SCIENCE CHINA
Information Sciences

• LETTER •

January 2026, Vol. 69, Iss. 1, 119201:1–119201:2
https://doi.org/10.1007/s11432-025-4691-7

H_∞ optimal output feedback control of an unknown linear system via adaptive dynamic programming

Yong-Sheng MA, Jian SUN* & Yong XU

State Key Laboratory of Autonomous Intelligent Unmanned Systems, School of Automation, Beijing Institute of Technology, Beijing 100081, China

Received 29 March 2025/Revised 23 June 2025/Accepted 22 July 2025/Published online 4 January 2026

Citation Ma Y-S, Sun J, Xu Y. H_∞ optimal output feedback control of an unknown linear system via adaptive dynamic programming. Sci China Inf Sci, 2026, 69(1): 119201, https://doi.org/10.1007/s11432-025-4691-7

Optimal control intends to optimize the user-defined performance index to determine an optimal control strategy for the control systems. Note that the conventional optimal control methods rely on precise model knowledge. This requirement presents a significant challenge in scenarios where precise model knowledge is unavailable. The adaptive dynamic programming (ADP) technique has emerged as a promising data-driven optimal control approach. Instead of relying on the system model, the ADP technique can learn an optimal controller from collected system data.

Two widely used algorithms in the ADP technique are value iteration (VI) and policy iteration (PI). The data-driven PI learning algorithm for optimal control of unknown systems is developed in [1]. The optimal output regulation problem is addressed by the ADP-based PI algorithm in [2]. The authors in [3] developed a VI-based learning algorithm to derive an optimal tracking controller from the system data. A robust optimal control learning algorithm is developed to guarantee the asymptotic stability of the power system in [4]. A data-driven event-triggered optimal control scheme is designed in [5] to reduce data transmission.

Note that the least-squares (LS) method is usually used during the execution of the learning algorithm. Specifically, the LS method solves a data equation formed by the collected system data. However, a full-rank condition is imposed on the coefficient matrix to ensure the solvability of the data equation. To satisfy this condition, past system data at multiple time instants are stored, consuming additional storage resources. In light of the above discussion, this study designs a novel model-free online learning algorithm to learn the optimal controller from system data. The main contributions of this study are summarized below. (1) Unlike existing model-free optimal control algorithms requiring storage of historical system data across multiple time instants, the proposed online learning algorithm leverages real-time system data during execution, significantly reducing memory resource consumption. (2) By constructing an invertible matrix-based data equation, we eliminate dependence on the persistence of excitation (PE) condition for convergence. Theoretical analysis demonstrates that the convergence is guaranteed under the milder interval excitation (IE) condition.

Problem description. This section presents the problem description of the H_∞ optimal control of the linear system with the following dynamics:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Dw(t), \\ y(t) = Cx(t), \end{cases} \tag{1}$$

where $x(t)$, $w(t)$, and $y(t)$ are the system state, disturbance, and system output, respectively. A, B, C and D are defined as the system matrices.

The H_∞ optimal control problem can be framed as a two-player zero-sum game, in which $u(t)$ and $w(t)$ represent the two players with opposing objectives. This problem can be formulated as a min-max optimization problem:

$$\max_{w(t)} \min_{u(t)} J(y(t), u(t), w(t)) = \int_0^\infty r(t) dt, \tag{2}$$

where $r(t) = y^{\mathrm{T}}(t)Qy(t) + u^{\mathrm{T}}(t)Ru(t) - \gamma^2 w^{\mathrm{T}}(t)w(t)$ is the performance function. T represents the transpose symbol. Q, R and γ are the weight parameters.

Based on the H_∞ optimal control theory, the optimal control policy and the worst disturbance policy can be designed separately as $u(t) = -Ky(t)$ and $w(t) = \gamma^{-2} D^{\mathrm{T}} Px(t)$, where K satisfies

$$KC = R^{-1}(B^{\mathrm{T}}P + E^{\mathrm{T}}P) \tag{3}$$

with P being the unique solution of the following game algebraic Riccati equation (GARE):

$$A^{\mathrm{T}}P + PA - PBR^{-1}B^{\mathrm{T}}P + PER^{-1}E^{\mathrm{T}}P + \gamma^{-2}PDD^{\mathrm{T}}P + C^{\mathrm{T}}QC = 0. \tag{4}$$

Note that the GARE cannot be solved directly, as it is a nonlinear matrix equation for P. Moreover, GARE is a model-based matrix equation, which implies that solving it requires a system model. This study proposes a model-free iterative learning algorithm that adaptively derives an optimal controller without prior model knowledge.

Main results. This section presents the main results of the proposed data-driven iterative learning algorithm. First, a model-based iterative scheme is presented to solve for P from the GARE (4), which involves solving the following iterative matrix equation:

$$A_i^{\mathrm{T}}P_{i+1} + P_{i+1}A_i + E_i^{\mathrm{T}}K_{i+1}C + (K_{i+1}C)^{\mathrm{T}}E_i + Q_i = 0, \tag{5}$$

* Corresponding author (email: sunjian@bit.edu.cn)

where $Q_i = C^{\mathrm{T}}QC + (K_iC)^{\mathrm{T}}RK_iC - E_i^{\mathrm{T}}K_iC - (K_iC)^{\mathrm{T}}E_i - \gamma^{-2}D_i^{\mathrm{T}}D_i$, $A_i = A - BK_iC + \gamma^{-2}DD_i^{\mathrm{T}}P_i$, and E_i, D_i and K_iC are updated by

$$E_{i+1} = E^{\mathrm{T}}P_{i+1}, \tag{6}$$

$$D_{i+1} = D^{\mathrm{T}}P_{i+1}, \tag{7}$$

$$K_{i+1}C = R^{-1}(B^{\mathrm{T}}P_{i+1} + E_{i+1}). \tag{8}$$

Note that matrix equation (5) is based on the system model. By implementing the equivalent transformation, we can obtain the following data equation:

$$\Psi_i(t)\Phi_{i+1} = -\Omega_i(t), \tag{9}$$

where $\Psi_i(t)$ and $\Omega_i(t)$ are data matrices. Φ_{i+1} is the solution of data equation. The specific form of $\Psi_i(t)$, $\Omega_i(t)$ and Φ_{i+1}, and the detailed transformation process are provided in Appendix A.

Remark 1. To ensure the data equation admits a unique solution, conventional methods employ data storage units to construct the coefficient matrix $\Psi_i(t)$, which needs to satisfy the column full-rank condition. Crucially, maintaining full rank demands storage of sufficient data, potentially consuming substantial memory resources.

Theorem 1. If there exists a time point $t_0 + T$ such that $\det[Z(t_0+T)] > 0$, where $\dot{Z}(t) = z(t)z^{\mathrm{T}}(t)$ with $z(t) = [I_{xx}^{\mathrm{T}}(t), I_{ux}^{\mathrm{T}}(t), I_{wx}^{\mathrm{T}}(t), I_{xy}^{\mathrm{T}}(t)]^{\mathrm{T}}$, $I_{xx}(t) = \int_{t_0}^{t} x(\tau) \otimes x(\tau)d\tau$, $I_{ux}(t) = \int_{t_0}^{t} u(\tau) \otimes x(\tau)d\tau$, $I_{wx}(t) = \int_{t_0}^{t} w(\tau) \otimes x(\tau)d\tau$ and $I_{xy}(t) = \int_{t_0}^{t} x(\tau) \otimes y(\tau)d\tau$, then Φ_{i+1} is uniquely solved by

$$\Phi_{i+1} = -\Psi_i^{-1}(t_0+T)\Omega_i(t_0+T). \tag{10}$$

Proof. The proof is presented in Appendix B.

Remark 2. Theorem 1 shows that Φ_{i+1} is solved using (10), where $t_0 + T$ is determined by $\det[Z(t_0+T)] > 0$. From $\dot{Z}(t) = z(t)z^{\mathrm{T}}(t)$, we get $Z(t_0+T) = \int_{t_0}^{t_0+T} z(\tau)z^{\mathrm{T}}(\tau)d\tau$. Thus, $\det[Z(t_0+T)] > 0$ implies that

$$Z(t_0+T) = \int_{t_0}^{t_0+T} z(\tau)z^{\mathrm{T}}(\tau)d\tau > \delta I > 0, \tag{11}$$

where $\delta > 0$ is a constant. Note that Eq. (11) is an IE condition. Different from the PE condition that requires sufficient energy over the entire time span, the established IE condition only requires energy for a finite duration, which relaxes the restriction of the PE condition.

Based on Theorem 1, the model-free online iterative learning algorithm is provided in Algorithm 1 with its convergence analysis being presented below.

First, observe that Eq. (10) is derived from (9), which itself constitutes an equivalent reformulation of the matrix equation (5). This matrix equation is satisfied by solutions K_{i+1} and P_{i+1} generated through the model-based iterative scheme (5)–(8). Theorem 1 establishes that Φ_{i+1} admits a unique solution via (10). Consequently, solving Φ_{i+1} from (10) is mathematically equivalent to solving K_{i+1} and P_{i+1} via (5)–(8). Furthermore, since K_{i+1} and P_{i+1} converge to optimal values K^* and P^* under this scheme, the proposed learning method likewise attains optimal convergence.

Algorithm 1 Data-driven online iterative learning algorithm.

```
Select u(t) ← K_0 y(t) + ϵ(t); {K_0: initial gain, ϵ: exploration noise};
Find t_0 + T satisfying det[Z(t_0 + T)] > 0;
Set i ← 0; Compute Φ_1 via (10);
while ‖Φ_i − Φ_{i−1}‖ ⩾ ε do
    i ← i + 1;
    Compute Φ_{i+1} via (10);
end while
```

Remark 3. The parameter $t_0 + T$ plays an important role in ensuring the convergence of the iteration learning algorithm. The theoretical results of Theorem 1 show that the data equation has a unique solution at t_0+T, which further guarantees that the control policy learned from the proposed iteration learning algorithm converges to its optimal value.

Simulation. A power system is used to verify the effectiveness of the proposed method. The detailed simulation result is presented in Appendix C.

Conclusion. In this work, the H_∞ optimal LFC problem is studied in the context of an unknown power system model. A novel online model-free iterative learning algorithm is developed to derive the optimal control policy from system data. The superiority of the proposed method is demonstrated through its application to the power system, where it is compared with existing methods.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62322305, 62495090, 62495095).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Ma Y S, Sun J, Xu Y, et al. Adaptive dynamic programming for optimal control of unknown LTI system via interval excitation. IEEE Trans Automat Contr, 2025, 70: 4896–4903
2 Xu Y, Wu Z-G, Che W-W, et al. Reinforcement learning-based unknown reference tracking control of HMASs with nonidentical communication delays. Sci China Inf Sci, 2023, 66: 170203
3 Jiang Y, Gao W, Wu J, et al. Reinforcement learning and cooperative H_∞ output regulation of linear continuous-time multi-agent systems. Automatica, 2023, 148: 110768
4 Zhao J, Yang C, Gao W, et al. Adaptive optimal output regulation of interconnected singularly perturbed systems with application to power systems. IEEE CAA J Autom Sin, 2024, 11: 595–607
5 Shen M, Wang X, Zhu S, et al. Data-driven event-triggered adaptive dynamic programming control for nonlinear systems with input saturation. IEEE Trans Cybern, 2024, 54: 1178–1188