POLICY ITERATION FOR HAMILTON-JACOBI-ISAACS EQUATIONS WITH CONTROL CONSTRAINTS AND COMPARISON WITH HAMILTON-JACOBI-BELLMAN EQUATIONS

SUDEEP KUNDU† AND KARL KUNISCH*

Abstract. Convergence of the bilevel policy iteration algorithm for the value function, satisfying the Hamilton-Jacobi-Isaacs (HJI) equation, is analyzed in the presence of control constraints. Both first and second-order HJI equations corresponding to the deterministic and stochastic systems are considered. For numerical tests, a semi-implicit upwind scheme for backward HJI PDEs is applied and a thorough comparison between the solutions to first and second-order Hamilton-Jacobi-Bellman (HJB) equations is presented for both the unconstrained and the constrained control cases.

Keywords: Optimal feedback control, $\mathcal{H}_2/\mathcal{H}_\infty$ control synthesis, Hamilton-Jacobi-Bellman equation, Isaacs' equation, policy iteration, upwind scheme.

AMS classification: 49J20, 49L20, 49N35, 93B52, 93B36.

1. Introduction

The Dynamic Programming Principle (DPP) is an important tool in optimal control theory since it opens the door for feedback or closed loop control. It evolves around the value function, which is the viscosity solution of a Hamilton-Jacobi (HJ) type partial differential equation (PDE) [7]. In open-loop control, the control depends only on time and the initial state. It is not given as a function of the state at time t. This has several disadvantages. If, for example, the initial state changes, the optimal control has to be recomputed from scratch. Also, open loop control is not designed to handle modeling errors or exogenous disturbances, or other variations of system parameters. In this scenario, feedback or closed-loop control has several advantages, since it explicitly depends on the state of the system. Moreover, robust optimal feedback control is essential in many scenarios of practical interest, including, for instance, robotics and flight control, due to its enhanced disturbance rejection properties. $\mathcal{H}_\infty$ robust control [22] - the name originated from the Hardy space of analytic functions bounded in the right half-plane - is a min-max optimization technique, also known as convex-concave games, which is the focus in this paper.

In the case of linear systems with quadratic cost, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems can be solved on the basis of the regulator or the game-theoretic Riccati equations. For nonlinear systems or non-quadratic cost, such complete solution strategies are not available. Linear-quadratic approximations may prove effective in a local analysis for certain cases, but that may step short in others since critical information of the nonlinearity gets ignored. In this paper, we are interested in directly controlling the nonlinear system. When there is only one player (control

†Department of Mathematical Sciences, Rajiv Gandhi Institute of Petroleum Technology, Jais-229304, Uttar Pradesh, India, (sudeep.kundu@rgipt.ac.in).

*Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria and Radon Institute for Computational and Applied Mathematics (RICAM), Altenbergerstraße 69, A-4040 Linz, Austria, (karl.kunisch@uni-graz.at).

or input) the optimal value function satisfies the Hamilton-Jacobi-Bellman (HJB) equation related to the $\mathcal{H}_2$ control problem. Two-person zero-sum noncooperative differential games, where one player or agent (control) wants to minimize the cost or payoff function and another player (disturbance) wants to maximize the cost, are closely related to the $\mathcal{H}_\infty$ control which in turn satisfies the Hamilton Jacobi Isaacs (HJI) equation [44]. Regarding details on $\mathcal{H}_\infty$ control, its relation to the Isaacs equation, and to two-player zero-sum differential game theory, we refer to [45, 46, 8]. Solving HJ equations is frequently addressed by successive approximation or policy iteration, which are essentially equivalent to applying a Newton iteration. In the case of infinite horizon or stabilization problems, a proper initialization in the form of a stabilizing control is required. Whenever an initial stabilizing control is not available, it is common to use a discount factor in the cost. Once it is solved, we obtain the control in the HJB case, and the control as well as the disturbance in the HJI case. For differentiability of the value function, some notable results are given in [25] for linear systems and in [17] for nonlinear systems.

A variety of numerical techniques are available in the literature to solve for HJB or HJI equations, including [19]. Policy iteration, mentioned already above, for the HJB equation was developed by several authors [4, 5, 10, 9] for unconstrained control and in [1, 30] for the case with control constraints. While in [30] the constraints are realized as hard constraints, they are approximated via non-quadratic penalty functionals in [1] and earlier in [37]. In [3] numerical techniques related to finite difference schemes are discussed. Other techniques include semi-Lagrangian schemes [23], domain decomposition methods [16], penalty approach with finite difference methods [49], the finite element methods [26], and level set methods [40]. Policy iteration for the HJI equation has likely first been addressed in [10]. More recent results include polynomial approximation methods [28, 29], tensor calculus techniques [6, 21] and deep neural network techniques [33, 20], reinforcement learning techniques [11], graph tree structures [5], sparse grids [24] and splitting methods [36].

HJI equations arise in stochastic differential games, providing the value function for zero-sum settings. In systems like the stochastic Lorenz equation and Van der Pol oscillator, they capture the minimax interaction between stabilizing control and destabilizing disturbances. This framework explains uncertainty in chaotic, noisy, and nonlinear dynamics, with applications ranging from engineering structures and secure communications to biological networks, finance, and climate systems. Another typical setting models a central agent (leader) influencing many followers, each striving to remain close while subject to random disturbances [50]. For stochastic systems, $\mathcal{H}_\infty$ control was discussed in [51] both for the finite and infinite horizon problems. Policy iteration, also known as Howard's algorithm, in the stochastic context, was considered in [48] for the unconstrained control case and in [30] for the case with constraints. Neural network based policy iteration for the second-order HJI equations was presented recently in [27]. A deep neural network technique to solve zero-sum stochastic differential games was applied in [41]. For second-order linear HJB, tensor calculus techniques were considered in [47]. The Markov chain approximation method was used in [13] to solve the stochastic HJB equation in conjunction with policy iteration and the finite difference method. Finally, we refer to the monograph [34] on numerical methods for stochastic control problems.

In this article, we address selected aspects concerning the convergence of policy iteration related to two-player zero-sum differential games, in the presence of control. In the unconstrained control case, this was addressed in [10] for deterministic systems. In [2] control constraints are realized by a penalty technique. Here, we employ a projection operator to strictly enforce convex control constraints. We show convergence of the value function for both first and second-order HJI equations. With a semi-implicit upwind scheme, we solve the HJI equation; and with

an application of the bisection algorithm, we obtain the smallest value of the disturbance attenuation level, to be defined below. We test our results for the controlled Van der Pol and Lorenz systems and solve them for both the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ formulations. A detailed comparison between the HJB and HJI synthesis in terms of the performance index, the evolution of the state, and the control trajectory is provided. To the best of our knowledge, such an analysis is not available in the literature for HJI equations with hard constraints.

The rest of the paper is divided as follows. Section 2 is dedicated to the convergence of the policy iteration algorithm for the first order HJI equation under input constraints. Section 3 presents the corresponding results for the stochastic case. Section 4 contains the numerical results, and Section 5 provides some concluding remarks.

2. Nonlinear $\mathcal{H}_\infty$ control for deterministic systems

We consider the following infinite horizon optimal robust control problem:

$$\min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u,w;x) := \int_0^\infty \Big(\ell(y(t)) + \|u(t)\|_R^2 - \gamma^2\|w(t)\|_P^2\Big)\,dt, \tag{2.1}$$

subject to the deterministic system

$$\dot{y}(t) = f(y(t)) + g(y(t))u(t) + h(y(t))w(t)\,, \quad y(0) = x, \tag{2.2}$$

where $y(t) = (y_1(t),\dots,y_d(t))^t \in \mathbb{R}^d$ is the state vector, $w(\cdot)\in\mathcal{W}$ is the disturbance signal with $\mathcal{W} = \mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^q)$, $u(\cdot)\in\mathcal{U}$ is the control input with $\mathcal{U} = \mathcal{L}^2(\mathbb{R}_+;U)$, where $U\subset\mathbb{R}^m$ is defined later. Further $\ell(y) > 0$, $y \neq 0$ is the state running cost, with $\ell(0) = 0$, $\|w\|_P^2 = w^tPw$, with $P\in\mathbb{R}^{q\times q}$, $P > 0$ a positive definite matrix. The disturbance attenuation level $\gamma > 0$ is known as the measure of robustness, the control penalization term $\|u\|_R^2 = u^tRu$ with $R\in\mathbb{R}^{m\times m}$, $R > 0$ is a positive definite matrix. We assume that the system dynamics $f(y) : \mathbb{R}^d \to \mathbb{R}^d$ with $f(0) = 0$, $h(y) : \mathbb{R}^d \to \mathbb{R}^{d\times q}$, and $g(y) : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are uniformly Lipschitz continuous on $\mathbb{R}^d$, and that g and h are globally bounded. Since $f(0) = 0$ the state $y = 0$ is an equilibrium of the above system for $u = w = 0$. Throughout, the initial states will be chosen from a fixed domain Ω containing the origin in its interior.

We assume that the control lies in a closed convex set U containing the origin in $\mathbb{R}^m$. As a particular example, U can be chosen as

$$U = \{u = (u_1,\dots,u_m) \mid \alpha_i \le u_i \le \beta_i, \quad i = 1,\dots,m\}, \tag{2.3}$$

where $\alpha = (\alpha_1,\dots,\alpha_m)\in\mathbb{R}^m$, $\beta = (\beta_1,\dots,\beta_m)\in\mathbb{R}^m$ and $\alpha_i \le 0$, $\beta_i \ge 0$.

In the time domain, the goal of $\mathcal{H}_\infty$ control consists in defining a feedback control law $u(t) = u(y(t))$ such that the system (2.2) is closed loop asymptotically stable and has $\mathcal{L}_2$ gain less than or equal to γ, i.e.

$$\int_0^T \Big(\ell(y(t)) + \|u(t)\|_R^2\Big)\,dt \le \gamma^2 \int_0^T \|w(t)\|_P^2\,dt, \tag{2.4}$$

for all $w\in\mathcal{L}^2(0,T;\mathbb{R}^q)$, and all $T\ge 0$, where y the solution of (2.2) with $x = 0$. The $\mathcal{H}_\infty$ norm γ^* is defined as the smallest value of γ such that for $\gamma\ge\gamma^*$, (2.4) holds under system (2.2). The optimal value function associated to (2.1), also known as available storage function, is defined by

$$V_\gamma(x) = \min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u,w;x).$$

Throughout we assume that:

Assumption 1. *The radially unbounded value function satisfies $V_\gamma \in C^1(\mathbb{R}^d)$. Further, for the policy iteration Algorithm 1, we assume that $V_\gamma^{(i,j)}(x) \in C^1(\mathbb{R}^d)$ for each $i,\ j \geq 0$.*

Remark 1. The regularity assumption for V_γ , $V_\gamma^{(i,j)}$ is common in the literature for related analysis. Similar assumptions were made earlier in [42, 44, 10, 1, 2, 24]. The investigation of the regularity of the value function is certainly an important research topic in its own right, see also [15] and [32]. In passing we note that the upper and lower solutions to the HJI equation (2.8) are known to be Lipschitz continuous and hence by Rademacher's theorem they are differentiable almost everywhere [39, Corollary 3.8].

Due to dynamic programming principle, under the Assumption 1, the optimal value function, $V_\gamma(x)$ satisfies the HJI equation [8, Chapter 8],

$$\min_{u\in U}\ \max_{w\in W}\{\nabla V_\gamma(x)^t(f(x)+g(x)u+h(x)w)+\ell(x)+\|u\|_R^2-\gamma^2\|w\|_P^2\}=0\,, \tag{2.5}$$

where $\nabla V_\gamma(x) = (\partial_{x_1}V_\gamma,\dots,\partial_{x_d}V_\gamma)^t$. Note that since the cost (2.1) and the system (2.2) are separable with respect to control and disturbance strategies, the so called Isaacs' condition holds [7, Chapter VIII] which guarantees the existence of the value function satisfying (2.5).

For $\gamma \geq \gamma^*$, (2.5) has a positive definite solution [7, Appendix], [46] (here positive definite means $V_\gamma(x) > 0$ for $x \neq 0$ and $V_\gamma(0) = 0$). If the initial state is not an equilibrium solution, then instead of (2.4) we get

$$\int_0^T \Big(\ell(y(t))+\|u(t)\|_R^2\Big)\,dt \leq \gamma^2\int_0^T \|w(t)\|_P^2\ dt + V_\gamma(x), \tag{2.6}$$

see e.g. [7, Appendix], [44].

Under Assumption 1, according to (2.5), the optimal control u is given in feedback strategic form by

$$u_\gamma^*(x) = \mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\Big),$$

where $\mathcal{P}_U$ is the orthogonal projection in the R-weighted inner product on $\mathbb{R}^m$ onto U, and V_γ is the corresponding HJI value function. For the particular case (2.3), the projection $\mathcal{P}_U$ is given by $u_\gamma^*(x) = \min\Big\{\beta, \max\{\alpha, -\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\}\Big\}$, where the min - and max - operations are considered coordinate wise. An optimal disturbance $w_\gamma \in \mathbb{R}^q$ is given in feedback strategic form by $w_\gamma^* = \frac{1}{2\gamma^2}P^{-1}h^t\nabla V_\gamma(x)$.

When there is no confusion we denote $(u_\gamma, w_\gamma) = (u_\gamma^*, w_\gamma^*)$. Moreover, in many places, we suppress the dependence of f , g and h on the state variable.

Using the values of u_γ and w_γ, the HJI equation (2.5) becomes

$$\begin{aligned}\nabla V_\gamma(x)^t\Big(f(x)+g(x)\mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\Big)\Big)+\frac{1}{4\gamma^2}\nabla V_\gamma(x)^t h(x)P^{-1}h(x)^t\nabla V_\gamma(x)\\ +\ell(x)+\Big\|\mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\Big)\Big\|_R^2=0\,.\end{aligned} \tag{2.7}$$

For an unconstrained control $u_\gamma \in \mathbb{R}^m$, the corresponding HJI equation reduces to

$$\nabla V_\gamma(x)^t f(x)+\frac{1}{4}\nabla V_\gamma(x)^t\left(\frac{1}{\gamma^2}h(x)P^{-1}h(x)^t-g(x)R^{-1}g(x)^t\right)\nabla V_\gamma(x)+\ell(x)=0\,. \tag{2.8}$$

Sufficient conditions for the local existence of a solution $V_\gamma \geq 0$ of (2.8) are given in [43]. When $\gamma \to \infty$ in the disturbance law, the HJI equations (2.7) and (2.8) result in the constrained and unconstrained HJB equations related to $\mathcal{H}_2$ control. Related details are given in [30]. The case with constraints is treated in the Appendix of [7], for example. In [7, Appendix B, Theorem 4.5] regularity conditions on the problem data are provided under which the existence of (viscosity) solutions to (2.7) implies the solvability of (2.1) and vice-versa. Here (strict)solvability refers to the property that the unperturbed system (2.2) is (asymptotically) stable and the L^2-gain property (2.6) holds [7, Appendix B.1, Theorem 4.5].

Throughout the paper, it is assumed that $\bar{\Omega}$ is contained in the region of stability related to (2.2) with $w = 0$, i.e. the region of initial conditions for which trajectories are asymptotically stabilizable to the origin. In certain cases, additional condition on initial data in Ω will be specified.

Solving (2.8) is a challenging task, and policy iteration is one possible approach. Alternative methods include polynomial approximations to the HJI equation [29] where the case without constraints has been discussed. The policy iteration or successive approximation procedure for solving the HJI equation (2.7) is presented in Algorithm 1 for a fixed value γ. Finding the smallest value of γ i.e. γ^* is an independent task which is typically achieved by a bisection algorithm in numerical practice, see e.g. [14]. By applying the bilevel policy iteration Algorithm 1, which can also be considered as a Newton iteration, equation (2.7) is reduced to solving a sequence of linear HJI equations (2.9), referred to as Generalized HJI (GHJI) equations.

Algorithm 1 Continuous Policy Iteration for HJI equations

Input: Let $u_\gamma^{(0)}(x)$ be an asymptotically stabilizing control law for the dynamics (2.2) with $w_\gamma = 0$ and $\gamma \geq \gamma^*$.

For $i = 0$ **to** ∞

Set $w_\gamma^{(i,0)}(x) \equiv 0$**,**

For $j = 0$ **to** ∞

Solve for $V_\gamma^{(i,j)}(x) \in C^1(\mathbb{R}^d)$ **:**

$$\begin{cases} \nabla V_\gamma^{(i,j)}(x)^t\big(f(x) + g(x)u_\gamma^{(i)} + h(x)w_\gamma^{(i,j)}\big) + \ell(x) + \|u_\gamma^{(i)}\|_R^2 - \gamma^2\|w_\gamma^{(i,j)}\|_P^2 = 0\,, \\ V_\gamma^{(i,j)}(0) = 0. \end{cases} \tag{2.9}$$

Update the disturbance:

$$w_\gamma^{(i,j+1)}(x) = \frac{1}{2\gamma^2}P^{-1}h(x)^t\nabla V_\gamma^{(i,j)}(x)\,, \tag{2.10}$$

j=j+1

End j loop

Update the control:

$$u_\gamma^{(i+1)}(x) = \mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma^{(i,\infty)}(x)\Big). \tag{2.11}$$

i=i+1

End i loop

Remark 2. In Algorithm 1 we tacitly assumed that the inner loop converges. Sufficient conditions for this to occur will be given in Theorem 2.1 and Remark 5 below.

Definition 1 (Asymptotic Stability in the sense of Lyapunov)**.** *The equilibrium solution* $y = 0$ *of* (2.2) *is locally asymptotically stable if:*

(1) *It is stable in the sense of Lyapunov i.e., for every $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ such that, if $\|x\| < \delta$, then $\|y(t)\| < \epsilon$ for all $t > 0$.*
(2) *There exists a $\delta' > 0$ such that, if $\|x\| < \delta'$, then $y(t) \to 0$ as $t \to \infty$.*

The equilibrium state $y = 0$ of (2.2) is globally asymptotically stable if it is asymptotically stable for any initial condition, meaning that the condition for asymptotic stability holds for any $\delta' > 0$.

Remark 3. In Algorithm 1 it is further assumed that $\gamma \geq \gamma^*$. There are no effective analytical methods to characterize the value γ^*, [7, page 509]. In numerical practice, γ is chosen sufficiently large at first and then reduced by a bisection algorithm, as in [14].

Next we establish a monotonicity result for the inner loop. For this purpose we define for $x \in \Omega$ and given a stabilizing control $u_\gamma^{(i)}$

$$\begin{cases} V_\gamma^{(i)}(x) := \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u_\gamma^{(i)}, w; x) \\ \dot{y}(t) = f(y(t)) + g(y(t))u_\gamma^{(i)}(y(t)) + h(y(t))w(t)\,, \quad y(0) = x. \end{cases} \tag{2.12}$$

Theorem 2.1. *If the system*

$$\dot{y}(t) = f(y(t)) + g(y(t))u_\gamma^{(i)}(t) + h(y(t))w_\gamma^{(i,j)}(t)\,, \quad y(0) = x, \tag{2.13}$$

is asymptotically stable in the sense of Lyapunov on Ω for all pairs (i, j), where $i \geq 0$ is fixed, and all $j \geq 0$, and $V_\gamma^{(i,j)} \in C^1(\mathbb{R}^d)$, then we obtain $0 \leq V_\gamma^{(i,j)}(x) \leq V_\gamma^{(i,j+1)}(x) \leq V_\gamma^{(i)}(x)$, $\forall x \in \Omega$. Moreover $V_\gamma^{(i,j)}(x) > 0$ for $x \neq 0$.

Proof. Since $\dot{y} = f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j)}$ is asymptotically stable for each pair (i, j) and utilizing that $V_\gamma^{(i,j)}(0) = 0$, the difference between $V_\gamma^{(i,j)}(x)$ and $V_\gamma^{(i,j+1)}(x)$ can be obtained by calculating the time derivative of $V_\gamma^{(i,j)}$, $V_\gamma^{(i,j+1)}$ along the trajectory of

$$\dot{y} = \Big(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j+1)}\Big),\ y(0) = x \in \Omega, \tag{2.14}$$

to obtain

$$\begin{aligned} &V_\gamma^{(i,j)}(x) - V_\gamma^{(i,j+1)}(x) \\ &= \int_0^\infty \Bigg(\Big(\nabla V_\gamma^{(i,j+1)}(y)\Big)^t \big(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j+1)}\big)\Big) \\ &\qquad - \Big(\nabla V_\gamma^{(i,j)}(y)\Big)^t \big(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j+1)}\big)\Big)\Bigg)\, dt, \end{aligned}$$

where in y, $u_\gamma^{(i)} = u_\gamma^{(i)}(y)$, and $w_\gamma^{(i,j+1)} = w_\gamma^{(i,j+1)}(y)$ the dependence on t is suppressed. From the GHJI equation (2.9), it follows that

$$\nabla V_\gamma^{(i,j+1)}(y)^t \Big(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j+1)}\Big) = -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j+1)}\right\|_P^2\Big), \tag{2.15}$$

and, suppressing the dependence of f, g and h on y,

$$\begin{aligned}\nabla V_\gamma^{(i,j)^t}&\Big(f+gu_\gamma^{(i)}+hw_\gamma^{(i,j+1)}\Big)\\ &=\nabla V_\gamma^{(i,j)^t}\Big(f+gu_\gamma^{(i)}+hw_\gamma^{(i,j)}\Big)+\langle h^t\nabla V_\gamma^{(i,j)},w_\gamma^{(i,j+1)}-w_\gamma^{(i,j)}\rangle\\ &=-\Big(\ell(y)+\left\|u_\gamma^{(i)}\right\|_R^2-\gamma^2\left\|w_\gamma^{(i,j)}\right\|_P^2\Big)+\langle 2\gamma^2Pw_\gamma^{(i,j+1)},w_\gamma^{(i,j+1)}-w_\gamma^{(i,j)}\rangle,\end{aligned}\tag{2.16}$$

where $\nabla V_\gamma^{(i,j)}$ is calculated along y. Subtracting (2.16) from (2.15) we obtain

$$V_\gamma^{(i,j)}(x)-V_\gamma^{(i,j+1)}(x)=-\int_0^\infty\gamma^2\left\|w_\gamma^{(i,j)}-w_\gamma^{(i,j+1)}\right\|_P^2\,dt.$$

Hence we get $V_\gamma^{(i,j)}(x)\le V_\gamma^{(i,j+1)}(x)$, and the desired monotonicity is established.

To obtain the upper bound, let $V_\gamma^{(i)}$ be as defined in (2.12). Calling on (2.9) we obtain

$$V_\gamma^{(i,j)}(x)=\int_0^\infty\left(\ell(y(t))+\|u_\gamma^{(i)}(y(t))\|_R^2-\gamma^2\|w_\gamma^{(i,j)}(y(t))\|_P^2\right)dt=\mathcal{J}(u_\gamma^{(i)},w_\gamma^{(i,j)};x),$$

where $y=y(u_\gamma^{(i)},w_\gamma^{(i,j)})$. Thus, we have $V_\gamma^{(i,j)}(x)\le V_\gamma^{(i)}(x)$ as desired. In an analogous way, utilizing that $w_\gamma^{(i,0)}=0$ we find

$$V_\gamma^{(i,0)}(x)=\int_0^\infty\left(\ell(y(t))+\|u_\gamma^{(i)}(y(t))\|_R^2\right)dt\ge 0,$$

where $y=y(u_\gamma^{(i)},w=0)$. Moreover $V_\gamma^{(i,0)}(x)>0$ for $x\ne 0$. □

Remark 4. Since the sequence $V_\gamma^{(i,j)}(x)$ is monotonically increasing and bounded above for each $x\in\Omega$, it follows that there exists a function denoted by $V_\gamma^{(i,\infty)}$ such that $\lim_{j\to\infty}V_\gamma^{(i,j)}(x)=V_\gamma^{(i,\infty)}(x)$ for every $x\in\Omega$.

If Ω is bounded and $V_\gamma^{(i,\infty)}(x)$ is continuous on $\bar\Omega$ then the convergence is uniform by Dini's theorem. Alternatively, if $V_\gamma^{(i,0)}$ is uniformly bounded from below on Ω, then by Lebesgue's monotone convergence theorem $\{V_\gamma^{(i,j)}\}_{j=1}^\infty$ converges in $L^1(\Omega)$ to $V_\gamma^{(i,\infty)}$.

Remark 5. We can show that $\{w_\gamma^{(i,j)}\}$ converges to $w_\gamma^{(i,\infty)}$ on Ω under the following assumptions. If Ω is bounded and $\{V_\gamma^{(i,j)}\}\in C^1(\Omega)$ satisfy (2.9), $V_\gamma^{(i,\infty)}\in C^1(\Omega)\cap C(\bar\Omega)$, and $\{\nabla V_\gamma^{(i,j)}\}$ is equicontinuous in Ω, then it can be shown that $\{\nabla V_\gamma^{(i,j)}\}_{j=1}^\infty$ converges pointwise to $\nabla V_\gamma^{(i,\infty)}$, $\{w_\gamma^{(i,j)}\}$ converges to $w_\gamma^{(i,\infty)}=\frac{1}{2\gamma^2}P^{-1}h(x)^t\nabla V_\gamma^{(i,\infty)}(x)$, and $V_\gamma^{(i,\infty)}(x)$ solves the HJI equation

$$\nabla V_\gamma^{(i,\infty)}(x)^t\big(f+gu_\gamma^{(i)}\big)+\frac{1}{4\gamma^2}\nabla V_\gamma^{(i,\infty)}(x)^thP^{-1}h^t\nabla V_\gamma^{(i,\infty)}(x)+\left\|u_\gamma^{(i)}\right\|_R^2+\ell(x)=0.\tag{2.17}$$

This can be argued similarly as Proposition 2 in [30]. Conversely, conditions for the existence of a solution $V_\gamma^{(i,\infty)}$ to (2.17), are given in [44], for example.

Next we provide sufficient conditions for the monotonicity for the sequence $V_\gamma^{(i,\infty)}$ which was introduced in Remark 2.

Theorem 2.2. *Let $V_\gamma^{(i,\infty)}\in C^1(\mathbb{R}^d)$ satisfy (2.17) on $\mathbb{R}^d$, and assume that*

$$\dot y=f(y)+gu_\gamma^{(i)}+hw_\gamma^{(i,\infty)},\quad y(0)=x,\tag{2.18}$$

is asymptotically stable on Ω for each i. Then we have $0 \le V_\gamma^{(i+1,\infty)}(x) \le V_\gamma^{(i,\infty)}(x)$ $\forall x \in \Omega$.

Remark 6. As a consequence of the previous theorem, the pointwise convergence of $V_\gamma^{(i,\infty)}$ follows. The reason why the regularity of $V_\gamma^{(i,\infty)}$ and equation (2.17) are assumed on all of $\mathbb{R}^d$ relates to the fact that the forced trajectories emanating from some $x \in \Omega$ need not remain in Ω in general. It could be an interesting issue of further work to provide conditions on the problem data which imply a-priori bounds on these forced trajectories.

Proof. Since (2.18) is asymptotically stable for each i, along the trajectory $\dot{y} = f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)}$, we obtain the difference between $V_\gamma^{(i+1,\infty)}(x)$ and $V_\gamma^{(i,\infty)}(x)$ as

$$\begin{aligned}
&V_\gamma^{(i+1,\infty)}(x) - V_\gamma^{(i,\infty)}(x)\\
&= \int_0^\infty \left(\left(\nabla V_\gamma^{(i,\infty)^t} (f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)}) \right) - \left(\nabla V_\gamma^{(i+1,\infty)^t} (f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)}) \right) \right) dt.
\end{aligned}$$

From the HJI equation (2.17), it follows that for $i \ge -1$

$$\nabla V_\gamma^{(i+1,\infty)^t} \left(f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)} \right) = -\left(\ell(y) + \left\| u_\gamma^{(i+1)} \right\|_R^2 - \gamma^2 \left\| w_\gamma^{(i+1,\infty)} \right\|_P^2 \right). \tag{2.19}$$

To estimate $\nabla V_\gamma^{(i,\infty)^t} (f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)})$, we rewrite it as

$$\begin{aligned}
&\nabla V_\gamma^{(i,\infty)}(y)^t \left(f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)} \right)\\
&\qquad = \nabla V_\gamma^{(i,\infty)}(y)^t \left(f + gu_\gamma^{(i)} + hw_\gamma^{(i,\infty)} \right) + \langle h^t \nabla V_\gamma^{(i,\infty)}(y), w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)} \rangle\\
&\qquad\quad + \langle g^t \nabla V_\gamma^{(i,\infty)}(y), u_\gamma^{(i+1)} - u_\gamma^{(i)} \rangle.
\end{aligned}$$

With the notation of $z = \frac{1}{2} R^{-1} g^t \nabla V_\gamma^{(i,\infty)}$ we obtain $g^t \nabla V_\gamma^{(i,\infty)} = 2Rz$ and $u_\gamma^{(i+1)} = \mathcal{P}_U(-z)$. Also from the worst disturbance, it follows that $h^t \nabla V_\gamma^{(i,\infty)} = 2\gamma^2 P w_\gamma^{(i,\infty)}$. Altogether, we have

$$\begin{aligned}
&\nabla V_\gamma^{(i,\infty)}(y)^t \left(f(y) + g(y) u_\gamma^{(i+1)}(y) + h(y) w_\gamma^{(i+1,\infty)}(y) \right)\\
&= -\ell(y) - \gamma^2 \left\| w_\gamma^{(i,\infty)} \right\|_P^2 + 2\gamma^2 P \langle w_\gamma^{(i,\infty)}, w_\gamma^{(i+1,\infty)} \rangle\\
&\quad - \left\| u_\gamma^{(i)} \right\|_R^2 + \langle 2Rz, \mathcal{P}_U(-z) - u_\gamma^{(i)} \rangle\\
&= -\ell(y) + \gamma^2 \left\| w_\gamma^{(i+1,\infty)} \right\|_P^2 - \gamma^2 \left\| w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)} \right\|_P^2 - \| \mathcal{P}_U(-z) \|_R^2\\
&\quad - \| u_\gamma^{(i)}(y) - \mathcal{P}_U(-z) \|_R^2 + 2(z + \mathcal{P}_U(-z))^t R (\mathcal{P}_U(-z) - u_\gamma^{(i)}(y)).
\end{aligned} \tag{2.20}$$

With these expressions we obtain

$$\begin{aligned}
&V_\gamma^{(i+1,\infty)}(x) - V_\gamma^{(i,\infty)}(x)\\
&= \int_0^\infty \Bigg(\Big(- \| u_\gamma^{(i)}(y) - \mathcal{P}_U(-z) \|_R^2 - \gamma^2 \left\| w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)} \right\|_P^2\\
&\qquad + 2(z + \mathcal{P}_U(-z))^t R (\mathcal{P}_U(-z) - u_\gamma^{(i)}(y)) \Big) \Bigg) dt.
\end{aligned}$$

Since $(z + \mathcal{P}_U(-z))^t R (\mathcal{P}_U(-z) - u_\gamma^{(i)}(y)) \le 0$, we get $V_\gamma^{(i+1,\infty)}(x) \le V_\gamma^{(i,\infty)}(x)$.

The lower bound $0 \le V_\gamma^{(i,\infty)}(x)$ is a consequence of Theorem 2.1. □

Remark 7. If Ω is bounded and $\{V_\gamma^{(i,\infty)}\} \in C^1(\Omega)$ satisfy (2.17), $V_\gamma \in C^1(\Omega) \cap C(\bar{\Omega})$, and $\{\nabla V_\gamma^{(i,\infty)}\}$ is equicontinuous in Ω, then the pair $(u_\gamma^{(i)}, V_\gamma^{(i,\infty)})$ of control and value function converges to $\Big(u^* = \mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\Big), V_\gamma(x)\Big)$, where $V_\gamma(x)$ solves the HJI equation (2.7). For a proof, we refer to [30].

At last we show the asymptotic stability property of $u_\gamma^{(i)}$ on Ω for the unperturbed trajectory at each iteration level. Note that $i = 0$ is assumed in the presentation of Algorithm 1.

Lemma 1. *If $V_\gamma^{(i,\infty)} \in C^1(\mathbb{R}^d)$ for $i \ge 0$ and (2.17) holds, then function $u_\gamma^{(i+1)}$ is an asymptotically stabilizing control on Ω for the unperturbed system $\dot{y} = f(y) + gu$.*

Proof. We show that $V_\gamma^{(i,\infty)}$ is a strict Lyapunov function for the system (2.2) with $w = 0$. By Theorem 2.1 the function $V_\gamma^{(i,\infty)}$ is positive definite. Next we take the time derivative of $t \to V_\gamma^{(i,\infty)}(y)(t)$ along the trajectory $\dot{y} = f(y) + gu_\gamma^{(i+1)}$. It can be expressed as

$$\begin{aligned}\frac{d}{dt}V_\gamma^{(i,\infty)}(y(t)) &= \nabla V_\gamma^{(i,\infty)}(y)^t\Big(f + gu_\gamma^{(i+1)}\Big)\\ &= \nabla V_\gamma^{(i,\infty)}(y)^t\Big(f + gu_\gamma^{(i)}\Big) + \langle g^t\nabla V_\gamma^{(i,\infty)}(y), u_\gamma^{(i+1)} - u_\gamma^{(i)}\rangle.\end{aligned}$$

Using (2.17), we obtain

$$\begin{aligned}\frac{d}{dt}V_\gamma^{(i,\infty)}(y(t)) = -\ell(y) - \frac{1}{4\gamma^2}\nabla V_\gamma^{(i,\infty)}(y)^t hP^{-1}h^t\nabla V_\gamma^{(i,\infty)}(y) - \Big\|u_\gamma^{(i)}\Big\|_R^2\\ + \langle g^t\nabla V_\gamma^{(i,\infty)}(y), u_\gamma^{(i+1)} - u_\gamma^{(i)}\rangle.\end{aligned} \tag{2.21}$$

Let $z = \frac{1}{2}R^{-1}g^t\nabla V_\gamma^{(i,\infty)}$, so that we have $g^t\nabla V_\gamma^{(i,\infty)} = 2Rz$. Consequently, from the control law (2.11) we obtain $u_\gamma^{(i+1)} = \mathcal{P}_U(-z)$. Hence, from (2.21) we obtain

$$\begin{aligned}\frac{d}{dt}V_\gamma^{(i,\infty)}(y(t)) &= \nabla V_\gamma^{(i,\infty)}(y)^t\Big(f + gu_\gamma^{(i+1)}\Big)\\ &= -\ell(y) - \frac{1}{4\gamma^2}\nabla V_\gamma^{(i,\infty)}(y)^t hP^{-1}h^t\nabla V_\gamma^{(i,\infty)}(y) - \|\mathcal{P}_U(-z)\|_R^2\\ &\quad - \|u_\gamma^{(i)}(y) - \mathcal{P}_U(-z)\|_R^2 + 2(z + \mathcal{P}_U(-z))^t R(\mathcal{P}_U(-z) - u_\gamma^{(i)}(y)).\end{aligned}$$

As $u_\gamma^{(i)}(y) \in U$, it follows that $(z + \mathcal{P}_U(-z))^t R(\mathcal{P}_U(-z) - u_\gamma^{(i)}(y)) \le 0$, and therefore we obtain finally $\frac{d}{dt}V_\gamma^{(i,\infty)}(y(t)) = \nabla V_\gamma^{(i,\infty)}(y)^t\Big(f + gu_\gamma^{(i+1)}\Big) < 0$, and the asymptotic stabilization property of the control follows. □

3. $\mathcal{H}_\infty$ FEEDBACK CONTROL FOR STOCHASTIC SYSTEMS

In this section, we consider the following infinite horizon stochastic optimal robust control problem:

$$\min_{u(\cdot)\in\mathcal{U}}\ \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u, w; x) := \mathbb{E}\int_0^\infty \Big(\ell(y(t)) + \|u(t)\|_R^2 - \gamma^2\|w(t)\|_P^2\Big)\,dt, \tag{3.1}$$

subject to the nonlinear stochastic system

$$dy = \Big(f(y) + g(y)u(t) + h(y)w(t)\Big)dt + \Big(g_1(y) + g_2(y)w(t)\Big)dW\,, \quad y(0) = x, \tag{3.2}$$

where $y(t) = (y_1(t), \dots, y_d(t))^t \in \mathbb{R}^d$ is the state vector, $W(t) \in \mathbb{R}^k$ is a standard k-dimensional Wiener process (i.e., with independent components) defined on a complete probability space. The exogenous disturbance $w(\cdot) \in \mathcal{W} = \mathcal{L}^2_{\mathcal{F}}(\mathbb{R}_+; \mathbb{R}^q)$ and the control input $u(\cdot) \in \mathcal{U} = \mathcal{L}^2_{\mathcal{F}}(\mathbb{R}_+; U)$, where the subscript $\mathcal{F}$ indicates that the processes are adapted to the natural filtration $\{\mathcal{F}_t\}$ generated by W and square integrable, $\mathbb{E}\int_0^\infty \|\cdot\|^2\, dt < \infty$, see [51]. Since we are interested in closed loop control, we note that the strong solution y of (3.2) is $\{\mathcal{F}_t\}$-adapted, so that feedback laws $u(y(t))$, $w(y(t))$ are admissible. Further U is a closed convex set containing the origin.

Moreover, for the dynamics, which are characterized by f, h, g (mentioned on page 3), $g_1(y) : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ and $g_2(y) = [g_2^1(y), \dots, g_2^q(y)]$ with $g_2^l(y) : \mathbb{R}^d \to \mathbb{R}^{d\times k}$, where $g_2(y)w := \sum_{l=1}^q w_l\, g_2^l(y)$ for $w \in \mathbb{R}^q$, it is assumed that these functions are Lipschitz continuous on $\mathbb{R}^d$ and that g, h, g_1 and g_2 are globally bounded. The analysis is carried out over a bounded domain $\Omega \subset \mathbb{R}^d$ containing the origin. Concerning the existence and uniqueness results for (3.2), we refer to e.g. [38, Chapter 2]. Throughout we assume that $f(0) = 0$, $g_1(0) = 0$ for $w = 0$ and $u = 0$.

We want to find a control $u \in \mathcal{L}^2_{\mathcal{F}}((0,T), U)$ such that the stochastic system (3.2) is closed loop asymptotically stable in probability and has an $\mathcal{L}_2$-gain less than or equal to the disturbance attenuation level $\gamma > 0$, i.e. for all $T \geq 0$ and $w \in \mathcal{L}^2_{\mathcal{F}}((0,T), \mathbb{R}^q)$,

$$\mathbb{E}\int_0^T \Big(\ell(y) + \|u(t)\|_R^2\Big)\, dt \leq \gamma^2\, \mathbb{E}\int_0^T \|w(t)\|_P^2\; dt, \tag{3.3}$$

where y is the solution to (3.2) with $x = 0$. We define the value function

$$V_\gamma(x) = \min_{u(\cdot)\in\mathcal{U}}\ \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u, w; x).$$

Here, for $V_\gamma \in C^2(\mathbb{R}^d)$, the traces involving g_2 are taken columnwise, i.e. $Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2] := \big(Tr[(g_2^l)^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2^m]\big)_{l,m=1}^q \in \mathbb{R}^{q\times q}$, $Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_1] := \big(Tr[(g_2^l)^t \frac{\partial^2 V_\gamma}{\partial x^2} g_1]\big)_{l=1}^q \in \mathbb{R}^q$ and $Tr[g_1^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2] := Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_1]^t$, and $2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2] > 0$ is understood in the sense of positive definite matrices.

Assumption 2. *The radially unbounded value function satisfies $V_\gamma \in C^2(\mathbb{R}^d)$. Further, for the policy iteration Algorithm 2, we assume that $V_\gamma^{(i,j)}(x) \in C^2(\mathbb{R}^d)$ for each i, $j \geq 0$.*

Similar assumptions were made earlier in [48, 51]. If in addition $V_\gamma \geq 0$ satisfies the HJI equation

$$\begin{cases} \nabla V_\gamma(x)^t\Big(f(x) + g(x)\mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g^t\nabla V_\gamma(x)\Big)\Big) \\ \qquad +\frac{1}{2}\Big(\nabla V_\gamma(x)^t h + Tr[g_1^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2]\Big)(2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2])^{-1}\Big(h^t\nabla V_\gamma(x) + Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_1]\Big) \\ \qquad\quad + \Big\|\mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g^t\nabla V_\gamma(x)\Big)\Big\|_R^2 + \ell(x) + \frac{1}{2}Tr[g_1^t \frac{\partial^2 V_\gamma}{\partial x^2} g_1] = 0, \\ \qquad V_\gamma(0) = 0, \quad 2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma}{\partial x^2} g_2] > 0, \end{cases} \tag{3.4}$$

then system (3.2) has a $\mathcal{L}_2$-gain less than or equal to $\gamma > 0$, with optimal control and optimal disturbance strategies given by

$$u_\gamma^*(x) = \mathcal{P}_U\Big(-\frac{1}{2}R^{-1}g(x)^t\nabla V_\gamma(x)\Big) \text{ and } w_\gamma^*(x) = \Big(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2]\Big)^{-1}\Big(h^t\nabla V_\gamma(x) + Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_1]\Big).$$

For the case without control constraints, we refer to [51, Theorem 3.1, Remark 3.1]. The proofs there can be adjusted to handle the control constrained case which we are treating. In the remainder of this section we simply write (u_γ, w_γ) for (u_γ^*, w_γ^*). Moreover, when there is no confusion, we suppress the dependence of f , g , h , g_1 and g_2 on the state variable.

The HJI equation related to unconstrained control satisfies

(3.5)

$$\begin{cases} \nabla V_\gamma(x)^t f(x) - \frac{1}{4}\nabla V_\gamma(x)^t g(x)R^{-1}g(x)^t\nabla V_\gamma(x) \\ \qquad +\frac{1}{2}\Big(\nabla V_\gamma(x)^t h + Tr[g_1^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2]\Big)(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2])^{-1}\Big(h^t\nabla V_\gamma(x) + Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_1]\Big) \\ \qquad\qquad +\ell(x) + \frac{1}{2}Tr[g_1^t\frac{\partial^2 V_\gamma}{\partial x^2}g_1] = 0\,, \\ V_\gamma(0) = 0\,, \quad 2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2] > 0. \end{cases}$$

For $h = 0 = g_2$ in (3.4)-(3.5), we obtain the corresponding constrained and unconstrained HJB equation which arise from stochastic control of dynamical systems respectively. The infinitesimal generator of (3.2) with optimal pair (u_γ, w_γ) is

$$\mathcal{L}_{(u_\gamma,w_\gamma)}V_\gamma = \nabla V_\gamma^t\Big(f + gu_\gamma + hw_\gamma\Big) + \frac{1}{2}Tr[(g_1 + g_2w_\gamma)^t\frac{\partial^2 V_\gamma}{\partial x^2}(g_1 + g_2w_\gamma)].$$

For the stochastic system (3.2), $y = 0$ is said to be stable in probability, or stochastically stable in the sense of Lyapunov, if for any $\epsilon > 0$, $\lim_{x\to 0} P(\sup_{t\geq 0}\|y(t)\| > \epsilon) = 0$. Further $y = 0$ in (3.2), is said to be globally asymptotically stable in the sense of Lyapunov in probability, if it is stable in probability and for all $x \in \mathbb{R}^d$, $P(\lim_{t\to\infty} y(t) = 0) = 1$. A sufficient condition for asymptotically stability in probability in the sense of Lyapunov is as follows: If there exists a positive definite decrescent Lyapunov function $V_\gamma \in C^2(\mathbb{R}^d)$ satisfying $\mathcal{L}_{(u_\gamma,w_\gamma)}V_\gamma < 0$ for all non-zero $y \in \mathbb{R}^d$, then the equilibrium point 0 of (3.2) is asymptotically stable in probability [35, 38]

The second order GHJI equations is of the form

(3.6)

$$\begin{cases} GHJI(V_\gamma, \nabla V_\gamma, \nabla^2 V_\gamma, u_\gamma; w_\gamma) = 0, \qquad V_\gamma(0) = 0, \quad 2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2] > 0 \quad \text{where} \\ GHJI(V_\gamma, \nabla V_\gamma, \nabla^2 V_\gamma, u_\gamma; w_\gamma) := \mathcal{L}_{(u_\gamma,w_\gamma)}V_\gamma + \Big(\ell + \|u_\gamma\|_R^2 - \gamma^2\|w_\gamma\|_P^2\Big), \end{cases}$$

and $\nabla V_\gamma(x) = (\partial_{x_1}V_\gamma, \ldots, \partial_{x_d}V_\gamma)^t$. Below we discuss policy iteration where we calculate w_γ and u_γ iteratively.

The following theorem establishes convergence related to the value function of policy iteration for the inner loop Algorithm 2. To obtain the upper bound for $V_\gamma^{(i,j)}$, similarly as in Theorem 2.1, we define for a stabilizing control $u_\gamma^{(i)}$ and $x \in \Omega$

(3.8)

$$\begin{cases} V_\gamma^{(i)}(x) := \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u_\gamma^{(i)}, w; x) \\ dy = \Big(f(y(t)) + g(y)u_\gamma^{(i)}(y(t)) + h(y)w(t)\Big)dt + \Big(g_1(y) + g_2(y)w(t)\Big)dW\,, \quad y(0) = x. \end{cases}$$

Algorithm 2 Policy iteration Algorithm

Let $u_\gamma^{(0)}$ be an asymptotically stabilizing control law for the system (3.2) with $w_\gamma \equiv 0$ and $\gamma \geq \gamma^*$.

For $i = 0$ **to** ∞

Set $w_\gamma^{(i,0)}(x) \equiv 0$**,**

For $j = 0$ **to** ∞

Solve for $V_\gamma^{(i,j)}(x) \in C^2(\mathbb{R}^d)$:

$$\begin{cases} \mathcal{L}_{(u_\gamma^{(i)}, w_\gamma^{(i,j)})} V_\gamma^{(i,j)}(x) + \left(\ell + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j)}\right\|_P^2\right) = 0\,, \\ V_\gamma^{(i,j)}(0) = 0\,, \quad 2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial x^2} g_2] > 0. \end{cases} \tag{3.7}$$

Update the Disturbance:

$$w_\gamma^{(i,j+1)}(x) = \left(2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial x^2} g_2]\right)^{-1} \left(h^t \nabla V_\gamma^{(i,j)}(x) + Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial x^2} g_1]\right),$$

j=j+1

End j loop

Update the Control:

$$u_\gamma^{(i+1)}(x) = \mathcal{P}_U\Big(- \frac{1}{2} R^{-1} g^t \nabla V_\gamma^{(i,\infty)}(x))\Big).$$

i=i+1

End i loop

Theorem 3.1. *If the stochastic system*

$$dy = \left(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j)}\right) dt + \left(g_1(y) + g_2(y)w_\gamma^{(i,j)}\right) dW, \quad y(0) = x, \tag{3.9}$$

is asymptotically stable in the sense of Lyapunov on Ω for all pairs (i, j), where $i \geq 0$ is fixed and all $j \geq 0$ and $V_\gamma^{(i,j)} \in C^2(\bar{\Omega})$, then we get $V_\gamma^{(i,j)}(x) \leq V_\gamma^{(i,j+1)}(x) \leq V_\gamma^{(i)}(x)$ for each $x \in \Omega$. Further $V_\gamma^{(i,j)}(x) > 0$ for $x \neq 0$.

Proof. We first calculate the difference $V_\gamma^{(i,j+1)} - V_\gamma^{(i,j)}$ by Itô's formula along the trajectory

$$dy = \left(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,j+1)}\right) dt + \left(g_1(y) + g_2(y)w_\gamma^{(i,j+1)}\right) dW. \tag{3.10}$$

Applying Itô's formula [35] to $V_\gamma^{(i,j+1)}(y)$ over a finite time interval $[0, T]$, we get

$$V_\gamma^{(i,j+1)}(y(T)) - V_\gamma^{(i,j+1)}(y(0)) = \int_0^T \mathcal{L}_{(u_\gamma^{(i)}, w_\gamma^{(i,j+1)})} V_\gamma^{(i,j+1)}(y)\, dt + M_T\,, \tag{3.11}$$

where M_T is the stochastic integral term

$$M_T = \int_0^T \nabla V_\gamma^{(i,j+1)}(y(t))^t \left(g_1(y(t)) + g_2(y(t))w_\gamma^{(i,j+1)}\right) dW(t).$$

By Assumption 2, $\nabla V_\gamma^{(i,j+1)}(y)$ and $\frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2}$ are bounded on $\bar{\Omega}$, and consequently on Ω. Further, $h(y)$, $g_1(y)$ and $g_2(y)$ are bounded on Ω by the assumption on the dynamics. As a consequence $\left(g_1(y(t)) + g_2(y(t))w_\gamma^{(i,j+1)}\right)$ is also bounded on Ω. Hence M_T is a true martingale. This implies

$\mathbb{E}[M_T] = 0$ for any finite T. Hence we obtain from (3.11)

$$\mathbb{E}V_\gamma^{(i,j+1)}(y(T)) - V_\gamma^{(i,j+1)}(x) = \mathbb{E}\left[\int_0^T \mathcal{L}_{(u_\gamma^{(i)}, w_\gamma^{(i,j+1)})} V_\gamma^{(i,j+1)}(y)\, dt\right].$$

Then, taking the limit as $T \to \infty$, we obtain $-V_\gamma^{(i,j+1)}(x) = \mathbb{E}\left[\int_0^\infty \mathcal{L}_{(u_\gamma^{(i)}, w_\gamma^{(i,j+1)})} V_\gamma^{(i,j+1)}(y)\, dt\right]$.

Since (3.9) is asymptotically stable on Ω for each pair (i,j), by Itô's formula we find

$$\begin{aligned}
&V_\gamma^{(i,j)}(x) - V_\gamma^{(i,j+1)}(x) \\
&= \mathbb{E}\int_0^\infty \Bigg(\Big(\nabla V_\gamma^{(i,j+1)^t}(f + gu_\gamma^{(i)} + hw_\gamma^{(i,j+1)}) + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,j+1)})^t \frac{\partial^2 V_\gamma^{(i,j+1)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,j+1)})]\Big) \\
&\quad - \Big(\nabla V_\gamma^{(i,j)^t}(f + gu_\gamma^{(i)} + hw_\gamma^{(i,j+1)}) + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,j+1)})^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,j+1)})]\Big)\Bigg) dt \\
&=: \mathbb{E}\int_0^\infty (A - B)\, dt.
\end{aligned}$$

Using the GHJI equation (3.7), it follows that $A = -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j+1)}\right\|_P^2\Big)$. Applying the update disturbance $w_\gamma^{(i,j+1)}$ i.e. using

$$-\Big(2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_2]\Big) w_\gamma^{(i,j+1)} + Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_1] = -h^t \nabla V_\gamma^{(i,j)}(y),$$

we get

$$\begin{aligned}
&\nabla V_\gamma^{(i,j)}(y)^t \big(f + hw_\gamma^{(i,j+1)} + gu_\gamma^{(i)}\big) + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,j+1)})^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,j+1)})] \\
&= -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j)}\right\|_P^2\Big) + \langle h^t \nabla V_\gamma^{(i,j)}, w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\rangle \\
&\quad + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,j+1)})^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,j+1)})] \\
&\quad - \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,j)})^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,j)})] \\
&\quad = -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j)}\right\|_P^2\Big) + \Big\langle (2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_2]) w_\gamma^{(i,j+1)}, \\
&\quad w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\Big\rangle + \Big\langle \frac{1}{2}Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_2](w_\gamma^{(i,j+1)} + w_\gamma^{(i,j)}), w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\Big\rangle \\
&= -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,j)}\right\|_P^2\Big) + \langle 2\gamma^2 P w_\gamma^{(i,j+1)}, w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\rangle \\
&\quad - \tfrac{1}{2}\big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big)^t Tr\big[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_2\big]\big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big).
\end{aligned} \tag{3.12}$$

Hence, an estimate for B is obtained from (3.12). Altogether, we arrive at

$$A - B = -\big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big)^t \Big(\gamma^2 P - \tfrac{1}{2}Tr\big[g_2^t \frac{\partial^2 V_\gamma^{(i,j)}}{\partial y^2} g_2\big]\Big)\big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big).$$

Therefore, it follows that

$$\begin{aligned}&V_\gamma^{(i,j)}(x) - V_\gamma^{(i,j+1)}(x)\\&\quad= -\tfrac{1}{2}\,\mathbb{E}\int_0^\infty \big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big)^t\Big(2\gamma^2 P - Tr\big[g_2^t\frac{\partial^2 V_\gamma^{(i,j)}}{\partial x^2}g_2\big]\Big)\big(w_\gamma^{(i,j+1)} - w_\gamma^{(i,j)}\big)\,dt \le 0,\end{aligned}$$

where (3.6) was used. Consequently, $V_\gamma^{(i,j)}(x) \le V_\gamma^{(i,j+1)}(x)$. The rest of the proof follows similarly as in Theorem 2.1. □

Remark 8. *If $V_\gamma^{(i,j)} \in C^2(\Omega)$ satisfy (3.7), $V_\gamma^{(i,\infty)} \in C^2(\Omega)\cap C^1(\bar\Omega)$ and $\{\nabla^m V_\gamma^{(i,j)}\}$ is equicontinuous in Ω, then we have $\{\nabla^m V_\gamma^{(i,j)}\}$ converges pointwise to $\nabla^m V_\gamma^{(i,\infty)}$ for $m = 1, 2$. Further, assuming the boundedness of $\Big(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,j)}}{\partial x^2}g_2]\Big)^{-1}$ and $\Big(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial x^2}g_2]\Big)^{-1}$ on Ω, $w_\gamma^{(i,j)}$ converges to $w_\gamma^{(i,\infty)}$ as $j\to\infty$. Moreover, $V_\gamma^{(i,\infty)}(x)$ solves the equation*

(3.13)

$$\begin{cases}\nabla V_\gamma(x)^t(f(x) + gu_\gamma^{(i)}) + \frac{1}{2}\Big(\nabla V_\gamma(x)^t h + Tr[g_1^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2]\Big)(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2])^{-1}\Big(h^t\nabla V_\gamma(x)\\ \qquad\qquad + Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_1]\Big) + \Big\|u_\gamma^{(i)}\Big\|_R^2 + \ell(x) + \frac{1}{2}Tr[g_1^t\frac{\partial^2 V_\gamma}{\partial x^2}g_1] = 0,\\ \qquad 2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma}{\partial x^2}g_2] > 0,\end{cases}$$

and

$$w_\gamma^{(i,\infty)} = \Big(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial x^2}g_2]\Big)^{-1}\Big(h^t\nabla V_\gamma^{(i,\infty)}(x) + Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial x^2}g_1]\Big). \tag{3.14}$$

Proof. The proof follows similarly from Proposition 4 in [30]. □

Theorem 3.2. *If the stochastic system*

$$dy = \Big(f(y) + g(y)u_\gamma^{(i)} + h(y)w_\gamma^{(i,\infty)}\Big)\,dt + \Big(g_1(y) + g_2(y)w_\gamma^{(i,\infty)}\Big)\,dW, \tag{3.15}$$

is asymptotically stable in probability on Ω for all i, and $V_\gamma^{(i,\infty)}(x) \in C^2(\mathbb{R}^d)$, then $0 \le V_\gamma^{(i+1,\infty)}(x) \le V_\gamma^{(i,\infty)}(x)$ holds for each $x \in \Omega$.

Proof. Due to asymptotical stability of (3.15) for each i, along the trajectory $dy = \Big(f(y) + g(y)u_\gamma^{(i+1)} + h(y)w_\gamma^{(i+1,\infty)}\Big)\,dt + \Big(g_1(y) + g_2(y)w_\gamma^{(i+1,\infty)}\Big)\,dW$, we obtain the difference between $V_\gamma^{(i+1,\infty)}(x)$ and $V_\gamma^{(i,\infty)}(x)$ as

$$\begin{aligned}&V_\gamma^{(i+1,\infty)}(x) - V_\gamma^{(i,\infty)}(x)\\&= -\mathbb{E}\int_0^\infty\Bigg(\Big(\nabla V_\gamma^{(i+1,\infty)^t}(f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)}) + \frac{1}{2}Tr[(g_1 + g_2w_\gamma^{(i+1,\infty)})^t\frac{\partial^2 V_\gamma^{(i+1,\infty)}}{\partial y^2}\\&\qquad(g_1 + g_2w_\gamma^{(i+1,\infty)})]\Big) - \Big(\nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)})\\&\qquad\qquad + \frac{1}{2}Tr[(g_1 + g_2w_\gamma^{(i+1,\infty)})^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1 + g_2w_\gamma^{(i+1,\infty)})]\Big)\Bigg)\,dt. =: -\mathbb{E}\int_0^\infty (A - B)\,dt.\end{aligned}$$

Using the HJI equation (3.13), it follows that $A = -\Big(\ell(y) + \left\|u_\gamma^{(i+1)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i+1,\infty)}\right\|_P^2\Big)$, where $w_\gamma^{(i+1,\infty)}$ satisfies (3.14). To obtain an estimate of B, we express B as follows

$$\begin{aligned}
&\nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i+1)} + hw_\gamma^{(i+1,\infty)}) + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i+1,\infty)})^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i+1,\infty)})] \\
&= \nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i)} + hw_\gamma^{(i,\infty)}) + \frac{1}{2}Tr[(g_1 + g_2 w_\gamma^{(i,\infty)})^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1 + g_2 w_\gamma^{(i,\infty)})] \\
&\quad + \nabla V_\gamma^{(i,\infty)^t} g(u_\gamma^{(i+1)} - u_\gamma^{(i)}) + \Big\langle h^t \nabla V_\gamma^{(i,\infty)} + Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_1], (w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)})\Big\rangle \\
&\quad + \Big\langle \tfrac{1}{2}Tr\big[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_2\big]\big(w_\gamma^{(i+1,\infty)} + w_\gamma^{(i,\infty)}\big),\ w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\Big\rangle \\
&= -\Big(\ell(y) + \left\|u_\gamma^{(i)}\right\|_R^2 - \gamma^2 \left\|w_\gamma^{(i,\infty)}\right\|_P^2\Big) + \langle 2Rz, \mathcal{P}_U(-z) - u_\gamma^{(i)}\rangle \\
&\quad + \Big\langle (2\gamma^2 P - Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_2])w_\gamma^{(i,\infty)} + \frac{1}{2}Tr[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_2](w_\gamma^{(i+1,\infty)} + w_\gamma^{(i,\infty)}), \\
&\qquad\qquad w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\Big\rangle \\
&= -\Big(\ell(y) + \|\mathcal{P}_U(-z)\|_R^2 - \gamma^2 \left\|w_\gamma^{(i+1,\infty)}\right\|_P^2\Big) - \left\|u_\gamma^{(i)} - \mathcal{P}_U(-z)\right\|_R^2 \\
&\quad + 2(z + \mathcal{P}_U(-z))^t R(\mathcal{P}_U(-z) - u_\gamma^{(i)}) \\
&\quad - \big(w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\big)^t \Big(\gamma^2 P - \tfrac{1}{2}Tr\big[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_2\big]\Big)\big(w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\big),
\end{aligned}$$

where as earlier we denote $z = \frac{1}{2}R^{-1}g^t \nabla V_\gamma^{(i,\infty)}$, $g^t \nabla V_\gamma^{(i,\infty)} = 2Rz$, and $u_\gamma^{(i+1)} = \mathcal{P}_U(-z)$. Finally, we arrive at

$$\begin{aligned}
&V_\gamma^{(i+1,\infty)}(x) - V_\gamma^{(i,\infty)}(x) \\
&= -\mathbb{E}\int_0^\infty \Big(\big(w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\big)^t \Big(\gamma^2 P - \tfrac{1}{2}Tr\big[g_2^t \frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2} g_2\big]\Big)\big(w_\gamma^{(i+1,\infty)} - w_\gamma^{(i,\infty)}\big)\Big) dt \\
&\quad + \mathbb{E}\int_0^\infty \Big(2(z + \mathcal{P}_U(-z))^t R(\mathcal{P}_U(-z) - u_\gamma^{(i)}) - \left\|u_\gamma^{(i)} - P_U(-z)\right\|_R^2\Big) dt \le 0.
\end{aligned}$$

Therefore, we get $V_\gamma^{(i+1,\infty)}(x) \le V_\gamma^{(i,\infty)}(x)$. Further $V_\gamma^{(i,\infty)}(x) \ge 0$ follows from Theorem 3.1. □

Remark 9. *If $V_\gamma^{(i,\infty)} \in C^2(\Omega)$ satisfy (3.7), $V_\gamma \in C^2(\Omega) \cap C^1(\bar{\Omega})$ and $\{\nabla^m V_\gamma^{(i,\infty)}\}$ is equicontinuous, then we have $\{\nabla^m V_\gamma^{(i,\infty)}\}$ converges pointwise to $\nabla^m V_\gamma$ for $m = 1, 2$ in Ω, where V_γ satisfies (3.4).*

Proof. See proposition 4 in [30] for the proof in an analogous situation. □

Lemma 2. *If $V_\gamma^{(i,\infty)} \in C^2(\mathbb{R}^d)$ for all $i \ge 0$ and (3.13) holds, then $u_\gamma^{(i+1)}$ is an asymptotically stabilizing control on Ω for the undisturbed system ($w = 0$), $dy = (f(y) + gu)\,dt + g_1(y)\,dW$.*

Proof. From Theorem 3.1, $V_\gamma^{(i,\infty)}$ is positive definite. Now along the trajectory $dy = (f(y) + gu_\gamma^{(i+1)})\,dt + g_1(y)\,dW$, the time derivative of $t \to V_\gamma^{(i,\infty)}(y)(t)$ can be obtained as

$$\begin{aligned}&\nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i+1)}) + \frac{1}{2}Tr[(g_1)^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1)]\\ &\quad = \nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i)}) + \frac{1}{2}Tr[(g_1)^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1)] + \nabla V_\gamma^{(i,\infty)^t} g(u_\gamma^{(i+1)} - u_\gamma^{(i)}).\end{aligned}$$

From (3.13), it follows that

$$\begin{aligned}&\nabla V_\gamma^{(i,\infty)}(y)^t(f(y) + gu_\gamma^{(i)}) + \frac{1}{2}Tr[g_1^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_1]\\ &= -\frac{1}{2}\Big(\nabla V_\gamma^{(i,\infty)}(y)^t h + Tr[g_1^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_2]\Big)(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_2])^{-1}\Big(h^t\nabla V_\gamma^{(i,\infty)}(y)\\ &\qquad + Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_1]\Big) - \left\|u_\gamma^{(i)}\right\|_R^2 - \ell(y).\end{aligned}$$

Also $\nabla V_\gamma^{(i,\infty)^t} g(u_\gamma^{(i+1)} - u_\gamma^{(i)})$ is estimated in Theorem 3.2. Hence, we get

$$\begin{aligned}&\nabla V_\gamma^{(i,\infty)^t}(f + gu_\gamma^{(i+1)}) + \frac{1}{2}Tr[(g_1)^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}(g_1)]\\ &\quad = -\ell(y) - \|\mathcal{P}_U(-z)\|_R^2 - \left\|u_\gamma^{(i)} - \mathcal{P}_U(-z)\right\|_R^2 + 2(z + \mathcal{P}_U(-z))^t R(\mathcal{P}_U(-z) - u_\gamma^{(i)})\\ &\qquad - \frac{1}{2}\Big(\nabla V_\gamma^{(i,\infty)}(y)^t h + Tr[g_1^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_2]\Big)(2\gamma^2 P - Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_2])^{-1}\Big(h^t\nabla V_\gamma^{(i,\infty)}(y)\\ &\qquad + Tr[g_2^t\frac{\partial^2 V_\gamma^{(i,\infty)}}{\partial y^2}g_1]\Big) < 0,\end{aligned}$$

and the asymptotic stabilizing property of $u_\gamma^{(i+1)}$ follows. □

4. Numerical examples

In this section we solve the HJI equation by policy iteration Algorithm 1 for the deterministic, and by Algorithm 2 for the stochastic case. An implicit upwind scheme is used for solving backward PDEs, and a bisection algorithm is employed for finding γ^*. For more details on the upwind scheme, see [3] and [30]. To address the difficulty arising from the need to initialize Algorithm 1 with an asymptotically stabilizing solution also in the case when the origin is not asymptotically stable, a discount factor $\lambda > 0$ is introduced in the cost. With this change the algorithm always converged satisfactorily. The result thus obtained can be used to initialize the algorithm without discount factor. An alternative possibility for initialization is provided by the choice given by the Riccati feedback law, which is obtained from considering the linearized system. For each of the numerical tests that we shall describe below, the stability properties of the steady state and its consequences for the initialization will be addressed.

For completeness, we briefly discuss the upwind scheme that is used, where for simplicity, we focus on the 1D deterministic model. For a fixed γ we describe the scheme to solve the HJI equation (2.9). Let (u, V) be the initial pair, suppressing the dependence on γ. Let n stand for the iteration/time loop, let i denote the numbering of the mesh points x_i, and let dt stand

for the time step. Set $V_i = V(x_i)$, $u_i = u(x_i)$ and $w_i = w(x_i)$ with $i = 1, \dots, I$, where I is the total number of mesh points. We use the forward difference operator $\nabla V_{i,F} = \frac{V_{i+1}-V_i}{\Delta x}$, $w_{i,F} = \frac{P^{-1}}{2\gamma^2}\nabla V_{i,F}$ whenever the drift $S_{i,F} = f(x_i) + g(x_i)u_i + h(x_i)w_{i,F} > 0$, and the backward difference operator $\nabla V_{i,B} = \frac{V_i - V_{i-1}}{\Delta x}$, if the drift $S_{i,B} = f(x_i) + g(x_i)u_i + h(x_i)w_{i,B} < 0$, and $D\bar{V}_i = 2P\gamma^2\bar{w}_i$ with $\bar{w}_i = -f(x_i)/h(x_i)$ for $S_{i,F} \le 0 \le S_{i,B}$, where $\Delta x = x_{i+1} - x_i$ is the mesh size. Then an approximation of the derivative of the value function can be constructed as

$$\nabla V^n_{iupwind} = \nabla V^n_{i,F}\mathbb{1}_{S_{i,F}>0} + \nabla V^n_{i,B}\mathbb{1}_{S_{i,B}<0} + D\bar{V}^n_i\mathbb{1}_{S_{i,F}\le 0\le S_{i,B}},$$

where $\mathbb{1}$ is the characteristic function. Finally, taking $\ell(x) = x^2$, we solve

$$\begin{aligned} -\frac{(V^n_i - V^{(n-1)}_i)}{dt} &- \lambda V^n_i + \nabla V^n_{iupwind}(f(x_i) + gu^n_i + hw^n_i) \\ &+ x_i^2 + \|u^n_i\|^2_R - \gamma^2 \|w^n_i\|^2_P = 0. \end{aligned} \tag{4.1}$$

Further, we get the disturbance update $w^{n+1}_i = \frac{1}{2\gamma^2}P^{-1}h(x_i)^t\nabla V^n_{iupwind}$. Once the disturbance loop is completed by satisfying the tolerance, we update the control similarly by considering the upwind derivative from the value function obtained after completing the disturbance loop. Subsequently, using a bisection algorithm we can get the smallest value of γ, and the corresponding value of u_{γ^*}, w_{γ^*} and V_{γ^*}. Similarly, for the stochastic case, the additional second order term is discretized by $D^2V^n_i = (V^n_{i+1} - 2V^n_i + V^n_{i-1})/\Delta x^2$, and an analogous upwind scheme is used.

4.1. **Test 1: 1D equation.** In the first example we test a linear unstable system. We first employ as discount factor $\lambda = 0.005$. Then the upwind solutions for $\lambda = 0.005$ and that for $\lambda = 0$ are compared, where in both cases we take the Riccati solution as the initialization. We also report on the difference between the upwind and the Riccati solutions.

We consider

$$\min_{u(\cdot)\in\mathcal{U}}\max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u(\cdot), w(\cdot); (x_0)) = \int_0^\infty e^{-\lambda t}\Big(\|x\|^2 + \|u\|^2_R - \gamma^2 \|w\|^2_P \Big)dt,$$

subject to the one dimensional equation

$$\frac{dx}{dt} = 0.5x + u + w, \qquad x(0) = x_0. \tag{4.2}$$

Further we choose $R = 0.1$, $P = 1$, $dt = 0.005$, $\Omega = (-2, 2)$ and $\gamma = 1$. Let $e_\infty :=$

TABLE 1. L^2 and L^∞ errors for the upwind scheme in Test 1 with $\lambda = 0$.

Meshpoints	$\|V - V_{upwind}\|_{L^2}$	Conv. order	$\|V - V_{upwind}\|_{L^\infty}$	Conv. order
200	0.07735		0.009425	
400	0.017396	2.15	0.0015028	2.64
600	0.0011153	6.77	7.87×10^{-5}	7.27
800	9.65×10^{-4}	0.503	5.9×10^{-5}	1.002
1000	8.63×10^{-4}	0.502	4.72×10^{-5}	1.001
1200	7.87×10^{-4}	0.502	3.93×10^{-5}	1.001
1400	7.29×10^{-4}	0.502	3.37×10^{-5}	1.001

$\|V - V_{upwind}\|_{L^\infty}$ and $e_2 := \|V - V_{upwind}\|_{L^2}$ denote the L^∞ and L^2 norm errors of the value

function, where V corresponds to the Riccati solution. We calculate the order of convergence $\alpha := ln(e_{new}/e_{old})/ln(h_{new}/h_{old})$, where e_{new} and e_{old} correspond to the current and previous steps, respectively, and h_{new}, h_{old} are the current and previous step mesh sizes respectively. Since h is inversely proportional to the number of mesh points I, the convergence order can be written as $\alpha = ln(e_{new}/e_{old})/ln(I_{old}/I_{new})$. Next we solve with $\lambda = 0$ in the value function, and compare the corresponding upwind solution with the $\lambda = 0.005$ case. From Table 1 we see that the L^2 and L^∞ order of convergence for $V - V_{upwind}$ are 0.5 and 1, respectively. Moreover, we observe that the difference between the value function for $\lambda = 0$ and $\lambda = 0.005$ is $\|V_0 - V_{0.005}\|_{L^2} = 5.93 \times 10^{-4}$ and $\|V_0 - V_{0.005}\|_{L^\infty} = 3.82 \times 10^{-5}$, where V_λ denotes the upwind value function with the corresponding λ.

4.2. **Test 2: 2D nonlinear equation.** We consider next the following minimization problem

$$\min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u(\cdot), w(\cdot); (x_0, y_0)) = \int_0^\infty e^{-\lambda t}\Big(\|x\|^2 + \|y\|^2 + \|u\|_R^2 - \gamma^2 \|w\|_P^2 \Big) dt,$$

associated to the following two-dimensional nonlinear equation

$$\begin{aligned} \frac{dx}{dt} &= -x - y, \quad x(0) = x_0, \\ \frac{dy}{dt} &= -(1 - x^2)y + u + w, \quad y(0) = y_0, \end{aligned} \tag{4.3}$$

where $R = 0.01$, $P = 1$, $\gamma = 1$, $\lambda = 0.005$ and $dt = 0.001$. Let us observe that the origin is a locally stable equilibrium for this system, but it is not globally stable. The HJI equation was solved for $V_{0.005}$ in $\Omega = (-2, 2)^2$ with 100×100 equidistant mesh points. Observe that this choice of Ω is not contained in the neighborhood of asymptotic stability for the uncontrolled system (4.3). Next the solution thus obtained was used as initialization to solve with zero discount factor. The associated value function is denoted by V_0. The L^∞ and L^2 errors were found to be $\|V_0 - V_{0.005}\|_{L^\infty} = 6.35 \times 10^{-5}$ and $\|V_0 - V_{0.005}\|_{L^2} = 0.0017$.

4.3. **Test 3: Lorenz system.** Here we treat the controlled Lorenz system, with both, stable and unstable parameter settings. We consider

$$\min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u(\cdot), w(\cdot); (x_0, y_0, z_0)) = \int_0^\infty e^{-\lambda t}\Big(\|x\|^2 + \|y\|^2 + \|z\|^2 + \|u\|_R^2 - \gamma^2 \|w\|_P^2 \Big) dt, \tag{4.4}$$

subject to the three-dimensional Lorenz system, with control and disturbance appearing in the second equation,

$$\frac{dx}{dt} = \sigma(y - x), \quad x(0) = x_0, \tag{4.5}$$

$$\frac{dy}{dt} = x(\rho - z) - y + u + w, \quad -1 \le u \le 1, \quad y(0) = y_0, \tag{4.6}$$

$$\frac{dz}{dt} = xy - \beta z, \quad z(0) = z_0, \tag{4.7}$$

where $\sigma > 1$, $\rho > 0$ and $\beta > 0$. The system (4.5)-(4.7) has 3 steady state solutions, $C^0 = (0, 0, 0)$, $C^+ = \Big(\sqrt{\beta(\rho - 1)}, \sqrt{\beta(\rho - 1)}, \rho - 1\Big)$ and $C^- = \Big(- \sqrt{\beta(\rho - 1)}, -\sqrt{\beta(\rho - 1)}, \rho - 1\Big)$. If $\rho < 1$, all the steady state solutions are stable.

We first consider such a stable unconstrained situation and choose $\rho = 0.5$, $\sigma = 10$, $\beta = 8/3$. The other parameter settings are chosen to be $R = 0.01$, $P = 1$, $\gamma = 2$, $dt = 0.01$, and

$40 \times 40 \times 40$ mesh points in the domain $\Omega = (-2, 2)^3$. We calculate the value function for $\lambda = 0.005$ and denote it by $V_{0.005}$. Moreover, using $V_{0.005}$ as initialization, we also obtain the solution for $\lambda = 0$ denoted by V_0. The L^∞ error between these two solutions is found to be $\|V_0 - V_{0.005}\|_{L^\infty} = 0.0042$.

We carried out many more experiments for the case of an unstable system with parameter settings $\sigma = 10$, $\beta = 8/3$, and $\rho = 2$. Again the HJI equation was solved over $\Omega = (-2, 2)^3$ with $40 \times 40 \times 40$ spatial discretized points, with $R = 0.1$, $P = 1$, $dt = 2$. To avoid difficulties with initialization we took $\lambda = 0.05$ and chose the Riccati solution as the initial guess for policy iteration. We find that $\gamma^* = 1.9196$ and $\gamma^* = 7.1136$ for the unconstrained and constrained cases respectively, with associated costs $\mathcal{J} = 1.86$ and $\mathcal{J} = 3.89$. When discussing the behavior of trajectories of system (4.5)-(4.7) below, we do this for $[x_0, y_0, z_0] = [-1.5, -1.5, 1.5]$ as initial conditions.

We next depict results for the unstable parameter choice settings: From Figure 1(i)-(ii) we observe that the uncontrolled state and the corresponding state cost do not settle towards zero. The states of the controlled system, both for the unconstrained and the constrained HJI controls, documented in Figure 1(iii), approach zero. As expected this is faster for unconstrained control case than for the constrained case. Similar behavior for the control cost and running cost are shown in Figure 1(iv) and Figure 1(vi) respectively. In Figure 1(v) the disturbance for the constrained control tends to zero faster compared to the disturbance for the unconstrained case since the constrained control cannot handle as much disturbance as the unconstrained one.

We show the behavior of the state cost $\|x(t)\|^2 + \|y(t)\|^2 + \|z(t)\|^2$ and the control trajectory with respect to different initial conditions in Figure 2 (i)-(ii).

We next show comparisons between the HJI and HJB synthesis in presence of more adverse noise and different initial conditions. If the HJB control is used within a disturbance of the form $w_1 = -0.6u_{HJB} + 0.1sin(10t)$ then the unconstrained trajectory for both HJB control and state cost oscillate, whereas the unconstrained HJI control and state cost are successful in noise rejection, and settle at zero, see Figure 3(i)-(ii). We mention that if we choose the disturbance of the form $w_1 = -0.5u_{HJB} + 0.1sin(10t)$, then there is no spike in the HJB case and the corresponding state and control trajectories converge towards zero.

Similarly, if the noise is of the form $w_2 = -0.6u_{HJI} + 0.1sin(10t)$, in Figure 4(i) we observe that the unconstrained HJI control is successful in handling the noise, whereas the unconstrained HJB control fails and starts to oscillate. Similar behavior is shown for the unconstrained running cost in Figure 4(ii).

To compare the constrained HJI and HJB synthesis in the presence of noise w_1 and w_2, we only modified the constraint in (4.6) to be $-4 \leq u \leq 4$. With disturbances of the form $w_1 = -0.6u_{HJB} + 0.01sin(10t)$, the HJI running cost tends to zero faster compared to the HJB cost, as seen in Figure 5. Further, with noise of the form $w_2 = -0.6u_{HJI} + 0.01sin(10t)$, the HJI synthesis is successful in driving the control and state cost to their equilibrium position, whereas the HJB synthesis fails as observed from Figure 6.

The study of stochastic zero-sum differential games is motivated by the need to design robust controls for chaotic systems such as the Lorenz system, where sensitivity to noise and modeling errors severely limits classical approaches. Reformulating the problem as a two-player stochastic game, with the controller minimizing and the disturbance maximizing, naturally leads to a second-order HJI equation. For the stochastic case, we put the additive noise $g_1 = 0.5$ in the third equation (4.7). The other parameters and the initial condition are the same as in the deterministic Lorenz system. We obtain $\gamma^* = 1.9135$ and $\gamma^* = 7.1136$ for the stochastic unconstrained, respectively constrained HJI cases. From Figure 7 we observe that the behavior

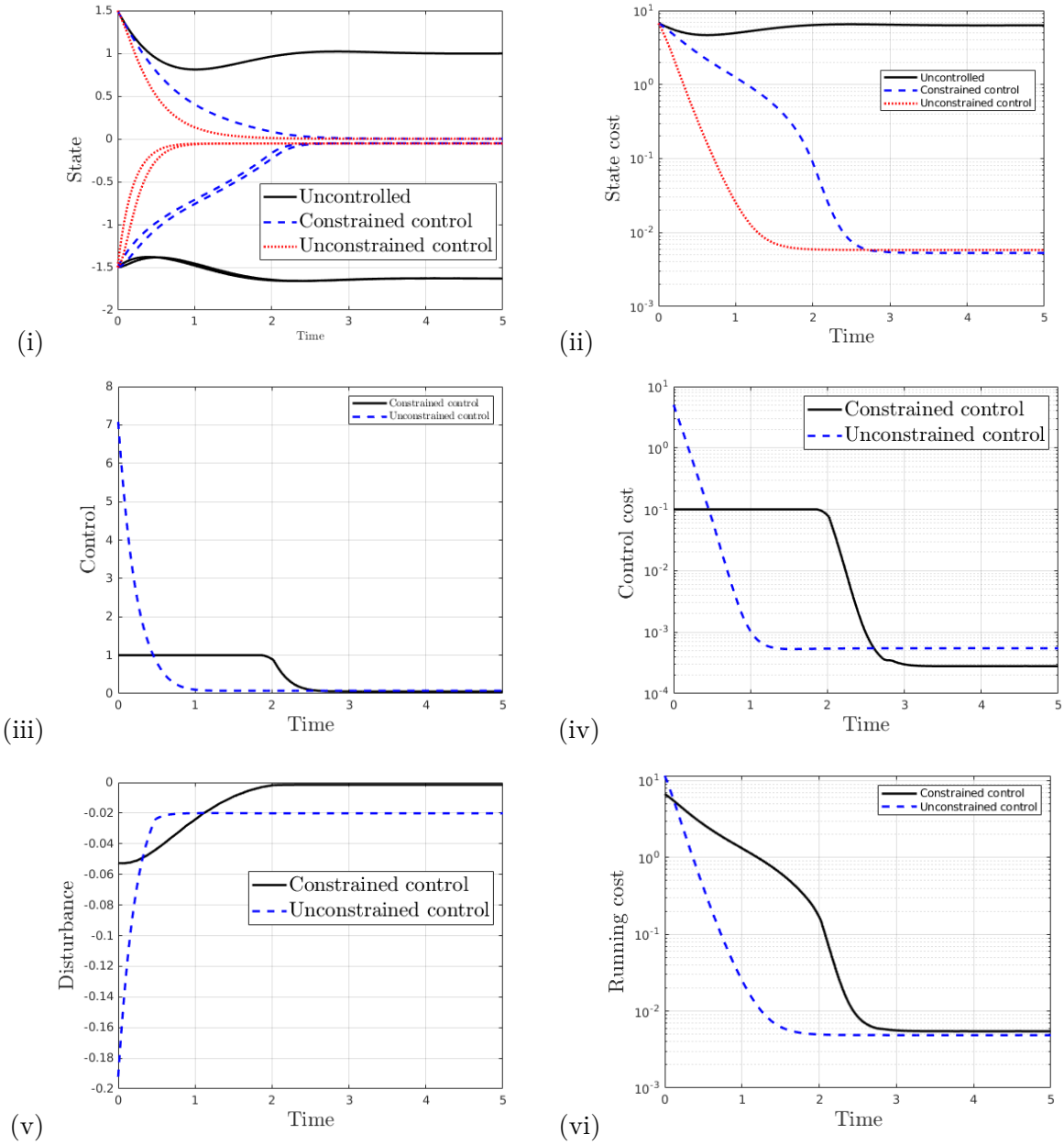


FIGURE 1. Test 3: Deterministic case. **i)** State, **ii)** State cost $\|x(t)\|^2 + \|y(t)\|^2 + \|z(t)\|^2$, **iii)** Control, **iv)** Control cost $\|u(t)\|_R^2$, **v)** Disturbance, **vi)** Running cost $\|x(t)\|^2 + \|y(t)\|^2 + \|z(t)\|^2 + \|u(t)\|_R^2 - \gamma^{*2}\|w(t)\|_P^2$.

of the stochastic Lorenz system is similar to the deterministic case except for the appearance of the randomness of the optimal state.

With the disturbance of the form $w_1 = -0.6u_{HJB} + 0.01sin(10t)$, comparisons between the HJB and HJI synthesis under stochastic dynamics are shown in Figure 8(i)-(ii) for the constrained case. The reader will notice that the HJI control and state cost both tend to settle

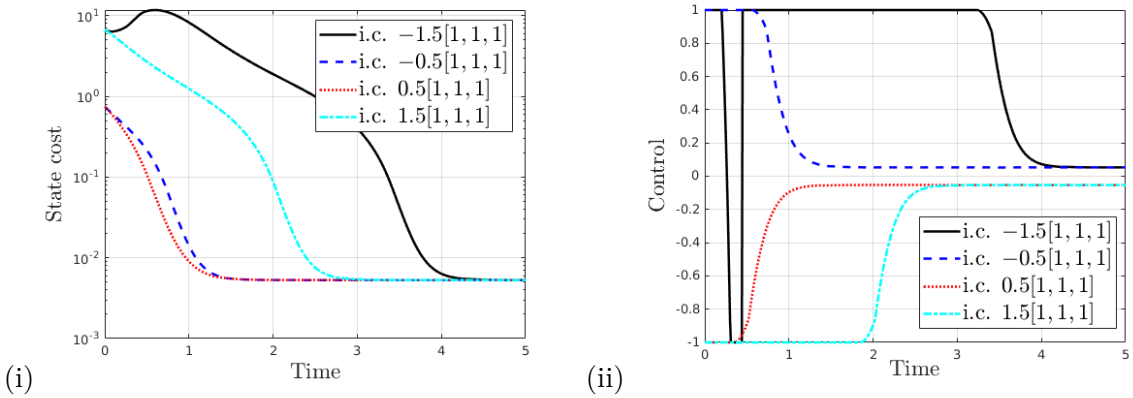


FIGURE 2. Test 3: Deterministic case. State cost and Control trajectory with different initial conditions. **i)** State cost **ii)** Control.

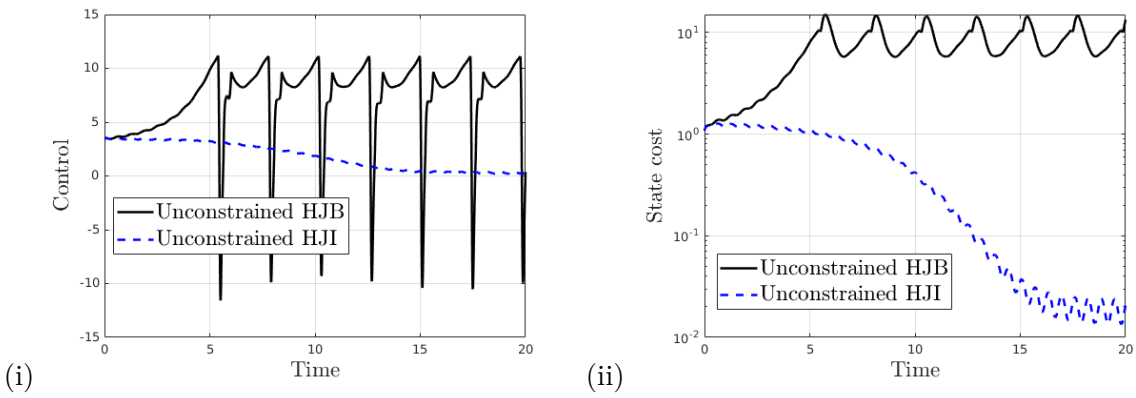
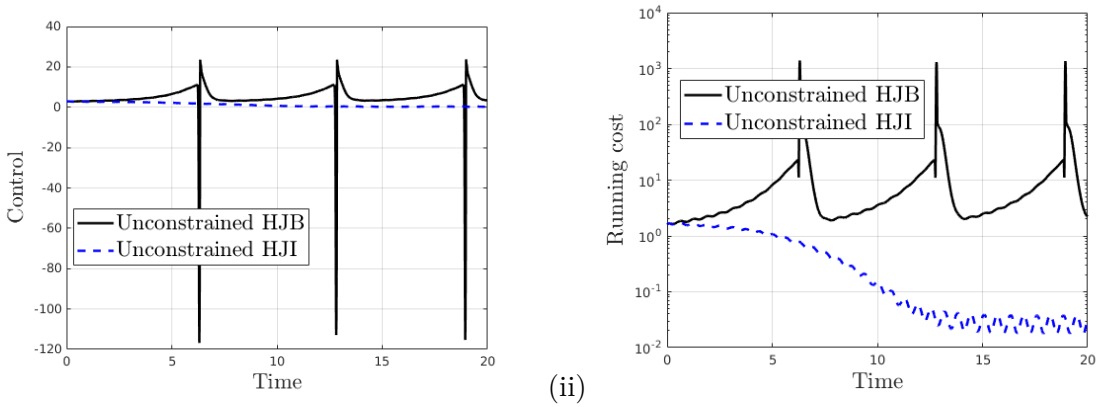


FIGURE 3. Test 3: Unconstrained case, Comparison between the HJI and HJB synthesis with noise $w_1 = -0.6u_{HJB} + 0.1sin(10t)$, $[x_0, y_0, z_0] = [-0.6, -0.6, -0.6]$. **i)** Control, **ii)** State cost.

to zero, while this does not happen in the HJB case. This difference occurs if the constraints are lifted.

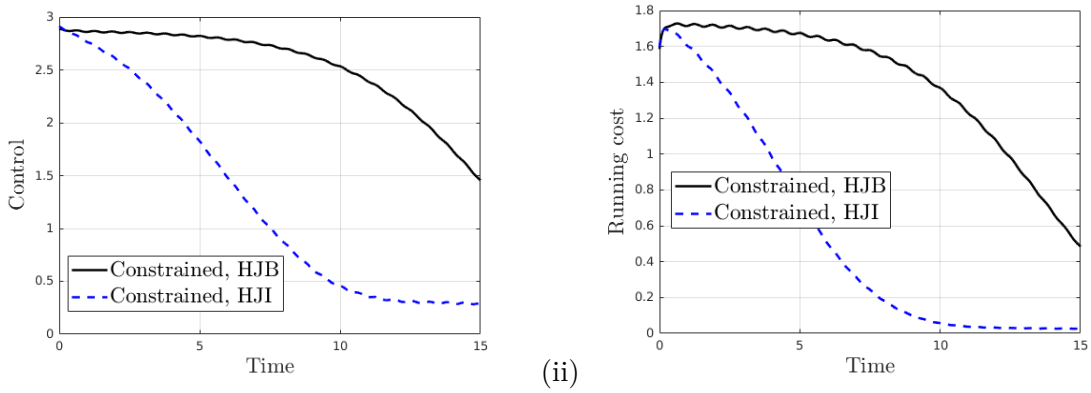
We also carried out tests as $\lambda \to 0$. In the unconstrained case, with the solution to $V_{0.05}$ as an initialization, $dt = 0.01$ and keeping the other parameters as before, we obtain for the L^∞ error $\|V_0 - V_{0.05}\|_{L^\infty(\Omega)} = 0.08$. If dt is reduced to $dt = 0.001$, then $\|V_0 - V_{0.05}\|_{L^\infty(\Omega)} = 0.03$. Further, for $\lambda = 0.005$, we obtained $\|V_0 - V_{0.005}\|_{L^\infty(\Omega)} = 3.1 \times 10^{-5}$. In the constrained case, with $-4 \leq u \leq 4$, then keeping $dt = 0.01$ and $\gamma = 10$, we obtain $\|V_0 - V_{0.005}\|_{L^\infty(\Omega)} = 0.02$ and $\|V_0 - V_{0.05}\|_{L^\infty(\Omega)} = 0.12$.

4.4. **Test 4: Van der Pol equation.** Here we consider the controlled Van der Pol equation. It is well-known that the unforced Van der Pol equation has a stable limit cycle, the origin is unstable. While admittedly we cannot find a stabilizing initial condition, it appears to be of



(i) (ii)

FIGURE 4. Test 3: Unconstrained case, Comparison between the HJI and HJB synthesis with noise $w_2 = -0.6u_{HJI} + 0.1sin(10t)$, $[x_0, y_0, z_0] = [-0.5, -0.5, -0.5]$. **i)** Control, **ii)** Running cost.



(i) (ii)

FIGURE 5. Test 3: Constrained case, comparison between the HJI and HJB synthesis with noise $w_1 = -0.6u_{HJB} + 0.01sin(10t)$, $-4 \leq u \leq 4$, $[x_0, y_0, z_0] = [-0.5, -0.5, -0.5]$, other parameters as before. **i)** Control, **ii)** Running cost.

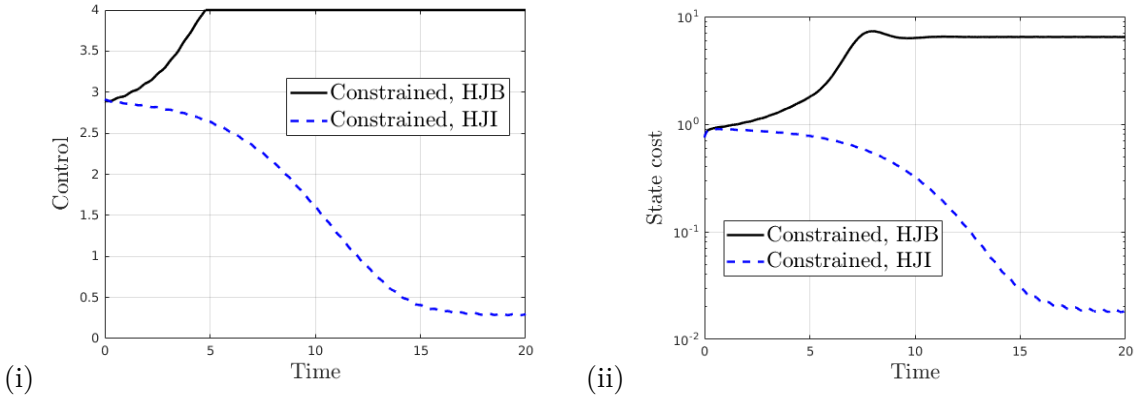
interest to test the proposed methodology for this challenging example. We consider

$$\min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u(\cdot), w(\cdot); (x_0, y_0)) = \int_0^\infty e^{-\lambda t}\Big(\|x\|^2 + \|y\|^2 + \|u\|_R^2 - \gamma^2 \|w\|_P^2 \Big) dt,$$

subject to the following two-dimensional Van der Pol equation

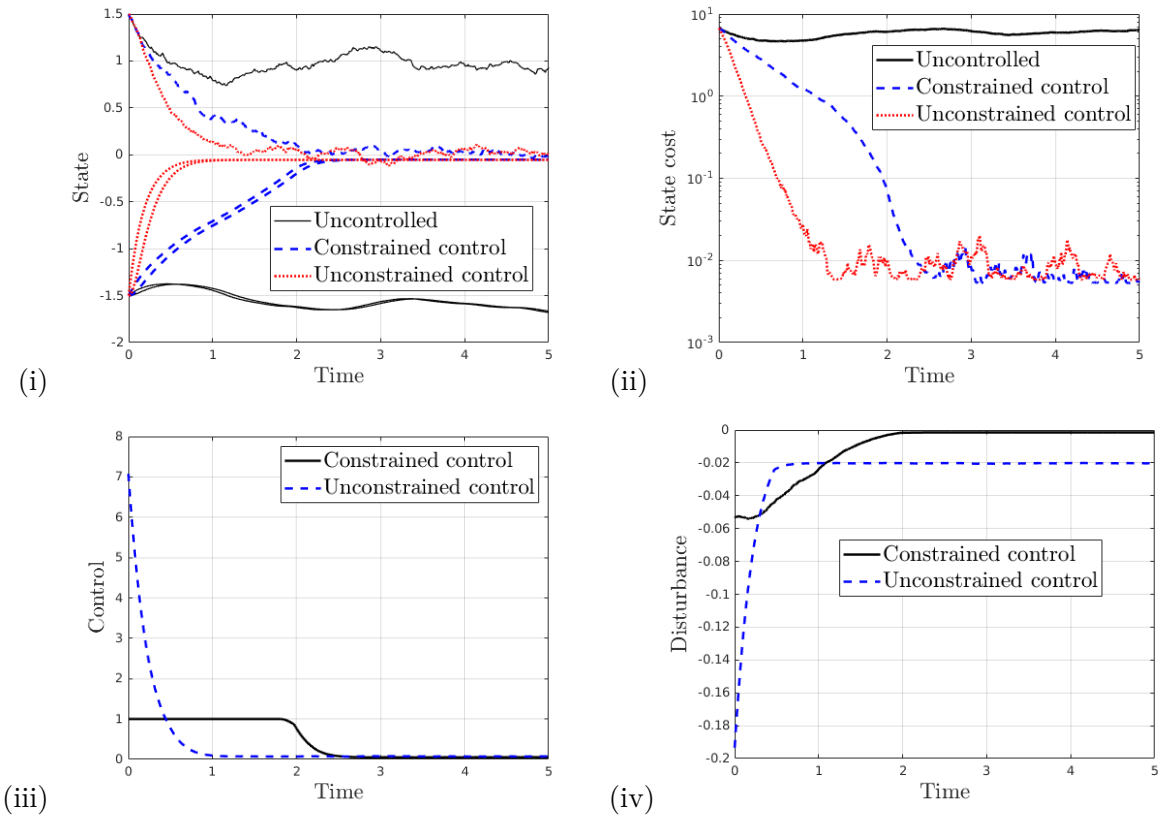
$$\begin{aligned} \frac{dx}{dt} &= y, \quad x(0) = x_0, \\ \frac{dy}{dt} &= (1 - x^2)y - x + u + w, \quad -1 \leq u \leq 1, \quad y(0) = y_0, \end{aligned} \tag{4.8}$$

where $R = 0.01$, $P = 1$ and $[x_0, y_0] = [1, -1]$. We solve the HJI and HJB equations over $\Omega = (-2, 2)^2$ choosing the time step $dt = 2$ for the artificial term in the upwind scheme along with 60×60 spatial discretization points. Solving the HJI equation with $\lambda = 0.05$ and the Riccati solution as an initialization, we obtain the lowest value of γ as $\gamma^* = 7.9$, respectively $\gamma^* = 0.65$,



(i) (ii)

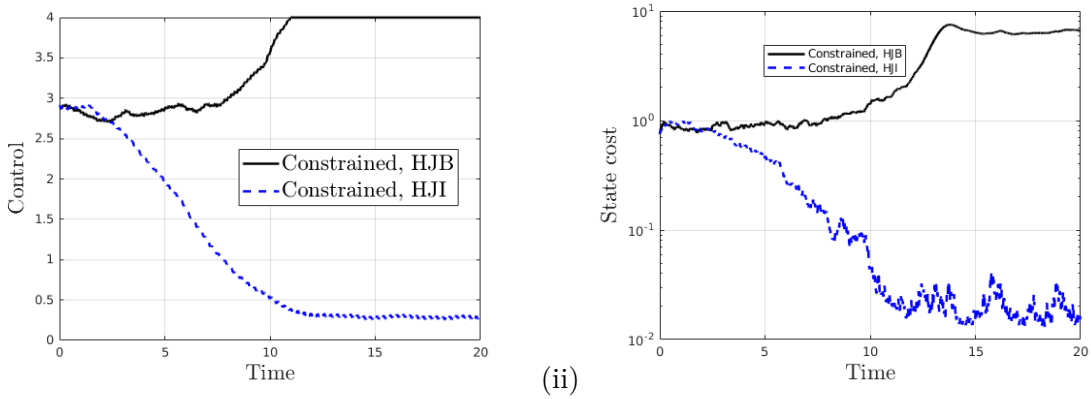
FIGURE 6. Test 3: Constrained case, comparison between the HJI and HJB synthesis with noise $w_2 = -0.6u_{HJI} + 0.01sin(10t)$, $-4 \leq u \leq 4$, $[x_0, y_0, z_0] = [-0.5, -0.5, -0.5]$, other parameters as before. **i)** Control, **ii)** State cost.



(i) (ii) (iii) (iv)

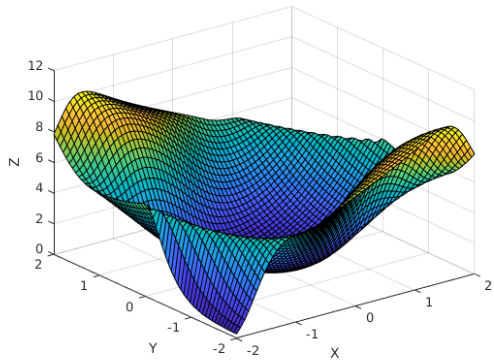
FIGURE 7. Test 3: Stochastic case, $[x_0, y_0, z_0] = [-1.5, -1.5, 1.5]$. **i)** State, **ii)** State cost $\|x(t)\|^2 + \|y(t)\|^2 + \|z(t)\|^2$, **iii)** Control, **iv)** Disturbance.

for the constrained and unconstrained control cases. The total cost becomes $\mathcal{J} = 4.69$ for the constrained control case, and $\mathcal{J} = 1.04$ for the unconstrained case.



(i) (ii)

FIGURE 8. Test 3: Constrained case, comparison between the HJI and HJB synthesis with noise $w_1 = -0.6u_{HJB} + 0.01sin(10t)$, $-4 \leq u \leq 4$, $[x_0, y_0, z_0] = [-0.5, -0.5, -0.5]$, other parameters as before. **i)** Control, **ii)** State cost.



(i)

FIGURE 9. Test 4: Deterministic case. **i)** Value function, constrained case.

The plot of the value function for the constrained case is documented in Figure 9. It suggests that the value function is not differentiable near boundary values of the domain, but that it is smooth in a neighborhood of the origin.

From Figure 10(i) we observe that the uncontrolled solution ($u = 0 = w$) of (4.8) does not settle at zero, but applying either constrained or unconstrained control, the controlled solutions converge to zero. The decay rate is faster for the unconstrained control case than for the constrained one. Constrained control is active up to time $t = 4$ and after that it steadily tends to zero, see Figure 10(ii). The state cost $\|x(t)\|^2 + \|y(t)\|^2$, documented in Figure 10(iii), does not settle to zero in the uncontrolled case, whereas both state and control cost change their transient behavior to settle at zero both for the constrained and unconstrained HJI synthesis given in Figure 10(iii)-(iv). The uppermost solid curve in Figure 10(v) is related to the convergence of the state cost for the constrained control case with $\gamma = \gamma^* = 7.9$. Further, if we increase γ i.e. for $\gamma > \gamma^*$, the state cost for the constrained HJI converges to the constrained HJB state cost which is also observed in Figure 10(v). Similar transient behavior for the convergence of the

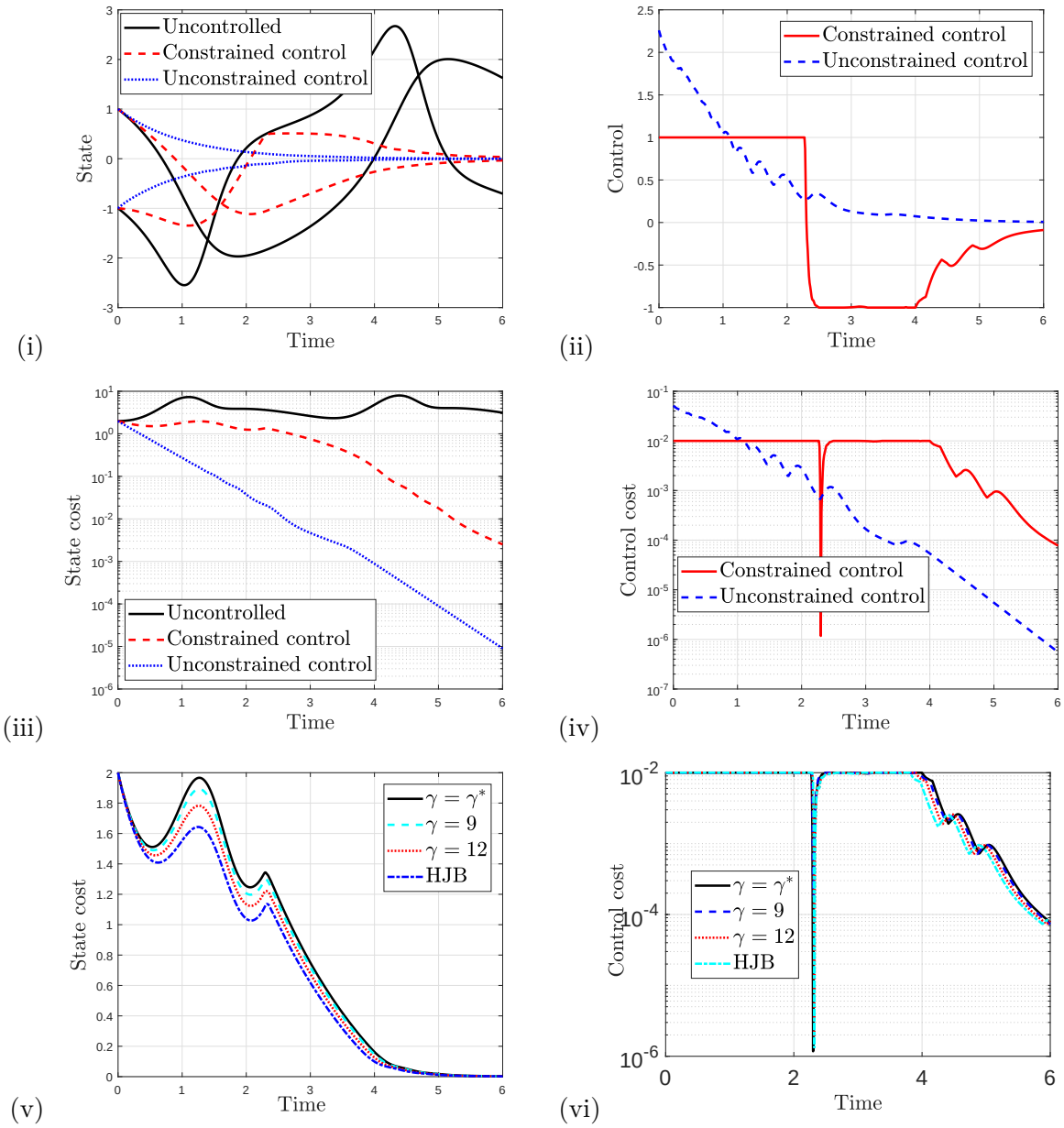
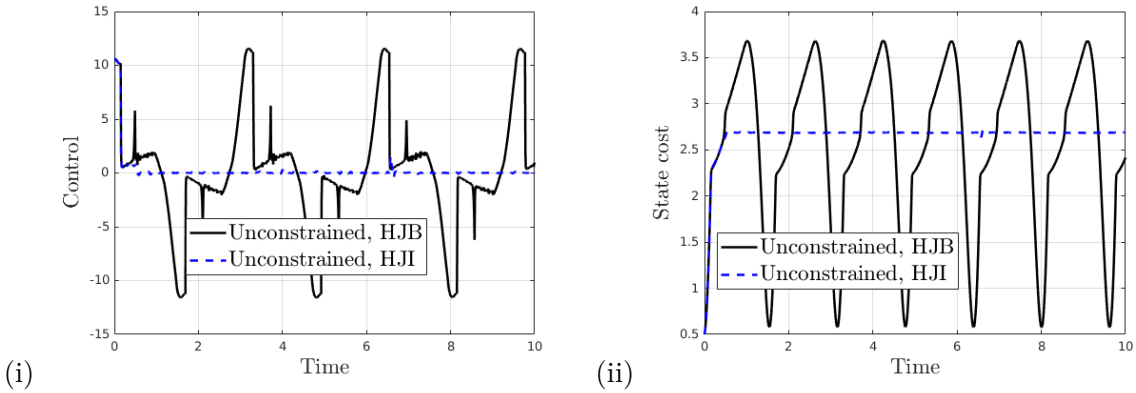


FIGURE 10. Test 4: Deterministic case. **i)** State, **ii)** Control, **iii)** State cost, **iv)** Control cost, **v)** State cost convergence of the HJI to HJB synthesis in the constrained case, **vi)** Control cost convergence of the HJI to HJB synthesis in the constrained case.

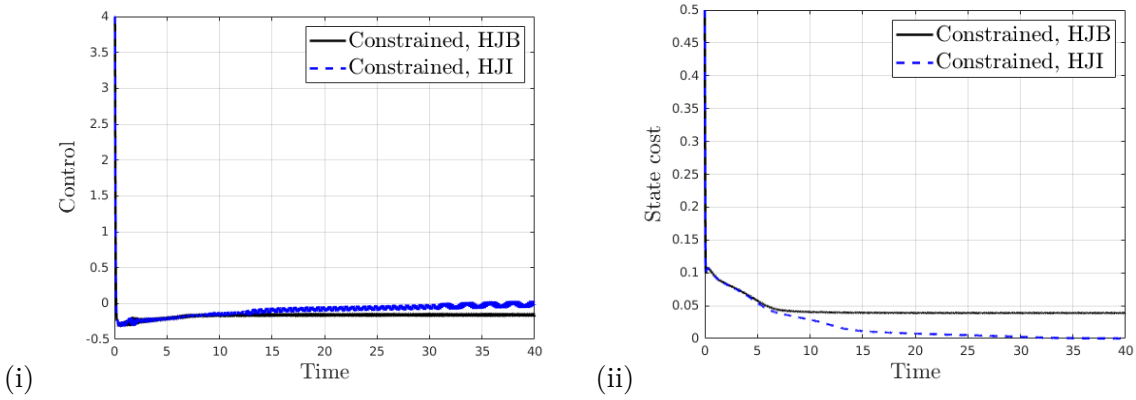
HJI constrained control cost to the HJB constrained control cost is shown in Figure 10(vi). We also observed similar behavior for the unconstrained control case.

Figures 11-12 show the temporal behavior of the HJI and HJB synthesis against the worst-case scenario in presence of different initial conditions. With different types of additional noise,



(i) (ii)

FIGURE 11. Test 4: Deterministic unconstrained case, comparison with the HJB case, $w_1 = -0.6u_{HJB} + 0.01sin(2t)$, $[x_0, y_0] = [-0.5, -0.5]$. **i)** Control, **ii)** State cost.



(i) (ii)

FIGURE 12. Test 4: Deterministic constrained case, comparison with the HJB case, $w_2 = 0.6u_{HJI} + 0.1sin(40t)$, $[x_0, y_0] = [-0.5, -0.5]$, $-4 \leq u \leq 4$. **i)** Control, **ii)** State cost.

comparisons of the HJI-HJB synthesis are discussed. Figure 11(i)-(ii) show that if the disturbance is of the form $w_1 = -0.6u_{HJB} + 0.01sin(2t)$, the control u and the running state cost $\|x\|^2 + \|y\|^2$ oscillate for the unconstrained HJB case; for the unconstrained HJI case they both are non-oscillatory. Concerning Figure 12, where the disturbance is of the form $w_2 = 0.6u_{HJI} + 0.1sin(40t)$ we set the control constraint as $-4 \leq u \leq 4$, since $-1 \leq u \leq 1$, for example, is too restrictive to observe the difference between the HJI and HJB case. We find that the constrained HJI synthesis tends to zero whereas the constrained HJB fails, which is documented in Figures 12(i)-(ii).

Concerning the second-order HJI equation, we consider

$$\min_{u(\cdot)\in\mathcal{U}} \max_{w(\cdot)\in\mathcal{W}} \mathcal{J}(u(\cdot), w(\cdot); (x_0, y_0)) = \mathbb{E}\int_0^\infty e^{-\lambda t}\Big(\|x\|^2 + \|y\|^2 + \|u\|_R^2 - \gamma^2 \|w\|_P^2 \Big) dt,$$

subject to the following stochastic Van der Pol system

$$
\begin{aligned}
&dx = y\,dt, \quad x(0) = x_0,\\
&dy = \Big((1-x^2)y - x + u + w\Big)\,dt + 0.1\,dW, \quad -1 \le u \le 1, \quad y(0) = y_0.
\end{aligned}
\tag{4.9}
$$

Here R, P, and dt are the same as in the deterministic case.
The dynamic behavior of engineering structures such as suspension bridges and offshore platforms, when subjected to random oscillations, is naturally described by second-order stochastic differential equations. Among these, variants of the Van der Pol oscillator play a crucial role, as their analysis is essential for developing robust control strategies in engineering and physical systems. These oscillators are central in the study of stability and bifurcations of stochastic dynamical systems. Within a game-theoretic framework, such systems can be interpreted as zero-sum stochastic differential games, where the controller seeks to stabilize the structure while nature or external disturbances act adversarially.

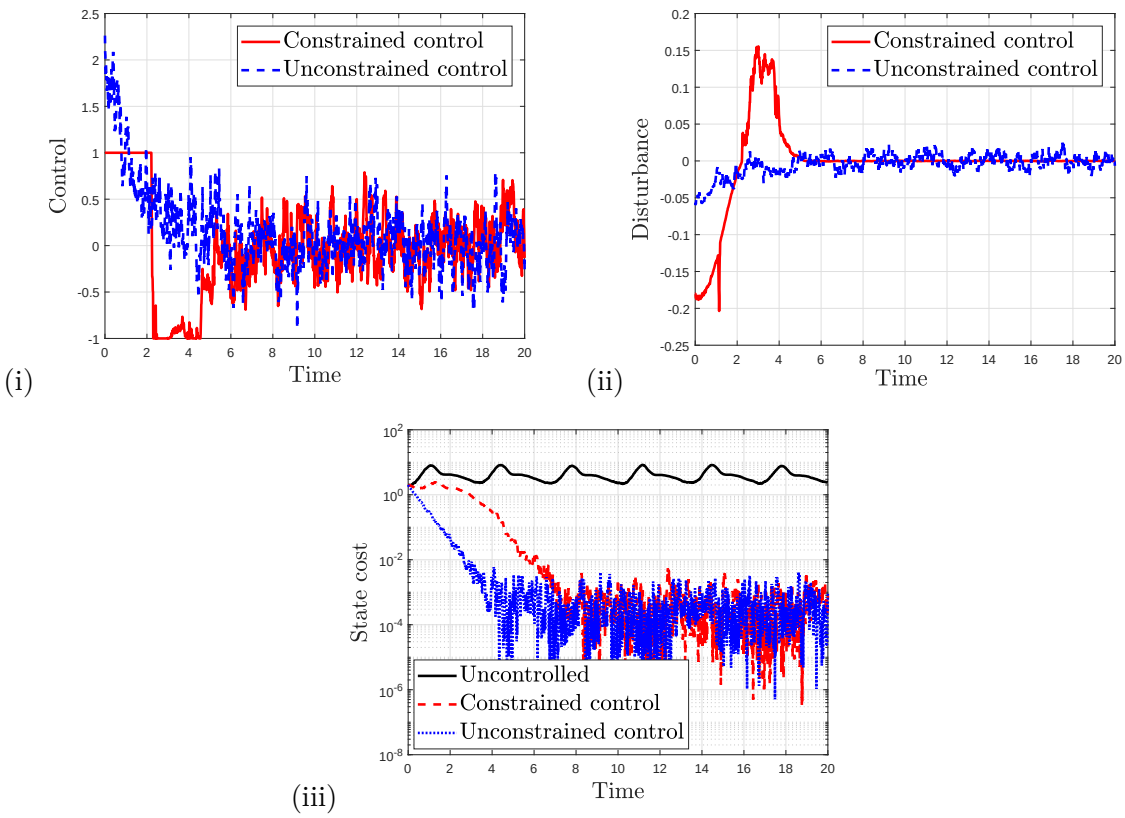


FIGURE 13. Test 4: Stochastic case, 2D Van der Pol equation. **i)** Control, **ii)** Disturbance, **iii)** State cost.

Solving the second-order HJI equation, we get $\gamma^* = 0.61$ for the unconstrained control case and $\gamma^* = 4.89$ for the constrained case. The constrained and unconstrained HJI controls under stochastic dynamics can be compared in Figure 13 (i). The corresponding state running cost converges to zero in both cases, with a faster decay rate for the unconstrained HJI case. The state cost for the uncontrolled and controlled cases can be compared in Figure 13 (iii). Figure

13 (ii) shows that the disturbance for both, the constrained and unconstrained HJI synthesis, settle to zero.

So far the discount factor was fixed at $\lambda = 0.005$. We next represent some results for the case when the solution for $\lambda = 0.005$ is used as initialization with $\lambda = 0$. For these calculations we choose $dt = 0.01$. In the unconstrained case, with $\gamma = 1$ we find $\|V_0 - V_{0.005}\|_{L^\infty(\Omega)} = 0.009$. Increasing λ to $\lambda = 0.05$, we found that the error increases to $\|V_0 - V_{0.05}\|_{L^\infty(\Omega)} = 0.08$. We also computed the optimal γ values, and obtained $\gamma^* = 0.59$ and $\gamma^* = 0.37$ for $\lambda = 0.005$ respectively $\lambda = 0$, along with the L^∞ error $\|V_0 - V_{0.005}\|_{L^\infty(\Omega)} = 0.07$. Similarly, in the constrained case we calculated V_0 and $V_{0.005}$ with $\gamma = 10$, keeping the other parameters the same as in the unconstrained case. We obtained $\|V_0 - V_{0.005}\|_{L^\infty(\Omega)} = 0.089$. Thus, at the expense of decreasing dt, successful results for $\lambda = 0$ were obtained when initializing the algorithm with solutions from $\lambda > 0$.

Concluding remarks

Policy iteration for two-player zero-sum differential games related to $\mathcal{H}_\infty$ control problem isdiscussed for both deterministic and stochastic systems. Using an implicit upwind scheme for backward PDEs, numerical results for the first and second-order HJI cases are presented with a detailed comparison for the constrained and unconstrained HJB synthesis. It would be of interest to extend the approach in the case of two-player nonzero-sum games, where instead of the HJI equation, we have coupled HJB equations corresponding to the so-called Nash equilibrium. Our convergence analysis relies on regularity assumptions (Assumptions 1 and 2), consistent with standard practice in the literature, though not generally verifiable from simple problem data. The example in Section 4 shows that the value function may fail to be differentiable everywhere, especially near domain boundaries. Establishing rigorous conditions that guarantee C^1 or C^2 differentiability of the value function is a promising research direction. Since only viscosity solutions can be expected in full generality, relaxing the C^1/C^2 requirements and establishing convergence of policy iteration in that framework remains a challenging open problem, which we identify as another interesting direction for future work.

Acknowledgments

The authors gratefully acknowledge support by the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research program. Sudeep Kundu also gratefully acknowledges the support of the Science & Engineering Research Board (SERB), Government of India, under the Start-up Research Grant, Project No. SRG/2022/000360.

References

[1] M. Abu-Khalaf and F. L. Lewis. *Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach*, Automatica J. IFAC **41**(5)(2005), 779–791.

[2] M. Abu-Khalaf, F. L. Lewis, and J. Huang. *Policy iterations on the Hamilton-Jacobi-Isaacs equation for H_∞ state feedback control with input saturation*, IEEE Trans. Automat. Control **51**(12)(2006), 1989–1995.

[3] Y. Achdou, J. Han, J. M. Lasry, P. L. Lions, and B. Moll. *Income and wealth distribution in macroeconomics: A continuous-time approach*, National Bureau of Economic Research, 2017.

[4] A. Alla, M. Falcone, and D. Kalise. *An efficient policy iteration algorithm for dynamic programming equations*, SIAM J. Sci. Comput. **37**(1)(2015), 181–200.

[5] A. Alla, M. Falcone, and L. Saluzzi. *An efficient DP algorithm on a tree-structure for finite horizon optimal control problems*, SIAM J. Sci. Comput. **41**(4)(2019), 2384–2406.

[6] M. Bachmayr, R. Schneider, and A. Uschmajew. *Tensor Networks and Hierarchical Tensors for the Solution of High-Dimensional Partial Differential Equations,* Found Comput Math. **16**(2016), 1423–1472.

[7] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, Birkhäuser Boston, 1997.

[8] T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory, 2nd Edition*, Society for Industrial and Applied Mathematics, (1998), doi:10.1137/1.9781611971132.

[9] R. W. Beard, G. N. Saridis, and J. T. Wen. *Approximate solutions to the Time-Invariant Hamilton-Jacobi-Bellman equation,* J. Optim. Theory Appl. **96**(3)(1998), 589–626.

[10] R. W. Beard and T. W. Mclain. *Successive Galerkin approximation algorithms for nonlinear optimal and robust control,* Internat. J. Control **71**(5)(1998), 717–743.

[11] D. P. Bertsekas. *Reinforcement Learning and Optimal Control* Athena Scientific, Belmont, Massachusetts, 2019.

[12] O. Bokanowski, S. Maroso, and H. Zidani. *Some Properties of Howards' Algorithm,* SIAM J. Numer. Anal. **47**(4)(2009), 3001–3026.

[13] J. F. Bonnans and H. Zidani. *Consistency of generalized finite difference schemes for the stochastic HJB equation*, SIAM J.Numer. Anal. **41**(3)(2003), 1008–1021.

[14] S. Boyd, V. Balakrishnan, and P. Kabamba. *A bisection method for computing the $\mathcal{H}_\infty$ norm of a transfer matrix and related problems,* Math. Control Signals Systems **2**(1989), 207–219.

[15] T. Breiten, K. Kunisch, and L. Pfeiffer. *Taylor expansions of the value function associated with a bilinear optimal control problem*, Ann. I. H. Poincare-AN **36**(2019), 1361–1399.

[16] S. Cacace, E. Cristiani, M. Falcone, and A. Picarelli. *A patchy dynamic programming scheme for a class of Hamilton-Jacobi-Bellman equations*, SIAM J. Sci. Comput. **34**(5)(2012), 2625–2649.

[17] P. Cannarsa and H. Frankowska. *Local regularity of the value function in optimal control*, Systems Control Lett. **62**(9)(2013), 791–794.

[18] T. Cheng, F. L. Lewis, and M. Abu-Khalaf. *Neural network solution for finite-horizon H_∞ state feedback control of nonlinear systems*, International Journal of Systems Science **38**(12)(2007), 1029–1041.

[19] M. G. Crandall and P. L. Lions. *Two approximations of solutions of Hamilton-Jacobi equations*, Math. Comp. **43**(167)(1984), 1–19.

[20] J. Darbon, G. P. Langlois, and T. Meng. *Overcoming the curse of dimensionality for some Hamilton–Jacobi partial differential equations via neural network architectures*, Res Math Sci **7**(3)(2020), https://doi.org/10.1007/s40687-020-00215-6.

[21] S. Dolgov, D. Kalise, and K. Kunisch. *Tensor Decompositions for High-dimensional Hamilton-Jacobi-Bellman Equations*, SIAM J. Sci. Comput. **43**(3)(2021), A1625–A1650.

[22] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A Francis. *State-space solutions to standard H_2 and H_∞ control problems*, IEEE Trans. Automat. Control **34**(8)(1989), 831–847.

[23] M. Falcone and R. Ferretti. *Semi-Lagrangian approximation schemes for linear and Hamilton-Jacobi equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2014.

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

[25] R. Goebel, *Convex optimal control problems with smooth Hamiltonians*, SIAM J. Control Optim. **43**(5)(2005), 1787–1811.

[26] R. Gonzalez and E. Rofman. *On deterministic control problems: an approximation procedure for the optimal cost. I. The stationary problem*, SIAM J. Control Optim. **23**(2)(1985), 242–266.

[27] K. Ito, C. Reisinger, and Y. Zhang. *A Neural Network-Based Policy Iteration Algorithm with Global H^2-Superlinear Convergence for Stochastic Games on Domains*, Found. Comput. Math. **21**(2)(2021), 331–374.

[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**(2)(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*, SIAM J. Appl. Dyn. Syst. **19**(2)(2020), 1496–1524.

[30] S. Kundu and K. Kunisch. *Policy iteration for Hamilton-Jacobi-Bellman equations with control constraints*, Comput. Optim. Appl. **87**(3) (2024), 785–809.

[31] A. Kröner, K. Kunisch, and H. Zidani. *Optimal feedback control of undamped wave equations by solving a HJB equation*, ESAIM: COCV, EDP Sciences **21**(2)(2014), 442–464.

[32] K. Kunisch and B. Priyasad. *Differentiability of the value function on $H^1(\Omega)$ of semilinear parabolic infinite time horizon optimal control problems under control constraints,* Math. Control Relat. Fields **14**(4)(2024), 1361–1390.

[33] K. Kunisch and D. Walter. *Semiglobal optimal feedback stabilization of autonomous systems via deep neural network approximation,* ESAIM Control Optim. Calc. Var. **27**(2021), https://doi.org/10.1051/cocv/2021009.

[34] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*, Springer-Verlag, New York (2001), xii+475.

[35] M. Lanchares and W. M. Haddad. *Nonlinear Optimal Control for Stochastic Dynamical Systems,* Mathematics **12**(2024), 647, https://doi.org/10.3390/math12050647.

[36] A. T. Lin, Y. T. Chow, and S. J. Osher. *A splitting method for overcoming the curse of dimensionality in Hamilton–Jacobi equations arising from nonlinear optimal control and differential games with applications to trajectory generation,* Commun. Math. Sci. **16**(7)(2018), 1933–1973.

[37] S. E. Lyshevski. *Optimal control of nonlinear continuous-time systems: design of bounded controllers via generalized nonquadratic functionals,* Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No. 98CH36207) **1**(1998), 205–209.

[38] X. Mao. *Stochastic differential equations and applications,* Elsevier, 2007.

[39] C. Pierre, *Introduction to differential games,* Universite de Brest, Place du Marchal De Lattre De Tassigny 75775 (2010).

[40] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces,* Applied Mathematical Sciences, Springer-Verlag, New York, 2003.

[41] C. Reisinger and Y. Zhang. *Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems,* Analysis and Applications **18**(6)(2020), 951–999.

[42] G. N. Saridis and C. S. G. Lee, *An approximation theory of optimal control for trainable manipulators,* IEEE Trans. Systems Man Cybernet **9**(3)(1979), 152–159.

[43] A. J. van der Schaft. *On a state space approach to nonlinear H_∞ control,* Systems Control Lett. **16**(1)(1991), 1–8.

[44] A. J. van der Schaft. *L_2-gain analysis of nonlinear systems and nonlinear state feedback H_∞ control,* IEEE Trans. Automat. Control **37**(6)(1992), 770–784.

[45] P. Soravia. *H_∞ control of nonlinear systems: differential games and viscosity solutions,* SIAM J. Control Optim. **34**(3)(1996), 1071–1097.

[46] P. Soravia. *Equivalence between nonlinear H_∞ control problems and existence of viscosity solutions of Hamilton-Jacobi-Isaacs equations,* Appl. Math. Optim. **39**(1)(1999), 17–32.

[47] E. Stefansson and Y. P. Leong. *Sequential Alternating Least Squares for Solving High Dimensional Linear Hamilton-Jacobi-Bellman Equations,* Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), doi: 10.1109/IROS.2016.7759553, 2016.

[48] F. Y. Wang and G. N. Saridis. *On successive approximation of optimal control of stochastic dynamic systems*, Modeling uncertainty, Internat. Ser. Oper. Res. Management Sci., Kluwer Acad. Publ., Boston, MA **46**(2002), 333–358.

[49] J. H. Witte and C. Reisinger. *A Penalty Method for the Numerical Solution of Hamilton-Jacobi-Bellman (HJB) Equations in Finance*, SIAM J. Numer. Anal. **49**(1)(2011), 213–231.

[50] H. J. Yang, M. Gim, and Y. Kim. *Solving nonconvex Hamilton–Jacobi–Isaacs equations with PINN-based policy iteration,* https://arxiv.org/abs/2507.15455.

[51] W. Zhang and B. S. Chen. *State feedback H_∞ control for a class of nonlinear stochastic systems*, SIAM J. Control Optim. **44**(6)(2006), 1973–1991.