SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Missile aerodynamic design using reinforcement learning and transfer learning

Xinghui YAN, Jihong ZHU*, Minchi KUANG & Xiangyang WANG

Department of Computer Science and Technology, Tsinghua University, Beijing 100084*, China;*
State Key Laboratory of Intelligent Technology and System, Tsinghua University, Beijing 100084*, China*

Appendix A Details of the application of SL-DDPG to missile aerodynamic design

The design parameters of the missile aerodynamic shape is defined in Table A1. The baseline model of the missile consists of a cone nose, a cylindrical body, four trapezoidal wings and tails. The wings and tails are arranged in "+" form and both have hexagonal airfoil. The wings and tails' trailing edges are set perpendicular to body axis. So SWEEP can be determined by CHORD_1, CHORD_2 and SSPAN_2. Now that LNOSE, LCENTER and DCENTER are mainly determined by components arrangement during missile overall design, the three parameters are kept constant during the optimization.

The architecture of DNNs in SL-DDPG is illustrated in Figure A1. We take the design parameter set as network input and the the parameters' change as output.

The iterative interaction of the agent and the hybrid environment is described as Figure A2. The Fluent calculation takes the Reynolds averaged navier-stokes (RANS) equations to describe the turbulent flow and Spalart-Allmaras (S-A) model to resolve the equations.

We train the DDPG network in Datcom environment first by maximizing the following reward function:

$$r_D(s_t, a_t) = \begin{cases} D(s_t + a_t) - D(s_0), & \text{XCP} \in [0.535, 0.635], \\ D(s_t + a_t) - D(s_0) - 10 \cdot \Delta\text{XCP}, & \text{otherwise.} \end{cases} \tag{A1}$$

Where D function denotes the lift-drag ratio calculation with Datcom, s_0 is the state correspond with baseline configuration and ΔXCP denotes the positron variation of pressure center compared with baseline. The network is trained to predict the best adjustment amount of design parameters in various conditions.

Based on the knowledge transfer discussed above, we extend the DDPG algorithm from Datcom environment to the hybrid environment using the reward function as:

$$r_H(s_t, a_t) = \begin{cases} H(s_t + a_t) - H(s_0), & \text{XCP} \in [0.535, 0.635], \\ H(s_t + a_t) - H(s_0) - 10 \cdot \Delta\text{XCP}, & \text{otherwise.} \end{cases} \tag{A2}$$

Where H function denotes the lift-drag ratio calculation with the hybrid model.

* Corresponding author (email: jhzhu@tsinghua.edu.cn)

Table A1 Definition and value bounds of design parameters

Design Parameter	Definition	Lower bound	Upper bound	Baseline
LNOSE	Nose length	-	-	0.49
LCENTER	Missile body length	-	-	3.17
DCENTER	Missile body diameter	-	-	0.18
XLE1	Distance from nose tip to wing leading edge	0.5	2.3	1.72
SWEEP1	Wing leading-edge sweep angle	-	-	-
CHORD1_1	Wing root chord	0.1	0.5	0.29
CHORD1_2	Wing tip chord	0	0.5	0.06
SSPAN1_2	Exposed semispan of wing	0.1	0.5	0.23
XLE2	Distance from nose tip to tail leading edge	2.9	3.2	3.2
SWEEP2	Wing trailing-edge sweep angle	-	-	-
CHORD2_1	Tail root chord	0.1	0.5	0.38
CHORD2_2	Tail tip chord	0	0.5	0.19
SSPAN2_2	Exposed semispan of tail	0.1	0.5	0.22

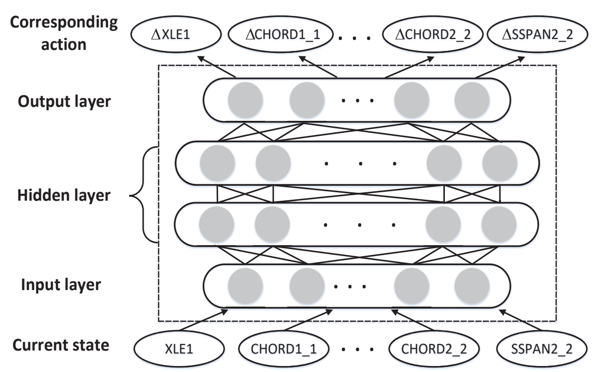

Figure A1 Network architecture of DNNs for actor and critic in SL-DDPG.

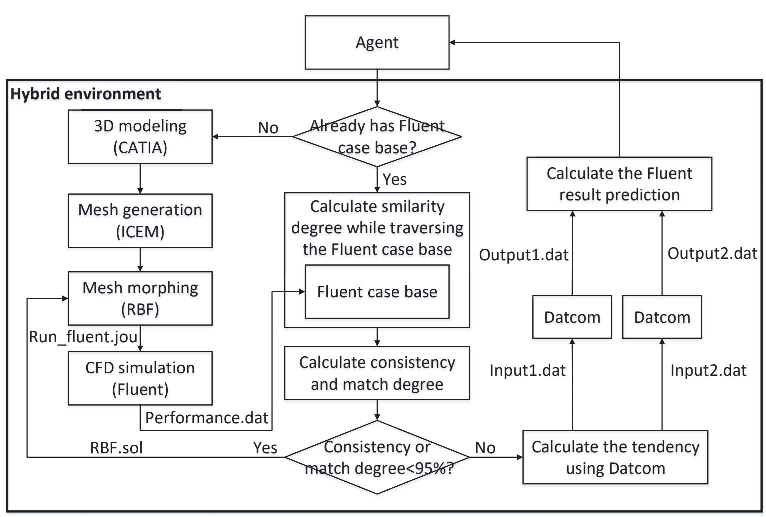

Figure A2 Reinforcement learning in the hybrid environment.