A Multi-step and Eligibility Trace Approach to Incremental Dual Heuristic Programming for Flight Control

W. Chan * and E. van Kampen †
Delft University of Technology, P.O. Box 5058, 2600GB Delft, The Netherlands

Incremental Dual Heuristic Programming (IDHP) is a succesor to the Dual Heuristic Programming (DHP) algorithm that uses an incremental system model, this algorithm showed promising flight control performance and tolerance of faults in simulation experiments. This paper studies the potential for extending IDHP through augmenting the computation of agent updates and returns, more specifically, by using eligibility trace updates and multi-step temporal difference error. This results in the IDHP(λ) and MIDHP(λ) algorithms, which are compared against IDHP in several simulated flight control scenarios with faults introduced mid-flight. The results demonstrate that the proposed algorithms have improved flight control performance and fault tolerance in terms of tracking errors when controlling a nominal aircraft and an aircraft with faults introduced.

Nomenclature

(Nomenclature entries should have the units identified)

I. Introduction

The civil aviation sector is undergoing many developments which will redefine what flight means, be it the advent of personal air vehicles, novel airliner designs, hydrogen fuelled concepts, or the increasing usage of drones [1–4]. Flight control on novel systems using existing control system design methods requires rigorous and extensive modelling and system identification campaigns, which cost a vast amount of time and resources, for instance in the case of gain-scheduled approaches to flight control system design [5]. Furthermore, the occurrence of faults causes aircraft to fly outside the normal flight envelope, resulting in loss of control [6], such an issue is exacerbated by a reliance on model-based controller synthesis techniques which focus on modelling nominal regions of the flight envelope. Fault tolerant and adaptive flight control is a trend in flight control design which directly addresses this issue, with promising control methods being actively researched [7–10].

Simultaneously, Reinforcement Learning (RL) has been developing at a rapid pace, from agents trained that can surpass human performance on games [11, 12], to ones with the ability of controlling real life systems [13, 14]. RL is a method of Machine Learning that predicates on the machine actively learning through sovereign actions, as opposed to other ML methods such as supervised learning where the machine is told what to do or what is correct. Actor-Critic Design (ACD) is a sub-field of RL which approaches the problem of RL from an optimal control perspective [15], where an agent is created comprising of an actor and a critic, both modelled using function approximators such as a neural network. The actor and critic are updated through steps known as policy improvement and policy evaluation respectively. ACD algorithms are generally classified into three categories: Heuristic Dynamic Programming (HDP, Dual Heuristic Programming (DHP), and Global Dual Heuristic Programming (GDHP) [16]. Incremental DHP (IDHP) is a successor to the DHP algorithm, which is extended with an online identified incremental model of the system dynamics to facilitate agent updates [17]. Parallel to ACD, there also exist Deep RL (DRL) algorithms where deep neural networks are used to model the agent, the present paper will, however, focus on ACD algorithms.

*MSc. Student, Faculty of Aerospace Engineering, Control and Simulation Division, Delft University of Technology.
†Associate Professor, Faculty of Aerospace Engineering, Control and Simulation Division, Delft University of Technology.

The intersection of these two developments has several promising potentials. These challenges could be overcome with the successful application of RL based flight controllers, which can learn to fly an aircraft without prior model knowledge, an advantage owed to the system agnostic nature of RL algorithms. This agnostic nature further implies the ability to use such controllers to fly systems which vary over time, or more importantly, experience faults during operation, e.g. a bird strike event on an aircraft. ACD algorithms are especially suited for this role due to their adaptiveness and high sample efficiency, which allows them to learn a stabilizing controller online during flight [17]. Success in the application of RL to flight control with faults introduced has been demonstrated in attitude control of the Innovative Control Effectors (ICE) model with a pure ADP controller [18] and an NDI hybrid ACD controller [19], as well as in attitude and velocity control of a business jet using a pure ACD controller [20].

ACD algorithms traditionally mainly use information from one timestep for agent training. However, by using information from more time steps it is possible to speed up agent learning. This can be done in two ways, by using eligibility traces where past function parameter updates are recorded and subsequently reused [21], or by using observed rewards from multiple time steps resulting in multi-step policy evaluations [22]. These ideas have been applied to the HDP and GDHP algorithms with success, improving their learning rate [21–24]. With improved learning rates, the speed at which a controller made from such algorithms recovers from sudden system faults could be improved as well. Thus, multi-step and eligibility traces for ACD are an interesting avenue to explore for fault tolerance RL based controllers. Additionally, such techniques have yet to be applied to IDHP.

Therefore, this paper's main contribution is in the extension of IDHP using the multi-step policy evaluations or temporal difference error, resulting in MIDHP, using eligibility trace updates, resulting in IDHP(λ), and a combination of the two, resulting in MIDHP(λ). These algorithms are evaluated both during nominal flight and with faults introduced to give an insight into these augmentation's effects.

This paper is outlined as follows: in Sec. II the background on IDHP, the application of IDHP to flight control, and the idea of multi-step policy evaluation and eligibility traces are presented; in Sec. III the proposed methodology is introduced; Section. IV introduces the setup used for the experiment. The main results are then presented and discussed in Sec. V and Sec. VI respectively, and the main conclusions are drawn in Sec. VII.

II. Background (Deliverable for the Mid-Term meeting at 15 weeks)

In this section you provide a background to your proposed methodology (how exactly is your method related to literature).

https://scite.ai/reports/10.1038/nature14236

III. Methodology (Deliverable for the Mid-Term meeting at 15 weeks and the Green-light meeting at 27 weeks)

In this section you present your methodology in detail. The detail presented should be such that the methodology can be reproduced by other students and researchers. Make sure you refer to literature where necessary, and emphasize any contributions you made (e.g. "by building on the method from ..., a new theory is introduced that not only takes into account addition, but also subtraction.")

Note that this section is a deliverable both at the Mid-Term meeting (week 15) and the Green-light meeting (week 27). At the Mid-Term meeting, an initial version of the methodology section should be presented. After the Mid-Term meeting, the methodology may be modified or extended, and hence the contents of this section can change as your project progresses.

A. Equations

A sample equation is included as follows:

$$\int_0^{r_2} F(r,\varphi)\mathrm{d}r\,\mathrm{d}\varphi = [\sigma r_2/(2\mu_0)] \int_0^{\infty} \exp(-\lambda|z_j - z_i|)\lambda^{-1} J_1(\lambda r_2) J_0(\lambda r_i\,\lambda \mathrm{d}\lambda). \tag{1}$$

Be sure that symbols in your equation are defined in the Nomenclature, or immediately following the equation. You can refer to equations as follows: Eq.1. Also define abbreviations and acronyms the first time they are used in the main text. (Very common abbreviations such as AIAA and NASA, do not have to be defined.)

IV. Experiment Setup (Deliverable for the Green-Light meeting at 27 weeks)

In this section, you introduce the experimental setup that you will use to answer your research questions. This can be a hardware facility, but may also consist of a simulation or other software.

V. Results (Deliverable for the Green-Light meeting at 27 weeks)

In this section, you present the main results from your experiments. Briefly discuss the results, and aid the reader in understanding your figures by providing descriptive figure captions as well as readable figures. Don't forget axis labels and legends!

Table 1 Transitions selected for thermometry

	Transition					
Line	v''		J''	Frequency, cm^{-1}	FJ, cm^{-1}	Gv, cm^{-1}
a	0	P_{12}	2.5	44069.416	73.58	948.66
b	1	R_2	2.5	42229.348	73.41	2824.76
c	2	R_{21}	805	40562.179	71.37	4672.68
d	0	R_2	23.5	42516.527	1045.85	948.76

All tables are numbered consecutively and must be cited in the text; give each table a definitive title. Be sure that you have a minimum of two columns (with headings) and two rows to constitute a proper table; otherwise reformat as a displayed list or incorporate the data into the text.

Below is a sample figure that includes axis labels:

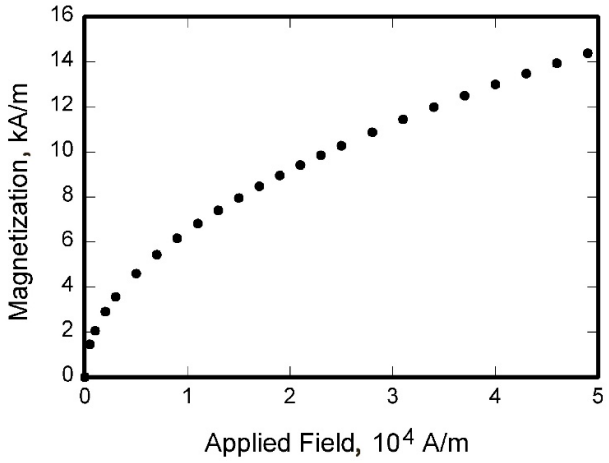

Fig. 1 Magnetization as a function of applied fields. This caption should allow the reader to understand the figure without having to read the main text.

VI. Discussion (Optional deliverable for the Green-Light meeting at 27 weeks)

In this section, you provide a discussion on the experiment results if the results where not fully conclusive. This section is optional and depends on the nature of your experiment.

VII. Conclusion (Deliverable for the Green-Light meeting at 27 weeks)

In this section, you briefly summarize the motivation, main contribution, and main result of your research. Although a conclusion may review the main points of the paper, it must not replicate the abstract. A conclusion might elaborate on the importance of the work, suggest applications and extensions, and provide hints to future work. Do not cite references in the conclusion.

Appendix

In the appendix you present methods, proofs of theorems, and results that are not of primary importance to the research, but which are nevertheless necessary to understand and reproduce the results.

Acknowledgments

An Acknowledgments section, if used, **immediately precedes** the References. Individuals other than the authors who contributed to the underlying research may be acknowledged in this section. The use of special facilities and other resources also may be acknowledged.

References

[1] McDonald, R. A., German, B. J., Takahashi, T., Bil, C., Anemaat, W., Chaput, A., Vos, R., and Harrison, N., "Future aircraft concepts and design methods," *The Aeronautical Journal*, Vol. 126, No. 1295, 2022, pp. 92–124.

[2] Hodgkinson, D., and Johnston, R., *Aviation law and drones: Unmanned aircraft and the future of aviation*, Routledge, 2018.

[3] Yusaf, T., Fernandes, L., Abu Talib, A. R., Altarazi, Y. S., Alrefae, W., Kadirgama, K., Ramasamy, D., Jayasuriya, A., Brown, G., Mamat, R., et al., "Sustainable aviation—Hydrogen is the future," *Sustainability*, Vol. 14, No. 1, 2022, p. 548.

[4] Hanover, D., Loquercio, A., Bauersfeld, L., Romero, A., Penicka, R., Song, Y., Cioffi, G., Kaufmann, E., and Scaramuzza, D., "Autonomous drone racing: A survey," *IEEE Transactions on Robotics*, 2024.

[5] Balas, G. J., "Flight control law design: An industry perspective," *European Journal of Control*, Vol. 9, No. 2-3, 2003, pp. 207–226.

[6] Kwatny, H., Dongmo, J.-E., Chang, B.-C., Bajpai, G., Yasar, M., and Belcastro, C., "Aircraft accident prevention: Loss-of-control analysis," *AIAA guidance, navigation, and control conference*, 2009, p. 6256.

[7] Sonneveldt, L., Van Oort, E., Chu, Q., and Mulder, J., "Nonlinear adaptive trajectory control applied to an F-16 model," *Journal of Guidance, control, and Dynamics*, Vol. 32, No. 1, 2009, pp. 25–39.

[8] Johnson, E., Calise, A., and De Blauwe, H., "In flight validation of adaptive flight control methods," *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2008, p. 6989.

[9] Bosworth, J., "Flight results of the NF-15B intelligent flight control system (IFCS) aircraft with adaptation to a longitudinally destabilized plant," *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2008, p. 6985.

[10] Zhang, S., and Meng, Q., "An anti-windup INDI fault-tolerant control scheme for flying wing aircraft with actuator faults," *ISA transactions*, Vol. 93, 2019, pp. 172–179.

[11] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T. P., Simonyan, K., and Hassabis, D., "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," *CoRR*, Vol. abs/1712.01815, 2017. URL `http://arxiv.org/abs/1712.01815`.

[12] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D., "Human-level control through deep reinforcement learning," *Nature*, Vol. 518, No. 7540, 2015, pp. 529–533. https://doi.org/10.1038/nature14236.

[13] Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A. A., Yogamani, S., and Pérez, P., "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 6, 2021, pp. 4909–4926.

[14] Kober, J., Bagnell, J. A., and Peters, J., "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, Vol. 32, No. 11, 2013, pp. 1238–1274.

[15] Khan, S. G., Herrmann, G., Lewis, F. L., Pipe, T., and Melhuish, C., "Reinforcement learning and optimal adaptive control: An overview and implementation examples," *Annual reviews in control*, Vol. 36, No. 1, 2012, pp. 42–59.

[16] Prokhorov, D. V., and Wunsch, D. C., "Adaptive critic designs," *IEEE transactions on Neural Networks*, Vol. 8, No. 5, 1997, pp. 997–1007.

[17] Zhou, Y., Van Kampen, E.-J., and Chu, Q., "Incremental model based online dual heuristic programming for nonlinear adaptive control," *Control Engineering Practice*, Vol. 73, 2018, pp. 13–25. https://doi.org/10.1016/j.conengprac.2017.12.011.

[18] Shayan, K., and Van Kampen, E.-J., "Online actor-critic-based adaptive control for a tailless aircraft with innovative control effectors," *AIAA Scitech 2021 Forum*, 2021, p. 0884.

[19] Li, H., Sun, L., Tan, W., Liu, X., and Dang, W., "Incremental dual heuristic dynamic programming based hybrid approach for multi-channel control of unstable tailless aircraft," *IEEE Access*, Vol. 10, 2022, pp. 31677–31691.

[20] Ferrari, S., and Stengel, R. F., "Online adaptive critic flight control," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 5, 2004, pp. 777–786.

[21] Li, T., Zhao, D., and Yi, J., "Heuristic Dynamic Programming strategy with eligibility traces," *2008 American Control Conference*, 2008, pp. 4535–4540. https://doi.org/10.1109/ACC.2008.4587210.

[22] Luo, B., Liu, D., Huang, T., Yang, X., and Ma, H., "Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems," *Information Sciences*, Vol. 411, 2017, pp. 66–83. https://doi.org/https://doi.org/10.1016/j.ins.2017.05.005.

[23] Wang, D., Wang, J., Zhao, M., Xin, P., and Qiao, J., "Adaptive Multi-Step Evaluation Design With Stability Guarantee for Discrete-Time Optimal Learning Control," *IEEE/CAA Journal of Automatica Sinica*, Vol. 10, No. 9, 2023, pp. 1797–1809. https://doi.org/10.1109/JAS.2023.123684.

[24] Ye, J., Bian, Y., Xu, B., Qin, Z., and Hu, M., "Online Optimal Control of Discrete-Time Systems Based on Globalized Dual Heuristic Programming with Eligibility Traces," *2021 3rd International Conference on Industrial Artificial Intelligence (IAI)*, 2021, pp. 1–6. https://doi.org/10.1109/IAI53119.2021.9619346.