CONVERGENCE OF MACHINE LEARNING METHODS FOR FEEDBACK CONTROL LAWS: AVERAGED FEEDBACK LEARNING SCHEME AND DATA DRIVEN METHODS

KARL KUNISCH* AND DONATO VÁSQUEZ-VARAS†

Abstract. This work addresses the synthesis of optimal feedback control laws via machine learning. In particular, the *Averaged Feedback Learning Scheme* (AFLS) and a data driven methods are considered. Hypotheses for each method ensuring the convergence of the evaluation of the objective function of the underlying control problem at the obtained feedback-laws towards the optimal value function are provided. These hypotheses are connected to the regularity of the value function and the stability of the dynamics. In the case of AFLS these hypotheses only require Hölder continuity of the value function, whereas for the data driven method the value function must be at least C^2. It is demonstrated that these methods are connected via their optimality conditions. Additionally, numerical experiments are provided by applying both methods to a family control problems, parameterized by a positive real number which controls the regularity of the value function. For small parameters the value function is smooth and in contrast for large parameters it is non-differentiable, but semi-concave. The results of the experiments indicate that both methods have a similar performance for the case that the value function is smooth. On the other hand, if the value function is not differentiable, AFLS has a better performance which is consistent with the obtained convergence results.

Key words. optimal feedback control, Hamilton-Jacobi-Bellman equation, learning approach, learning theory, non-linear system, non-smooth value function.

AMS subject classifications. 49L12, 49J15, 49N35, 68Q32, 93B52,

1. Introduction.

The problem of constructing feedback-laws via machine learning methods has attracted notable attention. This is due to the fact that classical method for synthesizing feedback laws suffer from the *curse of dimensionality.* Among the well-known methods for solving the HJB equation we mention finite difference schemes [10], semi-Lagrangian schemes [21], and policy iteration [3, 7, 47, 49]. The curse of dimensionality can be elevated by machine learning methods, since they are capable of efficiently approximating high dimensional functions. For instance we can mention the following contributions: representation formulas [13, 14, 15, 17], approximating the HJB equation by neural networks [26, 16, 42, 43, 28, 39, 48, 12, 52], data driven approaches [41, 40, 6, 32, 2, 20], max-plus methods [1, 24, 19], polynomial approximation [30, 29], tensor decomposition methods [27, 51, 25, 18, 45, 46], POD methods [5, 35], tree structure algorithms [4], and sparse grids techniques[8, 23, 33, 9], see also the proceedings volume [31]. Frequently, there is no proof concerning the question whether the feedback law constructed via machine learning approximates an optimal control, unless the value function is sufficiently smooth as is the case in [20], where the value functions is supposed to be an element of a reproducing kernel Hilbert space, where an interpolation method is used, or as in [39] and [38] where the convergence of the method is proved for control problems with C^2 value functions.

In the present work we analyze and study the convergence of two approaches and investigate the interconnections between them. The first one consists in finding a feedback law minimizing the average of the cost of the control problem with respect to

*Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences and Institute of Mathematics and Scientific Computing, University of Graz (karl.kunisch@uni-graz.at).

†Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, (donato.vasquez-varas@ricam.oeaw.ac.at.

a set of initial conditions. We call this method *Averaged Feedback Learning Scheme* (AFLS). It has been introduced and used in earlier work, see e.g. [39, 36, 36, 37, 43, 44, 48], but has not been attributed a name before. In the second approach the feedback law is obtained by using a regression or interpolation algorithm. We consider two versions of it. In all cases the feedback control is parameterized by means of the verification theorem.

The convergence of the methods is investigated in terms of the distance between the evaluation of the objective function obtained by the feedback law of the respective method and the value function of the underlying optimal control problem. This is explained in more detail this in Section 3. We are able to prove the convergence of the AFLS approach under the hypothesis of the existence of a stable sequence of consistent feedback laws, see [37]. This in turn depends on the regularity of the value function. For the data driven approach we are able to prove its convergence if the value function is at least C^2. In order to illustrate the dependence of the convergence in dependence of the regularity of the value function, we also study numerically the convergence of the different methods for a family of control problem indexed by a parameter $\gamma > 0$. This problem was first given in [37] where it was proved that its value function is semi-concave for all $\gamma > 0$ and is non-differentiable for γ large enough. Additionally, in the present article we prove that for γ small enough the value function is $C^\infty(\mathbb{R}^d)$. Thus applying the methods for this problem for different values of γ allows us to observe the influence of the regularity on the behavior of the proposed methods.

Let us briefly outline the structure of the paper. The problem statement is given in Section 2. Section 3 contains the description of the methods for its solution, and of the hypotheses needed for the convergence analysis, which is presented in Section 4. Section 5 is devoted to the description of the test problem. It depends on a parameter regulating the regularity of the associated value function. Numerical results are presented in Section 6. The Appendix contains some technical results.

2. Control problem.

We consider the task of finding an approximate optimal feedback-law for the following control problem

$$\min_{u\in L^2((0,\infty);\mathbb{R}^m)} \int_0^\infty \ell(y(t;y_0,u)) + \frac{\beta}{2}|u|^2 dt, \tag{2.1}$$

where $y(\cdot;y_0,u)$ is the solution of the equation

$$y' = f(y) + Bu, \; y(0) = y_0. \tag{2.2}$$

Throughout we assue that $B \in \mathbb{R}^{d\times m}$, that f and ℓ are Lipschitz continuous on bounded subsets of $\mathbb{R}^d$, and $\ell \geqslant 0$.

A classical way to find a feedback-law is based on dynamic programming. For this purpose the value function associated to (2.1) - (2.2) is utilized. It is defined by

$$V(y_0) = \min_{u\in L^2((0,\infty;\mathbb{R}^d))} \int_0^\infty \left(\ell(y(t;y_0,u)) + \frac{\beta}{2}|u(t)|^2\right) dt. \tag{2.3}$$

In the case that V is $C^{1,1}(\mathbb{R}^d)$ it is known that the feedback law given by

$$u(y) = -\frac{1}{\beta}B^\top \nabla V(y) \text{ for } y \in \mathbb{R}^d$$

is optimal, i.e., for $y_0 \in \mathbb{R}^d$, denoting by $y^* \in H^1_{loc}((0,\infty);\mathbb{R}^d)$ the unique solution of

$$y'(t) = f(y(t)) + Bu(y(t)),\ t > 0,\ y(0) = y_0$$

and setting $u^*(t) = u(y^*(t))$ for $t > 0$, we have that u^* is an optimal solution of (2.1). Further, in this case V is a classical solution of the following Hamilton-Jacobi equation

$$\max_{u\in\mathbb{R}^m} \left\{ -\nabla V(y)(f(y) + Bu) - \ell(y) - \frac{\beta}{2}|u|^2 \right\} = 0 \text{ for all } y \in \mathbb{R}^d. \tag{2.4}$$

Hence, for finding an optimal feedback law it is enough to solve (2.4). But there are two main obstacles for using this approach directly. The first one relates to the fact that the value functions is frequently merely locally Lipschitz continuous, and the second one is that solving (2.4) suffers from the curse of dimensionality. As was mentioned in the introduction, for this reason many authors have tried to overcome this difficulty by using machine learning techniques.

Problem (2.1)-(2.2) is posed over the infinite time horizon, and consequently special attention has to be paid to the well-posedness of this problem. Above we have not made a concrete choice for the running cost ℓ. If $\ell(y) = \frac{1}{2}|y|^2$, then (2.1)-(2.2) relates to optimal stabilization and appropriate stabilizability assumptions are required. Alternatively we can think of a running cost containing a discount factor. These aspects have received much attention in the control literature and we shall not focus on them here. Rather we make assumptions which guarantee that (2.1)-(2.2) admits a control u in feedback form such that the cost in (2.1) is finite for all initial conditions y_0 in a set $\omega \subset \mathbb{R}^d$.

3. Methods for the construction of approximating optimal feedback laws.

For enunciating precisely our goal and main results, we need first some definitions and hypotheses, which were previously introduced in [37, 38, 36]. Throughout let Ω be an open and bounded subset of $\mathbb{R}^d$ and let $\omega \subset \mathbb{R}^d$ be another open set with its closure contained in Ω.

DEFINITION 3.1. *For $u \in W^{1,\infty}_{loc}(\Omega;\mathbb{R}^m)$, $y_0 \in \Omega$, and $T \in (0,\infty]$ the function $y(\cdot;y_0,u) \in W^{1,\infty}(0,T;\mathbb{R}^d)$ is called the solution to (2.2) on $[0,T]$ if it is the only function satisfying*

$$y'(t) = f(y(t)) + Bu(y(t)),\ y(0) = y_0, \quad \textit{for } t \in [0,T]. \tag{3.1}$$

Frequently the function u in the above definition in the one associated to the verification theorem. For this purpose we introduce the following notation.

DEFINITION 3.2. *For $v \in C^{1,1}(\Omega)$ the mapping $u_v \in W^{1,\infty}_{loc}(\Omega;\mathbb{R}^d)$ is defined as*

$$u_v(y) = -\frac{1}{\beta}B^\top \nabla v(y) \textit{ for } y \in \Omega. \tag{3.2}$$

If u_v is evaluated along a trajectory $y(t)$ of (2.2) we refer to it as feedback-law. For $\delta > 0$ we further introduce

$$\Omega_\delta := \{x \in \Omega : \min_{z\in\partial\Omega} |x - z| > \delta\}. \tag{3.3}$$

.

Hypothesis 3.3. *For the tuple (T, u, δ, y_0) with $T > 0$, $u \in W^{1,\infty}_{loc}(\Omega; \mathbb{R}^m)$, $\delta > 0$, and $y_0 \in \Omega_\delta$, the solution $y(\cdot; y_0, u)$ to (2.2) exists on $[0, T]$ and $y(t; y_0, u) \in \Omega_\delta$ for all $t \in [0, T]$.*

If Hypothesis 3.3 is assumed for all $T > 0$, then it amounts to a stability assumption for (2.2) under the control u. Under a global Lipschitz condition on f and assumptions on V this assumption was addressed in [39, Proposition 3.5]. It can also be established by Ljapunov function techniques. For this purpose we consider the alternative Hypothesis 3.4 below and in the subsequent remark we discuss that it implies Hypothesis 3.3.

Hypothesis 3.4. *There exist $\tilde{\delta} > 0$, and $w \in C^1(\Omega)$ such that for*

$$\omega_{\tilde{\delta}} := \{y \in \Omega : w(y) < \sup_{y_0 \in \omega} w(y_0) + \tilde{\delta}\},$$

we have that $\omega \subset \omega_{\tilde{\delta}}$, $\overline{\omega_{\tilde{\delta}}} \subset \Omega$, and $\partial\omega_{\tilde{\delta}}$ is of class C^1. Moreover $\phi \in C(\Omega)$ is a viscosity super solution of

$$-\nabla w(y)^\top (f(y) + B(y) u_\phi(y)) = 0 \text{ in } \omega_{\tilde{\delta}},$$

i.e. for every $\bar{y} \in \omega_{\tilde{\delta}}$ and every $h \in C^1_{loc}(\Omega)$ such that $\phi - h$ attains a local minimum at $\bar{y}$ the following inequality holds

$$\nabla w(\bar{y})^\top (f(\bar{y}) + B(\bar{y}) u_h(\bar{y})) \leqslant 0. \tag{3.4}$$

Remark 3.1. *Let us consider $\tilde{\delta} > 0$, $w \in C^1$, and $\phi \in C^{1,1}(\Omega)$ as in Hypothesis 3.4. Under these conditions w is a classical Ljapunov function, see e.g. [34], and (3.4) is satisfied with $u_h = u_\phi$. Consequently we obtain that the trajectories originating from $y_0 \in \omega \subset \omega_{\tilde{\delta}}$ exist for all time $t \geqslant 0$ and cannot escape from $\overline{\omega_{\tilde{\delta}}}$. Let us set 2δ as the distance of $\overline{\omega_{\tilde{\delta}}}$ to $\partial\Omega$, so that $\overline{\omega_{\tilde{\delta}}} \subset \Omega_\delta$. Then for all $y_0 \in \omega$ and all $T > 0$ the tuple (T, u_ϕ, δ, y_0) satisfies Hypothesis 3.3, since $\omega_{\tilde{\delta}}$ is strictly included in Ω_δ.*

Next we introduce the functional $\mathcal{V}$ which evaluates the cost functional along the trajectory under the control u:

Definition 3.5. *Let (T, u, δ, y_0) be a tuple with $T > 0$, $u \in W^{1,\infty}_{loc}(\Omega; \mathbb{R}^m)$, $\delta > 0$, and $y_0 \in \Omega_\delta$, and let Hypothesis 3.3 hold. We define $\mathcal{V}_T(y_0, u)$ as*

$$\mathcal{V}_T(y_0, u) = \int_0^T \left(\ell(y(t; y_0, u)) + \frac{\beta}{2} |u(y(t; y_0, u))|^2 \right) dt.$$

Of course we always have $(T, u, \delta, y_0) \geqslant V(y_0)$, provided the two functionals are well-defined.

We study three methods with the aim of constructing an approximately optimal feedback law. To explain them we consider a finite time horizon $T > 0$, a finite dimensional Banach space $(\Theta, \|\cdot\|_\Theta)$, a continuous and coercive function $\mathcal{P} : \Theta \mapsto \mathbb{R}^+$, a penalty coefficient $\alpha > 0$, and a function $v : \Theta \mapsto C^2(\Omega)$. We will name the tuple $S = ((\Theta, \|\cdot\|_\Theta), \mathcal{P}, \alpha, v)$ the **setting** of the learning problems. We define the methods for the finite time horizon T. Subsequently will shall use these methods for a sequence of horizons T_n with $\lim_{n\to\infty} T_n = \infty$ to obtain an approximate feedback law for the infinite horizon problem (2.1)-(2.2). Note that involving finite horizon problems is natural for numerical realisations.

The function $v(\theta)$ will play the role of the approximation to the value function V which will henceforth be assumed to exist at least as an element in $L^1(\Omega)$.

The first method, which we refer to as Averaged Feedback Learning Scheme (AFLS), consists in minimizing the average cost of (2.1) plus a penalization term:

Method 3.6. *For a time horizon $T > 0$ and a setting $S = ((\Theta, \|\cdot\|_\Theta), \mathcal{P}, \alpha, v)$, we solve*

$$\min_{\theta \in \mathcal{O}_{T,\Theta}} \frac{1}{|\omega|} \int_\omega \mathcal{V}_T(y_0, u_{v(\theta)}) dy_0 + \alpha \mathcal{P}(\theta) \tag{3.5}$$

where

$$\mathcal{O}_{T,\Theta} := \{\theta \in \Theta : \exists \delta > 0 \text{ such that } u_{v_{(\theta)}} \text{ satifies Hypothesis 3.3 for all } y_0 \in \omega\}.$$

The second method consist in finding $\theta \in \Theta$ such that it minimizes the L^2 distance with respect to the optimal feedback controls along the trajectories originating from ω:

Method 3.7. *For $V \in C^1(\Omega)$, a given time horizon $T > 0$, and a setting $S = ((\Theta, \|\cdot\|_\Theta), \mathcal{P}, \alpha, v)$ we solve*

$$\min_{\theta \in \Theta} \frac{1}{2\beta|\omega|} \int_\omega \int_0^T |B^\top (\nabla v(\theta)(y^*(t; y_0)) - \nabla V(y^*(t; y_0)))|^2 dt dy_0 + \alpha \mathcal{P}(\theta), \tag{3.6}$$

where $y^* \in W^{1,\infty}((0,T) \times \omega; \mathbb{R}^d)$ is such that for almost all $y_0 \in \omega$, the control given by $u(t) = -\frac{1}{\beta} B^\top \nabla V(y^*(y_0, t))$ is an optimal solution of (2.1).

In a similar way, we consider the problem of minimizing the optimal feedback controls over the whole domain Ω instead of doing it along the trajectories. This leads to the following formulation:

Method 3.8. *For $V \in W^{1,2}(\Omega)$ and a given setting $S = ((\Theta, \|\cdot\|_\Theta), \mathcal{P}, \alpha, v)$ we solve*

$$\min_{\theta \in \Theta} \frac{1}{2\beta|\Omega|} \int_\Omega \left|B^\top \left(\nabla v(\theta)(y_0) - \nabla V(y_0)\right)\right|^2 dy_0 + \alpha \mathcal{P}(\theta). \tag{3.7}$$

Each of the methods is investigated for a sequence $S_n = ((\Theta_n, \|\cdot\|_n), \mathcal{P}_n, \alpha_n, v_n)$ of settings replacing S in the statements of these methods, with Θ_n finite dimensional, and a sequence of time horizons $T_n \to \infty$. Appropriate conditions will be established on $S_n = ((\Theta_n, \|\cdot\|_n), \mathcal{P}_n, \alpha_n, v_n)$ to allow an asymptotic analysis as $n \to \infty$.

Observe that Method 3.6 and Method 3.7 when applied to a setting S_n provide a solution $v_n(\theta_n^*)$ which involves initial conditions in ω. But $v_n(\theta_n^*)$ is also defined on all of Ω and we hope that it provides an approximation to the sought value function also outside of ω. In practice, of course, initial conditions will not be chosen in all of ω, but only at discrete sample points. These comments refer to the topic of generalization in the machine learning context. For a specific case of a feedback control problems, this was investigated in [38, Section 8].

To address the problem of approximating v in Methods 3.6-3.8 by sequences of finite dimensional settings $S_n = ((\Theta_n, \|\cdot\|_n), \mathcal{P}_n, \alpha_n, v_n)$ we require additional hypotheses which are explained next. This will need some patience on the side of the reader.

Let us consider sequences of time horizons $T_n > 0$ with $\lim_{n\to\infty} T_n = \infty$, of finite dimensional Bachach spaces $\{(\Theta_n, \|\cdot\|_n\}_{n \in \mathbb{N}}$, coercive and continuous penalty operators $\mathcal{P}_n : \Theta_n \mapsto \mathbb{R}_+$, functions $v_n \in C^1(\Theta_n; C^2(\Omega))$, and penalty parameters

$\alpha_n > 0$. Let $\theta_n^* \in \Theta_n$ be obtained by solving any of the three problems introduced above. Our goal then consists in finding assumptions which imply that

$$\lim_{n\to\infty} \left\| \mathcal{V}_{T_n}(\cdot, u_{v_n(\theta_n^*)}) - V \right\|_{L^p(\omega)} = 0, \tag{3.8}$$

for $p \in [1,\infty]$. For this, we will use the following hypotheses:

Hypothesis 3.9. *We have $V \in L^1(\Omega)$ and there exist an increasing sequence T_n of time horizons, a constant $\delta \in (0, dist(\omega, \partial\Omega))$, and a sequence $V_n \in C^2(\Omega)$ such that $\lim_{n\to\infty} T_n = \infty$, for all $y_0 \in \omega$ the tuple $(T_n, u_{V_n}, \delta, y_0)$ satisfies Hypothesis 3.3 with $T = T_n$, and*

$$\limsup_{n\to\infty} \int_\omega (\mathcal{V}_{\hat{T}_n}(y_0, u_{V_n}) - V(y_0)) dy_0 = 0. \tag{3.9}$$

Note that (3.8) involves the candidate for an approximately optimal feedback $u_{v_n(\theta_n^*)}$ while in (3.9) the feedback u_{V_n} involves the approximation V_n of the value function, which is supposed to exist. Thus this hypothesis serves the purpose of guaranteeing that V can in principle be approximated by controls of the form (3.2) with $v = V_n$ and utilized in the evaluation of $\mathcal{V}_{\hat{T}_n}(y_0, u_{V_n})$. We shall discuss the feasibility of (3.9) in Remark 4.2 right after Corollary 4.4 in Section 4. For this purpose we shall require that V is $\alpha-$Hölder continuous with $\alpha > \frac{1}{2}$.

The following hypothesis one will ensure the possibility to approximate these smooth functions by finitely parameterized functions, a property which is required for numerical realizations.

Hypothesis 3.10. *For each $g \in C^1(\Omega)$ there exist $\theta_n \in \Theta_n$ and a sequence of penalty coefficients $\alpha_n > 0$ with $\lim_{n\to\infty} \alpha_n = 0$ such that*

$$\lim_{n\to\infty} v_n(\theta_n) = g \text{ in } C^1_{loc}(\Omega) \text{ and } \lim_{n\to\infty} \alpha_n \mathcal{P}_n(\theta_n) = 0. \tag{3.10}$$

Hypotheses 3.9 and 3.10, and a proper choice of the time horizons T_n will imply the convergence of Method 3.6 in the sense of (3.8).

A more restrictive version of Hypothesis 3.10, requiring higher regularity of the value function, will allow us to prove that Methods 3.7 and 3.8 converge in the same manner as specified for Method 3.6 above. We first state this hypothesis and in the following remark we provide an illustrating example.

Hypothesis 3.11. *There exists a Banach space $(\Theta, \|\cdot\|_\Theta)$ such that $V \in \Theta \cap C^2(\Omega)$, Θ is compactly embedded in $C^1(\overline{\Omega})$ and there exists $\delta > 0$ such that for all $y_0 \in \omega$ the tuple (T, u_V, δ, y_0) satisfies Hypothesis 3.3 with $T = \infty$. Further, there exist sequences $\{(\Theta_n, \|\cdot\|_n)\}_{n\in\mathbb{N}}$ of finite dimensional Banach spaces, $v_n : \Theta_n \mapsto \Theta$ of continuous functions, and $P_n : \Theta_n \mapsto \mathbb{R}_+$ of continuous functions, such that for all $g \in \Theta$ there exists a sequence $\theta_n \in \Theta_n$ satisfying*

$$\lim_{n\to\infty} \|v_n(\theta_n) - g\|_\Theta = 0, \ \sup_{n\in\mathbb{N}} \mathcal{P}_n(\theta_n) < \infty.$$

Moreover there exist constants $C > 0$ and $\sigma > 0$ such that it holds

$$\|v_n(\theta)\|_\Theta \leqslant C\mathcal{P}_n(\theta)^\sigma, \text{ for all } n \in \mathbb{N} \text{ and } \theta \in \Theta_n.$$

Remark 3.2. *In order to illustrate the notation and give a situation where Hypothesis 3.11 holds, we consider the case of polynomials as ansatz functions. For* *the sake of simplicity we take $\Omega = (-1,1)^d$. For $k \in \mathbb{N}$ consider $H^k([-1,1])$ with the inner product $\langle \phi, \psi \rangle = \frac{1}{2}\sum_{i=0}^{k} \int_{-1}^{1} \phi^{(i)}(x)\psi^{(i)}(x)dx$ where ϕ^i stands for the i-th derivative of ϕ. Let $\{\phi_i^k\}_{i=1}^{\infty}$ be a orthonormal basis of $H^k((-1,1))$ obtained by applying the Gram-Schmidt process to the set of monomials $\{1, x, x^2, \ldots\}$. We set $\Theta = H^k_{mix}((-1,1)^d) := \otimes_{i=1}^{d} H^k((-1,1))$, i.e, Θ is the tonsorial product of d copies of $H^k((-1,1))$. The space $H^k_{mix}((-1,1)^d)$ is Hilbert with the tensorial inner product. For a multi-index $\alpha \in \mathbb{N}^d$ we denote $\phi_\alpha(y) := \prod_{i=1}^{d} \phi_i(y_i)^{\alpha_i}$. Hence the set $\{\phi_\alpha\}_{\alpha \in \mathbb{N}^d}$ is a basis for Θ. For $n \in \mathbb{N}$ we define $\Lambda_n \subset \mathbb{N}^d$ as the follows:*

$$\Lambda_n = \{\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d : \max_{i=1,\ldots,d} \alpha_i \leqslant n\}.$$

Let us consider an enumeration of Λ_n given by $\{\alpha^i\}_{i=1}^{|\Lambda_n|}$. We set $\Theta_n = \mathbb{R}^{|\Lambda_n|}$, $v_n(\theta) = \sum_{i=1}^{|\Lambda_n|} \theta_i \phi_{\alpha^i}$, $\|\theta\|_n = |\theta|$, and $\mathcal{P}_n(\theta) = \frac{1}{2}|\theta|^2$. With these choices Hypothesis 3.11 is met with $\sigma = \frac{1}{2}$ and $k \geqslant 3$, since $H^k_{min}((-1,1)^d)$ is continuously embedded in $W^{2,\infty}((-1,1)^d)$ (see [50]).

*For high-dimensional problems (i.e. large d) the **hyperbolic-cross** index set is a more efficient choice than Λ_n. This set of multi-indexes is given by*

$$\Gamma_n = \{\alpha = (\alpha_1, \ldots, \alpha_d) \in N^d : \prod_{i=1}^{d} (\alpha_i + 1) \leqslant n\}.$$

For $\Omega = \otimes_{i=1}^{d} (a_i, b_i)$ with $a_i < b_i$ for $i \in \{1, \ldots, d\}$ a change of variables is needed, that is, for $x \in \Omega$ we define $\tilde{\phi}_\alpha(x) = \prod_{i=1}^{d} \phi_{\alpha_i}\left(\frac{2}{(b_i - a_i)}\left(x_i - \frac{a_i+b_i}{2}\right)\right)$ and for $\theta \in \mathbb{R}^{|\Lambda_n|}$ we set $\tilde{v}_n(\theta) = \sum_{i=1}^{n} \theta_i \tilde{\phi}_{\alpha^i}$. With reference to Hypothesis 3.11 we choose $\Theta_n = \mathbb{R}^{|\Lambda_n|}$ and $\Theta = H^k_{mix}((-1,1)^d)$ as above. Again the embedding of Θ into $C^1(\bar{\Omega})$ is compact for $k \geqslant 3$.

In a more general perspective, Hypothesis 3.11 also encsompasses other interesting cases such as neural networks with Θ being the respective Barron space provided the activation function is smooth, or reproducing kernel Hilbert spaces with a C^2 kernel and Θ as the corresponding native space. In such cases Θ can then be chosen as the space of limits of functions of the form $v_n(\theta_n)$ for some $\theta_n \in \Theta_n$, endowed with the proper norm.

REMARK 3.3. *Note that in contrast to Method 3.6, knowledge of the value function is required for Method 3.7 and Method 3.8. In practice this can be realized by approximating the integrals by finite sums or by Monte Carlo methods with values for ∇V at sample points obtained by open loop control solves. This approach was investigated in [6] for instance. Recall the ∇V is related to the adjoint state associated to the control problem (3.1)-(2.2), see [6]. It is therefore obtained as a by-product in every numerical approach for open loop optimal control. But, as already pointed out above, approximating the feedback over the smaller set ω by globally defined functions also serves the purpose to utilize the approximating feedback laws outside of ω. This* generalization step*, of course, requires additional analysis and will depend on the underlying dynamics characterized by f.*

Clearly Method 3.7 and Method 3.8 are more closely related to each other, than either of them to Method 3.6. In practice data generation will be cheaper for Method 3.7 than for Method 3.8 since for the former, information is gathered all along the trajectory, and hence in the context of sampling fewer data points (initial conditions) will be necessary for Method 3.7 than for Method 3.8. Analytically Method 3.7 requires

C^1-regularity V and hence it required more regularity of the minimal value function than is needed for Method 3.8.

Remark 3.4. *Methods 3.7 and 3.8 minimize the L^2 distance with respect to the optimal feedback-law. While the first one explicitly involves information along all of the optimal trajectories, the second one utilizes the whole domain Ω. Next we provide an interesting relation between these two formulations. Namely, assuming that $\frac{1}{|det(D_{y_0}y^*(\cdot,\cdot))|} \in L^\infty((0,T)\times\omega)$, setting $\Gamma_T = \{z \in \mathbb{R}^d : \exists (t, y_0) \in [0,T]\times\omega, y^*(t,y_0) = z\}$, and defining $y^0(t,z)$ for $(t,z) \in [0,T]\times\Gamma_T$ as the inverse function of y^* with respect to its second argument, we obtain by means of the change of variables $z = y^*(t,y_0)$ that*

$$\int_\omega \int_0^T |B^\top(\nabla v(\theta)(y^*(t,y_0)) - \nabla V(y^*(t,y_0)))|^2 dt dy_0$$
$$= \int_{\Gamma_T} |B^\top(\nabla v(\theta)(z) - \nabla V(z))|^2 g(z) dz$$

where

$$g(z) = \int_{\{t\in[0,T]:\ \exists y_0\in\omega,\ y^*(t;y_0)=z\}} \frac{1}{|det(D_{y_0}y^*(t,y^0(t,z)))|} dt,$$

where $D_{y_0}y^*(\cdot,\cdot)$ denotes the derivative of y^* with respect to its second argument. This reflects that the main difference between Problems 3.7 and 3.8 consists in how they weigh the initial conditions. We also note that as consequence of classical ODE theory the hypotheses on $det(D_{y_0}y^*)$ expressed above hold if $V \in C^2(\Omega)$.

4. Convergence analysis.

This section is devoted to the convergence analysis for Methods 3.6 - 3.8. We first state the two main results and then provide their proofs below. For Method 3.6 we have the following convergence property.

Theorem 4.1. *Let Hypotheses 3.9 and 3.10 hold for sequences $(\Theta_n, \|\cdot\|_n)$, $\mathcal{P}_n$, v_n and T_n. Then there exist a sequence $\alpha_n > 0$ and a sub-sequence $((\Theta_{k_n}, \|\cdot\|_{k_n}), \mathcal{P}_{k_n}, v_{k_n})$ for which Method 3.6 with setting $S_n = ((\Theta_{k_n}, \|\cdot\|_{k_n}), \mathcal{P}_{k_n}, \alpha_n, v_{k_n})$ has at least one solution. Further, for every sequence of solutions $\theta_n^* \in \Theta_{k_n}$ to Method 3.6 with $S = S_n$ we have*

$$\lim_{n\to\infty} \mathcal{V}_{T_n}(\cdot, u_{v_{k_n}(\theta_n^*)}) = V \text{ in } L^1(\omega) \text{ and } \lim_{n\to\infty} \alpha_n \mathcal{P}_{k_n}(\theta_n^*) = 0.$$

Remark 4.1. *In the case of a nested setting, that is, if $\Theta_n \subset \Theta_{n+1}$ and $\mathcal{P}_{n+1}(\theta) = \mathcal{P}_n(\theta)$ for $\theta \in \Theta_n$, it is possible to get the convergence result for the whole sequence of settings and not only for a sub-sequence of them by associating the time horizons and penalty coefficients in an appropriate manner. For instance, for $j \in \mathbb{N}$ let us consider n_j the smallest $n \in \mathbb{N}$ such that $k_n \leqslant j \leqslant k_{k+1}$. Then the sequence of settings $\tilde{S}_j = ((\Theta_j, \|\cdot\|_j \mathcal{P}_j, \alpha_{n_j}, v_j)$ and the sequence of times horizons $\tilde{T}_j = T_{n_j}$ satisfy*

$$\lim_{j\to\infty} \mathcal{V}_{\tilde{T}_j}(\cdot, u_{v_n(\tilde{\theta}_j*)}) = V \text{ in } L^1(\omega) \text{ and } \lim_{n\to\infty} \alpha_{n_j} \mathcal{P}_j(\tilde{\theta}_j^*) = 0,$$

where $\tilde{\theta}_j^* \in \Theta_j$ is the solution to Method 3.6 with $S = \tilde{S}_j$.

In order to prove the convergence for Method 3.8, we shall need to assume Hypothesis 3.11 instead of Hypotheses 3.9 and 3.10. In this case, the convergence holds in $L^\infty(\omega)$. The regularity assumption on V expressed in Hypothesis 3.11, however, is more restrictive than the hypotheses for Theorem 4.1 where the assumptions on

V are only those involved in Hypothesis 3.9 . We shall see that the convergence for Method 3.8 also implies the convergence for Method 3.7, by using a change of variables as was already mentioned in Remark 3.4.

THEOREM 4.2. *Let V, $(\Theta, \|\cdot\|_\Theta)$, and a sequence $S_n = ((\Theta_n, \|\cdot\|_n), \mathcal{P}_n, \alpha_n, v_n)$ be such that Hypothesis 3.11 holds. Further assume that $\lim_{n\to\infty} \alpha_n = 0$ and there exists a constant $C > 0$ such that $\left\|B^\top(\nabla v_n(\theta_n) - V)\right\|^2_{L^2(\Omega;\mathbb{R}^m)} \leqslant C\alpha_n$ where $\theta_n \in \Theta_n$ is as in Hypothesis 3.11 with $g = V$. Then there exists a sequence of times $T_n > 0$ for which every sequence $\theta_n^* \in \overline{\mathcal{O}_{T_n}}$ of solutions of either Method 3.7 or Method 3.8 with $S = S_n$ and $T = T_n$ satisfies that $\theta_n^* \in \mathcal{O}_{T_n}$ and*

$$\lim_{n\to\infty} \mathcal{V}_{T_n}(\cdot, u_{v_n(\theta_n^*)}) = V \ in\ L^\infty(\omega).$$

To verify these two results we start by proving the equivalence between (3.9), and the following two properties

$$\lim_{n\to\infty} \mathcal{V}_{\hat{T}_n}(\cdot, u_{V_n}) = V \text{ in } L^1(\omega), \tag{4.1}$$

and

$$\lim_{n\to\infty} \int_\omega (\mathcal{V}_{\hat{T}_n}(y_0, u_{V_n}) - V(y_0))^+ dy_0 = 0. \tag{4.2}$$

LEMMA 4.3. *Let $T_n > 0$ be an increasing sequence with $\lim_{n\to\infty} T_n = \infty$, let δ be a positive constant, and let $V_n \in C^2(\overline{\Omega})$ be as sequence, such that the tuple $(T_n, u_{V_n}, \delta, y_0)$ satisfies Hypothesis 3.3 for all $y_0 \in \omega$ and $n \in \mathbb{N}$. Then for all $y_0 \in \omega$ the following identity holds*

$$\lim_{n\to\infty} (V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ = 0. \tag{4.3}$$

Proof. By contradiction, let us assume that there exists y_0 such that (4.3) does not hold. This implies that there exist $\kappa > 0$ and a sub-sequence still denoted by V_n such that

$$(V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ > \kappa \tag{4.4}$$

for all $n \in \mathbb{N}$. Here r^+ stands for the positive part of $r \in \mathbb{R}$. Let us denote $y_n = y(t; y_0, u_{V_n})$ and $u_n = u_{V_n} \circ y_n = -\frac{1}{\beta} B^\top \nabla V_n(y_n)$. By Hypothesis 3.3 and the boundedness of Ω we have that $\{\|y_n\|_{L^\infty((0,T_n);\mathbb{R}^d)}\}_{n=1}^\infty$ is bounded. By (4.4) and the definition of u_n we have that for all $n \in \mathbb{N}$

$$\int_0^{T_n} \left(\ell(y_n(t)) + \frac{\beta}{2}|u_n(t)|^2\right) dt \leqslant V(y_0) - \kappa. \tag{4.5}$$

This implies that $\{\|u_n\|_{L^2((0,T_n);\mathbb{R}^m)}\}_{n=1}^\infty$ is bounded. This, together with the fact that y_n is a solution of

$$\frac{d}{dt} y_n(t) = f(y_n(t)) + Bu_n(t),\ t \in (0, T_n),\ y_n(0) = y_0, \tag{4.6}$$

that f is Lipschitz continuous on bounded sets, and that $\{y_n\}_{n=1}^\infty$ is bounded in $L^\infty((0,T_n);\mathbb{R}^d)$, we obtain that $\{y_n\}_{n=1}^\infty$ is uniformly bounded in $H^1((0,T_n);\mathbb{R}^d)$. Therefore, by a diagonalization argument, with respect to $\{T_n\}_{n\in\mathbb{N}}$ and $\{(y_n, u_n)\}_{n\in\mathbb{N}}$, there exists $y^* \in H^1_{loc}((0,\infty);\mathbb{R}^d)$ and $u \in L^2_{loc}((0,\infty);\mathbb{R}^m)$ such that by passing to a sub-sequence, if needed, we have

$$\begin{aligned} &u_n \rightharpoonup u^* \text{ in } L^2_{loc}((0,\infty);\mathbb{R}^m),\ y_n \rightharpoonup y^* \text{ in } H^1_{loc}((0,\infty);\mathbb{R}^d), \\ &\text{and } y_n \to y^* \text{ in } C_{loc}([0,\infty);\mathbb{R}^d), \end{aligned} \tag{4.7}$$

as n tends to infinity. Using these facts together with (4.6) we have that

$$\frac{d}{dt}y^*(t) = f(y^*(t)) + Bu^*(t),\ t \in (0,\infty),\ y^*(0) = y_0. \tag{4.8}$$

Next, let $T > 0$ be a fixed time horizon and choose $n_0 \in \mathbb{N}$ such that for all $n > n_0$ we have $T < T_n$. Hence by (4.5) we get that for all $n > n_0$

$$\int_0^T \left(\ell(y_n(t)) + \frac{\beta}{2}|u_n(t)|^2\right) dt \leqslant V(y_0) - \kappa. \tag{4.9}$$

Using (4.7), the continuity of ℓ and the weak lower semi-continuity of the norm of $L^2((0,T);\mathbb{R}^m)$, letting n tend to infinity in (4.9) we obtain

$$\int_0^T \left(\ell(y^*(t)) + \frac{\beta}{2}|u^*(t)|^2\right) dt \leqslant V(y_0) - \kappa. \tag{4.10}$$

Since this holds for any arbitrary T we get that $u^* \in L^2((0,\infty);\mathbb{R}^d)$ and

$$\int_0^\infty \left(\ell(y^*(t)) + \frac{\beta}{2}|u^*(t)|^2\right) dt \leqslant V(y_0) - \kappa,$$

which contradicts the fact that V is the value function. Consequently, we have that for all $y_0 \in \omega$

$$\lim_{n\to\infty} (V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ = 0. \tag{4.11}$$

□

Corollary 4.4. *Under the hypotheses of the previous lemma* (3.9), (4.1) *and* (4.2) *are equivalent.*

Proof. Since $(V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ \leqslant V(y_0)$ we can apply the dominated convergence theorem together with (4.3) to get that

$$\lim_{n\to\infty} \int_\omega (V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ dy_0 = 0. \tag{4.12}$$

Rewriting the $L^1(\omega)$ norm of $V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n})$ we find that

$$\begin{aligned} &\int_\omega |V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n})| dy_0 \\ &= \int_\omega (\mathcal{V}_{T_n}(y_0, u_{V_n})) - V(y_0))^+ dy_0 + \int_\omega (V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ dy_0. \end{aligned}$$

Since the convergence to 0 of the last term in the previous inequality is always granted under the hypotheses, we get the equivalence of (4.1) and (4.2). Further, condition

(4.1) clearly implies (3.9). To conclude we only need to prove that (3.9) implies either (4.1) or (4.2). For this purpose we note that

$$\begin{aligned} 0 \quad &\leqslant \liminf \int_\omega (\mathcal{V}_{T_n}(y_0, u_{V_n}) - V(y_0))^+ dy_0 \\ &\leqslant \lim \int_\omega (V(y_0) - \mathcal{V}_{T_n}(y_0, u_{V_n}))^+ dy_0 + \limsup \int_\omega (\mathcal{V}_{T_n}(y_0, u_{V_n}) - V(y_0)) dy_0. \end{aligned}$$

Hence, if (3.9) holds, we can combine it with (4.12) to get (4.2). □

We are now prepared to discuss the feasibility of Hypothesis 3.9.

Remark 4.2. *Concerning the feasibility of Hypothesis 3.9, we assume that Hypothesis 3.4 holds and that V is $\alpha-$Hölder continuous with $\alpha > \frac{1}{2}$. Then we can use [37, Lemma 7.1c, Theorem 5.4], with $\phi = V \in C(\Omega)$, and choose $\delta > 0$, to construct sequences $\{u_{V_n}\}_{n\in\mathbb{N}}$ in $C^2(\Omega)$ and $\{T_n\}_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} T_n = \infty$ such that the tuple $(T_n, u_{V_n}, \delta, y_0)$ satisfies Hypothesis 3.3 for all $y_0 \in \omega$ and all n. This is the first requirement in Hypothesis 3.9.*

Regarding the second part, using again [37, Theorem 5.4] the tuple $(T_n, u_{V_n}, \delta, y_0)$ can be chosen such that

$$0 \leqslant \int_\omega \big(\mathcal{V}_{T_n}(\cdot, u_{V_n}) + V(y(T_n, u_{V_n}; \cdot) - V(\cdot)\big) dy_0 \to 0 \text{ for } n \to \infty. \tag{4.13}$$

Since $V(\cdot) \geqslant 0$ this implies that $0 \leqslant \int_\omega \big(\mathcal{V}_{T_n}(\cdot, u_{V_n}) - V(\cdot)\big)^+ dy_0 \to 0$ for $n \to \infty$. This together with (4.12) implies that $\lim_{n\to\infty} \int_\omega (\mathcal{V}_{T_n}(y_0, u_{V_n}) - V(y_0)) dy_0 = 0$, as required.

Before continuing we need to establish the following technical result concerning the stability of the solution to (3.1) if u is of the form (3.2) and perturbations occur in v. Its proof uses Lemma A.1.

Corollary 4.5. *Let $T > 0$ be a time horizon, $y_0 \in \omega$, $g_1 \in C^2(\Omega)$, $g_2 \in C^2(\Omega)$ and $\delta > 0$, be such that $(T, u_{g_2}, \delta, y_0)$ satisfies Hypothesis 3.3. Assume that*

$$\frac{|B| \, \|u_{g_1} - u_{g_2}\|_{L^\infty(\Omega_{\frac{\delta}{4}}; \mathbb{R}^m)}}{a} \left(e^{Ta} - 1\right) \leqslant \frac{\delta}{2}, \tag{4.14}$$

where

$$a = \|Df\|_{L^\infty(\Omega_{\frac{\delta}{4}}; \mathbb{R}^{d\times d})} + \frac{|B|^2}{\beta} \left\|\nabla^2 g_2\right\|_{L^\infty(\Omega_{\frac{\delta}{4}}; \mathbb{R}^{d\times d})}, \tag{4.15}$$

and Df stands for the derivative of f. Then $(T, u_{g_1}, \frac{\delta}{2}, y_0)$ satisfies Hypothesis 3.3.

Proof. For abbreviation we set $y_1(t) = y(t; y_0, u_{g_1})$ and $y_2(t) = y(t; y_0, u_{g_2})$ for all $t \geqslant 0$ such that these expressions are well defined. Let $\hat{T} > 0$ be the escape time of y_1 from $\Omega_{\frac{\delta}{2}}$, i.e. $\hat{T} = \sup\{\tilde{T} > 0 : y_1(t) \in \Omega_{\frac{\delta}{2}}$ for all $t \in [0, \tilde{T}]\}$. We need to prove that $\hat{T} \geqslant T$. By contradiction, assume that $\hat{T} < T$. Then there exists $\tilde{T} \in (\hat{T}, T)$ such that for all $t \in (\hat{T}, \tilde{T}]$ we have $y_1(t) \in \Omega_{\frac{\delta}{4}} \setminus \Omega_{\frac{\delta}{2}}$. Combining (A.1) (with δ replaced by $\frac{\delta}{4}$) from the Appendix, which establishes the Lipschitz continuous dependence of y on u_g, and (4.14) we obtain for all $t \in (\hat{T}, \tilde{T}]$ that

$$|y_1(t) - y_2(t)| \leqslant \frac{\delta}{2}.$$

Here we use that if $(T, u_{g_2}, \delta, y_0)$ satisfies Hypothesis 3.3 then it is also satisfied with $(T, u_{g_2}, \frac{\delta}{4}, y_0)$. Further again using that $(T, u_{g_2}, \delta, y_0)$ satisfies Hypothesis 3.3 and that $\tilde{T} < T$ we get for all $t \in (\hat{T}, \tilde{T}]$ and for all $z \in \partial\Omega$

$$\delta \leqslant |z - y_2(t)| < |z - y_1(t)| + |y_1(t) - y_2(t)| \leqslant \frac{\delta}{2} + |z - y_1(t)|.$$

This implies that for $t \in (\hat{T}, \tilde{T}]$ we have $|z - y_1(t)| > \frac{\delta}{2}$ for all $z \in \partial\Omega$ which implies that $y_1(t) \in \Omega_{\frac{\delta}{2}}$ for $t \in (\hat{T}, \tilde{T}]$. This contradicts the definition of $\hat{T}$ and therefore $\hat{T} \geqslant T$. □

Proof of Theorem 4.1. For proving the existence of a sub-sequences of α_n, T_n and $((\Theta_n, \|\cdot\|_n), \mathcal{P}_n, v_n)$, such that Method 3.6 has a solution we proceed similarly to the proof of Lemma 3 in [38]. Let T_n, V_n and δ be those appearing in Hypothesis 3.9. For $n \in \mathbb{N}$ fixed we use Hypothesis 3.10 with $g = V_n$ to define $\theta_{k,n} \in \Theta_k$ and $\alpha_{k,n} > 0$ such that $\lim_{k\to\infty} v_k(\theta_{k,n}) = V_n$ in $C^1_{loc}(\Omega)$ and $\lim_{k\to\infty} \alpha_{k,n}\mathcal{P}_k(\theta_{k,n}) = 0$.

Let $y_0 \in \omega$ be fixed and set $u_n(x) = u_{V_n}(x)$ and $u_{k,n}(x) = u_{v_k(\theta_{k,n})}(x)$ for all $x \in \Omega$ and $n, k \in \mathbb{N}$. Let $L_n = \left\|\nabla^2 V_n\right\|_{L^\infty\left(\Omega_{\frac{\delta}{4}};\mathbb{R}^{d\times d}\right)}$. Since for each $n \in \mathbb{N}$, $v_k(\theta_{k,n})$ converges to V_n in $C^1_{loc}(\Omega)$, there exists $k(n)$ such that for all $k > k(n)$ we have
(4.16)

$$\left(e^{(L_f+\frac{|B|^2}{\beta}L_n)T_n} - 1\right)\frac{|B|\|u^{k,n}-u^n\|_{L^\infty(\Omega_{\frac{\delta}{4}};\mathbb{R}^m)}}{(L_f+\frac{|B|^2}{\beta}L_n)} \leqslant \frac{\delta}{2} \text{ and } \left\|u^{k,n} - u^n\right\|_{L^\infty(\Omega_{\frac{\delta}{4}};\mathbb{R}^m)} \leqslant 1.$$

In view of (4.16) and the fact that (T_n, u_n, δ, y_0) satisfies Hypothesis 3.3 we can apply Corollary 4.5 with $g_1 = v_k(\theta_{k,n})$, $g_2 = V_n$ and $T = T_n$ to obtain that for all $y_0 \in \omega$ and $k \geqslant k(n)$ that the tuple $(T_n, u^{k,n}, \frac{\delta}{2}, y_0)$ satisfies Hypothesis 3.3. Consequently, we have that $\theta_{k,n} \in \mathcal{O}_{T_n,\Theta_k}$ for $k > k(n)$. Combining this, the fact that $\mathcal{P}_k$ is coercive and continuous, and that $\mathcal{V}_{T_n}$ is continuous, we get by the direct method that Method 3.6 with $S = ((\Theta_k, \|\cdot\|_k), \mathcal{P}_k, \alpha_{k,n}, v_k)$ and $T = T_n$ for $k > k(n)$ has an optimal solution.

For the second part we use (A.2). With this in mind we set

$$K_n = |B| \left[\|\nabla \ell\|_{L^\infty(\Omega;\mathbb{R}^d)} (e^{(L_f+\frac{|B|^2}{\beta}L_n)T_n} - 1)\frac{1}{(L_f + \frac{|B|^2}{\beta}L_n)^2} + \frac{T_n\beta}{2}(\|u^n\|_{L^\infty(\Omega_{\frac{\delta}{2}})} + 1)\right].$$

Using (A.2) we obtain for all $n \in \mathbb{N}$ and $k > k(n)$

$$\int_\omega \mathcal{V}_{T_n}(y_0, v_k(\theta_{k,n}))dy_0 \leqslant \int_\omega \mathcal{V}_{T_n}(y_0, V_n)dy_0 + |\omega|K_n \left\|u^{k,n} - u^n\right\|_{L^\infty(\Omega_{\frac{\delta}{4}};\mathbb{R}^m)} \tag{4.17}$$

By Hypothesis 3.10 we can choose $k_n \geqslant k(n)$ with $\lim_{n\to\infty} k_n = \infty$, such that

$$\lim_{n\to\infty} |\omega|K_n \left\|u^{k_n,n} - u^n\right\|_{L^\infty(\Omega_{\frac{\delta}{4}};\mathbb{R}^m)} + \alpha_{k_n,n}\mathcal{P}_{k_n}(\theta_{k_n,n}) = 0. \tag{4.18}$$

Let $\theta^*_n \in \Theta_{k_n}$ be an optimal solution of Method 3.6 for the choices $T = T_n$ and $S = ((\Theta_{k_n}, \|\cdot\|_{k_n}), \mathcal{P}_k, \alpha_{k_n,n}, v_{k_n})$. Then by (4.17)

$$\begin{aligned}\int_\omega \mathcal{V}_{T_n}(y_0, v_{k_n}(\theta^*_{k_n}))dy_0 + \alpha_{k_n}\mathcal{P}_{k_n}(\theta_{k_n}) &\leqslant \int_\omega \mathcal{V}_{T_n}(y_0, v_{k_n}(\theta_{k_n,n}))dy_0 + \alpha_{k_n}\mathcal{P}_{k_n}(\theta_{k_n,n}) \\ &\leqslant \int_\omega \mathcal{V}_{T_n}(y_0, u_{V_n})dy_0 + \alpha_{k_n}\mathcal{P}_{k_n}(\theta_{k_n,n}) + |\omega|K_n \left\|u^{k_n,n} - u^n\right\|_{L^\infty(\Omega_{\frac{\delta}{4}};\mathbb{R}^m)}.\end{aligned}$$

Combining the above inequality, (4.18) , and Hypothesis 3.9, we arrive at

$$\limsup_{n\to\infty}\int_\omega \mathcal{V}_{T_n}(y_0, v_{k_n}(\theta^*_{k_n}))dy_0 \leqslant \int_\omega V(y_0)dy_0.$$

Therefore, by Corollary 4.4 we conclude the proof. □

Proof of Theorem 4.2. Let us start by considering θ^*_n as solution of Method 3.8. From the optimality of θ^*_n we obtain that

$$\begin{aligned}&\tfrac{1}{2\beta}\left\|B^\top(\nabla v_n(\theta^*_n)-\nabla V)\right\|^2_{L^2(\Omega)}+\alpha_n\mathcal{P}_n(\theta^*_n)\\ &\leqslant \tfrac{1}{2\beta}\left\|B^\top(\nabla v_n(\theta_n)-\nabla V)\right\|^2_{L^2(\Omega)}+\alpha_n\mathcal{P}_n(\theta_n)\end{aligned} \tag{4.19}$$ □

with θ_n as in Hypothesis 3.11. By Hypothesis 3.11 we know that the right hand side of this inequality converges to 0 as n goes to infinity and consequently we obtain that

$$\lim_{n\to\infty} B^\top\nabla v_n(\theta^*_n) = B^\top\nabla V \text{ in } L^2(\Omega).$$

Moreover, again by Hypothesis 3.11 and the optimality of θ^*_n we obtain that $\|v_n(\theta^*_n)\|_\Theta$ is uniformly bounded for $n\in\mathbb{N}$. This together with the fact that Θ is compactly embedded in $C^1(\overline{\Omega})$ implies that $B^\top\nabla v_n(\theta^*_n)$ converges to $B^\top\nabla V$ in $C(\overline{\Omega})$. Let $T_n>0$ be a sequence such that $\lim_{n\to\infty} e^{T_n a}\left\|B^\top(\nabla v_n(\theta^*_n)-\nabla V)\right\|_{L^\infty(\Omega_{\frac{\delta}{2}};\mathbb{R}^m)}=0$, where a is given by (4.15) with $g_2=V$. Using Corollary 4.5 we know that for all $y_0\in\omega$ the tuple $(T_n, u_{v_n(\theta^*_n)}, \frac{\delta}{2}, y_0)$ satisfies Hypothesis 3.3 with δ replaced by $\frac{\delta}{2}$ for n large enough. With Lemma A.1 we conclude that

$$\lim_{n\to\infty}\mathcal{V}_{T_n}(\cdot, u_{v_n(\theta^*_n)}) = V \text{ in } L^\infty(\omega)$$

and thus the claim for Method 3.8 is verified.

Next we turn now our attention to Method 3.7. Let θ^*_n be an optimal solution for Method 3.7 with $S=S_n$ and $T=T_n$ as before. Since $D_{y_0}y^*(t,y_0)$ is invertible for all $(t,y_0)\in[0,T]\times\overline{\omega}$, by using the change of variables $z=y^*(t;y_0)$ as in Remark 3.4 we obtain

$$\int_\omega\int_0^T |B^\top(\nabla v(y^*(t;y_0))-\nabla V(y^*(t;y_0)))|^2 dtdy_0 = \int_{\Gamma_T}|B^\top(\nabla v(z)-\nabla V(z))|^2 g(z)dz$$

Moreover, we know that g is strictly positive in Γ_T and therefore it is equivalent to the Lebesgue measure in Γ_T. From this we can continue with the proof as in the case of Method 3.8 to obtain the desired convergence.

Remark 4.3. *Here we informally provide connections between Method 3.6 and Method 3.7 which are obtained through their optimality conditions. Let us consider a time horizon $T\in(0,\infty)$ and a setting $S=((\Theta,\|\cdot\|_\Theta),\mathcal{P},\alpha,v)$. The differentiability of the function $\mathcal{V}_T(y_0,u)$ for $y_0\in\Omega$ and $u\in C^2(\Omega)$ with respect to both arguments comes from [37] (see also the Appendix). Using this and (A.1) we get the optimality condition of Method 3.6 as*

$$\begin{aligned}\frac{1}{\beta|\omega|}\int_\omega\int_0^T &\left(\nabla_y v(\theta)(y(t;y_0,u_{v(\theta)}))-\nabla_y\mathcal{V}_T(y(t;y_0,u_{v(\theta)}),u_{v(\theta)})\right)^\top\\ &BB^\top\nabla\psi(y(t;y_0,u_{v(\theta)}))dtdy_0+\alpha D\mathcal{P}(\theta)\cdot\vartheta=0,\end{aligned} \tag{4.20}$$

where $\psi = (D_\theta v \cdot \vartheta)$, and $\vartheta \in \Theta$ is arbitrary. We note that (4.20) also arises as the optimality conditions of the regression problem given by minimizing over $\theta \in \Theta$ the following expression

$$\frac{1}{2\beta|\omega|}\int_\omega \int_0^T |B^\top(\nabla_y v(\vartheta)(y(t;y_0,u_{v(\theta)})) - \nabla_y \mathcal{V}_T(y(t;y_0,u_{v(\theta)}),u_{v(\theta)}))|^2 dt dy_0 + \alpha \mathcal{P}(\vartheta)$$

It is noteworthy to observe that this problem is similar to the problem arising in Method 3.7, replacing V by $\mathcal{V}_T$ and y^* by $y(\cdot;\cdot,u_{v(\vartheta)})$. This implies that if $\theta^* \in \Theta$ is an optimal solution of Method 3.7 and the trajectories $y(\cdot;y_0,u_{v(\theta^*)})$ are close to $y^*(\cdot;\cdot)$ in $W^{1,\infty}((0,T)\times\omega;\mathbb{R}^d)$, then $v(\theta^*)$ satisfies approximately (4.20). This holds in the opposite direction as well, that is, if θ^* is an optimal solution of Method 3.6 then it satisfies approximately the optimality condition associated to Method 3.7.

Remark 4.4. *There is still another important consequence of the optimality conditions of Method 3.6. In [36] it was proved that for finite horizon problems the optimality condition of this method becomes sufficient if the value function is C^2. In the following we will see that a similar situation holds for Method 3.6. Let us consider T tending to infinity. We formally set $\mathcal{V}_\infty$ as the limit of $\mathcal{V}_T$ and $\mathcal{O}_\infty$ as the superior limit of $\mathcal{O}_T$ as T tends to infinity.*

Let $\theta^* \in \mathcal{O}_\infty$ be such that it satisfies (4.20) and set $y(t;y_0) = y(t;y_0,u_{v(\theta^*)})$ for $(t,y_0) \in (0,\infty)\times\omega$. For $T>0$, by (A.7) and (A.5) we deduce that

$$\begin{aligned}\ell(y(t;y_0)) + \frac{1}{2\beta}|B^\top \nabla v(\theta^*)(y(t;y_0))|^2 + \nabla \mathcal{V}_T(y(t;y_0))^\top \\ \cdot(f(y(t;y_0)) - \tfrac{1}{\beta}BB^\top \nabla v(\theta^*)(y(t;y_0))) \\ = \ell(y(T;y_0)) + \tfrac{1}{2\beta}|B^\top \nabla v(\theta^*)(y(T;y_0))|^2.\end{aligned}$$

Further assuming that

$$\lim_{t\to\infty} \ell(y(t;y_0)) + \frac{1}{2\beta}|B^\top \nabla v(\theta^*)(y(t;y_0))|^2 = 0, \tag{4.21}$$

and letting T tend to infinity we obtain

$$\begin{aligned}\ell(y(t;y_0)) + \frac{1}{2\beta}|B^\top \nabla v(\theta^*)(y(t;y_0))|^2 + \nabla \mathcal{V}_\infty(y(t;y_0),u_{v(\theta^*)})^\top \\ \cdot(f(y(t;y_0)) - \frac{1}{\beta}BB^\top \nabla v(\theta^*)(y(t;y_0))) = 0.\end{aligned}$$

If there exists $\hat\vartheta \in \Theta$ such that $D_\theta v(\theta^*)\cdot\hat\vartheta = v(\theta^*) - \mathcal{V}_\infty(\cdot,u_{v(\theta^*)})$, then plugging $\hat\vartheta$ in (4.20) leads to

$$B^\top \nabla v(\theta^*)(y(t;y_0)) = B^\top \nabla \mathcal{V}_\infty(y(t;y_0),u_{v(\theta^*)}) \text{ for all } (t,y_0) \in (0,\infty)\times\omega. \tag{4.22}$$

We note that if $D_\theta v(\theta^*)$ is surjective then the existence of such a $\hat\vartheta$ is equivalent to assuming that $\mathcal{V}_\infty(\cdot,u_{v(\theta^*)}) \in \Theta$. Combining (4.22) and (4.4) we obtain

$$\begin{aligned}\ell(y(t;y_0)) + \tfrac{1}{2\beta}|B^\top \nabla \mathcal{V}_\infty(y(t;y_0),u_{v(\theta^*)})|^2 \\ +\nabla \mathcal{V}_\infty(y(t;y_0),u_{v(\theta^*)})^\top \cdot (f(y(t;y_0)) - \tfrac{1}{\beta}BB^\top \nabla \mathcal{V}_\infty(y(t;y_0),u_{v(\theta^*)})) = 0,\end{aligned} \tag{4.23}$$

which implies that $\mathcal{V}_\infty(\cdot, u_{v(\theta^*)})$ satisfies equation (2.4) *on the set* $\{y(t;y_0)\,|\,(t,y_0) \in (0,\infty)\times\omega\}$. *Further by the verification theorem (see Theorem I.7.1 in [22]), if ω is large enough in the sense that* $\{y(t;y_0)\,|\,(t,y_0) \in (0,\infty)\times\omega\} \supset \{y^*(t;y_0)\,|\,(t,y_0) \in (0,\infty)\times\omega\}$, *where y^* is optimal for* (2.1), *then* $\mathcal{V}_\infty(\cdot, u_{v(\theta^*)}) = V$. *In this case $u_{v(\theta^*)}$ provides an optimal feedback law for* (2.1) *and it is an optimal solution of Method* 3.6.

5. Obstacle problem. In this section we introduce a parameterized family of control problems with different degrees of smoothness of the value function depending on a penalty parameter. This example was already introduced in [37], for that reason we will follow the same definition.

For $\gamma \in [0,\infty)$, we consider the control problem

$$\min_{\substack{u\in L^2((0,\infty);\mathbb{R}^2),\\ y'=u,\ y(0)=y_0,}} \int_0^\infty \ell_\gamma(y(t))dt + \frac{\beta}{2}\int_0^\infty |u(t)|^2 dt, \tag{5.1}$$

where the running cost for the state variable is given by

$$\ell_\gamma(y) = \frac{1}{2}|y|^2\left(1+\gamma\psi\left(\frac{|y-z|}{\sigma}\right)\right), \text{ with } \psi(s) = \begin{cases} \exp\left(-\frac{1}{1-s^2}\right) & \text{if } |s|<1\\ 0 & \text{if } |s|\geqslant 1,\end{cases}$$

and $z\in\mathbb{R}^2$ satisfies $z_1 < -\sigma$, $z_2 = 0$, for some $\sigma > 0$.

The value function of this problem is denoted by V_γ. In [37], local Lipschitz regularity of V_γ for all $\gamma > 0$ and its lack of differentiability for γ sufficiently large were proved. Here we will prove that for γ close to 0 the function V_γ is in $C^\infty(\mathbb{R}^2)$. For this purpose we define

$$\gamma_s = \frac{1}{\sup_{y\in B(z,\sigma)}\left\{2|y|\cdot|\nabla\tilde\psi(y)| + \frac{1}{2}|y|\cdot|\nabla^2\tilde\psi(y)|\right\}} \tag{5.2}$$

with $\tilde\psi(y) = \psi\left(\frac{|y-z|}{\sigma}\right)$.

Proposition 5.1. *Let $\gamma\in[0,\gamma_s)$. The value function V_γ of* (5.1) *is $C^\infty(\mathbb{R}^2)$.*

Proof. By [37] we already know that for all $y_0\in\mathbb{R}^d$ and $\gamma\in[0,\infty)$ problem (5.1) has at least one solution and V_γ is locally Lipschitz continuous. Additionally, we observe that

$$\nabla^2\ell_\gamma(y) = I(1+\gamma\tilde\psi(y)) + 2\gamma y\otimes\nabla\tilde\psi(y) + \gamma\frac{|y|^2}{2}\nabla^2\tilde\psi(y)$$

where I is the 2×2 identity matrix. Multiplying both sides of the previous equality by $x\in\mathbb{R}^2\setminus(0,0)$ we see that

$$x^\top\cdot\nabla^2\ell_\gamma(y)\cdot x \geqslant \left((1+\gamma\tilde\psi(y)) - 2\gamma|y|\cdot|\nabla\tilde\psi(y)| - \gamma\frac{|y|^2}{2}\cdot|\nabla^2\tilde\psi(y)|\right)|x|^2.$$

Using that the support of $\tilde\psi$ is $B(z,\sigma)$ in the previous inequality, we obtain for $\gamma\in[0,\gamma_s)$ that

$$x^\top\cdot\nabla^2\ell_\gamma(y)\cdot x \geqslant \left(1-\gamma\sup_{y\in B(z,\sigma)}\left\{2|y|\cdot|\nabla\tilde\psi(y)| + \frac{1}{2}|y|\cdot|\nabla^2\tilde\psi(y)|\right\}\right)|x|^2 > 0, \tag{5.3}$$

which implies that ℓ_γ is a strictly convex function. Due to the linearity of the state equation of (5.1), we also get that the objective function of (5.1) is convex and consequently the optimal control of (5.1) is unique for $\gamma \in [0, \gamma_s)$.

In the following, we will use (5.3) to prove that V_γ is $C^2(\mathbb{R}^2)$ by means of the implicit function theorem. For this purpose we define the function $\Phi : X \times \mathbb{R}^2 \mapsto Y$ where $X = H^1((0,\infty);\mathbb{R}^2) \times L^2((0,\infty);\mathbb{R}^2) \times H^1((0,\infty);\mathbb{R}^2)$, and $Y = L^2((0,\infty);\mathbb{R}^2) \times L^2((0,\infty);\mathbb{R}^2) \times \mathbb{R}^2 \times L^2((0,\infty);\mathbb{R}^2)$ by

$$\Phi(y,u,p,y_0) = \begin{pmatrix} -p' + \nabla \ell_\gamma(y) \\ y' - u \\ y(0) - y_0 \\ \beta u - p \end{pmatrix}. \tag{5.4}$$

We note that $\Phi(y,u,p,y_0) = 0$ is the optimality condition for (5.1). Clearly the function Φ is of class C^∞ and its differential with respect to its first three variables is given by

$$D_{y,u,p}(y,u,p,y_0) \cdot (\delta_y, \delta_u, \delta_p) = \begin{pmatrix} -\delta_p' + \nabla^2 \ell_\gamma(y)\delta_y \\ \delta_y' - \delta_u \\ \delta_y(0) \\ \beta\delta_u - \delta_p \end{pmatrix}$$

for $((y,u,p),y_0) \in X \times \mathbb{R}^2$ and $(\delta_y,\delta_u,\delta_p) \in X$. In order to use the implicit function theorem we need to prove that for u^* the optimal solution of (5.1), y^* its optimal trajectory and p^* the respective adjoint state, the partial sub-differential $D_{y,u,p}(y^*,u^*,p^*)$ is invertible , i.e., for all $(z_p, z_y, z_{y_0}, z_u) \in Y$ there exist only one $(\delta_y,\delta_u,\delta_p) \in X$ such that

$$D_{y,u,p}(y^*,u^*,p^*) \cdot (\delta_y,\delta_u,\delta_p) = (z_p, z_y, z_{y_0}, z_u).$$

We note that this equation is equivalent to the optimality conditions of the following linear-quadratic problem

$$\begin{cases} \min \int_0^\infty \left(\frac{1}{2}\delta_y^\top(t) \nabla^2 \ell_\gamma(y^*(t))\delta_y(t) - z_p(t)\cdot\delta_y(t) + \frac{\beta}{2}|\delta_u(t)|^2 - \delta_u(t)\cdot z_u(t) \right) dt, \\ \text{where } \delta_u \in L^2((0,\infty);\mathbb{R}^2), \text{and } \delta_y' = \delta_u + z_y,\ \delta_y(0) = z_{y_0}. \end{cases} \tag{5.5}$$

Consider $\tilde{u}$ the optimal solution of (5.1) with initial condition z_{y_0} and $\tilde{y}$ the corresponding trajectory. We note that using $\delta_u = \tilde{u} - z_y$ we obtain that $\delta_y = \tilde{y}$ satisfies $\delta_y' - \delta_u = z_y$ and

$$\left| \int_0^\infty \left(\frac{1}{2}\delta_y^\top(t)\nabla^2 \ell_\gamma(y(t))\delta_y(t) - z_p(t)\cdot\delta_y(t) + \frac{\beta}{2}|\delta_u(t)|^2 - \delta_u(t)\cdot z_u(t) \right) dt \right| < +\infty.$$

This together with the positive definiteness of $\nabla^2\ell_\gamma$ and weak lower semi-continuity of the objective function imply that (5.5) has a unique optimal solution. Then by the implicit function theorem there exists an open neighborhood $\Gamma \subset \mathbb{R}^2$ of y_0 and functions $\hat{u} : \Gamma \mapsto L^2((0,\infty);\mathbb{R}^2)$, $\hat{y} : \Gamma \mapsto H^1((0,\infty);\mathbb{R}^2)$ and $\hat{p} : \Gamma \mapsto H^1((0,\infty);\mathbb{R}^2)$ of class C^∞, such that for all $z_0 \in \Gamma$ we have that $\Phi(y(z_0), u(z_0), p(z_0), z_0) = 0$.Since for $\gamma \in [0,\gamma_s)$ problem (5.1) is strictly convex, we have that $u(z_0)$ is its optimal solution, $y(z_0)$ its optimal trajectory, and $p(z_0)$ the corresponding adjoint state. Hence, $V_\gamma(z_0) = \int_0^\infty \left(\ell_\gamma(y(z_0)) + \frac{\beta}{2}|u(z_0)|^2 \right) dt$ and by the chain rule we obtain that V_γ is C^∞ in Γ, Since y_0 is arbitrary we deduce that V_γ is $C^\infty(\mathbb{R}^2)$. □

As mentioned before, in [37] it was proved that for γ large enough the value function V_γ is non differentiable. Specifically, for each such γ there exists x_γ such that V_γ is non-differentiable in $\{(x,0) \in \mathbb{R}^2 : x > x_\gamma\}$. However, it was also proved in [37] that for all $\gamma > 0$ Hypothesis 3.3 and Hypothesis 3.9 hold. Consequently, by using a sequence of setting satisfying Hypothesis 3.10 we obtain by means of Theorem 4.1 that convergence in the sense of (3.8) holds for Method 3.6. On the other hand, for Methods 3.7 and 3.8 and $\gamma \in [0, \gamma_s)$ we obtain that Hypothesis 3.11 holds provided the sequence of settings is chosen appropriately.

This together with Proposition 5.1 allows us to utilize problem (5.1) to numerically investigate the behavior of Methods 3.6 - 3.8 for different choices of γ, which account for different degrees of smoothness of V_γ.

We close this section by establishing another important property of the value function.

Proposition 5.2. *The value function associated to* (5.1) *is semiconcave on every compact subset of* $\mathbb{R}^2$.

Proof. Let $C \subset \mathbb{R}^2$ be compact and denote by $(y^*(\cdot; y_0, u^*(\cdot; y_0)))$ the optimal state-control pairs with $y_0 \in C$. Since the optimal trajectories associated to (5.1) are globally exponentially stable there exists $\hat{T}$ such that $y^*(t; y_0) < |z| - \sigma$ for all $(t, y_0) \in [\hat{T}, \infty) \times C$, and consequently $V_\gamma(y_0) = V_0(y_0) = \frac{\sqrt{\beta}}{2}|y_0|^2$, [37, Lemma 9.2, Lemma 9.3]. Here $B(0, |z| - \sigma)$ denotes the open ball with center at the origin and radius $|z| - \sigma$. Further $|u^*(t; y_0)| = \frac{2}{\beta}\ell_\gamma(y^*(t; y_0))$, [37, Proposition 9.1], and thus for a constant K we have $|u^*(t; y_0)| \leqslant K$, for all $(t, y_0) \in [0, \infty) \times C$.

Let us consider

(5.6)
$$\min_{\substack{u \in L^2((0,\hat{T}); \mathbb{R}^2),\ |u(t)| \leqslant K+1 \\ y' = u,\ y(0) = y_0 \in C}} \int_0^{\hat{T}} \ell_\gamma(y(t))dt + \frac{\beta}{2}\int_0^{\hat{T}} |u(t)|^2 dt + \hat{V}_0(y(\hat{T}, y_0)),$$

where $\hat{V}_0$ denotes the extension of V_0 outside of $B(0, |z| - \sigma)$ by $\hat{V}_0(y) = \frac{\sqrt{\beta}}{2}|y|^2$.

For $y_0 \in C$ this problem is equivalent to (5.1). Here we use Bellman's principle, the fact that $y^*(\hat{T}; y_0) \in B(0, |z| - \sigma)$, and that $|u^*|_{L^\infty(0,\infty)} \leqslant K$, so that the control constraint in (5.6) is not active at the solutions. The value function restricted to C coincides for these two problems. We next utilize that (5.6) can be considered as finite horizon problem with terminal penalty given by $\hat{V}_0$. It then follows from [11, Theorem 7.4.11] that the value function for (5.6) is semiconcave on C, provided that the assumptions of that theorem are satisfied. In this respect it suffices to recall that any function which is $C^1(A)$ in an open set A containing C is semiconcave on C, [11, Proposition 2.1.2]. This is directly applicable for $\ell \in C^1(\mathbb{R}^2)$ as well as $\hat{V}_0 \in C^\infty(\mathbb{R}^2)$. The remaining assumptions on the dynamical system and the running cost are trivially satisfied. □

6. Numerical experiments. Numerical tests for Methods 3.6 and 3.7 including high-dimensional problems are available in the literature, we refer to eg. [39, 48, 6, 37, 38, 43, 44]. Thus the strength and validity of them for practical realisation has already been established. Here we make an effort to compare these methods with respect to the influence of the regularity of the value function. For this purpose, we employ equation (5.1) and recall that the parameter γ acts as a control for the regularity of V_γ. Thus we are interested whether the performance of these methods significantly differs with respect to the regularity of the value function. It would be of interest to also compare these methods for high dimensional problems. This, however, is not within the scope of this research.

We consider problem 5.1 with $z_1 = -2$, $\sigma = 1$, and set $\omega = (-5, 5) \times (-2, 2)$. Since the dimensionality is low, a regular grid is used to approximate the integrals appearing in Methods 3.6 - 3.8. For high dimensional problems this is not feasible and other discretization method should be used instead, as for example Monte-Carlo integration, which is the usual choice for machine learning methods and has been used in [38, 36]. We report on the behavior of the obtained feedback laws for each of the methods with $\gamma \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$.

In the following we further describe the setting of the methods. For the Banach spaces Θ and the parametrization we follow Remark 3.2. Thus the setting depends on two parameters, which are the order k and the degree n. Here we consider $k \in \{1, 2\}$ and $n \in \{10, 20, 30, 40\}$. Recall that the choice of k connects the smoothness of the value function. The penalty function of the setting is $\mathcal{P}(\theta) = \frac{1}{2}\left(\frac{1}{2}|\theta|_2^2 + |\theta|_1\right)$ and the our tests were carried out with penalty parameter α chosen from the set $\{10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}\}$. In all the cases we consider the time horizon with $T = 1$.

The integrals over ω appearing in the learning problems are approximated over a regular grid of size 16×16. We call these points the **training set**. In order to study the performance of the feedback produced by each method we also consider a finer regular grid of size 32×32, which is the **test set**. In both the training and test sets the performance of each method is evaluated by the following error measures:

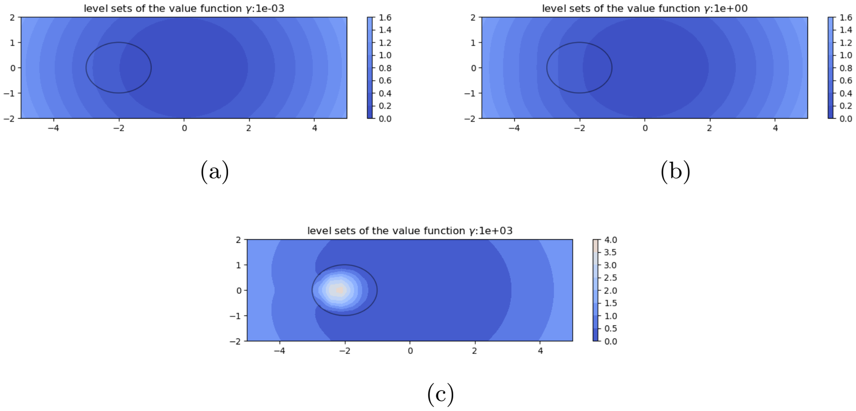
$$NMAE_V = \sum_{i=1}^{N} |V(y_0^i) - \mathcal{V}_T(y_0^i, v(\theta))| \cdot \frac{1}{\sum_{i=1}^{N} V(y_0^i)}, \tag{6.1}$$

$$NMRSE_c = \sum_{i=1}^{N} \|u_i^* - \hat{u}_i\|_{L^2((0,T);\mathbb{R}^2)} \cdot \frac{1}{\sum_{i=1}^{N} \|u_i^*\|_{L^2((0,T);\mathbb{R}^2)}}, \tag{6.2}$$

where $\{y_0^i\}_{i=1}^{n}$ is a set of initial conditions which in this case could be either the training or the test sets, $\{u_i^*\}_{i=1}^{N}$ are the optimal controls and $\{\hat{u}_i\}_{i=1}^{N}$ are the controls obtained by applying the respective feedback. For obtaining a reference value for the optimal value function V and the optimal controls for the chosen initial conditions we numerically solve the discretized open loop control problem. For this purpose the $ODEs$ and the integral of the running cost are discretized by an explicit Euler scheme with time step $\frac{1}{400}$. In accordance with Theorems 4.1 and 4.2 we expect that for small values of γ all three approaches will achieve a good performance since in this case V_γ is smooth. On the other hand, V_γ is non-smooth for large values of γ, and only Theorem 4.1 is applicable, which implies that we only have a guarantee of good performance for Method 3.6.

For solving the learning problems we employ a proximal gradient method together with the Barzilai-Borwein step size and a non-linear backtracking line search. This algorithm is described in [38] and [36] in the context of Method 3.6. Following the notation of Algorithm 1 in [36] we use $k = 3$, $tol = 10^{-6}$, $\kappa = 10^{-3}$ and $\xi = \frac{1}{2}$, with the same tolerances measures as in [36]. Concerning initialization, for each setting we first solve with $\gamma = 10^{-3}$ using as initial guess $\theta^0 = 0$, and subsequently for $\gamma > 10^{-3}$ the solution corresponding to the previous γ value is used as initial guess. The penalty parameter α was chosen such that it achieve approximately the best

$NMAE_V$ in the training set. The reported results were obtained on a computer with a Xeon E5-2630v3 (2x8Core 2,4Ghz) processor and 8 GB of RAM memory.



(a)

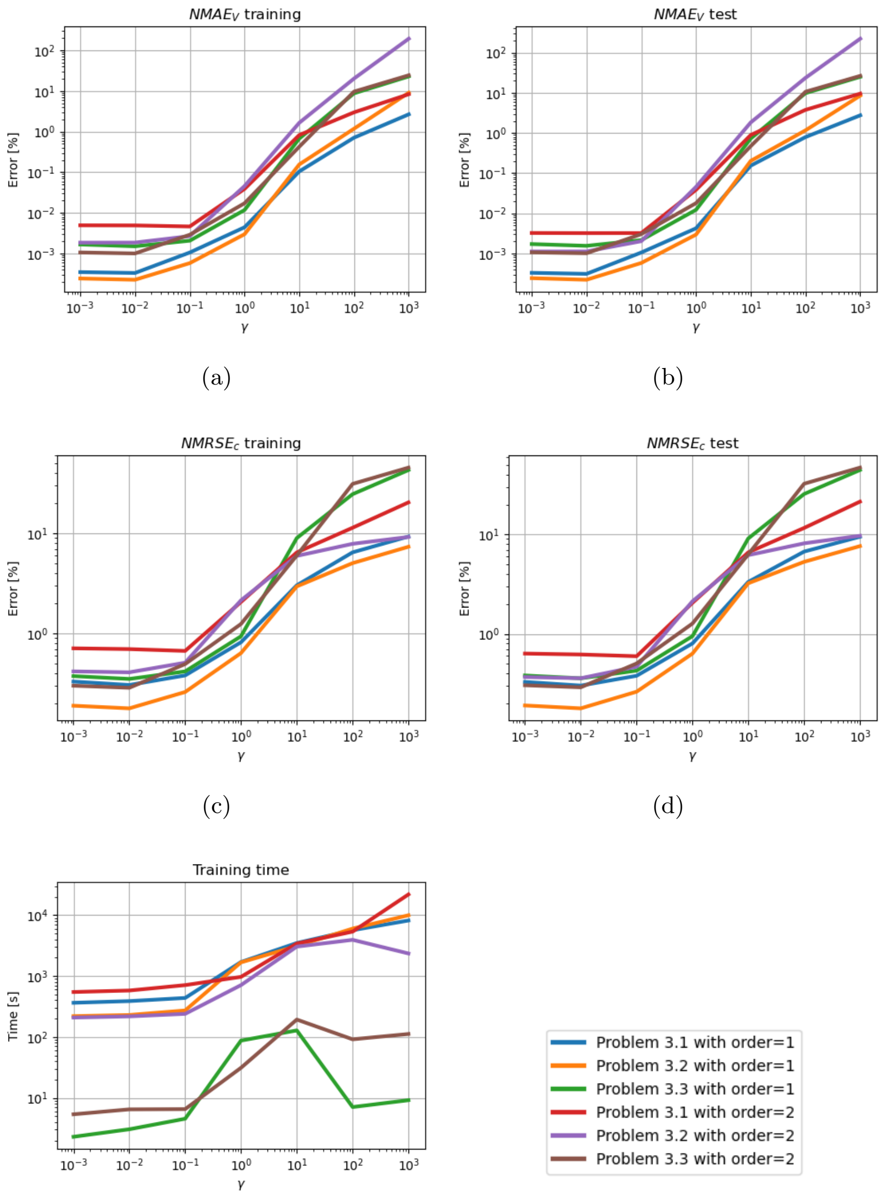
(b)

(c)

Fig. 1: Level sets of V_γ, the color of the level sets are in logarithmic scale.

We start the discussion of the experiments with the level sets of the value function for selected values of regularisation parameters $\gamma = 10^{-3}, 0, 10^3$. This is relevant to understand the distribution of the errors for each method which will be reported later on. In Figure 1 the level sets for selected γ-values together with the boundary of the obstacle are presented. We observe that on the right hand side of the $y-$axis the level sets do not change with γ. This is concordant with the fact that in this region the obstacle has no effect. On the other hand, on the left hand side we observe that in a neighborhood of the $x-$axis the level sets vary from being convex to non-convex. This is consistent with the fact that on this axis the value function is non-differentiable for large γ. Further, for $\gamma = 10^3$ the value function has a local maximum close to the center of the obstacle. Indeed, since V_γ converges to V_∞ it is expected that V_γ blows up inside the obstacle as γ tends to infinity.

To compare the performance graphically, in Figure 2 the errors and the training time for each method are depicted. We observe that typically $k = 1$ achieves a better performance than $k = 2$, especially for $\gamma \geqslant 10$. Additionally, we note that the error increases with γ. Methods 3.6 and 3.7 with $k = 1$ achieved the best performance in both error measures. To take a closer look we turn our attention to Figure 3 where the errors are normalized by dividing them by the error obtained by using Method 3.6 with k=1. From this we see that for $\gamma \leqslant 1$ the method with the best performance in terms of $NMAE_V$ and $NMRSE_c$ control is Method 3.7 with $k = 1$. In contrast, for $\gamma > 1$, Method 3.6 with $k = 1$ has the best $NMAE_V$ and moreover the distance with respect to Method 3.7 increases. We point out that also for $\gamma > 1$ the fact that Method 3.7 has best $NMRSE_c$ control does not imply the same with respect to $NMAE_V$. Concerning the efficiency of the methods in Figure 2(e) we provide the trainings time for each experiment, that is, the time that it takes to achieve the stooping criterion. In this figure we see that Method 3.8 is the one with the shortest training time. For Methods 3.6 and 3.7 the training times tend to increase with γ. This can be a consequence of decreasing regularity and the fact that the value function blows up inside the obstacle as γ tends to infinity.



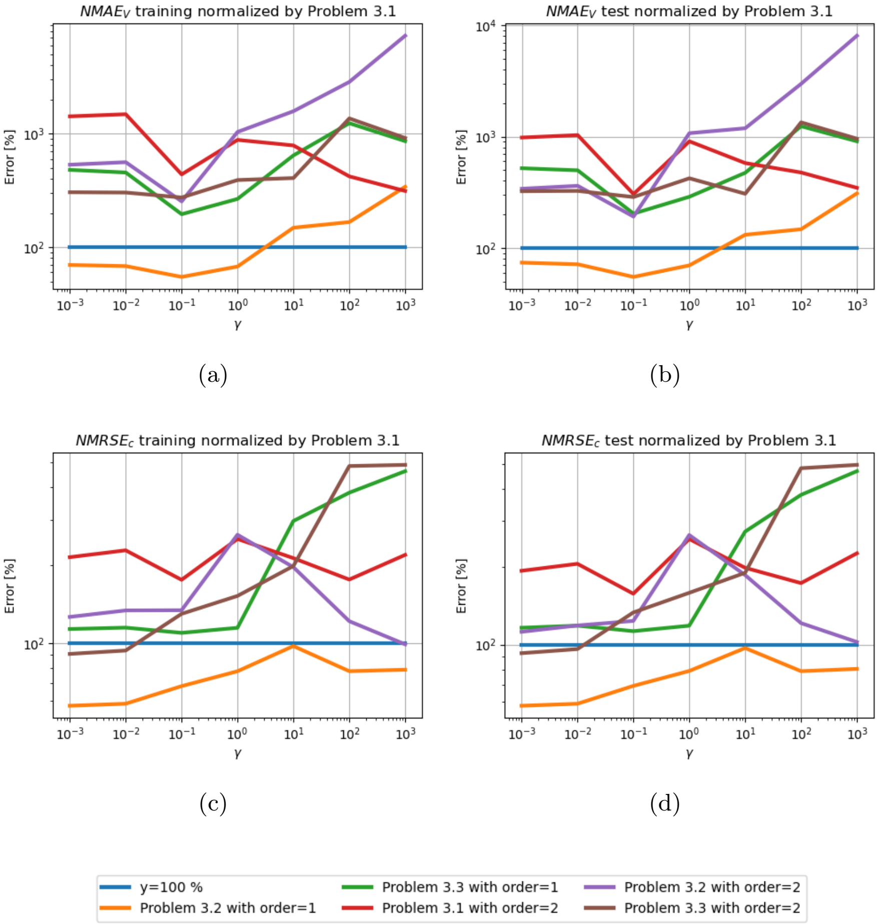
(a) (b)

(c) (d)

(e)

Fig. 2: Errors and training times. In all the figures both axes are in logarithmic scale.

The previous analysis does not address the spatial distribution of the error. This information is given in Figures 5-7 where the distribution of the error between the value functions V_γ and $\mathcal{V}_T(\cdot,\theta^*)$ is depicted, with θ^* the solution obtained by the respective experiment. From Figure 5 we observe that the error is uniformly smaller than 10^{-3} for $\gamma \leqslant 10^{-3}$, and we can report that this is also the case as long as $\gamma \leqslant 1$.



(a) (b)

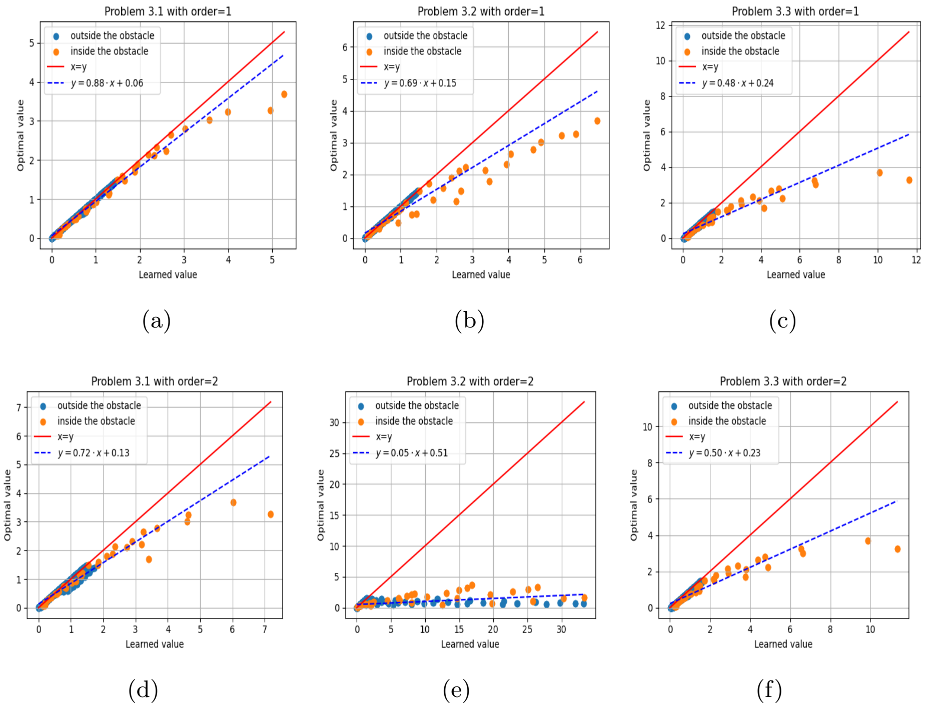
(c) (d)

Fig. 3: Errors normalized by the error of Method 3.1. In all the figures both axes are in logarithmic scale.

Starting from $\gamma = 10^1$ we see from Figures 6-7 that the error in the region left to the $y-$axis tends to increase with γ. Moreover, the error is higher for the points close to the $x-$axis which are to the left of the $y-$axis and also inside the obstacle than those outside. As explained before, this can be attributed to the fact that the value function is non smooth on the $x - axis$ for all x left of the obstacle and that it blows up inside the obstacle as γ tends to infinity.

Figures 4 show the scatter plot and the dispersion between V_γ and $\mathcal{V}_T(\cdot, \theta^*)$ evaluated at the test points for $\gamma = 10^3$. In all the cases the points in the interior of the obstacle tend to be further from the identity line than those outside of the obstacle. For Method 3.6 this effect is weaker than for the others.

7. Conclusions. In this work convergence in the sense of (3.8) is analyzed for three machine learning approaches. They differ with respect to the assumption on the regularity of the value function. Further, now that more techniques to compute suboptimal feedback controls guided by the Bellman principle and hence the HJB



(a) (b) (c)

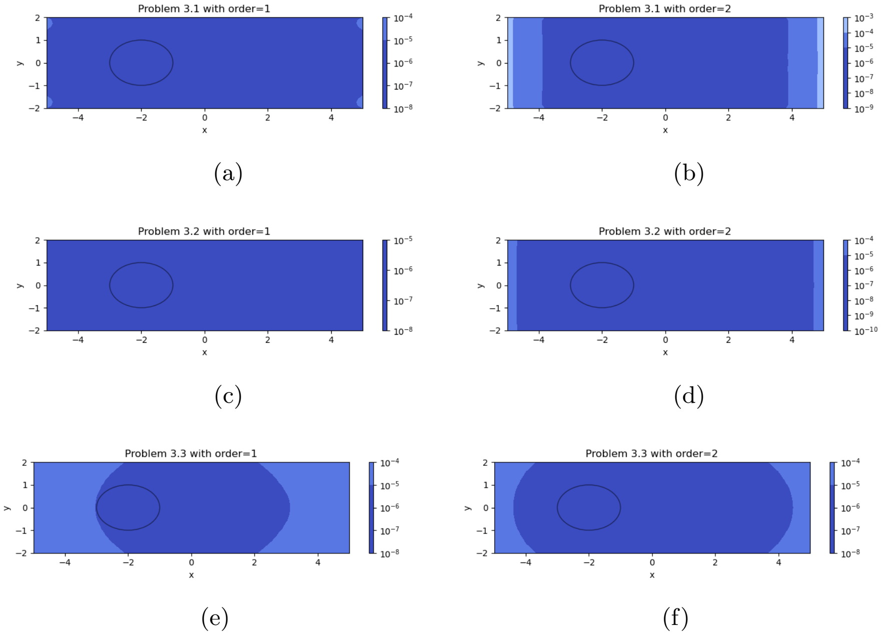
(d) (e) (f)

Fig. 4: Scatter plots for $\gamma = 10^3$.

equation, become available, it is of interest to start towards the challenging goal of comparing methods. For this purpose we present a family of control problems in Section 5 indexed by a penalty parameter $\gamma > 0$ for which the value function is smooth if γ close to 0, and while it is non-differentiable but semi-concave for γ large. This family of control problem was used to compare the performance of the three methods. From these experiments it was observed that for γ close to 0, which correspond to the case of smooth value functions, Method 3.7 has a better performance than the other approaches, but for larger γ Method 3.6 is the one with a better performance. This is coherent with the convergence results.

Further extensions of the work presented here are possible. For instance, the discretized version of the learning problems was not studied. This is particularly relevant for high-dimensional problems, since in this case the discretization is usually carried out by using Monte-Carlo integration. Problems with constraints on either the state or the control can be considered by first using an exterior penalty method as was used for the construction of problem (5.1). Finally, the extension of this problem to higher dimension is of interest, because it could help to study the behavior of the learning problems in the context of high-dimensions and non-smoothness of the value function.

Appendix A. Properties of $\mathcal{V}_T$. In this section we provide some important result concerning properties of $\mathcal{V}_T$ needed throughout this work. The proofs can be obtained with standard techniques and are therefore not presented here.

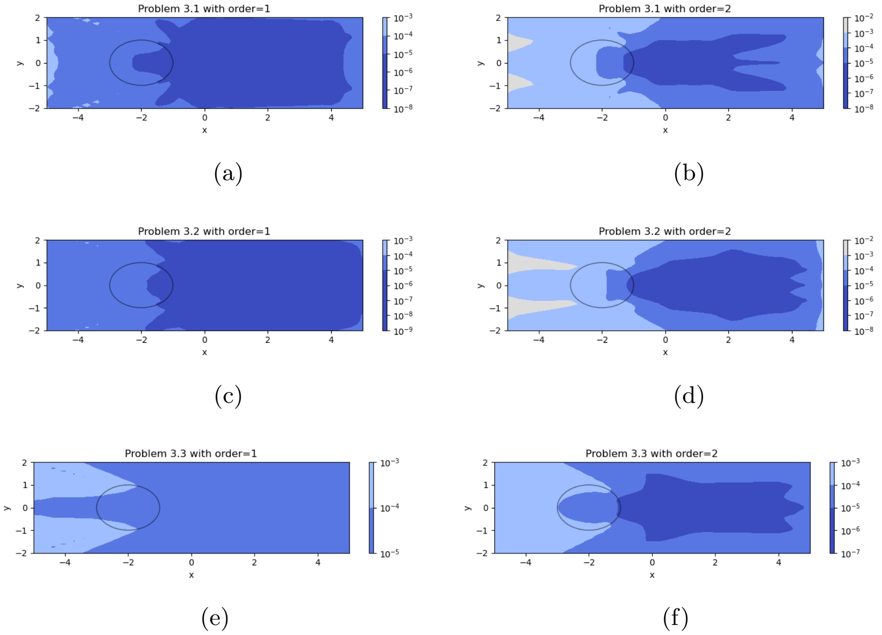


(a) (b)

(c) (d)

(e) (f)

Fig. 5: Error distribution for $\gamma = 10^{-3}$.

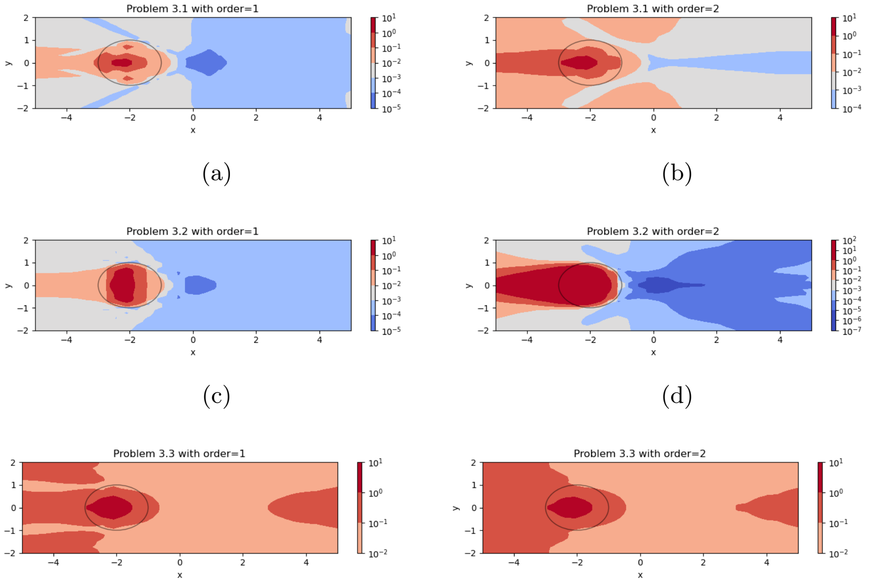


(a) (b)

(c) (d)

(e) (f)

Fig. 6: Error distribution for $\gamma = 1$.



(a) (b)

(c) (d)

(e) (f)

Fig. 7: Error distribution for $\gamma = 10^3$.

Lemma A.1. *Let $T > 0$ denote a finite time horizon, and let $y_0 \in \omega$, $g_1 \in C^2(\Omega)$, $g_2 \in C^2(\Omega)$, and $\delta > 0$ be such that $(T, u_{g_1}, \delta, y_0)$ and $(T, u_{g_2}, \delta, y_0)$ Hypothesis 3.3 holds. Then we have for all $t \in [0, T]$*

$$|y(t; y_0, u_{g_1}) - y(t; y_0, u_{g_2})| \leqslant \frac{|B|\, \|u_{g_1} - u_{g_2}\|_{L^\infty(\Omega_\delta;\mathbb{R}^m)}}{a} \left(e^{ta} - 1\right) \tag{A.1}$$

and

$$\begin{aligned} &|\mathcal{V}_T(y_0, u_{g_1}) - \mathcal{V}_T(y_0, u_{g_2})| \\ &\leqslant |B|\,[\frac{\|\nabla \ell\|_{L^\infty(\Omega_\delta);\mathbb{R}^d)}}{a} c_T + \frac{T\beta}{2} \|u_{g_1} + u_{g_2}\|_{L^\infty(\Omega_\delta;\mathbb{R}^d)}\,]\, \|u_{g_1} - u_{g_2}\|_{L^\infty(\Omega_\delta;\mathbb{R}^m)}\,, \end{aligned} \tag{A.2}$$

where

$$c_T = (\frac{e^{aT} - 1}{a} - T) \text{ and } a = \|Df\|_{L^\infty(\Omega_\delta;\mathbb{R}^{d\times d})} + \frac{|B|^2}{\beta} \left\|\nabla^2 g_2\right\|_{L^\infty(\Omega_\delta;\mathbb{R}^{d\times d})}.$$

Lemma A.2. *Let $T > 0$ be a time horizon, an let $y_0 \in \omega$, $u \in C^1(\Omega)$, and $\delta > 0$ be such that (T, u, δ, y_0) Hypothesis 3.3 holds for all $y_0 \in \omega$. For arbitrary $y_0 \in \omega$ let us define the adjoint state $p(\cdot; y_0, u)$ as the solution of the following adjoint equation*

$$-p'(t) + \nabla\ell(y(t)) + \beta Du(y(t))^\top \cdot u(y(t)) - (Df(y(t)) + BDu(y(t)))^\top \cdot p(t) = 0 \tag{A.3}$$

for all $t \in (0, T)$; $p(T) = 0$,with $y(t) = y(t; y_0, u)$. Then the function $\mathcal{V}_T$ its differentiable with respect to each of its variables and we have

$$D_u \mathcal{V}_T(y_0, u) \cdot \nu = \int_0^T (\beta u(y(t; y_0, u)) - B^\top p(t; y_0, u))^\top \cdot \nu(y(t; y_0, u)) dt \tag{A.4}$$

for all $\nu \in C^1(\Omega;\mathbb{R}^m)$*, and*

$$\nabla_y \mathcal{V}_T(y_0,u) = -p(0;y_0,u). \tag{A.5}$$

Further, for all $t \in (0,T]$ *we have*

$$\nabla_y \mathcal{V}_T(y(t),u) = -p(0;y(t),u) = -p(t;y_0,u). \tag{A.6}$$

Remark A.1. Under the hypotheses of the previous lemma and assuming $u = u_\phi$ the adjoint equation can be expressed in terms of ϕ as

$$-p'(t) + \nabla\ell(y(t)) - Df^\top(y(t))p(t) + \frac{1}{\beta}\nabla^2\phi(y(t))BB^\top(\nabla\phi(y(t)) + p(t)) = 0 \tag{A.7}$$

for all $t \in (0,T)$; $p(T) = 0$. Additionally, using (A.6) we can express the derivative of $\mathcal{V}_T(y_0,u_\phi)$ with respect to ϕ as

$$\begin{aligned} &D_\phi\left(\mathcal{V}_T(y_0,u_\phi)\right)\cdot\psi \\ &\quad = \tfrac{1}{\beta}\int_0^T (\nabla\phi(y(t;y_0,u_\phi)) - \nabla_y\mathcal{V}_T(y(t;y_0,u_\phi)))^\top BB^\top\nabla\psi(y(t;y_0,u_\phi))\,dt \end{aligned}$$

for $\psi \in C^2(\Omega)$.

REFERENCES

[1] M. Akian, S. Gaubert, and A. Lakhoua, *The max-plus finite element method for solving deterministic optimal control problems: Basic properties and convergence analysis*, SIAM J. Control Optim., 47 (2008), pp. 817–848.

[2] G. Albi, S. Bicego, and D. Kalise, *Gradient-augmented supervised learning of optimal feedback laws using state-dependent Riccati equations*, IEEE Control Systems Letters, 6 (2022), pp. 836–841.

[3] A. Alla, M. Falcone, and D. Kalise, *An efficient policy iteration algorithm for dynamic programming equations*, SIAM J. Control Optim., 37 (2015), pp. A181–A200.

[4] A. Alla, M. Falcone, and L. Saluzzi, *An efficient dp algorithm on a tree-structure for finite horizon optimal control problems*, SIAM J. Sci. Comput., 41 (2019), pp. A2384—A2406.

[5] A. Alla, M. Falcone, and S. Volkwein, *Error analysis for pod approximations of infinite horizon problems via the dynamic programming approach*, SIAM Journal on Control and Optimization, 55 (2017), pp. 3091—-3115.

[6] B. Azmi, D. Kalise, and K. Kunisch, *Optimal feedback law recovery by gradient-augmented sparse polynomial regression*, J. Mach. Learn. Res., 22 (2021), pp. 1—-32.

[7] R. W. Beard, G. N. Saridis, and J. T. Wen, *Galerkin approximation of the generalized Hamilton-Jacobi-Bellman equation*, Automatica, 33 (1997), pp. 2159–2177.

[8] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker, *An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations*, J. Sci. Comput., 55 (2013), pp. 575–605.

[9] O. Bokanowski, X. Warin, and A. Prost, *Neural networks for first order hjb equations and application to front propagation with obstacle terms*, 2022, https://arxiv.org/abs/2210.04300.

[10] F. Bonnans, P. Chartier, and H. Zidani, *Discrete approximation for a class of the Hamilton-Jacobi equation for an optimal control problem of a differential-algebraic system*, Control and Cybernetics, 32 (2003), pp. 33–55.

[11] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton—Jacobi Equations, and Optimal Control*, Birkhäuser Boston, 2004.

[12] G. Chen, *Deep neural network approximations for the stable manifolds of the hamilton-jacobi equations*, 2023, https://arxiv.org/abs/2007.15350.

[13] Y. T. Chow, J. Darbon, S. Osher, and W. Yin, *Algorithm for overcoming the curse of dimensionality for time-dependent non-convex Hamilton– Jacobi equations arising from optimal control and differential games problems*, J. Sci. Comput., 73 (2017), pp. 617–643.

[14] Y. T. Chow, J. Darbon, S. Osher, and W. Yin, *Algorithm for overcoming the curse of dimensionality for state-dependent Hamilton–Jacobi equations*, J. Comput. Phys, 387 (2019), pp. 376–409.

[15] Y. T. Chow, W. Li, S. Osher, and W. Yin, *Algorithm for Hamilton–Jacobi equations in density space via a generalized Hopf formula*, J. Sci. Comput., 80 (2019), pp. 1195–1239.

[16] J. Darbon, G. P. Langlois, and T. Meng, *Overcoming the curse of dimensionality for some Hamilton–Jacobi partial differential equations via neural network architectures*, Res. Math. Sci., 7 (2020), pp. 1–50.

[17] J. Darbon and S. Osher, *Algorithms for overcoming the curse of dimensionality for certain Hamilton–Jacobi equations arising in control theory and elsewhere*, Res. Math. Sci., 3 (2016), p. 19.

[18] S. Dolgov, D. Kalise, and K. Kunisch, *Tensor decomposition methods for high-dimensional Hamilton–Jacobi–Bellman equations*, SIAM J. Sci. Comput., 43 (2021), pp. A1625—A1650.

[19] P. M. Dower, W. M. McEneaney, and H. Zhang, *Max-plus fundamental solution semigroups for optimal control problems*, in 2015 Proceedings of the Conference on Control and its Applications, 2015, pp. 368—-375.

[20] T. Ehring and B. Haasdonk, *Hermite kernel surrogates for the value function of high-dimensional nonlinear optimal control problems*, Advances in Comp. Math., 50 (2023), pp. 1–33.

[21] M. Falcone and R. Ferretti, *Semi-Lagrangian approximation schemes for linear and Hamilton-Jacobi equations*, SIAM, Philadelphia, 2013.

[22] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Stochastic Modelling and Applied Probability, Springer, New York, second ed. ed., 2006.

[23] J. Garcke and A. Kröner, *Suboptimal feedback control of pdes by solving hjb equations on adaptive sparse grids*, J. Sci. Comput., 70 (2017), pp. 1–28.

[24] S. Gaubert, W. McEneaney, and Z. Qu, *Curse of dimensionality reduction in max-plus based approximation methods: Theoretical estimates and improved pruning algorithms*, in 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 1054–1061.

[25] A. Gorodetsky, S. Karaman, and Y. Marzouk, *High-dimensional stochastic optimal control using continuous tensor decompositions*, Int. J. Robot. Res., 37 (2018), pp. 340–377.

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

[27] M. B. Horowitz, A. Damle, and J. W. Burdick, *Linear Hamilton Jacobi Bellman equations in high dimensions*, in 2014 53rd IEEE Conf. Decis. Control, 2014, pp. 5880–5887.

[28] K. Ito, C. Reisinger, and Y. Zhang, *A neural network-based policy iteration algorithm with global h2-superlinear convergence for stochastic games on domains*, Found. Comput. Math., 21 (2021), pp. 331–374.

[29] D. Kalise, S. Kundu, and K. Kunisch, *Robust feedback control of nonlinear pdes by numerical approximation of high-dimensional Hamilton–Jacobi–Isaacs equations*, SIAM J. Appl. Dyn. Syst., 19 (2020), pp. 1496–1524.

[30] D. Kalise and K. Kunisch, *Polynomial approximation of high-dimensional Hamilton–Jacobi–Bellman equations and applications to feedback control of semilinear parabolic pdes*, SIAM J. Sci. Comput., 40 (2018), pp. A629–A652.

[31] D. Kalise, K. Kunisch, and Z. Rao, *Hamilton-Jacobi-Bellman Equations: Numerical Methods and Applications in Optimal Control*, Vol. 21 De Gruyter - Radon Series on Computational and Applied Mathematics, De Gruyter, Berlin, 2018.

[32] W. Kang, Q. Gong, and T. Nakamura-Zimmerer, *Algorithms of data development for deep learning and feedback design*, Phys. D: Nonlinear Phenom., 425 (2021), p. 132955.

[33] W. Kang and L. C. Wilcox, *Mitigating the curse of dimensionality: sparse grid characteristics method for optimal feedback control and HJB equations*, Computational Optimization and Applications, 68 (2017), pp. 289–315.

[34] H. K. Khalil, *Nonlinear systems; 3rd ed.*, Prentice-Hall, 2002, https://cds.cern.ch/record/1173048.

[35] K. Kunisch, S. Volkwein, and L. Xie, *Hjb-pod-based feedback design for the optimal control of evolution problems*, SIAM J. Appl. Dyn. Syst., 3 (2004), pp. 701–722.

[36] K. Kunisch and D. Vásquez-Varas, *Optimal polynomial feedback laws for finite horizon control problems*, Comput. Math. with Appl., 148 (2023), pp. 113–125.

[37] K. Kunisch and D. Vásquez-Varas, *Consistent smooth approximation of feedback laws for infinite horizon control problems with non-smooth value functions*, Journal of Differential Equations, 411 (2024), pp. 438–477.

[38] K. Kunisch, D. Vásquez-Varas, and D. Walter, *Learning optimal feedback operators and*

their sparse polynomial approximations, J. Mach. Learn. Res., 24 (2023), pp. 1–38.

[39] K. Kunisch and D. Walter, *Semiglobal optimal feedback stabilization of autonomous systems via deep neural network approximation*, ESAIM - Control Optim. Calc. Var., 27 (2021), p. 16.

[40] T. Nakamura-Zimmerer, Q. Gong, and W. Kang, *Adaptive deep learning for high-dimensional Hamilton–Jacobi–Bellman equations*, SIAM J. Sci. Comput., 43 (2021), pp. A1221—-A1247.

[41] T. Nakamura-Zimmerer, Q. Gong, and W. Kang, *Qrnet: Optimal regulator design with LQR-augmented neural networks*, EEE Control Syst. Lett., 5 (2021), pp. 1303–1308.

[42] N. Nüsken and L. Richter, *Solving high-dimensional Hamilton–Jacobi– Bellman pdes using neural networks: Perspectives from the theory of controlled diffusions and measures on path space*, (2020), https://arxiv.org/abs/2005.05409.

[43] D. Onken, L. Nurbekyan, X. Li, S. W. Fung, S. Osher, and L. Ruthotto, *A neural network approach applied to multi-agent optimal control*, in 2021 European Control Conference (ECC), 2021, pp. 1036–1041.

[44] D. Onken, L. Nurbekyan, X. Li, S. W. Fung, S. Osher, and L. Ruthotto, *A neural network approach for high-dimensional optimal control applied to multiagent path finding*, IEEE Trans. Control Syst. Technol., 31 (2023), pp. 235–251.

[45] M. Oster, L. Sallandt, and R. Schneider, *Approximating the stationary Hamilton–Jacobi–Bellman equation by hierarchical tensor products*, arXiv, (2019), https://arxiv.org/abs/1911.00279.

[46] M. Oster, L. Sallandt, and R. Schneider, *Approximating optimal feedback controllers of finite horizon control problems using hierarchical tensor formats*, SIAM J. Sci. Comput., 44 (2022), pp. B746–B770.

[47] M. Puterman and S. Brumelle, *On the convergence of policy iteration in stationary dynamic programming*, Math. Oper. Res., 4 (1979), pp. 60–69.

[48] L. Ruthotto, S. J. Osher, W. Li, L. Nurbekyan, and S. W. Fung, *A machine learning framework for solving high-dimensional mean field game and mean field control problems*, Proc. Natl. Acad. Sci., 117 (2020), pp. 9183–9193.

[49] M. Santos and J. Rust, *Convergence properties of policy iteration*, SIAM J. Control Optim., 42 (2004), pp. 2094–2115.

[50] W. Sickel and T. Ullrich, *Tensor products of sobolev–besov spaces and applications to approximation from the hyperbolic cross*, J. Approx. Theory, 161 (2009), pp. 748–786, https://doi.org/https://doi.org/10.1016/j.jat.2009.01.001, https://www.sciencedirect.com/science/article/pii/S0021904509000197.

[51] E. Stefansson and Y. P. Leong, *Sequential alternating least squares for solving high dimensional linear Hamilton–Jacobi–Bellman equation*, in 2016 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2016, pp. 3757—3764.

[52] Y. Zhao and J. Han, *Offline supervised learning v.s. online direct policy optimization: A comparative study and a unified training paradigm for neural network-based optimal feedback control*, 2023, https://arxiv.org/abs/2211.15930.