Optimal Stall Recovery via Deep Reinforcement Learning for a General Aviation Aircraft

Agustin Grillo *, Gabriel Torre † and Roberto A. Bunge‡
Universidad de San Andrés, San Fernando, Buenos Aires, Argentina

Stall is a hazardous region of the flight envelope for fixed-wing aircraft, in which a high angle of attack causes flow separation from the main wing. Indeed, inadvertent stalls are a major co-factor of fatal accidents in general aviation aircraft, and applying correct stall recovery actions is fundamental to minimizing altitude loss and thus the risk of ground collision. This paper studies the optimal stall recovery problem, whereby a control policy for control surface deflections and throttle is computed with the goal of minimizing total altitude loss for a general aviation aircraft. The optimal control problem is formulated and cast as an infinite horizon Dynamic Programming (DP) problem. To circumvent the curse of dimensionality, characteristic of exact DP algorithms, the Proximal Policy Optimization (PPO) algorithm is considered, a sampling based Deep Reinforcement Learning algorithm that can handle large continuous state-action spaces. The approach is first validated by comparing the results from PPO with those from Value-Iteration, an exact DP algorithm, on a reduced order problem, showing that indeed PPO converges to the optimal policy. PPO is then applied to the stall recovery problem computing the optimal policy for control surface deflection and throttle inputs. The results presented in this study confirm empirically developed guidelines for stall recovery of general aviation aircraft, and also shows the criticality of correct elevator and throttle management in the minimization of altitude loss, with as much as a 30% increase in altitude loss if application of full throttle is delayed by as little as 0.75 seconds. In addition to the quantitative results and analysis presented for the stall recovery problem, this work suggests that PPO could be applied to other challenging loss of control and upset recovery problems involving full aircraft degrees of freedom, opening up new lines of research in fixed wing loss of control.

Nomenclature

ρ	=	air density
b	=	wing span
c	=	chord length
S	=	wing surface area
p_x, p_y, p_z	=	northward, eastward and down position
h	=	altitude, from the ground
V	=	airspeed
V_s	=	stall airspeed
$\hat{V}$	=	non-dimensional airspeed, $\hat{V} = \frac{V}{V_s}$
α	=	angle of attack
β	=	sideslip angle
ϕ, θ, ψ	=	roll, pitch and yaw angles
γ	=	flight path angle
ξ	=	course over ground angle
μ	=	bank angle
p, q, r	=	roll, pitch and yaw rate
$\hat{p}, \hat{q}, \hat{r}$	=	dimensionless roll, pitch and yaw rate, $\hat{p} = \frac{pb}{2V}$, $\hat{q} = \frac{qc}{2V}$, $\hat{r} = \frac{rb}{2V}$

*Graduate Research Assistant, Department of Engineering
†Graduate Research Assistant, Department of Engineering
‡Assistant Professor, Department of Engineering, AIAA Senior Member

δe	=	elevator deflection, positive trailing edge down
δr	=	rudder deflection, positive trailing edge to the left
δa	=	aileron deflection, positive is trailing edge down of right aileron
δt	=	throttle position
L, D, Y	=	aerodynamic lift, drag and side force
T	=	thrust force
Mx, My, Mz	=	aerodynamic rolling, pitching and yawing moment about the c.g.
C_L, C_D, C_Y	=	aerodynamic lift, drag and side force coefficient
C_l, C_m, C_n	=	aerodynamic rolling, pitching and yawing moment coefficient about the c.g.
J	=	value function
r_t	=	step reward
γ_d	=	discount factor
x	=	vector of states
u	=	vector of control inputs

I. Introduction

A. Motivation

It is well established that stall is a hazardous region of the flight envelope for fixed-wing aircraft, in which a high angle of attack causes flow separation from the main wing. Depending on the aircraft aerodynamic characteristics and the severity of the stall, substantial lift may be lost, causing the aircraft to sink rapidly and possibly result in loss of control [1]. Indeed, inadvertent stalls are a major co-factor of fatal accidents in general aviation aircraft [2, 3], and applying correct stall recovery actions is fundamental to minimizing altitude loss and thus the risk of ground collision [4].

The US Federal Aviation Administration (FAA) guidelines for stall recovery prescribe application of nose down elevator to reduce angle of attack, followed by progressive application of full-throttle and aileron upon flow re-attachment to return the aircraft to return the aircraft to straight level flight [5], whereas UK's Civil Aviation Authority (CAA) guidelines indicate simultaneous nose down elevator and full-throttle, followed by ailerons for wings level upon flow re-attachment [6]. The present work aims at computationally studying the optimal stall recovery actions that minimize altitude loss. On the one hand this provides a means of assessing the aforementioned guidelines which were determined experimentally, by computational means. In addition, this work could provide the basis for the design of automatic stall recovery flight control systems.

B. Previous Work

The general problem of loss of control and upset recovery has received a lot of attention throughout aviation history, but especially in the past decades, with the ever-increasing desire for higher safety standards and the advent of modern flight computers and fly-by-wire systems which give way for automatic guidance and control systems, the topic became a focal point in aviation research [7–10]. Loss of control and upsets can have many causes, ranging from human factors, system failures and nonlinear aerodynamic flight regimes. Of the latter, stall recovery is probably the most common case of aerodynamics-induced loss of control in fixed-wing aircraft. The development of modern control techniques, coupled with availability of nonlinear flight models [11–13] have fostered research in this area.

Much work has been devoted to the design of state-space feedback controllers that attempt to reduce angle of attack, and other key flight states (such as angular rates and/or side-slip angle, in the case of spins), in a regulator type fashion where the states need to be returned to a trim condition [9, 14, 15]. Other work has focused on using trajectory planning and model-predictive control (MPC) algorithms to compute appropriate control inputs to return the aircraft back to safety [16, 17].

Despite the existence of some investigations on minimal altitude loss for off-nominal maneuvers in aviation [18, 19] most of the recent upset recovery investigations have focused on the recoverability problem of upset conditions, but not so much on the actual optimality of the maneuvers with respect to total altitude loss, which is a critical figure of merit for survivability. This may be due to the fact that most of these investigations focus on large general transport aircraft, where the main focus is returning the aircraft to a controllable flight envelope, assuming there is enough airspace to effect the complete recovery to straight level flight. While this may be appropriate for upsets that occur at high altitudes,

it is not acceptable at low altitudes, as is the case in general aviation, where stall upsets tend to occur while maneuvering in and out of the traffic pattern, and thus altitude minimization is of paramount importance, where every extra meter counts, and could be the difference between life and death.

As shown in [20] the "envelope recovery" phase accounts for only a small portion of the total altitude loss, and so neglecting the ensuing pullout (returning the aircraft to straight and level flight) may lead to excessively large altitude loss. It wasn't not until recently that total altitude loss has been systematically identified as the overarching figure of merit for a successful and safe recovery maneuver. [20] and [21] both independently state that the upset recovery problem should be posed as an altitude minimization optimal control problem. Beyond this, they both also show that the minimum altitude loss problem can be cast as an infinite horizon Dynamic Programming problem that can be solved exactly via value-iteration algorithms.

One of the features of the value-iteration DP approach is that it opens the possibility of computing the optimal policy over a wide range of states, rather than for a specific initial condition, as would be the case in an MPC type algorithm for example. In addition the optimal policy can be pre-computed off-line remooving the need for real time optimization, with the convergence issues that may arise. Moreover, obtaining the optimal policy also gives the full-picture with respect to optimality, suboptimality and recoverability of the problem for a wide range of states, allowing analysis beyond the control problem. The other feature of the value-iteration Dynamic Programming approach is that it is a gradient free algorithm that is guaranteed to converge on a globally optimal policy (if one exists) regardless of the nonlinearities of the system dynamics. As is known, high angle of attach aerodynamics and flight dynamics is plagued with strong nonlinearities and intricate couplings, which can prove challenging for gradient based optimization algorithms, that may lack convergence guarantees in finite time, as well as possibly guarantees of global optimality in non-convex problems.

The main obstacle though for value-iteration DP is that it is only computationally tractable on problems with a small number of state-action dimensions [22]. Thus, both [20] and [21] are forced to restrict their analysis to reduced order point mass models, effectively leaving out attitude and rotational dynamics, and thus cannot account for nonlinerities due to stall.

With this limitation in mind, and inspired by the fundamental link between DP, approximate DP and Reinforcement Learning (RL) shown in [23], the present work attempts to address the stall recovery problem by taking from the rapidly growing field of RL and its algorithms [24]. In particular, the proximal policy optimization (PPO) algorithm [25] is chosen, a sampling based policy-gradient algorithm, considered the state of the art for continous-space RL problems. The present paper shows that this approach can solve the stall recovery problem, providing a means for analysis of the problem and the synthesis of flight control laws. Moreover the results here presented are a strong indication that the same approach could be applied to more complex upset recovery problems, such as spin recovery, which involve even stronger nonlinearities and full aircraft degrees of freedom.

II. Problem Formulation

A. Optimal Control Problem and Dynamic Programming

The goal in a stall recovery maneuver is to minimize the probability of ground collision. This is equivalent to minimizing total altitude loss, which can be formulated mathematically as a fixed final state, variable final time optimal control problem, as shown in Equation 1, where the final state is level flight ($\gamma = 0$), and $f(x, u)$ is the state transition function.

$$h^*(x_0) = \left\{ \min_{u(t),t_f} \int_0^{t_f} -V \sin\gamma \, dt \;\middle|\; \gamma(t_f) = 0 \;,\; \dot{x} = f(x,u) \;,\; x(0) = x_0 \right\} \tag{1}$$

As noted in [20, 26] the minimum altitude loss optimal control problem stated in Equation 1, can be re-cast as an infinite horizon optimal control problem, with an absorbing state at level flight, as shown in Equation 2, where the stage cost is $g(x, u) = -V \sin\gamma$.

$$h^*(x_0) = \min_{u(t)} \int_0^{\infty} g(x,u) \, dt \;\ni\; \dot{x} = \begin{cases} 0 & \text{if } \gamma \geq 0 \\ f(x,u) & \text{if } \gamma < 0 \end{cases} \tag{2}$$

In principle, this problem can be solved via Dynamic Programming (DP) algorithms such as value-iteration [27], in which the state-action space is discretized and the Bellman equation is applied recursively until convergence. In practice, though, the curse of dimensionality dictates that this is only tractable for problems with low number of states

and actions, such as banked pullouts (see for example [26] and [21]). To address problems with higher dimensions, such as banked stall recovery or spin recovery, approximate DP methods are required. As shown in [23] the algorithms stemming from Reinforcement Learning (RL) [24] can be regarded as a case of approximate Dynamic Programming, and is the key insight that drives this investigation.

B. Reinforcement Learning and Proximal Policy Optimization

The central idea in RL is to learn an optimal policy, which maps states to control inputs in order to maximize the expected cumulative reward seen in equation 3 (note that RL usually is posed as a reward maximization problem, rather than cost minimization, so the reward signal is the negative of the stage cost in Equation 2). It should be noted that $\gamma_d \leq 1$ corresponds to the discount factor, not be confused with the flight path angle γ.

$$J(x) = \mathbb{E}\left\{\sum_{i=0}^{\infty} \gamma_d^i \, r(x_t, u_t) \,\middle|\, x_t = x\right\} \tag{3}$$

The main idea of RL can be represented by equation 4. Where $Q(x, u)$ is the action-value function representing the expected cumulative reward when applying control input u in state x.

$$Q(x, u) = \mathbb{E}\left\{r_{t+1} + \gamma_d \max_{u'} Q(x_{t+1}, u') \,\middle|\, x_t = x, u_t = u\right\} \tag{4}$$

1. Proximal Policy Optimization

Within the set of existing RL algorithms, proximal policy optimization (PPO) [25] is considered the state of the art for continuous state-action space optimal control tasks, making it ideal for the stall recovery problem.

PPO aims to strike a balance between exploration and exploitation by iteratively updating the policy using a trust region constraint. The objective function of PPO is given by:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left\{\min\left(r_t(\theta)\hat{A}_t, \text{clip}(R_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right\}$$

Where θ represents the policy parameters, $r_t(\theta)$ is the ratio of probabilities between the updated policy and the previous policy, $\hat{A}_t$ is the advantage estimate at time step t, and $\text{clip}(x, a, b)$ clips the value x between a and b. The objective function $L^{CLIP}(\theta)$ encourages policy updates that result in higher rewards while staying within the trust region defined by ϵ.

2. Reward signal

The problem is modelled in discrete time with $\Delta t = 0.1$. As is usual in RL formulations, the reward signal seen in equation 5 consists not only of the ultimate optimization objective, but of other signals which are useful to minimize reward hacking [28] and accelerate convergence. The presented reward aims to be informative without subtracting from the final objective.

The reward function selected for the problem can be seen in equations 5, 6

$$r(x_t, u_t) = r_h + r_{\delta_e} + r_\alpha \tag{5}$$

Where:

$$r_h = V \sin\gamma \Delta t \tag{6a}$$

$$r_{\delta_e} = -0.1 \, |\delta_e^t - \delta_e^{t-1}| \tag{6b}$$

$$r_\alpha = \begin{cases} 0 & -40 \; deg < \alpha < 40 \; deg \\ -20 & otherwise \end{cases} \tag{6c}$$

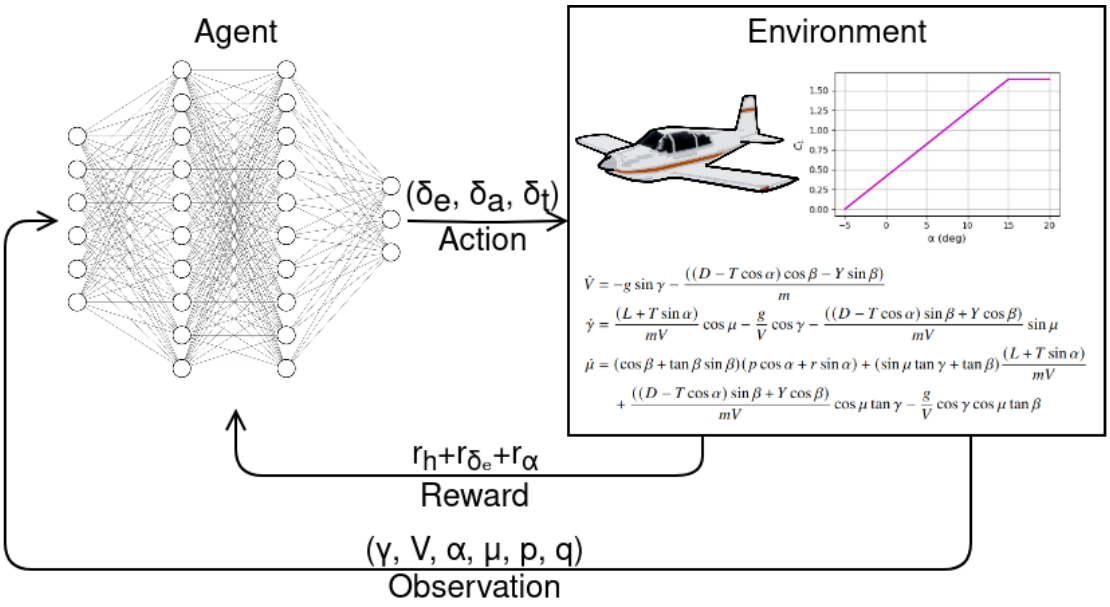


Fig. 1 system diagram

3. Neural network

The policy model is a two-layer deep, fully connected neural network with 64 neurons in each hidden layer. figure2 shows the architecture of the neural network. The input layer takes in 3-6 features, depending on the state space, and the output layer produces 2-3 outputs, depending on the control space. All inputs are normalized to a range of -1 to 1. The previously mentioned policy parameters θ are the connection weights.

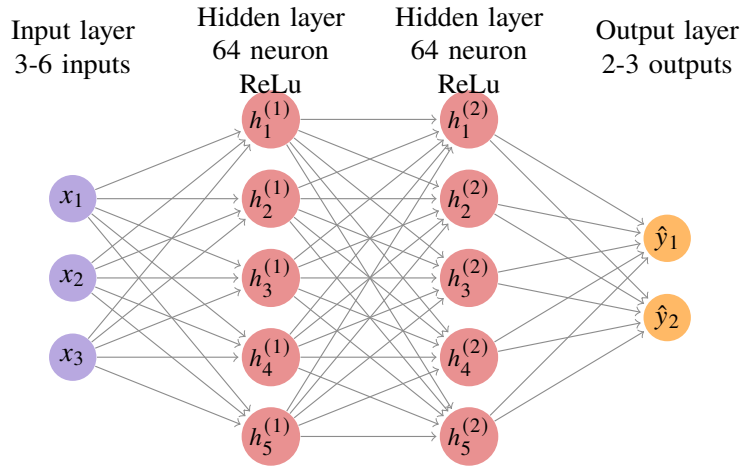


Fig. 2 Neural network architecture used as a Policy, two-layer deep, 64 neurons in each hidden layer

III. Aircraft Dynamics

A. Equations of Motion

For this investigation, it is useful to represent the aircraft equations of motion using the flight path (γ, μ) and flow angle (α, β) representation, as shown in [20]. Although altitude is of primary interest, for the optimal control problem as formulated in Equation 2 the position dynamic states (x, y, h) can be ignored, since they can be regarded simply as

outputs that don't affect the dynamics of the system. The same can be said of the course over ground angle ξ. It should be noted that although the altitude state has been removed, altitude loss is present in the objective function of the optimal control problem, as the integral of sink rate ($-V \sin\gamma$). This results in the full order system dynamics being a set of 8 nonlinear ordinary differential equations, shown in Equation 7. These equations of motion include four aerodynamic forces and three aerodynamic moments, namely lift, drag, side force and thrust (L, C, Y, T) and rolling, pitching and yawing moments (M_x, M_y, M_z).

$$\dot{V} = -g\sin\gamma - \frac{((D - T\cos\alpha)\cos\beta - Y\sin\beta)}{m} \tag{7a}$$

$$\dot{\gamma} = \frac{(L + T\sin\alpha)}{mV}\cos\mu - \frac{g}{V}\cos\gamma - \frac{((D - T\cos\alpha)\sin\beta + Y\cos\beta)}{mV}\sin\mu \tag{7b}$$

$$\dot{\mu} = (\cos\beta + \tan\beta\sin\beta)(p\cos\alpha + r\sin\alpha) + (\sin\mu\tan\gamma + \tan\beta)\frac{(L + T\sin\alpha)}{mV} \tag{7c}$$

$$+ \frac{((D - T\cos\alpha)\sin\beta + Y\cos\beta)}{mV}\cos\mu\tan\gamma - \frac{g}{V}\cos\gamma\cos\mu\tan\beta$$

$$\dot{\alpha} = q - \sec\beta(\frac{(L + T\sin\alpha)}{mV} - \frac{g}{V}\cos\gamma\cos\mu) - \tan\beta(p\cos\alpha + r\sin\alpha) \tag{7d}$$

$$\dot{\beta} = \frac{((D - T\cos\alpha)\sin\beta + Y\cos\beta)}{mV} + \frac{g}{V}\cos\gamma\sin\mu - (r\cos\alpha - p\sin\alpha) \tag{7e}$$

$$\dot{p} = \frac{M_x}{I_{xx}} - qr\frac{(I_{zz} - I_{yy})}{I_{xx}} \tag{7f}$$

$$\dot{q} = \frac{M_y}{I_{yy}} + pr\frac{(I_{zz} - I_{xx})}{I_{yy}} \tag{7g}$$

$$\dot{r} = \frac{M_z}{I_{zz}} - pq\frac{(I_{yy} - I_{xx})}{I_{zz}} \tag{7h}$$

$$\tag{7i}$$

B. Aerodynamic Model for the Grumman American AA-1 Yankee

In this paper we investigate stall recovery for the Grumman American AA-1 Yankee, a single-engine low-wing general aviation aircraft, which was extensively studied in the NASA Spin Recovery Research Program [29–32], and for which reliable high angle of attack aerodynamic data is available. The aerodynamic forces and moments can be expressed in terms of dynamic pressure and non-dimensional aerodynamic coefficients as follows:

$$L = \frac{\rho V^2 S C_L}{2} \tag{8a}$$

$$D = \frac{\rho V^2 S C_D}{2} \tag{8b}$$

$$Y = \frac{\rho V^2 S C_Y}{2} \tag{8c}$$

$$M_x = \frac{\rho V^2 S b C_l}{2} \tag{8d}$$

$$M_y = \frac{\rho V^2 S c C_m}{2} \tag{8e}$$

$$M_z = \frac{\rho V^2 S b C_n}{2} \tag{8f}$$

$$T = K_T \delta_t \tag{8g}$$

Based on the model developed in [33], the aerodynamic coefficients are modeled as shown in Equation 9 and tables 1 and 2. In particular, Figure 3 shows the lift vs. angle of attack. As should be noted, the aerodynamic coefficients depend on the elevator, aileron, rudder control surface deflections ($\delta_e, \delta_a, \delta_r$, respectively), and thrust depends on the throttle

command (δt). In the case of the thrust model, we assume a linear mapping between thrust and throttle command, where the constant K is estimated by assuming that maximum thrust is reached at the maximum cruise speed of the aircraft, as shown in Equation 10.

$$C_L(\alpha, \hat{q}, \delta e) = \begin{cases} C_{L_0} + C_{L_\alpha}\alpha_{Stall} & if\ \alpha > \alpha_{Stall}\ \ (\alpha_{stall} = 15\,\text{deg}) \\ C_{L_0} + C_{L_\alpha}\alpha + C_{L_{\delta_e}}\delta_e + C_{L_{\hat{q}}}\hat{q} & otherwise \end{cases} \tag{9a}$$

$$C_D(\alpha, \hat{q}, \delta e) = C_{D_0} + C_{D_\alpha}\alpha + C_{D_{\alpha^2}}\alpha^2 \tag{9b}$$

$$C_m(\alpha, \hat{q}, \delta e) = C_{m_0} + C_{m_\alpha}\alpha + C_{m_{\delta_e}}\delta_e + C_{m_{\hat{q}}}\hat{q} \tag{9c}$$

$$C_Y(\beta, \hat{p}, \hat{r}, \delta a, \delta_r) = C_{Y_\beta}\beta + C_{Y_{\hat{p}}}\hat{p} + C_{Y_{\hat{r}}}\hat{r} + C_{Y_{\delta_a}}\delta_a + C_{Y_{\delta_r}}\delta_r \tag{9d}$$

$$C_l(\beta, \hat{p}, \hat{r}, \delta a, \delta_r) = C_{l_\beta}\beta + C_{l_{\hat{p}}}\hat{p} + C_{l_{\hat{r}}}\hat{r} + C_{l_{\delta_a}}\delta_a + C_{l_{\delta_r}}\delta_r \tag{9e}$$

$$C_n(\beta, \hat{p}, \hat{r}, \delta a, \delta_r) = C_{n_\beta}\beta + C_{n_{\hat{p}}}\hat{p} + C_{n_{\hat{r}}}\hat{r} + C_{n_{\delta_a}}\delta_a + C_{n_{\delta_r}}\delta_r \tag{9f}$$

	C_0	α	α^2	q	δ_e
C_L	0.4100	4.6983	-	2.4200	0.3610
C_D	0.0525	0.2068	1.8712	-	-
C_m	0.0760	-0.8938	-	-7.1500	-1.0313

Table 1 Model parameters for C_L, C_D and C_m

	β	$\hat{p}$	$\hat{r}$	δ_a	δ_r
C_Y	-0.6303	0.01600	1.1000	-0.0057	0.1690
C_l	-0.1089	-0.5200	0.1900	-0.1031	0.0143
C_n	0.1003	-0.0600	-0.2000	0.0017	-0.0802

Table 2 Model parameters for C_Y, C_l and C_n

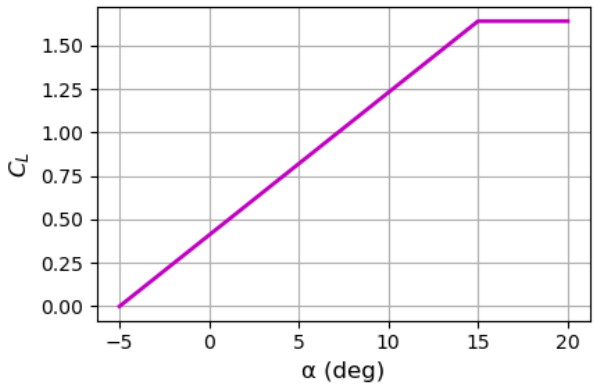


Fig. 3 C_L vs α

$$K_T = \frac{\rho S V_{max}^2 C_D(V_{max})}{2} \tag{10}$$

IV. Validation of PPO on Banked Pullout Problem

Although PPO has assured convergence, provided some reasonable assumptions are met [34], it is not guaranteed to converge on the true (global) optimal policy. In order to assess this for the case of aircraft recovery, in this section we apply the PPO algorithm to the banked pullout problem with idle power, a reduced order problem which, due to its low dimensionality, can also be solved via an exact Dynamic Programming algorithm, as shown in [26]. This provides a benchmark with which to validate the approach.

As shown in [26], for the case of banked pullouts with idle power, the equations of motion can be reduced to Equation 11, which has a three-dimensional state space (V, γ, μ) and a two-dimensional control space $(C_L^{cmd}, \dot{\mu}^{cmd})$.

It should be noted that in this problem the attitude dynamics are abstracted out, assuming there is an inner loop that controls them, and thus the control inputs are the commanded lift coefficient (C_L^{cmd}) (which can be viewed as an angle of attack command) and the commanded bank rate ($\dot{\mu}^{cmd}$). The range of the state and control space for this problem are shown in Tables 3 and 4.

$$\dot{V} = -g \sin\gamma - 1/2\rho \frac{S}{m} V^2 C_D(C_L^{cmd}) \tag{11a}$$

$$\dot{\gamma} = 1/2\rho \frac{S}{m} V C_L^{cmd} \cos\mu - \frac{g}{V} \cos\gamma \tag{11b}$$

$$\dot{\mu} = \dot{\mu}^{cmd} \tag{11c}$$

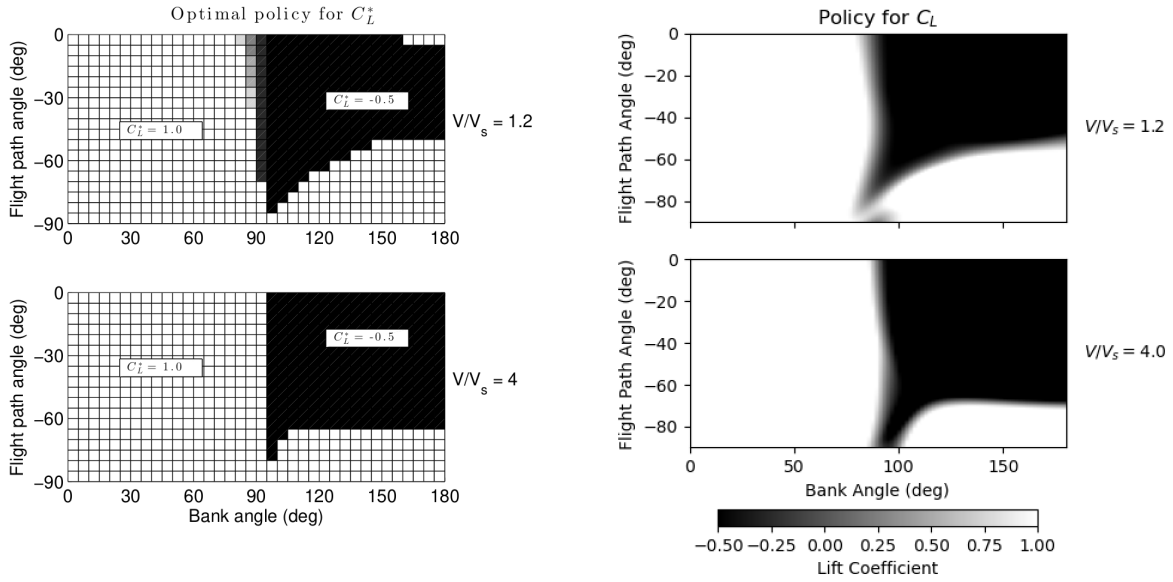
Table 3 Control Space for Banked Pullout with Idle Power

u	Min.	Max.
C_L^{cmd}	-0.5	1.0
$\dot{\mu}^{cmd}$ $(\frac{deg}{s})$	-30	30

Table 4 Initial State for Banked Pullout with Idle Power

x	Min.	Max.
γ (deg)	-180	0
$\hat{V}$	0.9	4.0
μ (deg)	-20	200

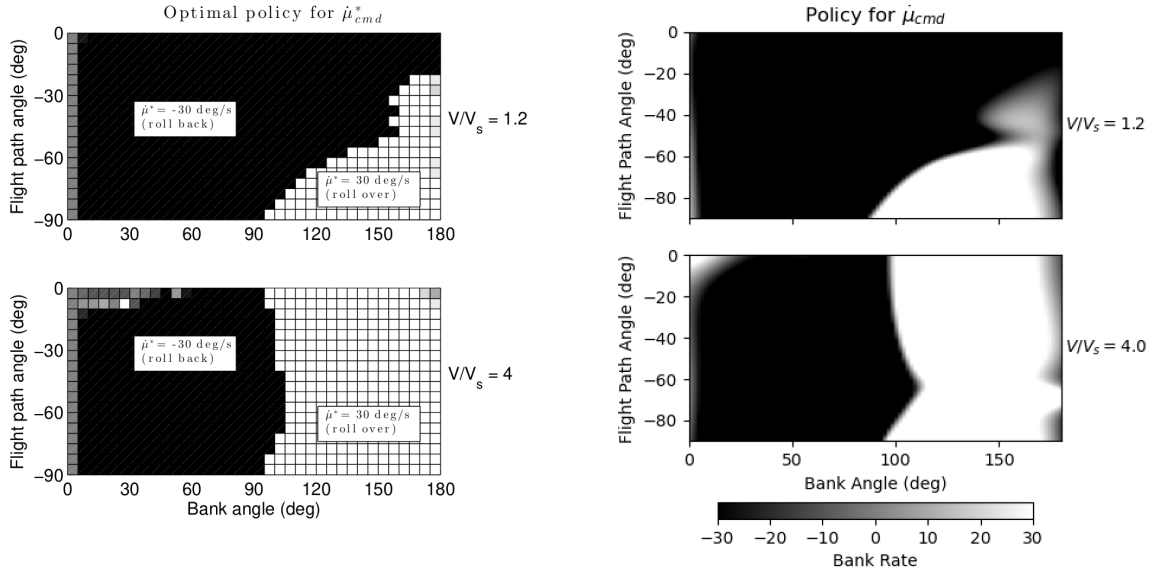
Figures 4 and 5 show a side-to-side comparison of the optimal policy obtained via DP (left) and PPO algorithm (right). As can be observed, the optimal policy computed via PPO is almost identical to that computed by DP showing that indeed PPO, although a sampling based approximate algorithm, indeed converges to the correct policy. In addition to this, PPO was run multiple times with different randomized seeds, and for all cases the algorithm converges to the same policy, providing further indication that indeed it is finding a global optimum.



(a) Optimal policy computed with DP (reproduced from [26])

(b) Optimal policy computed with RL PPO

Fig. 4 Lift Coefficient Policy Comparison



(a) Optimal policy computed with DP (reproduced from [26])

(b) Optimal policy computed with RL PPO

Fig. 5 Bank Rate Policy Comparison

V. Stall Recovery

In this section, we apply the PPO algorithm to the stall recovery problem, in which a nonlinear lift vs. angle of attack curve is used, as shown in Figure 3. We address the case of symmetric (wings level) stall recovery, which has four dynamic states (V, γ, α and q) and two control inputs ($\delta e, \delta t$). Equation 12 shows the equations of motion. Ranges for control and state space are shown in Tables 5 and 6 The other control inputs are set to a neutral position $\delta r = \delta a = 0$ and side-slip, roll rate and yaw rate are also zero ($\beta = p = r = 0$).

$$\dot{V} = -g \sin\gamma - \frac{D - T\cos\alpha}{m} \tag{12a}$$

$$\dot{\gamma} = \frac{L + T\sin\alpha}{mV} - \frac{g\cos\gamma}{V} \tag{12b}$$

$$\dot{\alpha} = q - \frac{L + T\sin\alpha}{mV} + \frac{g\cos\gamma}{V} \tag{12c}$$

$$\dot{q} = \frac{M_y}{I_{yy}} \tag{12d}$$

$$\dot{\xi} = \dot{\mu} = \dot{\beta} = \dot{p} = \dot{r} = 0 \tag{12e}$$

Table 5 Control Space for Symmetric Stall Recovery

u	Min.	Max.
δ_e (deg)	-25	15
δ_t	0	1

Table 6 State space for Symmetric Stall Recovery

x	Min.	Max.
γ (deg)	-90	0
$\hat{V} = V/V_s$	0.9	2.0
α (deg)	14	20
q $(\frac{deg}{s})$	-12	12

Figure 6 shows the optimal policy for elevator and throttle inputs as well as the minimal altitude loss as a function of flight path angle and angle of attack, for three different airspeeds (as measured relative to the stall speed). As can be seen,

the optimal policy shows that for stalled angles of attack ($\alpha > 15$ deg) and at speeds near stall ($\hat{V} \approx 1$), there is a region of low flight path angles for which the optimal action is to apply full-nose down elevator deflection and full-throttle. The lower the speed, the larger the range of angles of attack for which full pitch-down elevator is required. This is expected sin the nose down elevator deflection not only reduces the angle of attack but also increases airspeed. At angles of attack below stall, the optimal policy is full-pitch up elevator and full-throttle. These results are aligned with experimental results shown by [4], and the guidelines published by UK's CAA, which call for simultaneous application of full nose down elevator and full-throttle, followed by pitch up elevator once the angle of attack is reduced to effect the pullout.

Figure 7 shows a stall recovery trajectory, starting from a stalled angle of attack $\alpha = 20$ deg at level flight and low speed , as could be the case of a wings level stall. As can bee seen the PPO policy results in an initial full pitch down elevator with simultaneous application of full-throttle, until the angle of attack is reduced to close the stall angle, at which point the elevator reverses to a near full pitch-up position, and then subsides to a moderate pitch-up until the pullout is complete. The altitude loss is $\tilde{1}5$ meters, about twice the aircraft wing span, highlighting the hazardous consequence of stall.

Overlaid on Figure 7, is also shown the trajectory of a stall recovery with 0.75 seconds delay in the application of full throttle (this is the point at which the angle of attack and elevator deflection stabilize at their steady state values), which results in an altitude loss of $\tilde{2}0$ m, a 30% increase in altitude loss. This shows that throttle management is very important in minimizing altitude loss in a stall recovery, and also is aligned with the findings of [4]), which call for immediate application of full throttle, rather than delayed application. Given the limitations of the aerodynamic model used in this study in terms of power effects, further analyzes with more sophisticated models is warranted before drawing any further conclusions.

Another aspect that is worth noting, is that for both stall recovery trajectories shown in Figure 7 angle of attack stabilizes at a value a couple degrees above that of stall. This is somewhat counteractive, since one would expect that given the parabolic increase in drag with angle of attack, which in turn reduces speed, and thus reduces lift and consequently reduces turn rate $\dot{\gamma}$, there would be a trade-off between lift and drag, resulting in an angle of attack right at stall. This reasoning neglects the fact that turn rate $\dot{\gamma}$ also has a term of the form $T \sin\alpha$, thus having a higher angle of attack projects more of the thrust force onto the centripetal direction, resulting in a higher turn rate, which incentivizes having a higher angle of attack. Clearly, reducing angle of attack below stall is of paramount importance to ensure controllability, as indicated by both the FAA and CAA guidelines. To produce policies that avoid remaining in the stalled flight envelope one could add a cost that penalizes this , such as for example a log barrier at α_s. This is an aspect that requires further investigation.

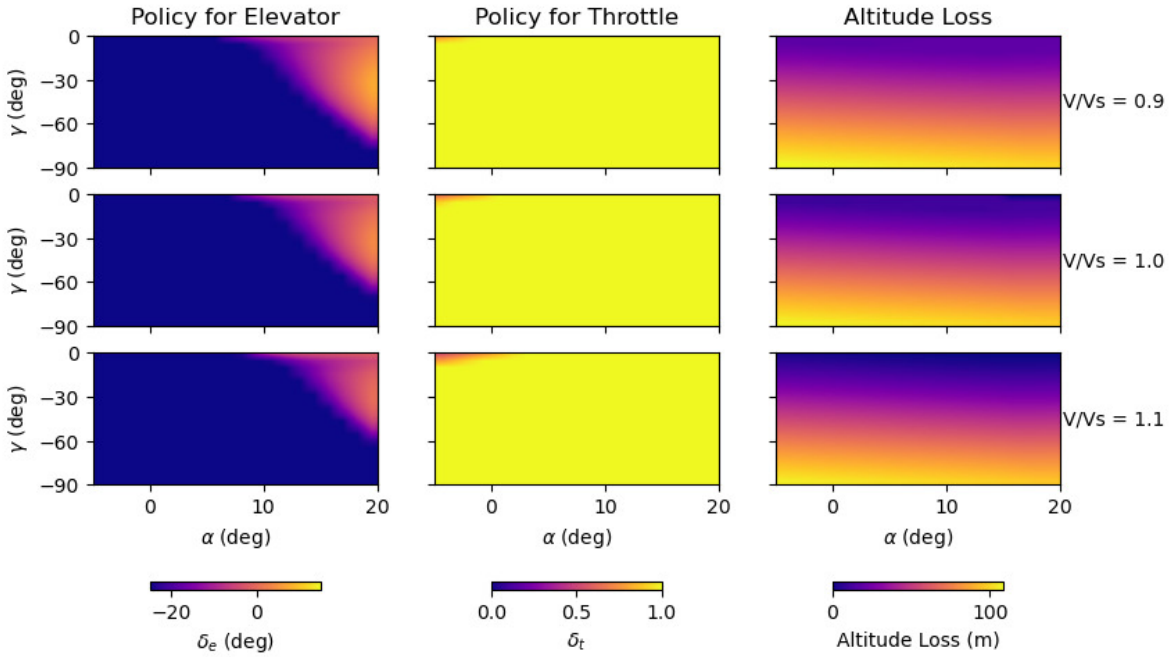


Fig. 6 Symmetric stall recovery results with PPO, for $q = 0$ deg/s.

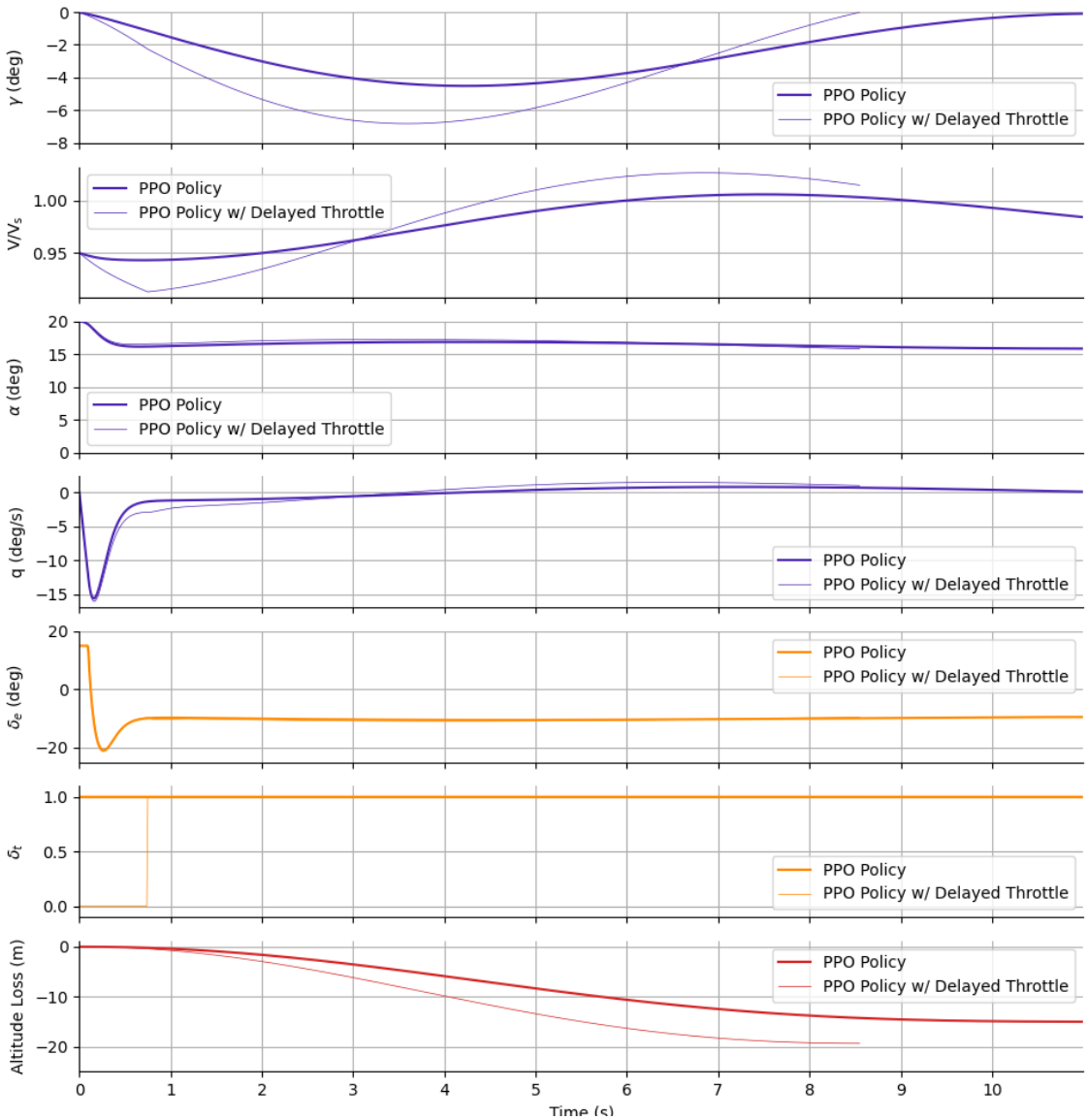


Fig. 7 Symmetric stall recovery trajectories with the PPO policy, and the PPO policy with delayed application of throttle.

VI. Conclusions

Stall is a major hazard in general aviation and having methodologies to compute optimal recovery actions that minimize altitude loss is fundamental to better assess guidelines for pilots regarding correct recovery actions and also enable the design of automatic stall recovery flight control systems. The results here presented indicate the effectiveness of the PPO algorithm in solving the optimal stall recovery problem. Firstly, it is shown that PPO indeed computes the optimal policy for altitude minimization by comparing it with that computed by value-iteration, an exact Dynamic Programming method, on the banked pullout problem, a benchmark reduced order problem. Secondly, PPO is applied to the symmetric stall recovery obtaining the optimal policy for elevator and throttle commands as a function of angle of attack, flight path angle and airspeed. These results show that indeed when stalled, and at low speeds, the optimal sequence is to first apply full nose-pitch down elevator and full throttle, and once flow re-attaches, apply full nose-pitch up elevator to recover straight and level flight, in agreement with standard pilot guidelines . These results also show that correct throttle management is of paramount importance to minimize total altitude loss, whereby delay of throttle

application may result in a as much as a 30% increase in altitude loss. This is aligned with experimental results reported in the literature.

Beyond the specific results presented for optimal stall recovery, the findings of this investigation suggest that the PPO algorithm could be applied to more complex upset recovery scenarios, such as the spin recovery problem, a higher dimensional problem involving not only high angle of attack but also high roll and yaw rates, and which still stands as an open problem in aeronautical research. Being able to systematically analyze these upset recovery problems could lead to significant advances in the study of loss of control and upset recovery in fixed-wing aircraft.

References

[1] FAA, "Airplane Flying Handbook (FAA-H-8083-3A)," , 2004. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Airplane+Flying+Handbook#1.

[2] Landsberg, B., "Stall/Spin: Entry point for crash and burn?" *Safe Pilots. Safe Skies.*, AOPA Air Safety Foundation, Frederick, MD, 2003.

[3] J Brownlow, N Everett, G Gratton, M Jackson, I Lee, and J Thorpe, "A Study of Fatal Stall or Spin Accidents to UK Registered Light Aeroplanes 1980 to 2008," 2010.

[4] Gratton, G. B., Hoff, R. I., Bromfield, M., Rahman, A., Harbour, C., and Williams, S., "Evaluating a set of stall recovery actions for single engine light aeroplanes," *The Aeronautical Journal*, 2014, pp. 1–28.

[5] FAA, "Stall and Spin Awareness Training (AC 61-67C)," , 2000.

[6] UK Civil Aviation Authority, "Handling Sense Leaflet Stall /Spin Awareness," 2009, pp. 2–3.

[7] Belcastro, C. M., "Validation of Safety-Critical Systems for Aircraft Loss-of-Control Prevention and Recovery," *AIAA Guidance Navigation and Control Conference*, 2012, pp. 1–30.

[8] Belcastro, C., "Loss of Control Prevention and Recovery: Onboard Guidance, Control, and Systems Technologies," *AIAA Guidance, Navigation, and Control Conference*, Minneapolis, MN, 2012.

[9] Kwatny, H. G., Dongmo, J. E., Chang, B.-C., Bajpai, G., Yasar, M., and Belcastro, C., "Nonlinear Analysis of Aircraft Loss of Control," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 1, 2013, pp. 149–162. https://doi.org/10.2514/1.56948, URL http://arc.aiaa.org/doi/abs/10.2514/1.56948.

[10] Engelbrecht, J., Pauck, S. J., and Peddle, I. K., "A Multi-Mode Upset Recovery Flight Control System for," *AIAA Guidance Navigation and Control Conference*, 2013, pp. 1–17.

[11] Jordan, T., Langford, W., Belcastro, C., Foster, J. V., Shah, G. H., Howland, G., and Kidd, R., "Development of a Dynamically Scaled Generic Transport Model Testbed for Flight Research Experiments," *AUVSI Unmanned Systems North America Symposium*, Anaheim, CA, 2004.

[12] Foster, J. V., Cunningham, K., Fremaux, C. M., Shah, G. H., Stewart, E. C., Rivers, R. A., Wilborn, J. E., and Gato, W., "Dynamics Modeling and Simulation of Large Transport," *AIAA Guidance Navigation and Control Conference*, 2005, pp. 1–13.

[13] Murch, A. M., and Foster, J. V., "Recent NASA Research on Aerodynamic Modeling of Post-Stall and Spin Dynamics of Large Transport Airplanes," *45th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, 2007.

[14] Lee, D.-C., and Nagati, M. G., "Optimal Control Strategy for Spin Recovery," *AIAA Guidance, Navigation and Control Conference*, New Orleans, LA, 1997.

[15] Crespo, L. G., Kenny, S. P., and Cox, D. E., "Analysis of Control Strategies for Aircraft Flight Upset Recovery," *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2012, pp. 1–31.

[16] Babl, M. L., and Engelbrecht, J. A., "Automatic Deep Stall Recovery using Optimal Trajectory Planning," *IFAC-PapersOnLine*, Vol. 53, No. 2, 2020, pp. 15508–15515. https://doi.org/10.1016/j.ifacol.2020.12.2377, URL https://doi.org/10.1016/j.ifacol.2020.12.2377.

[17] Cunis, T., Liao-Mcpherson, D., Condomines, J.-p., Burlion, L., and Kolmanovsky, I., "Economic Model-Predictive Control Strategies for Aircraft Deep-stall Recovery with Stability Guarantees," 2020.

[18] Jones, R. T., "Dynamics of Ultralight Aircraft - Dive Recovery of Hang Gliders (NASA-TM-X-73229)," Tech. rep., NASA Ames Research Center, 1977.

[19] Hyde, D., "Minimum Altitude Loss Gliding Turns with Terminal Constraints (Return To Runway After Engine Failure)," *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, 2005. https://doi.org/doi:10.2514/6.2005-5897, URL http://dx.doi.org/10.2514/6.2005-5897.

[20] Bunge, R. A., "Automatic Spin Detection and Recovery for Small Aircraft," Ph.D. thesis, Stanford University, Stanford, CA, 2017.

[21] Engelbrecht, J. A. A., "Automatic Flight Envelope Recovery for Large Transport Aircraft," Ph.D. thesis, University of Stellenbosch, 2016.

[22] Bertsekas, D. P., "Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC," *European Journal of Control*, 2005.

[23] Bertsekas, D., *Reinforcement Learning and Optimal Control*, Athena Scientific optimization and computation series, Athena Scientific, 2019. URL https://books.google.com.ar/books?id=ZlBIyQEACAAJ.

[24] Sutton, R. S., and Barto, A. G., *Reinforcement Learning: An Introduction*, 2nd ed., The MIT Press, 2018. URL http://incompleteideas.net/book/the-book-2nd.html.

[25] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O., "Proximal Policy Optimization Algorithms," *CoRR*, Vol. abs/1707.06347, 2017. URL http://arxiv.org/abs/1707.06347.

[26] Bunge, R. A., Pavone, M., and Kroo, I. M., "Minimal Altitude Loss Pullout Maneuvers," *AIAA Guidance, Navigation and Control Conference*, Kissimmee, FL, 2018.

[27] Bertsekas, D. P., *Dynamic Programming and Optimal Control*, Vol. 2, Athena Scientific, 2012.

[28] Amodei, D., Olah, C., Steinhardt, J., Christiano, P. F., Schulman, J., and Mané, D., "Concrete Problems in AI Safety," *CoRR*, Vol. abs/1606.06565, 2016. URL http://arxiv.org/abs/1606.06565.

[29] Newsom, W. A., Satran, D. R., and Johnson, J. L., "Effects of wing leading edge modifications on a low-wing full scale general aviation airplane: Wind Tunnel Investigation of High-Angle-of-Attack Aerodynamic Characteristics (NASA-TP-2011)," NASA Langley Research Center, Hampton, VA, 1982. URL http://www.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=N8226217AH.

[30] Stough, H. I., DiCarlo, D., and Patton, J. M. J., "Flight Investigation of the Effects of an Outboard Wing-Leading-Edge Modification on Stall/Spin Charcateristics of a Low-Wing, Single-Engine, T-Tail Light Airplane," Tech. rep., 1987.

[31] Manuel, G. S., DiCarlo, D., Stough, H. I., Brown, P., and Stuever, R., "Investigations of Modifications to Improve the Spin Resistance of a High-wing, Single-engine, Light Airplane," *SAE General Aviation Aircraft Meeting Exposition*, Wichita, KS, 1989. URL http://papers.sae.org/891039.

[32] Stough, H. I., Patton, J. M. J., and Sliwa, S. M., "Flight Investigation of the Effect of Tail Configuration on Stall, Spin and Recovery Characteristics of a Low-Wing General Aviation Research Airplane (NASA-TP-2644)," NASA Langley Research Center, Hampton, VA, 1987.

[33] Riley, D. R., "Simulator study of the stall departure characteristics of a light general aviatoin airplane with and without wing-leading edge modifcation (NASA-TM-86309)," NASA Langley Research Center, Hampton, VA, 1985.

[34] Holzleitner, M., Gruber, L., Arjona-Medina, J., Brandstetter, J., and Hochreiter, S., *Convergence Proof for Actor-Critic Methods Applied to PPO and RUDDER*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2021, pp. 105–130. https://doi.org/10.1007/978-3-662-63519-3_5, URL https://doi.org/10.1007/978-3-662-63519-3_5.