \), it holds with some constant \( C \geq 0 \) that, for all \( p \geq 1 \),

\[
\mathbb{E} \left[ \sup_{s \in [t,T]} e^{\alpha s} |Y_s^{t,x,\phi}|^p + \left( \int_t^T e^{\alpha s} |Y_s^{t,x,\phi}|^2 \, ds \right)^{p/2} \right] \leq C(\rho) \mathbb{E} \left[ e^{\alpha T} |\partial_x g(X_{T}^{t,x,\phi})|^p + \left( \int_t^T e^{\alpha s} |\partial_x g(X_{s}^{t,x,\phi}, \partial_x \phi(X_{s}^{t,x,\phi}))|^2 \, ds \right)^{p/2} \right] \leq C(\rho) \left( e^{\alpha T} C_g^2 + C_{fx}^2 \left( \int_t^T e^{\alpha s} \, ds \right)^{2p/\alpha} \right),
\]

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 \((t, x) \in [0, T] \times \mathbb{R}^n\),

\[
|Y_{t,x,\phi}^{t,x,\phi}| \leq C e^{-\frac{\alpha t}{2}} \left( e^{\alpha T} C_g + C_{fx} \int_t^T e^{\alpha s} \, ds \right) \leq C(C_g + C_{fx} T)e^{(\kappa_h - \rho + C)T}.
\] (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 \( a_0 \in A \) and \( z^{a_0} \in \mathbb{R}^k \) such that \( z^{a_0} \in \partial^\ell \ell(a_0) \). For each \( \phi^0 \in \mathcal{V}_A \), \( \tau > 0 \) and \( m \in \mathbb{N} \), let \( \phi_t^m \) be defined by (1.9). Then for all \( \phi^0 \in \mathcal{V}_A \) and \( \tau \in (0, \frac{2}{\mu + L_{fa} + \frac{1}{\mu}} \),

\[
\sup_{m \in \mathbb{N}, t \in [0,T]} |\phi_t^m(0)| \leq C(\phi^0) := \sup_{t \in [0,T]} |\phi_t^0(0)| + 2 \left( \frac{\mu + L_{fa}}{\mu + \nu} C_Y + \frac{2}{\mu + \nu} (C_{fa} + L_{fa} |a_0| + |z^{a_0}|) + |a_0| \right) + 4C_Y C_b \left( \frac{\mu + L_{fa}}{\mu + \nu} \right),
\] (3.11)

where 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 \( \phi^*_t[u] = \arg \min_{a \in \mathbb{R}^k} \langle H_t^e(0, a, u), \ell(a) \rangle \), with \( H^e \) defined as in (1.3). By (1.3) and (2.6),

\[
\phi_t^*[u] = \arg \min_{a \in \mathbb{R}^k} \langle \tilde{b}_t(0), u \rangle + h_t(0, a) = \partial_x \phi_t^* \left( 0, -\tilde{b}_t^*(0)u \right),
\] (3.12)
where $h^* : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^k$ is the convex conjugate function of $h$ defined by
\[
h^*_t(x, z) := \sup \{ \langle a, z \rangle - h_t(x, a) \mid a \in \mathbb{R}^k \}.
\] (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 $\partial_z h^*_t(x, z) = \arg \max_{a \in \mathbb{R}^k} \langle a, z \rangle - h_t(x, a)$.

Let $a_0 \in \text{dom} \ell$ and $z^{a_0} \in \mathbb{R}^k$ such that $z^{a_0} \in \partial^{\ell} f(a)$ $\neq \emptyset$. Then for all $t \in [0, T)$, by the differentiability and convexity of $a \mapsto f_t(0, a)$, $\partial_a f_t(0, a_0) + z^{a_0} \in \partial^h h_t(0, a_0)$ (see [35, 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 $a \mapsto h_t(0, a)$,

$$h_t(0, a_0) \geq h_t(0, \partial_z h^*_t(0, 0)) \geq h_t(0, a_0) + \langle \partial_a f_t(0, a_0), z^{a_0} \rangle, \quad \partial_z h^*_t(0, 0) - a_0 \rangle \geq 0, $$

which implies that
\[
|\partial_z h^*_t(0, 0) - a_0| \leq \frac{2}{\mu + \nu} |\partial_a f_t(0, a_0) + z^{a_0}| \leq \frac{2}{\mu + \nu} (C_f a + L_f a_0 | + |z^{a_0}|),
\]
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 $u \in \mathbb{R}^n$,
\[
|\phi_t^u| \leq |\partial_z h^*_t(0, -b^T(t) u) - \partial_z h^*_t(0, 0)| + |\partial_z h^*_t(0, 0)| \leq \frac{1}{\mu + \nu} C_B |u| + \frac{2}{\mu + \nu} (C_f a + L_f a_0 | + |z^{a_0}|) + |a_0|.
\] (3.14)

Let $\phi^0 \in \mathcal{V}_A$ and $\tau \in (0, \frac{2}{\mu + L_f a} \wedge \frac{1}{\nu})$ be fixed in the subsequent analysis. For each $t \in [0, T]$, let $c^0_t := \phi^0_t Y_t^{t, 0, \phi^0}$. Then for all $t \in [0, T]$, the fact that $c^0_t = \arg \min_{a \in \mathbb{R}^k} (H^r(t, 0, \phi^0) + \ell(a))$ implies that $c^0_t = \text{prox}_{\tau \ell}(c^0_t - \tau \partial_a H^r_t(0, c^0_t, Y_t^{t, 0, \phi^0})))$. Then for all $m \in \mathbb{N}_0$ and $t \in [0, T]$, (1.9) and Lemma 3.4 imply that
\[
|\phi_t^{m+1}(0) - c^0_t| = |\text{prox}_{\tau \ell}(\phi_t^m(0) - \tau \partial_a H^r_t(0, \phi^m_t, 0, Y_t^{t, 0, \phi^m})) - \text{prox}_{\tau \ell}(c^0_t - \tau \partial_a H^r_t(0, c^0_t, Y_t^{t, 0, \phi^0})))| \leq \left(1 - \frac{1}{2} \left(\frac{\mu L_f a + \nu}{\mu + L_f a} \right)\right) |\phi_t^m(0) - c^0_t| + \tau C_B |Y_t^{t, 0, \phi^m} - Y_t^{t, 0, \phi^0}|
\]
By Proposition 3.5, there exists an absolute constant $C \geq 0$ such that for all $t \in [0, T]$ and $\phi \in \mathcal{V}_A$, $|Y_t^{t, 0, \phi}| \leq C_Y := C(C_G + C_{fz} T)(e^\alpha + T)$, with $\alpha = \kappa_b + \rho \sigma + L_b + \sigma^2$, which implies that for all $m \in \mathbb{N}_0$,
\[
|\phi_t^{m}(0) - c^0_t| \leq |\phi_t^0(0) - c^0_t| + 2C_B \frac{\mu + L_f a}{(\mu + \nu) L_f a + \mu \nu} \sup_{m \in \mathbb{N}_0} |Y_t^{t, 0, \phi^m} - Y_t^{t, 0, \phi^0}|
\]
\[
\leq |\phi_t^0(0) - c^0_t| + 4C Y C_B \frac{\mu + L_f a}{(\mu + \nu) L_f a + \mu \nu}.
\]
By (3.14), for all $t \in [0, T]$, $|c^0_t| \leq \mu + \nu \frac{C_B C_Y + \frac{2}{\mu + \nu} (C_f a + L_f a_0 | + |z^{a_0}|) + |a_0|}$, from which for all $m \in \mathbb{N}_0$ and $t \in [0, T]$,
\[
|\phi_t^{m}(0)| \leq |\phi_t^0(0)| + 2 |c^0_t| + 4 C Y C_B \frac{\mu + L_f a}{(\mu + \nu) L_f a + \mu \nu}
\]
\[
\leq \sup_{t \in [0, T]} |\phi_t^0(0)| + 2 \left(\frac{\mu + \nu \frac{C_B C_Y + \frac{2}{\mu + \nu} (C_f a + L_f a_0 | + |z^{a_0}|) + |a_0|)}{\mu + \nu} \right) + 4 C Y C_B \frac{\mu + L_f a}{(\mu + \nu) L_f a + \mu \nu}.
\]
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 \( (\hat{\phi}^m)_{m \in \mathbb{N}_0} \) satisfy \( \sup_{m \in \mathbb{N}_0}|\hat{\phi}^m| < \infty \) if one of the conditions (i)-(vi) holds. The following proposition estimates the Lipschitz continuity of the function \( x \mapsto Y^t,x,\phi \) in terms of the Lipschitz continuity of a given feedback control \( \phi \in \mathcal{V}_A \).

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

\[
|Y_1^{t,x,\phi} - Y_1^{t,x',\phi}| \leq L_Y(\|\phi\|) |x - x'|,
\]

where for each \( M \geq 0 \), the constant \( L_Y(M) \geq 0 \) is defined by

\[
L_Y(M) := C \left[ L_g e^{(2L_bM + 2\kappa_1 + C) T} e^{\alpha T} + \left( (1 + L_hM) (C_g + C_f T) e^{(\kappa_1 - \rho + C) T} + L_f e^{(\alpha - 1) T} + \sqrt{T} \left(e^{\alpha T} C_g + C_f e^{(\alpha - 1) T}\right) \right) e^{(2L_bM + 2\kappa_1 + C) T} \right],
\]

\[
\alpha := \kappa_1 - \rho + L_b + L_2^2.
\]

**Proof.** Let \( f_1, f_2 : [t, T] \times \mathbb{R}^n \times \mathbb{R}^{n \times d} \to \mathbb{R}^n \) be such that for all \((s, \omega, y, z) \in [t, T] \times \mathbb{R}^n \times \mathbb{R}^{n \times d}\),

\[
\begin{align*}
\hat{f}_s = \partial_s H_s(X_s^{t,x,\phi}(\omega), \phi_s(X_s^{t,x,\phi}(\omega)), y, z), \\
\tilde{f}_s = \partial_s H_s(X_s^{t,x,\phi}(\omega), \phi_s(X_s^{t,x,\phi}(\omega)), y, z),
\end{align*}
\]

where \( H \) is defined by (1.4), and for each \( \phi \in \mathcal{V}_A \), \( X^{t,x,\phi} \in S^2(t, T; \mathbb{R}^n) \) is defined by (1.10). By using (3.9) and applying Lemma 3.2 with \( p = 1 \), \( f^1 = \hat{f}_1 \), \( \xi^1 = \partial_x g(X_t^{t,x,\phi}) \), \( f^2 = \tilde{f}_2 \), \( \xi^2 = \partial_x g(X_t^{t,x,\phi}) \), \( (Y^2, Z^2) = (Y^{t,x,\phi}, Z^{t,x,\phi}) \) and \( \varepsilon = 1/2 \), there exists an absolute constant \( C \geq 0 \), such that

\[
\mathbb{E} \left[ \sup_{s \in [t, T]} e^{\hat{\alpha} s} \left| Y^t_s^{t,x,\phi} - Y^{t,x',\phi}_s \right|^2 + \int_t^T e^{\hat{\alpha} T} \left| Z^{t,x,\phi}_s - Z^{t,x',\phi}_s \right|^2 ds \right] 
\]

\[
\leq C \left[ e^{\hat{\alpha} T} \left| \partial_x g(X_t^{t,x,\phi}) - \partial_x g(X_t^{t,x',\phi}) \right|^2 \right. 
\]

\[
+ \left. \left( \int_t^T e^{\hat{\alpha} T} \left| f_1^s(\cdot, Y^t_s^{t,x,\phi}, Z^{t,x,\phi}_s) - f_2^s(\cdot, Y^t_s^{t,x,\phi}, Z^{t,x',\phi}_s) \right| ds \right)^2 \right].
\]

where we defined \( \hat{\alpha} := 2(\kappa_1 - \rho + L_b + L_2^2) \) above and hereafter. Recall that in the subsequent analysis, \( C \) denotes a generic constant independent of \( T, \rho, \kappa_1, C_b, C_f, L_f, \mu, \nu, L_f, C_g, C_l \).

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 \( \Delta X^{t,x,\phi} = X^{t,x,\phi} - X^{t,x',\phi} \).

Then by (2.8) and (2.9), for all \( t \in [0, T] \) and \( x, x' \in \mathbb{R}^n \),

\[
\left( x - x', b^1_t(x) - b^2_t(x') \right) \leq \left( x - x', \hat{b}_t(x) - \tilde{b}_t(x') + \hat{b}_t(x') - \hat{b}_t(x) \phi_t(x) - \hat{b}_t(x') \phi_t(x') \right) 
\]

\[
\leq \kappa_b |x - x'|^2 + L_b |x - x'| (|x - x'| + |\phi_t(x) - \phi_t(x')|) 
\]

\[
\leq \left( \kappa_b + L_b (1 + |\phi|_1) \right) |x - x'|^2.
\]

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

\[
\|\Delta X^{t,x,\phi}\|_{L^p} \leq C(p) e^{T \left( 2(\kappa_1 + L_b (1 + |\phi|_1)) + C_p L_2^2 \right) + |x - x'|},
\]

(3.18)
from which by setting \( p = 2 \) and using (2.5), we obtain

\[
e^{-\frac{\hat{a}}{4} t} E \left[ e^{\tilde{\alpha} t} |\partial_x g(X_T^t,X^t_{x},\phi) - \partial_x g(X_T^{t,x'},\phi)|^2 \right]^{1/2} \leq C L_g e^{(T - 0) \frac{\hat{a}}{4} + T (2 (\kappa_0 + L_0 [\phi_1]) + C) + |x - x'|}. \tag{3.19}
\]

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

\[
|\partial_x H_t(x, \phi_t(x), y, z) - \partial_x H_t(x', \phi_t(x'))| \leq (L_0 |x - x'| + L_0 (|x - x'| + |\phi_t(x) - \phi_t(x')|)) |y| + L_0 |x - x'| |z| + L_{f_x} (|x - x'| + |\phi_t(x) - \phi_t(x')|)
\leq \left( (L_0 + L_0 (1 + |\phi_1|)) |y| + L_{f_x} (1 + |\phi_1| + L_0 |z|) \right) |x - x'|,
\]

which along with the Cauchy-Schwarz inequality implies that

\[
\mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} |f_s^t(\cdot, Y_s^t,x',\phi, Z_s^t,x',\phi) - f_s^t(\cdot, Y_s^t,x',\phi, Z_s^t,x',\phi)|^2 \, ds \right)^{1/2} \right]^{1/2} \leq \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} \left( (C + L_0 [\phi_1]) |Y_s^t,x',\phi| + L_{f_x} (1 + |\phi_1|) + C |Z_s^t,x',\phi| \right) |\Delta X_s^t,x' \, ds \right)^{2} \right]^{1/2} \]

\[
\leq \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} \left( (C + L_0 [\phi_1]) |Y_s^t,x',\phi| + L_{f_x} (1 + |\phi_1|) + C |Z_s^t,x',\phi| \right) \, ds \right)^{4} \right]^{1/4} \| \Delta X_s^t,x' \|_{S^4} \tag{3.20}
\]

\[
\leq \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} |Z_s^t,x',\phi| \, ds \right)^{4} \right]^{1/4} \| \Delta X_s^t,x' \|_{S^4}.
\]

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

\[
|Y_s^t,x',\phi| \leq C_Y := C (C_g + C_{f_x} T) e^{(\kappa_0 - \rho) + C) + T}. \tag{3.21}
\]

The Markovian property of \( Y^t,x',\phi \) implies that \( Y_s^t,x',\phi = Y_s^t,X_s^t,x',\phi \) (see e.g., [41, Theorem 5.1.3]), which subsequently shows that

\[
\mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} \left( (C + L_0 [\phi_1]) |Y_s^t,x',\phi| + L_{f_x} (1 + |\phi_1|) \right) \, ds \right)^{4} \right]^{1/4} \leq \left( (C + L_0 [\phi_1]) C_Y + L_{f_x} (1 + [\phi_1]) \right) \int_t^T e^{\tilde{\alpha} s} \, ds. \tag{3.22}
\]

On the other hand, by the Cauchy-Schwarz inequality and Proposition 3.5 with \( p = 2 \), there exists an absolute constant \( C \geq 0 \) such that

\[
\mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} |Z_s^t,x',\phi| \, ds \right)^{4} \right]^{1/4} \leq \sqrt{T - t} \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha} s} |Z_s^t,x',\phi|^2 \, ds \right)^{2} \right]^{1/4} \leq C \sqrt{T - t} \left( e^{\tilde{\alpha} T} C_g + C_{f_x} \int_t^T e^{\tilde{\alpha} s} \, ds \right). \tag{3.23}
\]
Combining (3.18) (with \(p = 4\)), (3.20), (3.22), and (3.23) gives that
\[
\mathbb{E}\left[\mu_2(\int_t^T e^{\frac{\hat{\kappa}}{2}s} |f^1_s(\cdot, Y_s, z_s, z_{s}', \phi) - f^2_s(\cdot, Y_s, z_s, z_{s}', \phi)| ds)^2\right]^{1/2} \\
\leq C \left[ \left( 1 + \Delta \right) C_\phi + \int_t^T e^{\frac{\hat{\kappa}}{2}s} ds \right] \\
+ \sqrt{T} \left( e^{\frac{\hat{\kappa}}{2}T} C_\phi + \int_t^T e^{\frac{\hat{\kappa}}{2}s} ds \right) e^{T(2\hat{\kappa} + \Delta) + |x - x'|},
\]
with \(C_\phi\) defined in (3.21). Consequently, by using (3.17), and (3.19) and the identity that
\[e^{-\frac{\hat{\kappa}}{2}t} \int_t^T e^{\frac{\hat{\kappa}}{2}s} ds = \frac{2}{\hat{\kappa}}(e^{\frac{\hat{\kappa}}{2}(T-t)} - 1),\]
\[|Y^t,x,\phi - Y^t,x',\phi| \leq C \left[ L_\phi e^{(T-t)\frac{\hat{\kappa}}{2} + T(2\hat{\kappa} + \Delta) + |x - x'|} \right] \]
\[\times \left( (1 + L_\phi)(C_\phi + \int_t^T e^{(\hat{\kappa} - \rho + C) + T} + L_\phi(1 + [\phi]_1)) \frac{2}{\hat{\kappa}}(e^{\frac{\hat{\kappa}}{2}(T-t)} - 1) \right) \]
\[+ \sqrt{T} \left( e^{\frac{\hat{\kappa}}{2}(T-t)} C_\phi + \int_t^T e^{\frac{\hat{\kappa}}{2}s} ds \right) e^{T(2\hat{\kappa} + \Delta) + |x - x'|}.
\]
Then the facts that for all \(\alpha \in \mathbb{R}\), \([0, T] \ni t \mapsto \frac{2}{\alpha}(e^{\frac{\alpha}{2}(T-t)} - 1)\) is maximized at \(t = 0\), and \(\alpha = 2(\hat{\kappa} - \rho + L_\phi + L_\phi^2)\) 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 \(\phi^0 \in \mathcal{V}_A\), the sequence of feedback controls \((\phi^m)_{m \in \mathbb{N}}\) 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, \(C_\phi\) and \(\alpha\) be defined in (3.8) and (3.16), respectively, and for each \(\phi^0 \in \mathcal{V}_A\), define
\[A_1 := C_\phi C \left( L_\phi + \sqrt{T} C_\phi e^{\alpha + T} + \frac{e^{\alpha T - 1}}{\alpha} \left( (C_\phi + L_\phi T)e^{\alpha T} + L_\phi + \sqrt{T} C_\phi \right) \right) \]
\[A_2 := C_\phi \frac{e^{\alpha T - 1}}{\alpha} \left( (C_\phi + L_\phi T)e^{\alpha T} L_\phi + L_\phi \right), \quad \mu_0 := \frac{1}{4} \left( \frac{\mu L_\phi}{\mu + L_\phi} + \nu \right), \quad K := \max \left\{ \frac{2TL_\phi[\phi^0]_1 + 2TL_\phi A_1 + 2TL_\phi (L_\phi C_\phi + L_\phi)}{\mu_0} + 1 \right\}. \quad \text{(3.24)}
\]

**Theorem 3.8.** Suppose (H.1) holds. For each \(\phi^0 \in \mathcal{V}_A\), \(\tau > 0\) and \(m \in \mathbb{N}\), let \(\phi^m\) be defined by (1.9) with the initial guess \(\phi^0\) and stepsize \(\tau\). Let \(C_\phi \geq 0\) be a constant such that (3.8) and (3.15) hold, let \(C_\phi \geq 0\) be defined in (3.8), let \(\alpha \in \mathbb{R}\) be defined in (3.16), and let \(A_1, A_2, \mu_0, K \geq 0\) be defined in (3.24). Then for all \(\phi^0 \in \mathcal{V}_A\) satisfying
\[2TL_\phi A_1 e^{(2\hat{\kappa} + C) + K} \leq \mu_0, \quad \text{and} \quad A_2 e^{(2\hat{\kappa} + C) + K + 1} \leq \mu_0, \quad \text{(3.25)}
\]
and for all \(\tau \in (0, \frac{2}{\mu + L_\phi} \wedge \frac{1}{\mu})\) and \(m \in \mathbb{N}_0\),
\[\left[\phi^m\right]_1 \leq L(\phi^0) := \left[\phi^0\right]_1 + \frac{1}{\mu_0} \left( L_\phi C_\phi + L_\phi + A_1 (e^{(2\hat{\kappa} + C) + K + 1}) \right). \quad \text{(3.26)}
\]

**Proof.** Throughout this proof, let \(\phi^0 \in \mathcal{V}_A\) and \(\tau \in (0, \frac{2}{\mu + L_\phi} \wedge \frac{1}{\mu})\) be fixed. Suppose that \(\phi^m \in \mathcal{V}_A\) 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 \(a = \phi^m(x),\)
\( a' = \phi_t^{m}(x'), y = Y_t^{t,x,\phi^m} \) and \( y' = Y_t^{t,x',\phi^m} \),

\[
|\phi_t^{m+1}(x) - \phi_t^{m+1}(x')| \leq \left( 1 - \frac{1}{2} \left( \frac{\mu L_{fa}}{L_{fa} + \nu} \right) \right) |\phi_t^{m}(x) - \phi_t^{m}(x')| + \tau C_0 |Y_t^{t,x,\phi^m} - Y_t^{t,x',\phi^m} | + \tau (L_b |Y_t^{t,x',\phi^m} | + L_{fa}) |x - x'|.
\]

By using the Lipschitz continuity of \( \phi^m \), 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

\[
|\phi_t^{m+1}(x) - \phi_t^{m+1}(x')| \leq \left( [\phi^m] \left( 1 - \frac{1}{2} \left( \frac{\mu L_{fa}}{L_{fa} + \nu} \right) \right) + \tau C_0 Y (\beta) + \tau (L_b C_Y + L_{fa}) \right) |x - x'|,
\]

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

\[
[\phi^{m+1}]_1 \leq [\phi^m]_1 \left( 1 - \frac{1}{2} \left( \frac{\mu L_{fa}}{L_{fa} + \nu} \right) \right) + \tau (L_b C_Y + L_{fa}) + \tau C_b Y (\beta).
\]

In the sequel, 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

\[
L_Y([\phi^m]_1) := C \left[ L_g e^{2L_b[\phi^m]_1 + 2\kappa + C + T} e^{\alpha + T} \right. \\
+ L_{fa} \left. \left( 1 + [\phi^m]_1 \right) \frac{e^{\alpha + T}}{\alpha + T} + \sqrt{T} \left( e^{\alpha + T} C_g + C_{fx} \frac{e^{\alpha - T}}{\alpha - T} \right) \right] e^{2L_b[\phi^m]_1 + 2\kappa + C + T} \\
= C \left[ \left( L_g + \sqrt{T} C_g \right) e^{\alpha + T} + \left( C_g + C_{fx} \right) e^{\alpha + T} \right. \\
\times \left. \left( e^{2L_b[\phi^m]_1 + 2\kappa + C} + 1 \right) + \left( C_g + C_{fx} \right) e^{\alpha + T} \right] \frac{e^{\alpha + T}}{\alpha + T} \left[ \left( e^{2L_b[\phi^m]_1 + 2\kappa + C} T + 1 \right) \right] \\
+ \left. \left( C_g + C_{fx} \right) e^{\alpha + T} \right] \frac{e^{\alpha + T}}{\alpha + T} \left[ \left( e^{2L_b[\phi^m]_1 + 2\kappa + C} T + 1 \right) \right].
\]

Let \( A_1, A_2, \mu_0, K \) be defined as in (3.24). Then by writing \( \tilde{A}_1 := 2TL_k A_1 \) and \( [\tilde{\phi}^m]_1 := 2TL_k[\phi^m]_1 \) for all \( m \in \mathbb{N}_0 \), multiplying both sides of (3.27) by \( 2TL_k \) and using (3.28), we have

\[
[\tilde{\phi}^{m+1}]_1 \leq [\tilde{\phi}^m]_1 \left( 1 - 2\mu_0 \tau \right) + 2\tau TL_k (L_b C_Y + L_{fa}) + \tau \left( \tilde{A}_1 \left( e^{(2\kappa + C) T + [\tilde{\phi}^m]_1} + 1 \right) + A_2 [\tilde{\phi}^m]_1 \left( e^{(2\kappa + C) T + [\tilde{\phi}^m]_1} + 1 \right) \right).
\]

Now we prove by induction that \( \sup_{m \in \mathbb{N}_0} [\tilde{\phi}^m]_1 \leq K \) under the conditions that

\[
\tilde{A}_1 e^{(2\kappa + C) T + K} \leq \mu_0, \quad \text{and} \quad A_2 e^{(2\kappa + C) T + K + 1} \leq \mu_0.
\]

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

\[
[\tilde{\phi}^{m+1}]_1 \leq [\tilde{\phi}^m]_1 \left( 1 - 2\mu_0 \tau \right) + 2\tau TL_k (L_b C_Y + L_{fa}) + \tau \left( \tilde{A}_1 + \mu_0 + \mu_0 [\tilde{\phi}^m]_1 \right)
\]

\[
= [\tilde{\phi}^m]_1 \left( 1 - \mu_0 \tau \right) + \tau \left( 2TL_k (L_b C_Y + L_{fa}) + \tilde{A}_1 + \mu_0 \right),
\]

\[
\leq K \left( 1 - \mu_0 \tau \right) + \tau \mu_0 K \leq K.
\]
This finishes the proof of the fact $\sup_{m \in \mathbb{N}_0} [\tilde{m}]_1 \leq K$. Substituting this a-priori bound into (3.29) and using (3.30) give that

$$[\tilde{m}]_1 \leq [\tilde{m}]_1 (1 - 2\mu_0 \tau) + 2\tau TL_b(L_bC_Y + L_f) + \tau \left( A_1 \left( e^{(2\kappa_b+C)T+K} + 1 \right) + A_2 \left( e^{(2\kappa_b+C)T+K} + 1 \right) \right)$$

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

$$[\tilde{m}]_1 \leq \frac{1}{\mu_0} \left( 2TL_b(L_bC_Y + L_f) + A_1 \left( e^{(2\kappa_b+C)T+K} + 1 \right) \right).$$

Dividing both sides of the above inequality by $2TL_b$ shows

$$[\tilde{m}]_1 \leq \frac{1}{\mu_0} \left( L_bC_Y + L_f + A_1 \left( e^{(2\kappa_b+C)T+K} + 1 \right) \right).$$

with constants $A_1, K, \mu_0$ defined as in (3.24).

### 3.4 Contraction in a weighted sup-norm

Based on the uniform Lipschitz continuity of $(\tilde{m})_{m \in \mathbb{N}_0}$ in Theorem 3.8, we prove the contractivity of the iterates $(\tilde{m})_{m \in \mathbb{N}_0}$ with respect to the weighted sup-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 $\phi \in \mathcal{V}_A$.

**Proposition 3.9.** Suppose (H.1) holds. For each $\phi \in \mathcal{V}_A$ 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 $\phi' \in \mathcal{V}_A$ and $(t,s,x) \in [0,T] \times [t,T] \times \mathbb{R}^n$, it holds with some constant $C_Z^{\phi'} \geq 0$ that $|Z^{t,x,\phi'}_s| \leq C_Z^{\phi'}$ for $dt \otimes d\mathbb{P}$-a.e. Then there exists a constant $C_0 \geq 0$ such that for all $\phi, \phi' \in \mathcal{V}_A$ and $(t,x) \in [0,T] \times \mathbb{R}^n$,

$$|Y^{t,x,\phi}_t - Y^{t,x,\phi'}_t| \leq B[\phi, \phi', C_Z^{\phi'}](1 + |x|) |\phi - \phi'|_0, \quad (3.32)$$

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

$$B[\phi, \phi', C_Z^{\phi'}] := CC_b \left( 1 + T + TC_b \sup_{t \in [0,T]} |\phi'_t(0)| \right) e^{T\beta} \left( L_g m_{(\alpha,\beta)}^{1/2} + \frac{e^{T\frac{\alpha-1}{\alpha} \left( (C_Y + L_f) (1 + |\phi|_1) + C_Z^{\phi'} \right) T} e^{T\beta} + L_f + C_Y \right),$$

$$\beta := 2\kappa_b + 2L_b \max\{|[\phi]|_1, |[\phi']|_1\} + C, \quad m_{(\alpha,\beta)} := \sup_{t \in [0,T]} e^{2\alpha(T-t)} \int_t^T e^{(T-s)\beta} ds,$$

with $C_Y$ and $\alpha$ 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 \((s, \omega, y, z) \in [t, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d}, \)

\[
\begin{align*}
\bar{f}^1_s(\omega, y, z) = & \partial_x H_s(X^{t,x,\phi}_s(\omega), \phi_s(X^{t,x,\phi}_s(\omega)), y, z), \\
\bar{f}^2_s(\omega, y, z) = & \partial_x H_s(X^{t,x,\phi'}_s(\omega), \phi'_s(X^{t,x,\phi'}_s(\omega)), y, z),
\end{align*}
\]

where \( H \) is defined in (1.4), and for each \( \phi \in \mathcal{V}_A \) 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

\[
\begin{align*}
\mathbb{E} \left[ \sup_{s \in [t,T]} e^{\tilde{\alpha} s} |Y^t_s - Y^{t,x,\phi'}_s|^2 + \int_t^T e^{\tilde{\alpha} s} |Z^{t,x,\phi}_s - Z^{t,x,\phi'}_s|^2 \, ds \right] \\
\leq C \mathbb{E} \left[ e^{\tilde{\alpha} T} |\partial_x g(X^{t,x,\phi}_T) - \partial_x g(X^{t,x,\phi'}_T)|^2 \\
+ \left( \int_t^T e^{\tilde{\alpha} s} \left| \bar{f}^1_s(Y^{t,x,\phi}_s, Z^{t,x,\phi}_s) - \bar{f}^2_s(Y^{t,x,\phi'}_s, Z^{t,x,\phi'}_s) \right| \, ds \right)^2 \right],
\end{align*}
\]

(3.34)

where we defined \( \tilde{\alpha} := 2(\kappa_0 + \rho + L_0 + L_0^2) \) above and hereafter. In the subsequent analysis, we denote by \( C \geq 1 \) a generic constant independent of \( T, \rho, \kappa_0, C_{f_x}, L_{f_x}, \mu, \nu, L_{f_0}, C_g, L_g \).

To estimate the right-hand side of (3.34), we first quantify the dependence of \( X^{t,x,\phi} \) on \( \phi \). 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 \((t, x) \in [0, T] \times \mathbb{R}^n, \phi, \phi' \in \mathcal{V}_A \),

\[
\| \Delta X^{t,x} \|_{\mathcal{S}^2} \leq C_b \sqrt{T} M_{(\phi_1)} \| \phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'}) \|_{\mathcal{H}^2} \\
\leq C_b T M_{(\phi_1)} \| \phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'}) \|_{\mathcal{S}^2},
\]

(3.35)

where the constant \( M_{(\phi_1)} \) is defined by

\[
M_{(\phi_1)} := C e^{T \left( 2(\kappa_0 + L_0(1 + [\phi_1]) + C L_0^2) \right)} \leq C e^{2T \left( \kappa_0 + L_0 [\phi_1] + C \right)}. \tag{3.36}
\]

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 \( \|\phi'(0)\|_{\infty} = \sup_{0 \leq t \leq T} |\phi'_t(0)| \) 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 \( \sigma^2_t(x) = 0 \),

\[
\|X^{t,x,\phi'}\|_{\mathcal{S}^2} \leq M_{(\phi_1)} (|x| + \sqrt{T} \|b(0, \phi'(0))\|_{\mathcal{H}^2} + \|\sigma(0)\|_{\mathcal{H}^2}) \\
\leq C M_{(\phi_1)} (|x| + T + C_b \|\phi'(0)\|_{\infty} + \sqrt{T}) \tag{3.38}
\]

which along with (3.35), (3.37) and \( M_{(\phi_1)} \geq 1 \) shows

\[
\|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{\mathcal{S}^2} \leq C M_{(\phi_1)} \left( 1 + T + C_b \|\phi'(0)\|_{\infty} \right) |\phi - \phi'|_0,
\]

\[
\|\Delta X^{t,x}\|_{\mathcal{S}^2} \leq C C_b T M_{(\phi_1)} M_{(\phi_1)} \left( 1 + T + C_b \|\phi'(0)\|_{\infty} \right) (1 + |x|) |\phi - \phi'|_0. \tag{3.39}
\]

20
Similarly, by setting \( \tilde{\beta} = 2(\kappa_{\tilde{b}} + L_{\tilde{b}}(1 + [\phi]_1)) + L^2_{\alpha} + 1 \) and using (3.2) with \( b^1_t(x) = b_t(x, \phi_t(x)) \) and \( b^2_t(x) = b_t(x, \phi'_t(x)) \), we have

\[
\text{Eq. (3.40)}
\]

\[
\mathbb{E} \left[ |\Delta X^t_s|^2 \right]^{\frac{1}{2}} \leq CC_b \left( \int_t^T e^{(T-s)\tilde{\beta}} \right) \frac{1}{2} \|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{S^2} 
\]

\[
\leq CC_b \left( \int_t^T e^{(T-s)\beta} \right) M_{[\phi]_1}(1 + T + TC_b\|\phi'(0)\|_{\infty})(1 + |x|)\|\phi - \phi'\|_0,
\]

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

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

\[
\text{Eq. (3.41)}
\]

\[
e^{-\frac{\tilde{\alpha}}{2}t} \mathbb{E} \left[ e^{\tilde{\alpha}T} |\partial_x g(X^{t,x,\phi}) - \partial_x g(X^{t,x,\phi'})|^2 \right]^{\frac{1}{2}} \leq L_g e^{\tilde{\alpha}(T-t)} \mathbb{E} \left[ |\Delta X^t_s|^2 \right]^{\frac{1}{2}}
\]

where we recall \( m_{(\alpha, \beta)} = \sup_{t \in [0, T]} \mathbb{E} \left[ |\Delta X^t_s|^2 \right]^{\frac{1}{2}} \). On the other hand, by Lemma 3.3, for all \( t \in [0, T], x, x' \in \mathbb{R}^n, \phi, \phi' \in \mathcal{V}_A, (y, z) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}, \)

\[|\partial_x H_t(x, \phi_t(x), y, z) - \partial_x H_t(x', \phi'_t(x'), y, z)| \leq (L_{\tilde{b}}|x - x'| + L_{\tilde{b}}(|x - x'| + |\phi_t(x) - \phi'_t(x')|))|y| + L_{\alpha}|x - x'| + L_{fx}(|x - x'| + |\phi_t(x) - \phi'_t(x')|)
\]

\[\leq \left((L_{\tilde{b}} + L_{\tilde{b}}(1 + [\phi]_1))|y| + L_{\alpha}|x - x'| + L_{fx} + L_{\tilde{b}}y)\right)|\phi_t(x') - \phi'_t(x')|.
\]

This along with (3.8) and the assumption that \( |Z^{t,x,\phi}_s| \leq C^{\phi}_Z \) implies that

\[
\text{Eq. (3.42)}
\]

\[
\mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha}s} |f^1_s(\cdot, Y^{t,x,\phi'}, Z^{t,x,\phi'}) - f^2_s(\cdot, Y^{t,x,\phi'}, Z^{t,x,\phi'})| ds \right)^{\frac{1}{2}} \right]^{\frac{1}{2}}
\]

\[
\leq \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha}s} \left( (C + L_{\tilde{b}}[\phi]_1)|Y^{t,x,\phi'}_s| + L_{fx}(1 + [\phi]_1) + C|Z^{t,x,\phi'}_s| \right) |\Delta X^t_s| ds \right)^{\frac{1}{2}}
\]

\[
+ \mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha}s} \left( (L_{fx} + L_{\tilde{b}}Y^{t,x,\phi'})|\phi_s(X^{t,x,\phi'}) - \phi'_s(X^{t,x,\phi'})| \right) ds \right)^{\frac{1}{2}} \right]^{\frac{1}{2}}
\]

\[
\leq C \left( \int_t^T e^{\tilde{\alpha}s} ds \right) \left( (1 + [\phi]_1)C_Y + L_{fx}(1 + [\phi]_1) + C^{\phi'}_Z \right) |\Delta X^t_s|_{S^2}
\]

\[+ (L_{fx} + C_Y)\|\phi(X^{t,x,\phi'}) - \phi'(X^{t,x,\phi'})\|_{S^2}.
\]

Substituting (3.39) into the above estimate yields

\[
\text{Eq. (3.43)}
\]

\[
\mathbb{E} \left[ \left( \int_t^T e^{\tilde{\alpha}s} |f^1_s(\cdot, Y^{t,x,\phi'}, Z^{t,x,\phi'}) - f^2_s(\cdot, Y^{t,x,\phi'}, Z^{t,x,\phi'})| ds \right)^{2^\frac{1}{2}} \right]^{\frac{1}{2}}
\]

\[
\leq CC_b \left( \int_t^T e^{\tilde{\alpha}s} ds \right) \left( (C_Y + L_{fx})(1 + [\phi]_1) + C^{\phi'}_Z TM_{[\phi]_1}M_{[\phi]_1}(1 + T + TC_b\|\phi'(0)\|_{\infty})
\]

\[+ (L_{fx} + C_Y)M_{[\phi]_1}(1 + T + TC_b\|\phi'(0)\|_{\infty})(1 + |x|)\|\phi - \phi'\|_0,
\]

21
Combining with above estimate with (3.34) and (3.41), and using $e^{-\frac{\alpha}{2}t} \int_t^T e^{\frac{\alpha}{2}s} ds \leq \frac{2}{\alpha} (e^{\frac{\alpha}{2}T} - 1)$ with $\alpha = 2(\kappa - \rho + L + L_\sigma^2)$, we conclude the desired estimate

$$\|Y^{t,x,\phi}_t - Y^{t,x,\phi\prime}_t\| \leq CC_\beta(1 + |x|)\|\phi - \phi\prime\|_0(1 + T + TC_\beta\|\phi'\|_\infty)M(\|\phi\|_1)\left(Lg_\lambda^{1/2}(\alpha,\beta) + \frac{e^{\alpha T} - 1}{\alpha}\left((Cy + Lf_\varepsilon)(1 + |\phi\|_1) + C_Z^\phi TM(\|\phi\|_1) + (Lf_\varepsilon + Cy)\right)\right)$$

with $\alpha = \kappa - \rho + L + L_\sigma^2$, $\beta = 2(\kappa - L_\sigma[|\phi\|_1]) + C$, 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 and extends the arguments of [25, 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 $\phi \in \mathcal{V}_A$ and $(t, x) \in [0, T) \times \mathbb{R}^n$, let $(Y^{t,x,\phi}, Z^{t,x,\phi}) \in S^2(t,T;\mathbb{R}^n) \times H^2(t, T; \mathcal{P}[\mathbb{R}^{n \times d}])$ be defined by (1.11). Then for all $\phi \in \mathcal{V}_A$ and $(t, x) \in [0, T) \times \mathbb{R}^n$,

$$\|Z^{t,x,\phi}\| \leq C_\phi := C_\sigma L_Y(|\phi\|_1), \quad \text{for } dt \otimes dP\text{-a.e.},$$

where the constant $L_Y(|\phi\|_1) \geq 0$ 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 $\phi^0 \in \mathcal{V}_A$, 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 $\phi^0 \in \mathcal{V}_A$, $\tau > 0$ and $m \in \mathbb{N}$, let $\phi^m$ be defined by (1.9) with the initial guess $\phi^0$ and stepsize $\tau$. Let $C \geq 0$ be a constant such that (3.8), (3.15) and (3.32) hold, let $C_\gamma \geq 0$ be defined in (3.8), and let $\alpha \in \mathbb{R}$ be defined in (3.16). For each $\phi^0 \in \mathcal{V}_A$ and $M \geq 0$, let $C_\phi(\phi^0)$ be defined in (3.11), let $L_\phi(M) \geq 0$ be defined in (3.26), and let $L_Y(M) \geq 0$ be defined in (3.16). Then for all $\phi^0 \in \mathcal{V}_A$, if we assume further that (3.25) holds and

$$C(1 + T + TC_\phi C_\phi^0) e^{T^\beta} + (Te^{T^\beta} + 1)B(\phi^0) < \frac{1}{2} \left(\frac{\mu L_\phi}{\mu + L_\phi} + \nu\right),$$

with the constants $\beta \in \mathbb{R}$, $m_{\alpha,\beta} > 0$ and $B^0(\phi^0) \geq 0$ defined by

$$\beta := 2\kappa + 2L_\phi(\phi^0) + C, \quad m_{\alpha,\beta} := \sup_{t \in [0,T]} e^{2a(T-t)} \int_t^T e^{(T-s)\beta} ds,$$

$$B^0(\phi^0) := C_\phi^2 \left[ Lg_{\lambda}^{1/2}(\alpha,\beta) + \frac{e^{\tau_\alpha - 1}}{\alpha}\left((Cy + Lf_\varepsilon)(1 + L(\phi^0)) + C_Y L_y(L(\phi^0))\right)\right],$$

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

$$|\phi^{m+1} - \phi^m|_0 \leq c|\phi^m - \phi^{m-1}|_0, \quad \forall m \in \mathbb{N}_0.$$

**Remark 3.1.** Theorem 3.11 shows that if (3.25) and (3.41) 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) $\frac{\mu L_\phi}{\mu + L_\phi} + \nu$ is sufficiently large or (b) $B(\phi^0)$
is sufficiently small. The former case holds if either \( \mu \) or \( \nu \) is sufficiently large (note that (2.2) and (2.3) imply that \( \mu \leq Lfa \)). The latter case holds for (a) small \( C_b \), or (b) small \( \frac{e^{T\alpha-1}}{\alpha} \) and \( m_{(\alpha, \beta)} \), or (c) small \( Lg, CY, Lfx \) and \( L_Y(L(\phi^\rho)) \). By the definitions of \( \alpha \) and \( \beta \), \( \frac{e^{T\alpha-1}}{\alpha} \) and \( m_{(\alpha, \beta)} \) tend to 0, as \( T \to 0 \) or \( \kappa_b \to -\infty \) or \( \rho \to \infty \) (see Lemma A.2 in Appendix A), while by (3.8) and (3.15), \( CY \) and \( L_Y(L(\phi^\rho)) \) scale linearly in \( C_g, Lg, C_fx, Lfx \), and hence \( B(\phi^\rho) \) is close to zero if \( C_g, Lg, C_fx, Lfx \) are sufficiently small.

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

\[
|\phi^{m+1}_t(x) - \phi^m_t(x)| \leq \left( 1 - \frac{1}{2} \left( \frac{\mu Lfa + \nu}{\mu + Lfa} \right) \right) |\phi^m_t(x) - \phi^{m-1}_t(x)| + \tau \bar{C}_b |Y_t^{t.x, \phi^m} - Y_t^{t.x, \phi^{m-1}}|.
\]

(3.46)

Applying Proposition 3.9 further gives

\[
\frac{|\phi^{m+1}_t(x) - \phi^m_t(x)|}{1 + |x|} \leq \left( 1 - \frac{1}{2} \left( \frac{\mu Lfa + \nu}{\mu + Lfa} \right) \right) \frac{|\phi^m_t(x) - \phi^{m-1}_t(x)|}{1 + |x|} + \tau \bar{C}_b B[\phi^m, \phi^{m-1}, \phi^{m-1}_Z] |\phi^m - \phi^{m-1}|_0
\]

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

\[
|\phi^{m+1} - \phi^m|_0 \leq \left( 1 + \tau \left( \bar{C}_b B[\phi^m, \phi^{m-1}, \phi^{m-1}_Z] - \frac{1}{2} \left( \frac{\mu Lfa + \nu}{\mu + Lfa} \right) \right) \right) |\phi^m - \phi^{m-1}|_0,
\]

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

\[
\bar{C}_b B[\phi^m, \phi^{m-1}, \phi^{m-1}_Z] \leq C\bar{C}_b^2 \left( 1 + T + T \bar{C}_b \sup_{t \in [0, T]} |\phi^{m-1}_t(0)| \right) e^{T\beta+} \left( Lg m^{1/2}_{(\alpha, \beta)} \right)
\]

\[
+ \frac{e^{T\alpha-1}}{\alpha} ((CY + Lfx)(1 + [\phi^m]_1) + C^{\phi^{m-1}_Z})(Te^{T\beta+} + 1)
\]

\[
\leq C(1 + T + T \bar{C}_b C(L(\phi^\rho)) e^{T\beta+} (Te^{T\beta+} + 1) B(\phi^\rho),
\]

with \( \beta \), \( m_{(\alpha, \beta)} \) and \( B(\phi^\rho) \) defined as in (3.44). Then under (3.43), the desired estimate holds with

\[
c = 1 + \tau \left( C(1 + T + T \bar{C}_b C(L(\phi^\rho)) e^{T\beta+} (Te^{T\beta+} + 1) B(\phi^\rho) - \frac{1}{2} \left( \frac{\mu Lfa + \nu}{\mu + Lfa} \right) \right) \in [0, 1).
\]

Note that (3.43) implies \( c < 1 \), and \( \tau \leq \frac{2}{\mu + Lfa} \wedge \frac{1}{\nu} \) implies that \( c \geq 1 - \frac{1}{2} \left( \frac{\mu Lfa + \nu}{\mu + Lfa} \right) \geq 0 \).}

**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 \( (\alpha^m)_m \in \mathbb{N}_0 \) 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 $\| \cdot \|_X$, $F : X \to \mathbb{R}$ be a Fréchet differentiable function, $G : X \to \mathbb{R} \cup \{ \infty \}$ be a proper, lower semicontinuous, convex function, and $x^* \in \text{dom} \ G$. Then $x^*$ is a stationary point of $F + G$ if and only if for some $\tau > 0$,

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

where for all $x \in X$, $\text{prox}_{\tau G}(x) = \min_{z \in X} \left( \frac{1}{\tau} \| z - x \|_X + \tau G(z) \right)$.

Proof. By [29, Proposition 1.107], the Fréchet differentiability of $F$ implies that $\partial(F + G)(x^*) = \nabla F(x^*) + \partial G(x^*)$. Hence $x^*$ is a stationary point of $F + G$ if and only if $-\nabla F(x^*) \in \partial G(x^*)$. By the properties of $G$, $\partial G$ agrees with the convex subdifferential of $G$ (see [29, Equation 1.51 and Theorem 1.93]), which along with the definition of $\text{prox}$ shows that for all $x, u \in X$ and $\tau > 0$, $u = \text{prox}_{\tau G}(x)$ if and only if $x - u \in \partial(\tau G)(u)$. Hence by $-\nabla F(x^*) \in \partial G(x^*)$, for all $\tau > 0$, $(x^* - \tau \nabla F(x^*)) - x^* \in \partial(\tau G)(x^*)$, which leads to the desired result. \hfill \Box

The 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 $(\alpha^m)_{m \in \mathbb{N}_0}$ based on Proposition 3.12.

Theorem 3.13. Assume the same notation as in Theorem 3.11. For each $\phi \in \mathcal{V}_A$, let $\alpha^\phi \in \mathcal{H}^2(\mathbb{R}^k)$ be the associated control process. Then for all $\phi^0 \in \mathcal{V}_A$ satisfying (3.25) and (3.43), and for all $\tau \in (0, \frac{2}{\mu + L_f A} \cup \frac{1}{\rho})$, there exists $c \in [0, 1)$, $C \geq 0$ and $\phi^* \in \mathcal{V}_A$ such that

1. for all $m \in \mathbb{N}_0$, $|\phi^{m+1} - \phi^*|_0 \leq c|\phi^m - \phi^*|_0$,
2. for all $m \in \mathbb{N}_0$, $\|\alpha^m - \alpha^*\|_{\mathcal{H}^2} \leq Cc^m |\phi^0 - \phi^*|_0$,
3. $\alpha^\phi$ 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 $\phi^0 \in \mathcal{V}_A$ satisfy (3.25) and (3.43), and $\tau \in (0, \frac{2}{\mu + L_f A} \cup \frac{1}{\rho})$. 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 $(B([0, T] \times \mathbb{R}^n; \mathbb{R}^k), | \cdot |_0)$. As $(B([0, T] \times \mathbb{R}^n; \mathbb{R}^k), | \cdot |_0)$ is a Banach space, the Banach fixed point theorem shows that there exists $\phi^* \in B([0, T] \times \mathbb{R}^n; \mathbb{R}^k)$ such that $\lim_{m \to \infty} |\phi^m - \phi^*|_0 = 0$. 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 $[\phi^*]_1 \leq L(\phi^0)$. Hence, to show $\phi^* \in \mathcal{V}_A$, it remains to prove $\phi^*$ takes values in $A$ a.e.

By Proposition 3.9 and $\sup_{m \in \mathbb{N}_0, t \in [0, T]} \left( |\phi^m_t(0) + [\phi^m(t)]_1 < \infty \right)$, there exists $C \geq 0$ such that for all $(t, x) \in [0, T] \times \mathbb{R}^n$ and $m \in \mathbb{N}_0$, $|Y^{t,x,\phi^m} - Y^{t,x,\phi^m}|_0 \leq C(1 + \| x \| ) |\phi^m - \phi^m|_0$. This along with the fact that $(\phi^m)_{m \in \mathbb{N}}$ is a Cauchy sequence in $(B([0, T] \times \mathbb{R}^n; \mathbb{R}^k), | \cdot |_0)$ shows that for all $(t, x) \in [0, T] \times \mathbb{R}^n$, $(Y^t_{l,x,\phi^m})_{m \in \mathbb{N}_0}$ is a Cauchy sequence in $\mathbb{R}^n$. Hence there exists a function $\mathcal{Y} : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ such that $\lim_{m \to \infty} Y^t_{l,x,\phi^m} = \mathcal{Y}_l(x)$ 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 $\text{prox}_{\tau \ell}$ and $\partial \alpha H^{\ell \epsilon}_t$ and the pointwise convergence of $(\phi^m)_{m \in \mathbb{N}_0}$ and $(Y^t_{l,x,\phi^m})_{m \in \mathbb{N}_0}$, one can pass $m$ to infinity in (1.9) and show for a.e. $(t, x) \in [0, T] \times \mathbb{R}^n$,

$$\phi^*_t(x) = \lim_{m \to \infty} \phi^{m+1}_t(x) = \lim_{m \to \infty} \text{prox}_{\tau \ell}(\phi^m_t(x) - \tau \partial \alpha H^{\ell \epsilon}_t(x, \phi^m_t(x), Y^t_{l,x,\phi^m})) = \text{prox}_{\tau \ell}(\phi^*_t(x) - \tau \partial \alpha H^{\ell \epsilon}_t(x, \phi^*_t(x), \mathcal{Y}_t))(3.47)$$

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

$$\phi^*_t(x) = \text{prox}_{\tau \ell}(\phi^*_t(x) - \tau \partial \alpha H^{\ell \epsilon}_t(x, \phi^*_t(x), Y^t_{l,x,\phi^*}))$$

(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 \(x' = x\), \(y' = Y_t^{t,x,\phi_0^m}\), \(a = \phi_t^m(x)\) and \(a' = \phi_t^*(x)\) and (3.48) immediately yield that for all \(\tau \in (0, \frac{1}{\mu + L_{fa}})\),

\[
|\phi_t^{m+1}(x) - \phi^*(x)| \leq \left(1 - \frac{1}{2} \left(\frac{\mu L_{fa}}{\mu + L_{fa}} + \nu\right)\right) |\phi_t^m(x) - \phi_t^*(x)| + \tau C_\delta |Y_t^{t,x,\phi_0^m} - Y_t^{t,x,\phi^*}|.
\]

Now, following the exact same lines as the proof of Theorem 3.11 (cf. (3.46)) and using the above facts that \(\phi^* \in \mathcal{V}_A\), \(\sup_{t \in [0,T]}|\phi_t^*(0)| \leq C(\phi_0)\) and \([\phi^*]_1 \leq L(\phi_0)\), we deduce \(|\phi_t^{m+1} - \phi^*|_0 \leq C|\phi_t^m - \phi^*|_0\) with the same constant \(C \in [0, 1]\) as in Theorem 3.11.

To prove Item (2), observe that for each \(m \in \mathbb{N}_0\), \(\alpha^m = \phi_t^m(X_{\xi_0,\phi_0^m})\) and \(\alpha^* = \phi^*(X_{\xi_0,\phi^*})\), which implies that

\[
||\alpha^{m+1} - \alpha^*||_{\mathcal{H}^2} \leq ||\alpha^{m+1}(X_{\xi_0,\phi_0^m}) - \alpha^{m+1}(X_{\xi_0,\phi^*})||_{\mathcal{H}^2} + ||\alpha^{m+1}(X_{\xi_0,\phi^*}) - \alpha^*(X_{\xi_0,\phi^*})||_{\mathcal{H}^2} \leq [\phi^{m+1}]_1 ||X_{\xi_0,\phi_0^m} - X_{\xi_0,\phi^*}||_{\mathcal{H}^2} + ||\phi^{m+1} - \phi^*||_0 (1 + ||X_{\xi_0,\phi^*}||_{\mathcal{H}^2}).
\]

By using \(\phi^* \in \mathcal{V}_A\) and \(\sup_{m \in \mathbb{N}}([\phi_0^m]_1) < \infty\) and Lemma 3.1, one can easily show that there exists \(C \geq 0\) such that for all \(m \in \mathbb{N}_0\), \(||X_{\xi_0,\phi_0^m}||_{\mathcal{H}^2} \leq C\) and \(||X_{\xi_0,\phi^m} - X_{\xi_0,\phi^*}||_{\mathcal{H}^2} \leq C|\phi^m - \phi^*|_0\), which along with (3.49) leads to the desired estimate \(||\alpha^{m+1} - \alpha^*||_{\mathcal{H}^2} \leq Cm|\phi_t^0 - \phi^*|_0\) for all \(m \in \mathbb{N}_0\), with some constant \(\bar{C} \geq 0\) independent of \(m\).

It remains to prove Item (3). Let \(\bar{H}^{'re} : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \rightarrow \mathbb{R}\) and \(\bar{H} : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^{n \times d} \rightarrow \mathbb{R}\) be such that for all \((t, x, a, y, z) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}^{n \times d}\), \(\bar{H}^{'re}(x, a, y) := (b_t(x, a), y) + e^{-\rho t} f(x, a)\) and \(\bar{H}_t(x, a, y, z) := \bar{H}_t^{'re}(x, a, y) + \langle \sigma_t(x, z) \rangle\). For each \((t, x) \in [0, T] \times \mathbb{R}^n\), let \(X_t^{t,x,\phi^*} \in \mathcal{S}^2(t, T; \mathbb{R}^n)\) satisfy (1.10) with \(\phi_t^* \in \mathcal{V}_A\), and \((\bar{Y}_t^{t,x,\phi^*}, \bar{Z}_t^{t,x,\phi^*}) \in \mathcal{S}^2(t, T; \mathbb{R}^n) \times \mathcal{H}^2(t, T; \mathbb{R}^{n \times d})\) be such that \(\delta \bar{Y}_s^{t,x,\phi^*} = e^{-\rho s} \bar{Y}_s^{t,x,\phi^*}\) 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 \(\eta > 0\),

\[
a = \text{prox}_\ell(a - \eta u) \iff 0 \in (a - (a - \eta u)) + \partial \ell(a) \iff 0 \in u + \partial (\eta^{-1} \ell)(a) \iff 0 \in (a - (a - \eta u)) + \partial (\eta^{-1} \ell)(a) \iff a = \text{prox}_{\eta^{-1} \ell}(a - u).
\]

Hence by (3.48) and the affineness of \(H^{'re} \bar{H}\) in \(y\) and \(z\) implies that \(\bar{Y}_t^{t,x,\phi^*} = e^{-\rho t} Y_s^{t,x,\phi^*}\) 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 \(\eta > 0\),

\[
\phi_t^*(x) = \text{prox}_\ell(\phi_t^*(x) - \tau e^{\rho t} \partial_a \bar{H}_t^{t,x,\phi^*}(x, \phi_t^*(x), e^{-\rho t} Y_t^{t,x,\phi^*}))
\]

\[
= \text{prox}_\ell(e^{-\rho t} \phi_t^*(x) - \tau \partial_a \bar{H}_t^{t,x,\phi^*}(x, \phi_t^*(x), \bar{Y}_t^{t,x,\phi^*}))
\]

(3.50)

Now consider the solution \((X_{\xi_0,\phi^*}, Y_{\xi_0,\phi^*}, Z_{\xi_0,\phi^*}) \in \mathcal{S}^2(\mathbb{R}^n) \times \mathcal{S}^2(\mathbb{R}^n) \times \mathcal{H}^2(\mathbb{R}^{n \times d})\) to the following FBSDE: for all \(t \in [0, T]\),

\[
\begin{align*}
\text{d}X_t^{\xi_0,\phi^*} &= b_t(X_t^{\xi_0,\phi^*}, \phi_t^*(X_t^{\xi_0,\phi^*})) \text{d}t + \sigma_t(X_t^{\xi_0,\phi^*}) \text{d}W_t,
\text{d}Y_t^{\xi_0,\phi^*} &= -\partial_x \bar{H}_t(X_t^{\xi_0,\phi^*}, \phi_t^*(X_t^{\xi_0,\phi^*}), Y_t^{\xi_0,\phi^*}, Z_t^{\xi_0,\phi^*}) \text{d}t + \bar{Z}_t^{\xi_0,\phi^*} \text{d}W_t,
\text{d}Z_t^{\xi_0,\phi^*} &= e^{-\rho t} \partial_x g(X_T^{\xi_0,\phi^*}) \text{d}t \otimes \text{d}\mathbb{P}\ a.e.,
\end{align*}
\]

The Markov property in [41, Theorem 5.1.3] implies \(\bar{Y}_t^{t,x,\phi^*} = \bar{Y}_t^{X_t^{\xi_0,\phi^*},\phi_t^*(X_t^{\xi_0,\phi^*}), \bar{Y}_t^{\xi_0,\phi^*}}\).

(3.51)
Observe that for all $\alpha \in \mathcal{H}^2(\mathbb{R}^k)$, $J(\alpha; \xi_0) = F(\alpha) + G(\alpha)$, where
\[
F(\alpha) := \mathbb{E} \left[ \int_0^T e^{-\rho t} f_t(X_t^{\xi_0, \alpha}, \alpha_t) \, dt + e^{-\rho T} g(X_T^{\xi_0, \alpha}) \right], \quad G(\alpha) := \mathbb{E} \left[ \int_0^T e^{-\rho t} \ell(\alpha_t) \, dt \right],
\]
where $X^{\xi_0, \alpha}$ satisfies (1.1). Then the regularity of coefficients and [1, Lemma 3.1] imply that $F$ is Fréchet differentiable, and the derivative $\nabla F$ at $\alpha^* = \phi^*(X^{\xi_0, \phi^*})$ is given by
\[
\nabla F(\alpha^*)_t = \partial_\alpha \tilde{H}_t^{\xi_0, \phi^*}(X_t^{\xi_0, \phi^*}, \phi^*_t(X_t^{\xi_0, \phi^*}, \tilde{Y}_t^{\xi_0, \phi^*}), \, dt \otimes d\mathbb{P} \text{ a.e.}
\]
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 $\tau > 0$, $\text{prox}_{\tau G}(\alpha) = \text{prox}_{\tau e^{-\rho t} \ell}(\alpha)$ for $dt \otimes d\mathbb{P}$ a.e. Hence, Proposition 3.12 shows that $\alpha^* \in \mathcal{H}^2(\mathbb{R}^k)$ is a stationary point of $J(\cdot; \xi_0)$. \square

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 $\phi^0 \in \mathcal{V}_A$, 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, $\tau \in (0, \frac{2}{\mu + L_{fa}} \wedge \frac{1}{\beta}]$, and there exist constants $\tilde{C}, \tilde{L} \geq 0$ such that $\sup_{t \in [0, T]} |\tilde{\phi}^m_t(0)| \leq \tilde{C}$ and $[\tilde{\phi}^m]_1 \leq \tilde{L}$ for all $m \in \mathbb{N}_0$, and
\[
C(1 + T + TC_b \tilde{C}) e^{T\beta} (Te^{T\beta} + 1) \tilde{B} \leq \frac{1}{2} \left( \frac{\mu L_{fa}}{\mu + L_{fa}} + \nu \right), \quad (3.52)
\]
with the constants $\beta \in \mathbb{R}$, $m_{(\alpha, \beta)} > 0$ and $\tilde{B} \geq 0$ defined by
\[
\beta := 2\kappa_b + 2L_b \max\{L(\phi^0), \tilde{L}\} + C, \quad m_{(\alpha, \beta)} := \sup_{t \in [0, T]} e^{2\alpha(T-t)} \int_t^T e^{(T-s)\beta} \, ds, \quad \tilde{B} := C^2_b \left[ L_a \mathbb{m}_{(\alpha, \beta)}^{1/2} + \frac{e^{T\alpha - \beta}}{\alpha} \left( (C_Y + L_{fa})(1 + L(\phi^0)) + C_{\sigma} L_Y(\tilde{L}) \right) \right].
\]
Then for
\[
c = 1 + \tau \left( C(1 + T + TC_b \tilde{C}) e^{T\beta} (Te^{T\beta} + 1) \tilde{B} - \frac{1}{2} \left( \frac{\mu L_{fa}}{\mu + L_{fa}} + \nu \right) \right) \in [0, 1), \quad (3.54)
\]
and for all $m \in \mathbb{N}_0$,
\[
|\phi^* - \tilde{\phi}^m|_0 \leq c^m |\phi^0 - \phi^*|_0 + \sum_{j=0}^{m-1} c^{m-1-j} T \kappa_b \mathbb{m}_{(\alpha, \beta)} \sup_{(t,x) \in [0, T] \times \mathbb{R}^n} \left| \frac{Y_t^{x, \phi^0} - \tilde{Y}_t^{\phi^0}(x)}{1 + |x|} \right|, \quad (3.55)
\]
where $\phi^* \in \mathcal{V}_A$ is the limit function in Theorem 3.13.

**Proof.** First, observe that the conditions $\sup_{t \in [0, T]} |\tilde{\phi}^m_t(0)| \leq \tilde{C}$ and $[\tilde{\phi}^m]_1 \leq \tilde{L}$ for all $m \in \mathbb{N}_0$ guarantee $\tilde{\phi}^m \in \mathcal{V}_A$. We continue by quantifying $|\phi^m_{t+1}(x) - \tilde{\phi}^m_{t+1}(x)|$ for any $(t, x) \in [0, T] \times \mathbb{R}^n$, where $\tilde{\phi}^m_{t+1}(x)$ is defined by (2.14). By Lemma 3.4 with $x = x'$, $a = \phi^m_t(x)$, $a' = \phi^{m+1}_t(x)$, $y = Y_t^{x, \phi^0}$ and $y' = \tilde{Y}_t^{\phi^m}(x)$,
\[
|\phi^m_{t+1}(x) - \tilde{\phi}^m_{t+1}(x)| \leq \left( 1 - \tau \frac{1}{2} \left( \frac{\mu L_{fa}}{\mu + L_{fa}} + \nu \right) \right) |\phi^m_t(x) - \tilde{\phi}^m_t(x)| + \tau C_b Y_t^{x', \phi^0} - \tilde{Y}_t^{\phi^m}(x). \quad (3.55)
\]
Now, using the definition of $\tilde{B}$ given in (3.53), the estimates in the proof of Theorem 3.11 and recalling $\phi^0 = \tilde{\phi}^0$, observe that

$$|Y_t^{t,x,\phi^m} - \tilde{Y}_t^{\tilde{\phi}^m}(x)| \leq |Y_t^{t,x,\phi^m} - Y_t^{t,x,\tilde{\phi}^m}| + |Y_t^{t,x,\tilde{\phi}^m} - \tilde{Y}_t^{\tilde{\phi}^m}(x)|$$

$$\leq C(1 + T + TC_B e^{2T\beta} + \beta) |\phi^m_t(x) - \tilde{\phi}^m_t(x)| + |Y_t^{t,x,\tilde{\phi}^m} - \tilde{Y}_t^{\tilde{\phi}^m}(x)|.$$  

Now let $c$ be defined as in (3.54). The condition (3.52) implies $c < 1$, and $\tau \leq \frac{2}{\mu + L_{fa}} + \frac{1}{l}$ implies that $c \geq 1 - \frac{\tau}{\mu + L_{fa}}$ and

$$|\phi^m_t(x) - \tilde{\phi}^m_t(x)| \leq c|\phi^m_t(x) - \tilde{\phi}^m_t(x)| + \tau C_b |Y_t^{t,x,\tilde{\phi}^m} - \tilde{Y}_t^{\tilde{\phi}^m}(x)|,$$

which along with $\phi^0 = \tilde{\phi}^0$ implies that

$$|\phi^{m+1}_t(x) - \tilde{\phi}^{m+1}_t(x)| \leq \sum_{j=0}^{m} c^{m-j} \tau C_b |Y_t^{t,x,\phi^j} - \tilde{Y}_t^{\phi^j}(x)|.$$

The desired conclusion then follows from Theorem 3.13 Item (1).

\[\Box\]

A Proofs of technical results

This section is devoted to the proofs of Proposition 2.1 and Lemmas 3.1, 3.3, 3.10.

To prove Proposition 2.1, we first establish a general well-posedness result for BSDEs with non-Lipschitz and unbounded coefficients. Although the result does not follow directly from [7, Proposition 3.5] due to the unboundedness of $A_t$ (i.e., the condition (A1(3)) in [7] fails), the techniques there can be extended to the present setting.

**Lemma A.1.** Let $T > 0$, $\kappa \in \mathbb{R}$, $L \geq 0$, $\xi \in L^\infty(F_T; \mathbb{R}^n)$, let $A \in \mathcal{H}^2(\mathbb{R}^{n \times n})$ be such that $y^\top A_t(\omega)y \leq \kappa |y|^2$ for all $(t, \omega, y) \in [0, T] \times \Omega \times \mathbb{R}^n$, and let $f : [0, T] \times \Omega \times \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^n$ satisfy for all $(t, \omega) \in [0, T] \times \Omega$ and $z, z' \in \mathbb{R}^{n \times d}$, $(f_t(\cdot, z))_{t \in [0,T]}$ is progressively measurable, $|f_t(\omega, z) - f_t(\omega, z')| \leq L|z - z'|$ and $\sup_{(t,\omega) \in [0,T] \times \Omega} |f_t(\omega, 0)| < \infty$. Then the following BSDE

$$dY_t = -(A_tY_t + f_t(\cdot, Z_t)) dt + Z_t dW_t, \quad t \in [0, T]; \quad Y_T = \xi, \quad (A.1)$$

admits a unique solution $(Y, Z) \in S^2(\mathbb{R}^n) \times \mathcal{H}^2(\mathbb{R}^{n \times d})$.

**Proof of Lemma A.1.** Throughout this proof, let $h_t(\omega, y, z) = A_t(\omega)y + f_t(\omega, z)$ for all $(t, \omega, y, z) \in [0, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d}$. Then for all $(t, \omega) \in [0, T] \times \Omega$ and $y, y' \in \mathbb{R}^n$ and $z, z' \in \mathbb{R}^{n \times d}$, $|h_t(\omega, y, z) - h_t(\omega, y', z')| \leq |y - y'|^2$ and $|h_t(\omega, y, z) - h_t(\omega, y, z')| \leq L|z - z'|$. Hence the a-priori estimate in Lemma 3.2 shows that (A.1) admits at most one solution in the space $S^2(\mathbb{R}^n) \times \mathcal{H}^2(\mathbb{R}^{n \times d})$.

To establish the existence of solutions, we construct a sequence of Lipschitz functions $(h^m)_{m \in \mathbb{N}}$ approximating $h$ as in [7, Proposition 3.5]. Without loss of generality, we assume that $\kappa = 0$, which in general can be achieved with an exponential time scaling of the solution. For each $m \in \mathbb{N}$, let $h : [0, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^n$ be such that

$$h^m_t(\omega, y, z) := A^m_t(\omega)y + f_t(\omega, z), \quad \forall (t, \omega, y, z) \in [0, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d},$$

where $A^m_t(\omega) := \min(m, |A_t(\omega)|) A_t(\omega)$. It is clear that for each $m \in \mathbb{N}$, $h^m$ is uniformly Lipschitz continuous in $y$ and $z$. Hence by [41, Theorem 4.3.1], there exist unique processes $(Y^m, Z^m) \in \mathcal{H}^2(\mathbb{R}^n) \times \mathcal{H}^2(\mathbb{R}^{n \times d})$ satisfying

$$dY_t = -h^m_t(Y_t, Z_t) dt + Z_t dW_t, \quad t \in [0, T]; \quad Y_T = \xi.$$

27
Observe from $\kappa = 0$ that for all $(t, \omega, y, z) \in [0, T] \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n \times d}$,
\[
\langle y, h^m_t(\omega, y, z) \rangle = \langle y, A^m_t(\omega)y \rangle + \langle y, f_t(\omega, z) \rangle \leq |y|(C_f + L|z|),
\]
where $C_f = \sup_{(t, \omega) \in [0, T] \times \Omega} |f_t(\omega, 0)|$. Hence by [7, Proposition 2.1], for $dt \otimes d\mathbb{P}$ a.e.,
\[
\sup_{t \in [0, T]} |Y^m|^2 \leq \|\xi\|^2_{L^\infty} e^{\beta T} + \frac{C_f^2}{\beta}(e^{\beta T} - 1), \quad \text{with } \beta = 1 + L^2.
\]
We now denote by $C$ a generic constant independent of $m$. By Lemma 3.2, for all $m, m' \in \mathbb{N}$,
\[
\|Y^{m'} - Y^m\|_{S^2}^2 + \|Z^{m'} - Z^m\|_{H^2}^2 \leq C E \left[ \left( \int_0^T |h^{m'}_{t}(\cdot, Y^{m}_{t}, Z^{m}_{t}) - h^{m}_{t}(\cdot, Y^{m}_{t}, Z^{m}_{t})| dt \right)^2 \right]
\]
\[
\leq C E \left[ \left( \int_0^T |A^{m'}_{t} - A^{m}_{t}| |Y^m_t| dt \right)^2 \right] \leq C \|A^{m'} - A^m\|_{H^2}^2,
\]
where the last inequality follows from the uniform bound of $(Y^m)_{m \in \mathbb{N}}$. Lebesgue’s dominated convergence theorem shows that $\|A^{m'} - A^m\|_{H^2}$ tends to zero as $m, m' \to \infty$ with $m' \geq m$. This implies that $(Y^m, Z^m)_{m \in \mathbb{N}}$ is a Cauchy sequence in $S^2(\mathbb{R}^n) \times H^2(\mathbb{R}^{n \times d})$, and hence converges to some processes $(Y, Z)$ in $S^2(\mathbb{R}^n) \times H^2(\mathbb{R}^{n \times d})$. Moreover, by Fatou’s lemma, there exits $C \geq 0$ such that $|Y| \leq C$ for $dt \otimes d\mathbb{P}$ a.e.

It remains to verify that $(Y, Z)$ is a solution to (A.1). As $(Z^m)_{m \in \mathbb{N}}$ converges to $Z$ in $H^2(\mathbb{R}^{n \times d})$, for all $t \in [0, T]$, $\lim_{m \to \infty} \| \int_0^T Z^m dW_s - \int_0^T Z_s dW_s \|_{L^2} = 0$. Moreover, for all $t \in [0, T]$,
\[
\left\| \int_0^T (h^m_t(\cdot, Y^m_t, Z^m_t) - h_t(\cdot, Y_t, Z_t)) dt \right\|_{L^1} \leq \left\| \int_0^T (|A^m_t - A_t| |Y^m_t| + |A_t| |Y^m_t - Y_t|) dt \right\|_{L^1} + \left\| \int_0^T |f_t(\cdot, Z^m_t) - f_t(\cdot, Z_t)| dt \right\|_{L^1},
\]
which converge to zero as $m \to \infty$, due to the Cauchy-Schwarz inequality, Lebesgue’s dominated convergence theorem, the boundedness of $(Y^m)_{m \in \mathbb{N}}$ and the convergence of $(Y^m, Z^m)_{m \in \mathbb{N}}$. This proves the desired existence result.

Armed with Lemma A.1, we prove the well-posedness of the iterates $(\phi^m)_{m \in \mathbb{N}_0}$ for general stepsizes $\tau > 0$.

*Proof of Proposition 2.1.* For any given $\phi^0 \in \mathcal{V}_A$ and $\tau > 0$, we prove the statement with an induction argument. Due to the assumption $\phi^0 \in \mathcal{V}_A$, the statement clearly holds for $m = 0$.

Suppose that $\phi^m \in \mathcal{V}_A$ for some $m \in \mathbb{N}_0$. Then by (H.1), $(t, x) \mapsto (b_t(x, \phi^m_t(x)), \sigma_t(x))$ is locally Lipschitz and monotone in $x$. Hence, for each $(t, x) \in [0, T] \times \mathbb{R}^n$, by [26, Theorem 3.6, p. 58], (1.10) admits a unique solution $X^{t, x, \phi^m} \in S^2(t, T; \mathbb{R}^n)$.

We then apply Lemma A.1 for the well-posedness of $(Y^{t, x, \phi^m}, Z^{t, x, \phi^m})$. By Lemma 3.3 and the fact that $\phi^m$ takes values in $A$, $|\partial_x H_t(x, \phi_t(x), y, z) - \partial_x H_t(x, \phi_t(x), y, z)| \leq L_\phi |z - z'|$, and
\[
(y - y', \partial_x H_t(x, \phi^m_t(x), y, z) - \partial_x H_t(x, \phi^m_t(x), y', z)) \leq (\kappa_b + \rho + L_b)|y - y'|^2.
\]
Moreover, by (2.1) and (2.5), $\partial_x g(X^{t, x, \phi^m}_T)$ and $\partial_x f_s(X^{t, x, \phi^m}_s, \phi^m_s(X^{t, x, \phi^m}_s))$ are uniformly bounded $dt \otimes d\mathbb{P}$-a.e., and
\[
\|\partial_x b(X^{t, x, \phi^m}, \phi^m(X^{t, x, \phi^m}))\|_{H^2} \leq C(1 + \|X^{t, x, \phi^m}\|_{S^2}) < \infty,
\]

28
where we used the linear growth of \( \partial_x \hat{b} \) and \( \phi^m \) and the boundedness of \( \partial_x \hat{b} \). Hence by Lemma A.1, (1.11) admits a unique solution in the space \( S^2(t, T; \mathbb{R}^n) \times H^2(t, T; \mathbb{R}^{n \times d}) \). The fact that \( (t, x) \mapsto Y_t^{t, x, \phi^m} \) can be identified as a deterministic function follows from \([32, \text{Remark 2.1}].\)

Finally, similar to Propositions 3.5 and 3.7, one can prove that \( (t, x) \mapsto Y_t^{t, x, \phi^m} \) is bounded and \( x \mapsto Y_t^{t, x, \phi^m} \) is Lipschitz continuous uniformly in \( t \). As \( \tau \ell : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\} \) is proper, lower semicontinuous and convex, \( z \mapsto \text{prox}_{\tau \ell}(z) \) is Lipschitz continuous and takes values in \( \mathbf{A} \). Hence, one can easily deduce from (H.1) that \( \phi^{m+1} \in \mathcal{Y}_A \), which completes the induction argument. \( \square \)

**Proof of Lemma 3.1.** Throughout this proof, let \( C \) be a generic constant depending only on \( p \), let \( t \in [0, T], x_1, x_2 \in \mathbb{R}^n \) and write \( X_s^1 = X_{s,t^1}^1, X_s^2 = X_{s,t^2}^2 \) and \( \Delta X_s = X_s^1 - X_s^2 \). By applying Itô’s formula to \( (|\Delta X_s|^p)_{s \in [t, T]} \), for \( t \leq s \leq T \),

\[
|\Delta X_s|^p \leq |x_1 - x_2|^p + \int_t^s (p|\Delta X_r|^p - 2\langle \Delta X_r, b_r(X_r^1) - b_r(X_r^2) \rangle) dr \\
+ \frac{p(p-1)}{2} |\Delta X_r|^{p-4} \left| b_r(X_r^2) - b_r(X_r^1) \right|^2 dr \\
+ p \int_t^s |\Delta X_r|^{p-2} \left| \Delta X_r, \sigma_r(X_r^1) - \sigma_r(X_r^2) \right|^2 dW_r, \tag{A.3}
\]

which along with the assumptions of \( b^1 \) and \( \sigma^1 \) gives

\[
|\Delta X_s|^p \leq |x_1 - x_2|^p + \int_t^s \left( (p\mu_1 + p(p-1)\nu_1^2)|\Delta X_r|^p \\
+ p|\Delta X_r|^{p-1}|b_r(X_r^1) - b_r(X_r^2)| + p(p-1)|\Delta X_r|^{p-2} |\sigma_r(X_r^1) - \sigma_r(X_r^2)|^2 \\
+ p \int_t^s |\Delta X_r|^{p-2} \left| \Delta X_r, \sigma_r(X_r^1) - \sigma_r(X_r^2) \right|^2 dW_r \right) \right). \tag{A.4}
\]

Observe that by the Burkholder-Davis-Gundy inequality (see [41, Theorem 2.4.1]) and Young’s inequality, for all \( \varepsilon > 0 \),

\[
pE \left[ \sup_{s \in [t, T]} \left| \int_t^s |\Delta X_r|^{p-2} \left( \Delta X_r, \sigma_r(X_r^1) - \sigma_r(X_r^2) \right) dW_r \right|^p \right] \\
\leq Cp^2 \left( \int_t^T |\Delta X_r|^{p-4} |\Delta X_r|^2 |\sigma_r(X_r^1) - \sigma_r(X_r^2)|^2 dr \right)^{1/2} \\
\leq Cp^2 \left[ \sup_{s \in [t, T]} |\Delta X_s|^{p/2} \left( \int_t^T |\Delta X_r|^{p-2} |\sigma_r(X_r^1) - \sigma_r(X_r^2)|^2 dr \right)^{1/2} \right] \\
\leq \varepsilon \|\Delta X\|_{S^p}^p + C\nu^2 \varepsilon^{-1} p^2 \mathbb{E} \left[ \int_t^T |\Delta X_r|^p dr \right] + C\varepsilon^{-1} p^2 \int_t^T |\Delta X_r|^{p-2} |\sigma_r(X_r^1) - \sigma_r(X_r^2)|^2 dr.
\]

Hence, by taking supremum over \( s \in [t, T] \) and expectations on both sides of (A.4),

\[
\|\Delta X\|_{S^p}^p \\
\leq |x_1 - x_2|^p + (p\mu_1 + p(p-1)\nu_1^2) \mathbb{E} \left[ \int_t^T |\Delta X_r|^p dr \right] + \epsilon \|\Delta X\|_{S^p}^p + C\nu^2 \varepsilon^{-1} p^2 \mathbb{E} \left[ \int_t^T |\Delta X_r|^p dr \right] \\
+ E \left[ \int_t^T (p|\Delta X_r|^{p-1}|b_r(X_r^1) - b_r(X_r^2)| + (p(p-1) + C\varepsilon^{-1} p^2)|\Delta X_r|^{p-2} |\sigma_r(X_r^1) - \sigma_r(X_r^2)|^2 dr \right].
\]

29
Then by Young’s inequality,

\[
\mathbb{E} \left[ \int_t^T (p|\Delta X_r|^{p-1}|b_r^1(X_r^2) - b_r^2(X_r^2)| + (p(p-1) + C\varepsilon^{-1}p^2)|\Delta X_r|^{p-2}\sigma_r^1(X_r^2) - \sigma_r^2(X_r^2)|^2 \right] dr \\
\leq \mathbb{E} \left[ \sup_{s \in [t,T]} |\Delta X_s|^{p-1} \int_t^T p|b_r^1(X_r^2) - b_r^2(X_r^2)| dr \right] \\
+ \mathbb{E} \left[ \sup_{s \in [t,T]} |\Delta X_s|^{p-2} \int_t^T (p(p-1) + C\varepsilon^{-1}p^2)|\sigma_r^1(X_r^2) - \sigma_r^2(X_r^2)|^2 dr \right] \\
\leq \varepsilon \|\Delta X\|_{S^p}^p + \varepsilon^{-1}T^{p/2}C(p)\mathbb{E} \left[ \left( \int_t^T |b_r^1(X_r^2) - b_r^2(X_r^2)|^2 dr \right)^{p/2} \right] \\
+ \varepsilon^{-1}C(p)(p(p-1) + C\varepsilon^{-1}p^2)^{p/2} \mathbb{E} \left[ \left( \int_t^T |\sigma_r^1(X_r^2) - \sigma_r^2(X_r^2)|^2 dr \right)^{p/2} \right],
\]

which gives us that

\[
\|\Delta X\|_{S^p}^p \leq \|x_1 - x_2\|^p + (p\mu_1 + p(p-1)\nu_1^2)\mathbb{E} \left[ \int_t^T |\Delta X_r|^p dr \right] + \varepsilon \|\Delta X\|_{S^p}^p \\
+ C\nu_1^2\varepsilon^{-1}p^2 \mathbb{E} \left[ \int_t^T |\Delta X_r|^p dr \right] + \varepsilon^{-1}T^{p/2}C(p)||b^1(X^{t,x_2,2}) - b^2(X^{t,x_2,2})||_{H^p}^p \\
+ \varepsilon^{-1}C(p)(p(p-1) + C\varepsilon^{-1}p^2)^{p/2} \|\sigma^1(X^{t,x_2,2}) - \sigma^2(X^{t,x_2,2})\|_{H^p}^p.
\]

The desired estimate follows by choosing \(\varepsilon = 1/2\), by applying Grönwall’s inequality and by taking the \(p\)-th root on both sides.

We now prove the inequality (3.2) by assuming that \(\sigma^1 \equiv \sigma^2\) and \(x^1 = x^2 = x\). Let \(\beta := 2\mu_1 + \nu_1^2 + 1\). Applying Itô’s formula to \((e^{-s\beta}|\Delta X_s|^2)_{s \in [t,T]}\) and taking expectations on both sides yield that

\[
\mathbb{E}[e^{-T\beta}|\Delta X_T|^2] \\
\leq \mathbb{E} \left[ \int_t^T (e^{-r\beta}(-\beta|\Delta X_r|^2 + 2(\Delta X_r, b_r^1(X_r^2) - b_r^2(X_r^2)) + |\sigma_r^1(X_r^2) - \sigma_r^2(X_r^2)|^2) dr \right] \\
\leq \mathbb{E} \left[ \int_t^T (e^{-r\beta}(-\beta|\Delta X_r|^2 + 2\mu_1|\Delta X_r|^2 + 2|\Delta X_r||b_r^1(X_r^2) - b_r^2(X_r^2)| + \nu_1^2|\Delta X_r|^2) dr \right] \\
\leq \mathbb{E} \left[ \int_t^T e^{-r\beta}|b_r^1(X_r^2) - b_r^2(X_r^2)|^2 dr \right],
\]

where the second inequality follows from the assumptions of \(b^1\) and \(\sigma^1\), and the last inequality follows from the Cauchy–Schwarz inequality and the fact that \(\beta = 2\mu_1 + \nu_1^2 + 1\).

**Proof of Lemma 3.3.** By (2.8) and (2.9),

\[
\langle y - y', \partial_x H_t(x, a, y, z) - \partial_x H_t(x, a, y', z) \rangle = \langle y - y', \partial_x b_t(x, a)\top(y - y') - \rho(y - y') \rangle \\
= \langle y - y', (\partial_x b_t(x) + \partial_x b_t(x) a)\top(y - y') - \rho(y - y') \rangle \leq (\kappa_b - \rho + L_b)|y - y'|^2.
\]

Moreover, by (2.8), (2.9) and (2.1),

\[
|\partial_x H_t(x, a, y, z) - \partial_x H_t(x', a', y, z')| \\
\leq |\partial_x b_t(x, a) - \partial_x b_t(x', a')||y| + |\text{tr}(\partial_x \sigma_t(x)\top z) - \text{tr}(\partial_x \sigma_t(x')\top z')| + |\partial_x f_t(x, a) - \partial_x f_t(x', a')| \\
\leq (L_b|x - x'| + L_b(|x - x'| + |a - a'|))|y| + L_\sigma|x - x'||z| + L_\sigma|z - z'| + L_{fx}(|x - x'| + |a - a'|).
\]

30
We start by fixing $\alpha, \beta > 0$, then the weak limit of the sequences $(\hat{Z}^t_x, \hat{\phi}_t)_{\varepsilon > 0}$ in $\mathbb{H}^2(t, T; \mathbb{R}^{n \times d})$, where $(\hat{u}^\varepsilon(t))_{\varepsilon > 0}$ is a standard mollification of $u$. Then by Proposition 3.7, $|\partial u| \leq L_Y(\hat{\phi}_1)$, which along with (2.10) leads to the desired result.

**Lemma A.2.** Let $m_{(\alpha, \beta)}$ be defined as in (3.33), where $\alpha$ and $\beta$ are given by (3.16) and (3.33), respectively. Then $m_{(\alpha, \beta)} \to 0$, as $T \to 0$, or $\rho \to \infty$ or $\kappa_0 \to -\infty$.

**Proof.** We start by fixing $T > 0$ and $\kappa_0$ and considering $\rho \to \infty$. Then $\beta$ is fixed and $\alpha$ tends to $-\infty$. By definition of $m_{(\alpha, \beta)}$,

$$m_{(\alpha, \beta)} = \sup_{t \in [0, T]} e^{2\alpha(T-t)} \frac{e^{\beta(T-t)} - 1}{\beta} = \sup_{t \in [0, T]} \frac{e^{(2\alpha+\beta)t} - e^{2\alpha t}}{\beta}.$$
Assume without loss of generality that \( \alpha \) is sufficiently negative such that \( 2\alpha + \beta < 0 \). A straightforward computation shows that the supremum is attained for \( t^* = \frac{1}{\beta} \ln \left( \frac{2\alpha}{2\alpha + \beta} \right) \), and hence
\[
m(\alpha, \beta) = e^{2\alpha} \frac{\ln \left( \frac{2\alpha}{2\alpha + \beta} \right)}{\beta} \leq e^{2\alpha} \frac{\ln \left( \frac{2\alpha}{2\alpha + \beta} \right)}{\beta} - 1 = -e^{2\alpha} \frac{2\alpha}{2\alpha + \beta} \frac{1}{\beta}.
\]

By L’Hospital’s rule, \( \lim_{\alpha \to -\infty} \frac{2\alpha}{\beta} \ln \left( \frac{2\alpha}{2\alpha + \beta} \right) = -1 \), which shows that \( \lim_{\rho \to \infty} m(\alpha, \beta) = 0 \).

Now we fix \( \rho \) and send \( \kappa \to -\infty \) (i.e., \( \beta \to -\infty \)) or \( T \to 0 \). Observe that
\[
m(\alpha, \beta) = \sup_{t \in [0, T]} e^{2\alpha(T-t)} \int_t^T e^{(T-s)\beta} \, ds \leq \sup_{t \in [0, T]} e^{2\alpha(T-t)} \int_0^T e^{(T-s)\beta} \, ds = e^{2\alpha + T} \frac{T\beta}{\beta} - 1,
\]
from which one can easily deduce that \( \lim_{T \to 0} m(\alpha, \beta) = \lim_{\kappa \to -\infty} m(\alpha, \beta) = 0 \).

### Acknowledgements

Wolfgang Stockinger is supported by a special Upper Austrian Government grant.

### References

[1] B. Acciaio, J. Backhoff-Veraguas, and R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM Journal on Control and Optimization*, 57(6):3666–3693, 2019.

[2] V. Barbu and T. Precupanu. *Convexity and optimization in Banach spaces*. Springer Science & Business Media, 2012.

[3] G. Barles and E. R. Jakobsen. On the convergence rate of approximation schemes for hamilton-jacobi-bellman equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 36(1):33–54, 2002.

[4] 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 Journal on Control and Optimization*, 56(6):4017–4044, 2018.

[5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[6] C. Bender and J. Zhang. Time discretization and markovian iteration for coupled FBSDEs. *The Annals of Applied Probability*, 18(1):143–177, 2008.

[7] P. Briand and R. Carmona. BSDEs with polynomial growth generators. *Journal of Applied Mathematics and Stochastic Analysis*, 13(3):207–238, 2000.

[8] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical finance*, 7(1):1–71, 1997.

[9] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476. PMLR, 2018.
[10] H. Gu, X. Guo, X. Wei, and R. Xu. Mean-field multi-agent reinforcement learning: A decentralized network approach. *arXiv preprint arXiv:2108.02731*, 2021.

[11] X. Guo, A. Hu, and J. Zhang. Theoretical guarantees of fictitious discount algorithms for episodic reinforcement learning and global convergence of policy gradient methods. *arXiv preprint arXiv:2109.06362*, 2021.

[12] X. Guo, A. Hu, and Y. Zhang. Reinforcement learning for linear-convex models with jumps via stability analysis of feedback controls. *arXiv preprint arXiv:2104.09311*, 2021.

[13] B. Hambly, R. Xu, and H. Yang. Policy gradient methods for the noisy linear quadratic regulator over a finite horizon. *SIAM Journal on Control and Optimization*, 59(5):3359–3391, 2021.

[14] J. Han and W. E. Deep learning approximation for stochastic control problems. *arXiv preprint arXiv:1611.07422*, 2016.

[15] R. Hu. Deep fictitious play for stochastic differential games. *arXiv preprint arXiv:1903.09376*, 2019.

[16] C. Hure, H. Pham, and X. Warin. Deep backward schemes for high-dimensional nonlinear pdes. *Mathematics of Computation*, 89(324):1547–1579, 2020.

[17] P. Imkeller, G. Dos Reis, and W. Salkeld. Differentiability of sdes with drifts of super-linear growth. *Electronic journal of probability*, 24:1–43, 2019.

[18] 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. *Foundations of Computational Mathematics*, 21(2):331–374, 2021.

[19] Y. Jia and X. Y. Zhou. Policy gradient and actor-critic learning in continuous time and space: Theory and algorithms. *arXiv preprint arXiv:2111.11232*, 2021.

[20] B. Kerimkulov, J.-M. Leahy, D. Šiška, and L. Szpruch. Convergence of policy gradient for entropy regularized MDPs with neural network approximation in the mean-field regime. *arXiv preprint arXiv:2201.07296*, 2022.

[21] B. Kerimkulov, D. Šiška, and L. Szpruch. Exponential convergence and stability of Howard’s policy improvement algorithm for controlled diffusions. *SIAM Journal on Control and Optimization*, 58(3):1314–1340, 2020.

[22] B. Kerimkulov, D. Šiška, and L. Szpruch. A modified MSA for stochastic control problems. *Applied Mathematics & Optimization*, 84(10):3417–3436, 2021.

[23] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Science & Business Media, 2001.

[24] Q. Li, L. Chen, and C. Tai. Maximum principle based algorithms for deep learning. *Journal of Machine Learning Research*, 18:1–29, 2018.

[25] A. Lionnet, G. Dos Reis, and L. Szpruch. Time discretization of FBSDE with polynomial growth drivers and reaction–diffusion PDEs. *The Annals of Applied Probability*, 25(5):2563–2625, 2015.
[26] X. Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

[27] B. Maslowski and P. Veverka. Sufficient stochastic maximum principle for discounted control problem. *Applied Mathematics & Optimization*, 70(2):225–252, 2014.

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

[29] B. S. Mordukhovich. *Variational analysis and generalized differentiation I: Basic theory*, volume 330. Springer Science & Business Media, 2006.

[30] R. Munos. Policy gradient in continuous time. *Journal of Machine Learning Research*, 7:771–791, 2006.

[31] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

[32] 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*, pages 79–127. Springer, 1998.

[33] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

[34] C. Reisinger, W. Stockinger, and Y. Zhang. A fast iterative PDE-based algorithm for feedback controls of nonsmooth mean-field control problems. *arXiv preprint arXiv:2108.06740*, 2021.

[35] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[36] D. Šiška and L. Szpruch. Gradient flows for regularized stochastic control problems. *arXiv preprint arXiv:2006.05956*, 2020.

[37] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[38] L. Szpruch, T. Treetanthiploet, and Y. Zhang. Exploration-exploitation trade-off for continuous-time episodic reinforcement learning with linear-convex models. *arXiv preprint arXiv:2112.10264*, 2021.

[39] C. Tallec, L. Blier, and Y. Ollivier. Making deep q-learning methods robust to time discretization. In *International Conference on Machine Learning*, pages 6096–6104. PMLR, 2019.

[40] L. Wang, Q. Cai, Z. Yang, and Z. Wang. Neural policy gradient methods: Global optimality and rates of convergence. *arXiv preprint arXiv:1909.01150*, 2019.

[41] J. Zhang. *Backward Stochastic Differential Equations*. Springer, 2017.