



# A Neural Network-Based Policy Iteration Algorithm with Global $H^2$ -Superlinear Convergence for Stochastic Games on Domains

Kazufumi Ito<sup>1</sup> · Christoph Reisinger<sup>2</sup> · Yufei Zhang<sup>2</sup>

Received: 14 June 2019 / Revised: 23 January 2020 / Accepted: 30 March 2020 / Published online: 18 May 2020  
© The Author(s) 2020

## Abstract

In this work, we propose a class of numerical schemes for solving semilinear Hamilton–Jacobi–Bellman–Isaacs (HJBI) boundary value problems which arise naturally from exit time problems of diffusion processes with controlled drift. We exploit policy iteration to reduce the semilinear problem into a sequence of linear Dirichlet problems, which are subsequently approximated by a multilayer feedforward neural network ansatz. We establish that the numerical solutions converge globally in the  $H^2$ -norm and further demonstrate that this convergence is superlinear, by interpreting the algorithm as an inexact Newton iteration for the HJBI equation. Moreover, we construct the optimal feedback controls from the numerical value functions and deduce convergence. The numerical schemes and convergence results are then extended to oblique derivative boundary conditions. Numerical experiments on the stochastic Zermelo navigation problem are presented to illustrate the theoretical results and to demonstrate the effectiveness of the method.

**Keywords** Hamilton–Jacobi–Bellman–Isaacs equations · Neural networks · Policy iteration · Inexact semismooth Newton method · Global convergence ·  $q$ -superlinear convergence

**Mathematics Subject Classification** 82C32 · 91A15 · 65M12

---

Communicated by Endre Suli.

---

✉ Yufei Zhang  
yufei.zhang@maths.ox.ac.uk

Kazufumi Ito  
kito@ncsu.edu

Christoph Reisinger  
christoph.reisinger@maths.ox.ac.uk

<sup>1</sup> Department of Mathematics, North Carolina State University, Raleigh, NC 27607, USA

<sup>2</sup> Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

## 1 Introduction

In this article, we propose a class of numerical schemes for solving Hamilton–Jacobi–Bellman–Isaacs (HJBI) boundary value problems of the following form:

$$-a^{ij}(x)\partial_{ij}u + G(x, u, \nabla u) = 0, \quad \text{in } \Omega \subset \mathbb{R}^n; \quad Bu = g, \quad \text{on } \partial\Omega, \quad (1.1)$$

where  $\Omega$  is an open bounded domain and  $G$  is the (nonconvex) Hamiltonian defined as

$$G(x, u, \nabla u) = \max_{\alpha \in \mathbf{A}} \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta)\partial_i u(x) + c(x, \alpha, \beta)u(x) - f(x, \alpha, \beta)), \quad (1.2)$$

with given nonempty compact sets  $\mathbf{A}, \mathbf{B}$ , and  $B$  is a boundary operator, i.e., if  $B$  is the identity operator, (1.1) is an HJBI Dirichlet problem, while if  $Bu = \gamma^i \partial_i u + \gamma^0 u$  with some functions  $\{\gamma^i\}_{i=0}^n$ , (1.1) is an HJBI oblique derivative problem. Above and hereafter, when there is no ambiguity, we shall adopt the summation convention as in [20], i.e., repeated equal dummy indices indicate summation from 1 to  $n$ .

It is well known that the value function of zero-sum stochastic differential games in domains satisfies the HJBI equation (1.1), and the optimal feedback controls can be constructed from the derivatives of the solutions (see e.g. [32] and references within; see also Sect. 6 for a concrete example). In particular, the HJBI Dirichlet problem corresponds to exit time problems of diffusion processes with controlled drift (see e.g. [10, 32, 37]), while the HJBI oblique derivative problem corresponds to state constraints (see e.g. [34, 36]). A nonconvex HJBI equation as above also arises from a penalty approximation of hybrid control problems involving continuous controls, optimal stopping and impulse controls, where the HJB (quasi-)variational inequality can be reduced to an HJBI equation by penalizing the difference between the value function and the obstacles (see e.g. [27, 39, 40, 47]). As (1.1) in general cannot be solved analytically, it is important to construct effective numerical schemes to find the solution of (1.1) and its derivatives.

The standard approach to solving (1.1) is to first discretize the operators in (1.1) by finite difference or finite element methods, and then solve the resulting nonlinear discretized equations by using policy iteration, also known as Howard’s algorithm, or generally (finite-dimensional) semismooth Newton methods (see e.g. [8, 18, 39, 45]). However, this approach has the following drawbacks, as do most mesh-based methods: (1) it can be difficult to generate meshes and to construct consistent numerical schemes for problems in domains with complicated geometries; (2) the number of unknowns in general grows exponentially with the dimension  $n$ , i.e., it suffers from Bellman’s *curse of dimensionality*, and hence, this approach is infeasible for solving high-dimensional control problems. Moreover, since policy iteration is applied to a fixed finite-dimensional equation resulting from a particular discretization, it is difficult to infer whether the same convergence rate of policy iteration remains valid as the mesh size tends to zero [8, 42]. We further remark that, for a given discrete HJBI equation, it can be difficult to determine a good initialization of policy iteration to

ensure fast convergence of the algorithm; see [2] and references therein on possible accelerated methods.

Recently, numerical methods based on deep neural networks have been designed to solve high-dimensional partial differential equations (PDEs) (see e.g. [6,15,16,26,35,44]). Most of these methods reformulate (1.1) into a nonlinear least-squares problem:

$$\inf_{u \in \mathcal{F}} \| -a^{ij} \partial_{ij} u + G(\cdot, u, \nabla u) \|_{L^2(\Omega)}^2 + \| Bu - g \|_{L^2(\partial\Omega)}^2, \quad (1.3)$$

where  $\mathcal{F}$  is a collection of neural networks with a smooth activation function. Based on collocation points chosen randomly from the domain, (1.3) is then reduced into an empirical risk minimization problem, which is subsequently solved by using stochastic optimization algorithms, in particular the stochastic gradient descent (SGD) algorithm or its variants. Since these methods avoid mesh generation, they can be adapted to solve PDEs in high-dimensional domains with complex geometries. Moreover, the choice of smooth activation functions leads to smooth numerical solutions, whose values can be evaluated everywhere without interpolations. In the following, we shall refer to these methods as the direct method, due to the fact that there is no policy iteration involved.

We observe, however, that the direct method also has several serious drawbacks, especially for solving nonlinear nonsmooth equations including (1.1). Firstly, the nonconvexity of both the deep neural networks and the Hamiltonian  $G$  leads to a nonconvex empirical minimization problem, for which there is no theoretical guarantee on the convergence of SGD to a minimizer (see e.g. [43]). In practice, training a network with a desired accuracy could take hours or days (with hundreds of thousands of iterations) due to the slow convergence of SGD. Secondly, each SGD iteration requires the evaluation of  $\nabla G$  (with respect to  $u$  and  $\nabla u$ ) on sample points, but  $\nabla G$  is not necessarily defined everywhere due to the nonsmoothness of  $G$ . Moreover, evaluating the function  $G$  (again on a large set of sample points) can be expensive, especially when the sets  $\mathbf{A}$  and  $\mathbf{B}$  are of high dimensions, as we do not require more regularity than continuity of the coefficients with respect to the controls, so that approximate optimization may only be achieved by exhaustive search over a discrete coverage of the compact control set. Finally, as we shall see in Remark 4.2, merely including an  $L^2(\partial\Omega)$ -norm of the boundary data in the loss function (1.3) does not generally lead to convergence of the derivatives of numerical solutions or the corresponding feedback control laws.

In this work, we propose an efficient neural network-based policy iteration algorithm for solving (1.1). At the  $(k + 1)$ th iteration,  $k \geq 0$ , we shall update the control laws  $(\alpha^k, \beta^k)$  by performing pointwise maximization/minimization of the Hamiltonian  $G$  based on the previous iterate  $u^k$ , and obtain the next iterate  $u^{k+1}$  by solving a linear boundary value problem, whose coefficients involve the control laws  $(\alpha^k, \beta^k)$ . This reduces the (nonconvex) semilinear problem into a sequence of *linear* boundary value problems, which are subsequently approximated by a multilayer neural network ansatz. Note that compared to Algorithm Ho-3 in [8] for discrete HJBI equations, which requires to solve a nonlinear HJB subproblem (involving minimization over the set  $\mathbf{B}$ ) for each iteration, our algorithm only requires to solve a linear subproblem for each iteration; hence, it is in general more efficient, especially when the dimension of  $\mathbf{B}$  is high.

Policy iteration (or successive Galerkin approximation) was employed in [4,5,28,30] to solve *convex HJB equations* on the whole space  $\mathbb{R}^n$ . Specifically, [4,5,28] approximate the solution to each linear equation via a separable polynomial ansatz (without concluding any convergence rate), while [30] assumes each linear equation is solved sufficiently accurately (without specifying a numerical method), and deduces pointwise *linear* convergence. The continuous policy iteration in [28] has also been applied to solve HJBI equations on  $\mathbb{R}^n$  in [29], which is a direct extension of Algorithm Ho-3 in [8] and still requires to solve a nonlinear HJB subproblem at each iteration. In this paper, we propose an easily implementable accuracy criterion for the numerical solutions of the *linear* PDEs which ensures the numerical solutions converge superlinearly in a suitable function space for nonconvex HJBI equations from an arbitrary initial guess.

Our algorithm enjoys the main advantage of the direct method, i.e., it is a mesh-free method and can be applied to solve high-dimensional stochastic games. Moreover, by utilizing the superlinear convergence of policy iteration, our algorithm effectively reduces the number of pointwise maximization/minimization over the sets  $\mathbf{A}$  and  $\mathbf{B}$ , and significantly reduces the computational cost of the direct method, especially for high dimensional control sets. The superlinear convergence of policy iteration also helps eliminate the oscillation caused by SGD, which leads to smoother and more rapidly decaying loss curves in both the training and validation processes (see Fig. 7). Our algorithm further allows training of the feedback controls on a separate network architecture from that representing the value function, or adaptively adjusting the architecture of networks for each policy iteration.

A major theoretical contribution of this work is the proof of global superlinear convergence of the policy iteration algorithm for the HJBI equation (1.1) in  $H^2(\Omega)$ , which is novel even for HJB equations (i.e., one of the sets  $\mathbf{A}$  and  $\mathbf{B}$  is singleton). Although the (local) superlinear convergence of policy iteration for discrete equations has been proved in various works (e.g. [8,18,38,39,42,45,47]), to the best of our knowledge, there is no published work on the superlinear convergence of policy iteration for HJB PDEs in a function space, nor on the global convergence of policy iteration for solving nonconvex HJBI equations.

Moreover, this is the first paper which demonstrates the convergence of neural network-based methods for the solutions and their (first and second order) derivatives of nonlinear PDEs with merely measurable coefficients (cf. [22,23,26,44]). We will also prove the pointwise convergence of the numerical solutions and their derivatives, which subsequently enables us to construct the optimal feedback controls from the numerical value functions and deduce convergence.

Let us briefly comment on the main difficulties encountered in studying the convergence of policy iteration for HJBI equations. Recall that at the  $(k+1)$ th iteration, we need to solve a linear boundary value problem, whose coefficients involve the control laws  $(\alpha^k, \beta^k)$ , obtained by performing pointwise maximization/minimization of the Hamiltonian  $G$ . The uncountability of the state space  $\Omega$  and the nonconvexity of the Hamiltonian require us to exploit several technical measurable selection arguments to ensure the measurability of the controls  $(\alpha^k, \beta^k)$ , which is essential for the well-definedness of the linear boundary value problems and the algorithm.

Moreover, the nonconvexity of the Hamiltonian prevents us from following the arguments in [8,18,42] for discrete HJB equations to establish the *global* convergence of our inexact policy iteration algorithm for HJBI equations. In fact, a crucial step in the arguments for discrete HJB equations is to use the discrete maximum principle and show the iterates generated by policy iteration converge monotonically with an arbitrary initial guess, which subsequently implies the global convergence of the iterates. However, this monotone convergence is in general false for the iterates generated by the inexact policy iteration algorithm, due to the nonconvexity of the Hamiltonian and the fact that each linear equation is only solved approximately. We shall present a novel analysis technique for establishing the global convergence of our inexact policy iteration algorithm, by interpreting it as a fixed point iteration in  $H^2(\Omega)$ .

Finally, we remark that the proof of *superlinear* convergence of our algorithm is significantly different from the arguments for discrete equations. Instead of working with the sup-norm for (finite-dimensional) discrete equations as in [8,18,39,42,47], we employ a two-norm framework to establish the generalized differentiability of HJBI operators, where the norm gap is essential as has already been pointed out in [24,45,46]. Moreover, by taking advantage of the fact that the Hamiltonian only involves low-order terms, we further demonstrate that the inverse of the generalized derivative is uniformly bounded. Furthermore, we include a suitable fractional Sobolev norm of the boundary data in the loss functions used in the training process, which is crucial for the  $H^2(\Omega)$ -superlinear convergence of the neural network-based policy iteration algorithm.

We organize this paper as follows. Section 2 states the main assumptions and recalls basic results for HJBI Dirichlet problems. In Sect. 3, we propose a policy iteration scheme for HJBI Dirichlet problems and establish its global superlinear convergence. Then, in Sect. 4, we shall introduce the neural network-based policy iteration algorithm, establish its various convergence properties, and construct convergent approximations to optimal feedback controls. We extend the algorithm and convergence results to HJBI oblique derivative problems in Sect. 5. Numerical examples for two-dimensional stochastic Zermelo navigation problems are presented in Sect. 6 to confirm the theoretical findings and to illustrate the effectiveness of our algorithms. Appendix collects some basic results which are used in this article and gives a proof for the main result on the HJBI oblique derivative problem.

## 2 HJBI Dirichlet Problems

In this section, we introduce the HJBI Dirichlet boundary value problems of our interest, recall the appropriate notion of solutions, and state the main assumptions on its coefficients. We start with several important spaces used frequently throughout this work.

Let  $n \in \mathbb{N}$  and  $\Omega$  be a bounded  $C^{1,1}$  domain in  $\mathbb{R}^n$ , i.e., a bounded open connected subset of  $\mathbb{R}^n$  with a  $C^{1,1}$  boundary. For each integer  $k \geq 0$  and real  $p$  with  $1 \leq p < \infty$ , we denote by  $W^{k,p}(\Omega)$  the standard Sobolev space of real functions with their weak derivatives of order up to  $k$  in the Lebesgue space  $L^p(\Omega)$ . When  $p = 2$ , we use  $H^k(\Omega)$  to denote  $W^{k,2}(\Omega)$ . We further denote by  $H^{1/2}(\partial\Omega)$  and  $H^{3/2}(\partial\Omega)$  the spaces

of traces from  $H^1(\Omega)$  and  $H^2(\Omega)$ , respectively (see [21, Proposition 1.1.17]), which can be equivalently defined by using the surface measure  $\sigma$  on the boundaries  $\partial\Omega$  as follows (see e.g. [19]):

$$\begin{aligned}\|g\|_{H^{1/2}(\partial\Omega)} &= \left[ \int_{\partial\Omega} |g|^2 d\sigma + \iint_{\partial\Omega \times \partial\Omega} \frac{|g(x) - g(y)|^2}{|x - y|^n} d\sigma(x) d\sigma(y) \right]^{1/2}, \\ \|g\|_{H^{3/2}(\partial\Omega)} &= \left[ \int_{\partial\Omega} (|g|^2 + \sum_{i=1}^n |\partial_i g|^2) d\sigma + \sum_{i=1}^n \iint_{\partial\Omega \times \partial\Omega} \frac{|\partial_i g(x) - \partial_i g(y)|^2}{|x - y|^n} d\sigma(x) d\sigma(y) \right]^{1/2}.\end{aligned}\quad (2.1)$$

We shall consider the following HJBI equation with nonhomogeneous Dirichlet boundary data:

$$F(u) := -a^{ij}(x)\partial_{ij}u + G(x, u, \nabla u) = 0, \quad \text{a.e. } \Omega, \quad (2.2a)$$

$$\tau u = g, \quad \text{on } \partial\Omega. \quad (2.2b)$$

where the nonlinear Hamiltonian is given as in (1.1):

$$G(x, u, \nabla u) = \max_{\alpha \in \mathbf{A}} \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta)\partial_i u(x) + c(x, \alpha, \beta)u(x) - f(x, \alpha, \beta)). \quad (2.3)$$

Throughout this paper, we shall focus on the strong solution to (2.2), i.e., a twice weakly differentiable function  $u \in H^2(\Omega)$  satisfying the HJBI equation (2.2a) almost everywhere in  $\Omega$ , and the boundary values on  $\partial\Omega$  will be interpreted as traces of the corresponding Sobolev space. For instance,  $\tau u = g$  on  $\partial\Omega$  in (2.2b) means that the trace of  $u$  is equal to  $g$  in  $H^{3/2}(\partial\Omega)$ , where  $\tau \in \mathcal{L}(H^2(\Omega), H^{3/2}(\partial\Omega))$  denotes the trace operator (see [19, Proposition 1.1.17]). See Sect. 5 for boundary conditions involving the derivatives of solutions.

We now list the main assumptions on the coefficients of (2.2).

**H.1** Let  $n \in \mathbb{N}$ ,  $\Omega \subset \mathbb{R}^n$  be a bounded  $C^{1,1}$  domain,  $\mathbf{A}$  be a nonempty finite set, and  $\mathbf{B}$  be a nonempty compact metric space. Let  $g \in H^{3/2}(\partial\Omega)$ ,  $\{a^{ij}\}_{i,j=1}^n \subseteq C(\bar{\Omega})$  satisfy the following ellipticity condition with a constant  $\lambda > 0$ :

$$\sum_{i,j=1}^n a^{ij}(x)\xi_i\xi_j \geq \lambda \sum_{i=1}^n \xi_i^2, \quad \text{for all } \xi \in \mathbb{R}^n \text{ and } x \in \Omega,$$

and  $\{b^i\}_{i=1}^n, c, f \in L^\infty(\Omega \times \mathbf{A} \times \mathbf{B})$  satisfy that  $c \geq 0$  on  $\Omega \times \mathbf{A} \times \mathbf{B}$ , and that  $\phi(x, \alpha, \cdot) : \mathbf{B} \rightarrow \mathbb{R}$  is continuous, for all  $\phi = b^i, c, f$  and  $(x, \alpha) \in \Omega \times \mathbf{A}$ .

As we shall see in Theorem 3.3 and Corollary 3.5, the finiteness of the set  $\mathbf{A}$  enables us to establish the semismoothness of the HJBI operator (2.2a), whose coefficients involve a general nonlinear dependence on the parameters  $\alpha$  and  $\beta$ . If all coefficients of (2.2a) are in a separable form, i.e., it holds for all  $\phi = b^i, c, f$  that  $\phi(x, \alpha, \beta) = \phi_1(x, \alpha) + \phi_2(x, \beta)$  for some functions  $\phi_1, \phi_2$  (e.g. the penalized equation for variational inequalities with bilateral obstacles in [27]), then we can relax the finiteness of  $\mathbf{A}$  to the same conditions on  $\mathbf{B}$ .

Finally, in this work we focus on boundary value problems in a  $C^{1,1}$  domain to simplify the presentation, but the numerical schemes and their convergence analysis can be extended to problems in nonsmooth convex domains with sufficiently regular coefficients (see e.g. [21,45]).

We end this section by proving the uniqueness of solutions to the Dirichlet problem (2.2) in  $H^2(\Omega)$ . The existence of strong solutions shall be established constructively via policy iteration below (see Theorem 3.7).

**Proposition 2.1** *Suppose (H.1) holds. Then, the Dirichlet problem (2.2) admits at most one strong solution  $u^* \in H^2(\Omega)$ .*

**Proof** Let  $u, v \in H^2(\Omega)$  be two strong solutions to (2.2), we consider the following linear homogeneous Dirichlet problem:

$$-a^{ij}(x)\partial_{ij}w + \tilde{b}^i(x)\partial_i w + \tilde{c}(x)w = 0, \quad \text{a.e. in } \Omega; \quad \tau w = 0, \quad \text{on } \partial\Omega, \quad (2.4)$$

where we define the following measurable functions: for each  $i = 1, \dots, n$ ,

$$\tilde{b}^i(x) = \begin{cases} \frac{G(x, v, ((\partial_j v)_{1 \leq j < i, \partial_i u, (\partial_j u)_{i < j \leq n})) - G(x, v, ((\partial_j v)_{1 \leq j < i, \partial_i v, (\partial_j u)_{i < j \leq n}))}{(\partial_i u - \partial_i v)(x)}, & \text{on } \{x \in \Omega \mid \partial_i(u - v)(x) \neq 0\}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\tilde{c}(x) = \begin{cases} \frac{G(x, u, \nabla u) - G(x, v, \nabla u)}{(u - v)(x)}, & \text{on } \{x \in \Omega \mid (u - v)(x) \neq 0\}, \\ 0, & \text{otherwise,} \end{cases}$$

with the Hamiltonian  $G$  defined as in (2.3). Note that the boundedness of coefficients implies that  $\{\tilde{b}^i\}_{i=1}^n \subseteq L^\infty(\Omega)$ , and  $\tilde{c} \in L^\infty(\Omega)$ . Moreover, one can directly verify that the following inequality holds for all parametrized functions  $(f^{\alpha, \beta}, g^{\alpha, \beta})_{\alpha \in \mathbf{A}, \beta \in \mathbf{B}}$ : for all  $x \in \mathbb{R}^n$ ,

$$\begin{aligned} \inf_{(\alpha, \beta) \in \mathbf{A} \times \mathbf{B}} f^{\alpha, \beta}(x) - g^{\alpha, \beta}(x) &\leq \inf_{\alpha \in \mathbf{A}} \sup_{\beta \in \mathbf{B}} f^{\alpha, \beta}(x) - \inf_{\alpha \in \mathbf{A}} \sup_{\beta \in \mathbf{B}} g^{\alpha, \beta}(x) \\ &\leq \sup_{(\alpha, \beta) \in \mathbf{A} \times \mathbf{B}} f^{\alpha, \beta}(x) - g^{\alpha, \beta}(x), \end{aligned}$$

which together with (H.1) leads to the estimate that  $\tilde{c}(x) \geq \inf_{(\alpha, \beta) \in \mathbf{A} \times \mathbf{B}} c(x, \alpha, \beta) \geq 0$  on the set  $\{x \in \Omega \mid (u - v)(x) \neq 0\}$ , and hence, we have  $\tilde{c} \geq 0$  a.e.  $\Omega$ . Then, we can deduce from Theorem A.1 that the Dirichlet problem (2.4) admits a unique strong solution  $w^* \in H^2(\Omega)$  and  $w^* = 0$ . Since  $w = u - v \in H^2(\Omega)$  satisfies (2.4) a.e. in  $\Omega$  and  $\tau w = 0$ , we see that  $w = u - v$  is a strong solution to (2.4) and hence  $u - v = w^* = 0$ , which subsequently implies the uniqueness of strong solutions to the Dirichlet problem (2.2).  $\square$

### 3 Policy Iteration for HJBI Dirichlet Problems

In this section, we propose a policy iteration algorithm for solving the Dirichlet problem (2.2). We shall also establish the global superlinear convergence of the algorithm,

which subsequently gives a constructive proof for the existence of a strong solution to the Dirichlet problem (2.2).

We start by presenting the policy iteration scheme for the HJBI equations in Algorithm 1, which extends the policy iteration algorithm (or Howard's algorithm) for discrete HJB equations (see e.g. [8, 18, 39]) to the continuous setting.

---

**Algorithm 1** Policy iteration algorithm for Dirichlet problems

---

1. Choose an initial guess  $u^0$  in  $H^2(\Omega)$ , and set  $k = 0$ .
2. Given the iterate  $u^k \in H^2(\Omega)$ , update the following control laws: for all  $\alpha \in \mathbf{A}$ ,  $x \in \Omega$ ,

$$\alpha^k(x) \in \arg \max_{\alpha \in \mathbf{A}} \left[ \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta) \partial_i u^k(x) + c(x, \alpha, \beta) u^k(x) - f(x, \alpha, \beta)) \right], \quad (3.1)$$

$$\beta^k(x) \in \arg \min_{\beta \in \mathbf{B}} (b^i(x, \alpha^k(x), \beta) \partial_i u^k(x) + c(x, \alpha^k(x), \beta) u^k(x) - f(x, \alpha^k(x), \beta)). \quad (3.2)$$

3. Solve the linear Dirichlet problem for  $u^{k+1} \in H^2(\Omega)$ :

$$-a^{ij} \partial_{ij} u + b_k^i \partial_i u + c_k u - f_k = 0, \quad \text{in } \Omega; \quad \tau u = g, \quad \text{on } \partial\Omega, \quad (3.3)$$

where  $\phi_k(x) := \phi(x, \alpha^k(x), \beta^k(x))$  for  $\phi = b^i, c, f$ .

4. If  $\|u^{k+1} - u^k\|_{H^2(\Omega)} = 0$ , then terminate with outputs  $u^{k+1}$ ,  $\alpha^k$  and  $\beta^k$ , otherwise increment  $k$  by one and go to step 2.
- 

The remaining part of this section is devoted to the convergence analysis of Algorithm 1. For notational simplicity, we first introduce two auxiliary functions: for each  $(x, \mathbf{u}, \alpha, \beta) \in \Omega \times \mathbb{R}^{n+1} \times \mathbf{A} \times \mathbf{B}$  with  $\mathbf{u} = (z, p) \in \mathbb{R} \times \mathbb{R}^n$ , we shall define the following functions

$$\ell(x, \mathbf{u}, \alpha, \beta) := b^i(x, \alpha, \beta) p_i + c(x, \alpha, \beta) z - f(x, \alpha, \beta), \quad (3.4)$$

$$h(x, \mathbf{u}, \alpha) := \min_{\beta \in \mathbf{B}} \ell(x, \mathbf{u}, \alpha, \beta). \quad (3.5)$$

Note that for all  $k \geq 0$  and  $x \in \Omega$ , by setting  $\mathbf{u}^k(x) = (u^k(x), \nabla u^k(x))$ , we can see from (3.1) and (3.2) that

$$\ell(x, \mathbf{u}^k(x), \alpha^k(x), \beta^k(x)) = \min_{\beta \in \mathbf{B}} \ell(x, \mathbf{u}^k(x), \alpha^k(x), \beta) = \max_{\alpha \in \mathbf{A}} \min_{\beta \in \mathbf{B}} \ell(x, \mathbf{u}^k(x), \alpha, \beta). \quad (3.6)$$

We then recall several important concepts, which play a pivotal role in our subsequent analysis. The first concept ensures the existence of measurable feedback controls and the well-posedness of Algorithm 1.

**Definition 3.1** Let  $(S, \Sigma)$  be a measurable space, and let  $X$  and  $Y$  be topological spaces. A function  $\psi : S \times X \rightarrow Y$  is a Carathéodory function if:



1. for each  $x \in X$ , the function  $\psi_x = \psi(\cdot, x) : S \rightarrow Y$  is  $(\Sigma, \mathcal{B}_Y)$ -measurable, where  $\mathcal{B}_Y$  is the Borel  $\sigma$ -algebra of the topological space  $Y$ ; and
2. for each  $s \in S$ , the function  $\psi_s = \psi(s, \cdot) : X \rightarrow Y$  is continuous.

**Remark 3.1** It is well known that if  $X, Y$  are two complete separable metric spaces, and  $\psi : S \times X \rightarrow Y$  is a Carathéodory function, then for any given measurable function  $f : S \rightarrow X$ , the composition function  $s \rightarrow \psi(s, f(s))$  is measurable (see e.g. [3, Lemma 8.2.3]). Since any compact metric space is complete and separable, it is clear that (H.1) implies that the coefficients  $b^i, c, f$  are Carathéodory functions (with  $S = \Omega$  and  $X = \mathbf{A} \times \mathbf{B}$ ). Moreover, one can easily check that both  $\ell$  and  $h$  are Carathéodory functions, i.e.,  $\ell$  (resp.  $h$ ) is continuous in  $(\mathbf{u}, \alpha, \beta)$  (resp.  $(\mathbf{u}, \alpha)$ ) and measurable in  $x$  (see Theorem A.3 for the measurability of  $h$  in  $x$ ).

We now recall a generalized differentiability concept for nonsmooth operators between Banach spaces, which is referred as semismoothness in [46] and slant differentiability in [12, 24]. It is well known (see e.g. [8, 39, 45]) that the HJBI operator in (2.2a) is in general non-Fréchet-differentiable, and this generalized differentiability is essential for showing the superlinear convergence of policy iteration applied to HJBI equations.

**Definition 3.2** Let  $F : V \subset Y \mapsto Z$  be defined on a open subset  $V$  of the Banach space  $Y$  with images in the Banach space  $Z$ . In addition, let  $\partial^* F : V \rightrightarrows \mathcal{L}(Y, Z)$  be a given a set-valued mapping with nonempty images, i.e.,  $\partial^* F(y) \neq \emptyset$  for all  $y \in V$ . We say  $F$  is  $\partial^* F$ -semismooth in  $V$  if for any given  $y \in V$ , we have that  $F$  is continuous near  $y$ , and

$$\sup_{M \in \partial^* F(y+s)} \|F(y+s) - F(y) - Ms\|_Z = o(\|s\|_Y), \quad \text{as } \|s\|_Y \rightarrow 0.$$

The set-valued mapping  $\partial^* F$  is called a generalized differential of  $F$  in  $V$ .

**Remark 3.2** As in [46], we always require that  $\partial^* F$  has a nonempty image, and hence the  $\partial^* F$ -semismooth of  $F$  in  $V$  shall automatically imply that the image of  $\partial^* F$  is nonempty on  $V$ .

Now we are ready to analyze Algorithm 1. We first prove the semismoothness of the Hamiltonian  $G$  defined as in (2.3), by viewing it as the composition of a pointwise maximum operator and a family of HJB operators parameterized by the control  $\alpha$ . Moreover, we shall simultaneously establish that, for each iteration, one can select measurable control laws  $\alpha^k, \beta^k$  to ensure the measurability of the controlled coefficients  $b_k^i, c_k, f_k$  in the linear problem (3.3), which is essential for the well-posedness of strong solutions to (3.3), and the well-definedness of Algorithm 1.

The following proposition establishes the semismoothness of a parameterized family of first-order HJB operators, which extends the result for scalar-valued HJB operators in [45]. Moreover, by taking advantage of the fact that the operators involve only first-order derivatives, we are able to establish that they are semismooth from  $H^2(\Omega)$  to  $L^p(\Omega)$  for some  $p > 2$  (cf. [45, Theorem 13]), which is essential for the superlinear convergence of Algorithm 1.

**Proposition 3.1** Suppose (H.1) holds. Let  $p$  be a given constant satisfying  $p \geq 1$  if  $n \leq 2$  and  $p \in [1, \frac{2n}{n-2}]$  if  $n > 2$ , and let  $F_1 : H^2(\Omega) \rightarrow (L^p(\Omega))^{|A|}$  be the HJB operator defined by

$$F_1(u) := \left( \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta) \partial_i u + c(x, \alpha, \beta)u - f(x, \alpha, \beta)) \right)_{\alpha \in \mathbf{A}}, \quad \forall u \in H^2(\Omega).$$

Then,  $F_1$  is Lipschitz continuous and  $\partial^* F_1$ -semismooth in  $H^2(\Omega)$  with a generalized differential

$$\partial^* F_1 : H^2(\Omega) \rightarrow \mathcal{L}(H^2(\Omega), (L^p(\Omega))^{|A|})$$

defined as follows: for any  $u \in H^2(\Omega)$ , we have

$$\partial^* F_1(u) := \left( b^i(\cdot, \alpha, \beta^u(\cdot, \alpha)) \partial_i + c(\cdot, \alpha, \beta^u(\cdot, \alpha)) \right)_{\alpha \in \mathbf{A}}, \quad (3.7)$$

where  $\beta^u : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$  is any jointly measurable function such that for all  $\alpha \in \mathbf{A}$  and  $x \in \Omega$ ,

$$\beta^u(x, \alpha) \in \arg \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta) \partial_i u(x) + c(x, \alpha, \beta)u(x) - f(x, \alpha, \beta)). \quad (3.8)$$

**Proof** Since  $\mathbf{A}$  is a finite set, we shall assume without loss of generality that, the Banach space  $(L^p(\Omega))^{|A|}$  is endowed with the usual product norm  $\|\cdot\|_{p, \mathbf{A}}$ , i.e., for all  $u \in (L^p(\Omega))^{|A|}$ ,  $\|u\|_{p, \mathbf{A}} = \sum_{\alpha \in \mathbf{A}} \|u(\cdot, \alpha)\|_{L^p(\Omega)}$ . Note that the Sobolev embedding theorem shows that the following injections are continuous:  $H^2(\Omega) \hookrightarrow W^{1,q}(\Omega)$ , for all  $q \geq 2$ ,  $n \leq 2$ , and  $H^2(\Omega) \hookrightarrow W^{1,2n/(n-2)}(\Omega)$ , for all  $n > 2$ . Thus for any given  $p$  satisfying the conditions in Proposition 3.1, we can find  $r \in (p, \infty)$  such that the injection  $H^2(\Omega) \hookrightarrow W^{1,r}(\Omega)$  is continuous. Then, the boundedness of  $b^i, c, f$  implies that the mappings  $F_1$  and  $\partial^* F_1$  are well defined, and  $F_1 : H^2(\Omega) \rightarrow (L^p(\Omega))^{|A|}$  is Lipschitz continuous.

Now we show that the mapping  $\partial^* F_1$  has a nonempty image from  $W^{1,r}(\Omega)$  to  $(L^p(\Omega))^{|A|}$ , where we choose  $r \in (p, \infty)$  such that the injection  $H^2(\Omega) \hookrightarrow W^{1,r}(\Omega)$  is continuous, and naturally extend the operators  $F_1$  and  $\partial^* F_1$  from  $H^2(\Omega)$  to  $W^{1,r}(\Omega)$ . For each  $u \in W^{1,r}(\Omega)$ , we consider the Carathéodory function  $g : \Omega \times \mathbf{A} \times \mathbf{B} \rightarrow \mathbb{R}$  such that  $g(x, \alpha, \beta) := \ell(x, (u, \nabla u)(x), \alpha, \beta)$  for all  $(x, \alpha, \beta) \in \Omega \times \mathbf{A} \times \mathbf{B}$ , where  $\ell$  is defined by (3.5). Theorem A.3 shows there exists a function  $\beta^u : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$  satisfying (3.8), i.e.,

$$\beta^u(x, \alpha) \in \arg \min_{\beta \in \mathbf{B}} \ell(x, (u(x), \nabla u(x)), \alpha, \beta), \quad \forall (x, \alpha) \in \Omega \times \mathbf{A},$$

and  $\beta^u$  is jointly measurable with respect to the product  $\sigma$ -algebra on  $\Omega \times \mathbf{A}$ . Hence  $\partial^* F_1(u)$  is nonempty for all  $u \in W^{1,r}(\Omega)$ .

We proceed to show that the operator  $F_1$  is in fact  $\partial^* F_1$ -semismooth from  $W^{1,r}(\Omega)$  to  $(L^p(\Omega))^{|A|}$ , which implies the desired conclusion due to the continuous embedding

$H^2(\Omega) \hookrightarrow W^{1,r}(\Omega)$ . For each  $\alpha \in \mathbf{A}$ , we denote by  $F_{1,\alpha} : W^{1,r}(\Omega) \rightarrow L^p(\Omega)$  the  $\alpha$ -th component of  $F_1$ , and by  $\partial^* F_{1,\alpha}$  the  $\alpha$ -th component of  $\partial^* F_1$ . Theorem A.4 and the continuity of  $\ell$  in  $\mathbf{u}$  show that for each  $(x, \alpha) \in \Omega \times \mathbf{A}$ , the set-valued mapping

$$\mathbf{u} \in \mathbb{R}^{n+1} \mapsto \arg \min_{\beta \in \mathbf{B}} \ell(x, \mathbf{u}, \alpha, \beta) \subseteq \mathbf{B},$$

is upper hemicontinuous, from which, by following precisely the steps in the arguments for [45, Theorem 13], we can prove that  $F_{1,\alpha} : W^{1,r}(\Omega) \rightarrow L^p(\Omega)$  is  $\partial^* F_{1,\alpha}$ -semismooth. Then, by using the fact that a direct product of semismooth operators is again semismooth with respect to the direct product of the generalized differentials of the components (see [46, Proposition 3.6]), we can deduce that  $F_1 : W^{1,r}(\Omega) \rightarrow (L^p(\Omega))^{|A|}$  is semismooth with respect to the generalized differential  $\partial^* F_1$  and finishes the proof.  $\square$

We then establish the semismoothness of a general pointwise maximum operator, by extending the result in [24] for the max-function  $f : x \in \mathbb{R} \rightarrow \max(x, 0)$ .

**Proposition 3.2** *Let  $p \in (2, \infty)$  be a given constant,  $A$  be a finite set, and  $\Omega$  be a bounded subset of  $\mathbb{R}^n$ . Let  $F_2 : (L^p(\Omega))^{|A|} \rightarrow L^2(\Omega)$  be the pointwise maximum operator such that for each  $u = (u(\cdot, \alpha))_{\alpha \in A} \in (L^p(\Omega))^{|A|}$ ,*

$$F_2(u)(x) := \max_{\alpha \in A} u(x, \alpha), \quad \text{for a.e. } x \in \Omega. \quad (3.9)$$

*Then,  $F_2$  is  $\partial^* F_2$ -semismooth in  $(L^p(\Omega))^{|A|}$  with a generalized differential*

$$\partial^* F_2 : (L^p(\Omega))^{|A|} \rightarrow \mathcal{L}((L^p(\Omega))^{|A|}, L^2(\Omega))$$

*defined as follows: for any  $u = (u(\cdot, \alpha))_{\alpha \in A}$ ,  $v = (v(\cdot, \alpha))_{\alpha \in A} \in (L^p(\Omega))^{|A|}$ , we have*

$$(\partial^* F_2(u)v)(x) := v(x, \alpha^u(x)), \quad \text{for } x \in \Omega,$$

*where  $\alpha^u : \Omega \rightarrow A$  is any measurable function such that*

$$\alpha^u(x) \in \arg \max_{\alpha \in A} (u(x, \alpha)), \quad \text{for } x \in \Omega. \quad (3.10)$$

*Moreover,  $\partial^* F_2(u)$  is uniformly bounded (in the operator norm) for all  $u \in (L^p(\Omega))^{|A|}$ .*

**Proof** Let the Banach space  $(L^p(\Omega))^{|A|}$  be endowed with the product norm  $\|\cdot\|_{p,A}$  defined as in the proof of Proposition 3.1. We first show the mappings  $F_2$  and  $\partial^* F_2$  are well defined,  $\partial^* F_2$  has nonempty images, and  $\partial^* F_2(u)$  is uniformly bounded for  $u \in (L^p(\Omega))^{|A|}$ .

The finiteness of  $A$  implies that any  $u \in (L^p(\Omega))^{|A|}$  can also be viewed as a Carathéodory function  $u : \Omega \times A \rightarrow \mathbb{R}$ . Hence for any given  $u \in (L^p(\Omega))^{|A|}$ , we

can deduce from Theorem A.3 the existence of a measurable function  $\alpha^u : \Omega \rightarrow \mathbf{A}$  satisfying (3.10). Moreover, for any given measurable function  $\alpha^u : \Omega \rightarrow \mathbf{A}$  and  $v \in (L^p(\Omega))^{|A|}$ , the function  $\partial^* F_2(u)v$  remains Lebesgue measurable (see Remark 3.1). Then, for any given  $u \in (L^p(\Omega))^{|A|}$  with  $p > 2$ , one can easily check that  $F_2(u) \in L^2(\Omega)$ , and  $\partial^* F_2(u) \in \mathcal{L}((L^p(\Omega))^{|A|}, L^2(\Omega))$ , which subsequently implies that  $F_2$  and  $\partial^* F_2$  are well defined, and the image of  $\partial^* F_2$  is nonempty on  $(L^p(\Omega))^{|A|}$ . Moreover, for any  $u, v \in (L^p(\Omega))^{|A|}$ , Hölder's inequality leads to the following estimate:

$$\int_{\Omega} |v(x, \alpha^u(x))|^2 dx \leq \int_{\Omega} \sum_{\alpha \in \mathbf{A}} |v(x, \alpha)|^2 dx \leq \sum_{\alpha \in \mathbf{A}} |\Omega|^{(p-2)/p} \|v(\cdot, \alpha)\|_{L^p(\Omega)}^2,$$

which shows that  $\|\partial^* F_2(u)\|_{\mathcal{L}((L^p(\Omega))^{|A|}, L^2(\Omega))} \leq |\Omega|^{(p-2)/(2p)}$  for all  $u \in (L^p(\Omega))^{|A|}$ .

Now we prove by contradiction that the operator  $F_2$  is  $\partial^* F_2$ -semismooth. Suppose there exists a constant  $\delta > 0$  and functions  $u, \{v_k\}_{k=1}^{\infty} \in (L^p(\Omega))^{|A|}$  such that  $\|v_k\|_{p, \mathbf{A}} \rightarrow 0$  as  $k \rightarrow \infty$ , and

$$\|F_2(u + v_k) - F_2(u) - \partial^* F_2(u + v_k)v_k\|_{L^2(\Omega)} / \|v_k\|_{p, \mathbf{A}} \geq \delta > 0, \quad k \in \mathbb{N}, \quad (3.11)$$

where for each  $k \in \mathbb{N}$ ,  $\partial^* F_2(u + v_k)$  is defined with some measurable function  $\alpha^{u+v_k} : \Omega \rightarrow \mathbf{A}$ . Then, by passing to a subsequence, we may assume that for all  $\alpha \in \mathbf{A}$ , the sequence  $\{v_k(\cdot, \alpha)\}_{k \in \mathbb{N}}$  converges to zero pointwise a.e. in  $\Omega$ , as  $k \rightarrow \infty$ .

For notational simplicity, we define  $\Sigma(x, u) := \arg \max_{\alpha \in \mathbf{A}} (u(x, \alpha))$  for all  $u \in (L^p(\Omega))^{|A|}$  and  $x \in \Omega$ . Then, for a.e.  $x \in \Omega$ , we have  $\lim_{k \rightarrow \infty} v_k(x, \alpha) = 0$  for all  $\alpha \in \mathbf{A}$ ,  $\alpha^{u+v_k}(x) \in \Sigma(x, u + v_k)$  for all  $k \in \mathbb{N}$ . By using the finiteness of  $\mathbf{A}$  and the convergence of  $\{v_k(\cdot, \alpha)\}_{k \in \mathbb{N}}$ , it is straightforward to prove by contradiction that for all such  $x \in \Omega$ ,  $\alpha^{u+v_k}(x) \in \Sigma(x, u)$  for all large enough  $k$ .

We now derive an upper bound of the left-hand side of (3.11). For a.e.  $x \in \Omega$ , we have

$$\begin{aligned} & F_2(u + v_k)(x) - F_2(u)(x) - (\partial^* F_2(u + v_k)v_k)(x) \\ & \leq (u + v_k)(x, \alpha^{u+v_k}(x)) - u(x, \alpha^{u+v_k}(x)) - v_k(x, \alpha^{u+v_k}(x)) = 0, \\ & F_2(u + v_k)(x) - F_2(u)(x) - (\partial^* F_2(u + v_k)v_k)(x) \\ & \geq (u + v_k)(x, \alpha^u(x)) - u(x, \alpha^u(x)) - v_k(x, \alpha^{u+v_k}(x)) \\ & = v_k(x, \alpha^u(x)) - v_k(x, \alpha^{u+v_k}(x)), \end{aligned}$$

from any  $\alpha^u(x) \in \Sigma(x, u)$ . Thus, for each  $k \in \mathbb{N}$ , we have for a.e.  $x \in \Omega$  that,

$$\begin{aligned} & |F_2(u + v_k)(x) - F_2(u)(x) - (\partial^* F_2(u + v_k)v_k)(x)| \\ & \leq \phi_k(x) := \inf_{\alpha^u \in \Sigma(x, u)} |v_k(x, \alpha^u) - v_k(x, \alpha^{u+v_k}(x))|, \end{aligned}$$

where, by applying Theorem A.3 twice, we can see that both the set-valued mapping  $x \mapsto \Sigma(x, u)$  and the function  $\phi_k$  are measurable.

We then introduce the set  $\Omega_k = \{x \in \Omega \mid \alpha^{u+v_k}(x) \notin \Sigma(x, u)\}$  for each  $k \in \mathbb{N}$ . The measurability of the set-valued mapping  $x \mapsto \Sigma(x, u)$  implies the associated distance function  $\rho(x, \alpha) := \text{dist}(\alpha, \Sigma(x, u))$  is a Carathéodory function (see [1, Theorem 18.5]), which subsequently leads to the measurability of  $\Omega_k$  for all  $k$ . Hence, we can deduce that

$$\begin{aligned} & \|F_2(u + v_k) - F_2(u) - \partial^* F_2(u + v_k)v_k\|_{L^2(\Omega)}^2 \\ & \leq \int_{\Omega_k} \inf_{\alpha^u \in \Sigma(x, u)} |v_k(x, \alpha^u) - v_k(x, \alpha^{u+v_k}(x))|^2 dx \\ & \leq 2 \int_{\Omega_k} \sum_{\alpha \in A} |v_k(x, \alpha)|^2 dx \leq 2 \sum_{\alpha \in A} |\Omega_k|^{(p-2)/p} \|v_k(\cdot, \alpha)\|_{L^p(\Omega)}^2, \end{aligned}$$

which leads to the following estimate:

$$\begin{aligned} & \|F_2(u + v_k) - F_2(u) - \partial^* F_2(u + v_k)v_k\|_{L^2(\Omega)} / \|v_k\|_{p, A} \\ & \leq \sqrt{2} |\Omega_k|^{(p-2)/(2p)} \rightarrow 0, \quad \text{as } k \rightarrow \infty, \end{aligned}$$

where we have used the bounded convergence theorem and the fact that for a.e.  $x \in \Omega$ ,  $1_{\Omega_k}(x) = 0$  for all large enough  $k$ . This contradicts to the hypothesis (3.11) and hence finishes our proof.  $\square$

Now we are ready to conclude the semismoothness of the HJBI operator. Note that the argument in [45] does not apply directly to the HJBI operator, due to the nonconvexity of the Hamiltonian  $G$  defined as in (2.3).

**Theorem 3.3** *Suppose (H.1) holds, and let  $F : H^2(\Omega) \rightarrow L^2(\Omega)$  be the HJBI operator defined as in (2.2a). Then,  $F$  is semismooth in  $H^2(\Omega)$ , with a generalized differential  $\partial^* F : H^2(\Omega) \rightarrow \mathcal{L}(H^2(\Omega), L^2(\Omega))$  defined as follows: for any  $u \in H^2(\Omega)$ ,*

$$\partial^* F(u) := -a^{ij}(\cdot) \partial_{ij} + b^i(\cdot, \alpha(\cdot), \beta^u(\cdot, \alpha(\cdot))) \partial_i + c(\cdot, \alpha(\cdot), \beta^u(\cdot, \alpha(\cdot))), \quad (3.12)$$

where  $\beta^u : \Omega \times A \rightarrow B$  is any jointly measurable function satisfying (3.8), and  $\alpha : \Omega \rightarrow A$  is any measurable function such that for a.e.  $x \in \Omega$ .

$$\alpha(x) \in \arg \max_{\alpha \in A} \left[ \min_{\beta \in B} \left( b^i(x, \alpha, \beta) \partial_i u(x) + c(x, \alpha, \beta) u(x) - f(x, \alpha, \beta) \right) \right], \quad (3.13)$$

**Proof** Note that we can decompose the HJBI operator  $F : H^2(\Omega) \rightarrow L^2(\Omega)$  into  $F = F_0 + F_2 \circ F_1$ , where  $F_0 : H^2(\Omega) \rightarrow L^2(\Omega)$  is the linear operator  $u \mapsto -a^{ij} \partial_{ij} u$ ,  $F_1 : H^2(\Omega) \rightarrow (L^p(\Omega))^{|A|}$  is the HJB operator defined in Proposition 3.1,  $F_2 : (L^p(\Omega))^{|A|} \rightarrow L^2(\Omega)$  is the pointwise maximum operator defined in Proposition 3.2, and  $p$  is a constant satisfying  $p > 2$  if  $n \leq 2$ , and  $p \in (2, 2n/(n-2))$  if  $n > 2$ .

Proposition 3.1 shows that  $F_1$  is Lipschitz continuous and semismooth with respect to the generalized differential  $\partial^* F_1$  defined by (3.7), while Proposition 3.2 shows

that  $F_2$  is semismooth with respect to the uniformly bounded generalized differential  $\partial^* F_2$  defined by (3.9). Hence, we know the composed operator  $F_2 \circ F_1$  is semismooth with respect to the composition of the generalized differentials (see [46, Proposition 3.8]), i.e.,  $\partial^*(F_2 \circ F_1)(u) = \partial^* F_2(F_1(u)) \circ \partial^* F_1(u)$  for all  $u \in H^2(\Omega)$ . Consequently, by using the fact that  $F_0$  is Fréchet differentiable with the derivative  $-a^{ij} \partial_{ij} \in \mathcal{L}(H^2(\Omega), L^2(\Omega))$ , we can conclude from Propositions 3.1 and 3.2 that  $F : H^2(\Omega) \rightarrow L^2(\Omega)$  is semismooth on  $H^2(\Omega)$ , and that (3.12) is a desired generalized differential of  $F$  at  $u$ .  $\square$

Note that the above characterization of the generalized differential of the HJBI operator involves a jointly measurable function  $\beta^u : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$ , satisfying (3.8) for all  $(x, \alpha) \in \Omega \times \mathbf{A}$ . We now present a technical lemma, which allows us to view the control law  $\beta^k$  in (3.2) as such a feedback control on  $x \in \Omega$  and  $\alpha \in \mathbf{A}$ .

**Lemma 3.4** *Suppose (H.1) holds. Let  $h, \{h_i\}_{i=1}^n : \Omega \rightarrow \mathbb{R}$ ,  $\alpha^h : \Omega \rightarrow \mathbf{A}$  be given measurable functions, and  $\beta^h : \Omega \rightarrow \mathbf{B}$  be a measurable function such that for all  $x \in \Omega$ ,*

$$\beta^h(x) \in \arg \min_{\beta \in \mathbf{B}} \left( b^i(x, \alpha^h(x), \beta) h_i(x) + c(x, \alpha^h(x), \beta) h(x) - f(x, \alpha^h(x), \beta) \right). \quad (3.14)$$

*Then, there exists a jointly measurable function  $\tilde{\beta}^h : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$  such that  $\beta^h(x) = \tilde{\beta}^h(x, \alpha^h(x))$  for all  $x \in \Omega$ , and it holds for all  $x \in \Omega$  and  $\alpha \in \mathbf{A}$  that*

$$\tilde{\beta}^h(x, \alpha) \in \arg \min_{\beta \in \mathbf{B}} \left( b^i(x, \alpha, \beta) h_i(x) + c(x, \alpha, \beta) h(x) - f(x, \alpha, \beta) \right). \quad (3.15)$$

**Proof** Let  $\beta^h : \Omega \rightarrow \mathbf{B}$  be a given measurable function satisfying (3.14) for all  $x \in \Omega$  (see Remark 3.1 and Theorem A.3 for the existence of such a measurable function). As shown in the proof of Proposition 3.1, there exists a jointly measurable function  $\tilde{\beta} : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$  satisfying the property (3.15) for all  $(x, \alpha) \in \Omega \times \mathbf{A}$ . Now suppose that  $\mathbf{A} = \{\alpha_i\}_{i=1}^{|\mathbf{A}|}$  with  $|\mathbf{A}| < \infty$  (see (H.1)), we shall define the function  $\tilde{\beta}^h(x, \alpha) : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$ , such that for all  $(x, \alpha) \in \Omega \times \mathbf{A}$ ,

$$\tilde{\beta}^h(x, \alpha) = \begin{cases} \beta^h(x), & (x, \alpha) \in \mathcal{C} := \bigcup_{i=1}^{|\mathbf{A}|} (\{x \in \Omega \mid \alpha^h(x) = \alpha_i\} \times \{\alpha_i\}), \\ \tilde{\beta}(x, \alpha), & \text{otherwise.} \end{cases}$$

The measurability of  $\alpha^h$  and the finiteness of  $\mathbf{A}$  imply that the set  $\mathcal{C}$  is measurable in the product  $\sigma$ -algebra on  $\Omega \times \mathbf{A}$ , which along with the joint measurability of  $\tilde{\beta}$  leads to the joint measurability of the function  $\tilde{\beta}^h$ .

For any given  $x \in \Omega$ , we have  $(x, \alpha^h(x)) \in \{y \in \Omega \mid \alpha^h(y) = \alpha^h(x)\} \times \{\alpha^h(x)\}$ , from which we can deduce from the definition of  $\tilde{\beta}^h$  that  $\tilde{\beta}^h(x, \alpha^h(x)) = \beta^h(x)$  for all  $x \in \Omega$ . Finally, for any given  $\alpha_i \in \mathbf{A}$ , we shall verify (3.15) for all  $x \in \Omega$  and  $\alpha = \alpha_i$ . Let  $x \in \Omega$  be fixed. If  $\alpha^h(x) = \alpha_i$ , then the fact that  $(x, \alpha_i) \in \mathcal{C}$  and the definition of  $\tilde{\beta}^h$  imply that  $\tilde{\beta}^h(x, \alpha_i) = \beta^h(x)$ , which along with (3.14) and  $\alpha^h(x) = \alpha_i$  shows that

(3.15) holds for the point  $(x, \alpha_i)$ . On the other hand, if  $\alpha^h(x) \neq \alpha_i$ , then  $(x, \alpha_i) \notin \mathcal{C}$  and  $\tilde{\beta}^h(x, \alpha_i) = \tilde{\beta}(x, \alpha_i)$  satisfies the condition (3.15) due to the selection of  $\tilde{\beta}$ .  $\square$

As a direct consequence of the above extension result, we now present an equivalent characterization of the generalized differential of the HJBI operator.

**Corollary 3.5** *Suppose (H.1) holds, and let  $F : H^2(\Omega) \rightarrow L^2(\Omega)$  be the HJBI operator defined as in (2.2a). Then,  $F$  is semismooth in  $H^2(\Omega)$ , with a generalized differential  $\partial^* F : H^2(\Omega) \rightarrow \mathcal{L}(H^2(\Omega), L^2(\Omega))$  defined as follows: for any  $u \in H^2(\Omega)$ ,*

$$\partial^* F(u) := -a^{ij}(\cdot) \partial_{ij} + b^i(\cdot, \alpha^u(\cdot), \beta^u(\cdot)) \partial_i + c(\cdot, \alpha^u(\cdot), \beta^u(\cdot)), \quad (3.16)$$

where  $\alpha^u : \Omega \rightarrow \mathbf{A}$  and  $\beta^u : \Omega \rightarrow \mathbf{B}$  are any measurable functions satisfying for all  $x \in \Omega$  that

$$\begin{aligned} \alpha^u(x) &\in \arg \max_{\alpha \in \mathbf{A}} \left[ \min_{\beta \in \mathbf{B}} \left( b^i(x, \alpha, \beta) \partial_i u(x) + c(x, \alpha, \beta) u(x) - f(x, \alpha, \beta) \right) \right], \\ \beta^u(x) &\in \arg \min_{\beta \in \mathbf{B}} \left( b^i(x, \alpha^u(x), \beta) \partial_i u(x) + c(x, \alpha^u(x), \beta) u(x) - f(x, \alpha^u(x), \beta) \right). \end{aligned} \quad (3.17)$$

**Proof** Let  $u \in H^2(\Omega)$ , and let  $\alpha^u$  and  $\beta^u$  be given measurable functions satisfying (3.17) (see Remark 3.1 and Theorem A.3 for the existence of such measurable functions). Then, by using Lemma 3.4, we know there exists a jointly measurable function  $\tilde{\beta}^u : \Omega \times \mathbf{A} \rightarrow \mathbf{B}$  such that  $\tilde{\beta}^u$  satisfies (3.8) for all  $(x, \alpha) \in \Omega \times \mathbf{A}$ , and  $\tilde{\beta}^u(x, \alpha^u(x)) = \beta^u(x)$  for all  $x \in \Omega$ . Hence, we see the linear operator defined in (3.16) is equal to the following operator

$$\begin{aligned} &-a^{ij}(\cdot) \partial_{ij} + b^i(\cdot, \alpha^u(\cdot), \tilde{\beta}^u(\cdot, \alpha^u(\cdot))) \partial_i \\ &+ c(\cdot, \alpha^u(\cdot), \tilde{\beta}^u(\cdot, \alpha^u(\cdot))) \in \mathcal{L}(H^2(\Omega), L^2(\Omega)), \end{aligned}$$

which is a generalized differential of the HJBI operator  $F$  at  $u$  due to Theorem 3.3.  $\square$

The above characterization of the generalized differential of the HJBI operator enables us to demonstrate the superlinear convergence of Algorithm 1 by reformulating it as a semismooth Newton method for an operator equation.

**Theorem 3.6** *Suppose (H.1) holds and let  $u^* \in H^2(\Omega)$  be a strong solution to the Dirichlet problem (2.2). Then, there exists a neighborhood  $\mathcal{N}$  of  $u^*$ , such that for all  $u^0 \in \mathcal{N}$ , Algorithm 1 either terminates with  $u^k = u^*$  for some  $k \in \mathbb{N}$  or generates a sequence  $\{u^k\}_{k \in \mathbb{N}}$  that converges  $q$ -superlinearly to  $u^*$  in  $H^2(\Omega)$ , i.e.,  $\lim_{k \rightarrow \infty} \|u^{k+1} - u^*\|_{H^2(\Omega)} / \|u^k - u^*\|_{H^2(\Omega)} = 0$ .*

**Proof** Note that the Dirichlet problem (2.2) can be written as an operator equation  $\tilde{F}(u) = 0$  with the following operator

$$\tilde{F} : u \in H^2(\Omega) \rightarrow (F(u), \tau u - g) \in L^2(\Omega) \times H^{3/2}(\partial\Omega),$$

where  $F$  is the HJBI operator defined as in (2.2a), and  $\tau : H^2(\Omega) \rightarrow H^{3/2}(\partial\Omega)$  is the trace operator. Moreover, one can directly check that given an iterate  $u^k \in H^2(\Omega)$ ,  $k \geq 0$ , the next iterate  $u^{k+1}$  solves the following Dirichlet problem:

$$L_k(u^{k+1} - u^k) = -F(u^k), \quad \text{in } \Omega; \quad \tau(u^{k+1} - u^k) = -(\tau u^k - g), \quad \text{on } \partial\Omega.$$

with the differential operator  $L_k \in \partial^* F(u^k)$  defined as in (3.16). Since  $F : H^2(\Omega) \rightarrow L^2(\Omega)$  is  $\partial^* F$ -semismooth (see Corollary 3.5) and  $\tau \in \mathcal{L}(H^2(\Omega), H^{3/2}(\partial\Omega))$ , we can conclude that Algorithm 1 is in fact a semismooth Newton method for solving the operator equation  $\tilde{F}(u) = 0$ .

Note that the boundedness of coefficients and the classical theory of elliptic regularity (see Theorem A.1) imply that under condition (H.1), there exists a constant  $C > 0$ , such that for any  $u \in H^2(\Omega)$  and any  $L \in \partial^* F(u)$ , the inverse operator  $(L, \tau)^{-1} : L^2(\Omega) \times H^{3/2}(\partial\Omega) \rightarrow H^2(\Omega)$  is well defined, and the operator norm  $\|(L, \tau)^{-1}\|$  is bounded by  $C$ , uniformly in  $u \in H^2(\Omega)$ . Hence, one can conclude from [46, Theorem 3.13] (see also Theorem A.5) that the iterates  $\{u^k\}_{k \in \mathbb{N}}$  converges superlinearly to  $u^*$  in a neighborhood  $\mathcal{N}$  of  $u^*$ .  $\square$

The next theorem strengthens Theorem 3.6 and establishes a novel global convergence result of Algorithm 1 applied to the Dirichlet problem (2.2), which subsequently provides a constructive proof for the existence of solutions to (2.2). The following additional condition is essential for our proof of the global convergence of Algorithm 1:

**H.2** Let the function  $c$  in (H.1) be given as:  $c(x, \alpha, \beta) = \bar{c}(x, \alpha, \beta) + \underline{c}_0$ , for all  $(x, \alpha, \beta) \in \Omega \times A \times B$ , where  $\underline{c}_0$  is a sufficiently large constant, depending on  $\Omega$ ,  $\{a^{ij}\}_{i,j=1}^n$ ,  $\{b^i\}_{i=1}^n$  and  $\|\bar{c}\|_{L^\infty(\Omega \times A \times B)}$ .

In practice, (H.2) can be satisfied if (2.2) arises from an infinite-horizon stochastic game with a large discount factor (see e.g. [10]), or if (2.2) stems from an implicit (time-)discretization of parabolic HJBI equations with a small time stepsize.

**Theorem 3.7** Suppose (H.1) and (H.2) hold, then the Dirichlet problem (2.2) admits a unique strong solution  $u^* \in H^2(\Omega)$ . Moreover, for any initial guess  $u^0 \in H^2(\Omega)$ , Algorithm 1 either terminates with  $u^k = u^*$  for some  $k \in \mathbb{N}$ , or generates a sequence  $\{u^k\}_{k \in \mathbb{N}}$  that converges  $q$ -superlinearly to  $u^*$  in  $H^2(\Omega)$ , i.e.,  $\lim_{k \rightarrow \infty} \|u^{k+1} - u^*\|_{H^2(\Omega)} / \|u^k - u^*\|_{H^2(\Omega)} = 0$ .

**Proof** If Algorithm 1 terminates in iteration  $k$ , we have  $F(u^k) = L_k u^k - f_k = 0$  and  $\tau u^k = g$ , from which we obtain from the uniqueness of strong solutions to (2.2) (Proposition 2.1) that  $u^k = u^*$  is the strong solution to the Dirichlet problem (2.2). Hence, we shall assume without loss of generality that Algorithm 1 runs infinitely.

We now establish the global convergence of Algorithm 1 by first showing the iterates  $\{u^k\}_{k \in \mathbb{N}}$  form a Cauchy sequence in  $H^2(\Omega)$ . For each  $k \geq 0$ , we deduce from (3.6) and (H.2) that  $\tau u^{k+1} = g$  on  $\partial\Omega$  and



$$\begin{aligned}
 & -a^{ij}\partial_{ij}u^{k+1} + b_k^i\partial_i u^{k+1} + c_k u^{k+1} - f_k \\
 & = -a^{ij}\partial_{ij}u^{k+1} + b_k^i\partial_i u^{k+1} + (\bar{c}_k + \underline{c}_0)u^{k+1} - f_k \\
 & = -a^{ij}\partial_{ij}u^{k+1} + \underline{c}_0 u^{k+1} + b_k^i\partial_i(u^{k+1} - u^k) \\
 & \quad + \bar{c}_k(u^{k+1} - u^k) + \bar{G}(\cdot, u^k, \nabla u^k) = 0,
 \end{aligned} \tag{3.18}$$

for a.e.  $x \in \Omega$ , where the function  $\bar{c}_k(x) := \bar{c}(x, \alpha^k(x), \beta^k(x))$  for all  $x \in \Omega$ , and the modified Hamiltonian is defined as:

$$\bar{G}(x, u, \nabla u) = \max_{\alpha \in \mathbf{A}} \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta)\partial_i u(x) + \bar{c}(x, \alpha, \beta)u(x) - f(x, \alpha, \beta)). \tag{3.19}$$

Hence, by taking the difference of equations corresponding to the indices  $k - 1$  and  $k$ , one can obtain that

$$\begin{aligned}
 & -a^{ij}\partial_{ij}(u^{k+1} - u^k) + \underline{c}_0(u^{k+1} - u^k) = -b_k^i\partial_i(u^{k+1} - u^k) - \bar{c}_k(u^{k+1} - u^k) \\
 & \quad + b_{k-1}^i\partial_i(u^k - u^{k-1}) + \bar{c}_{k-1}(u^k - u^{k-1}) - [\bar{G}(\cdot, u^k, \nabla u^k) - \bar{G}(\cdot, u^{k-1}, \nabla u^{k-1})],
 \end{aligned} \tag{3.20}$$

for  $x \in \Omega$ , and  $\tau(u^{k+1} - u^k) = 0$  on  $\partial\Omega$ .

It has been proved in Theorem 9.14 of [20] that there exist positive constants  $C$  and  $\gamma_0$ , depending only on  $\{a^{ij}\}_{i,j=1}^n$  and  $\Omega$ , such that it holds for all  $u \in H^2(\Omega)$  with  $\tau u = 0$ , and for all  $\gamma \geq \gamma_0$  that

$$\|u\|_{H^2(\Omega)} \leq C\| -a^{ij}\partial_{ij}u + \gamma u \|_{L^2(\Omega)},$$

which, together with the identity that  $\gamma u = (-a^{ij}\partial_{ij}u + \gamma u) + a^{ij}\partial_{ij}u$  and the boundedness of  $\{a^{ij}\}_{i,j}$ , implies that the same estimate also holds for  $\|u\|_{H^2(\Omega)} + \gamma\|u\|_{L^2(\Omega)}$ :

$$\|u\|_{H^2(\Omega)} + \gamma\|u\|_{L^2(\Omega)} \leq C\| -a^{ij}\partial_{ij}u + \gamma u \|_{L^2(\Omega)}.$$

Thus, by assuming  $\underline{c}_0 \geq \gamma_0$  and using the boundedness of the coefficients, we can deduce from (3.20) that

$$\begin{aligned}
 & \|u^{k+1} - u^k\|_{H^2(\Omega)} + \underline{c}_0\|u^{k+1} - u^k\|_{L^2(\Omega)} \leq C \left( \| -b_k^i\partial_i(u^{k+1} - u^k) - \bar{c}_k(u^{k+1} - u^k) \right. \\
 & \quad \left. + b_{k-1}^i\partial_i(u^k - u^{k-1}) + \bar{c}_{k-1}(u^k - u^{k-1}) - [\bar{G}(\cdot, u^k, \nabla u^k) - \bar{G}(\cdot, u^{k-1}, \nabla u^{k-1})] \|_{L^2(\Omega)} \right) \\
 & \leq C(\|u^{k+1} - u^k\|_{H^1(\Omega)} + \|u^k - u^{k-1}\|_{H^1(\Omega)}),
 \end{aligned} \tag{3.21}$$

for some constant  $C$  independent of  $\underline{c}_0$  and the index  $k$ .

Now we apply the following interpolation inequality (see [20, Theorem 7.28]): there exists a constant  $C$ , such that for all  $u \in H^2(\Omega)$  and  $\varepsilon > 0$ , we have  $\|u\|_{H^1(\Omega)} \leq \varepsilon\|u\|_{H^2(\Omega)} + C\varepsilon^{-1}\|u\|_{L^2(\Omega)}$ . Hence, for any given  $\varepsilon_1 \in (0, 1)$ ,  $\varepsilon_2 > 0$ , we have

$$\begin{aligned}
& (1 - \varepsilon_1) \|u^{k+1} - u^k\|_{H^2(\Omega)} + \underline{c}_0 \|u^{k+1} - u^k\|_{L^2(\Omega)} \\
& \leq \varepsilon_2 \|u^k - u^{k-1}\|_{H^2(\Omega)} + C\varepsilon_1^{-1} \|u^{k+1} - u^k\|_{L^2(\Omega)} + C\varepsilon_2^{-1} \|u^k - u^{k-1}\|_{L^2(\Omega)}.
\end{aligned}$$

Then, by taking  $\varepsilon_1 \in (0, 1)$ ,  $\varepsilon_2 < 1 - \varepsilon_1$ , and assuming that  $\underline{c}_0$  satisfies  $(\underline{c}_0 - C/\varepsilon_1)/(1 - \varepsilon_1) \geq C/\varepsilon_2^2$ , we can obtain for  $c' = C/\varepsilon_2^2$  that

$$\begin{aligned}
& \|u^{k+1} - u^k\|_{H^2(\Omega)} + c' \|u^{k+1} - u^k\|_{L^2(\Omega)} \\
& \leq \frac{\varepsilon_2}{1 - \varepsilon_1} (\|u^k - u^{k-1}\|_{H^2(\Omega)} + c' \|u^k - u^{k-1}\|_{L^2(\Omega)}),
\end{aligned}$$

which implies that  $\{u^k\}_{k \in \mathbb{N}}$  is a Cauchy sequence with the norm  $\|\cdot\|_{c'} := \|\cdot\|_{H^2(\Omega)} + c' \|\cdot\|_{L^2(\Omega)}$ .

Since  $\|\cdot\|_{c'}$  is equivalent to  $\|\cdot\|_{H^2(\Omega)}$  on  $H^2(\Omega)$ , we can deduce that  $\{u^k\}_{k \in \mathbb{N}}$  converges to some  $\bar{u}$  in  $H^2(\Omega)$ . By passing  $k \rightarrow \infty$  in (3.18) and using Proposition 2.1, we can deduce that  $\bar{u} = u^*$  is the unique strong solution of (2.2). Finally, for a sufficiently large  $K_0 \in \mathbb{N}$ , we can conclude the superlinear convergence of  $\{u^k\}_{k \geq K_0}$  from Theorem 3.6.  $\square$

We end this section with an important remark that if one of the sets **A** and **B** is a singleton, and  $a^{ij} \in C^{0,1}(\bar{\Omega})$  for all  $i, j$ , then Algorithm 1 applied to the Dirichlet problem (2.2) is in fact monotonically convergent with an arbitrary initial guess. Suppose, for instance, that **A** is a singleton, then for each  $k \in \mathbb{N} \cup \{0\}$ , we have that

$$0 = L_k u^{k+1} - f_k \geq F(u^{k+1}) = -L_{k+1}(u^{k+2} - u^{k+1}), \quad \text{for a.e. } x \in \Omega.$$

Hence, we can deduce that  $w^{k+1} := u^{k+1} - u^{k+2}$  is a weak subsolution to  $L_{k+1}w = 0$ , i.e., it holds for all  $\phi \in C_0^1(\Omega)$  with  $\phi \geq 0$  that

$$\int_{\Omega} \left[ a^{ij} \partial_j w^{k+1} \partial_i \phi + ((\partial_i a^{ij} + b_{k+1}^i) \partial_i w^{k+1} + c_{k+1} w^{k+1}) \phi \right] dx \leq 0,$$

Thus, the weak maximal principle (see [19, Theorem 1.3.7]) and the fact that  $u^{k+1} - u^{k+2} = 0$  a.e.  $x \in \partial\Omega$  (with respect to the surface measure) leads to the estimate  $\text{ess sup}_{\Omega} u^{k+1} - u^{k+2} \leq 0$ , which consequently implies that  $u^k \leq u^{k+1}$  for all  $k \in \mathbb{N}$  and a.e.  $x \in \Omega$ .

## 4 Inexact Policy Iteration for HJBI Dirichlet Problems

Note that at each policy iteration, Algorithm 1 requires us to obtain an exact solution to a linear Dirichlet boundary value problem, which is generally infeasible. Moreover, an accurate computation of numerical solutions to linear Dirichlet boundary value problems could be expensive, especially in a high-dimensional setting. In this section, we shall propose an inexact policy iteration algorithm for (2.2), where we compute

an approximate solution to (3.3) by solving an optimization problem over a family of trial functions, while maintaining the superlinear convergence of policy iteration.

We shall make the following assumption on the trial functions of the optimization problem.

**H.3** *The collections of trial functions  $\{\mathcal{F}_M\}_{M \in \mathbb{N}}$  satisfy the following properties:  $\mathcal{F}_M \subset \mathcal{F}_{M+1}$  for all  $M \in \mathbb{N}$ , and  $\mathcal{F} = \{\mathcal{F}_M\}_{M \in \mathbb{N}}$  is dense in  $H^2(\Omega)$ .*

It is clear that (H.3) is satisfied by any reasonable  $H^2$ -conforming finite element spaces (see e.g. [9]) and high-order polynomial spaces or kernel-function spaces used in global spectral methods (see e.g. [4,5,13,28,29]). We now demonstrate that (H.3) can also be easily satisfied by the sets of multilayer feedforward neural networks, which provides effective trial functions for high-dimensional problems. Let us first recall the definition of a feedforward neural network.

**Definition 4.1** (*Artificial neural networks*) Let  $L, N_0, N_1, \dots, N_L \in \mathbb{N}$  be given constants, and  $\varrho : \mathbb{R} \rightarrow \mathbb{R}$  be a given function. For each  $l = 1, \dots, L$ , let  $T_l : \mathbb{R}^{N_{l-1}} \rightarrow \mathbb{R}^{N_l}$  be an affine function given as  $T_l(x) = W_l x + b_l$  for some  $W_l \in \mathbb{R}^{N_l \times N_{l-1}}$  and  $b_l \in \mathbb{R}^{N_l}$ . A function  $F : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_L}$  defined as

$$F(x) = T_L \circ (\varrho \circ T_{L-1}) \circ \dots \circ (\varrho \circ T_1), \quad x \in \mathbb{R}^{N_0},$$

is called a feedforward neural network. Here, the activation function  $\varrho$  is applied componentwise. We shall refer the quantity  $L$  as the depth of  $F$ ,  $N_1, \dots, N_{L-1}$  as the dimensions of the hidden layers, and  $N_0, N_L$  as the dimensions of the input and output layers, respectively. We also refer to the number of entries of  $\{W_l, b_l\}_{l=1}^L$  as the complexity of  $F$ .

Let  $\{L^{(M)}\}_{M \in \mathbb{N}}, \{N_1^{(M)}\}_{M \in \mathbb{N}}, \dots, \{N_{L^{(M)}-1}^{(M)}\}_{M \in \mathbb{N}}$  be some nondecreasing sequences of natural numbers, we define for each  $M$  the set  $\mathcal{F}_M$  of all neural networks with depth  $L^{(M)}$ , input dimension being equal to  $n$ , output dimension being equal to 1, and dimensions of hidden layers being equal to  $\{N_1^{(M)}, \dots, N_{L^{(M)}-1}^{(M)}\}_{M \in \mathbb{N}}$ . It is clear that if  $L^{(M)} \equiv L$  for all  $M \in \mathbb{N}$ , then we have  $\mathcal{F}_M \subset \mathcal{F}_{M+1}$ . The following proposition is proved in [25, Corollary 3.8], which shows neural networks with one hidden layer are dense in  $H^2(\Omega)$ .

**Proposition 4.1** *Let  $\Omega \subset \mathbb{R}^n$  be an open bounded star-shaped domain, and  $\varrho \in C^2(\mathbb{R})$  satisfying  $0 < |D^l \varrho|_{L^1(\Omega)} < \infty$  for all  $l = 1, 2$ . Then, the family of all neural networks with depth  $L = 2$  is dense in  $H^2(\Omega)$ .*

Now we discuss how to approximate the strong solutions of Dirichlet problems by reformulating the equations into optimization problems over trial functions. The idea is similar to least squares finite-element methods (see e.g. [7]) and has been employed previously to develop numerical methods for PDEs based on neural networks (see e.g. [6,35,44]). However, compared to [6,35], we do not impose additional constraints on the trial functions by requiring that the networks exactly agree with the boundary conditions, due to the lack of theoretical support that the constrained neural networks

are still dense in the solution space. Moreover, to ensure the convergence of solutions in the  $H^2(\Omega)$ -norm, we include the  $H^{3/2}(\partial\Omega)$ -norm of the boundary data in the cost function, instead of the  $L^2(\partial\Omega)$ -norm used in [44] (see Remark 4.2 for more details).

For each  $k \in \mathbb{N} \cup \{0\}$ , let  $u^{k+1} \in H^2(\Omega)$  be the unique solution to the Dirichlet problem (3.3):

$$L_k u - f_k = 0, \text{ in } \Omega; \quad \tau u = g, \text{ on } \partial\Omega,$$

where  $L_k$  and  $f_k$  denote the linear elliptic operator and the source term in (3.3), respectively. For each  $M \in \mathbb{N}$ , we shall consider the following optimization problems:

$$J_{k,M} := \inf_{u \in \mathcal{F}_M} J_k(u), \quad \text{with } J_k(u) = \|L_k u - f_k\|_{L^2(\Omega)}^2 + \|\tau u - g\|_{H^{3/2}(\partial\Omega)}^2. \quad (4.1)$$

The following result shows that the cost function  $J_k$  provides a computable indicator of the error.

**Proposition 4.2** *Suppose (H.1) and (H.3) hold. For each  $k \in \mathbb{N} \cup \{0\}$  and  $M \in \mathbb{N}$ , let  $u^{k+1} \in H^2(\Omega)$  be the unique solution to (3.3), and  $J_k, J_{k,M}$  be defined as in (4.1). Then, there exist positive constants  $C_1$  and  $C_2$ , such that we have for each  $u \in H^2(\Omega)$  and  $k \in \mathbb{N} \cup \{0\}$  that*

$$C_1 J_k(u) \leq \|u - u^{k+1}\|_{H^2(\Omega)}^2 \leq C_2 J_k(u).$$

Consequently, it holds for each  $k \in \mathbb{N} \cup \{0\}$  that  $\lim_{M \rightarrow \infty} J_{k,M} = 0$ .

**Proof** Let  $k \in \mathbb{N} \cup \{0\}$  and  $u \in H^2(\Omega)$ . The definition of  $J_k(u)$  implies that  $L_k u - f_k = f^e \in L^2(\Omega)$ ,  $\tau u - g = g^e \in H^{3/2}(\partial\Omega)$  and  $J(u) = \|f^e\|_{L^2(\Omega)}^2 + \|g^e\|_{H^{3/2}(\partial\Omega)}^2$ . Then, by using the assumption that  $u^{k+1}$  solves (3.3), we deduce that the residual term satisfies the following Dirichlet problem:

$$L_k(u - u^{k+1}) = f^e, \text{ in } \Omega; \quad \tau(u - u^{k+1}) = g^e, \text{ on } \partial\Omega.$$

Hence, the boundedness of coefficients and the regularity theory of elliptic operators (see Theorem A.1) lead to the estimate that

$$\begin{aligned} C_1 (\|f^e\|_{L^2(\Omega)}^2 + \|g^e\|_{H^{3/2}(\partial\Omega)}^2) &\leq \|u - u^{k+1}\|_{H^2(\Omega)}^2 \\ &\leq C_2 (\|f^e\|_{L^2(\Omega)}^2 + \|g^e\|_{H^{3/2}(\partial\Omega)}^2), \end{aligned}$$

where the constants  $C_1, C_2 > 0$  depend only on the  $L^\infty(\Omega)$ -norms of  $a^{ij}, b_k^i, c_k, f_k$ , which are independent of  $k$ . The above estimate, together with the facts that  $\{\mathcal{F}_M\}_{M \in \mathbb{N}}$  is dense in  $H^2(\Omega)$  and  $\mathcal{F}_M \subset \mathcal{F}_{M+1}$ , leads us to the desired conclusion that  $\lim_{M \rightarrow \infty} J_{k,M} = 0$ .  $\square$

We now present the inexact policy iteration algorithm for the HJBI problem (2.2), where at each policy iteration, we solve the linear Dirichlet problem within a given accuracy. <sup>1</sup>

**Algorithm 2** Inexact policy iteration algorithm for Dirichlet problems

1. Choose a family of trial functions  $\mathcal{F} = \{\mathcal{F}_M\}_{M \in \mathbb{N}} \subset H^2(\Omega)$ , an initial guess  $u^0$  in  $\mathcal{F}$ , a sequence  $\{\eta_k\}_{k \in \mathbb{N} \cup \{0\}}$  of positive scalars, and set  $k = 0$ .
2. Given the iterate  $u^k$ , update the control laws  $\alpha^k$  and  $\beta^k$  by (3.1) and (3.2), respectively.
3. Find  $u^{k+1} \in \mathcal{F}$  such that<sup>1</sup>

$$J_k(u^{k+1}) = \|L_k u^{k+1} - f_k\|_{L^2(\Omega)}^2 + \|\tau u^{k+1} - g\|_{H^{3/2}(\partial\Omega)}^2 \leq \eta_{k+1} \min(\|u^{k+1} - u^k\|_{H^2(\Omega)}^2, \eta_0), \quad (4.2)$$

where  $L_k$  and  $f_k$  denote the linear operator and the source term in (3.3), respectively.

4. If  $\|u^{k+1} - u^k\|_{H^2(\Omega)} = 0$ , then terminate with outputs  $u^{k+1}$ ,  $\alpha^k$  and  $\beta^k$ , otherwise increment  $k$  by one and go to step 2.

**Remark 4.1** In practice, the evaluation of the squared residuals  $J_k$  in (4.2) depends on the choice of trial functions. For trial functions with linear architecture, e.g. if  $\{\mathcal{F}_M\}_{M \in \mathbb{N}}$  are finite element spaces, high-order polynomial spaces, and kernel-function spaces (see [9,13,28,29]), one may evaluate the norms by applying high-order quadrature rules to the basis functions involved.

For trial functions with nonlinear architecture, such as feedforward neural networks, we can replace the integrations in  $J_k$  by the empirical mean over suitable collocation points in  $\Omega$  and on  $\partial\Omega$ , such as pseudorandom points or quasi-Monte Carlo points (see Sect. 6; see also [6,35,44]). In particular, due to the existence of local coordinate charts of the boundaries, we can evaluate the double integral in the definition of the  $H^{3/2}(\partial\Omega)$ -norm (see (2.1)) by first generating points in  $\mathbb{R}^{2(n-1)}$  and then mapping the samples onto  $\partial\Omega \times \partial\Omega$ . The resulting empirical least-squares problem for the  $k+1$ -th policy iteration step (cf. (4.1)) can then be solved by stochastic gradient descent (SGD) algorithms; see Sect. 6. We remark that instead of pre-generating all the collocation points in advance, one can perform gradient descent based on a sequence of mini-batches of points generated at each SGD iteration. This is particularly useful in higher dimensions, where many collocation points may be needed to cover the boundary, and using mini-batches avoids having to evaluate functions at all collocation points in each iteration.

It is well known (see e.g. [14,46]) that the residual term  $\|u^{k+1} - u^k\|_{H^2(\Omega)}$  is crucial for the superlinear convergence of inexact Newton methods. This next theorem establishes the global superlinear convergence of Algorithm 2.

**Theorem 4.3** Suppose (H.1), (H.2) and (H.3) hold, and  $\lim_{k \rightarrow \infty} \eta_k = 0$  in Algorithm 2. Let  $u^* \in H^2(\Omega)$  be the solution to the Dirichlet problem (2.2). Then, for any initial guess  $u^0 \in \mathcal{F}$ , Algorithm 2 either terminates with  $u^k = u^*$  for some  $k \in \mathbb{N}$ , or generates a sequence  $\{u^k\}_{k \in \mathbb{N}}$  that converges  $q$ -superlinearly to  $u^*$  in  $H^2(\Omega)$ , i.e.,  $\lim_{k \rightarrow \infty} \|u^{k+1} - u^*\|_{H^2(\Omega)} / \|u^k - u^*\|_{H^2(\Omega)} = 0$ . Consequently, we have  $\lim_{k \rightarrow \infty} (u^k, \partial_i u^k, \partial_{ij} u^k)(x) = (u^*, \partial_i u^*, \partial_{ij} u^*)(x)$  for a.e.  $x \in \Omega$ , and for all  $i, j = 1, \dots, n$ .

<sup>1</sup> With a slight abuse of notation, we denote by  $u^{k+1}$  the inexact solution to the Dirichlet problem (3.3).

**Proof** Let  $u^0 \in \mathcal{F}$  be an arbitrary initial guess. We first show that Algorithm 2 is always well defined. For each  $k \in \mathbb{N} \cup \{0\}$ , if  $u^k \in \mathcal{F}$  is the strong solution to (3.3), then we can choose  $u^{k+1} = u^k$ , which satisfies (4.2) and terminates the algorithm. If  $u^k$  does not solve (3.3), the fact that  $\mathcal{F}$  is dense in  $H^2(\Omega)$  enables us to find  $u^{k+1} \in \mathcal{F}$  satisfying the criterion (4.2).

Moreover, one can clearly see from (4.2) that if Algorithm 2 terminates at iteration  $k$ , then  $u^k$  is the exact solution to the Dirichlet problem (2.2). Hence in the sequel we shall assume without loss of generality that Algorithm 2 runs infinitely, i.e.,  $\|u^{k+1} - u^k\|_{H^2(\Omega)} > 0$  and  $u^k \neq u^*$  for all  $k \in \mathbb{N} \cup \{0\}$ .

We next show the iterates converge to  $u^*$  in  $H^2(\Omega)$  by following similar arguments as those for Theorem 3.7. For each  $k \geq 0$ , we can deduce from (4.2) that there exists  $f_k^e \in L^2(\Omega)$  and  $g_k^e \in H^{3/2}(\partial\Omega)$  such that

$$L_k u^{k+1} - f_k = f_k^e, \quad \text{in } \Omega; \quad \tau u^{k+1} - g = g_k^e, \quad \text{on } \partial\Omega, \quad (4.3)$$

and  $J_k(u^{k+1}) = \|f_k^e\|_{L^2(\Omega)}^2 + \|g_k^e\|_{H^{3/2}(\partial\Omega)}^2 \leq \eta_{k+1}(\|u^{k+1} - u^k\|_{H^2(\Omega)}^2)$  with  $\lim_{k \rightarrow \infty} \eta_k = 0$ . Then, by taking the difference between (4.3) and (2.2), we obtain that

$$\begin{aligned} -a^{ij} \partial_{ij}(u^{k+1} - u^*) + \underline{c}_0(u^{k+1} - u^*) &= -b_k^i \partial_i(u^{k+1} - u^k) - \bar{c}_k(u^{k+1} - u^k) \\ &\quad - [\bar{G}(\cdot, u^k, \nabla u^k) - \bar{G}(\cdot, u^*, \nabla u^*)] + f_k^e, \quad \text{in } \Omega, \end{aligned}$$

and  $\tau(u^{k+1} - u^*) = g_k^e$  on  $\partial\Omega$ , where  $\bar{G}$  is the modified Hamiltonian defined as in (3.19). Then, by proceeding along the lines of Theorem 3.7, we can obtain a positive constant  $C$ , independent of  $\underline{c}_0$  and the index  $k$ , such that

$$\begin{aligned} &\|u^{k+1} - u^*\|_{H^2(\Omega)} + \underline{c}_0 \|u^{k+1} - u^*\|_{L^2(\Omega)} \\ &\leq C(\|u^{k+1} - u^*\|_{H^1(\Omega)} + \|u^{k+1} - u^k\|_{H^1(\Omega)} + \|u^k - u^*\|_{H^1(\Omega)}) \\ &\quad + o(\|u^{k+1} - u^k\|_{H^2(\Omega)}) \\ &\leq C(\|u^{k+1} - u^*\|_{H^1(\Omega)} + \|u^k - u^*\|_{H^1(\Omega)}) \\ &\quad + o(\|u^{k+1} - u^*\|_{H^2(\Omega)} + \|u^k - u^*\|_{H^2(\Omega)}) \end{aligned}$$

as  $k \rightarrow \infty$ , where the additional high-order terms are due to the residuals  $f_k^e$  and  $g_k^e$ . Then, by using the interpolation inequality and assuming  $\underline{c}_0$  is sufficiently large, we can deduce that  $\{u^k\}_{k \in \mathbb{N}}$  converge linearly to  $u^*$  in  $H^2(\Omega)$ .

We then reformulate Algorithm 2 into a quasi-Newton method for the operator equation  $\tilde{F}(u) = 0$ , with the operator  $\tilde{F} : u \in H^2(\Omega) \rightarrow (F(u), \tau u - g) \in L^2(\Omega) \times H^{3/2}(\partial\Omega)$  defined in the proof of Theorem 3.6. Let  $H^2(\Omega)^*$  denote the strong dual space of  $H^2(\Omega)$ , and  $\langle \cdot, \cdot \rangle$  denote the dual product on  $H^2(\Omega)^* \times H^2(\Omega)$ . For each  $k \in \mathbb{N} \cup \{0\}$ , by using the fact that  $\|u^{k+1} - u^k\|_{H^2(\Omega)} > 0$ , we can choose  $w_k \in H^2(\Omega)^*$  satisfying  $\langle w_k, u^{k+1} - u^k \rangle = -1$ , and introduce the following linear operators  $\delta L_k \in \mathcal{L}(H^2(\Omega), L^2(\Omega))$  and  $\delta \tau_k \in \mathcal{L}(H^2(\Omega), H^{3/2}(\partial\Omega))$ :

$$\begin{aligned}\delta L_k : v \in H^2(\Omega) &\mapsto \langle w_k, v \rangle f_k^e \in L^2(\Omega), \\ \delta \tau_k : v \in H^2(\Omega) &\mapsto \langle w_k, v \rangle g_k^e \in H^{3/2}(\partial\Omega).\end{aligned}$$

Then, we can apply the identity  $F(u^k) = L_k u^k - f_k$  and rewrite (4.3) as:

$$\begin{aligned}(L_k + \delta L_k)(u^{k+1} - u^k) &= -F(u^k), \quad \text{in } \Omega; \\ (\tau + \delta \tau_k)(u^{k+1} - u^k) &= -(\tau u^k - g), \quad \text{on } \partial\Omega,\end{aligned}$$

with  $(L_k, \tau) \in \partial^* \tilde{F}(u^k)$  as shown in Theorem 3.6. Hence, one can clearly see that (4.3) is precisely a Newton step with a perturbed operator for the equation  $\tilde{F}(u) = 0$ .

Now we are ready to establish the superlinear convergence of  $\{u^k\}_{k \in \mathbb{N}}$ . For notational simplicity, in the subsequent analysis we shall denote by  $Z := L^2(\Omega) \times H^{3/2}(\partial\Omega)$  the Banach space with the usual product norm  $\|z\|_Z := \|z_1\|_{L^2(\Omega)} + \|z_2\|_{H^{3/2}(\partial\Omega)}$  for each  $z = (z_1, z_2) \in Z$ . By using the semismoothness of  $\tilde{F} : H^2(\Omega) \rightarrow Z$  (see Theorem 3.6) and the strong convergence of  $\{u^k\}_{k \in \mathbb{N}}$  in  $H^2(\Omega)$ , we can directly infer from Theorem A.5 that it remains to show that there exists a open neighborhood  $V$  of  $u^*$ , and a constant  $L > 0$ , such that

$$\|v - u^*\|_{H^2(\Omega)} / L \leq \|\tilde{F}(v) - \tilde{F}(u^*)\|_Z \leq L \|v - u^*\|_{H^2(\Omega)}, \quad \forall v \in V, \quad (4.4)$$

and also

$$\lim_{k \rightarrow \infty} \|(\delta L_k s^k, \delta \tau_k s^k)\|_Z / \|s^k\|_{H^2(\Omega)} = 0, \quad \text{with } s^k = u^{k+1} - u^k \text{ for all } k \in \mathbb{N}. \quad (4.5)$$

The criterion (4.2) and the definitions of  $\delta L_k$  and  $\delta \tau_k$  imply that (4.5) holds:

$$\begin{aligned}\left( \frac{\|(\delta L_k s^k, \delta \tau_k s^k)\|_Z}{\|s^k\|_{H^2(\Omega)}} \right)^2 &= \left( \frac{\|f_k^e\|_{L^2(\Omega)} + \|g_k^e\|_{H^{3/2}(\partial\Omega)}}{\|s^k\|_{H^2(\Omega)}} \right)^2 \\ &\leq \frac{2J_k(u^{k+1})}{\|s^k\|_{H^2(\Omega)}^2} \leq 2\eta_0 \eta_{k+1} \rightarrow 0,\end{aligned}$$

as  $k \rightarrow \infty$ . Moreover, the boundedness of the coefficients  $a^{ij}, b^i, c, f$  shows that  $\tilde{F}$  is Lipschitz continuous. Finally, the characterization of the generalized differential of  $\tilde{F}$  in Theorem 3.6 and the regularity theory of elliptic operators (see Theorem A.1) show that for each  $v \in H^2(\Omega)$ , we can choose an invertible operator  $M_v = (L_v, \tau) \in \partial^* \tilde{F}(v)$  such that  $\|M_v^{-1}\|_{\mathcal{L}(Z, H^2(\Omega))} \leq C < \infty$ , uniformly in  $v$ . Thus, we can conclude from the semismoothness of  $\tilde{F}$  at  $u^*$  that

$$\begin{aligned}\|\tilde{F}(v) - \tilde{F}(u^*)\|_Z &= \|M_v(v - u^*) + o(\|v - u^*\|_{H^2(\Omega)})\|_Z \\ &\geq \|M_v(v - u^*)\|_Z - o(\|v - u^*\|_{H^2(\Omega)}) \\ &\geq \|v - u^*\|_{H^2(\Omega)} / C - o(\|v - u^*\|_{H^2(\Omega)}) \geq \|v - u^*\|_{H^2(\Omega)} / (2C),\end{aligned}$$

for all  $v$  in some neighborhood  $V$  of  $u^*$ , which completes our proof for  $q$ -superlinear convergence of  $\{u^k\}_{k \in \mathbb{N}}$ .

Finally, we establish the pointwise convergence of  $\{u^k\}_{k=1}^\infty$  and their derivatives. For any given  $\gamma \in (0, 1)$ , the superlinear convergence of  $\{u^k\}_{k=1}^\infty$  implies that there exists a constant  $C > 0$ , depending on  $\gamma$ , such that  $\|u^k - u^*\|_{H^2(\Omega)}^2 \leq C\gamma^{2k}$  for all  $k \in \mathbb{N}$ . Taking the summation over the index  $k$ , we have

$$\begin{aligned} & \int_{\Omega} \sum_{k=1}^{\infty} \left( |u^k - u^*|^2 + \sum_{i,j=1}^n [|\partial_i u^k - \partial_i u^*|^2 + |\partial_{ij} u^k - \partial_{ij} u^*|^2] \right) dx \\ &= \sum_{k=1}^{\infty} \|u^k - u^*\|_{H^2(\Omega)}^2 \leq \frac{C\gamma^2}{1 - \gamma^2} < \infty, \end{aligned}$$

where we used the monotone convergence theorem in the first equality. Thus, we have

$$\sum_{k=1}^{\infty} \left( |u^k - u^*|^2 + \sum_{i,j=1}^n [|\partial_i u^k - \partial_i u^*|^2 + |\partial_{ij} u^k - \partial_{ij} u^*|^2] \right)(x) < \infty, \quad \text{for a.e. } x \in \Omega,$$

which leads us to the pointwise convergence of  $u^k$  and its partial derivatives with respect to  $k$ .  $\square$

**Remark 4.2** We reiterate that merely including the  $L^2(\partial\Omega)$ -norm of the boundary data in the cost functional (4.2) in general cannot guarantee the convergence of the derivatives of the numerical solutions  $\{u^k\}_{k=1}^\infty$ , which can be seen from the following simple example. Let  $\{g_k\}_{k=1}^\infty \subseteq H^{3/2}(\partial\Omega)$  be a sequence such that  $g_k \rightarrow 0$  in  $L^2(\partial\Omega)$  but not in  $H^{1/2}(\partial\Omega)$ , and for each  $k \in \mathbb{N}$ , let  $h^k \in H^2(\Omega)$  be the strong solution to  $-\Delta h^k = 0$  in  $\Omega$  and  $h^k = g_k$  on  $\partial\Omega$ .

The fact that  $g_k \not\rightarrow 0$  in  $H^{1/2}(\partial\Omega)$  implies that  $h^k \not\rightarrow 0$  in  $H^1(\Omega)$  as  $k \rightarrow \infty$ . We now show  $\lim_{k \rightarrow \infty} h^k = 0$  in  $L^2(\Omega)$ . Let  $w \in H^2(\Omega)$  be the solution to  $-\Delta w = h^k$  in  $\Omega$  and  $w = 0$  on  $\partial\Omega$ , we can deduce from the integration by parts and the *a priori* estimate  $\|w\|_{H^2(\Omega)} \leq C\|h^k\|_{L^2(\Omega)}$  that

$$\begin{aligned} \|h^k\|_{L^2(\Omega)}^2 &= \int_{\Omega} (-\Delta w) h^k dx = \int_{\Omega} w (-\Delta h^k) dx + \int_{\partial\Omega} w \partial_n h^k d\sigma - \int_{\partial\Omega} h^k \partial_n w d\sigma \\ &\leq C\|g_k\|_{L^2(\partial\Omega)} \|w\|_{H^2(\Omega)} \leq C\|g_k\|_{L^2(\partial\Omega)} \|h^k\|_{L^2(\Omega)}, \end{aligned}$$

which shows that  $\|h^k\|_{L^2(\Omega)} \leq C\|g_k\|_{L^2(\partial\Omega)} \rightarrow 0$  as  $k \rightarrow \infty$ . Now let  $\mathcal{F}$  be a given family of trial functions, which is dense in  $H^2(\Omega)$ . One can find  $\{u^k\}_{k=1}^\infty \subseteq \mathcal{F}$  satisfying  $\lim_{k \rightarrow \infty} \|u^k - h^k\|_{H^2(\Omega)} = 0$ , and consequently  $u^k \not\rightarrow 0$  in  $H^1(\Omega)$  as  $k \rightarrow \infty$ . However, we have

$$\begin{aligned} & \| -\Delta u^k \|_{L^2(\Omega)}^2 + \| u^k \|_{L^2(\partial\Omega)}^2 \\ &= \| -\Delta(u^k - h^k) \|_{L^2(\Omega)}^2 + \| u^k - h^k + g_k \|_{L^2(\partial\Omega)}^2 \rightarrow 0, \quad \text{as } k \rightarrow \infty. \end{aligned}$$



Similarly, one can construct functions  $\{u^k\}_{k=1}^\infty \subseteq \mathcal{F}$  such that  $\| - \Delta u^k \|_{L^2(\Omega)}^2 + \| u^k \|_{H^{1/2}(\partial\Omega)}^2 \rightarrow 0$  as  $k \rightarrow \infty$ , but  $\{u^k\}_{k=1}^\infty$  does not converge to 0 in  $H^2(\Omega)$ .

We end this section with a convergent approximation of the optimal control strategies based on the iterates  $\{u^k\}_{k=1}^\infty$  generated by Algorithm 2. For any given  $u \in H^2(\Omega)$ , we denote by  $\mathbf{A}^u(x)$  and  $\mathbf{B}^u(x, \alpha)$  the set of optimal control strategies for all  $\alpha \in \mathbf{A}$  and for a.e.  $x \in \Omega$ , such that

$$\begin{aligned} \mathbf{B}^u(x, \alpha) &= \arg \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta) \partial_i u(x) + c(x, \alpha, \beta) u(x) - f(x, \alpha, \beta)), \\ \mathbf{A}^u(x) &= \arg \max_{\alpha \in \mathbf{A}} \min_{\beta \in \mathbf{B}} (b^i(x, \alpha, \beta) \partial_i u(x) + c(x, \alpha, \beta) u(x) - f(x, \alpha, \beta)). \end{aligned}$$

As an important consequence of the superlinear convergence of Algorithm 2, we now conclude that the feedback control strategies  $\{\alpha^k\}_{k=1}^\infty$  and  $\{\beta^k\}_{k=1}^\infty$  generated by Algorithm 2 are convergent to the optimal control strategies.

**Corollary 4.4** *Suppose the assumptions of Theorem 4.3 hold, and let  $u^* \in H^2(\Omega)$  be the solution to the Dirichlet problem (2.2). Assume further that there exist functions  $\alpha^* : \Omega \rightarrow \mathbf{A}$  and  $\beta^* : \Omega \rightarrow \mathbf{B}$  such that  $\mathbf{A}^{u^*}(x) = \{\alpha^*(x)\}$  and  $\mathbf{B}^{u^*}(x, \alpha^*(x)) = \{\beta^*(x)\}$  for a.e.  $x \in \Omega$ . Then, the measurable functions  $\alpha^k : \Omega \rightarrow \mathbf{A}$  and  $\beta^k : \Omega \rightarrow \mathbf{B}$ ,  $k \in \mathbb{N}$ , generated by Algorithm 2 converge to the optimal feedback control  $(\alpha^*, \beta^*)$  pointwise almost everywhere.*

**Proof** Let  $\ell$  and  $h$  be the Carathéodory functions defined by (3.4) and (3.5), respectively, and we consider the following set-valued mappings:

$$\begin{aligned} \Gamma_1 : (x, \mathbf{u}) \in \Omega \times \mathbb{R}^{n+1} &\rightrightarrows \Gamma_1(x, \mathbf{u}) := \arg \max_{\alpha \in \mathbf{A}} h(x, \mathbf{u}, \alpha), \\ \Gamma_2 : (x, \mathbf{u}, \alpha) \in \Omega \times \mathbb{R}^{n+1} \times \mathbf{A} &\rightrightarrows \Gamma_2(x, \mathbf{u}, \alpha) := \arg \min_{\beta \in \mathbf{B}} \ell(x, \mathbf{u}, \alpha, \beta). \end{aligned} \quad (4.6)$$

Theorem A.4 implies that the set-valued mappings  $\Gamma_1(x, \cdot) : \mathbb{R}^{n+1} \rightrightarrows \mathbf{A}$  and  $\Gamma_2(x, \cdot, \cdot) : \mathbb{R}^{n+1} \times \mathbf{A} \rightrightarrows \mathbf{B}$  are upper hemicontinuous. Then, the result follows directly from the pointwise convergence of  $(u^k, \nabla u^k)_{k=1}^\infty$  in Theorem 4.3, and the fact that  $\mathbf{A}^{u^*}(x) = \{\alpha^*(x)\}$  and  $\mathbf{B}^{u^*}(x, \alpha^*(x)) = \{\beta^*(x)\}$  are singleton for a.e.  $x \in \Omega$ .  $\square$

**Remark 4.3** If we assume in addition that  $\mathbf{A} \subset X_A$  and  $\mathbf{B} \subset Y_B$  for some Banach spaces  $X_A$  and  $Y_B$ , then by using the compactness of  $\mathbf{A}$  and  $\mathbf{B}$  (see (H.1)), we can conclude from the dominated convergence theorem that  $\alpha^k \rightarrow \alpha^*$  in  $L^p(\Omega; X_A)$  and  $\beta^k \rightarrow \beta^*$  in  $L^p(\Omega; Y_B)$ , for any  $p \in [1, \infty)$ .

## 5 Inexact Policy Iteration for HJBI Oblique Derivative Problems

In this section, we extend the algorithms introduced in previous sections to more general boundary value problems. In particular, we shall propose a neural network-based

policy iteration algorithm with global  $H^2$ -superlinear convergence for solving HJBI boundary value problems with oblique derivative boundary conditions. Similar arguments can be adapted to design superlinear convergent schemes for mixed boundary value problems with both Dirichlet and oblique derivative boundary conditions.

We consider the following HJBI oblique derivative problem:

$$F(u) := -a^{ij}(x)\partial_{ij}u + G(x, u, \nabla u) = 0, \quad \text{a.e. } x \in \Omega, \quad (5.1a)$$

$$Bu := \gamma^i \tau(\partial_i u) + \gamma^0 \tau u - g = 0, \quad \text{on } \partial\Omega. \quad (5.1b)$$

where (5.1a) is the HJBI equation given in (2.2a), and (5.1b) is an oblique boundary condition. Note that the boundary condition  $Bu$  on  $\partial\Omega$  involves the traces of  $u$  and its first partial derivatives, which exist almost everywhere on  $\partial\Omega$  (with respect to the surface measure).

The following conditions are imposed on the coefficients of (5.1):

**H.4** Assume  $\Omega$ ,  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $(a^{ij})_{i,j=1}^n$ ,  $(b^i)_{i=1}^n$ ,  $c$ ,  $f$  satisfy the same conditions as those in (H.1). Let  $g \in H^{1/2}(\partial\Omega)$ ,  $\{\gamma^i\}_{i=0}^n \subseteq C^{0,1}(\partial\Omega)$ ,  $\gamma^0 \geq 0$  on  $\partial\Omega$ , and assume there exists a constant  $\mu > 0$ , such that  $c \geq \mu$  on  $\Omega \times \mathbf{A} \times \mathbf{B}$ , and  $\sum_{i=1}^n \gamma^i v_i \geq \mu$  on  $\partial\Omega$ , where  $\{v_i\}_{i=1}^n$  are the components of the unit outer normal vector field on  $\partial\Omega$ .

The next proposition establishes the well-posedness of the oblique derivative problem.

**Proposition 5.1** Suppose (H.4) holds. Then, the oblique derivative problem (2.2) admits a unique strong solution  $u^* \in H^2(\Omega)$ .

**Proof** We shall establish the uniqueness of strong solutions to (5.1) in this proof, and then explicitly construct the solution in Theorem 5.2 with the help of policy iteration; see also Theorem 3.7. Suppose that  $u, v \in H^2(\Omega)$  are two strong solutions to (5.1), then we can see  $w = u - v$  is a strong solution to the following linear oblique derivative problem:

$$-a^{ij}\partial_{ij}w + \tilde{b}^i\partial_iw + \tilde{c}w = 0, \quad \text{a.e. in } \Omega; \quad \gamma^i\tau(\partial_iw) + \gamma^0\tau w = 0, \quad \text{on } \partial\Omega, \quad (5.2)$$

where  $\tilde{b}^i$  is defined as in Proposition 2.1, and

$$\tilde{c}(x) = \begin{cases} \frac{G(x,u,\nabla u) - G(x,v,\nabla v)}{(u-v)(x)}, & \text{on } \{x \in \Omega \mid (u-v)(x) \neq 0\}, \\ \mu, & \text{otherwise.} \end{cases}$$

By following the same arguments as the proof of Proposition 2.1, we can show that  $\tilde{b}^i, \tilde{c} \in L^\infty(\Omega)$ , and  $\tilde{c} \geq \mu > 0$  a.e. in  $\Omega$ , which, along with Theorem A.2, implies that  $w^* = 0$  is the unique strong solution to (5.2), and consequently  $u = v$  in  $H^2(\Omega)$ .  $\square$

Now we present the neural network-based policy iteration algorithm for solving the oblique derivative problem and establish its rate of convergence.

---

**Algorithm 3** Inexact policy iteration algorithm for oblique derivative problems

---

1. Choose a family of trial functions  $\mathcal{F} = \{\mathcal{F}_M\}_{M \in \mathbb{N}} \subset H^2(\Omega)$ , an initial guess  $u^0$  in  $\mathcal{F}$ , a sequence  $\{\eta_k\}_{k \in \mathbb{N} \cup \{0\}}$  of positive scalars, and set  $k = 0$ .
2. Given the iterate  $u^k$ , update the control laws  $\alpha^k$  and  $\beta^k$  by (3.1) and (3.2), respectively.
3. Find  $u^{k+1} \in \mathcal{F}$  such that

$$J_k(u^{k+1}) = \|L_k u^{k+1} - f_k\|_{L^2(\Omega)}^2 + \|B u^{k+1}\|_{H^{1/2}(\partial\Omega)}^2 \leq \eta_{k+1} \min(\|u^{k+1} - u^k\|_{H^2(\Omega)}^2, \eta_0), \quad (5.3)$$

where  $L_k$ ,  $f_k$ , and  $B$  denote the linear operator in (3.3), the source term in (3.3) and the boundary operator in (5.1b), respectively.

4. If  $\|u^{k+1} - u^k\|_{H^2(\Omega)} = 0$ , then terminate with outputs  $u^{k+1}$ ,  $\alpha^k$  and  $\beta^k$ , otherwise increment  $k$  by one and go to step 2.
- 

Note that the  $H^{1/2}(\partial\Omega)$ -norm of the boundary term is included in the cost function  $J_k$ , instead of the  $H^{3/2}(\partial\Omega)$ -norm as in Algorithm 2. It is straightforward to see that Algorithm 3 is well defined under (H.3) and (H.4). In fact, for each  $k \in \mathbb{N} \cup \{0\}$ , given the iterate  $u^k \in \mathcal{F} \subset H^2(\Omega)$ , Corollary 3.5 shows that one can select measurable control laws  $(\alpha^k, \beta^k)$  such that the following linear oblique boundary value problem has measurable coefficients:

$$L_k u - f_k = 0, \text{ in } \Omega; \quad B u = 0, \text{ on } \partial\Omega,$$

and hence admits a unique strong solution  $\bar{u}^k$  in  $H^2(\Omega)$  (see Theorem A.2). If  $u^k = \bar{u}^k$ , then  $u^k$  solve the HJB oblique derivative problem (5.1), and we can select  $u^{k+1} = u^k$  and terminate the algorithm. Otherwise, the facts that  $J_k(u^{k+1}) \leq C \|\bar{u}^k - u^{k+1}\|_{H^2(\Omega)}^2$  and  $\mathcal{F}$  is dense in  $H^2(\Omega)$  allows us to choose  $u^{k+1} \in \mathcal{F}$  sufficiently closed to  $\bar{u}$  such that the criterion (5.3) is satisfied, and proceed to the next iteration.

The following result is analogue to Theorem 4.3 and shows the global superlinear convergence of Algorithm 3 for solving the oblique derivative problem (5.1). The proof follows precisely the lines given in Theorem 4.3; hence, we shall only present the main steps in “Appendix B” for the reader’s convenience. The convergence of feedback control laws can be concluded similarly to Corollary 4.4 and Remark 4.3.

**Theorem 5.2** *Suppose (H.2), (H.3) and (H.4) hold, and  $\lim_{k \rightarrow \infty} \eta_k = 0$  in Algorithm 3. Let  $u^* \in H^2(\Omega)$  be the solution to the oblique derivative problem (5.1). Then, for any initial guess  $u^0 \in \mathcal{F}$ , Algorithm 3 either terminates with  $u^k = u^*$  for some  $k \in \mathbb{N}$ , or generates a sequence  $\{u^k\}_{k \in \mathbb{N}}$  that converges  $q$ -superlinearly to  $u^*$  in  $H^2(\Omega)$ , i.e.,  $\lim_{k \rightarrow \infty} \|u^{k+1} - u^*\|_{H^2(\Omega)} / \|u^k - u^*\|_{H^2(\Omega)} = 0$ . Consequently, we have  $\lim_{k \rightarrow \infty} (u^k, \partial_i u^k, \partial_{ij} u^k)(x) = (u^*, \partial_i u^*, \partial_{ij} u^*)(x)$  for a.e.  $x \in \Omega$ , and for all  $i, j = 1, \dots, n$ .*

## 6 Numerical Experiments: Zermelo’s Navigation Problem

In this section, we illustrate the theoretical findings and demonstrate the effectiveness of the schemes through numerical experiments. We present a two-dimensional

convection-dominated HJBI Dirichlet boundary value problem in an annulus, which is related to stochastic minimum time problems.

In particular, we consider the stochastic Zermelo navigation problem (see e.g. [37]), which is a time-optimal control problem where the objective is to find the optimal trajectories of a ship/aircraft navigating a region of strong winds, modeled by a random vector field. Given a bounded open set  $\Omega \subset \mathbb{R}^n$  and an adaptive control strategy  $\{\alpha_t\}_{t \geq 0}$  taking values in  $\mathbf{A}$ , we assume the dynamics  $X^{x,\alpha}$  of the ship is governed by the following controlled dynamics:

$$dX_t = b(X_t, \alpha_t) dt + \sigma dW_t, \quad t \in [0, \infty); \quad X_0 = x \in \Omega,$$

where the drift coefficient  $b : \Omega \times \mathbf{A} \rightarrow \mathbb{R}^n$  is the sum of the velocity of the wind and the relative velocity of the ship, the nondegenerate diffusion coefficient  $\sigma : \Omega \rightarrow \mathbb{R}^{n \times n}$  describes a random perturbation of the velocity field of the wind, and  $W$  is an  $n$ -dimensional Brownian motion defined on a probability space  $(\tilde{\Omega}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ .

The aim of the controller is to minimize the expected exit time of the region  $\Omega$ , taking model ambiguity into account in the spirit of [41]. More generally, we consider the following value function:

$$\begin{aligned} u(x) &:= \inf_{\alpha \in \mathcal{A}} \mathcal{E} \left[ \int_0^{\tau_{x,\alpha}} f(X_t^{x,\alpha}) dt + g(X_{\tau_{x,\alpha}}^{x,\alpha}) \right] \\ &= \inf_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}} \left[ \int_0^{\tau_{x,\alpha}} f(X_t^{x,\alpha}) dt + g(X_{\tau_{x,\alpha}}^{x,\alpha}) \right] \end{aligned} \quad (6.1)$$

over all admissible choices of  $\alpha \in \mathcal{A}$ , where  $\tau_{x,\alpha} := \inf\{t \geq 0 \mid X_t^{x,\alpha} \notin \Omega\}$  denotes the first exit time of the controlled dynamics  $X^{x,\alpha}$ , the functions  $f$  and  $g$  denote the running cost and the exit cost, respectively, which indicate the desired destinations, and  $\mathcal{M}$  is a family of absolutely continuous probability measures with respect to  $\mathbb{P}$  with density  $M_t = \exp\left(\int_0^t \beta_t dW_t - \frac{1}{2} \int_0^t \beta_t^2 dt\right)$ , where  $\{\beta_t\}_{t \geq 0}$  is a predictable process satisfying  $\|\beta_t\|_\infty = \max_i |\beta_{t,i}| \leq \kappa$  for all  $t$  and a given parameter  $\kappa \geq 0$ . In other words, we would like to minimize a functional of the trajectory up to the exit time under the worst-case scenario, with uncertainty arising from the unknown law of the random perturbation.

By using the dual representation of  $\mathcal{E}[\cdot]$  and the dynamic programming principle (see e.g. [10,41]), we can characterize the value function  $u$  as the unique viscosity solution to an HJBI Dirichlet boundary value problem of the form (2.2). Moreover, under suitable assumptions, one can further show that  $u$  is the strong (Sobolev) solution to this Dirichlet problem (see e.g. [33]).

For our numerical experiments, we assume that the domain  $\Omega$  is an annulus, i.e.,  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid r^2 < x^2 + y^2 < R^2\}$ , the wind blows along the positive  $x$ -axis with a magnitude  $v_c$ :

$$v_c(x, y) = 1 - a \sin\left(\pi \frac{x^2 + y^2 - r^2}{R^2 - r^2}\right), \quad \text{for some constant } a \in [0, 1),$$

which decreases in terms of the distance from the bank, and the random perturbation of the wind is given by the constant diffusion coefficient  $\sigma = \text{diag}(\sigma_x, \sigma_y)$ . We also assume that the ship moves with a constant velocity  $v_s$ , and the captain can control the boat's direction instantaneously, which leads to the following dynamics of the boat in the region:

$$\begin{pmatrix} dX_t^{x,\alpha} \\ dY_t^{x,\alpha} \end{pmatrix} = \begin{pmatrix} v_c(X_t^{x,\alpha}, Y_t^{x,\alpha}) + v_s \cos(\alpha_t) \\ v_s \sin(\alpha_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix} dW_t, \quad t \geq 0; \quad \begin{pmatrix} X_0^{x,\alpha} \\ Y_0^{x,\alpha} \end{pmatrix} = x,$$

where  $\alpha_t \in \mathbf{A} = [0, 2\pi]$  represents the angle (measured counter-clockwise) between the positive  $x$ -axis and the direction of the boat. Finally, we assume the exit cost  $g \equiv 0$  on  $\partial B_r(0)$  and  $g \equiv 1$  on  $\partial B_R(0)$ , which represents that the controller prefers to exit the domain through the inner boundary instead of the outer one (see Fig. 1). Then, the corresponding Dirichlet problem for the value function  $u$  in (6.1) is given by:  $u \equiv 0$  on  $\partial B_r(0)$ ,  $u \equiv 1$  on  $\partial B_R(0)$ , and

$$\begin{aligned} F(u) &= -\frac{1}{2}(\sigma_x^2 u_{xx} + \sigma_y^2 u_{yy}) - v_c u_x \\ &\quad - v_s \inf_{\alpha \in \mathbf{A}} [(\cos(\alpha), \sin(\alpha))^T \nabla u] - \sup_{\|\beta\|_\infty \leq \kappa} [\beta^T (\sigma \nabla u)] - f \\ &= -\frac{1}{2}(\sigma_x^2 u_{xx} + \sigma_y^2 u_{yy}) - v_c u_x + v_s \|\nabla u\|_{\ell^2} - \kappa \|\sigma \nabla u\|_{\ell^1} - f = 0, \quad \text{in } \Omega, \end{aligned} \quad (6.2)$$

where  $\|\cdot\|_{\ell^1}$  and  $\|\cdot\|_{\ell^2}$  denote the  $\ell^1$ -norm and  $\ell^2$ -norm on  $\mathbb{R}^2$ , respectively. The optimal feedback control laws can be further computed as

$$\alpha^* = \pi + \theta, \quad \beta^* = \kappa (\text{sgn}(\sigma_x u_x), \text{sgn}(\sigma_y u_y))^T, \quad \text{a.e. in } \Omega, \quad (6.3)$$

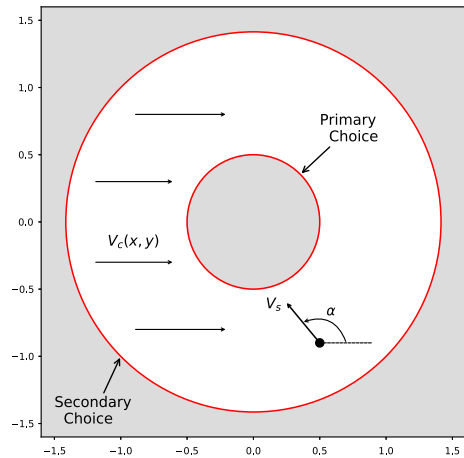
where  $u \in H^2(\Omega)$  is the strong solution to (6.2), and  $\theta \in (-\pi, \pi]$  is the angle between  $\nabla u$  and the positive  $x$  direction. Note that Eq. (6.2) is neither convex nor concave in  $\nabla u$ .

## 6.1 Implementation Details

In this section, we discuss the implementation details of Algorithm 2 for solving (6.2) with multilayer neural networks (see Definition 4.1) as the trial functions. We shall now introduce the architecture of the neural networks, the involved hyper-parameters, and various computational aspects of the training process.

For simplicity, we shall adopt a fixed set of trial functions  $\mathcal{F}_M$  for all policy iterations, which contains fully connected networks with the activation function  $\varrho(y) = \tanh(y)$ , the depth  $L$ , and the dimension of each hidden layer  $H$ . The hyper-parameters  $L$  and  $H$  will be chosen depending on the complexity of the problem, which ensures that  $\mathcal{F}_M$  admits sufficient flexibility to approximate the solutions within the desired accuracy. More complicated architectures of neural networks with shortcut

**Fig. 1** Zermelo navigation problem in an annulus



connections can be adopted to further improve the performance of the algorithm (see e.g. [16,44]).

We then proceed to discuss the computation of the cost functional  $J_k$  in (4.2) for each policy iteration. It is well known that Sobolev norms of functions on sufficiently smooth boundaries can be explicitly computed via local coordinate charts of the boundaries (see e.g. [21]). In particular, due to the annulus shaped domain and the constant boundary conditions used in our experiment, we can express the cost functional  $J_k$  as follows: for all  $k \in \mathbb{N}$  and  $u \in \mathcal{F}_M$ ,

$$J_k(u) = \|L_k u - f_k\|_{L^2(\Omega)}^2 + \sum_{l=r,R} \left[ \|u - g\|_{L^2(\partial B_l(0))}^2 + \gamma \left( \int_{-\pi}^{\pi} |D_\theta(u \circ \Phi_l)|^2 d\theta + \int_{(-\pi,\pi)^2} \frac{|D_\theta(u \circ \Phi_l)(\theta_1) - D_\theta(u \circ \Phi_l)(\theta_2)|^2}{|\theta_1 - \theta_2|^2} d\theta_1 d\theta_2 \right) \right], \quad (6.4)$$

where we define the map  $\Phi_l : \theta \in (-\pi, \pi) \rightarrow (l \cos(\theta), l \sin(\theta)) \in \partial B_l(0)$  for  $l = r, R$ . Note that we introduce an extra weighting parameter  $\gamma > 0$  in (6.4), which helps achieve the optimal balance between the residual of the PDE and the residuals of the boundary data. We set the parameter  $\gamma = 0.1$  for all the computations.

The cost functional (6.4) is further approximated by an empirical cost via the collocation method (see [6,35]), where we discretize  $\Omega$  and  $\Theta = (-\pi, \pi)^2$  by sets of collocation points  $\Omega_d = \{x_i \in \Omega \mid 1 \leq i \leq N_d\}$  and  $\Theta_d = \{\theta = (\theta_{1,i}, \theta_{2,i}) \in \Theta \mid 1 \leq i \leq N_b\}$ , respectively, and write the discrete form of (6.4) as follows: for all  $k \in \mathbb{N}$  and  $u \in \mathcal{F}_M$ ,

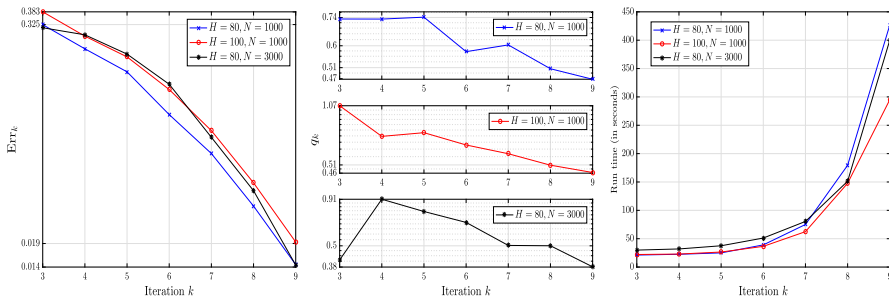
$$J_{k,d}(u) = \frac{|\Omega|}{N_d} \sum_{x_i \in \Omega_d} |L_k u(x_i) - f_k(x_i)|^2 + \sum_{l=r,R} \left[ \frac{|\partial B_l(0)|}{N_b} \sum_{\theta \in \Theta_d} |(u - g) \circ \Phi_l(\theta_{1,i})|^2 \right. \\ \left. + \gamma \left( \frac{2\pi}{N_b} \sum_{\theta \in \Theta_d} |D_\theta(u \circ \Phi_l)|^2(\theta_{1,i}) + \frac{(2\pi)^2}{N_b} \sum_{\theta \in \Theta_d} \frac{|D_\theta(u \circ \Phi_l)(\theta_{1,i}) - D_\theta(u \circ \Phi_l)(\theta_{2,i})|^2}{|\theta_{1,i} - \theta_{2,i}|^2} \right) \right], \quad (6.5)$$

where  $|\Omega| = \pi(R^2 - r^2)$ , and  $|\partial B_l(0)| = 2\pi l$  for  $l = r, R$  are, respectively, the Lebesgue measures of the domain and boundaries. Note that the choice of the smooth activation function  $\varrho(y) = \tanh(y)$  implies that every trial function  $u \in \mathcal{F}_M$  is smooth, hence all its derivatives are well defined at any given point. For simplicity, we take the same number of collocation points in the domain and on the boundaries, i.e.,  $N_d = N_b = N$ .

It is clear that the choice of collocation points is crucial for the accuracy and efficiency of the algorithm. Since the total number of points in a regular grid grows exponentially with respect to the dimension, such a construction is infeasible for high-dimensional problems. Moreover, it is well known that uniformly distributed pseudorandom points in high dimensions tend to cluster on hyperplanes and lead to a suboptimal distribution by relevant measures of uniformity (see e.g. [6, 11]). Therefore, we shall generate collocation points by a quasi-Monte Carlo (QMC) method based on low-discrepancy sequences. In particular, we first define points in  $[0, 1]^2$  from the generalized Halton sequence (see [17]) and then map those points into the annulus via the polar map  $(x, y) \mapsto (l \cos(\psi), l \sin(\psi))$ , where  $l = \sqrt{(R^2 - r^2)x + r^2}$  and  $\psi = 2\pi y$  for all  $(x, y) \in [0, 1]^2$ . The above transformation preserves fractional area, which ensures that a set of well-distributed points on the square will map to a set of points spread evenly over the annulus. We also use Halton points to approximate the (one-dimensional) boundary segments.

Now we are ready to describe the training process, i.e., how to optimize (6.5) over all trial functions in  $\mathcal{F}_M$ . The optimization is performed by using the well-known Adam stochastic gradient descent (SGD) algorithm [31] with a decaying learning rate schedule. At each SGD iteration, we randomly draw a mini-batch of points with size  $B = 25$  from the collection of collocation points, and perform gradient descent based on these samples. We initialize the learning rate at  $10^{-3}$  and decrease it by a factor of 0.5 for every 2000 SGD iterations for the examples with analytic solutions in Sect. 6.2, while for the examples without analytic solutions in Sect. 6.3 we decrease the learning rate by a factor of 0.5 once the total number of iterations reaches one of the milestones 2000, 4000, 6000, 10,000, 20,000, and 30,000.

We implement Algorithm 2 using PyTorch and perform all computations on a NVIDIA Tesla K40 GPU with 12 GB memory. The entire algorithm can be briefly summarized as follows. Let  $\{\eta_k\}_{k=0}^\infty$  be a given sequence, denoting the accuracy requirement for each policy iteration. For each  $k \in \mathbb{N} \cup \{0\}$ , given the previous iterate  $u^k$ , we compute the feedback controls as in (6.3) and obtain the controlled coefficients as defined in (3.3). Then we apply the SGD method with analytically derived gradient to optimize  $J_{k,d}$  over  $\mathcal{F}_M$  until we obtain a solution  $u^{k+1}$  satisfying  $J_{k,d}(u^{k+1}) \leq \eta_k \min(\|u^{k+1} - u^k\|_{2,d}^2, \eta_0)$ , where  $\|\cdot\|_{2,d}$  denotes the discrete  $H^2$ -norm



**Fig. 2** Impact of the training sample size  $N$  and the hidden width  $H$  on the performance of Algorithm 2; from left to right: relative errors (plotted in a log scale),  $q$ -factors and the overall runtime for all policy iterations

evaluated based on the training samples in  $\Omega_d$ . We then proceed to the next policy iteration and terminate Algorithm 2 once the desired accuracy is achieved.

## 6.2 Examples with Analytical Solutions

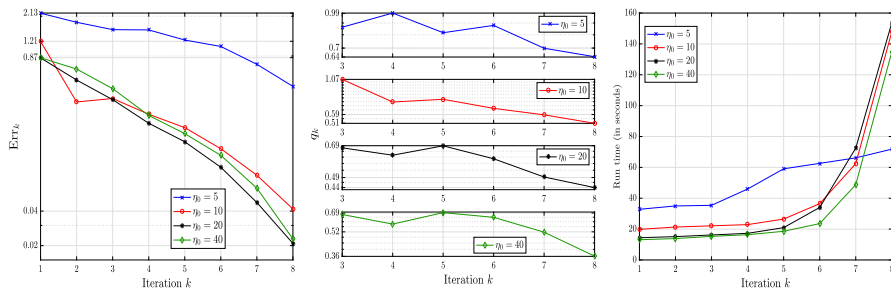
In this section, we shall examine the convergence of Algorithm 2 for solving Dirichlet problems of the form (6.2) with known solutions. In particular, we shall choose a running cost  $f$  such that the analytical solution to (6.2) is given by  $u^*(x, y) = \sin(\pi r^2/2) - \sin(\pi(x^2 + y^2)/2)$  for all  $(x, y) \in \Omega$ . To demonstrate the generalizability and the superlinear convergence of the numerical solutions obtained by Algorithm 2, we generate a different set of collocation points in  $\Omega$  of the size  $N_{\text{val}} = 2000$ , and use them to estimate the relative error and the  $q$ -factor of the numerical solution  $u^k$  obtained from the  $k$ -th policy iteration for all  $k \in \mathbb{N}$ :

$$\text{Err}_k = \frac{\|u^k - u^*\|_{2,\Omega,\text{val}}}{\|u^*\|_{2,\Omega,\text{val}}} \quad \text{and} \quad q_k = \frac{\|u^k - u^*\|_{2,\Omega,\text{val}}}{\|u^{k-1} - u^*\|_{2,\Omega,\text{val}}}.$$

We use neural networks with depth  $L = 4$  and varying  $H$  as trial functions, initialize Algorithm 2 with  $u^0 = 0$ , and perform experiments with the following model parameters:  $a = 0.04$ ,  $\sigma_x = 0.5$ ,  $\sigma_y = 0.2$ ,  $r = 0.5$ ,  $R = \sqrt{2}$ ,  $\kappa = 0.1$  and  $v_s = 0.6$ .

Figure 2 depicts the performance of Algorithm 2 with different sizes of training samples and the dimensions of hidden layers, which are denoted by  $N$  and  $H$ , respectively. The hyper-parameters  $\{\eta_k\}_{k=0}^\infty$  are chosen as  $\eta_0 = 10$  and  $\eta_k = 2^{-k}$  for all  $k \in \mathbb{N}$ . One can clearly see from Fig. 2 (left) and (middle) that, despite the fact that Algorithm 2 is initialized with a relatively poor initial guess, the numerical solutions converge superlinearly to the exact solution in the  $H^2$ -norm for all these combinations of  $H$  and  $N$ , which confirms the theoretical result in Theorem 4.3. It is interesting to observe from Fig. 2 that even though increasing either the complexity of the networks (the red lines) or the size of training samples (the black lines) seems to accelerate the training process slightly (right), neither of them ensures a higher generalization accuracy on the testing samples (left). In all our computations, the accuracies of numerical solutions in the  $L^2$ -norm and the  $H^1$ -norm are in general higher than the accuracy in the  $H^2$ -norm.



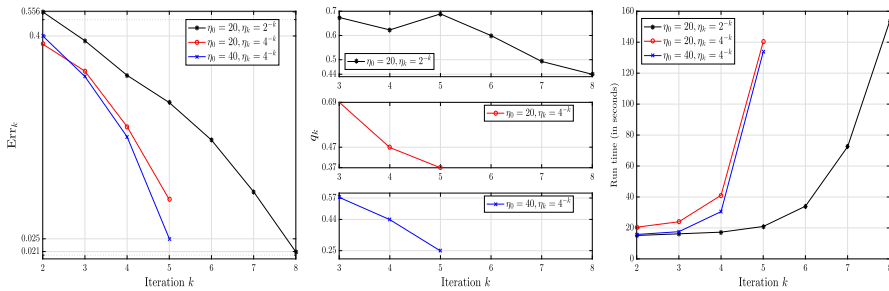


**Fig. 3** Impact of  $\eta_0$  on the performance of Algorithm 2; from left to right: relative errors (plotted in a log scale),  $q$ -factors and the overall runtime for all policy iterations

For example, both the  $L^2$ -relative error and the  $H^1$ -relative error of the numerical solution obtained at the 9th policy iteration with  $H = 80$ ,  $N = 1000$  are 0.0045.

We then proceed to analyze the effects of the hyper-parameters  $\{\eta_k\}_{k=0}^\infty$ . Roughly speaking, the magnitude of  $\eta_0$  indicates the accuracy of the iterates  $\{u^k\}_{k=1}^\infty$  to the linear Dirichlet problems in the initial stage of Algorithm 2, while the decay of  $\{\eta_k\}_{k=1}^\infty$  determines the speed at which the  $q$ -factors  $\{q_k\}_{k=1}^\infty$  converge to 0, at an extra cost of solving the optimization problem in a given iteration more accurately for smaller  $q_k$ . Figure 3 presents the numerical results for different choices of  $\eta_0$  with a fixed training sample size  $N = 1000$ , hidden width  $H = 100$  and  $\eta_k = 2^{-k}$  for all  $k \geq 1$ . Note that solving each linear equation extremely accurate in the initial stage, i.e., by choosing  $\eta_0$  to be a small value (the blue line), may not be beneficial for the overall performance of the algorithm in terms of both the accuracy and computational efficiency. This is due to the fact that the initialization of the algorithm is in general far from the exact solution to the semilinear boundary value problem, and so are the solutions of the linear equations arising from the first few policy iterations. In fact, it appears in our experiments that the choices of  $\eta_0 = 20, 40$  lead to the optimal performance of Algorithm 2, which solves the initial equations sufficiently accurately, and leverages the superlinear convergence of policy iteration to achieve a higher accuracy with a similar computational cost.

We further perform computations with different choices of  $\{\eta_k\}_{k=1}^\infty$  by fixing the training sample size  $N = 1000$  and the hidden width  $H = 100$ . Numerical results are shown in Fig. 4, from which we can clearly observe that the iterates obtained with  $\eta_k = 4^{-k}$ ,  $k \in \mathbb{N}$ , converge more rapidly to the exact solution. Note that for  $\eta_k = 4^{-k}$ , the optimal performance of the algorithm is achieved at  $\eta_0 = 40$  instead of  $\eta_0 = 20$ . This is due to the fact that we solve the first linear Dirichlet problem up to the accuracy  $\eta_0\eta_1$  (if we ignore the requirement that  $J_0(u^1) \leq \eta_1\|u^1 - u^0\|_{H^2(\Omega)}^2$  in (4.2)), hence one needs to enlarge  $\eta_0$  for a smaller  $\eta_k$ , such that  $\eta_0\eta_1$  is of the same magnitude as before. We observe that the rapid convergence of policy iteration indeed improves the efficiency of the algorithm, in the sense that, to achieve the same accuracy, Algorithm 2 with  $\eta_k = 4^{-k}$  requires slightly less computational time than Algorithm 2 with  $\eta_k = 2^{-k}$ , even though Algorithm 2 with  $\eta_k = 4^{-k}$  takes more time to solve the linear equations for each policy iteration; see the last few iterations of the blue line



**Fig. 4** Impact of  $\eta_k$  on the performance of Algorithm 2; from left to right: relative errors (plotted in a log scale),  $q$ -factors and the overall runtime for all policy iterations

**Table 1** Numerical results with different parameters  $\{\eta_k\}_{k=0}^\infty$  for the scenario where the ship is slower than the wind ( $v_s = 0.5$ )

$H = 80, \eta_0 = 80, \eta_k = 2^{-k}$				$H = 80, \eta_0 = 40, \eta_k = 1/k$			
PI Itr	HJBI residual	SGD Itr	Run time (s)	PI Itr	HJBI residual	SGD Itr	Run time (s)
9	0.0466	11,030	922	58	0.0252	32,500	2729
10	0.0144	20,330	1710	59	0.0201	39,080	3275
11	0.0046	45,510	3820	60	0.0156	45,770	3836

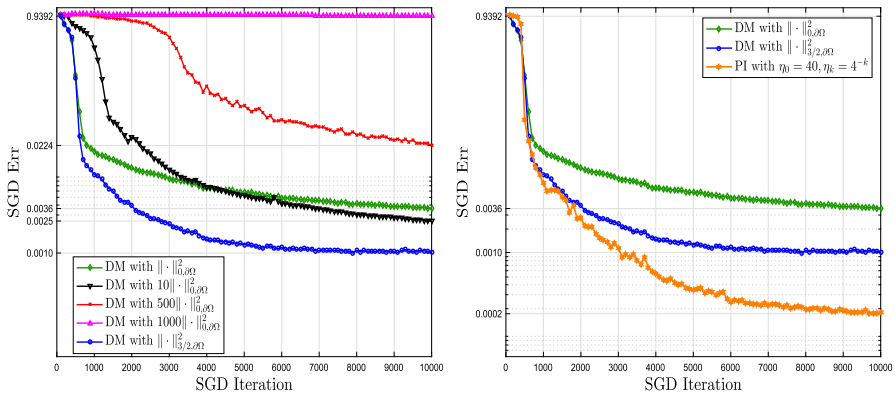
and the black line. This efficiency improvement is more pronounced for the practical problems with complicated solutions in Sect. 6.3; see Fig. 7 and Table 1.

Finally, we shall compare the efficiency of Algorithm 2 (with  $\eta_0 = 40, \eta_k = 4^{-k}$ ) to that of the direct methods (see e.g. [6, 16, 35, 44]) by fixing the trial functions (4-layer networks with hidden width  $H = 100$ ), the training samples (with size  $N = 1000$ ) and the learning rate of the SGD algorithm. In the direct methods, we shall directly apply the SGD method to minimize the following (discretized) squared residual of the semilinear boundary value problem (6.2):<sup>2</sup>

$$\|F(u)\|_{0,\Omega,\text{tra}}^2 + \|u - g\|_{X,\partial\Omega,\text{tra}}^2, \quad (6.6)$$

where  $\|\cdot\|_{0,\Omega,\text{tra}}$  is the discrete  $L^2$  interior norm evaluated from the training samples in  $\Omega$ , and  $\|\cdot\|_{X,\partial\Omega,\text{tra}}$  is a certain discrete boundary norm evaluated from samples on the boundary. In particular, we shall perform computations by setting  $\|\cdot\|_{X,\partial\Omega,\text{tra}}^2 = \|\cdot\|_{3/2,\partial\Omega,\text{tra}}^2$  (defined as in (6.5)) and  $\|\cdot\|_{X,\partial\Omega,\text{tra}}^2 = \vartheta \|\cdot\|_{0,\partial\Omega,\text{tra}}^2$  with different choices of  $\vartheta > 0$  ( $\vartheta = 1$  in [44] and  $\vartheta \in \{500, 1000\}$  in [16]), which will be referred to as “DM with  $\|\cdot\|_{3/2,\partial\Omega}^2$ ” and “DM with  $\vartheta \|\cdot\|_{0,\partial\Omega}^2$ ”, respectively, in the following discussion. For both the direct methods and Algorithm 2, we shall estimate the  $H^2$ -

<sup>2</sup> Strictly speaking, the squared residual (6.6) is not differentiable (with respect to the network parameters) at the samples where one of the first partial derivatives of the current iterate  $u$  is zero, due to the nonsmooth functions  $\|\cdot\|_{\ell^1}, \|\cdot\|_{\ell^2} : \mathbb{R}^2 \rightarrow [0, \infty)$  in the HJBI operator  $F$  (see (6.2)). In practice, PyTorch will assign 0 as partial derivatives of  $\|\cdot\|_{\ell^1}$  and  $\|\cdot\|_{\ell^2}$  functions at their nondifferentiable points and use it in the backward propagation.



**Fig. 5** Relative errors of the direct methods and Algorithm 2 with different numbers of SGD iterations (plotted in a log scale); from left to right: improvements caused by the  $H^{3/2}$ -boundary norm and by policy iteration

relative error of the numerical solution  $\hat{u}_i$  obtained from the  $i$ -th SGD iteration by using the same testing samples in  $\Omega$  of the size  $N_{\text{val}} = 2000$  as follows:

$$\text{SGD Err}_i = \|\hat{u}_i - u^*\|_{2,\Omega,\text{val}} / \|u^*\|_{2,\Omega,\text{val}},$$

where  $u^*$  denotes the analytical solution to (6.2).

Figure 5 (left) depicts the  $H^2$ -convergence of “DM with  $\|\cdot\|_{3/2,\partial\Omega}^2$ ” and “DM with  $\vartheta \|\cdot\|_{0,\partial\Omega}^2$ ” (with various choices of  $\vartheta > 0$ ) as the number of SGD iterations tends to infinity, which clearly shows that, compared with using the  $L^2$ -boundary norm as in [16,44], incorporating the  $H^{3/2}$ -boundary norm in the loss function helps achieve a higher  $H^2$ -accuracy of the numerical solutions. It is interesting to point out that even though penalizing the  $L^2$ -norm of the boundary term with a suitable parameter  $\vartheta$  helps improve the accuracy of “DM with  $\vartheta \|\cdot\|_{0,\partial\Omega}^2$ ” as suggested in [16], in our experiments,  $\vartheta = 10$  leads to the best  $H^2$ -convergence of “DM with  $\vartheta \|\cdot\|_{0,\partial\Omega}^2$ ” (after  $10^4$  SGD iterations) among other choices of  $\vartheta \in \{0.1, 1, 5, 10, 20, 50, 100, 500, 1000\}$ .

Figure 5 (right) presents the decay of  $H^2$ -relative errors with respect to the number of SGD iterations used in “DM with  $\|\cdot\|_{0,\partial\Omega}^2$ ”, “DM with  $\|\cdot\|_{3/2,\partial\Omega}^2$ ” and Algorithm 2, which clearly demonstrates that the superlinear convergence of policy iteration significantly accelerates the convergence of the algorithm. In particular, the accuracy enhancement of Algorithm 2 over “DM with  $\|\cdot\|_{0,\partial\Omega}^2$ ” (or equivalently the deep Galerkin method proposed in [44]) is of a factor of 20 with  $10^4$  SGD iterations. We remark that the training time of Algorithm 2 is only slightly longer than that of “DM with  $\|\cdot\|_{0,\partial\Omega}^2$ ” (the runtimes of Algorithm 2 and “DM with  $\|\cdot\|_{0,\partial\Omega}^2$ ” with  $10^4$  SGD iterations are 333 and 308 s, respectively), since Algorithm 2 requires to determine whether a given iterate solves the policy evaluation equations sufficiently accurate (see (4.2)), in order to proceed to the next policy iteration step.

### 6.3 Examples Without Analytical Solutions

In this section, we shall demonstrate the performance of Algorithm 2 by solving (6.2) with  $f \equiv 1$ . This corresponds to a minimum time problem with preferred targets, whose solution in general is not known analytically. Numerical simulations will be conducted with the following model parameters:  $a = 0.2$ ,  $\sigma_x = 0.5$ ,  $\sigma_y = 0.2$ ,  $r = 0.5$ ,  $R = \sqrt{2}$  and  $\kappa = 0.1$  but two different values of  $v_s$ ,  $v_s = 0.5$  and  $v_s = 1.2$ , which are associated with the two scenarios where the ship moves slower than and faster than the wind, respectively. The algorithm is initialized with  $u^0 = 0$ .

We remark that this is a numerically challenging problem due to the fact that the convection term in (6.2) dominates the diffusion term, which leads to a sharp change of the solution and its derivatives near the boundaries. However, as we shall see, these boundary layers can be captured effectively by the numerical solutions of Algorithm 2.

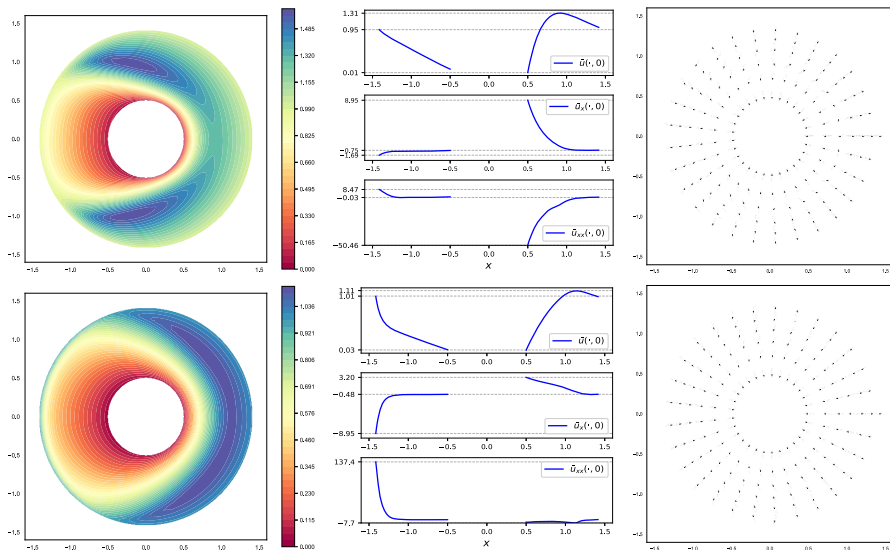
Figure 6 presents the numerical results for the two different scenarios obtained by Algorithm 2 with  $N = 2000$ ,  $\eta_0 = 40$  and  $\eta_k = 1/k$  for all  $k \in \mathbb{N}$ . The set of trial functions consists of all fully connected neural networks with depth  $L = 7$  and hidden width  $H = 50$  (the total number of parameters in this network is 12,951). We can clearly see from Fig. 6 (left) and (middle) that for both scenarios, the numerical solution  $\bar{u}$  and its derivatives are symmetric with respect to the axis  $y = 0$ , and change rapidly near the boundaries.

The feedback control strategies, computed by (6.3), are depicted in Fig. 6 (right). If the ship starts from the left-hand side and travels toward the inner boundary, then the expected travel time to  $\partial B_r(0)$  is around  $\frac{R-r}{v_s+v_c}$ , which is smaller than the exit cost along  $\partial B_R(0)$ . Hence, the ship would move in the direction of the positive  $x$ -axis for both cases,  $v_s < v_c$  and  $v_s > v_c$ . However, the optimal control is different for the two scenarios if the ship is on the right-hand side. For the case where the ship's speed is less than the wind ( $v_s = 0.5$ ), if the ship is closed to  $\partial B_r(0)$ , then it would move in the direction of the negative  $x$ -axis, hoping the random perturbation of the wind would bring it to the preferred target, while if it is far from  $\partial B_r(0)$ , then it has less chance to reach  $\partial B_r(0)$ , so it would move along the positive  $x$ -axis. On the other hand, for the case where the ship's speed is larger than the wind ( $v_s = 1.2$ ), the ship would in general try to reach the inner boundary. However, if the ship is sufficiently close to  $\partial B_R(0)$  in the right-hand half-plane, then the expected travel time to  $\partial B_r(0)$  is around  $\frac{R-r}{v_s-v_c}$ , which is larger than the exit cost along  $\partial B_R(0)$ . Hence, the ship would choose to exit directly from the outer boundary.

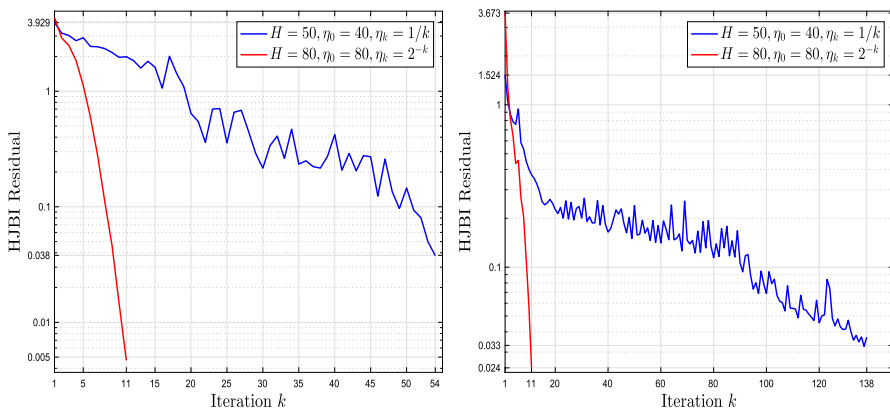
We then analyze the convergence of Algorithm 2 by performing computations with 7-layer networks with different hidden width  $H$  (networks with wider hidden layers are employed such that every linear Dirichlet problem can be solved more accurately) and parameters  $\{\eta_k\}_{k=0}^\infty$ . For any given iterate  $u^k$ , we shall consider the following (squared) residual of the semilinear boundary value problem (6.2) :

$$\text{HJBI Residual} := \|F(u^k)\|_{0,\Omega,\text{val}}^2 + \|u^k - g\|_{3/2,\partial\Omega,\text{val}}^2, \quad (6.7)$$

which will be evaluated similar to (6.5) based on testing samples in  $\Omega$  and on  $(\partial\Omega)^2$  of the same size  $N_{\text{val}} = 2000$ . Figure 7 presents the decay of the residuals in terms



**Fig. 6** Numerical results for the two different scenarios; from top to bottom: the ship moves slower than the wind ( $v_s = 0.5$ ) and faster than the wind ( $v_s = 1.2$ ); from left to right: the value function  $\bar{u}$ , the numerical solutions along  $y = 0$ , and feedback control strategies



**Fig. 7** Residuals of the HJBI Dirichlet problems for the two different scenarios with respect to the number of policy iterations (plotted in a log scale); from left to right: the ship moves slower than the wind ( $v_s = 0.5$ ) and faster than the wind ( $v_s = 1.2$ )

of the number of policy iterations, which suggests the  $H^2$ -superlinear convergence of the iterates  $\{u^k\}_{k=0}^\infty$  (the  $H^2$ -norms of the last iterates for  $v_s = 0.5$  and  $v_s = 1.2$  are 20.3 and 31.8, respectively). Note that the parameter  $\eta_k = 1/k$ ,  $k \in \mathbb{N}$  leads to a slower and more oscillating convergence of the iterates  $\{u^k\}_{k=0}^\infty$ , due to the fact that we apply a mini-batch SGD method to optimize the discrete cost functional  $J_{k,d}$  for each policy iteration. A faster and smoother convergence can be achieved by choosing a more rapidly decaying  $\{\eta_k\}_{k=1}^\infty$ .

We further investigate the influence of the parameters  $\{\eta_k\}_{k=1}^\infty$  on the accuracy and efficiency of Algorithm 2 in detail. Algorithm 2 is carried out with the same trial functions (7-layer networks with hidden width  $H = 80$  and complexity 32,721) but different  $\{\eta_k\}_{k=1}^\infty$  (we choose different  $\eta_0$  to keep the quantity  $\eta_0\eta_1$  constant), and the numerical results are summarized in Table 1. One can clearly see that a more rapidly decaying  $\{\eta_k\}_{k=1}^\infty$  results in a better overall performance (in terms of accuracy and computational time), even though it requires more time to solve the linear equation for each policy iteration. The rapid decay of  $\{\eta_k\}_{k=1}^\infty$  not only accelerates the superlinear convergence of the iterates  $\{u^k\}_{k=1}^\infty$ , but also helps to eliminate the oscillation caused by the randomness in the SGD algorithm (see Fig. 7), which enables us to achieve a higher accuracy with less total computational effort. However, we should keep in mind that smaller  $\{\eta_k\}_{k=1}^\infty$  means that we need to solve all linear equations with higher accuracy, which subsequently requires more complicated networks and more careful choices of the optimizers for  $J_{k,d}$ . Therefore, in general, we need to tune the balance between the superlinear convergence rate of policy iteration and the computational costs of the linear solvers, in order to achieve optimal performance of the algorithm.

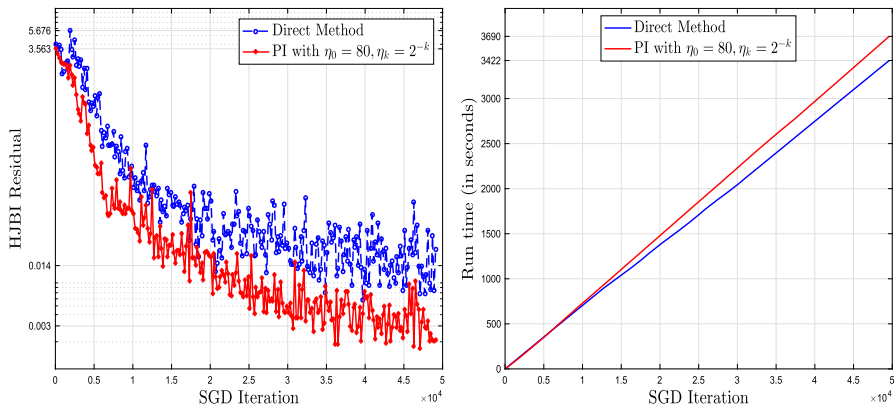
Finally, we shall compare the performance of Algorithm 2 (with  $\eta_0 = 80$ ,  $\eta_k = 2^{-k}$ ) and the direct method (with  $\|\cdot\|_{X,\partial\Omega,\text{tra}} = \|\cdot\|_{3/2,\partial\Omega,\text{tra}}$  in (6.6)) by fixing the trial functions (7-layer networks with hidden width  $H = 80$ ), the training samples and the learning rates of the SGD algorithms. For both methods, we shall consider the following squared residual for each iterate  $\hat{u}_i$  obtained from the  $i$ -th SGD iteration (see (6.7)):

$$\text{HJBI Residual} := \|F(\hat{u}_i)\|_{0,\Omega,\text{val}}^2 + \|\hat{u}_i - g\|_{3/2,\partial\Omega,\text{val}}^2.$$

Figure 8 (left) presents the decay of the residuals as the number of SGD iterations tends to infinity, which demonstrates the efficiency improvement of Algorithm 2 over the direct method. The superlinear convergence of policy iteration helps to provide better initial guesses of the SGD algorithm, which leads to a more rapidly decaying loss curve with smaller noise (on the validation samples); the HJBI residuals obtained in the last 1000 SGD iterations of Algorithm 2 (resp. the direct method) oscillates around the value 0.0028 (resp. 0.0144) with a standard derivation 0.00096 (resp. 0.0093).

## 7 Conclusions

This paper develops a neural network-based policy iteration algorithm for solving HJBI boundary value problems arising in stochastic differential games of diffusion processes with controlled drift and state constraints. We establish the  $q$ -superlinear convergence of the algorithm in  $H^2(\Omega)$  with an arbitrary initial guess and also the pointwise (almost everywhere) convergence of the numerical solutions and their (first and second order) derivatives, which subsequently leads to convergent approximations of optimal feedback controls. The convergence results also hold for general trial functions, including kernel functions and high-order separable polynomials used in global spectral methods. Numerical examples for stochastic Zermelo navigation problems are presented to illustrate the theoretical findings.



**Fig. 8** Performance comparison of the direct method and Algorithm 2 for the scenario where the ship is slower than the wind ( $v_s = 0.5$ ); from left to right: residuals (plotted in a log scale) and overall runtime for all SGD iterations

To the best of our knowledge, this is the first paper which demonstrates the global superlinear convergence of policy iteration for nonconvex HJBI equations in function spaces and proposes convergent neural network-based numerical methods for solving the solutions of nonlinear boundary value problems and their derivatives. Natural next steps would be to extend the inexact policy iteration algorithm to parabolic HJBI equations, and to employ neural networks with tailored architectures to enhance the efficiency of the algorithm for solving high-dimensional problems.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Some Fundamental Results

Here, we collect some well-known results which are used frequently in the paper.

We start with the well-posedness of strong solutions to Dirichlet boundary value problems. In the sequel, we shall denote by  $\tau$  the trace operator.

**Theorem A.1** ([19, Theorem 1.2.19]) *Let  $\Omega$  be a bounded  $C^{1,1}$  domain. Suppose that for all  $1 \leq i, j \leq n$ ,  $a^{ij}$  is in  $C(\bar{\Omega})$ , and  $b^i, c$  are in  $L^\infty(\Omega)$ , satisfying  $c \geq 0$  and*

$$\sum_{i,j=1}^n a^{ij}(x) \xi_i \xi_j \geq \lambda |\xi|^2, \quad \text{for all } \xi \in \mathbb{R}^n \text{ and for almost every } x \in \Omega, \quad (\text{A.1})$$

for some constant  $\lambda > 0$ . Then, for every  $f \in L^2(\Omega)$  and  $g \in H^{3/2}(\partial\Omega)$ , there exists a unique strong solution  $u \in H^2(\Omega)$  to the Dirichlet problem

$$-a^{ij}\partial_{ij}u + b^i\partial_iu + cu = f, \text{ in } \Omega; \quad \tau u = g, \text{ on } \partial\Omega,$$

and the following estimate holds with a constant  $C$  independent of  $f$  and  $g$ :

$$\|u\|_{H^2(\Omega)} \leq C(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)}).$$

The next theorem shows the well-posedness of oblique boundary value problems.

**Theorem A.2** ([19, Theorem 1.2.20]) *Let  $\Omega$  be a bounded  $C^{1,1}$  domain. Suppose that for all  $1 \leq i, j \leq n$ ,  $a^{ij}$  is in  $C(\bar{\Omega})$ , and  $b^i, c$  are in  $L^\infty(\Omega)$ , satisfying  $c \geq 0$  and the uniform elliptic condition (A.1) for some constant  $\lambda > 0$ .*

*Assume in addition that  $\{\gamma^j\}_{j=0}^n \subseteq C^{0,1}(\partial\Omega)$ ,  $\gamma^0 \geq 0$  on  $\partial\Omega$ ,  $\text{ess sup}_{\partial\Omega} c + \max_{\partial\Omega} \gamma^0 > 0$ , and  $\sum_{j=1}^n \gamma^j v_j \geq \mu$  on  $\partial\Omega$  for some constant  $\mu > 0$ , where  $\{v_j\}_{j=1}^n$  are the components of the unit outer normal vector field on  $\partial\Omega$ . Then, for every  $f \in L^2(\Omega)$  and  $g \in H^{1/2}(\partial\Omega)$ , there exists a unique strong solution  $u \in H^2(\Omega)$  to the following oblique derivative problem:*

$$-a^{ij}\partial_{ij}u + b^i\partial_iu + cu = f, \text{ in } \Omega; \quad \gamma^j\tau(\partial_ju) + \gamma^0\tau u = g, \text{ on } \partial\Omega,$$

and the following estimate holds with a constant  $C$  independent of  $f$  and  $g$ :

$$\|u\|_{H^2(\Omega)} \leq C\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right).$$

We then recall several important measurability results. The following measurable selection theorem follows from Theorems 18.10 and 18.19 in [1] and ensures the existence of a measurable selector maximizing (or minimizing) a Carathéodory function.

**Theorem A.3** *Let  $(S, \Sigma)$  a measurable space and  $X$  be a separable metrizable space. Let  $\Gamma : S \rightrightarrows X$  be a measurable set-valued mapping with nonempty compact values, and suppose  $g : S \times X \rightarrow \mathbb{R}$  is a Carathéodory function. Define the value function  $m : S \rightarrow \mathbb{R}$  by  $m(s) = \max_{x \in \Gamma(s)} g(s, x)$ , and the set-valued map  $\mu : S \rightrightarrows X$  by  $\mu(s) = \{x \in \Gamma(s) \mid g(s, x) = m(s)\}$ . Then, we have*

1. *The value function  $m$  is measurable.*
2. *The set-valued mapping  $\mu$  is measurable and has nonempty and compact values. Moreover, there exists a measurable function  $\psi : S \rightarrow X$  satisfying  $\psi(s) \in \mu(s)$  for each  $s \in S$ .*

The following theorem shows the arg max set-valued mapping is upper hemicontinuous.

**Theorem A.4** ([1, Theorem 17.31]) *Let  $X, Y$  be topological spaces,  $\Gamma \subset Y$  be a nonempty compact subset, and  $g : X \times \Gamma \rightarrow \mathbb{R}$  be a continuous function. Define the value function  $m : X \rightarrow \mathbb{R}$  by  $m(x) = \max_{y \in \Gamma} g(x, y)$ , and the set-valued map*



$\mu : X \rightrightarrows Y$  by  $\mu(x) = \{y \in \Gamma \mid g(x, y) = m(x)\}$ . Then,  $\mu$  has nonempty and compact values. Moreover, if  $Y$  is Hausdorff, then  $\mu$  is upper hemicontinuous, i.e., for every  $x \in X$  and every neighborhood  $U$  of  $\mu(x)$ , there is a neighborhood  $V$  of  $x$  such that  $z \in V$  implies  $\mu(z) \subset U$ .

Finally, we present a special case of [14, Theorem 2], which characterizes  $q$ -superlinear convergence of quasi-Newton methods for a class of semismooth operator-valued equations.

**Theorem A.5** *Let  $Y, Z$  be two Banach spaces, and  $F : Y \rightarrow Z$  be a given function with a zero  $y^* \in Y$ . Suppose there exists an open neighborhood  $V$  of  $y^*$  such that  $F$  is semismooth with a generalized differential  $\partial^* F$  in  $V$ , and there exists a constant  $L > 0$  such that*

$$\|y - y^*\|_Y / L \leq \|F(y) - F(y^*)\|_Z \leq L\|y - y^*\|_Y, \quad \forall y \in V.$$

*For some starting point  $y^0$  in  $V$ , let the sequence  $\{y^k\}_{k \in \mathbb{N}} \subset V$  satisfy  $y^k \neq y^*$  for all  $k$ , and be generated by the following quasi-Newton method:*

$$B_k s^k = -F(y^k), \quad y^{k+1} = y^k + s^k, \quad k = 0, 1, \dots$$

*where  $\{B_k\}_{k \in \mathbb{N}}$  is a sequence of bounded linear operators in  $\mathcal{L}(Y, Z)$ . Let  $\{A_k\}_{k \in \mathbb{N}}$  be a sequence of generalized differentials of  $F$  such that  $A_k \in \partial^* F(y^k)$  for all  $k$ , and let  $E_k = B_k - A_k$ . Then,  $y^k \rightarrow y^*$   $q$ -superlinearly if and only if  $\lim_{k \rightarrow \infty} y^k = y^*$  and  $\lim_{k \rightarrow \infty} \|E_k s^k\|_Z / \|s^k\|_Y = 0$ .*

## B Proof of Theorem 5.2

Let  $u^0 \in \mathcal{F}$  be an arbitrary initial guess, we shall assume without loss of generality that Algorithm 3 runs infinitely, i.e.,  $\|u^{k+1} - u^k\|_{H^2(\Omega)} > 0$  and  $u^k \neq u^*$  for all  $k \in \mathbb{N} \cup \{0\}$ .

We first show  $\{u^k\}_{k \in \mathbb{N}}$  converges to the unique solution  $u^*$  in  $H^2(\Omega)$ . For each  $k \geq 0$ , we can deduce from (5.3) that there exists  $f_k^e \in L^2(\Omega)$  and  $g_k^e \in H^{1/2}(\partial\Omega)$  such that

$$L_k u^{k+1} - f_k = f_k^e, \quad \text{in } \Omega; \quad B u^{k+1} = g_k^e, \quad \text{on } \partial\Omega, \quad (\text{B.1})$$

and  $\|f_k^e\|_{L^2(\Omega)}^2 + \|g_k^e\|_{H^{1/2}(\partial\Omega)}^2 \leq \eta_{k+1}(\|u^{k+1} - u^k\|_{H^2(\Omega)}^2)$  with  $\lim_{k \rightarrow \infty} \eta_k = 0$ . Then, we can proceed as in the proof of Theorem 4.3, and conclude that if  $c \geq c_0$  with a sufficiently large  $c_0$ , then  $\{u^k\}_{k \in \mathbb{N}}$  converges to the solution  $u^*$  of (5.1).

The  $q$ -superlinear convergence of Algorithm 3 can then be deduced by interpreting the algorithm as a quasi-Newton method for the operator equation  $\bar{F}(u) = 0$ , with the operator  $\bar{F} : u \in H^2(\Omega) \rightarrow (F(u), Bu) \in Z$ , where we introduce the Banach space  $Z := L^2(\Omega) \times H^{1/2}(\partial\Omega)$  with the usual product norm  $\|z\|_Z := \|z_1\|_{L^2(\Omega)} + \|z_2\|_{H^{1/2}(\partial\Omega)}$  for each  $z = (z_1, z_2) \in Z$ . Since  $B \in \mathcal{L}(H^2(\Omega), H^{1/2}(\partial\Omega))$ , we can directly infer

from Corollary 3.5 that  $\bar{F} : H^2(\Omega) \rightarrow Z$  is semismooth in  $H^2(\Omega)$ , with a generalized differential  $M_k = (L_k, \gamma^i \tau(\partial_i) + \gamma^0 \tau) \in \partial^* \bar{F}(u^k) \subset \mathcal{L}(H^2(\Omega), Z)$  for all  $k \in \mathbb{N} \cup \{0\}$ . Then, for each  $k \geq 0$ , by following the same arguments as in Theorem 4.3, we can construct a perturbed operator  $\delta M_k \in \mathcal{L}(H^2(\Omega), Z)$ , such that (B.1) can be equivalently written as  $(M_k + \delta M_k)s_k = -\bar{F}(u^k)$  with  $s_k = u^{k+1} - u^k$ , and  $\|\delta M_k s_k\|/\|s_k\|_{H^2(\Omega)} \leq \sqrt{2\eta_0\eta_{k+1}} \rightarrow 0$ , as  $k \rightarrow \infty$ . Finally, the regularity theory of elliptic oblique derivative problems (see Theorem A.2) shows that  $M_k$  is nonsingular for each  $k$ , and  $\|M_k^{-1}\|_{\mathcal{L}(Z, H^2(\Omega))} \leq C$  for some constant  $C$  independent of  $k$ . Hence, we can verify that there exist a neighborhood  $V$  of  $u^*$  and a constant  $L > 0$ , such that

$$\|u - u^*\|_{H^2(\Omega)}/L \leq \|\bar{F}(u) - \bar{F}(u^*)\|_Z \leq L\|u - u^*\|_{H^2(\Omega)}, \quad \forall u \in V,$$

which allows us to conclude from Theorem A.5 the  $q$ -superlinear convergence of  $\{u^k\}_{k \in \mathbb{N}}$ .

## References

1. C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed., Springer-Verlag, Berlin, 2006.
2. A. Alla, M. Falcone, and D. Kalise, *An efficient policy iteration algorithm for dynamic programming equations*, SIAM J. Sci. Comput., 37 (2015), pp. A181–A200.
3. J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, Basel, 1990.
4. 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.
5. R. W. Beard and T. W. McClain, *Successive Galerkin approximation algorithms for nonlinear optimal and robust control*, Internat. J. Control, (71) 1998, pp. 717–743.
6. J. Berg and K. Nyström, *A unified deep artificial neural network approach to partial differential equations in complex geometries*, Neurocomputing, (317) 2018, pp. 28–41.
7. P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods*, Springer, New York, 2009.
8. O. Bokanowski, S. Maroso, and H. Zidani, *Some convergence results for Howard's algorithm*, SIAM J. Numer. Anal., 47 (2009), pp. 3001–3026.
9. S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, New York, 1994.
10. R. Buckdahn and T. Y. Nie, *Generalized Hamilton–Jacobi–Bellman equations with Dirichlet boundary condition and stochastic exit time optimal control problem*, SIAM J. Control Optim., 54 (2016), pp. 602–631.
11. C. Cervellera and M. Muselli, *Deterministic design for neural network learning: An approach based on discrepancy*, IEEE Trans. Neural Networks, 15 (2004), pp. 533–544.
12. X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.
13. K.-C. Cheung, L. Ling, R. Schaback,  *$H^2$ -convergence of least-squares kernel collocation methods*, SIAM J. Numer. Anal. 56 (2018) 614–633.
14. A. L. Dontchev, *Generalizations of the Dennis–Moré theorem*, SIAM J. Optim., 22 (2012), pp. 821–830.
15. W. E, J. Han, and A. Jentzen, *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, Commun. Math. Stat., 5 (2017), pp. 349–380.
16. W. E and B. Yu, *The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems*, Commun. Math. Stat., 6 (2018), pp. 1–12.
17. H. Faure and C. Lemieux, *Generalized Halton sequence in 2008: a comparative study*, ACM Trans. Model. Comput. Simul., 19 (2009).

18. P. Forsyth and G. Labahn, *Numerical methods for controlled Hamilton–Jacobi–Bellman PDEs in finance*, J. Computational Finance, 11 (2007/2008, Winter), pp. 1–43.
19. M. G. Garroni and J. L. Menaldi, *Second order elliptic integro-differential problems*, Chapman & Hall/CRC, Boca Raton, FL, 2002.
20. D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd edition, Springer-Verlag, Berlin, New York, 1983.
21. E. Grisvard, *Elliptic problems in nonsmooth domains*, Pitman, Boston, MA, 1985.
22. J. Han and W. E, *Deep learning approximation for stochastic control problems*, preprint, [arXiv:1611.07422](https://arxiv.org/abs/1611.07422), 2016.
23. J. Han and J. Long, *Convergence of the deep BSDE method for coupled FBSDEs*, preprint, [arXiv:1811.01165v1](https://arxiv.org/abs/1811.01165v1), 2018.
24. M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.
25. K. Hornik, M. Stinchcombe, and H. White, *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*, Neural Networks, 3 (1990), pp. 551–560.
26. C. Huré, H. Pham, A. Bachouch, and N. Langrené, *Deep neural networks algorithms for stochastic control problems on finite horizon, part I: convergence analysis*, preprint, [arXiv:1812.04300](https://arxiv.org/abs/1812.04300), 2018.
27. K. Ito and K. Kunisch, *Semismooth Newton methods for variational inequalities of the first kind*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 41–62.
28. 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), A629–A652.
29. D. Kalise, S. Kundu, and K. Kunisch, *Robust feedback control of nonlinear PDEs by numerical approximation of high-dimensional Hamilton–Jacobi–Isaacs equations*, preprint, [arXiv:1905.06276](https://arxiv.org/abs/1905.06276), 2019.
30. B. Kerimkulov, D. Šiška, and Ł. Szpruch, *Exponential convergence and stability of Howard’s policy improvement algorithm for controlled diffusions*, preprint, [arXiv:1812.07846](https://arxiv.org/abs/1812.07846), 2018.
31. D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, CoRR preprint, [arxiv:1412.6980](https://arxiv.org/abs/1412.6980), 2014.
32. N. V. Krylov, *On the dynamic programming principle for uniformly nondegenerate stochastic differential games in domains and the Isaacs equations*, Probab. Theory Related Fields, 158 (2014), pp. 751–783.
33. N.V. Krylov, *Sobolev and Viscosity Solutions for Fully Nonlinear Elliptic and Parabolic Equations*, Mathematical Surveys and Monographs, 233, Amer. Math. Soc., Providence, RI, 2018.
34. H. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York, 1991.
35. E. Lagaris, A. Likas, and D. I. Fotiadis, *Artificial neural networks for solving ordinary and partial differential equations*, IEEE Trans. Neural Netw., 9 (1998) pp. 987–1000.
36. P.-L. Lions and A.-S. Sznitman, *Stochastic differential equations with reflecting boundary conditions*, Comm. Pure Appl. Math., 37 (1984) pp. 511–553.
37. P. Mohajerin Esfahani, D. Chatterjee, and J. Lygeros, *The stochastic reach-avoid problem and set characterization for diffusions*, Automatica, 70 (2016), pp. 43–56.
38. M.L. Puterman and S.L. Brumelle, *On the convergence of policy iteration in stationary dynamic programming*, Math. Oper. Res., 4 (1979), pp. 60–69.
39. C. Reisinger and Y. Zhang, *A penalty scheme and policy iteration for nonlocal HJB variational inequalities with monotone drivers*, preprint, [arXiv:1805.06255](https://arxiv.org/abs/1805.06255), 2018.
40. C. Reisinger and Y. Zhang, *Error estimates of penalty schemes for quasi-variational inequalities arising from impulse control problems*, preprint, [arXiv:1901.07841](https://arxiv.org/abs/1901.07841), 2019.
41. M. Royer, *Backward stochastic differential equations with jumps and related non-linear expectations*, Stochastic Process. Appl., 116 (2006), pp. 1358–1376.
42. M.S. Santos and J. Rust, *Convergence properties of policy iteration*, SIAM J. Control Optim., 42 (2004), pp. 2094–2115.
43. O. Shamir and T. Zhang, *Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes*, in Proceedings of the International Conference on Machine Learning, 2013.
44. J. Sirignano and K. Spiliopoulos, *DGM: A deep learning algorithm for solving partial differential equations*, J. Comput. Phys., 375 (2018), pp. 1339–1364.

45. I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of Hamilton-Jacobi-Bellman equations with Cordes coefficients*, SIAM J. Numer. Anal., 52 (2014), pp. 993–1016,
46. M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Ser. Optim. 11, SIAM, Philadelphia, 2011.
47. J. H. Witte and C. Reisinger, *Penalty methods for the solution of discrete HJB equations: Continuous control and obstacle problems*, SIAM J. Numer. Anal., 50 (2012), pp. 595–625.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.