面向固定翼飞机的智能容错姿态控制

Towards Intelligent Fault-Tolerant Attitude Control of Fixed-Wing Aircraft

Authors: Alex Zongo, Li Qing.

International Conference on Guidance Navigation and Control, Changsha, China

Tsinghua University

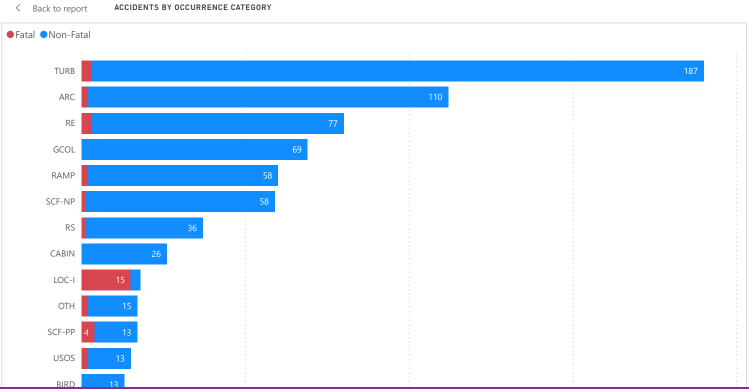
August 2024

Contents

Overview

Overview

Figure 2: 2023 ICAO Safety Report [国际民航组织安全报告]
https://www.icao.int/safety/iStars/Pages/Accident-Statistics.aspx



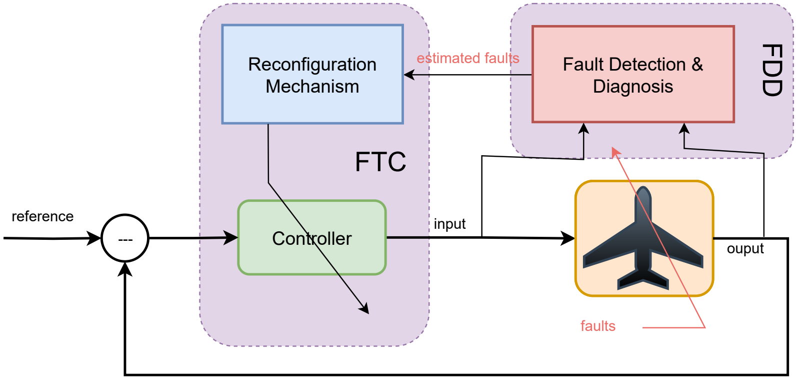
Overview

Loss Of Control (LOC) can be caused by:

- Saturated or Partial loss of the Actuators (Elevator, ailerons, rudder) 饱和或部分损失的作动器（升降舵、副翼、方向舵)
- Shifted Center of Gravity (重心偏移)
- Iced Wings 机翼结冰
- Or heavy wind or gust, noisy sensors 或强风、阵风、传感器噪声
- All of these faults combined or any other not listed here or NEVER SEEN before or anticipated (所有这些故障结合在一起，或其他未列出或从未见过或预料到的故障)

Conventional FTC

Conventional Active Fault-Tolerant Control FTC (传统容错控制)

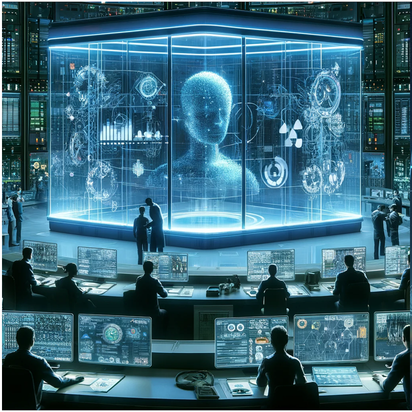


Shortcomings

Shortcomings of Current FTC Strategies

- Limited information Available (Faults Detection Challenges) 可用信息有限（故障检测挑战）
- High Computational Complexity 高计算复杂度
- Difficulty in the design to withstand multiple components failure 难以设计以承受多个组件故障
- Performance Limitation due to the use of adaptive strategies in certain situation 由于在某些情况下使用自适应策略而导致的性能限制

AI in Flight Control



AI in Fault Tolerant Flight Control (FTC)

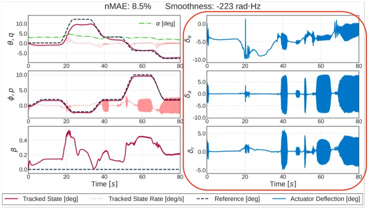
The following represent recent studies in Fully Neural Network-based FTC.

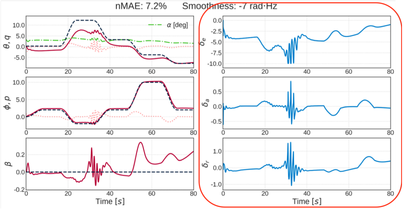
Literature

- Soft Actor Critic (SAC) RL method: (Dally Killian and Van Kampen Erik-Jan, «Soft Actor-Critic Deep Reinforcement Learning for Fault Tolerant Flight Control», presented at the AIAA SCITECH 2022 Forum). [1]
- Evolutionary Reinforcement Learning: (Vlad Gavra, «Evolutionary Reinforcement Learning: A hybrid Approach for Safety-Informed Intelligent Fault-tolerant Flight Control» [2]

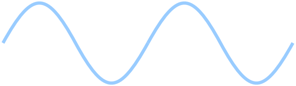
Challenge in RL Application

Emerging or Recurrent Challenge in RL application: action smoothness [2]





rough actuating signal



Smooth actuating signal

Control Problem Formulation

This study focuses on optimizing an aircraft's attitude control, aiming to minimize tracking errors and ensure action smoothness.

Problem Formulation 控制问题的公式化

Given a state vector $s(t)$, a control input $u(t)$, and a reference input vector $r(t)$, the optimization problem is defined by:

$$u^* = arg\min_u \int_{t_0}^{t_f} L_s(s,u,r) + L_u(u)dt \text{ s.t. } |u| \leq u_{max} \text{ and } |\Delta u| \leq \frac{u_{max}}{\Delta t}$$

Where L_s minimizes the deviation from the reference trajectory and L_u optimizes for smooth changes between consecutive control inputs. (L_s 最小化与参考轨迹的偏差，L_u 优化连续控制输入之间的平滑变化。)

Assumptions

States Definition:

- High-fidelity 6-DOF non-linear dynamics Fixed-Wing Aircraft Model developed by TU-Delft [3, 4]
- p = roll rate 滚转角速率
- q = pitch rate 俯仰角速率
- r = yaw rate 偏航角速率
- $\theta, \Delta\theta$ = pitch angle 俯仰角, pitch angle error 俯仰角误差
- $\phi, \Delta\phi$ = bank angle and roll angle error (滚转角和滚转角误差)
- β = yaw (side-slip) angle 偏航（侧滑）角
- $\delta_e, \delta_a, \delta_r$ = deflection angle of the elevator, ailerons and rudder 升降舵、副翼和方向舵的偏转角

Assumptions

Assumptions:

- Unknown Fault and model dynamics: given by the simulation. 未知故障和模型动力学（由仿真提供）
- Training on Nominal Flight 在正常飞行中训练
- Attitude Control 姿态控制
- Thrust is handled by an inner PID auto-throttle
- State-Action Configuration (状态-动作配置):

$$
\begin{aligned}
s &:= [\Delta\theta, \Delta\phi, 0-\beta, p, q, r] && (1)\\
a_t &:= [\delta_e, \delta_a, \delta_r]^T \in [-1, 1]^3 && (2)\\
u &:= u_{min} + (a_t + 1)\frac{u_{max} - u_{min}}{2} && (3)
\end{aligned}
$$

Reward Function Definition

Reward Function Design

Given the angular $rate := [p, q, r]^T$, the attitude reference signal $s_r = [\theta_r, \phi_r, 0]^T$, the controlled angles $s_{ctrl} = [\theta, \phi, \beta]^T$ and the control actions $u = [\delta_e, \delta_a, \delta_r]^T$;

$$r_1 = -\frac{1}{3}||rate||_1 \text{ (最小化角速度)} \quad (4)$$

$$r_2 = -\frac{1}{3}||clip(c_r \odot [s_r - s_{ctrl}], -1, 1)||_1, \; c_r = \frac{6}{\pi}[1, 1, 4] \text{ (最小化跟踪误差)} \quad (5)$$

$$r_3 = -\frac{C_p}{\Delta t}(T_{max} - T), \; C_p = 2 \text{ (奖励更长时间的飞行)} \quad (6)$$

$$s_m := -\frac{C_{s_m}}{T}\sqrt{\frac{2}{n}\sum_{u=1}^{|U|}\sum_{k=1}^{n/2}(S_{uu}[k] \cdot f_k)_u}, \; C_{s_m} = 1 \text{ (测量连续控制动作的平滑度)} \quad (7)$$

Reward Function Definition

Reward Function Design

- Total Reward

$$r \quad := \quad W \odot [r_1, r_2, r_3] + s_m, \;\; W = [w_1, w_2, w_3] \tag{8}$$

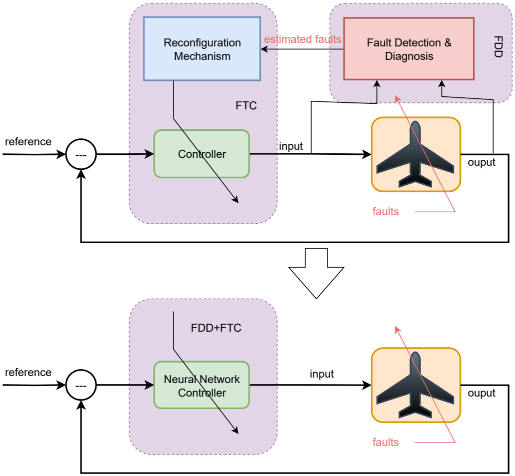
where W = [0.2, 0.6, 0.2]

- Avoiding UNSAFE MANEUVERS, an episode ends when:

$$\theta > \theta_{max}, \;\; \phi > \phi_{max} \; or \; altitude < 50 \; or \; T = T_{max} \tag{9}$$

Overall Control Approach

Blending FDD, FTC into a single Deep Neural Network Model.



General Control Framework

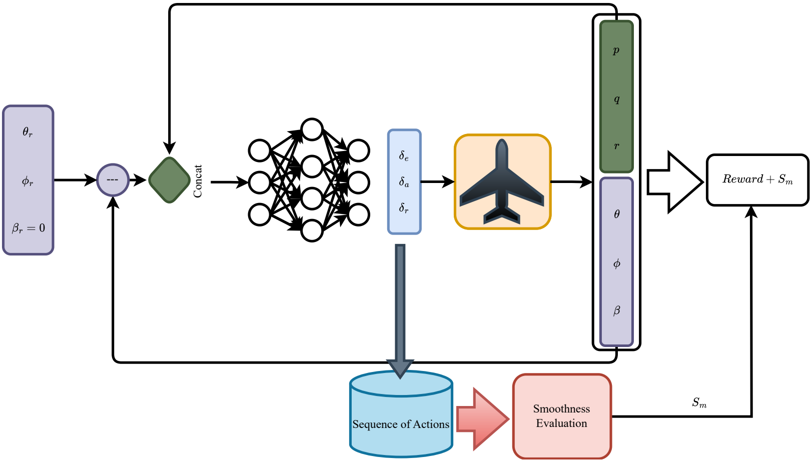


Figure 4: Agent-Environment Control Framework Architecture

Summary of state-actions-reward configuration

Table 1: Summary of state-actions-reward function

state s	action a_t	reward function
$\Delta\theta = \theta_r - \theta$: pitch error	δ_e: elevator deflection	$r_1 = -\frac{1}{3}\|\|rate\|\|_1$
$\Delta\phi = \phi_r - \phi$: roll error	δ_a: ailerons deflection	$r_2 = -\frac{1}{3}\|\|clip(c_r \odot [s_r - s_{ctrl}], -1, 1)\|\|_1$
$0 - \beta$: side-slip error	δ_r: rudder deflection	$r_3 = -\frac{C_p}{\Delta t}(T_{max} - T)$
p: rolling rate	-	$s_m := -\frac{C_{s_m}}{T}\sqrt{\frac{2}{n}\sum_{u=1}^{\|U\|}\sum_{k=1}^{n/2}(S_{uu}[k] \cdot f_k)_u}$
q: pitch rate	-	$r := W \odot [r_1, r_2, r_3] + s_m$
r: yaw rate	-	-
$s_r = [\theta_r, \phi_r, \beta_r = 0]^T$	-	-
$s_{ctrl} = [\theta, \phi, \beta]^T$	-	-
$rate = [p, q, r]^T$	-	-

Proximal Policy Optimization

Definition

- PPO: Proximal Policy Optimization [5]
- known for its learning stability and successful application in robotics

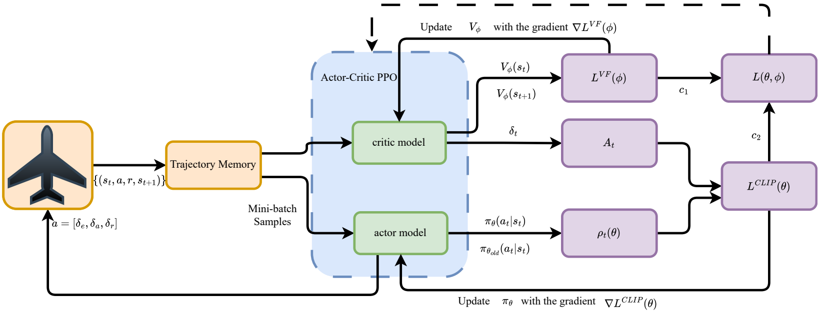


Figure 5: PPO overall learning framework

Cross Entropy Method-Reinforcement Learning

Definition

- CEM-TD3: Cross-Entropy Methods Twin Delayed Deep Deterministic Policy Gradient. [6]
- It is an Evolutionary Distribution Algorithm (EDA)
- that benefits from TD3 gradient based improvements and the search efficiency of CEM (它受益于基于TD3梯度的改进和CEM的搜索效率)

Cross Entropy Method-Reinforcement Learning

Changes in CEM update Strategy

The following changes in the update strategy improve learning stability, exploration, and convergence of the learning process. 以下更新策略的变化在训练期间提高了学习稳定性、探索性和收敛性。

$$\mu \leftarrow \mu + \alpha\left(\sum_{i=1}^{\lambda} w_i z_i - \mu\right) \tag{10}$$

$$\Sigma \leftarrow \sigma \cdot \frac{\sum_{i=1}^{\lambda} w_i (z_i - \mu_{old})^2}{||\Sigma||_2} \tag{11}$$

$$\sigma \leftarrow \sigma * \sigma_{decay} \ \textit{while}\ \sigma > \sigma_{lim} \tag{12}$$

Cross Entropy Method-Reinforcement Learning

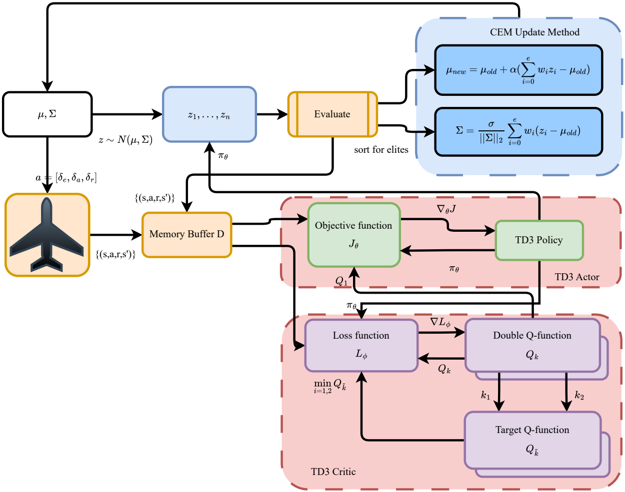


Figure 6: General overview of CEM-TD3 Algorithm

Key-Elements of The Online Adaptation Framework

An Online Adaptation Framework proposed here, consist of the following key-elements:

Key-Elements

- Database of Trained Policies (训练策略数据库): Collection of trained control agents
- Model Identification Unit (模型识别单元): A Gaussian Process model
- Soft Switching Mechanism (软切换机制)

Online Adaptation Framework Overall Structure

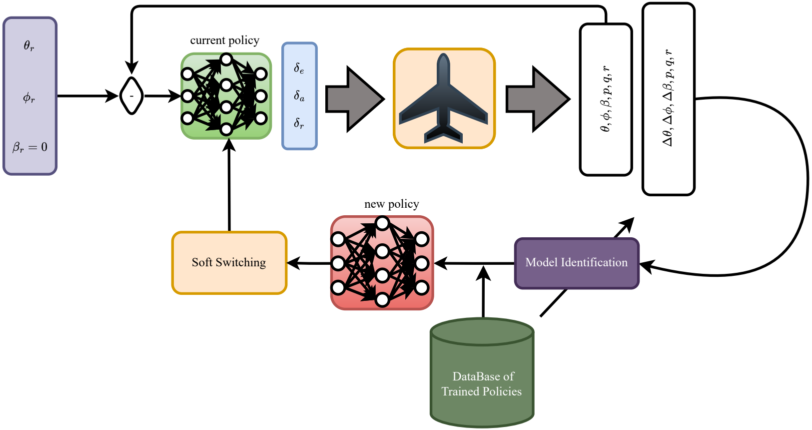


Figure 7: High-Level Overview of the Adaptation Framework

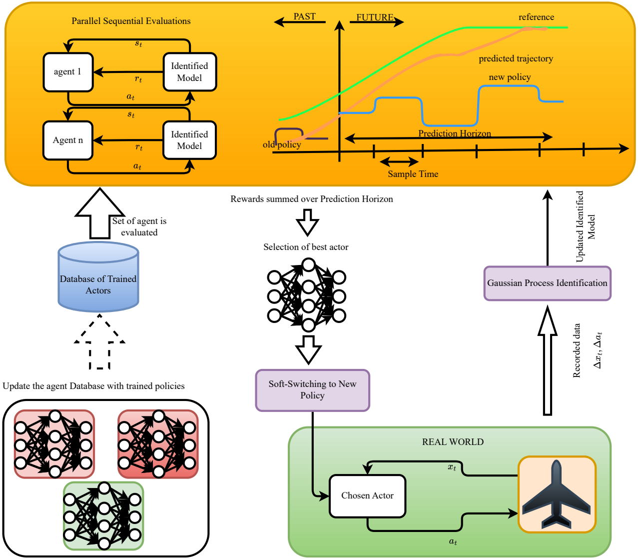


Figure 8: Extended Online Adaptation Framework. adapted from [2]

Training Configuration and Evaluation Scenarios

Training

- algorithms, frameworks and simulation are implemented in Python programming language 算法、框架和仿真都是用Python 编程语言实现的。
- Each algorithm is trained on a nominal flight without any faults.
- The complexity and infinite possibility of faults that can occur renders practically impossible to model and train for each individual situation, in addition to the limited computational resources. 每个算法都在没有故障的正常飞行中进行训练。故障的复杂性和无限可能性使得几乎不可能为每个单独的情况建模和训练，此外还有有限的计算资源.
- Computational Device: 12 Intel(R) i7-5930K 3.5GHz CPU cores with NVIDIA GeForce GTX TITAN X graphics computer.

Training Configuration and Evaluation Scenarios

Training

- The algorithms are trained offline for 2000 timesteps per episode (20 s) with cosine smoothed reference signals uniformly sampled ($\theta \in [-25°, 25°]$, $\phi \in [-45°45°]$, $\beta_r = 0$) [1] 算法在每个回合2000个时间步（20秒）离线训练，使用余弦平滑的参考信号均匀采样($\theta \in [-25°, 25°]$, $\phi \in [-45°45°]$, $\beta_r = 0$) [1]
- The generalizability and adaptability of Deep Neural Network is expected to perform well even under unseen and unexpected conditions and states. 预计深度神经网络的泛化性和适应性在未见过和意外的情况下和状态下也能表现良好。

Training Configuration and Evaluation Scenarios

Table 2: Evaluation Cases and Trim Conditions borrowed from [2] and Used in Simulation

Name	Description
Nominal	Trimmed at H=2,000m and $V_{tas} = 90m/s$
Iced Wings	$0.7\alpha_{max}$ and the $C_D + 0.6$
Aft Shifted CG	The center of gravity is shifted aft by 0.25 m
Saturated Aileron	Deflection clipped at $\pm 1^\circ$
Saturated Elevator	Deflection clipped at $\pm 2.5^\circ$
Partial Loss of Elevator	A 70% reduction in effectiveness coefficient
Jammed Rudder	Rudder stuck at 15°
High Dynamic Pressure	Trimmed at H=2,000m and $V_{tas} = 150m/s$
Low Dynamic Pressure	Trimmed at H=10,000m, $V_{tas} = 90m/s$
Wind and Sensor Noise	Identified from flight tests [2] and isolated from [7]

Stability Analysis Approach

Stability of the trained controllers is performed via a linear model that is trained on experiences collected from a simulated agent-environment (for various flight conditions) to approximated the A and B matrices in Eq. 13 via mean squared error loss function. The system takes as input a reference signal and outputs the state (pitch, roll and side-slip angle) of the aircraft.

$$g(X) = X' = A \cdot X + B \cdot u \tag{13}$$

where $u = [\theta_r, \phi_r, 0]^T$ while $X = [\theta, \phi, \beta]$. Hence A and B are 3-by-3 matrices.

Stability Analysis Approach

Definition

Given experience data from the simulation $ref = [\theta_r, \phi_r, 0]$ and $y = [\theta_t, \phi_t, \beta_t]$, $g(ref) = \hat{y} = [\theta', \phi', \beta']$, a gradient descent algorithm is used to update A and B. $\eta = 0.001$ is chosen as learning rate.

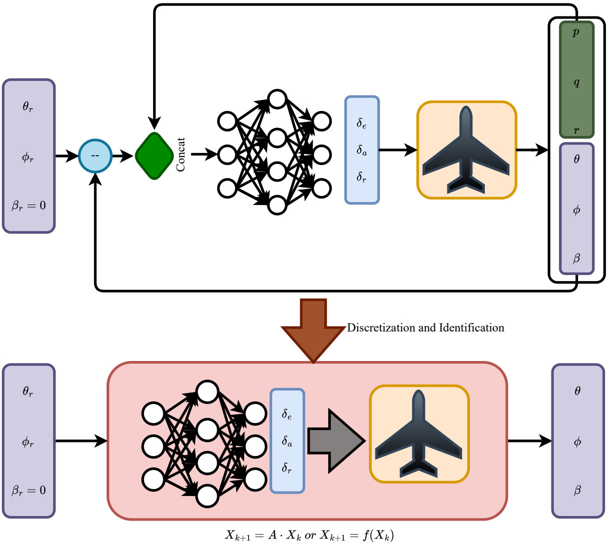
$$L(X, u) = ||y - \hat{y}||_2^2 \tag{14}$$

$$A \leftarrow A - \eta \nabla_X L \tag{15}$$

$$B \leftarrow B - \eta \nabla_u L \tag{16}$$

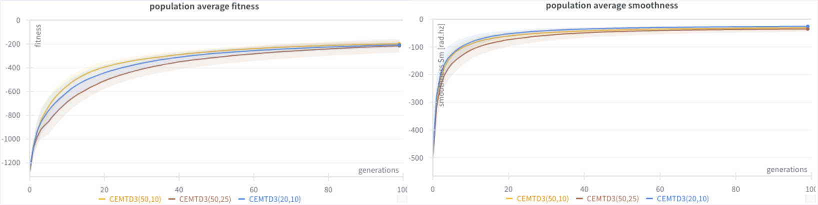
The eigenvalues of A are then used to indicate the stability of the overall system.

Stability Analysis Framework



Training Results

Training Results - CEM-TD3

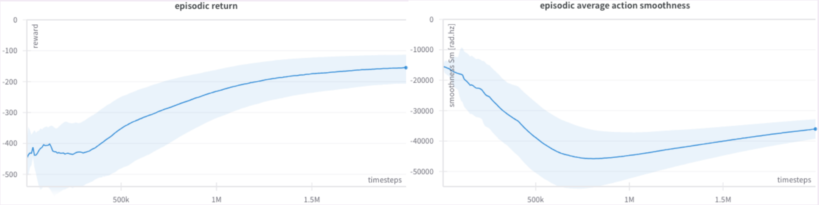


(a) (b)

Figure 10: CEM-TD3 Training over 3 different seeds. (a) Average Population Fitness (b) Average Population Action Smoothness

Training Results and Application on Fault-scenarios

Training Results - PPO



(a) (b)

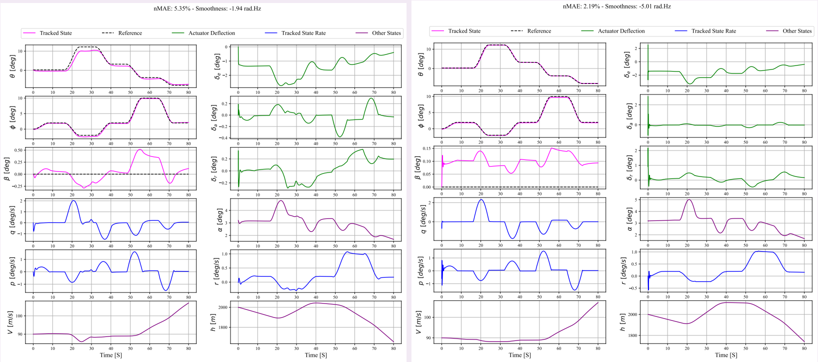
Figure 11: PPO Training over 3 different seeds. (a) Episodic Rewards (b) Episodic Action Smoothness

Performance on Nominal and Fault-scenarios

The following presents flight performance of the trained agent in simulation, subjected to a reference trajectory.

Case 1 - Normal Flight

Nominal Flight

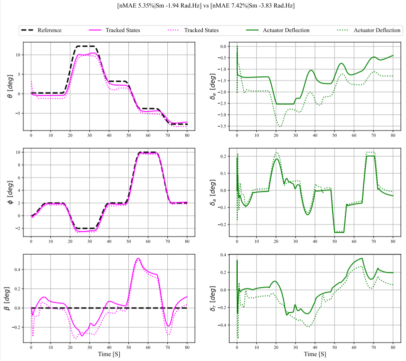


(a) (b)

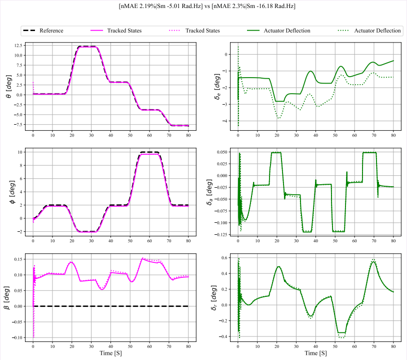
Figure 12: Performance on Nominal Flight (a) CEM-TD3 (b) PPO agents

Case 2 - Iced Wings

Iced Wings: $0.7C_{L_{max}}$ and $C_D + 0.06$



(a)

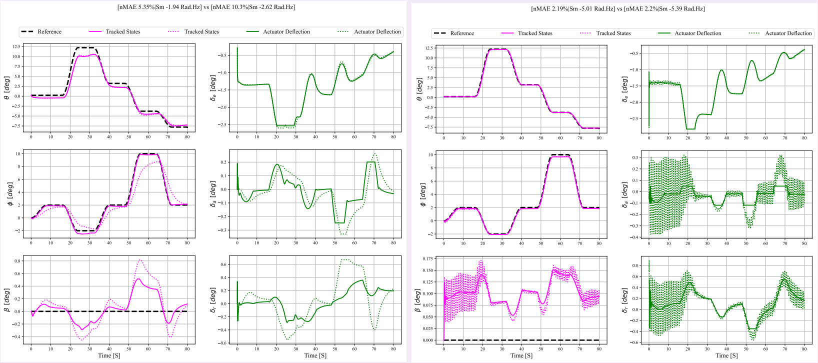


(b)

Figure 13: Iced Wings Flight [·] (a) CEM-TD3 (b) PPO agents

Case 3 - Saturated Ailerons

Saturated Ailerons: $[-1^\circ, 1^\circ]$

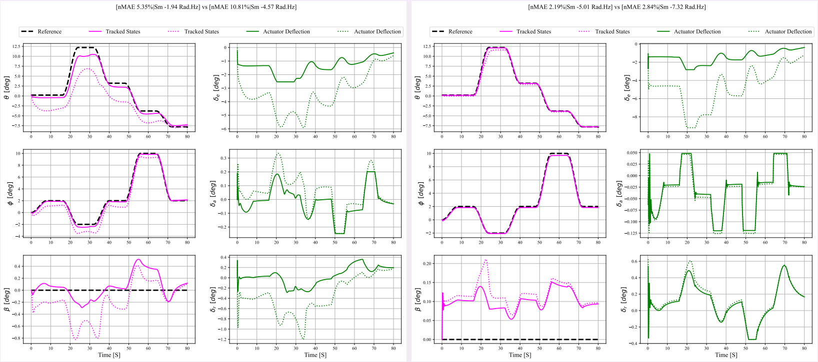


(a) (b)

Figure 14: SA [·] (a) CEM-TD3 (b) PPO agents

Case 4 - Partial Loss of Elevator

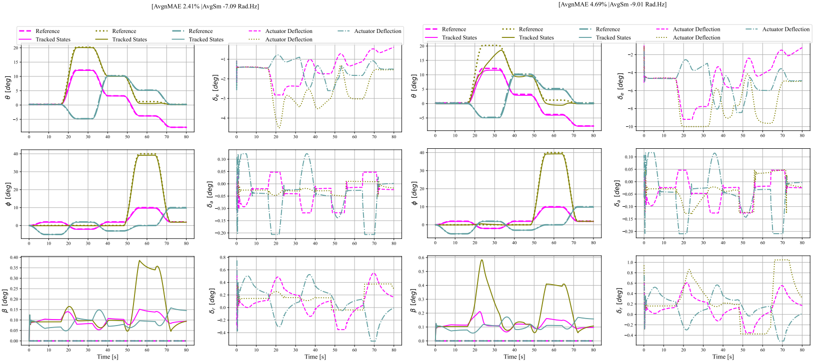
Partial Loss of Elevator: $0.3C_{\frac{L\cdot\delta_e}{m\cdot D}}$ *and* $0.3C_{m_q}$



(a) (b)

Figure 15: LE [·] (a) CEM-TD3 (b) PPO agents

How do the trained agents perform on different sets of trajectory?

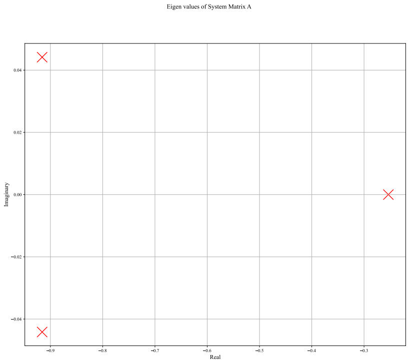


(a) (b)

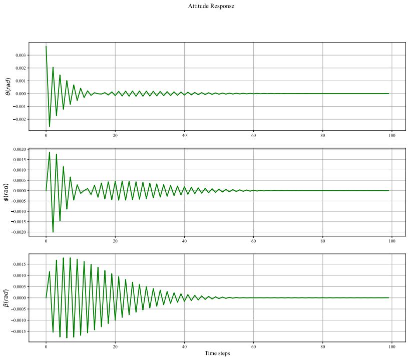
Figure 16: (a) Normal Flight (b) Partial Loss of Elevator

Example of Eigen values and Attitude time-series response on Partial Loss of Elevator Model with PPO controller.

With or without fault, the sharp oscillations are mitigated with very low amplitude.



(a)



(b)

Overall Tracking Error Comparison with Literature Benchmark

PPO is better in terms of tracking error.

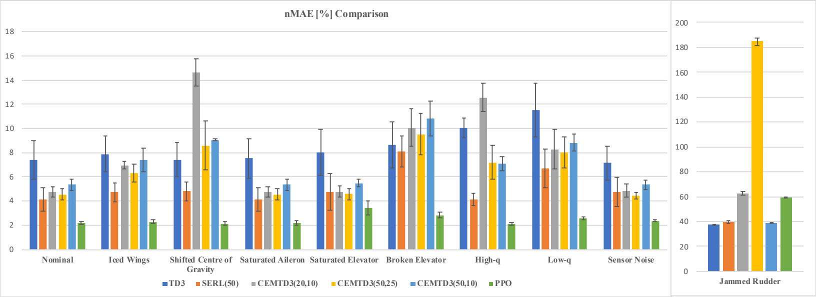


Figure 18: **Tracking Error** nMAE[%] of the designed controllers CEM-TD3 and PPO compared to the state-of-art (SOTA) SERL and TD3

Overall Action Smoothness Comparison with the Literature Benchmark

CEM-TD3 shows better action policy smoothness. It balances tracking error and actuator control smoothness.

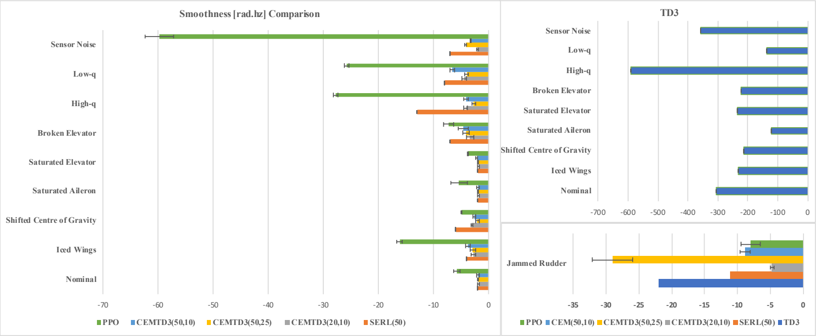


Figure 19: **Action Smoothness** [rad.Hz] of the designed controllers CEM-TD3 and PPO compared to the state-of-art SERL and TD3

Research Relevance and Contribution

- A modular Online Adaptation Framework based on the literature and using advances AI techniques is proposed and performs well. 提出了一种模块化在线适应框架，且表现良好。
- CEM-RL is improved, leading to faster convergence and a training-efficient CEM-TD3 policy gradient algorithm. CEM-TD3 balances tracking error and action smoothness. CEM-RL 得到改进，导致更快的收敛速度和高效训练的CEM-TD3 策略梯度算法。CEM-TD3 平衡了跟踪误差和动作平滑性。
- PPO is, for the first time, applied to fault-tolerant control and shows state-of-the-art adaptability and robustness to faults. PPO 首次应用于容错控制，并显示出最先进的适应性和故障的鲁棒性.
- The overall code repo: `https://github.com/Alex-Zongo/rl_for_ftc.git`

Future work

- Improve the computational complexity of Evolutionary Reinforcement Algorithms to speed up the learning process. 改进进化强化算法的计算复杂度以加速学习过程
- Scale up the adaptability of the controllers with more complex fault-induced scenarios. 通过更复杂的故障诱导情景扩大控制器的适应性
- Dive into the field of Explainable Reinforcement Learning for more insights into underlining mechanism of NN-based controllers, which should contributes to their certification and trust. 深入研究可解释的强化学习领域，以获得更多关于基于神经网络控制器的基本机制的见解，这将有助于它们的认证和信任。
- Move from simulation to real-world flight tests.

Bibliography I

[1] Dally K, Kampen E-J V. Soft Actor-Critic Deep Reinforcement Learning for Fault Tolerant Flight Control[C/OL] // AIAA SCITECH 2022 Forum, [S.l.]: American Institute of Aeronautics and Astronautics, 2022. https://doi.org/10.2514%2F6.2022-2078.

[2] Gavra V. Evolutionary Reinforcement Learning: A hybrid Approach for Safety-Informed Intelligent Fault-tolerant Flight Control[D/OL]. [S.l.]: TU Delft, 2022. http://repository.tudelft.nl/.

[3] Linden V d. DASMAT-Delft University Aircraft Simulation Model and Analysis tool[C/OL] // , [S.l.], . http://resolver.tudelft.nl/uuid:25767235-c751-437e-8f57-0433be609cc1.

Bibliography II

[4] van den Hoek M. A, de Visser C. C, D. M P. Identification of a Cessna Citation II Model Based on Flight Test Data[M/OL] // Advances in Aerospace Guidance, Navigation and Control. 2018 : 259-277. http://dx.doi.org/10.1007/978-3-319-65283-2_14.

[5] Filip J S, Prafulla W, Alec R D, et al. Proximal Policy Optimization Algorithms[J/OL]. https://arxiv.org/abs/1707.06347, 2017. http://dx.doi.org/https://doi.org/10.48550/arXiv.1707.06347.

[6] Pourchot A, Sigaud O. CEM-RL: Combining evolutionary and gradient-based methods for policy search[J/OL], 2019. http://dx.doi.org/https://doi.org/10.48550/arXiv.1810.01222.

Bibliography III

[7] Moorhouse D, Woodcock R. Background information and user guide for MIL-F-8785C, military specification-flying qualities of piloted airplanes[R]. [S.l.] : Air Force Wright Aeronautical Labs Wright-Patterson AFB OH, 1982.

Thanks for your attention!

Q & A