SCIENCE CHINA
Information Sciences

• Supplementary File •

H_∞ Optimal Output Feedback Control of an Unknown Linear System via Adaptive Dynamic Programming

Yong-Sheng Ma[1], Jian Sun[1*] & Yong Xu[1]

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

Appendix A The transformation process from matrix equation to data equation

By multiplying both sides of the matrix equation (5) on the left and right by the vectors $x^{\mathrm{T}}(t)$ and $x(t)$, respectively, we obtain

$$2x^{\mathrm{T}}(t)P_{i+1}[\dot{x}(t)-Bu(t)-Dw(t)-BK_iCx(t)+\gamma^{-2}DD_ix(t)]-x^{\mathrm{T}}(t)(P_{i+1}A_i+A_iP_{i+1})x(t)=0, \tag{A1}$$

where $\dot{x}(t)$ is consistent with system (1).

Substituting (6)-(8) into (A1) and integrating both sides of (A1) yield

$$\begin{aligned}&x^{\mathrm{T}}(t)P_{i+1}x(t)-x^{\mathrm{T}}(t_0)P_{i+1}x(t_0)+\int_{t_0}^{t}x^{\mathrm{T}}(\tau)Q_ix(\tau)d\tau-2\int_{t_0}^{t}x^{\mathrm{T}}(\tau)(RK_{i+1}C-E_{i+1})(u(\tau)+K_iy(\tau))d\tau\\&-2\int_{t_0}^{t}x^{\mathrm{T}}(\tau)D_{i+1}^{\mathrm{T}}(w(\tau)-\gamma^{-2}D_ix(\tau))d\tau-\int_{t_0}^{t}x^{\mathrm{T}}(\tau)(E_i^{\mathrm{T}}K_{i+1}C+(K_{i+1}C)^{\mathrm{T}}E_i)x(\tau)d\tau=0.\end{aligned} \tag{A2}$$

For the convenience of subsequent derivation, define following two operators: $vec(A)=[a_1^{\mathrm{T}},\dots,a_N^{\mathrm{T}}]^{\mathrm{T}}$ with a_i being the ith column of A; $vech\left(\begin{bmatrix}a_1 & a_2\\ a_2^{\mathrm{T}} & a_3\end{bmatrix}\right)=[a_1,a_2,a_3]^{\mathrm{T}}$.

Using the Kronecker product, (A2) can be expressed in the following compact form:

$$\psi_i^{\mathrm{T}}(t)\Phi_{i+1}=-I_{xx}^{\mathrm{T}}(t)vec(Q_i), \tag{A3}$$

where

$$\begin{aligned}&\psi_i(t)=[\psi_i^{[11]}(t),\psi_i^{[12]}(t),\psi_i^{[13]}(t),\psi_i^{[14]}(t)],\Phi_{i+1}=[vech^{\mathrm{T}}(P_{i+1}),vec^{\mathrm{T}}(K_{i+1}),vec^{\mathrm{T}}(E_{i+1}),vec^{\mathrm{T}}(D_{i+1})]^{\mathrm{T}},\\&\psi_i^{[11]}(t)=\chi(t)-\chi(t_0),\psi_i^{[13]}(t)=2I_{ux}^{\mathrm{T}}(t)+2I_{xy}^{\mathrm{T}}(t)k_i,\psi_i^{[12]}(t)=2I_{xy}^{\mathrm{T}}(t)e_i-2I_{uy}^{\mathrm{T}}(t)r-2I_{yy}^{\mathrm{T}}(t)b_i,\\&\psi_i^{[14]}(t)=-2I_{wx}^{\mathrm{T}}(t)+2I_{xx}^{\mathrm{T}}(t)f_i,I_{xx}(t)=\int_{t_0}^{t}x(\tau)\otimes x(\tau)d\tau,I_{xy}(\tau)=\int_{t_0}^{t}x(\tau)\otimes y(\tau)d\tau,\\&I_{uy}(t)=\int_{t_0}^{t}u(\tau)\otimes y(\tau)d\tau,I_{ux}(t)=\int_{t_0}^{t}u(\tau)\otimes x(\tau)d\tau,I_{wx}(t)=\int_{t_0}^{t}w(\tau)\otimes x(\tau)d\tau,I_{yy}(t)=\int_{t_0}^{t}y(\tau)\otimes y(\tau)d\tau,\\&e_i=E_i^{\mathrm{T}}\otimes I,r=I\otimes R,k_i=I\otimes K_i^{\mathrm{T}},b_i=I\otimes K_i^{\mathrm{T}}R,f_i=I\otimes\gamma^2D_i^{\mathrm{T}},\chi(t)=vech(2x(t)x^{\mathrm{T}}(t)-\mathrm{diag}^2(x(t))).\end{aligned}$$

Rather than storing historical system data and using the LS method, we construct a data equation based on online system data and propose a new method to solve it. To achieve this objective, multiplying both sides of the data equation (A3) by $\psi_i(t)$ yields

$$\psi_i(t)\psi_i^{\mathrm{T}}(t)\Phi_{i+1}=-\psi_i(t)I_{xx}^{\mathrm{T}}(t)vec(Q_i). \tag{A4}$$

Then, integrating both sides of data equation (A4) yields

$$\Psi_i(t)\Phi_{i+1}=-\Omega_i(t), \tag{A5}$$

where

$$\Psi_i(t)=\int_{t_0}^{t}\psi_i(\tau)\psi_i^{\mathrm{T}}(\tau)d\tau,\Omega_i(t)=\int_{t_0}^{t}\psi_i(\tau)I_{xx}^{\mathrm{T}}(\tau)vec(Q_i)d\tau.$$

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

Appendix B Proof of Theorem 1

Consider the data equation (A5) with $t = t_0 + T$, i.e.,

$$\Psi_i(t_0+T)\Phi_{i+1} = -\Omega_i(t_0+T). \tag{B1}$$

If $\Psi_i(t_0+T)$ is invertible, then (10) can be obtained by multiplying both sides of (B1) by $\Psi_i^{-1}(t_0+T)$. We now prove this, which is equivalent to proving that $\zeta = 0$ is the only solution of the following equation:

$$\zeta^{\mathrm{T}}\Psi_i^P(t_0+T)\zeta = 0, \tag{B2}$$

where $\zeta = [vech^{\mathrm{T}}(U), vec^{\mathrm{T}}(V), vec^{\mathrm{T}}(W), vec^{\mathrm{T}}(X)]^{\mathrm{T}}$ with U, V, W and X being arbitrary matrices.

This statement is proven by contradiction. Assume that ζ is a nonzero solution of (B2).

Substituting $\psi_i(t)$ into $\Psi_i(t)$ yields

$$\Psi_i(t) = \int_{t_0}^{t} \begin{bmatrix} \Xi_{11} & \Xi_{12} & \Xi_{13} & \Xi_{14} \\ * & \Xi_{22} & \Xi_{23} & \Xi_{24} \\ * & * & \Xi_{33} & \Xi_{34} \\ * & * & * & \Xi_{44} \end{bmatrix} d\tau, \tag{B3}$$

where $\Xi_{mm} = \psi_i^{[1m]}(\tau)(\psi_i^{[1m]}(\tau))^{\mathrm{T}}, m = 1,2,3,4$.

Moreover, by the property of the Kronecker product, it follows from (A1) that

$$\begin{aligned} &[\chi(t) - \chi(t_0)]^{\mathrm{T}} vech(P_{i+1}) \\ =&I_{xx}^{\mathrm{T}}(t)vec(P_{i+1}A_i + A_i^{\mathrm{T}}P_{i+1}) + 2[I_{ux}^{\mathrm{T}}(t)b + I_{xy}^{\mathrm{T}}(t)d_i + I_{wx}^{\mathrm{T}}(t)f - I_{xx}^{\mathrm{T}}(t)g_i]vec(P_{i+1}), \end{aligned} \tag{B4}$$

where $d_i = I_n \otimes K_i^{\mathrm{T}}B^{\mathrm{T}}$, $b = I_n \otimes B^{\mathrm{T}}$, $f = I_n \otimes D^{\mathrm{T}}$ and $g_i = I_n \otimes D_i^{\mathrm{T}}D^{\mathrm{T}}$.

Combining with (B2)-(B4), we can get

$$\Upsilon^T \int_{t_0}^{t_0+T} z(\tau)z^{\mathrm{T}}(\tau)d\tau\Upsilon = 0, \tag{B5}$$

where $\Upsilon = [\Upsilon_1^{\mathrm{T}}, \Upsilon_2^{\mathrm{T}}, \Upsilon_3^{\mathrm{T}}, \Upsilon_4^{\mathrm{T}}]^{\mathrm{T}}$. $\Upsilon_1 = vec(UA_i + A_i^{\mathrm{T}}U) + 2\gamma^2 vec(D_i^{\mathrm{T}}X - D_i^{\mathrm{T}}D^{\mathrm{T}}U)$, $\Upsilon_2 = 2vec(B^{\mathrm{T}}U - RV + W)$, $\Upsilon_3 = 2vec(D^{\mathrm{T}}U - X)$ and $\Upsilon_4 = 2vec(K_i^{\mathrm{T}}B^{\mathrm{T}}U + E_i^{\mathrm{T}}V + K_i^{\mathrm{T}}W - K_i^{\mathrm{T}}RV)$.

The explicit solution to $\dot{Z}(t) = z(t)z^{\mathrm{T}}(t)$ is $Z(t) = \int_{t_0}^{t} z(\tau)z^{\mathrm{T}}(\tau)d\tau$. Since $\det[Z(t_0+T)] > 0$ (where $\det[\cdot]$ represents the determinant), it follows that $Z(t_0+T)$ is positive definite, implying $Z(t_0+T) = \int_{t_0}^{t_0+T} z(\tau)z^{\mathrm{T}}(\tau)d\tau > \delta I > 0$, where $\delta > 0$ is constant. Thus, the unique solution of (B5) is $\Upsilon_1 = 0$, $\Upsilon_2 = 0$, $\Upsilon_3 = 0$, and $\Upsilon_4 = 0$. With $\Upsilon_2 = 0$ and $\Upsilon_4 = 0$, we get $E_i^{\mathrm{T}}V = 0$, implying that $V = 0$ since $E_i \neq 0$. Note that the Hurwitz property of A_i is guaranteed in the model-based iterative scheme and consequently preserved in the data-driven approach, as it constitutes an equivalent formulation of the model-based approach. Additionally, with $\Upsilon_1 = 0$ and $\Upsilon_3 = 0$, we get $UA_i + A_i^{\mathrm{T}}U = 0$, implying that $U = 0$ since A_i is Hurwitz. Since $V = 0$ and $U = 0$, it follows from $\Upsilon_2 = 0$ and $\Upsilon_3 = 0$ that $W = 0$ and $X = 0$. Thus, we have $\zeta = [\mathrm{vech}^{\mathrm{T}}(U), \mathrm{vec}^{\mathrm{T}}(V), \mathrm{vec}^{\mathrm{T}}(W), \mathrm{vec}^{\mathrm{T}}(X)]^{\mathrm{T}} = 0$, which contradicts the assumption that ζ is a nonzero solution to (B2). Therefore, $\Psi_i(t_0+T)$ is invertible.

Appendix C Simulation

A power system is used to verify the effectiveness of the proposed method and demonstrate its superiority by comparing it with existing methods. The framework diagram of power system under the proposed data-driven learning method is illustrated in Figure C1, where the power system with the following dynamics [2]:

$$\begin{cases} \Delta\dot{f}(t) = \dfrac{1}{M}[\Delta P_m(t) - \Delta P_d(t) - F\Delta f(t)] \\ \Delta\dot{P}_m(t) = \dfrac{1}{T_t}[\Delta P_v(t) - \Delta P_m(t)] \\ \Delta\dot{P}_v(t) = \dfrac{1}{T_g}[u(t) - \dfrac{1}{S}\Delta f(t) - \Delta P_v(t)] \\ ACE(t) = \beta\Delta f(t) \end{cases} \tag{C1}$$

where the physical meaning and values of power parameters refer to [2]. Defining system state $x(t) = [\Delta f(t), \Delta P_m(t), \Delta P_v(t)]^{\mathrm{T}}$, the disturbance $w(t) = \Delta P_d(t)$, and the system output $y(t) = ACE(t)$, a state space representation of the power system (C1) can be described as

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} \frac{-F}{M} & \frac{1}{M} & 0 \\ 0 & \frac{-1}{T_t} & \frac{1}{T_t} \\ \frac{-1}{ST_g} & 0 & \frac{-1}{T_g} \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T_g} \end{bmatrix} u(t) + \begin{bmatrix} \frac{-1}{M} \\ 0 \\ 0 \end{bmatrix} w(t) \\ y(t) = \begin{bmatrix} \beta & 0 & 0 \end{bmatrix} x(t). \end{cases} \tag{C2}$$

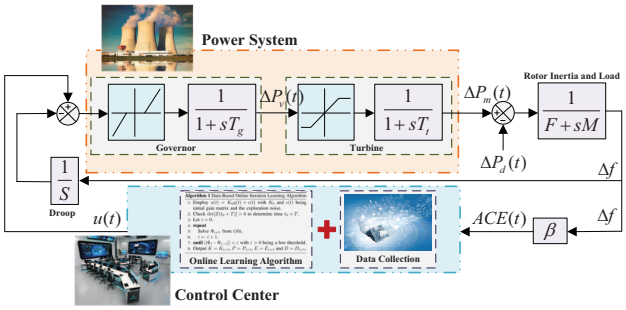

Figure C1 The framework diagram of power system under the proposed data-driven learning method.

Appendix C.1 Simulation Results

The learning algorithm parameters are set to $R = 1$, $Q = \mathrm{diag}\{1, 1, 1\}$, $\gamma = 4$, $K_0 = 0$, and $\varepsilon = 10^{-3}$. The design of exploration noise poses a significant challenge for data-driven learning, especially in high-dimensional systems. Commonly used profiles include random noise (convenient but potentially suboptimal), exponentially decaying probing signals (balancing ease and effectiveness), and sinusoidal signals (providing directional, periodic forces). For our simulations, we implement composite exploration noise $\epsilon(t) = e^{-t}(-\sin(6.7t) + \cos(15.2t))$ combining exponentially decaying probing and sinusoidal signals to facilitate more efficient strategy exploration. By checking the condition $\det\{Z(t_0 + T)\} > 0$, it can be confirmed that $\det\{Z(1)\} > 0$, which indicates that the time point $t_0 + T$ is determined as $t_0 + T = 1s$. The optimal control gain derived at $t = 1, \mathrm{s}$ is $K = [-0.8177, 0.3821, 0.5069]$. Figures C2 and C3 depict the learning process: Figure C2 shows $||K_i - K^*||$ gradually converging to zero, while Figure C3 demonstrates the performance function $V_i(t) = x^{\mathrm{T}}(t)P_i x(t)$ reaching its optimal value through iterative learning. The learned optimal controller is then implemented in the power system to stabilize the system, with state/input trajectories in Figures C4 and C5. Here, learning occurs during $t \in [0, 1]s$, after which the system switches to control mode ensuring asymptotic stability.

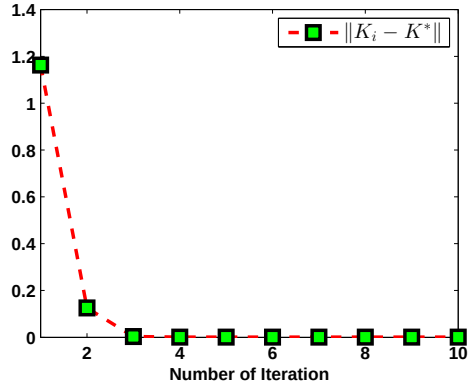

Figure C2 Trajectory of the learning error $\|K_i - K^*\|$.

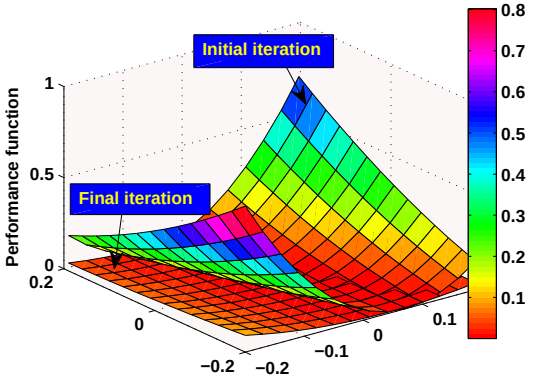

Figure C3 Evolution of performance function.

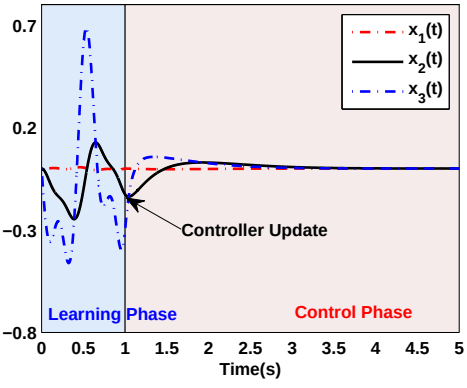

Figure C4 Trajectory of system state under different stages.

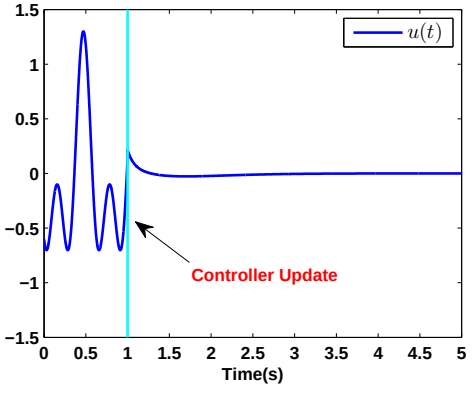

Figure C5 Trajectory of optimal control input.

To examine the impact of parameters on algorithm performance, we select different values and compare the learning error $||K_{10} - K^*||$ after 10 iterations, as shown in Table C1. Notably, the algorithm achieves excellent convergence after 10 iterations. Table C1 demonstrates that the learning error decreases with increasing parameters R and γ, while increases with larger Q. These findings indicate that higher R and γ values combined with smaller Q enhance algorithm performance.

Table C1 Comparison of learning error $||K_{10} - K^*||$ under different parameters

(a) Different parameter R

$R = 1$	1.53×10^{-4}
$R = 2$	9.54×10^{-5}
$R = 3$	7.38×10^{-5}

(b) Different parameter Q

$Q = \text{diag}\{2, 2, 2\}$	1.77×10^{-4}
$Q = \text{diag}\{3, 3, 3\}$	3.67×10^{-4}
$Q = \text{diag}\{4, 4, 4\}$	6.90×10^{-4}

(c) Different parameter γ

$\gamma = 4$	2.47×10^{-4}
$\gamma = 6$	1.46×10^{-4}
$\gamma = 8$	1.22×10^{-4}

We further validate the robustness of the learned controller against external disturbances. Figures C6 and C7 depict the trajectories of external disturbances, disturbance-corrupted control inputs, and perturbed system states, with disturbances generated by step signals and random signals. Figure C7 confirms that the perturbed states exhibit bounded fluctuations, demonstrating robust disturbance rejection. Moreover, the results in Figures C8 and C9 verify the H∞ performance requirement. Figure C8 shows the energy consumption $\int_0^t (x^{\mathrm{T}}Qx + u^{\mathrm{T}}Ru)dt$ and disturbance energy $\int_0^t w^{\mathrm{T}}wdt$. Figure C9 displays the energy ratio trajectory $\frac{\int_0^t (x^{\mathrm{T}}Qx+u^{\mathrm{T}}Ru)dt}{\int_0^t w^{\mathrm{T}}wdt}$, which remains below γ^2, confirming satisfaction of the H∞ performance requirement.

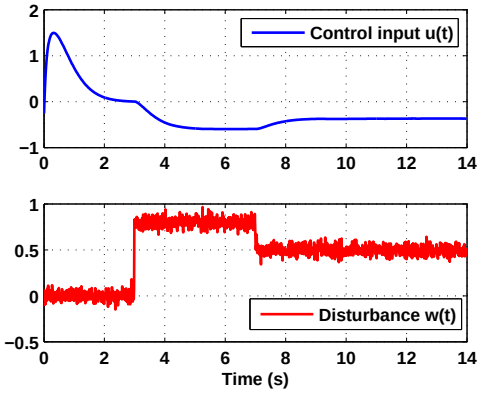

Figure C6 Trajectories of control input and external disturbance.

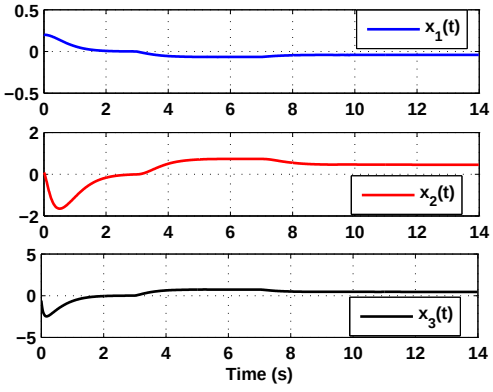

Figure C7 Trajectory of system state under the influence of external disturbance.

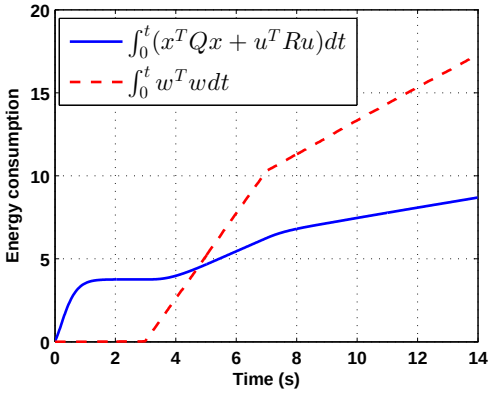

Figure C8 Trajectory of the energy consumption.

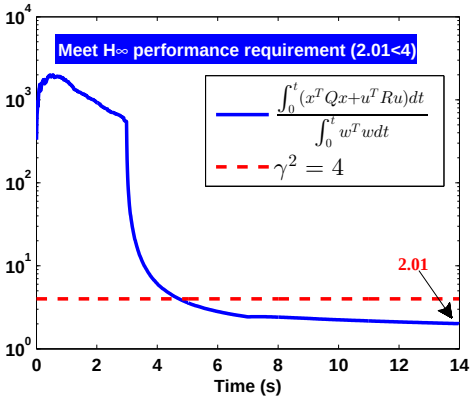

Figure C9 Trajectory of energy consumption ratio.

Appendix C.2 Comparison Results

To validate the superiority of the proposed method, we conduct comparative simulations against existing data-driven approaches: the concurrent learning algorithm [1] and model-free iterative learning method [2]. Table C2 provides theoretical

comparisons across three critical aspects: full-rank condition, historical data storage, and PE condition. It is worth noting that the advantage of [2] over [1] is that it gets rid the limitation of PE condition. However, both the full-rank condition and data storage are necessary prerequisites to ensure the learning algorithm convergence in [1] and [2]. In contrast, our method constructs a novel data equation structure using real-time system data and establishes an online verification condition, which eliminates the full-rank requirement, obviates the historical data storage and relaxes PE condition to interval excitation (IE) condition. Furthermore, the numerical quantitative comparison result is shown in Figures C10-C12. Note that the numerical validation is examined by single-input systems ($m = 1$) of varying dimensions ($n = 1, 2, 3$), where a stable system is generated via MATLAB's rss command. Figure C10 presents the comparative data utilization results, where our method demonstrates superior data efficiency by eliminating reliance on data storage. We evaluate computational efficiency by measuring the average time required for algorithms to complete 10 iterations. It is noteworthy that accurate controller parameters are attained after 10 iterations for all methods owing to their quadratic convergence rates. Figure C11 reveals that our approach consumes significantly less computation time than the other two methods. This efficiency stems primarily from eliminating repeated verification of full-rank conditions and avoiding historical data storage. Although our method eliminates the requirements (full-rank condition, data storage, and PE condition), it maintains convergence speed without significant reduction, as demonstrated in Figure C12.

Table C2 Comparison for different methods

	Our Method	Ref. [1]	Ref. [2]
Need Full-Rank Condition	✗	✓	✓
Need Data Storage	✗	✓	✓
Need PE Condition	✗	✗	✓

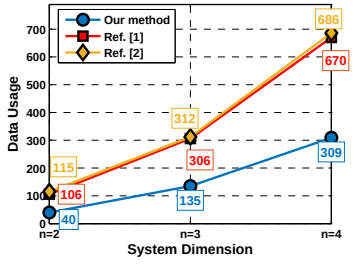

Figure C10 Comparison of data usage under different methods.

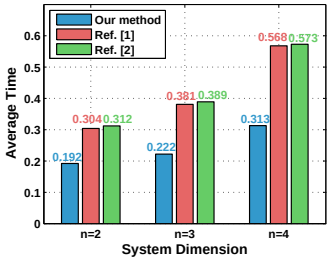

Figure C11 Comparison of computational efficiency under different methods.

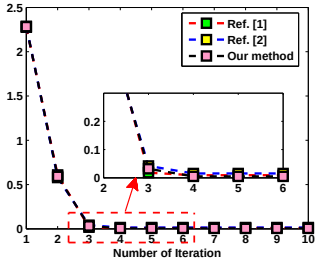

Figure C12 Comparison of convergence speed under different methods.

References

1 Tan L, Duc D. Integral reinforcement learning-based event-triggered $H\infty$ control algorithm for affine nonlinear systems with asymmetric input saturation and external disturbances. *Franklin Open*, 2024, 8:100132.

2 Hu S, Luo Y, Xie X et al. $H\infty$ optimal load frequency control of power system: A novel model-free approach. *IEEE Trans. Circuits Syst. II, Exp. Briefs*, 2025, 72:228–232.