A Novel Attitude Controller for Hypersonic Aircraft Based on FR-PI2

Jiajun Fan[1], Tianyi Li[1], Xian Guo[1], Mingrui Hao[2], and Mingwei Sun[1]

[1] College of Artificial Intelligence, Nankai University, Tianjin 300350, China
guoxian@nankai.edu.cn
[2] Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, Bejing 100074, China

Abstract. Due to the uncertainties and various noises of the Hypersonic Aircraft, traditional method of designing attitude controllers based on accurate models faces immense challenges. To address these problems effectively, in this paper, a novel data-driven RL algorithm called FR-PI2 is proposed for the task of Hypersonic Aircraft attitude control, wherein three kernel techniques are introduced. Firstly, Hypersonic Aircraft attitude control is transformed into a stochastic optimal control problem, and a generalized path-integral-control approach is proposed to obtain the numerical solution of the stochastic dynamical system. Secondly, a model-free parameterization method using PID controller is introduced into the RL algorithm to ensure a low dependency of the system model. Thirdly, a brand-new sample filtration method is proposed to guarantee rapid convergence, and an online optimization technology called rolling optimization is applied to guarantee optimality. Simulation and experimental results are included and analyzed to illustrate the superior performance of the proposed algorithm.

Keywords: Attitude control, Hypersonic aircraft, PID controller, Path integral, Reinforcement learning (RL), Rolling optimization, Sample filtration

1 INTRODUCTION

Hypersonic Aircraft (HA) has been the focus of competition for air and space rights owing to its important strategic position. One key issue to ensure the practicality of Hypersonic Aircraft is the design of the Hypersonic Aircraft attitude control system. Because of the characteristic of Hypersonic Aircraft system[6] such as highly non-linear, strong coupling, uncertainty, fast time-varying, etc., few traditional control methods can be applied directly to the design of Hypersonic Aircraft attitude control system. How to achieve effective attitude control of Hypersonic Aircraft has become a very challenging problem in the field of control science.

There is a large body of work on attitude control for Hypersonic Aircraft, wherein some model-based controllers using a model predictive control (MPC)

scheme [3][4] can not only ensure system stability but also achieve its optimality. However, these algorithms highly depend on an accurate system model. Moreover, involving a large amount of calculation makes it unpractical for real-time control. In contrast, the proportional-integral-derivative (PID) controller, as a conventional model-free method, is widely used for different control tasks. We can apply it to the attitude control of Hypersonic Aircraft because of its low dependency on the accurate system model and strong robustness. Unfortunately, massive time has been spent on seeking out the optimal parameters of the PID controller to ensure good performance, which brings difficulty for real-time implementation. By contrast, controller designed with the RL algorithm, can not only maintain low dependence on the system model but also guarantee the optimality.

Reinforcement learning has been recently employed to deal with various control problems[2]. It first maps situations to actions via an offline exploration and exploitation method to maximize the reward function closely associated with a given task automatically. Afterward, the offline obtained optimal controls can be used online for the investigated system[8]. Different from traditional RL, path-integral-based RL, say PI2, within the framework of stochastic optimal control, requires far fewer iterations and guarantees optimality and reliable training convergence[5]. Nevertheless, the PI2 algorithm can easily fall into local optimum when the optimization function contains massive saddle points, which also result in degradation of convergent speed and real-time performance.

The task of this paper is to achieve completely automatic control of Hypersonic Aircraft pitch attitude. To fulfill this task, in this paper, a generalized path-integral-control approach is used for the stochastic optimal control of Hypersonic Aircraft attitude. Moreover, a novel learning algorithm called FR-PI2 is designed to find optimal control parameters for Hypersonic Aircraft attitude controller.

Compared with existing results, the main contributions of this paper lie in the following aspects.

1. A novel parameterization using PID techniques is introduced to RL algorithm. This combination bridges traditional control and intelligent learning control, which simultaneously guarantees the robustness and the optimality.
2. A data-driven RL algorithm called FR-PI2 is used to seek out the optimal PID controller, thereby addressing the attitude control problem of Hypersonic Aircraft system.
3. A sample-filtering PI2 method is used to obtain optimal control which can extensively increase convergence rate and avoid local optimum during training.
4. Similar to the MPC algorithm for calculating the optimal control online according to the system states, optimum parameters for different control stages are tuned online with real-time system states, leading to improved global optimality. Specifically, compared with PI2, the proposed method called FR-PI2 requires less computation and has strong global stability.

Additionally, the parameterization method using PID techniques and the policy update method based on the FR-PI2 algorithm can be generalized for a class of control systems, such as mobile robot systems, in addition to Hypersonic Aircraft.

The remainder of this paper is organized in the following manner. In Section 2, we have established a stochastic optimal control model for Hypersonic Aircraft pitch attitude control, and a path-integral-control method is proposed to solve this stochastic optimal control problem. In Section 3, A model-free parameterization is used for RL. Subsequently, the design of the FR-PI2 RL algorithm, specifically for the pitch attitude control problem of Hypersonic Aircraft, is presented. A simulation comparison and reliability analysis of the proposed algorithm are given in Section 4. Section 5 gives some concluding remarks and discusses future work.

2 Problem Formulation

Winged-cone model is a standard platform for studying Hypersonic Aircraft control systems[1], which however brings unnecessary difficulty and calculation pressure to the designing work and analysis because of its complexity. In this section, a linear-simplified longitudinal pitch angle of Hypersonic Aircraft system dynamic model is proposed first, and then, the analysis for stochastic optimal control for general Hypersonic Aircraft system is illustrated, and a generalized path-integral-control approach is employed to obtain the numerical solution of the stochastic dynamical system.

2.1 Dynamic equation

To simplify the Winged-cone model, a linear model is proposed of the independent channel based on the principle of small disturbance linearization by the premise of guarantee of the same control effect[6], which is expressed as follows

$$\begin{cases} \ddot{\theta} + a_{\omega_z}\dot{\theta} + a_\alpha \alpha = -a_{\delta_z}\delta_z \\ \dot{\vartheta} - b_\alpha \alpha = b_{\delta_z}\delta_z \\ \theta = \vartheta + \alpha \\ q = \dot{\theta} \end{cases} \tag{1}$$

After differentiating the third equation and substituting the second equation into the obtained differential equation to eliminate ϑ, a second-order partial differential equations is presented as follows

$$\begin{cases} \ddot{\theta} + a_{\omega_z}\dot{\theta} + a_\alpha \alpha = -a_{\delta_z}\delta_z \\ \dot{\alpha} = \dot{\theta} - b_\alpha \alpha - b_{\delta_z}\delta_z \\ q = \dot{\theta} \end{cases} \tag{2}$$

Organizing **formula** (2), a liner-simplified model is given as follows

$$\begin{cases} \dot{\alpha} = q - b_{\alpha}\alpha - b_{\delta_z}\delta_z \\ \dot{\theta} = q \\ \dot{q} = -a_{\omega_z}q - a_{\alpha}\alpha - a_{\delta_z}\delta_z \end{cases} \tag{3}$$

Table 1 gives the notations used in this paper, and constant in **formula** (3) can be found in [6].

Table 1. NOTATIONS USED

Notation	Description
θ	Pitching angle
ϑ	Trajectory inclination angle
α	Attack angle
δ_z	Elevator angle
q	Pitching angular velocity
x	State vector
u	Control vector (input vector)
e	The difference between θ and θ_{desire}
$R(\tau_i)$	Path cost
$P(\tau_i)$	Path association probability
σ	Overshoot
t_s	Adjusting time or stability time
t_f	Total simulation time
$\mathcal{K}$	Parameter vector Learned for PID controller
N	Total number of noisy trajectories or roll-outs
M	Total number of updates
Σ	Variance matrix
β	Variance update factor
M_β	Variance update interval
M_O	Rolling optimization interval
$N(0,\sigma)$	Gaussian distribution with variance σ and zero expectation
λ	PI2 hyper-parameters

2.2 Stochastic Dynamical System

The general stochastic dynamical system is expressed as follows

$$\begin{aligned} \dot{\boldsymbol{x}}_{\boldsymbol{t}} &= \boldsymbol{f}(\boldsymbol{x}_{\boldsymbol{t}}, t) + \boldsymbol{G}(\boldsymbol{x}_{\boldsymbol{t}})(\boldsymbol{u}_{\boldsymbol{t}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}}) \\ &= \boldsymbol{f}_{\boldsymbol{t}} + \boldsymbol{G}_{\boldsymbol{t}}(\boldsymbol{u}_{\boldsymbol{t}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}}) \end{aligned} \tag{4}$$

where $\boldsymbol{u}_t \in \Re^{p\times 1}$ represents the control vector, $\boldsymbol{f}_t \in \Re^{n\times 1}$ stands for the system dynamic equation, $\boldsymbol{G}_t \in \Re^{n\times p}$ donates the control matrix, $\boldsymbol{x_t} \in \Re^{n\times 1}$ is the systems state, and $\boldsymbol{\varepsilon_t} \in \Re^{p\times 1}$ is the Gaussian noise submitting to N(0, Σ_ε).

For the proposed Hypersonic Aircraft system, its general stochastic dynamical system is given as **formula** (3), wherein $\boldsymbol{x_t} = \left[\alpha, \theta, q\right]^T$, $\boldsymbol{G_t} = \left[-b_{\delta_z}, 0, -a_{\delta_z}\right]^T$, $\boldsymbol{u_t} = \delta_z$, $\boldsymbol{\varepsilon_t} = \varepsilon_t$, $\boldsymbol{f_t} = \left[-b_\alpha \alpha + q, q, -a_{\omega_z} q - a_\alpha \alpha\right]^T$.

2.3 Stochastic Optimal Control

The performance criterion function for a path τ_i starting at time t_i in state x_t^i and ending at time t_f is defined as follows[8]

$$J\left(\tau_i\right) = \phi_{t_f} + \int_{t_i}^{t_f} L\left[x_t, u_t, t\right] dt \tag{5}$$

Wherein $\phi_{t_f} = \phi\left(\boldsymbol{x}_{t_f}, t_f\right)$ is terminal value, $\int_{t_i}^{t_f} L\left[x_t, u_t, t\right] dt$ stands for a process value. The immediate cost is defined as follows

$$L_t = L\left[x_t, u_t, t\right] = q_t + \frac{1}{2} u_t^T \boldsymbol{R} u_t \tag{6}$$

where $q_t = q\left(\boldsymbol{x}_t, t\right)$ is the cost function related to states, $\boldsymbol{R}$ represents coefficient matrix. In this paper, our task is to make the pitch angle of Hypersonic Aircraft follow its expected value swiftly and stably, therefore $J\left(\tau_i\right) = \sigma + t_s$. The goal of stochastic optimal control is to seek out the control u_t to minimize the following performance criterion function

$$\min \quad V\left(\boldsymbol{x}_t^i\right) = V_t^i = \min_{u_t^i : t_f} E_{\tau_i}\left[J\left(\tau_i\right)\right] = \min_{u_t^i : t_f} E_{\tau_i}\left[\sigma + t_s\right] \tag{7}$$

where $E_{\tau_i}[\cdot]$ is the expectation of all trajectories starting from $\boldsymbol{x}_t^i$. It is very difficult to obtain the analytical solution for optimization problem (7), instead, with path integral algorithm, which is a numerical method used to solve stochastic optimal control problems, for which the goal is to minimize a performance criterion for a stochastic dynamical system[8], we can propose its numerical solutions after iteration convergence of **formula** (8).

$$u_t^i = \int P\left(\tau_i\right) u\left(\tau_i\right) d\tau_i \tag{8}$$

where $P\left(\tau_i\right)$ is defined[5] as follows

$$P\left(\tau_i\right) = \frac{e^{-\frac{1}{\lambda} S(\tau_i)}}{\int e^{-\frac{1}{\lambda} S(\tau_i)} d\tau_i} \tag{9}$$

For the convenience of the implement, **formula** (9) can be approximate to (10)

$$P\left(\tau_i\right) = \frac{e^{-\frac{1}{\lambda} S(\tau_i)}}{\sum_{i=1}^{N} e^{-\frac{1}{\lambda} S(\tau_i)} d\tau_i} \tag{10}$$

wherein $S(\tau_i)$ is a normalized version of the path cost defined as follows

$$S(\tau_i) = \frac{R(\tau_i) - \min(\boldsymbol{R})}{\max(\boldsymbol{R}) - \min(\boldsymbol{R})} \tag{11}$$

where $R(\tau_i)$ represents the loss function similar to $J(\tau_i)$, $\boldsymbol{R}$ represents loss function vector for N trajectories. According to the literature[6], Hypersonic Aircraft system state x, as well as its control input δ_z, has to satisfy the constraint condition. Thus the constrained Hypersonic Aircraft attitude stochastic optimal control problem can be defined as follows

Definition 1. *Constrained Stochastic Optimal Control Problem*

$$\begin{aligned} \min \quad & V(\boldsymbol{x}_{t_t}) = V_{t_t} = \min_{u_{t_i}:t_f} E_{\tau_i}\left[J(\tau_i)\right] = \min_{u_{t_i}:t_f} E_{\tau_i}\left[\sigma + t_s\right] \\ s.t. & \begin{cases} \dot{q} = -a_{\omega_z} q - a_\alpha \alpha - a_{\delta_z}(\delta_z + d\omega_z) \\ \dot{\theta} = q \\ \dot{\alpha} = q - b_\alpha \alpha - b_{\delta_z}(\delta_z + d\omega_z) \end{cases} \\ & \alpha \in [-1^\circ, 10^\circ], \quad \delta_z \in [-20^\circ, 20^\circ], |\dot{\delta}_z| \leq 50 \end{aligned} \tag{12}$$

Wherein ω_z is the Wiener noise, $d\omega_z$ is Gaussian noise.

It is obviously perceived from **formula** (7) to (11) that the generalized path-integral-control approach for the numerical solution of a stochastic dynamical problem like **formula** (12) avoids the calculation for the gradient and matrix inversion[8], thereby ensuring fast and reliable convergence, which is demonstrated in Section 3.

3 Reinforcement Learning

As a data-driven RL method, FR-PI2 is used to reduce trajectory loss function through exploration and exploitation. Instead of directly randomly parameterizing the control $u(\tau_i)$, which requires long interaction with the system, $u(\tau_i)$ is parameterized by using the PID method, to deal with various noises. Furthermore, only three parameters are able to completely represent $u(\tau_i)$, which makes much faster convergence during RL training.

3.1 Policy Parameterization

It is very difficult to obtain an accurate model for Hypersonic Aircraft system, which causes great performance depreciation for many model-based controllers. Thus, a model-free controller called PID is proposed to parameterize control policy as **formula** (13).

$$u_t = k_p * e + k_d * \dot{e} + k_i * \int e \, dt \tag{13}$$

wherein $e = \theta_{desire} - \theta$, $\mathcal{K} = \left[k_p, k_d, k_i\right]^T$. $\mathcal{K}$ represents the control parameters, and PI2 is a learning mechanism to seek out an optimal combination of control parameters, say $\mathcal{K}^*$.

3.2 PI2 Algorithm

In PI2 implementation here the parameter vector to be learned, say $\mathcal{K}$, is updated at the end of every update. Each update consists of N noisy trajectories or roll-outs. M updates are performed to obtain the optimal parameter vector.

Loss Function Overshoot and adjusting time is introduced to the design of loss function as the evaluation index of the stability of and rapidity of a path. However, multi-objective optimization is generally difficult to treat for normal optimization methods. Therefore, a target-oriented loss function is defined as follows

$$R(\tau_i) = \frac{\sigma - \sigma_{target}}{\sigma_{target}} + \frac{t_s - t_s^{target}}{t_s^{target}} \tag{14}$$

where t_s^{target} represents the desired adjusting time setting as $3s$ and σ_{target} stands for the desired overshoot, which is set as $0.001°$.

Implementation of PI2 More details of PI2 implementation will be discussed here, such as the process of sample exploration and exploitation.

In the starting state x_t^i at time t_i. which is set as zero here, the controls u_t of the ith path are randomly generated using the random parameter vector $\mathcal{K} + \boldsymbol{\varepsilon_i}$ with $\varepsilon_i = \left[\varepsilon_i^{(1)}\ \varepsilon_i^{(2)}\ \varepsilon_i^{(3)}\right]^T$ being the Gaussian noise vector of the ith path and Hypersonic Aircraft pitch state is defined as

$$u_t = \left[e, \dot{e}, \int e\right] \begin{bmatrix} k_p + \varepsilon_i^{(1)} \\ k_d + \varepsilon_i^{(2)} \\ k_i + \varepsilon_i^{(3)} \end{bmatrix} \tag{15}$$

with the constraints of $|u_t| \leq u_t^{\max}, |\dot{u_t}| \leq \dot{u_t}^{\max}$. In the initial state $x_0 = \left[\alpha^{(0)}\ \theta^{(0)}\ q^{(0)}\right]^T$, N paths are generated based on the Gaussian noise with variance as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \tag{16}$$

each path is an exploration-and-exploitation possibility, wherein the path cost is evaluated by **formula** (14). There are M iterations in total, each of which includes N paths; consequently, $M \times N$ paths are generated during the whole RL process. The parameter vector $\mathcal{K}$ is updated every N paths as follows:

$$\begin{aligned} \mathcal{K}^{(\text{new})} &= \mathcal{K}^{(\text{old})} + \varepsilon \\ \varepsilon &= \sum_{i=1}^{N} P(\tau_i)\, \varepsilon_i \end{aligned} \tag{17}$$

where ε is the probability-weighted averaging of N Gaussian noise vectors, which are used to generate N different paths and $P(\tau_i)$ is calculated by **formula** (10).

3.3 FR-PI2 Algorithm

With the implementation of PI2, the stochastic optimal control problem defined as (12) can be solved. However, traditional PI2 algorithm fails to guarantee fast convergence and real-time performance with a complicated loss function. To address this problem, two kernel techniques are proposed in this part. Firstly, we apply sample-filtering and variance-enhancing methods to ensure fast convergence and stability. Furthermore, a real-time optimization technology called rolling optimization is proposed to improve the real-time performance of PI2. Consequently, the procedure of FR-PI2 is presented to solve the proposed stochastic optimal control problem and improve the defects of PI2.

Sample Filtration and Variance Enhance The results of control experiment, which will be presented in section 4, demonstrates the fact that massive worse experience will be learned if we update parameters $\mathcal{K}$ directly using **formula** (17), leading to more instability and decrease of convergence rate during training. Thus, we combine a novel sample filtration method with PI2 as **Algorithm** 1 to guarantee the stability while the variance is enhanced as **formula** (18) every M_β iterations to improve the convergence rate. Specifically, the learning rate β is set greater than one, which is different from traditional PI2 proposed in [8] where β is set less than one to attenuate the variance. In **Algorithm** 1, $R(\tau_i)$ stands for loss function of the path τ_i defined as **formula** (14), and R_i is current optimal loss function.

Algorithm 1 Sample Filtration Algorithm

```
if R(τ_i)>R_i then
    reject the path τ_i, and set ε_i as zero
else
    accept the path τ_i
end if
```

$$\Sigma = \beta^2 \Sigma = \begin{bmatrix} (\beta\sigma_1)^2 & 0 & 0 \\ 0 & (\beta\sigma_2)^2 & 0 \\ 0 & 0 & (\beta\sigma_3)^2 \end{bmatrix} \tag{18}$$

Rolling Optimization Rolling optimization transforms single-stage optimization into multi-stage optimization to reduce whole computational complexity, thereby enhance the optimality and stability. After each optimization, the optimized parameter vectors are directly used as initial settings for next optimization to reduce training iterations. Combining with sample filtration algorithm and variance enhance, the FR-PI2 algorithm is illustrated in **Algorithm** 2

Algorithm 2 FR-PI2 Algorithm

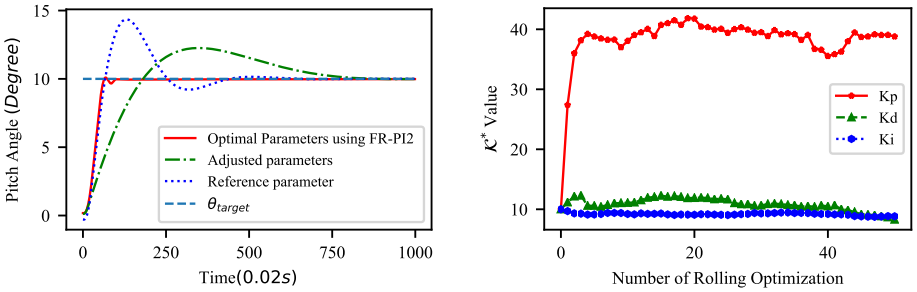
initialize the optimal parameter vector $\mathcal{K}_t^*$, and Gaussian variance Σ
initialize the system state x_t
for $t \in [0, t_f]$ **do**
  if $t\%M_O == 0$ **then**
    set PI2 initial parameter $\mathcal{K}$ as $\mathcal{K}_t^*$, and initialize proper parameters for PI2
    for each $i \in [1, M]$ **do**
      for each $j \in [1, N]$ **do**
        generate a path τ_i using Gaussian noise ε_i with variance Σ
        obtain the path cost $R(\tau_i)$using **formula** (14)
        filter the sample using **Algorithm** 1
      end for
      update parameter vector $\mathcal{K}$ using **formula** (10) and(17)
      record current optimal path cost with R_i
      if $i\%M_\beta == 0$ **then**
        enhance Gaussian variance Σ using **formula** (18)
      end if
    end for
    set $\mathcal{K}_t^*$ as $\mathcal{K}^*$
  end if
  update x_t using **formula** (13) and (3) with $\mathcal{K}_t^*$
  $\mathcal{K}_t^*$ keeps invariant till next optimization
end for

4 Experiments and Results

To validate the efficiency of the FR-PI2 algorithm, two groups of experiments are displayed in this section. First, we present a real-time control simulation experiment of Hypersonic Aircraft pitch attitude to demonstrate that FR-PI2 algorithm can solve multi-objective stochastic optimal control problem defined as (12) with good performance. Second, a group of algorithm performance control experiments are proposed to illustrate a better real-time performance, rapider convergence rate and good robustness of proposed FR-PI2 algorithm compared with Sample-Filter PI2, say F-PI2, and PI2.

4.1 Simulation Experiment

An online simulation comparison among optimized parameters, reference parameters, and constant adjusted parameters is implemented with same initial states. The results are shown in **Fig.** 1(a), where the initial reference parameters are $k_p = 10.0$, $k_d = 10.0$, and $k_i = 10.0$, the artificially adjusted parameters are $k_p = 1.5$, $k_d = 2.5$, $k_i = 0.5$ and the optimal parameters are tuned using FR-PI2 algorithm with the real-time system states shown in **Fig.** 1(b). As can be seen from the results in **Fig.** 1(a), compared with the method using the reference parameters, the stability time of the other two methods is less. Moreover, the overshoot using the constant adjusted parameters is greater than that using the optimized parameters.

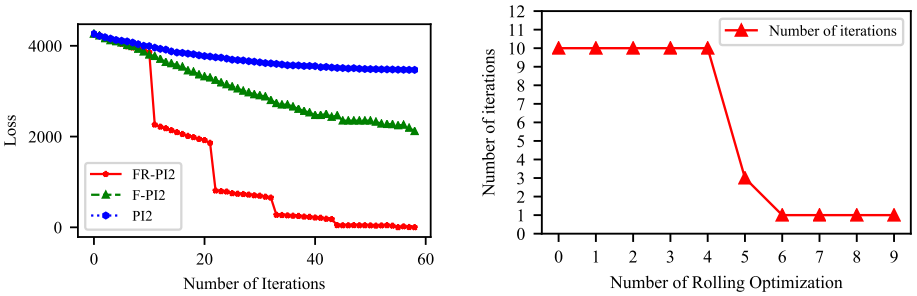


(a) pitch Control Curve (b) $\mathcal{K}^*$ Iteration Graph

Fig. 1. Simulation comparison among optimized parameters, reference parameters, and constant adjusted parameters.

4.2 Performance Test and Comparision

Two control experiments are displayed, in this part, to verify the rapider convergence rate, high real-time performance and robustness of FR-PI2 in comparison with PI2 and F-PI2 (without rolling optimization). For all following experiments, test environment and irrelevant variables such as training times, system initial state, variance update interval, variance update coefficient etc., are set the same value to reveal algorithm performance.



(a) Loss Function Curve (b) Iterations Every Rolling Optimization

Fig. 2. Convergence and real-time performance control experiment among FR-PI2, F-PI2, and PI2.

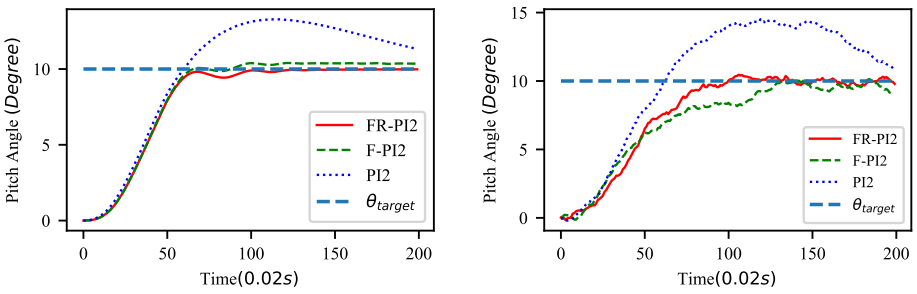
Convergence Rate and Real-time Performance Control Experiment Convergence rate and real-time performance among proposed three algorithms

will be verified in this experiment to demonstrate the ascendant efficiency of FR-PI2. The results are shown in **Fig.** 2, where $M = 57$, $M_O = 10$, $t_f = 200$, $N = 20$, $M_\beta = 5$, $x_0 = [0\ 0\ 0]^T$, $\beta = 1.17$ (for PI, the learning rate can set as $0.85 \approx \frac{1}{1.17}$). The whole convergence process is demonstrated in **Fig.** 2(a). Obviously, it is concluded that the convergence rate of FR-PI2 is predominant among proposed three algorithms, with which the loss functions converge to zero within 42 iterations. The number of rolling optimizations per optimization is shown in **Fig.** 2(b) to illustrate the descending calculation of FR-PI2 and the superior real-time performance compared with PI2 and F-PI2, where M (in this experiment is set 57 as the total training time of FR-PI2) iterations are required for every optimization.

Robustness Performance Control Experiment Robustness performance among proposed three algorithms will be tested in this experiment to demonstrate the superior noise immunity of FR-PI2. The results are shown in **Fig.** 3, where $M = 100$, $M_\beta = 5$, $x_0 = [0\ 0\ 0]^T$, $\beta = 1.17$(for PI, the learning rate can set as 0.85), $M_O = 10$, $t_f = 200$, $N = 20$.The pitch control curve of a noise-free Hypersonic Aircraft system is shown in **Fig.** 3(a), while **Fig.** 3(b) illustrate the control curve of a high-noise Hypersonic Aircraft system, whose stochastic control system equation is similar to (4) defined as follow

$$\dot{\boldsymbol{x}}_{\boldsymbol{t}} = \boldsymbol{f}_{\boldsymbol{t}} + \boldsymbol{G}_{\boldsymbol{t}}\left(\boldsymbol{u}_{\boldsymbol{t}}\right) + \varepsilon^* \tag{19}$$

wherein ε^* is the Gaussian noise submitting to N(0, 5). By comparison of **Fig.** 3(a) and **Fig.** 3(b), we can find that FR-PI2 and F-PI2 has a superior anti-noise ability and robustness compared with PI2, which both make the high-noise system reach the steady-state within 120 steps(2.4 seconds) without overshoot.



(a) θ Control Curve of noise-free HA system(b) θ Control Curve of high-noise HA system

Fig. 3. Anti-noise performance control experiment among FR-PI2, F-PI2, and PI2.

5 Conclusion

In this paper, a novel path-integral-based RL algorithm for the pitch attitude stochastic optimal control of Hypersonic Aircraft is proposed. Specifically, to improve the learning efficiency and real-time performance, the FR-PI2 algorithm, rather than the traditional PI2, is adopted to obtain the numerical solution of the constrained stochastic optimal control problem. Moreover, a model-free parameterization is introduced into RL to guarantee the stability and robustness of the system with respect to various uncertainties and high noises. Finally, the parameters for different stages are tuned online with real-time system states, which ensures the real-time performance and the optimality of the whole path. To further verify the performance of the proposed method, another two groups of control experiments are implemented. The high learning efficiency, real-time performance and optimality of the proposed algorithm are verified via simulation and control experiments.

In the future, we will apply this method to more control domains, such as robot control, thereby further demonstrating the generalizability of the proposed method.

Acknowledgment

This work was supported by Science Foundation of Science and Technology on Complex System Control and Intelligent Agent Cooperative Laboratory (192003).

References

1. Shahriar Keshmiri, Richard Colgren, and Maj Mirmirani. Six dof nonlinear equations of motion for a generic hypersonic vehicle. 2007.
2. Bahare Kiumarsi, Kyriakos G Vamvoudakis, Hamidreza Modares, and Frank L Lewis. Optimal and autonomous control using reinforcement learning: A survey. *IEEE transactions on neural networks and learning systems*, 29(6):2042–2062, 2017.
3. JJ Recasens, QP Chu, and JA Mulder. Robust model predictive control of a feedback linearized system for a lifting-body re-entry vehicle. page 6147, 2005.
4. Arthur Richards and Jonathan How. Implementation of robust decentralized model predictive control. page 6366, 2005.
5. Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, 11:3137–3181, 2010.
6. Sun Mingwei Ma Shunjian Park Minnan Wang Yongkun Li Yi. *Hypersonic aircraft auto-disturbance control method*, pages 118–128. Science Press, 2017.
7. Wei Zhu, Xian Guo, Yongchun Fang, and Zhang. Path-integrated reinforcement learning snake robot targeted motion. *Pattern recognition and artificial intelligence*, pages 1–9, 2019.
8. Wei Zhu, Xian Guo, Yongchun Fang, and Xueyou Zhang. A path-integral-based reinforcement learning algorithm for path following of an autoassembly mobile robot. *IEEE Transactions on Neural Networks and Learning Systems*, PP:1–13, 12 2019.