Delft University of Technology
Department of Control & Simulation

Technical Report

W. Chan

Date:	insert date here
Subject:	RL for Flying-V
Supervisors:	Dr. E.J. van Kampen
Project term:	01/2024 - 08/2024
E-mail:	w.y.chan@student.tudelft.nl

IDHP pseudo code

Table 1: IDHP step by step, α tracking task for short period model.

Variable	t_0	t_1	t	
Env reward, c	$c_0 = -\frac{1}{2}(\alpha_{0,ref}-\alpha_0)^2$	$c_1 = -\frac{1}{2}(\alpha_{1,ref}-\alpha_1)^2$	$c_t = -\frac{1}{2}(\alpha_{t,ref}-\alpha_t)^2$	(1.1)
Env obs, s	$s_0 = (\alpha_{0,ref}-\alpha_0)$	$s_1 = (\alpha_{1,ref}-\alpha_1)$	$s_t = (\alpha_{t,ref}-\alpha_t)$	(1.2)
Model state, x	$x_0 = \begin{bmatrix}\alpha_0 & q_0\end{bmatrix}^\top$	$x_1 = \begin{bmatrix}\alpha_1 & q_1\end{bmatrix}^\top$	$x_t = \begin{bmatrix}\alpha_t & q_t\end{bmatrix}^\top$	(1.3)
Env/Model action, a	$a_0 = [\delta_{0,e}]$	$a_1 = [\delta_{1,e}]$	$a_t = [\delta_{t,e}]$	(1.4)
RL				
Critic/TD error, e	$e_0 = None$	$e_1 = \lambda(x_0) - \frac{\partial c_0}{\partial x_0} - \gamma\lambda'(x_1)\frac{\partial x_1}{\partial x_0}$	$e_t = \lambda(x_{t-1}) - \frac{\partial c_{t-1}}{\partial x_{t-1}} - \gamma\lambda'(x_t)\frac{\partial x_t}{\partial x_{t-1}}$ [1]	(1.5)
Critic weight, w_c	$w_{0,c} = init$	$w_{1,c} = w_{0,c} - \eta_c e_0^\top \frac{\partial\lambda(x_0)}{\partial w_{0,c}}$	$w_{t,c} = w_{t-1,c} - \frac{\eta_c}{2} e_{t-1}^\top \frac{\partial\lambda(x_{t-1})}{\partial w_{t-1,c}}$	(1.6)
Target critic	$w_{0,c'} = w_{0,c}$	$w_{1,c'} = \tau w_{1,c} + (1-\tau)w_{0,c'}$	$w_{t,c'} = \tau w_{t,c} + (1-\tau)w_{t,c'}$	(1.7)
Actor loss,	$L_0 = c_0 + \gamma J(x_0)$	$L_1 = c_1 + \gamma J(x_1)$	$L_t = c_t + \gamma J(x_t)$	(1.8)
Actor weight, w_a	$w_{0,a} = init$	$w_{1,a} = w_{0,a} - \eta_a(\frac{\partial c_0}{\partial x_1} + \gamma\lambda(x_1))\frac{\partial x_1}{\partial a_0}\frac{\partial a_0}{\partial w_{0,a}}$	$w_{t,a} = w_{t-1,a} - \eta_a(\frac{\partial c_{t-1}}{\partial x_t} + \gamma\lambda(x_t))\frac{\partial x_t}{\partial a_{t-1}}\frac{\partial a_{t-1}}{\partial w_{t-1,a}}$	(1.9)
RLS				
δx	$\delta x_0 = None$	$\delta x_1 = x_1 - x_0$	$\delta x_t = x_t - x_{t-1}$	(1.10)
δa	$\delta a_0 = None$	$\delta a_1 = a_1 - a_0$	$\delta a_t = a_t - a_{t-1}$	(1.11)
X	$X_0 = None$	$X_1 = \begin{bmatrix}\partial x_1 & \partial a_1\end{bmatrix}^\top$	$X_t = \begin{bmatrix}\partial x_t & \partial a_t\end{bmatrix}^\top$	(1.12)

[1] Using Zhou's TD convention, reward denoted as the earlier timestep.

RLS gain, k	$k_0 = None$	$k_1 = \frac{\Sigma_0 X_1}{\gamma_R + X_1^\top \Sigma_0 X_1}$	$k_t = \frac{\Sigma_{t-1} X_t}{\gamma_R + X_t^\top \Sigma_{t-1} X_t}$	(1.13)
Estimated $\hat{\delta x}$	$\hat{\delta x}_0 = None$	$\hat{\delta x}_1 = X_1^\top \theta_0$	$\hat{\delta x}_t = X_t^\top \theta_{t-1}$	(1.14)
RLS error, ϵ	$\epsilon_0 = None$	$\epsilon_1 = \delta x_1 - \hat{\delta x}_1$	$\epsilon_t = \delta x_t - \hat{\delta x}_t$	(1.15)
Estimated param, θ	$\theta_0 = init$	$\theta_1 = \theta_0 + k_1 \epsilon_1$	$\theta_t = \theta_{t-1} + k_t \epsilon_t$	(1.16)
RLS cov, Σ	$\Sigma_0 = init$	$\Sigma_1 = \frac{1}{\gamma_R}(\Sigma_0 - k_1 X_1^\top \Sigma_0)$	$\Sigma_t = \frac{1}{\gamma_R}(\Sigma_{t-1} - k_t X_t^\top \Sigma_{t-1})$	(1.17)