INSTITUT DES SYSTEMES INTELLIGENTS ET DE ROBOTIQUE
OFFRE DE THÈSE

Sujet de thèse/*Thesis topic*

Titre de la thèse : Apprentissage par renforcement frugal pour le contrôle de l'exosquelette Atalante X
Thesis title: Frugal reinforcement learning for the control of the Atalante X exoskeleton

Directrice ou directeur de thèse (*PhD supervisor*) : Nicolas Perrin-Gilbert

Laboratoire d'accueil (*Laboratory*) : ISIR (*Institut des Systèmes Intelligents et de Robotique*), Campus Pierre et Marie Curie, 4 place Jussieu, 75005 Paris.

Personne à contacter/*Person to contact*

Nicolas Perrin-Gilbert

perrin@isir.upmc.fr

Envoyer votre candidature par mail, avec [Apprentissage par renforcement frugal pour le contrôle de l'exosquelette Atalante X] en objet, un CV et une lettre de motivation.
Send your application by e-mail, with [Frugal reinforcement learning for the control of the Atalante X exoskeleton] in the subject line, a CV and a covering letter.

Date limite de dépôt de la candidature : 06/06/2025
Application deadline : June 6, 2025

Le début de la thèse est prévu pour septembre ou octobre 2025.
The thesis is scheduled to start in September or October 2025.

Description du sujet (en français)

Contexte :
L'usage des exosquelettes pour assister la marche des personnes atteintes de handicap est une solution prometteuse pour améliorer leur qualité de vie. Toutefois, pour que ces dispositifs soient efficaces et acceptés par les utilisateurs, il est crucial qu'ils soient personnalisés en fonction des besoins et des capacités de chaque individu. Dans cette perspective, l'apprentissage par renforcement peut jouer un rôle important en permettant l'adaptation de la commande de l'exosquelette aux spécificités de chaque utilisateur, ceci afin d'obtenir des mouvements fluides, réactifs, et n'interférant pas avec les mouvements de l'utilisateur. En effet, l'apprentissage par renforcement permet au système contrôlé d'affiner progressivement ses décisions en interagissant avec son environnement, de s'adapter à des conditions d'utilisation imprévues et de s'améliorer au fil du temps. Cependant, son efficacité concrète peut être limitée par le besoin de collecter un grand nombre de données avant d'obtenir une politique de commande performante.

Objectif scientifique et description du projet :
L'objectif de cette thèse sera d'étudier et de mettre en œuvre une approche frugale basée sur des *a priori robotiques* qui augmenteront l'efficacité de l'apprentissage et permettront une personnalisation rapide de la commande du système, c'est-à-dire au plus en quelques heures d'utilisation.

Ces a priori robotiques sont un ensemble de biais prédéfinis provenant d'une connaissance experte de la tâche de commande robotique à résoudre. Ils permettront de guider l'agent dans son exploration de l'espace des actions et de réduire la durée des interactions nécessaires à l'obtention d'une politique de commande performante et adaptée. Au moins deux types d'a priori seront considérés et serviront de point de départ aux travaux menés :

- Le premier type d'a priori est temporel, il se base sur la séquentialité des mouvements. En effet, pour les exosquelettes d'assistance à la marche, les mouvements peuvent être décomposés en plusieurs phases connues à l'avance, et synchronisées au mouvements de marche de l'utilisateur. La connaissance de ces phases permet de guider l'apprentissage vers des solutions acceptables, et résout en partie le "credit assignment problem" propre à l'apprentissage par renforcement, qui désigne la difficulté d'interpréter des signaux de récompense lorsqu'ils sont une conséquence non immédiate des actions.
Cet axe s'appuiera sur des travaux précédents [1] basés sur une approche de type "diviser pour régner", qui ont permis d'atteindre une efficacité d'apprentissage nettement supérieur à l'état de l'art pour des tâches de locomotion bipède en simulation à partir de démonstrations uniques.

- Le second type d'a priori est géométrique, il découle de la connaissance de l'espace de travail tridimensionnel ainsi que de certaines mesures clés. Par exemple, avec une centrale inertielle, il est possible d'avoir une mesure précise et réactive de l'orientation par rapport à la verticale, qui est une donnée cruciale dans les tâches de contrôle de l'équilibre. En combinant cette information à d'autres mesures et caractéristiques du système (proprioception et dimensions, cinématique directe et inverse), l'objectif sera de structurer et pondérer les observations utilisées par l'agent et définir des signaux de récompenses annexes qui guideront progressivement l'apprentissage vers la réussite de la tâche globale. Un curriculum sera défini, c'est-à-dire une séquence de sous-tâches de difficulté croissante qui mèneront à un ajustement par étapes des paramètres de la politique de commande. Il s'agira par exemple de traiter en premier l'équilibre statique, puis le mouvement de l'exosquelette pour un pas isolé, puis les enchaînements de pas, en veillant à chaque étape de ne pas diminuer les performances de la commande sur les sous-tâches précédentes.

Ces approches accéléreront l'apprentissage en ligne, mais le point de départ sera toujours une politique apprise en simulation grâce à un modèle dynamique relativement fidèle de l'exosquelette. Cet apprentissage préalable suivra l'approche bien connue de "domain randomization" et fournira une politique de commande initiale qui sera raffinée par l'apprentissage par renforcement en ligne. Pour pouvoir mettre en œuvre cette approche basée sur l'apprentissage en ligne, nous veillerons à définir très précisément des contraintes cinématiques et dynamiques qui assureront la sécurité des expériences.

Enfin, L'approche proposée sera spécifique à la commande d'exosquelettes d'assistance à la marche, mais elle devra également être générique, au moins dans une certaine mesure. Afin de vérifier que la méthode proposée n'est pas adaptée à un système unique, nous envisagerons de tester l'approche sur d'autres dispositifs plus simples qu'un exosquelette de marche.

[1] Chenu, A., Serris, O., Sigaud, O., & Perrin-Gilbert, N. (2022). Leveraging sequentiality in reinforcement learning from a single demonstration. arXiv preprint arXiv:2211.04786.

Profil recherché : Nous recherchons un(e) candidat(e) titulaire d'un Master 2, avec une solide expérience en développement Python et de bonnes connaissances en apprentissage par renforcement profond. Un fort intérêt pour l'expérimentation sur systèmes robotiques réels est indispensable, une expérience concrète dans ce domaine étant un atout majeur. Autonome, persévérant(e) et rigoureux(se), le/la candidat(e) devra également avoir un goût prononcé pour l'ingénierie et le travail pratique.

Compétences requises :

Compétences techniques :
- Développement Python : maîtrise avancée de Python, avec une bonne structuration du code (POO, modularité, tests, versioning).
- Apprentissage par renforcement profond (Deep RL) : connaissance des algorithmes classiques (DQN, PPO, SAC...), de leurs principes théoriques et de leur mise en œuvre pratique.
- Frameworks de machine learning : expérience avec des bibliothèques comme PyTorch, TensorFlow ou JAX.
- Manipulation de systèmes robotiques (*idéalement*) : expérience en contrôle de robots réels (bras robotisés, robots mobiles, etc.)
- RL et simulation robotique : connaissances en outils comme Mujoco et Gymnasium.

Compétences comportementales :
- Autonomie : capacité à avancer seul(e), à identifier les blocages et à chercher des solutions proactivement.
- Persévérance : goût pour les défis techniques complexes et patience face aux expérimentations longues et parfois instables sur robots réels.
- Sens de l'ingénierie : goût pour le prototypage, le debugging hardware/software, la mise au point de systèmes robustes.
- Capacité à documenter et partager ses résultats et ses approches.

Description du sujet (en anglais)

Context:
The use of exoskeletons to assist walking in individuals with disabilities is a promising solution for improving their quality of life. However, for these devices to be effective and accepted by users, it is crucial that they are personalized to each individual's needs and capabilities. In this context, reinforcement learning can play an important role by enabling the exoskeleton control to adapt to the specific characteristics of each user, with the goal of achieving smooth, responsive movements that do not interfere with the user's own actions. Indeed, reinforcement learning allows the controlled system to gradually refine its decisions through interaction with the environment, adapt to unforeseen usage conditions, and improve over time. However, its practical effectiveness may be limited by the need to collect a large amount of data before obtaining a high-performing control policy.

Scientific Objective and Project Description:
The objective of this PhD thesis will be to study and implement a frugal approach based on robotic priors that increase learning efficiency and enable rapid personalization of the system's control policy — ideally within just a few hours of use.

These robotic priors refer to a set of predefined biases derived from expert knowledge of the robotic control task at hand. They will help guide the agent's exploration of the action space and reduce the number of interactions required to obtain an effective and personalized control policy. At least two types of priors will be considered as starting points for this work:

- The first type of prior is temporal, based on the sequential nature of movements. For example, in the case of exoskeletons assisting gait, movements can be decomposed into several predefined phases that are synchronized with the user's walking pattern. Knowing these phases helps guide the learning process toward acceptable solutions and partially addresses the credit assignment problem (i.e. the difficulty of attributing delayed rewards to the correct actions).
This direction will build upon previous work [1] using a "divide and conquer" approach, which has led to significantly improved learning efficiency over the state of the art in simulated biped locomotion tasks from single demonstrations.

- The second type of prior is geometric, based on knowledge of the three-dimensional workspace and certain key measurements. For instance, using an inertial measurement unit (IMU), one can obtain accurate and responsive data about orientation relative to the vertical axis — a crucial variable in balance control. By combining this information with other sensory and system characteristics (proprioception, dimensions, forward and inverse kinematics), the goal will be to structure and weight the agent's observations and define auxiliary reward signals that progressively guide learning toward successful task completion. A curriculum will be defined — that is, a sequence of subtasks of increasing difficulty — to incrementally adjust the control policy parameters. For example, the curriculum may begin with static balance, proceed to executing a single step with the exoskeleton, and then to chaining multiple steps, ensuring at each stage that performance on earlier subtasks is not degraded.

These approaches will accelerate online learning, but the starting point will always be a policy pretrained in simulation, using a reasonably accurate dynamic model of the exoskeleton. This preliminary training will follow the well-known domain randomization approach and produce an initial control policy, which will then be refined through online reinforcement learning. To safely implement this online learning process, precise kinematic and dynamic constraints will be defined to ensure the safety of all experiments.

Finally, while the proposed approach will be specifically designed for controlling exoskeletons for gait assistance, it should also aim to be generalizable, at least to some extent. To verify that the method is not overly tailored to a single system, we will consider testing it on other devices that are simpler than a full walking exoskeleton.

[1] Chenu, A., Serris, O., Sigaud, O., & Perrin-Gilbert, N. (2022). Leveraging sequentiality in reinforcement learning from a single demonstration. arXiv preprint arXiv:2211.04786.

Required Profile: We are looking for a candidate with a Master's degree (Master 2) and solid experience in Python development, along with good knowledge of deep reinforcement learning. A strong interest in experimentation on real robotic systems is essential, and hands-on experience in this area is a major asset. The ideal candidate will be autonomous, persevering, and rigorous, with a strong interest in engineering and practical work.

Required skills:

Technical Skills:
- Python Development: Advanced proficiency in Python with well-structured code practices (OOP, modularity, testing, version control).
- Deep Reinforcement Learning (Deep RL): Familiarity with standard algorithms (DQN, PPO, SAC…), their theoretical foundations, and practical implementation.
- Machine Learning Frameworks: Experience with libraries such as PyTorch, TensorFlow, or JAX.
- Robotic System Handling (ideally): Experience in controlling real robotic platforms (robotic arms, mobile robots, etc.).

- RL and Robotic Simulation: Knowledge of tools such as Mujoco and Gymnasium.

Soft Skills:
- Autonomy: Ability to work independently, identify challenges, and proactively seek solutions.
- Perseverance: Enthusiasm for tackling complex technical problems and patience in the face of long, sometimes unstable, real-world robotic experiments.
- Engineering Mindset: Interest in prototyping, debugging (both hardware and software), and building robust systems.
- Communication: Ability to document and share results and methodologies effectively.