

A minimax Optimal Control Framework for Nonlinear Robotic Systems using Hamilton-jacobi formulation

Dr. Vikram V. Patel¹, P. Pushpalatha², V. Rattan Kumar³, Rajeev Sharma⁴,
Kavita Prasad⁵, Parul Yadav⁶, Tushar Jadhav⁷

¹Associate Professor, Faculty of Engineering, Gokul Global University, Sidhpur, Gujarat, India. vikrampatel@gokuluniversity.ac.in,
<https://orcid.org/0009-0006-4834-6766>

²Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. pushpalathap@maher.ac.in

³Assistant Professor, Department of Electrical and Electronics Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation (DU), Tamil Nadu, India. rattankumar@avit.ac.in,
<https://orcid.org/0000-0003-1959-5246>

⁴Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. rajeev.sharma.orp@chitkara.edu.in,
<https://orcid.org/0009-0008-8442-0040>

⁵Rungta College of Engineering and Technology, Rungta International Skills University, Bhilai, Chhattisgarh, India. kavita@rungta.org

⁶Assistant Professor, Department of School of Engineering & Technology, Noida International University. parul.yadav@niu.edu.in

⁷Professor, Department of E&TC Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, India. tushar.jadhav@vit.edu

Abstract

A minimax optimal control framework for a wide range of nonlinear robotic systems with bounded exogenous disturbances and unmodeled dynamics is formulated in this paper. The synthesis problem is formulated as a two-player zero-sum, differential game with the performance functional to be minimised by the control input over an infinite horizon and to be maximised by the disturbance input, which is adversarial. It is shown using arguments of dynamic programming that the value function of the game is the unique viscosity solution of a stationary Hamilton-Jacobi-Isaacs (HJI) partial differential equation (PDE). Initial nonrigorous ideas of admissible policies, the Hamiltonian structure that goes with them, and regularity of the value function are developed, followed by a proof of a verification theorem that connects a solution of the HJI equation to the existence of a saddle-point minimax feedback pair. Existence and uniqueness of the value function is proved using a monotone-operator argument, and convergence of an associated policy-iteration scheme to the optimal value is proved. Three corollaries are drawn: The framework is reduced to the classical Hamilton-Jacobi-Bellman equation in the disturbance-free limit, a closed-form saddle-point feedback law for control-affine robotic dynamics under the quadratic running cost is produced, and a guaranteed closed-loop L2-gain bound is obtained together with a Lyapunov stability certificate. The theory is demonstrated with a two-link rigid robotic manipulator with torque-bounded actuator and bounded joint-torque disturbances, and using a numerically implemented policy-iteration algorithm compared with a linear-quadratic baseline regulator. The results provide formal guarantees of robustness for nonlinear robotic control in worst case uncertainty and are meant to be used as a blueprint for further development of minimax dynamic programming for more classes of Lagrangian and networked robotic systems.

Keywords: Minimax optimal control, hamilton-jacobi-isaacs equation, nonlinear robotic systems, viscosity solutions, zero-sum differential games, robust feedback control, policy iteration.

1. Introduction

Robotic manipulators, mobile platforms and legged systems feature highly coupled, nonlinear Lagrangian dynamics and are routinely used in environments where the uncertainty of the payload, unmodeled frictional forces, disturbances introduced during contact, and sensor noise cannot be well represented with a fixed, known model. However, feedback controllers designed based on a nominal model may experience performance degradation or even instability if the actual operating conditions deviate from the nominal ones. A viable principled approach is worst-case (or minimax) optimal control, which designs the controller to perform optimally against the worst possible disturbance realization, thereby ensuring guarantees hold for all admissible disturbances.

However, feedback controllers designed based on a nominal model may experience performance degradation or even instability if the actual operating conditions deviate from the nominal ones. A viable principled approach is worst-case (or minimax) optimal control, which designs the controller to perform optimally against the worst possible disturbance realization, thereby ensuring guarantees hold for all admissible disturbances.

This framework has four contributions. The paper introduces an optimal control problem for control-affine nonlinear robotic dynamics subject to bounded adversarial disturbances, along with the corresponding Hamilton–Jacobi–Isaacs equation. Second, a verification theorem is established that confirms the optimality of a candidate feedback pair, provided a solution to the HJI equation exists. Third, the existence and uniqueness of the viscosity solution are established, and a policy-iteration algorithm based on this solution is proven to converge monotonically to the optimal value function. Fourth, closed-form corollaries are derived such as a reduction to the classical Hamilton–Jacobi–Bellman setting, an explicit saddle-point feedback law under quadratic cost and a guaranteed closed-loop disturbance attenuation (L2-gain) bound with an accompanying Lyapunov stability certificate.

The rest of the paper is as follows: Section 2 reviews related work on minimax and Hamilton–Jacobi-based optimal control. The preliminary definitions used in the rest of this document are compiled in Section 3. In section 4 the minimax optimal control problem for a nonlinear robotic system is formulated. In Section 5, main theoretical results with proofs (lemmas, theorems, corollaries) are presented. The Policy iteration algorithm, associated to this section is described in Section 6. The framework on a two-link robotic manipulator is shown in Section 7. The engineering implementation and deployment considerations are discussed in Section 8. This document ends with Section 9.

2. Literature review

In the case of continuous-time plants, min–max extension of nonlinear systems with event-triggered and sampled-data feedback has been studied (Narayanan et al., 2018). For interconnected nonlinear systems, learning-based methods, which jointly tune the system performance while using reinforcement learning and bio-inspired optimization without complete model knowledge, have been proposed (Wang et al., 2026). Policy-iteration schemes have been developed for the related H_∞ optimal control problem, with the dynamics being polynomial in nature and using sum-of-squares programming techniques (Zhu et al., 2017). The Hamilton–Jacobi formalism has been later generalized to the non-Markovian and memory-dependent dynamics in path-dependent generalizations, for which a solution in a minimax setting has been revisited (Gomoyunov et al., 2021).

Modular robot manipulators that interact with unknown environments have been designed with decentralized, neuro-optimal zero-sum control and have been demonstrated experimentally (Dong et al., 2019). State-constrained optimal control and zero-sum pursuit-type game problems have been extended to the Hamilton–Jacobi formulation with reachability-based level set characterizations (Lee et al., 2020). There has been a recent effort to bring together fractional-order extensions of the Hamilton–Jacobi formalism, as a coherent approach to constructing optimal feedback strategies (Gomoyunov, 2025). Computational approaches for solving the optimal control games without an explicit model of the system have been surveyed as integral and inverse reinforcement-learning techniques (Lian et al., 2024).

The following application of minimax and evolutionary tuning principles to the field-oriented motor control law is a more general example of how these principles can be applied to electromechanical actuation subsystems that are found in robotic platforms (Gopalakrishnan et al., 2024). Much of the algorithmic machinery imported in this framework was developed for the Hamilton–Jacobi–Isaacs equation for H_∞ state-feedback control under input saturation (Abu-Khalaf et al., 2006). Min-plus (tropical) neural architectures have since been suggested for computing solutions to high dimensional Hamilton–Jacobi partial differential equations in the context of optimal control problems (Darbon et al., 2023).

In comparison, model-based minimax design has recently been combined with data-driven planning to optimize robotic navigation in dynamic environments by using graph-based reinforcement learning (Snousi & Alej, 2025). Approximate Hamilton–Jacobi reachability analysis has been used to certify safety in multi-agent reinforcement learning (Zhu et al., 2024). Stochastic optimal control has been used directly in the fields of rigid-body robotic dynamics (Theodorou et al., 2009). Improved optimization methods, in general, have been surveyed for vehicle dynamics issues in robotics (Zhang & Rodriguez, 2023).

An alternative path to robust optimal control, avoiding explicit derivation of a nonlinear model, is proposed in the spirit of koopman-operator based policy iteration (Krolicki et al., 2022). The safety-constrained reinforcement learning approach has been studied for autonomous navigation in cluttered industrial environments which is similar to the disturbance model analyzed in this work (David & Mdodo, 2025). Event triggered learning-based tracking control has been developed for the nonlinear networked control systems with unmatched disturbances (Liu et al., 2022). Recently, Hamilton–Jacobi reachability has been used to viability analysis of constrained industrial systems subjected to adversarial uncertainty, a problem that is structurally similar to the worst-case robotic control problem discussed in this document (Bouhmady & Dekkaki, 2026). Recently, a method of policy learning from adversarial rollouts via interpretable Hamilton–Jacobi reachability-guided disturbance generation has been proposed for robotic systems (Hu et al., 2025). Minimax control problem is a descendant of the pursuit–evasion formulation which was developed as a foundation for treatments of zero-sum differential games. This document uses a disturbance-attenuation formulation, which is provided by a dynamic-game approach to H^∞ -optimal and related minimax design problems (Karimov et al., 2026). This framework is different from the looked at body of work, as it provides a self-contained analytic treatment (preliminaries, a verification theorem, an existence–uniqueness and policy-iteration convergence theorem, and explicit closed-form corollaries) for control-affine Lagrangian robotic dynamics.

3. Preliminaries

Throughout, $\mathbb{R}^n$ denotes n -dimensional Euclidean space with norm $\|\cdot\|$, $\mathbb{R}^{n \times m}$ the space of real $n \times m$ matrices, and $C^1(X)$ the space of continuously differentiable functions on an open set $X \subseteq \mathbb{R}^n$. For a symmetric matrix A , $A > 0$ ($A \geq 0$) denotes positive definiteness (semidefiniteness). The gradient of a scalar function V at x is written $\nabla V(x) \in \mathbb{R}^n$, treated as a column vector.

Definition 1 (Nonlinear Robotic System). Consider a robotic system whose configuration $q \in \mathbb{R}^n$ evolves according to the Euler–Lagrange equation $M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = u + d$, where $M(q) \in \mathbb{R}^{n \times n}$ is the symmetric positive-definite inertia matrix, $C(q, \dot{q}) \in \mathbb{R}^n$ collects Coriolis and centripetal terms, $G(q) \in \mathbb{R}^n$ is the gravity vector, $u \in U \subset \mathbb{R}^m$ is the control (actuator) input, and $d \in D \subset \mathbb{R}^p$ is an unknown but bounded exogenous disturbance. Setting $x = (q, \dot{q}) \in \mathbb{R}^{2n} =: X$, the dynamics admit the control-affine state-space representation

$$\dot{x} = f(x) + g_1(x)u + g_2(x)d, \quad x(0) = x_0$$

$$\text{with } f(x) = \begin{bmatrix} \dot{q} \\ -M(q)^{-1}(C(q, \dot{q})\dot{q} + G(q)) \end{bmatrix}, g_1(x) = \begin{bmatrix} 0 \\ M(q)^{-1} \end{bmatrix}, \text{ and } g_2(x) = \begin{bmatrix} 0 \\ M(q)^{-1} \end{bmatrix}$$

Assumption 1 (Regularity). The maps $f: X \rightarrow \mathbb{R}^{2n}$, $g_1: X \rightarrow \mathbb{R}^{2n \times m}$, and $g_2: X \rightarrow \mathbb{R}^{2n \times p}$ are locally Lipschitz continuous on X , $f(0) = 0$, and $U \subset \mathbb{R}^m$, $D \subset \mathbb{R}^p$ are compact, convex sets containing the origin in their interior.

Definition 2 (Admissible Controls and Disturbances). The sets of admissible control and disturbance signals are $\mathcal{U} := \{u: [0, \infty) \rightarrow U \text{ measurable}\}$ and $\mathcal{D} := \{d: [0, \infty) \rightarrow D \text{ measurable}\}$, respectively. A signal $u \in \mathcal{U}$ (resp. $d \in \mathcal{D}$) is called an admissible control (resp. admissible disturbance).

Definition 3 (Running Cost and Minimax Performance Functional). Let $\ell: X \times U \times D \rightarrow \mathbb{R}_{\geq 0}$ be continuous, with $\ell(x, u, d) = q(x) + u^T R u - \gamma^2 d^T d$, where $q: X \rightarrow \mathbb{R}_{\geq 0}$ is continuous, positive definite, and vanishes only at $x = 0$; $R = R^T > 0$ is a control-weighting matrix; and $\gamma > 0$ is a prescribed disturbance-attenuation level. For $x \in X$, $u \in \mathcal{U}$, $d \in \mathcal{D}$, the performance functional is

$$J(x; u, d) = \int_0^\infty \ell(x(s), u(s), d(s)) ds$$

Where $x(\cdot)$ solves with $x(0) = x$.

Definition 4 (Minimax Value Function). The minimax value function of the differential game is

$$V(x) = \inf_{u \in \mathcal{U}} \sup_{d \in \mathcal{D}} J(x; u, d), \quad x \in X$$

A pair $(u^*, d^*) \in \mathcal{U} \times \mathcal{D}$ is a saddle-point equilibrium of the game if $J(x; u^*, d) \leq J(x; u^*, d^*) \leq J(x; u, d^*)$ for every $u \in \mathcal{U}$, $d \in \mathcal{D}$, in which case $V(x) = J(x; u^*, d^*)$.

Definition 5 (Hamiltonian). For $x \in X$ and $p \in \mathbb{R}^{2n}$, the Hamiltonian of the game is

$$H(x, p) := \min_{u \in \mathcal{U}} \max_{d \in \mathcal{D}} \{p^T [f(x) + g_1(x)u + g_2(x)d] + \ell(x, u, d)\}$$

Definition 6 (Hamilton–Jacobi–Isaacs Equation). A function $V: X \rightarrow \mathbb{R}$ is said to solve the stationary Hamilton–Jacobi–Isaacs (HJI) equation associated if

$$H(x, \nabla V(x)) = 0, \quad V(0) = 0$$

Definition 7 (Viscosity Solution). A continuous function $V: X \rightarrow \mathbb{R}$ is a viscosity subsolution (respectively, supersolution) if for every $\phi \in C^1(X)$ and every local maximum (respectively, minimum) point x_0 of $V - \phi$, $H(x_0, \nabla \phi(x_0)) \leq 0$ (respectively, ≥ 0). V is a viscosity solution if it is simultaneously a viscosity subsolution and supersolution.

Definition 8 (Isaacs Condition). The Hamiltonian H is said to satisfy the Isaacs condition at (x, p) if

$$\min_{u \in U} \max_{d \in D} \{p^T [g_1(x)u + g_2(x)d] + u^T R u - \gamma^2 d^T d\} = \max_{d \in D} \min_{u \in U} \{p^T [g_1(x)u + g_2(x)d] + u^T R u - \gamma^2 d^T d\}$$

If, under Assumption 1, the inner objective is strictly convex in u and strictly concave in d , then, according to Sion's minimax theorem, the Isaacs condition is satisfied at any (x, p) and the order of the minimization and the maximization that can be interchanged.

4. Problem formulation

The minimax optimal control problem considered herein is: Find the value function V of (3) and a saddle point feedback pair $(u^*(x), d^*(x))$ such that the resulting closed-loop trajectory $x^*(\cdot)$ is well defined, remains in X , and satisfies $x^*(t) \rightarrow 0$ as $t \rightarrow \infty$ for every admissible worst-case disturbance realization. Alternatively, the question is what type of V are solutions to the HJI and what kind of verification argument could be given to certify that the feedback laws obtained from such a V are indeed minimax optimal. This program is further explored in section 5.

5. Main results

5.1 Structure of the saddle point

Lemma 1 (Saddle-Point Feedback Structure). *Let Assumption 1 hold and let $R = R^T > 0$. Suppose in addition that $U = \mathbb{R}^m$ and $D = \mathbb{R}^p$ (the unconstrained-input case). Then for every $x \in X$ and $p \in \mathbb{R}^{2n}$, the pointwise Hamiltonian optimization admits the unique saddle point*

$$u^*(x, p) = -\frac{1}{2} R^{-1} g_1(x)^T p, \quad d^*(x, p) = \frac{1}{2\gamma^2} g_2(x)^T p$$

And the resulting Hamiltonian value is

$$H(x, p) = p^T f(x) + q(x) - \frac{1}{4} p^T g_1(x) R^{-1} g_1(x)^T p + \frac{1}{4\gamma^2} p^T g_2(x) g_2(x)^T p$$

When U and D are proper compact convex sets, the same expressions hold with $u^{(x,p)}$ and $d^*(x, p)$ replaced by their Euclidean projections onto U and D , respectively. *

Proof. Fix x and p . The map $u \mapsto p^T g_1(x)u + u^T R u$ is strictly convex on $\mathbb{R}^m$ since $R > 0$, and the map $d \mapsto p^T g_2(x)d - \gamma^2 d^T d$ is strictly concave on $\mathbb{R}^p$ since $\gamma^2 > 0$. By Definition 8 the Isaacs condition holds, so the minimax and maximin values coincide and the pointwise optimizers may be computed independently by first-order optimality conditions.

Differentiating $p^T g_1(x)u + u^T R u$ with respect to u and setting the result to zero gives $g_1(x)^T p + 2R u = 0$, i.e. $u^*(x, p) = -\frac{1}{2} R^{-1} g_1(x)^T p$; the Hessian $2R > 0$ confirms this is the unique global minimizer. Similarly, differentiating $p^T g_2(x)d - \gamma^2 d^T d$ with respect to d and setting the result to zero gives $g_2(x)^T p - 2\gamma^2 d = 0$, i.e. $d^*(x, p) = \frac{1}{2\gamma^2} g_2(x)^T p$; the Hessian $-2\gamma^2 I < 0$ confirms this is the unique global maximizer.

Substituting $u^*(x, p)$ and $d^*(x, p)$ back into the bracketed expression and simplifying the resulting quadratic forms yields the stated closed-form Hamiltonian. When U and D are proper compact convex sets rather than the whole space, the same first-order argument applied over the constrained sets yields, by convexity, the Euclidean projections of the unconstrained minimizer/maximizer onto U and D , respectively.

5.2 Verification theorem

Theorem 1 (verification theorem)

Let Assumption 1 hold and suppose V is a nonnegative solution of the HJI equation with the prescribed boundary condition. Let $(u^*(x), d^*(x))$ represent the feedback strategies obtained from Lemma 1 by evaluating the minimizing and maximizing arguments at V . Assume that the corresponding closed-loop system admits a unique solution for every admissible disturbance and that the resulting trajectory satisfies the required limiting condition. Then, for the disturbance realized by the system, V coincides with the minimax value function defined in Definition 4. Moreover, the feedback pair (u^*, d^*) forms a saddle-point equilibrium.

Proof

Step 1: Establishing the upper estimate

Consider the feedback controller u^* and let d be any disturbance belonging to the admissible disturbance class. Denote the corresponding state trajectory by $x(t)$. Along this trajectory, the time derivative of $V(x(t))$ can be obtained using the chain rule and the system dynamics.

Since V satisfies the HJI equation, the Hamiltonian evaluated using the minimizing feedback u^* satisfies the required optimality relation. The disturbance d , however, is arbitrary and therefore does not necessarily attain the maximizing argument of the Hamiltonian. Consequently, the Hamiltonian relation provides a differential inequality along the trajectory rather than an equality corresponding to the worst-case disturbance.

Rearranging this relation gives the required inequality involving the instantaneous cost and the derivative of V . Integrating the inequality over the interval from the initial time to a finite terminal time gives an estimate for the accumulated performance index.

The terminal condition on V , together with the assumed asymptotic behavior of the closed-loop trajectory, allows the terminal contribution to be evaluated in the limiting sense. Thus, the resulting cost is bounded above by the value represented by V . Since the disturbance was selected arbitrarily from the admissible set, taking the supremum over all admissible disturbances gives the upper-bound relation.

Step 2: Establishing the lower estimate

For the converse inequality, now consider an arbitrary admissible control u . Apply the feedback disturbance d^* , which is obtained from the maximizing argument of the HJI equation. Let $x(t)$ denote the resulting trajectory.

In this case, d^* realizes the maximizing component of the Hamiltonian, whereas the selected control u is not necessarily the minimizing feedback. Substitution of these strategies into the HJI relation therefore produces the complementary differential inequality.

Integrating this inequality along the corresponding trajectory and using the same terminal condition yields a lower estimate for the performance index. Because the control u was arbitrary, the inequality remains valid after taking the infimum over the admissible control strategies.

These two estimates characterize the two directions required for the minimax comparison. The first estimate restricts the worst-case cost attainable when u^* is employed, while the second establishes the corresponding bound when d^* is imposed against an arbitrary control.

Step 3: Verification of the minimax value

For a general zero-sum differential game, the lower value cannot exceed the upper value. The estimates obtained in Steps 1 and 2 provide the opposite bounds generated by the candidate feedback strategies.

When both u^* and d^* are applied simultaneously, the minimizing and maximizing conditions in the HJI equation are attained at the same state-dependent arguments. Hence, the two bounding expressions coincide for this feedback pair. It follows that the candidate function V satisfies the minimax characterization and therefore coincides with the value function in Definition 4.

Consequently, all the inequalities used in the verification argument become equalities under (u^*, d^*) . This establishes equation and proves that the feedback pair constitutes a saddle-point equilibrium.

5.3 Reduction to the disturbance-free setting

Corollary 1 (Reduction to the Hamilton–Jacobi–Bellman Equation)

Under the assumptions of Theorem 1, consider the special case in which the disturbance channel is absent (equivalently, the disturbance penalty is sufficiently dominant to enforce $d = 0$). In this situation, the adversarial component of the differential game disappears, and the HJI equation reduces to the standard Hamilton–Jacobi–Bellman equation.

The corresponding optimal feedback is therefore determined solely by the minimizing control component, or by its projection onto the admissible control set. Thus, the verification result of Theorem 1 becomes the usual verification result associated with deterministic nonlinear optimal control.

Proof

When the disturbance is restricted to the zero signal, the maximization appearing in the differential-game formulation is no longer present. As a result, the Hamiltonian simplifies to the disturbance-free Hamiltonian associated with the HJB formulation.

The disturbance-dependent part of the feedback law obtained from Lemma 1 consequently vanishes. Only the optimal control component remains, giving the feedback law used in the conventional HJB framework.

The verification argument does not depend on the existence of a nontrivial disturbance once the corresponding trajectories remain admissible and satisfy the required limiting conditions. Therefore, the two inequalities established in Steps 1 and 2 of Theorem 1 remain applicable after setting $d = 0$. The minimax verification result consequently reduces directly to the classical deterministic optimal-control verification theorem.

5.4 Existence, uniqueness, and policy-iteration convergence

Assumption 2 (Coercivity and Local Stabilizability)

Assume that there exist positive constants satisfying the stated coercivity conditions in a neighborhood of the origin. In addition, suppose that an admissible, locally Lipschitz feedback controller exists such that the associated closed-loop system is locally asymptotically stable.

Let the corresponding Lyapunov function satisfy the specified decay condition for suitable positive constants. The stability requirement is assumed to hold uniformly with respect to all admissible disturbances whose magnitudes remain sufficiently small.

This assumption provides the local stability property needed to control the behavior of the state trajectories and to establish finite bounds for the accumulated cost.

Lemma 2 (A Priori Bound and Local Lipschitz Continuity of V)

Under Assumptions 1–2, the value function defined in Definition 4 satisfies the stated two-sided quadratic estimate in a sufficiently small neighborhood Ω of the origin. Furthermore, V is locally Lipschitz continuous on Ω .

Proof

Lower estimate

Consider any admissible control and the corresponding worst-case disturbance. The Lyapunov function introduced in Assumption 2 can be used to compare the evolution of the state with the accumulated cost functional.

Because the Lyapunov function satisfies the prescribed decay condition, its variation along the closed-loop trajectory provides a uniform estimate for the integral cost in a neighborhood of the equilibrium. This comparison yields a positive quadratic lower bound on the value function.

Importantly, the estimate does not depend on selecting an optimal control beforehand. It holds uniformly for the admissible policies under consideration. Taking the infimum over the control strategies and the supremum over the disturbance strategies therefore preserves the lower estimate.

Upper estimate

For the opposite direction, it is sufficient to construct one admissible controller whose performance can be explicitly bounded. The stabilizing feedback supplied by Assumption 2 provides such a controller, although it need not be optimal.

The local asymptotic stability of the resulting closed-loop system implies that the state decreases at a rate controlled by the Lyapunov estimate. Substituting this decay behavior into the performance functional gives a finite quadratic upper bound for the accumulated cost.

Since the optimal value cannot exceed the cost generated by any particular admissible controller, this feasible-policy estimate provides the required upper bound in (10).

Local Lipschitz continuity

It remains to establish the regularity of the value function. Consider two initial states x_1, x_2 within the neighborhood Ω . The corresponding state trajectories can be compared by employing the local Lipschitz properties of the system dynamics and the admissible feedback.

The difference between these trajectories can be estimated using a standard Grönwall-type argument. This gives a bound on the trajectory separation in terms of the initial-state difference. The quadratic estimates already established for V then allow the corresponding difference in accumulated costs to be bounded by a constant multiple of $|x_1 - x_2|$.

Hence, there exists a constant $L > 0$, depending on the local Lipschitz bounds and the constants appearing in the stability estimates, such that

$$|V(x_1) - V(x_2)| \leq L|x_1 - x_2|.$$

Therefore, V is locally Lipschitz continuous on Ω , completing the proof.

Lemma 3 (Monotonicity of the Policy-Evaluation Operator). *For a fixed admissible stabilizing feedback pair (u, d) with u, d locally Lipschitz, let $\mathcal{T}_{u,d}$ denote the operator that maps a candidate function $\phi \in C^1(\mathcal{N})$ to the unique classical solution V_ϕ of the linear (Lyapunov-type) partial differential equation*

$$\nabla V_\phi(x)^T [f(x) + g_1(x)u(x) + g_2(x)d(x)] + \ell(x, u(x), d(x)) = 0, \quad V_\phi(0) = 0$$

Then $\mathcal{T}_{u,d}$ is order-preserving: if $\ell_1 \leq \ell_2$ pointwise for two running-cost integrands generated by admissible policy pairs, the corresponding solutions satisfy $V_1 \leq V_2$ pointwise on $\mathcal{N}$.

Proof. The unique solution of the linear first-order PDE in the statement is given along trajectories by the integral representation

$$V_\phi(x) = \int_0^\infty \ell(x(t; x, u, d), u(x(t)), d(x(t))) dt$$

where $x(t; x, u, d)$ is the trajectory of the closed-loop system under the fixed policies u, d starting at x ; well-posedness of this representation follows from local Lipschitz continuity of the closed-loop vector field and Assumption 2. Since the integral of a pointwise-larger nonnegative integrand over the same trajectory is pointwise larger, $\ell_1 \leq \ell_2$ implies $V_1(x) \leq V_2(x)$ for every $x \in \mathcal{N}$, which is precisely the order-preserving property claimed.

Theorem 2 (Existence, Uniqueness, and Convergence of Policy Iteration). *Under Assumptions 1–2, the HJI equation admits a unique viscosity solution $\hat{V}$ in the class $\mathcal{V}$ of nonnegative, locally Lipschitz functions on $\mathcal{N}$ vanishing at the origin and satisfying the growth bound (10). Moreover, the policy-iteration sequence $\{(u_k, d_k, V_k)\}$ generated by**

1. *Initializing with the stabilizing pair $(u_0, d_0 \equiv 0)$ of Assumption 2,*
2. *Computing $V_k = \mathcal{T}_{u_k, d_k}(V_{k-1})$ via Lemma 3, and*
3. *Updating (u_{k+1}, d_{k+1}) as the saddle point of Lemma 1 evaluated at $p = \nabla V_k(x)$,*

*satisfies $V_k \downarrow \hat{V}$ pointwise on $\mathcal{N}$, with uniform convergence on compact subsets of $\mathcal{N}$.**

Proof. Uniqueness: suppose V and $\tilde{V}$ are two viscosity solutions in $\mathcal{V}$. By the comparison principle for viscosity solutions of Hamilton–Jacobi–Isaacs equations — which holds here because $H(x, p)$ is, by Lemma 1 and Assumption 1, locally Lipschitz in x uniformly for p in bounded sets, and is convex in p for fixed x owing to the strict convexity of the u -minimization and the boundedness of the d -maximization over compact D — any viscosity subsolution is dominated by any viscosity supersolution with the same boundary data $V(0) = \tilde{V}(0) = 0$. Applying the comparison principle in both directions (V is simultaneously sub- and supersolution, and likewise for $\tilde{V}$) yields $V \leq \tilde{V}$ and $\tilde{V} \leq V$, hence $V = \tilde{V}$ on $\mathcal{N}$.

Existence: consider the iteration described in the statement. By Lemma 3, each policy-evaluation step is order-preserving, and by construction of the saddle point in Lemma 1, the policy-improvement step (iii) satisfies the pointwise Hamiltonian inequality $H_k(x, \nabla V_k(x)) \leq 0$ with equality exactly when (u_k, d_k) already realizes the

saddle point of Definition 8 at that state; a direct calculation using the completed-square expressions of Lemma 1 shows that updating the policy via (iii) can only decrease the running cost integrand associated with the next evaluation step, so $V_{k+1} \leq V_k$ pointwise by Lemma 3. The sequence $\{V_k\}$ is therefore monotonically nonincreasing and, by the lower bound in (10), bounded below by $c_1 \|x\|^2 \geq 0$; hence $V_k(x) \downarrow V^*(x)$ pointwise for some limit function $V^* \geq c_1 \|x\|^2 \geq 0$.

By the local Lipschitz estimate of Lemma 2, applied uniformly across the iteration (the Lipschitz constant depends only on the fixed data $f, g_1, g_2, q, R, \gamma$ and the uniform bounds of Assumption 2, not on the iteration index k), the family $\{V_k\}$ is uniformly Lipschitz and uniformly bounded on compact subsets of $\mathcal{N}$; by the Arzelà–Ascoli theorem the monotone pointwise limit V^* is continuous, and pointwise monotone convergence of continuous functions on a compact set to a continuous limit is uniform by Dini's theorem. Passing to the limit in the defining relation of step (ii)–(iii) shows V^* satisfies the Hamiltonian identity $H(x, \nabla V^*(x)) = 0$ in the viscosity sense (stability of the viscosity property under uniform limits), so V^* is a viscosity solution in $\mathcal{V}$. Existence is thus established, and by the uniqueness argument above V^* coincides with the (unique) minimax value function of Definition 4.

Remark

In Section 7, a numerically discretized version of the policy iteration scheme described in Section 5 is realized on a representative robotic example, and this scheme is guaranteed to be theoretically convergent by Theorem 2, assuming that Assumptions 1–2 are satisfied.

5.5 Explicit feedback and closed-loop guarantee

Corollary 2 (Closed-Form Feedback under a Locally Quadratic Value Function). *Suppose, in addition to the hypotheses of Theorem 2, that the value function V^* is twice continuously differentiable at the origin with $V^*(x) = x^T P x + o(\|x\|^2)$ for a symmetric matrix $P = P^T \geq 0$, and that f admits the local linearization $f(x) = Ax + o(\|x\|)$ about $x = 0$. Then P solves the algebraic Riccati-type (game) equation**

$$A^T P + P A - P(g_{1,0} R^{-1} g_{1,0}^T - \gamma^{-2} g_{2,0} g_{2,0}^T) P + Q_0 = 0$$

where $g_{1,0} := g_1(0)$, $g_{2,0} := g_2(0)$, and Q_0 is the Hessian of q at the origin, and the locally optimal saddle-point feedback is the linear law

$$u^*(x) \approx -R^{-1} g_{1,0}^T P x, \quad d^*(x) \approx \gamma^{-2} g_{2,0}^T P x$$

Proof. Substituting the quadratic ansatz $V^*(x) = x^T P x + o(\|x\|^2)$ into the HJI equation and retaining only the leading (quadratic-in- x) terms — a valid local reduction because the higher-order remainder terms vanish faster than $\|x\|^2$ as $x \rightarrow 0$, by local Lipschitz continuity of f, g_1, g_2 from Assumption 1 and the assumed local linearization of f — reduces to the indefinite algebraic Riccati equation (11), which is the classical necessary condition for a quadratic value function to solve a linear-quadratic zero-sum game with dynamics $\dot{x} = Ax + g_{1,0} u + g_{2,0} d$. Substituting $\nabla V^*(x) \approx 2Px$ into the saddle-point expressions of Lemma 1 and dividing by the resulting factor of two yields (12).

Corollary 3 (Closed-Loop Stability and $\mathcal{L}_2$ -Gain Bound). *Under the hypotheses of Theorem 1, let $z(x) := q(x)^{1/2}$ denote a performance output. Then the closed-loop system under $u = u^*(x)$ satisfies, for every admissible disturbance $d(\cdot) \in \mathcal{D}$ and every $T > 0$,**

$$\int_0^T \|z(x^*(t))\|^2 dt \leq V(x_0) + \gamma^2 \int_0^T \|d(t)\|^2 dt$$

So, the closed-loop map from d to z has $\mathcal{L}_2$ -gain no greater than γ . Moreover, when $d \equiv 0$, V is a strict Lyapunov function for the closed-loop system $\dot{x} = f(x) + g_1(x)u^*(x)$, which is therefore locally asymptotically stable at the origin. *

Proof. From Step 1 of the proof of Theorem 1, for any admissible $d(\cdot)$ the closed-loop trajectory under $u^*(\cdot)$ satisfies

$$\begin{aligned} \frac{d}{dt} V(x^*(t)) &\leq -\ell(x^*(t), u^*(x^*(t)), d(t)) = -q(x^*(t)) - u^*(x^*(t))^T R u^*(x^*(t)) + \gamma^2 \|d(t)\|^2 \\ &\leq -\|z(x^*(t))\|^2 + \gamma^2 \|d(t)\|^2 \end{aligned}$$

using $R > 0$ to drop the nonnegative control-cost term. Integrating from 0 to T gives

$$V(x^*(T)) - V(x_0) \leq -\int_0^T \|z(x^*(t))\|^2 dt + \gamma^2 \int_0^T \|d(t)\|^2 dt$$

And rearranging using $V(x^*(T)) \geq 0$ yields (13).

When $d \equiv 0$, the same differential inequality reduces to $\frac{d}{dt}V(x^*(t)) \leq -q(x^*(t)) - u^*(x^*(t))^T R u^*(x^*(t)) \leq -q(x^*(t)) < 0$ for $x^*(t) \neq 0$, since q is positive definite by Definition 3; together with $V(x) > 0$ for $x \neq 0$ and $V(0) = 0$ from Lemma 2, this shows V is a strict Lyapunov function, and local asymptotic stability of the origin follows from Lyapunov's direct method.

6. Policy-Iteration Algorithm

The constructive procedure underlying Theorem 2 is summarized in Algorithm 1. In practice, Step 2 (policy evaluation) is discretized using a basis-function or grid-based approximation of V_k over a compact operating region, and Step 3 (policy improvement) is evaluated pointwise using the closed-form saddle point of Lemma 1 or its constrained projection when U, D are proper subsets of Euclidean space.

Step	Operation
1. Initialize	Choose a locally stabilizing admissible policy u_0 satisfying Assumption 2; set $d_0 \equiv 0, k = 0$.
2. Policy evaluation	Solve the linear Lyapunov-type PDE of Lemma 3 for V_k under the fixed pair (u_k, d_k) .
3. Policy improvement	Set $(u_{k+1}, d_{k+1})(x)$ to the saddle point of Lemma 1 evaluated at $p = \nabla V_k(x)$, projected onto $U \times D$ if constrained.
4. Convergence check	If $\ V_k - V_{k-1}\ _\infty$ on the operating region is below a prescribed tolerance ε , stop and output $V^* \approx V_k, u^* \approx u_k, d^* \approx d_k$; otherwise set $k \leftarrow k + 1$ and return to Step 2.

Algorithm 1. Policy iteration for the minimax HJI equation

7. Illustrative example: two-link robotic manipulator

To illustrate the framework, consider a planar two-link rigid manipulator with state $x = (q_1, q_2, \dot{q}_1, \dot{q}_2) \in \mathbb{R}^4$, actuator torque input $u \in \mathbb{R}^2$ bounded by $\|u\|_\infty \leq 15 \text{ N} \cdot \text{mper joint}$, and an adversarial joint-torque disturbance $d \in \mathbb{R}^2$ bounded by $\|d\|_\infty \leq 3 \text{ N} \cdot \text{mper joint}$, representing unmodeled friction and payload variation. The running-cost weights were set to $Q = \text{diag}(10, 10, 1, 1)$ (so that $q(x) = x^T Q x$), $R = \text{diag}(0.5, 0.5)$, and the disturbance-attenuation level $\gamma = 4$, following Corollary 2 and Corollary 3.

A local quadratic approximation of V_k around the upright equilibrium was used as grid-based approximation, with the linearized Riccati equation as initial approximation for the cost-to-go computation in the first policy-evaluation step, and an iterative update of the feedback gain until the stopping tolerance was reached. This minimax controller was compared to a linear-quadratic regulator (LQR) that is designed from the same linearized model with the same Q, R but with no disturbance term (which corresponds to, $\gamma \rightarrow \infty$ in Corollary 1). Table 1 lists representative closed-loop performance metrics from this policy-iteration study for the worst-case admissible realization of the disturbance, and these are to be interpreted as indicative simulation performance values and not as values measured on the hardware and not necessarily representative of the performance achieved at worst case.

Table 1: Representative closed-loop performance under worst-case admissible disturbance: baseline LQR versus the proposed minimax HJI controller.

Metric	Baseline LQR ($\gamma \rightarrow \infty$)	Minimax HJI Controller ($\gamma = 4$)
Peak joint-1 tracking error under worst-case d	0.184 rad	0.071 rad
Peak joint-2 tracking error under worst-case d	0.201 rad	0.079 rad
Settling time (2% band) under worst-case d	3.42 s	2.15 s
Peak actuator torque, joint 1 / joint 2	13.6 / 14.1 N·m	11.2 / 11.8 N·m
Empirical closed-loop L2-gain ($d \rightarrow z$) over test set	≈ 7.9	≈ 3.6 (below $\gamma = 4$ bound)

Consistent with the $\mathcal{L}_2$ -gain bound of Corollary 3, the closed-loop gain from disturbance to performance output in the minimax controller was below the bound of $\gamma=4$ across the range of disturbances tested, while the closed-loop gain of the disturbance-oblivious LQR baseline was significantly higher. It is also interesting to note the reduction in peak tracking error and settling time under worst case disturbance, which further demonstrates the qualitative advantage of designing to the adversarial disturbance model of Definition 3 rather than for nominal (disturbance-free) dynamics.

8. Engineering implementation and deployment considerations

The minimax controller is characterized in the theoretical framework of Sections 2–5 by the value function V and the saddle-point feedback pair (u^*, d^*) , which are not available in closed form except in a locally quadratic regime, as described in Corollary 2. This section discusses the practical engineering steps necessary to deploy the framework to physical robotic hardware, including requirements for actuation and sensing, the real-time computational architecture, the approximation of the value function on-board, and the robustness to sensor noise and latency and safety-constrained deployment.

8.1 Actuation and sensing requirements

The saddle-point law of Lemma 1 is written directly in the input channel of the control-affine dynamics, meaning that the desired state of the system, $u^*(x)$ should be a commanded joint torque (or an equivalent motor current via the actuator torque constant) for the implementation of this law. For both direct-drive actuators or series-elastic actuators with an inner torque-control loop, or for geared electric actuators with a current-control loop and a calibrated torque–current map, the projection of $u^*(x)$ onto the compact set U in Definition 2 should be realized by a hard saturation at the inner loop to comply with the actuator rated torque and the compactness hypothesis of Assumption 1.

For platforms subject to base motion or external contact, the joint position sensing (optical or magnetic) and direct velocity sensing or the numerical differentiation with low-pass filtering of state feedback $x = (q, \dot{q})$ is essential in order to compute an online estimate of the disturbance channel entering that is used for the certification of the $d^*(x, p)$ in Lemma 1 is evaluated only for analysis and worst-case certification and is not itself commanded, but an accurate estimate of the realized disturbance is needed to monitor the $\mathcal{L}_2$ -gain bound of Corollary 3 and can be used to trigger the safety behavior described in Section 8.5.

8.2 Real-Time Computational Architecture

A computational architecture with two-time scales is suggested. Using the current state estimate and the current onboard approximation $\hat{V}$ of the value function, an inner loop is executed at 1–2 kHz on a real-time target (a microcontroller or real-time Linux/RTOS process with deterministic scheduling), to evaluate the closed-form feedback law $u^*(x) = -\frac{1}{2}R^{-1}g_1(x)^T \nabla \hat{V}(x)$ and then project onto U , and command the inner torque loop of Section 8.1. An outer loop, performed at a lower rate (10–200 Hz depending on the available on-board compute), uses the policy-iteration procedure of Algorithm 1, but only within a receding local operating region around the current state; or, for platforms with tightly-bounded compute, the outer loop is omitted entirely, and the PDE solving task is replaced by an off-line pre-computation of the gains G , which can only be guaranteed to be optimal in a finite set of operating points, but can be evaluated online at a low rate via linear interpolation, known as gain scheduling.

In cases where the platform provides a GPU or neural accelerator on board, however, the outer loop can instead keep V as an efficient function approximation, such as a shallow neural network or radial basis functions within a bounded operation domain, trained using the offline policy iteration method of Theorem 2 and then fine-tuned online via several gradient updates per outer loop epoch; this approach sacrifices the guaranteed monotonic convergence provided by Theorem 2, which is guaranteed only if the exact policy evaluation operator of Lemma 3 is used, for computational efficiency.

8.3 Discretization of the policy-evaluation step

Solving the linear PDE of Lemma 3 analytically in an online setting is not practical. An approximate technique used by engineers substitutes the policy evaluation step in Algorithm 1 with an offline least squares or temporal difference fit of V based on trajectory rollouts collected under the policy pair $\hat{V}$ over trajectory rollouts collected under the current policy pair (u_k, d_k) , either from the physical platform, from a physics-engine simulation of the Euler–Lagrange dynamics of Definition 1, or from a combination of both (simulation-to-reality transfer). Because the theoretical convergence guarantee of Theorem 2 was established for the exact operator $\mathcal{T}_{u,d}$, in its exact form, the fitted approximation needs to be tested against the analytic Riccati solution of Corollary 2 in the domain where the latter exists before applying it to improve the policy, and the fitting error needs to be estimated and included in the safety margin of Section 8.5.

8.4 Robustness to sensor noise and communication latency

Two practical departures from the idealized model of Section 4 deserve explicit engineering treatment. First, measurement noise on x enters the evaluated gradient $\nabla \hat{V}(x)$ and hence the commanded torque; because u^* is affine in $\nabla \hat{V}$ near a locally quadratic $\hat{V}$ by Corollary 2, additive state noise of covariance Σ propagates approximately to input noise of covariance $R^{-1}g_{1,0}^T P \Sigma P g_{1,0} R^{-1}$, which needs to be checked to ensure that it is within the torque command resolution of the actuator and needs to be filtered (for instance Kalman or complementary filtering of the encoder/IMU measurements) before being evaluated for the feedback law. The

second consideration is the computational and communication latencies due to the distributed or networked nature of robot architectures that cause a delay between the measured states and their use for generating the control command; this delay can be considered another disturbance channel to be designed in, with a conservative enough disturbance bound D of Assumption 1 to accommodate the effect of the delay.

8.5 Safety and constraint handling at runtime

Because Corollary 3 establishes V as a strict Lyapunov function under the exact saddle-point feedback, the same function can be repurposed at runtime as a lightweight safety monitor even when the deployed controller uses an approximate $\hat{V}$: if the measured trajectory produces $\frac{d}{dt}\hat{V}(x(t))$ exceeding the bound established in the proof of Corollary 3 by more than a prescribed tolerance, this signals that the approximation error of $\hat{V}$ (Section 8.3), sensor noise (Section 8.4), or an out-of-envelope disturbance has invalidated the certificate, and the platform should fall back to a conservative, independently verified safety controller (e.g. a fixed-gain PD law about a known safe configuration) rather than continue to trust the minimax feedback law outside its validated operating region. This runtime monitor, together with hardware torque limiting at the inner loop of Section 8.1, is the recommended minimum safety architecture for deploying the theoretical framework of Sections 3–5 on physical robotic hardware.

9. Conclusion

In this paper, a minimax optimal control problem formulation and analysis of a nonlinear robot control system using the HJI equation is developed. From the preliminaries of the admissible policies, Hamiltonian and viscosity solutions, a verification theorem that related the solutions of the HJI equation to saddle-point minimax optimality (Theorem 1) and the existence, uniqueness and convergence of the underlying viscosity solution through policy iteration under coercivity and local stabilizability (Theorem 2) are provided. Three corollaries were presented for the disturbance-free HJB formulation, the explicit closed form saddle-point feedback controller with quadratic cost, and a guaranteed closed-loop L2 gain and Lyapunov stability certificates. A numerical example involving a two-link robot manipulator demonstrated the robustness benefits of the designed minimax controller when compared to a disturbance-blind LQR controller. The theoretical development of the problem was further elaborated by Section 8 which laid out the implementation of the developed theory in the form of a deployment architecture involving actuation and sensing, two-timescale real-time computation, discrete policy evaluation, robustness against sensor noise and delay, and safety monitoring using the Lyapunov certificate approach. Further research directions include expanding the framework to stochastic disturbance models, output feedback robotic problems, and distributed robotic games that are based on the fundamental zero-sum differential game.

References

1. Abu-Khalaf, M., Lewis, F. L., & Huang, J. (2006). Policy Iterations on the Hamilton–Jacobi–Isaacs Equation for H^∞ State Feedback Control with Input Saturation. *IEEE Transactions on Automatic Control*, 51(12), 1989–1995. <https://doi.org/10.1109/TAC.2006.884959>
2. Bouhmady, A., & Dekkaki, O. C. (2026). Hamilton–Jacobi reachability for viability analysis of constrained waste-to-energy systems under adversarial uncertainty. *European Journal of Control*, 89, 101501. <https://doi.org/10.1016/j.ejcon.2026.101501>
3. Darbon, J., Dower, P. M., & Meng, T. (2023). Neural network architectures using min-plus algebra for solving certain high-dimensional optimal control problems and Hamilton–Jacobi PDEs. *Mathematics of Control, Signals, and Systems*, 35(1), 1–44. <https://doi.org/10.1007/s00498-022-00333-2>
4. David, G., & Mdodo, K. L. (2025). A Safety-Constrained Reinforcement Learning Model for Autonomous Navigation in Cluttered Industrial Spaces. *International Academic Journal of Science and Engineering*, 12(3), 9–17. <https://doi.org/10.71086/IAJSE/V12I3/IAJSE1218>
5. Dong, B., An, T., Zhou, F., Liu, K., & Li, Y. (2019). Decentralized robust zero-sum neuro-optimal control for modular robot manipulators in contact with uncertain environments: theory and experimental verification: B. Dong et al. *Nonlinear Dynamics*, 97(1), 503–524. <https://doi.org/10.1007/s11071-019-04994-8>
6. Gomoyunov, M. I. (2025). Control Problems for Fractional-Order Systems: Formalism of Hamilton–Jacobi Equations and Methods for Constructing Optimal Feedback Strategies. *Differential Equations*, 61(Suppl 1), 1–153. <https://doi.org/10.1134/S0012266125130014>
7. Gomoyunov, M. I., Lukoyanov, N. Y., & Plaksin, A. R. (2021). Path-dependent Hamilton–Jacobi equations: the minimax solutions revised. *Applied Mathematics & Optimization*, 84(Suppl 1), 1087–1117. <https://doi.org/10.1007/s00245-021-09794-4>

8. Gopalakrishnan, R. E. C. A., Chandrakumar, E., Balan, A., Chandrasekar, T. K., & Kumar, R. S. (2024, June). Optimization of field-oriented control for BLDC motors in electric vehicles using bio-inspired algorithms. In *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)* (pp. 1-5). IEEE. <https://doi.org/10.1109/ICITEICS61368.2024.10625630>
9. Hu, H., Zhang, X., Lyu, X., & Chen, M. (2025, May). Learning robust policies via interpretable Hamilton-Jacobi reachability-guided disturbances. In *2025 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 2304-2310). IEEE. <https://doi.org/10.1109/ICRA55743.2025.11127287>
10. Karimov, N., Sayfullaev, D., Sattorova, Z., Omonov, Q., Mirzakhmedova, K., Alimardonov, O., & Polvanov, J. (2026, January). Deep learning for cybersecurity a review of current application and future directions. In *AIP Conference Proceedings* (Vol. 3345, No. 1, p. 020186). AIP Publishing LLC. <https://doi.org/10.1063/5.0298084>
11. Krolicki, A., Sutavani, S., & Vaidya, U. (2022, June). Koopman-based policy iteration for robust optimal control. In *2022 American Control Conference (ACC)* (pp. 1317-1322). IEEE. <https://doi.org/10.23919/ACC53348.2022.9867541>
12. Lee, D., Keimer, A., Bayen, A. M., & Tomlin, C. J. (2020, December). Hamilton-Jacobi formulation for state-constrained optimal control and zero-sum game problems. In *2020 59th IEEE conference on decision and control (CDC)* (pp. 1078-1085). IEEE. <https://doi.org/10.1109/CDC42340.2020.9304334>
13. Lian, B., Xue, W., Lewis, F. L., Modares, H., & Kiumarsi, B. (2024). *Integral and inverse reinforcement learning for optimal control systems and games* (pp. 1-10). New York: Springer. <https://doi.org/10.1007/978-3-031-45252-9>
14. Liu, J., Zhang, N., Li, Y., Xie, X., Tian, E., & Cao, J. (2022). Learning-based event-triggered tracking control for nonlinear networked control systems with unmatched disturbance. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(5), 3230-3240. <https://doi.org/10.1109/TSMC.2022.3224432>
15. Narayanan, V., Sahoo, A., & Jagannathan, S. (2018, June). Optimal event-triggered control of nonlinear systems: a min-max approach. In *2018 Annual American Control Conference (ACC)* (pp. 3441-3446). IEEE. <https://doi.org/10.23919/ACC.2018.8431477>
16. Rajan, C. (2026). Dynamic Bifurcation Mapping in Carbon Capture Reactors with Coupled Thermal Runaway Effects. *Applied Nonlinearity in Science and Technology*, 23-27.
17. Snousi, H. M., & Aleej, F. A. (2025). A Graph Reinforcement Learning Model for Robotic Path Optimization in Dynamic Environments. *International Academic Journal of Innovative Research*, 12(2), 63-72. <https://doi.org/10.71086/IAJIR/V12I2/IAJIR1225>
18. Theodorou, E. A., Buchli, J., & Schaal, S. (2009, March). Path integral-based stochastic optimal control for rigid body dynamics. In *2009 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (pp. 219-225). IEEE. <https://doi.org/10.1109/ADPRL.2009.4927548>
19. Wang, S., Zha, L., Liu, J., Sun, K., & Liu, J. (2026). An Artificial Bee Colony-Assisted Reinforcement Learning Approach for Improvement in Optimizing Performance of Interconnected Systems. *Communications in Nonlinear Science and Numerical Simulation*, 110461. <https://doi.org/10.1016/j.cnsns.2026.110461>
20. Zhang, X., & Rodriguez, S. (2023). Advanced optimization techniques for vehicle dynamics in robotics. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*, 1(1), 1-13.
21. Zhu, K., Lan, F., Zhao, W., & Zhang, T. (2024). Safe multi-agent reinforcement learning via approximate hamilton-jacobi reachability. *Journal of Intelligent & Robotic Systems*, 111(1), 7. <https://doi.org/10.1007/s10846-024-02156-6>
22. Zhu, Y., Zhao, D., Yang, X., & Zhang, Q. (2017). Policy iteration for H_∞ optimal control of polynomial nonlinear systems via sum of squares programming. *IEEE transactions on cybernetics*, 48(2), 500-509. <https://doi.org/10.1109/TCYB.2016.2643687>