Journal of Southeast University (English Edition) Vol. 42, No. 1, pp. 112-120 Mar. 2026 ISSN 1003-7985

Temperature regulation of an optomechanical frame based on reinforcement learning active disturbance rejection control

GU Yanping[1], ZHANG Hao[1], XU Tao[1], QIAN Bin[2]

(1. Shanghai Institute of Satellite Engineering, Shanghai 201109, China; 2. Shanghai Academy of Spaceflight Technology, Shanghai 201109, China)

Abstract: Spaceborne optomechanical systems face the dual challenges of extreme thermal disturbances and millikelvin-level temperature control precision during orbital operations, demanding robust control strategies. To address the performance limitations of conventional fixed-parameter active disturbance rejection control (ADRC) under complex operating conditions, this work proposes a Q-learning-enhanced adaptive ADRC framework. A thermal-transfer model incorporating multisource disturbances (solar radiation, structural conduction, and contact thermal resistance) is established, coupled with a reinforcement learning-driven parameter optimization mechanism. The ε-greedy policy dynamically adjusts observer bandwidth ($\omega_o \in [0.01, 0.2]$) and controller bandwidth ($\omega_c \in [0.01, 0.1]$) to enable real-time estimation and compensation of total disturbances. Simulation results demonstrate significant improvements over fixed-parameter ADRC and a self-tuning internal model control proportional-integral (SIMC-PI) controller: 31.3% and 15.4% reduction in settling time during setpoint responses, respectively; 21.8% lower integral absolute error (IAE) than the fixed-parameter ADRC during setpoint step responses; 12.7% and 52.5% enhancement in control precision over conventional fixed-parameter and SIMC-PI controllers, respectively, under ±10 K periodic and step thermal disturbances. Monte Carlo robustness tests reveal smaller fluctuation ranges of IAE, settling time, and overshoot under ±5% parameter perturbations. This methodology establishes a new paradigm for millikelvin-level thermal control in space optical payloads.

Key words: optomechanical system; active disturbance rejection controller; Q-learning; high precision temperature control

Received 2025-04-27, **Revised** 2025-10-17.

Biography: GU Yanping (1986—), female, doctor, senior research fellow, gyp0523@163.com.

Foundation items: The National Key R&D Program of China (No. 2022YFB3902902), the National Natural Science Foundation of China (No. 52276003).

Citation: GU Yanping, ZHANG Hao, XU Tao, et al. Temperature regulation of an optomechanical frame based on reinforcement learning active disturbance rejection control[J]. Journal of Southeast University (English Edition), 2026, 42(1): 112-120. DOI: 10.3969/j.issn.1003-7985.2026.01.011.

DOI: 10.3969/j.issn.1003-7985.2026.01.011

In satellite-based Earth observation and cosmic exploration missions, spaceborne optical systems have overcome the limitations of ground-based observations, significantly advancing fields including terrestrial monitoring, atmospheric studies, oceanographic surveys, and astronomical exploration[1]. During orbital operations, spacecraft experience variable thermal radiation from solar exposure, planetary albedo, and deep-space background radiation (with an equivalent temperature of approximately 4 K), causing periodic thermal fluctuations in optomechanical structures[2]. Particularly for geostationary orbiting cameras, their optical systems endure approximately 4 h of solar radiation per orbital cycle, followed by extreme cold immersion in shadow periods, resulting in surface temperature variations exceeding ±200 K[3]. These thermal gradients induce changes in lens curvature, thickness, and refractive index through conductive and radiative heat transfer, severely compromising imaging fidelity[4]. Consequently, precision thermal regulation of optical components has emerged as a critical technology for maintaining satellite imaging stability[5].

Global efforts in spacecraft thermal management have achieved significant advances. International advanced aerospace systems have attained thermal control precision of ±0.02 K[6], while China's GF-2 satellite achieved ±0.3 K accuracy[7]. Critical components of the Hubble Space Telescope maintain temperature within ±0.1 K[8]. Tong et al.[9] enhanced thermal regulation through an improved proportional-integral (PI) algorithm, elevating temperature control precision from ±0.8 to ±0.5 K for lens barrels and from ±0.15 to ±0.05 K for primary mirror mounts.

Emerging requirements present new challenges: the Herschel Space Observatory's far-infrared optics demanded ±3 mK stability within 10-s intervals[10], while space interferometers necessitate ±1 mK precision[5]. Active disturbance rejection control (ADRC) has shown exceptional promise in thermal management due to its model-independent architecture and algorithmic simplic-

ity[11]. Pan et al.[12] developed an ADRC-Smith composite system, achieving a 0.14 ℃ temperature tracking error for shroud temperature control. Yun et al.[13] implemented discrete ADRC on CompactRIO hardware, attaining 0.03 ℃/10 min stability for blackbody radiation sources.

Reinforcement learning (RL) is categorized into value-based methods (e.g., Q-learning, deep Q-networks) and policy-based methods (e.g., policy gradients and actor-critic algorithms)[14]. Compared with modern algorithms that rely on large-scale neural networks and extensive training data, classical Q-learning, characterized by its lightweight implementation, stability, and discrete action mechanism[15], exhibits greater suitability for ADRC parameter-tuning tasks with limited state-action spaces and strict real-time requirements. It is noteworthy that existing research has rarely explored the integration of RL and ADRC in spacecraft thermal control. This work pioneers the application of Q-learning-ADRC to optomechanical thermal regulation. Comprehensive simulations comparing adaptive and fixed-parameter controllers, along with time-frequency domain analyses, validate the proposed system's effectiveness for precision thermal management in spaceborne optical systems.

1 Problem Formulation

1.1 Optomechanical thermal control system overview

The optomechanical system's core function is information acquisition. Photoelectric conversion transforms optical signals (reflected or radiated from target objects) into processable electrical signals for subsequent image processing and data analysis. The imaging performance is fundamentally governed by the inherent properties of the optical lenses and the thermal regulation precision.

Fig. 1 is a schematic of a typical optomechanical system. In this configuration, the lens exterior interfaces with external thermal loads from solar radiation, planetary albedo, and cold-space background radiation (about 3 K), while the interior interacts with thermal emissions from the system's internal environment. Heating elements are integrated into the lens mount for precise temperature stabilization, providing a thermal conduction framework that enables millikelvin-level temperature control through advanced algorithmic optimization of the controlled plant[16].

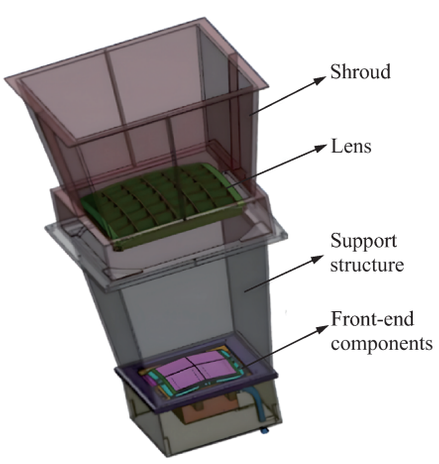

Fig. 1 Schematic of the optomechanical structure[17]

1.2 Simplified thermal dynamics model

This work adopts the thermal control model proposed by Li[17], which simplifies the mathematical representation of the optomechanical assembly to a first-order inertial system with pure delay. This simplification effectively captures the dominant low-frequency thermal behavior relevant to temperature regulation, while neglecting the higher-order transient processes that have a limited impact on closed-loop performance. Moreover, the ADRC framework uses an extended state observer (ESO) to compensate for unmodelled dynamics and external disturbances in real time, thereby ensuring robust and accurate control. Given that this simplified thermal model has been experimentally validated[17], this work focuses on optimizing the control strategy rather than re-identifying or remodeling the plant.

The energy-balance equation governing the lens-mount temperature regulation can be expressed as

$$C_{\mathrm{m}}\frac{\mathrm{d}T_{\mathrm{m}}}{\mathrm{d}t}=Q_{\mathrm{p}}+\alpha_1(T_1-T_{\mathrm{m}})+\alpha_2(T_2-T_{\mathrm{m}})+\frac{T_3-T_{\mathrm{m}}}{R_{\mathrm{jc}}} \tag{1}$$

where T_{m} represents the lens-mount temperature; Q_{p} denotes heater power; $C_{\mathrm{m}}=262.3$ J/(kg·℃) is the total heat capacity of the lens mount; $\alpha_1=1.515\times10^{-9}(T_{\mathrm{m}}^3+T_{\mathrm{m}}^2T_1+T_{\mathrm{m}}T_1^2+T_1^3)$ is the heat-transfer coefficient between T_{m} and the shroud temperature T_1, in W/K; $\alpha_2=0.055$ W/K is the heat-transfer coefficient between T_{m} and the main support structure temperature T_2; $R_{\mathrm{jc}}=1.897$ K/W gives the contact thermal resistance between T_{m} and the lens temperature T_3.

Applying the Laplace transformation to Eq. (1) yields the transfer function from heater power Q_{p} to lens-mount temperature T_{m} as

$$G_{\mathrm{p}}(s)=\frac{T_{\mathrm{m}}}{Q_{\mathrm{p}}}=\frac{1}{C_{\mathrm{m}}s+\alpha_1+\alpha_2+R_{\mathrm{jc}}^{-1}}\mathrm{e}^{-3s} \tag{2}$$

Disturbance terms from the shroud temperature T_1, main support structure temperature T_2, and lens temperatures T_3 to T_{m} are defined as

$$\begin{cases}G_1(s)=\dfrac{T_{\mathrm{m}}}{T_1}=\dfrac{\alpha_1}{C_{\mathrm{m}}s+\alpha_1+\alpha_2+R_{\mathrm{jc}}^{-1}}\mathrm{e}^{-3s}\\ G_2(s)=\dfrac{T_{\mathrm{m}}}{T_2}=\dfrac{\alpha_2}{C_{\mathrm{m}}s+\alpha_1+\alpha_2+R_{\mathrm{jc}}^{-1}}\mathrm{e}^{-3s}\\ G_3(s)=\dfrac{T_{\mathrm{m}}}{T_3}=\dfrac{R_{\mathrm{jc}}^{-1}}{C_{\mathrm{m}}s+\alpha_1+\alpha_2+R_{\mathrm{jc}}^{-1}}\mathrm{e}^{-3s}\end{cases} \tag{3}$$

1.3 Control system design objectives

The system is modeled as a four-input single-output configuration composed of first-order transfer functions. The ADRC framework regulates the heater power input, treating the remaining three thermal pathways as exogenous disturbances as defined in Eq. (3). The control system dynamically optimizes ADRC parameters via RL, outperforming fixed-parameter ADRC and SIMC-PI in the following scenarios:

① Setpoint tracking. Achieve engineering requirements for transient-response metrics (settling time t_s, overshoot σ, integral absolute error (IAE), and maximum sensitivity M_S) during temperature setpoint transitions.

② Disturbance rejection. Maintain lens-mount temperature stability within ±50 mK precision when subjected to periodic ±10 K variations in adjacent component temperatures.

③ Model robustness. Ensure rapid setpoint tracking and stable operation with minimal performance degradation under ±5% parameter perturbations in the controlled plant[18].

2 First-Order ADRC Scheme

Although nonlinear ADRC (NLADRC) can theoretically improve tracking accuracy and disturbance rejection, its implementation suffers from tuning complexity and uncertainty in stability analysis. Conversely, linear ADRC (LADRC) is more suitable for engineering applications, offering a simpler and analytically tractable structure. Therefore, this work employs a first-order LADRC as the baseline controller in the lens-mount heater-regulation system.

2.1 System architecture

Fig. 2 illustrates the Q-learning-ADRC framework. The state-space model of the first-order plant is

$$\begin{cases} y = x_1 \\ \dot{x}_1 = x_2 + b_0 u \\ \dot{x}_2 = \dot{f} \end{cases} \tag{4}$$

where x_1 and x_2 denote the system states; b_0 represents the estimated high-frequency gain; and f is the lumped disturbance, which is regarded as an extended state to be estimated in real time.

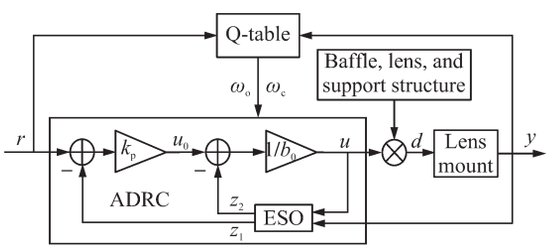

Fig. 2 Q-learning-ADRC architecture

To achieve dynamic estimation and compensation of the total disturbance[19], an ESO is designed as

$$\begin{cases} \hat{y} = z_1 \\ \dot{\hat{x}}_1 = \dot{z}_1 = \beta_1 (y - z_1) + z_2 + b_0 u \\ \dot{\hat{x}}_2 = \dot{z}_2 = \beta_2 (y - z_1) \end{cases} \tag{5}$$

where z_1 and z_2 represent the estimated system states; β_1 and β_2 denote the observer gains.

Based on the estimated states, the control law is formulated in the form of disturbance compensation as

$$\begin{cases} u_0 = k_p (r - z_1) \\ u = \dfrac{1}{b_0} (u_0 - z_2) \end{cases} \tag{6}$$

where r denotes the reference input and k_p denotes the proportional gain.

Key tunable parameters in this ADRC system include b_0, β_1, β_2, k_p, along with the Q-learning optimization framework. This architecture enables a systematic disturbance estimation and compensation process while maintaining analytical tractability.

2.2 Parameter tuning methodology

Conventional ADRC relies on manual pole placement to configure the observer and controller parameters. For a first-order LADRC, three parameters are particularly critical:

(1) Estimated high-frequency gain b_0. It influences the system's stability margin and control authority[20].

(2) Observer bandwidth ω_o. It determines the disturbance estimation speed and noise sensitivity[21].

(3) Controller bandwidth ω_c. It governs the closed-loop response speed and the degree of overshoot[22].

Due to nonlinear coupling among parameters, manual tuning cannot yield a global optimum. Therefore, a Q-learning algorithm is used to adaptively adjust the parameters, ensuring optimal operation under various conditions.

2.3 Performance-evaluation metrics

To quantitatively evaluate the control performance, the following metrics are selected:

(1) IAE[23]. It measures the cumulative deviation between the output and the reference.

(2) Maximum sensitivity[24]. It evaluates the system's robustness against model uncertainties.

(3) Gain margin. It quantifies how much the loop gain can increase before the system reaches instability.

(4) Phase margin. It measures the additional phase lag the system can tolerate before becoming unstable.

For comparison, the fixed-parameter ADRC controller employs median values from the RL action space. The proposed method's superiority is validated through IAE

reduction, maximum sensitivity optimization, and robustness testing under ±5% parameter perturbations. This evaluation framework ensures a comprehensive assessment of transient response, disturbance rejection, and model uncertainty tolerance.

3 Q-Learning-Based Controller Parameter Adaptation

Q-learning, an off-policy RL algorithm, updates a Q-table that maps state-action pairs to weight estimates for adaptive parameter adjustment[25], as illustrated in Fig. 3. In this figure, Δe represents instantaneous temperature deviation and Δr represents setpoint variation; ω_o and ω_c are the two parameters to be optimized in the ADRC framework; $\boldsymbol{s}_i$ and $\boldsymbol{a}_i$ represent the state vector and action vector at simulation interval i, respectively.

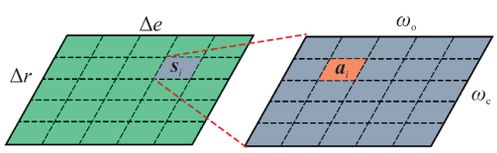

Fig. 3 Q-table (state-action mapping)

3.1 RL framework for ADRC tuning

The ADRC parameter optimization control is formulated as a Markov decision process, whose three fundamental components are defined as follows[26]:

(1) The state space captures real-time thermal regulation as

$$\boldsymbol{S} = [\Delta e, \Delta r]^{\mathrm{T}} \tag{7}$$

(2) The action space governs parametric adjustments according to the bandwidth tuning principle.

$$\boldsymbol{A} = [\omega_o, \omega_c]^{\mathrm{T}} \tag{8}$$

(3) Reward function is expressed as

$$R = \sum_{i=0}^{20}[-k_1(J - J_d) - k_2(|M_S - 1.4| + 0.8)^2 + k_3 G_M + k_4 P_M] \tag{9}$$

where J and J_d denote the IAE and deviation-compensated IAE, respectively, with J penalizing the cumulative tracking error and J_d accounting for deviations during setpoint transitions; M_S enforces robustness via the maximum sensitivity near the ideal range $[1.2, 1.6]$; G_M and P_M represent the gain margin and phase margin, serving as the rewards for stability margins; k_1, k_2, k_3, and k_4 are weighting coefficients, which ensure dimensional consistency and stable training. They can be adjusted to emphasize different control objectives according to the application scenario. Initially, the coefficients are normalized to balance the contribution of each term to the total reward, with typical magnitudes around 1, 1, 10^{-3}, and 10^{-3}. Although these coefficients have little impact on training speed, proper initialization helps prevent divergence during the early training stages. Fine-tuning is based on the final convergence of the training to ensure stable and effective learning.

This formulation enables autonomous adaptation to dynamic thermal loads while maintaining strict stability constraints, demonstrating superior control compared to fixed-parameter configurations in subsequent analyses. The Q-learning hyperparameters and training specifications are summarized in Table 1.

Table 1 Q-learning parameters

Parameter	Value
Learning rate α	0.99
Discount factor γ	0.5
Episodes	6 000
Initial ε_0	1
Decay rate of ε	1.98×10^{-4}
Interval i	60
Duration t/s	10
ω_o action range	[0.01, 0.2]
ω_c action range	[0.01, 0.1]

The procedure of the Q-learning-ADRC algorithm is compiled as follows.

Algorithm 1 Q-learning-ADRC

1 Initialize Q-table $Q(\boldsymbol{s}_0, \boldsymbol{a}_0)$
2 For episode = 1 to 6 000
3 Reset system state $\boldsymbol{s}_0 = [\Delta e_0, \Delta r_0]$
4 For each interval i = 1 to 10
5 Select action $\boldsymbol{a}_i = [\omega_{o,i}, \omega_{c,i}]$ via ε-greedy policy
6 Execute action $\boldsymbol{a}_i$ for 60 s, observe reward r_i and next state $\boldsymbol{s}_{i+1}$
7 Update Q-table: $Q(\boldsymbol{s}_i, \boldsymbol{a}_i) \leftarrow Q(\boldsymbol{s}_i, \boldsymbol{a}_i) + \alpha[r_i + \gamma \max Q(\boldsymbol{s}_{i+1}, \boldsymbol{a}) - Q(\boldsymbol{s}_i, \boldsymbol{a}_i)]$
8 Update state: $\boldsymbol{s}_i \leftarrow \boldsymbol{s}_{i+1}$

The reward moving average (calculated over every 20 episodes) demonstrates progressive policy optimization, achieving convergence after 6 000 training episodes, as shown in Fig. 4.

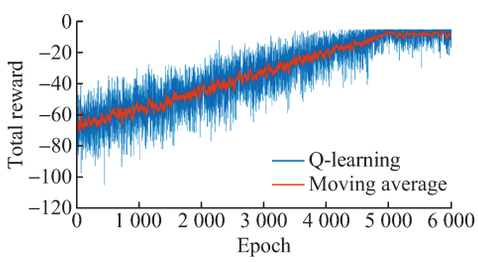

Fig. 4 Cumulative reward curve

3.2 Offline simulation in a composite training scenario

Fig. 5 presents the time-domain performance of the Q-learning-ADRC system under multicondition testing. A multisegment reference trajectory is designed to simulate alternating heating and cooling phases, representing orbital thermal load variations such as solar heating and deep-space cooling.

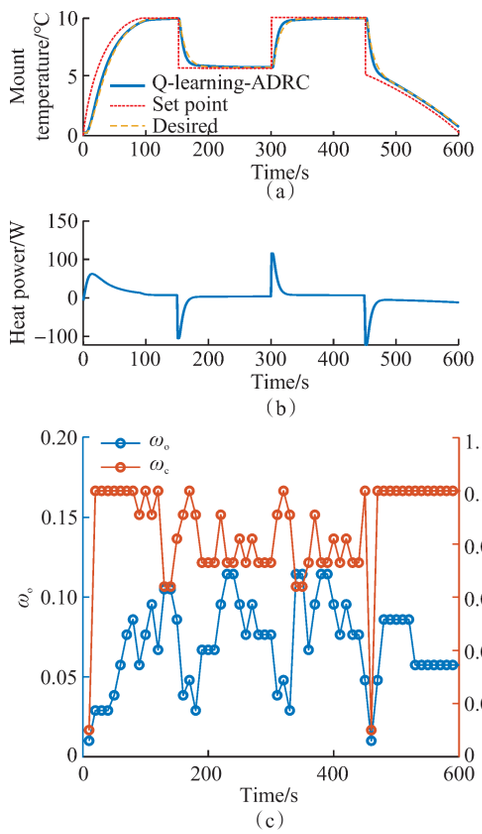

Fig. 5 Offline simulation results. (a) Multisegment tracking performance; (b) Control inputs; (c) Adaptive adjustment of ADRC parameters

As shown in Fig. 5, the ADRC system with a Q-learning parameter tuning strategy exhibits superior control performance, with a total IAE value of 331.77 calculated from the simulation results.

3.3 Stability and sensitivity analysis

The graph of the test scenario frequency-domain metrics in Fig. 6 validates the controller's robust stability. It can be observed from Fig. 6 that the maximum sensitivity remains within the ideal range [1.2, 2.0]. The gain margin (>10 dB) and phase margin (>30°) are maintained throughout the training process. The Q-learning-ADRC controller exhibits moderate sensitivity during dynamic parameter changes, while not being overly sensitive to disturbance, and has a phase margin and gain margin in line with engineering applications.

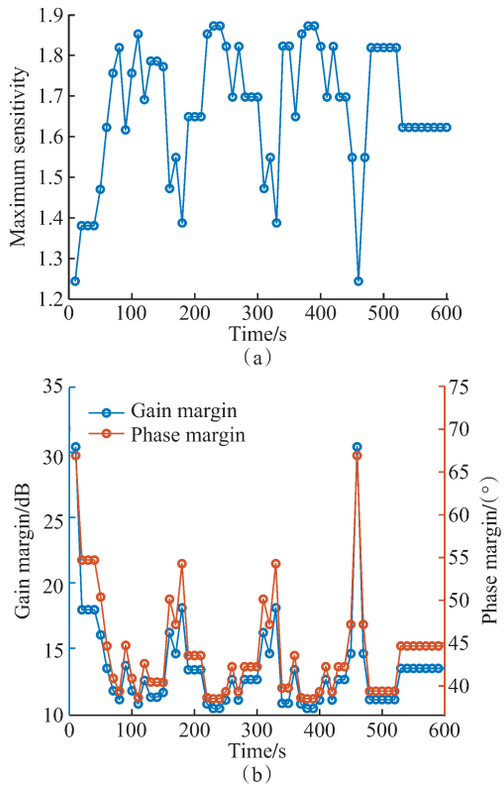

Fig. 6 Dynamic changes during controller parameter adaptation of maximum sensitivity, gain margin, and phase margin in training scenarios. (a) Maximum sensitivity variation; (b) Gain margin and phase margin variations

3.4 ESO observer performance

Fig. 7 illustrates the estimation performance of the ESO for both total and output disturbances. As shown in Fig. 7, the Q-learning-ADRC controller achieves smaller estimation errors than the fixed-parameter ADRC, with its error trajectories are closer to zero.

To quantify the estimation accuracy, the estimation error statistics are summarized in Table 2. The Q-learning-ADRC controller achieves a total disturbance

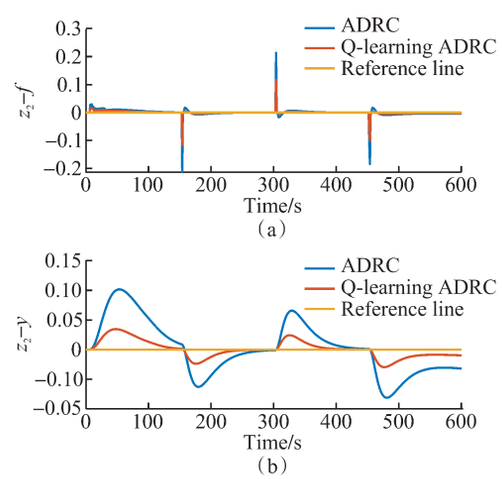

Fig. 7 ESO tracking performance comparison. (a) Comparison of total disturbance estimation error; (b) Comparison of output estimation error

Table 2 ESO estimation error performance comparison

Controller	Disturbance $z_1 - y$	Output $z_2 - f$
Q-learning-ADRC	5.86	1.50
Fixed-parameter ADRC	17.57	2.28

estimation error of 5. 86 and an output estimation error of 1. 50, which are significantly lower than those of the fixed-parameter ADRC. These results further confirm that integrating the ESO with Q-learning effectively improves disturbance estimation accuracy.

4 Comparative Simulation

4. 1 Tracking under setpoint steps

During orbital operations, the lens-mount temperature requires dynamic adjustments to accommodate varying imaging tasks. Fig. 8 compares the performance of the Q-learning-ADRC controller with the fixed-parameter ADRC controller (using median action values) and a SIMC-PI controller under setpoint step changes.

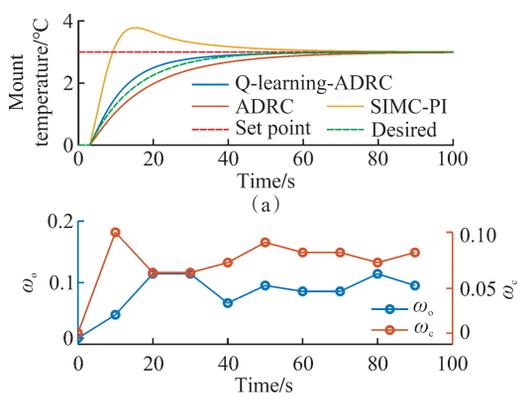

Fig. 8 Setpoint tracking performance. (a) Tracking performance under setpoint changes; (b) Adaptive adjustment of ADRC parameters

The key performance metrics summarized in Tables 3 and 4 demonstrate the superiority of the Q-learning-ADRC controller.

Table 3 Dynamic performance comparison

Controller	IAE	t_s/s	σ/%
Q-learning-ADRC	39.77	49.4	0.00
Fixed-parameter ADRC	58.18	72.0	0.00
SIMC-PI	34.30	58.4	25.95

Table 4 Robustness comparison

Controller	Maximum sensitivity	Gain margin/dB	Phase margin/(°)
Q-learning-ADRC	1.708	12.507	42.155
Fixed-parameter ADRC	1.730	11.709	41.872
SIMC-PI	1.694	9.469	47.763

A comparative analysis of Table 3 and Fig. 8 shows that under the setpoint step condition, the Q-learning-ADRC controller has significantly better control performance than the traditional fixed-parameter ADRC and SIMC-PI:

(1) The settling time is shortened by 31. 3% and 15. 4%, respectively, and the dynamic response is faster.

(2) The overshooting amount is reduced by 26% compared to the SIMC-PI system, and the stability of the system is strengthened.

(3) The IAE is reduced by 21. 8% compared to the fixed-parameter ADRC system, and the control accuracy is improved.

In addition, the robustness comparison in Table 4 demonstrates that the system preserves adequate stability margins under varying operating conditions, confirming the effectiveness of the proposed Q-learning-based parameter-tuning algorithm in real time during setpoint tracking.

4. 2 Stability against periodic and step disturbances

During the satellite's operation in orbit, the shroud temperature exhibits cyclical changes: the initial temperature is 250 K, and under the cyclic influence of solar irradiation, its temperature change behavior can be described as follows: for the first 40 min, it rises by 10 K, followed by a drop of 10 K in the next 80 min, and then returns to the initial temperature of 250 K for the last 40 min, forming a complete 160-min cycle of change, with a temperature fluctuation amplitude of ± 10 K. This periodic temperature fluctuation can be approximated as a sinusoidal function with an amplitude of 10 K. The temperature of the hood is maintained in the range of 250 ± 10 K, and the heat is transferred to the frame by thermal radiation, which has a significant effect on the frame temperature. Based on the above analysis, the temperature perturbation can be modeled as the following sinusoidal function form:

$$T_1 = 250 \pm 10\sin\left(\frac{\pi}{80}t\right) \tag{10}$$

In addition, time series plots of the two-parameter trajectories for the three controllers are shown in Fig. 9, where black dashed vertical lines indicate the times at which −30 and 5 K steps occur in T_2 and T_3, respectively.

The maximum amplitude and deviation integrals of the three controllers are shown in Table 5. As shown in Fig. 9 and Table 5, the ADRC system based on Q-learning exhibits superior control performance under the effect of periodic external perturbations: the system response time is significantly shortened, and the maximum amplitude of the temperature fluctuations is effectively controlled within the range of ±40 mK. Compared with the fixed-parameter ADRC and SIMC-PI control systems, the Q-learning-ADRC controller achieves a substantially lower IAE value. In addition, the system's phase mar-

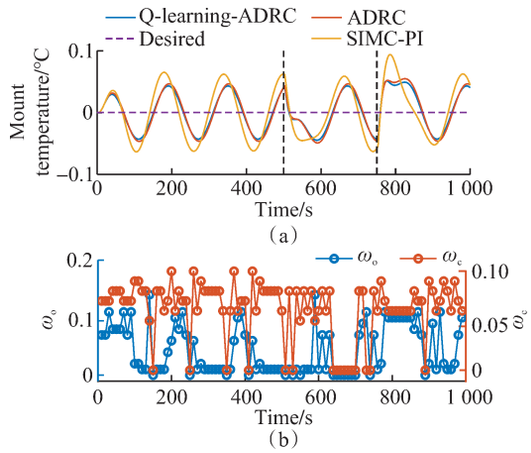

Fig. 9 Disturbance rejection performance. (a) Tracking performance under disturbances; (b) Adaptive adjustment of ADRC parameters

Table 5 Tracking accuracy comparison

Controller	IAE	Peak deviation/K
Q-learning-ADRC	27.53	0.040 4
Fixed-parameter ADRC	29.64	0.046 9
SIMC-PI	38.28	0.097 2

gin and gain margin are verified by stability analysis to meet the requirements of engineering applications, verifying the reliability of the proposed method.

The simulation experiment results show that there is a significant difference in the control accuracy of the three systems under the fluctuating working conditions of shroud temperatures: the traditional fixed-parameter ADRC control system has an accuracy of ±46 mK, and the SIMC-PI system has an accuracy of ±97 mK, while the Q-learning-ADRC control system improves the control accuracy to ±40 mK, which is 12.7% and 52.5% higher, respectively. This result proves that the proposed Q-learning-ADRC control system can effectively achieve the target control accuracy requirements and provides a reliable technical solution for orbital thermal control of satellites.

4.3 Robustness analysis through perturbation simulations

To comprehensively evaluate controller robustness, we implement Monte Carlo simulations with ±5% perturbations in the following parameters[27]: (1) Lens-mount total heat capacity C_m; (2) Frame-shroud heat transfer coefficient α_1; (3) Frame-support heat transfer coefficient α_2; (4) Frame-lens contact thermal resistance R_{jc}; (5) System time delay L.

Fig. 10 illustrates the parameter variation ranges across 100 randomized trials, while Fig. 11 statistically analyzes the performance metrics. Based on the experimental results, the distribution of the total system error IAE, settling time t_s, and overshoot σ is plotted in Fig. 11.

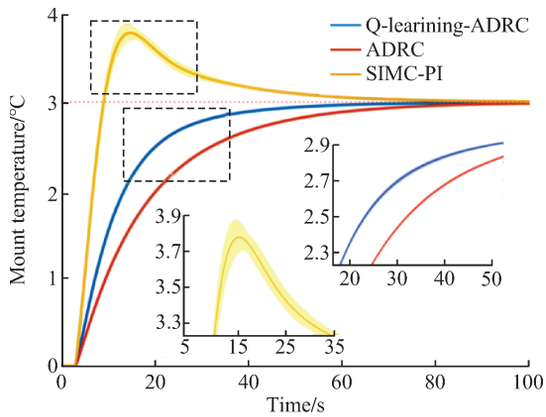

Fig. 10 Parameter perturbation ranges of ±5%

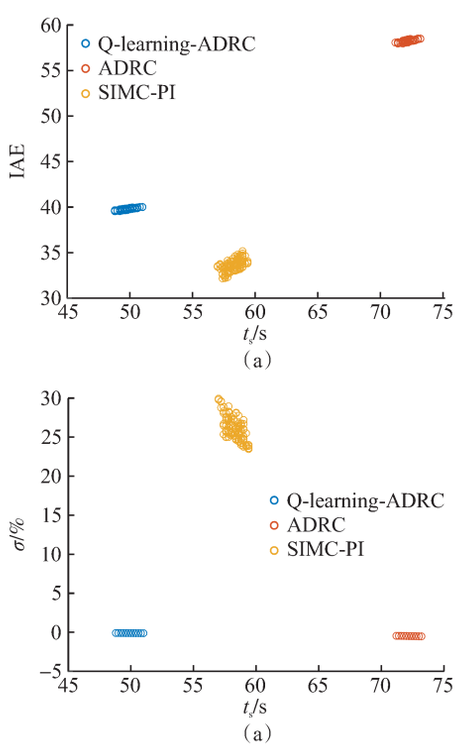

Fig. 11 Performance metric distributions under model uncertainties. (a) IAE-settling time distribution; (b) Overshoot-settling time distribution

Statistical analysis of Fig. 11 reveals that the Q-learning-based ADRC control system exhibits superior robustness when the controlled object parameters vary within ±5%: the fluctuation ranges of the settling time and IAE are reduced, compared to the fixed-parameter ADRC controller and the SIMC-PI controller. The experimental results confirm that the Q-learning mechanism effectively enhances the ADRC controller's ability to observe and compensate for the unmodelled system dynamics and significantly improves the tolerance of the control system to model uncertainty. Specifically, the introduction of RL enables the control system to adaptively compensate for parameter perturbation effects, effectively suppress performance fluctuations caused by the unmod-

elled dynamics, and maintain stable control despite the presence of model uncertainty. These properties make the proposed method particularly suitable for orbital thermal control scenarios with complex operating conditions and uncertainties.

5 Conclusions

To address the high-precision temperature control requirements for optomechanical systems in space exploration satellites, this work proposes a Q-learning-ADRC scheme. Through systematic theoretical analysis and simulation validation, the method achieves real-time adaptive optimization of ADRC parameters, significantly enhancing control performance under complex operating conditions. The three main conclusions are as follows:

(1) The proposed Q-learning-ADRC control system demonstrates superior dynamic characteristics. Compared with fixed-parameter ADRC and SIMC-PI under the setpoint step condition, the proposed system exhibits superior dynamic performance: the settling time is reduced by 31.3% and 15.4% relative to the fixed-parameter ADRC and SIMC-PI, respectively, while the overshoot decreases by 26% compared with SIMC-PI and the IAE by 21.8% compared with the fixed-parameter ADRC.

(2) The system maintains excellent stability under periodic external disturbances with control precision within ±40 mK, representing a 12.7% and 52.5% improvement over conventional fixed-parameter methods and SIMC-PI methods, respectively.

(3) Monte Carlo simulations indicate that the proposed Q-learning-ADRC controller yields a 3%-6% reduction in the fluctuation ranges of settling time and IAE compared with fixed-parameter ADRC and SIMC-PI controllers under ±5% variations in system parameters, indicating enhanced robustness to model uncertainties.

Although the proposed Q-learning-ADRC framework exhibits strong potential in simulation studies, its practical deployment remains challenging. The online exploration and high-frequency updates inherent to the Q-learning algorithm may impose real-time constraints on satellite hardware with limited computational resources. Future work will focus on lightweight algorithm design, state-space discretization strategies, and hardware-in-the-loop testing to evaluate the framework's real-time performance in embedded systems. This work presents an innovative solution for high-precision temperature control of optomechanical systems, offering both significant theoretical insights and practical engineering value for advancing spacecraft thermal management technologies.

References

[1] GAO J X, SONG Y S, LIU Y. Application of nonlinear PID self-immunity control in temperature control system of fast mirror[J]. Laser & Optoelectronics Progress, 2023, 60(5): 0523001. (in Chinese)

[2] WEN M X, LI J, WANG C, et al. Summary of high precision temperature sensing, measurement and control technology[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2021, 60(S1): 146-155. (in Chinese)

[3] YU F, XU N N, ZHAO Y, et al. Design and validation of thermal control system of Gaofen-4 satellite camera[J]. Space Return and Remote Sensing, 2016, 37(4): 72-79. (in Chinese)

[4] HIETA T, MERIMAA M. Spectroscopic measurement of air temperature[J]. International Journal of Thermophysics, 2010, 31(8): 1710-1718.

[5] AARON K M, HASHEMI A, MORRIS P A, et. Space Interferometry Mission thermal design[C]//Astronomical Telescopes and Instrumentation. Waikoloa, HI, USA, 2003: 279.

[6] TONG Y L, LI G Q, GENG L Y. Current status of research on precision temperature control technology for spacecraft[J]. Space Return and Remote Sensing, 2016, 37(2): 1-8. (in Chinese)

[7] ZHAO Z M, LU P, SONG X Y. Design and validation of thermal control system for Gaofen-2 satellite camera [J]. Space Return and Remote Sensing, 2015, 36(4): 34-40. (in Chinese)

[8] GILMORE D. Spacecraft thermal control handbook, Volume Ⅰ: Fundamental technologies[M]. Washington, DC, USA: American Institute of Aeronautics and Astronautics, Inc., 2002.

[9] TONG Y L, LI G Q, YU L, et al. Application of PI control for precision temperature control of space camera [J]. Space Return and Remote Sensing, 2012, 33(4): 42-49. (in Chinese)

[10] DE PALO S, CAIROLA M, COMPASSI M, et al. Herschel heaters control modeling and correlation[J]. SAE International Journal of Aerospace, 2009, 4(1): 29-39.

[11] HAN J Q. From PID to active disturbance rejection control[J]. IEEE Transactions on Industrial Electronics, 2009, 56(3): 900-906.

[12] PAN C, YE Y, GU B Z, et al. Temperature control of the extinction cylinder of a 2.5 m large-field-of-view high-resolution telescope[J]. Infrared and Laser Engineering, 2023, 52(9): 20230024. (in Chinese)

[13] YUN Z R, WANG Z G, WANG J H. ADRC-based temperature control system for blackbody radiation sources [J]. Infrared Technology, 2019, 41(3): 232-238. (in Chinese)

[14] SIVAMAYIL K, RAJASEKAR E, ALJAFARI B, et al. A systematic study on reinforcement learning based applications[J]. Energies, 2023, 16(3): 1512.

[15] WILSON C, RICCARDI A. Improving the efficiency of reinforcement learning for a spacecraft powered descent with Q-learning[J]. Optimization and Engineering, 2023, 24(1): 223-255.

[16] YU B, LI C L, YANG T, et al. A high-precision temperature control method based on thermal characteristics

of space camera[J]. Aerospace Return and Remote Sensing. 2014, 35(3): 84-89. (in Chinese)
[17] LI S. Research on high stability temperature control technology for optical machines[D]. Shanghai: University of Chinese Academy of Sciences, 2021. (in Chinese)
[18] ZHAO S, SHI H W, LIU X S, et al. Hydraulic servo flow control with third-order linear self-immunity controller[J]. Hydraulic and Pneumatic 2021, 45(5): 149-156. (in Chinese)
[19] ZHAO X J, ZHU J, LUO X. Application of ADRC in lower limb rehabilitation training apparatus[J]. Journal of Southeast University (Natural Science Edition), 2019, 49(6): 1026-1032. (in Chinese)
[20] JIN H Y, SONG J C, LAN W Y, et al. On the characteristics of ADRC: A PID interpretation[J]. Science China Information Sciences, 2020, 63(10): 209201.
[21] WANG X P, ZHAO J, WANG B H, et al. Predictive current control system of PMSM based on LADRC[J]. Journal of Southeast University (English Edition), 2022, 38(3): 227-234.
[22] BAE Y, LEE S, YOON K J, et al. Three-dimensional dynamic modeling and transport analysis of solid oxide fuel cells under electrical load change[J]. Energy Conversion and Management, 2018, 165: 405-418.
[23] DAI W. Structural design and numerical simulation for high-precision sounding temperature sensor[J]. Transducer and Microsystem Technologies, 2022, 41(11): 5-8, 17. (in Chinese)
[24] CHENG D X, CHEN Z F, SU D W, et al. Stability analysis and robustness improvement of high-precision thermostat[J]. Journal of Hefei University of Technology (Natural Science Edition), 2022, 45(9): 1160-1164. (in Chinese)
[25] TARUN A K, CHUNDAWAT V S, MANDAL M, et al. Fast yet effective machine unlearning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 13046-13055.
[26] ZHANG J, JIANG X, SHI X Y, et al. Offline reinforcement learning for eco-driving control at signalized intersections[J]. Journal of Southeast University (Natural Science Edition), 2022, 52(4): 762-769. (in Chinese)
[27] ZHANG Y Q, LI D H. Active disturbance rejection control on a bubbling fluidized bed[J]. Journal of University of Science and Technology of China, 2012, 42(5): 391-397. (in Chinese)

基于强化学习自抗扰控制的光学机架温度调节

顾燕萍[1], 张好[1], 徐涛[1], 钱斌[2]

(1.上海卫星工程研究所,上海 201109; 2.上海航天技术研究院,上海 201109)

摘要: 空间光机系统在轨运行过程中面临极端热扰动与毫开尔文级温度控制精度的双重挑战,对控制策略的鲁棒性提出了极高要求。针对传统固定参数自抗扰控制(ADRC)在复杂工况下性能受限的问题,本文提出一种基于Q学习的自适应ADRC框架。结合强化学习驱动的参数优化机制,建立了包含多源热扰动(太阳辐射、结构传导及接触热阻)的热传递模型。通过ε-贪婪策略实时调整观测器带宽($\omega_o \in [0.01, 0.2]$)与控制器带宽($\omega_c \in [0.01, 0.1]$),实现对总扰动的动态估计与补偿。仿真结果表明,与固定参数ADRC及SIMC-PI控制相比,本方法在设定值响应过程中可分别缩短31.3%和15.4%的调节时间;在阶跃响应下,积分绝对误差(IAE)较固定参数ADRC降低21.8%;在±10 K周期扰动与阶跃扰动下,控制精度分别提高12.7%和52.5%。蒙特卡洛鲁棒性试验结果显示,在±5%参数摄动下,IAE、调节时间及超调量的波动范围显著减小。该方法为空间光学载荷的毫开尔文级高精度热控提供了一种新的控制范式。

关键词: 光机系统; 自抗扰控制; Q学习; 高精度温度控制

中图分类号: TP273.2