RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ONERA
THE FRENCH AEROSPACE LAB

www.onera.fr

PROPOSITION DE SUJET DE THESE

Intitulé : Apprentissage par renforcement profond basé modèle pour la commande de drones avec intégration de connaissances physiques a priori

Référence : **TIS-DTIS-2026-08**
(à rappeler dans toute correspondance)

Début de la thèse : Dernier trimestre 2026	**Date limite de candidature :**

Thématiques scientifiques : Robotique & Autonomie, Intelligence Artificielle et Décision

Mots clés : Apprentissage par Renforcement ; Réseaux de neurones informés par les modèles ; Commande de robots et drones

Profil et compétences recherchées :
Master 2 Recherche ou Diplôme d'ingénieur avec spécialisation en apprentissage et/ou robotique
Bon niveau rédactionnel, bon niveau d'anglais, expérience et goût pour la recherche appliquée.

Compétences : Connaissances en apprentissage automatique, notamment en apprentissage par renforcement ; Expérience avec la modélisation et la commande de systèmes dynamiques ; Intérêt pour la robotique en particulier aérienne

Outils : TensorFlow, PyTorch, Programmation Objet (Python, C++)

Présentation du projet doctoral, contexte et objectif :

L'apprentissage par renforcement profond (*deep RL*) est une approche de plus en plus utilisée pour commander des robots, soit pour la planification de tâches, soit pour la commande à plus bas niveau. Cependant, l'application d'outils de *deep RL* pour apprendre des politiques efficaces directement sur des plates-formes robotiques fait encore face à de nombreux défis, tels que l'évitement de mouvements non sûrs lors de l'exploration, l'obtention de performances stables, et un volume d'échantillons et de temps pour l'apprentissage encore élevé.

Les algorithmes d'apprentissage par renforcement dits *model-based* [1] (MBRL) exploitent un modèle explicite de la dynamique et de l'environnement pour prédire le comportement et déterminer la politique d'un agent robotique. La mise en œuvre de ce type de méthodes est intéressante en particulier pour la commande des systèmes robotiques autonomes, notamment les drones, puisque le modèle peut aider à faire face aux défis mentionnés ci-dessus : anticipation de mouvements non sûrs pour les éviter, garanties plus fortes de stabilité par la théorie de la commande, et accélération de l'exploration et de l'apprentissage grâce au guidage par le modèle. Parmi les cibles applicatives, les drones à voilure fixe, les véhicules terrestres et les systèmes poly-articulés pourraient particulièrement en bénéficier.

Les modèles utilisés dans ces approches peuvent être établis a priori à partir de la physique du système, appris entièrement à partir des données, ou hybrides, à mi-chemin entre ces deux catégories. La modélisation hybride des robots est un sujet en plein essor dans la littérature, notamment pour l'identification de systèmes [2]. Du côté de la commande, l'application de MBRL hybride a déjà été validée en simulation avec des systèmes de commande classiques sur des problèmes standards, pour lesquels l'utilisation de *neural ODEs* [12] comme source d'a priori physique permet une meilleure efficacité en échantillons des approches [3]. En parallèle, l'algorithme TD-MPC (*Temporal Difference Model Predictive Control*) qui comprend un schéma prédictif exploitant le modèle appris en ligne sans a priori physique, présente des performances supérieures aux algorithmes de RL classiques sur un benchmark de référence [4], et une première application de cette méthode pour la commande d'un drone à voilure fixe en présence de perturbations a été réalisée [5].

Dans la lignée de ces travaux, la thèse s'intéresse à l'apport d'un a priori physique dans la constitution du modèle au sein d'algorithmes de MBRL pour la commande des robots autonomes.

Après une bibliographie sur les travaux récents de MBRL et l'utilisation d'a priori physiques en apprentissage, un premier travail consistera à porter la méthodologie à base de *neural ODEs* proposée dans [3] à un drone à voilure fixe, en utilisant le modèle et le simulateur introduits par [5], pour la comparer à l'existant en MBRL sans a priori physique, dont les algorithmes de l'état de l'art issus de TD-MPC [4] ou PETS [6].

Dans un second temps, il s'agira d'aborder un certain nombre de problèmes bien identifiés et ouverts pour ce type d'application à l'aide de l'existant :

- La capacité à gérer les tâches à horizon temporel long pour le système commandé, notamment lorsque celui-ci comprend des dynamiques à constantes de temps différentes (boucle d'attitude et de guidage pour un drone), par exemple en introduisant des séquences d'objectifs à horizons plus courts [14],
- La recherche de la méthode la plus adaptée pour gérer la sparsité du signal de récompense des environnements, entre d'un côté l'exploitation du mécanisme de *Hindsight Experience Replay* [10][11], et de l'autre les techniques d'enrichissement direct du signal de récompense [15]
- La recherche de stratégies pour gérer l'observabilité partielle subie par les systèmes, en utilisant des RNNs [13] ou des *Deep State Space Models* [9], afin d'améliorer la résilience face à des conditions opérationnelles incertaines ou changeantes.
- La généralisation de la méthode à différentes tâches robotiques, telles que rallier plusieurs points de passage successifs sous contrainte, pour des systèmes pouvant présenter des incertitudes paramétriques (variabilité dans une même population).

Les stratégies proposées seront évaluées et comparées sur divers scénarios de simulation représentatifs des défis rencontrés. Ces travaux pourront s'appuyer sur des outils de simulation déjà disponibles au sein du laboratoire ou en *open-source* (par exemple JSBSim [7] ou Flycraft [8] pour les drones à voilure fixe). On cherchera dans un second temps à étudier la généricité des méthodes proposées en les confrontant à un corpus plus large de benchmarks pour la commande de systèmes robotiques.

Cette thèse s'inscrit dans le cadre du projet ANR HAMMER du PEPR Accélération Robotique portant sur l'hybridation données-modèles pour la locomotion robotique. Des échanges ponctuels avec les autres laboratoires impliqués dans le projet pourront venir enrichir le travail réalisé dans la thèse.

Références :

[1] Moerland T. M., Broekens J., Plaat A., Jonker C. M. (2023). Model-based reinforcement learning: A survey. Foundations and Trends in Machine Learning, 16(1), 1-118.
[2] M. Lutter and J. Peters. Combining physics and deep learning to learn continuous-time dynamics models. arXiv preprint arXiv:2110.01894.
[3] El Asri, Z., Sigaud, O., & Thome, N. (2024). Physics-Informed Model and Hybrid Planning for Efficient Dyna-Style Reinforcement Learning. Reinforcement Learning Conference, 2024.
[4] Hansen N., Wang X., Su H. (2022). Temporal Difference Learning for Model Predictive Control, ICML
[5] D. Olivares, P. Fournier, P. Vasishta, J. Marzat (2024). Model-Free versus Model-Based Reinforcement Learning for Fixed-Wing UAV Attitude Control Under Varying Wind Conditions, ICINCO
[6] Chua, Kurtland, et al. "Deep reinforcement learning in a handful of trials using probabilistic dynamics models." Advances in neural information processing systems 31 (2018).
[7] https://jsbsim.sourceforge.net/
[8] https://github.com/GongXudong/fly-craft
[9] Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. Advances in neural information processing systems, 31.
[10] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., ... & Zaremba, W. (2017). Hindsight experience replay. Advances in neural information processing systems, 30.
[11] Chaffre, T., Santos, P. E., Le Chenadec, G., Chauveau, E., Sammut, K., & Clement, B. (2023). Learning adaptive control of a UUV using a bio-inspired experience replay mechanism. IEEE Access, 11, 123505-123518.
[12] Chen, Ricky TQ, et al. "Neural ordinary differential equations." Advances in neural information processing systems 31 (2018).
[13] Ni, Tianwei, Benjamin Eysenbach, and Ruslan Salakhutdinov. "Recurrent model-free rl can be a strong baseline for many pomdps." arXiv preprint arXiv:2110.05038 (2021).
[14] Serris, Olivier, Stéphane Doncieux, and Olivier Sigaud. "A tale of two goals: leveraging sequentiality in multi-goal scenarios." arXiv preprint arXiv:2503.21677 (2025).
[15] Lo, C., Roice, K., Panahi, P. M., Jordan, S. M., White, A., Mihucz, G., ... & White, M. (2024). Goal-space planning with subgoal models. *Journal of Machine Learning Research*, *25*(330), 1-57.

Co-encadrement : Olivier Sigaud, ISIR MLIA, Sorbonne Université, Paris

Laboratoire d'accueil à l'ONERA
Département : Traitement de l'information et Systèmes
Lieu (centre ONERA) : Palaiseau
Contact : Pierre Fournier, Julien Marzat
Tél. : 01 80 38 65 73 Email : pierre.fournier@onera.fr

Directeur de thèse
Nom : Julien Marzat
Laboratoire : ONERA DTIS
Tél. : 01 80 38 66 50
Email : julien.marzat@onera.fr

Pour plus d'informations : https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche