

## Soutenance de thèse

**Erwan LECARPENTIER** soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral ISAE-ONERA CSDV et intitulée «*Apprentissage par renforcement en environnement non stationnaire*»

**Le 06 juillet 2020 à 10h00, salle des thèses ISAE-SUPAERO**

devant le jury composé de

M. Guillaume INFANTES	Ingénieur JoliBrain Toulouse	Directeur de thèse
M. Bruno ZANUTTINI	Professeur Université de Caen	Rapporteur
M. Olivier BUFFET	Chargé de Recherche INRIA Nancy	Rapporteur
M. Emmanuel RACHELSON	Professeur ISAE-SUPAERO	Co-directeur de thèse
Mme Émilie KAUFMANN	Chargé de Recherche INRIA Lille	
M. Régis SABBADIN	Directeur de Recherche INRA Toulouse	
M. Tristan CAZENAVE	Professeur Université Paris-Dauphine	
M. Aurélien GARIVIER	Professeur ENS Lyon	

### Résumé :

Comment un agent doit-il agir étant donné l'incertitude qu'il a sur l'évolution de son environnement? Dans cette thèse, nous adoptons la perspective de l'apprentissage par renforcement pour fournir une réponse à cette question. Le problème est vu sous trois aspects différents. Premièrement, nous étudions le compromis planification vs. re-planification des algorithmes de recherche arborescente dans les Processus Décisionnels Markoviens stationnaires. Nous proposons une méthode pour réduire la complexité de calcul d'un tel algorithme, tout en conservant des garanties théoriques sur la performance. Deuxièmement, nous étudions le cas des environnements évoluant graduellement au cours du temps. Cette hypothèse est formulée dans un cadre mathématique appelé Processus de Décision Markoviens Non-Stationnaires Lipschitziens. Dans ce cadre, nous proposons un algorithme de planification robuste aux évolutions possibles. Troisièmement, nous considérons le cas de l'évolution temporelle abrupte dans le cadre du "lifelong learning" (apprentissage tout au long de la vie). Nous proposons une méthode de transfert non-négatif basée sur l'étude théorique de la continuité de Lipschitz de la Q-fonction optimale par rapport à l'espace des tâches. L'approche permet d'accélérer l'apprentissage dans de nouvelles tâches. Dans l'ensemble, cette dissertation propose des réponses à la question de la résolution des Processus de Décision Markoviens Non-Stationnaires dans trois cadres d'hypothèses.

**Mots-clés :** Apprentissage par Renforcement, Planification, Apprentissage tout au long de la vie, Processus Décisionnel de Markov, Processus Décisionnel de Markov Non Stationnaire, Apprentissage par Renforcement Basé Modèle

### Summary:

How should an agent act in the face of uncertainty on the evolution of its environment? In this thesis, I give a Reinforcement Learning perspective on the resolution of dynamical problems, i.e., problems changing through time. Three answers, relying on different sets of hypothesis, are proposed on that

topic. First, by considering stationary environments, I propose an approach for computing an optimal behaviour with a controlled probability, taking advantage of the fact that we know a stationary model of the world. Secondly, I study the case where the environment evolves smoothly, but in an unknown manner. In that setting, a robust approach to the possible evolutions is proposed. Thirdly, non-smooth evolution is considered in a setting best known as Lifelong Reinforcement Learning. I propose a knowledge transfer method based on the theoretical study of the continuity of a certain objective function in the space of tasks. The approach allows to speed-up the learning in new encountered tasks.

**Keywords:** Reinforcement Learning, Planning, Lifelong Learning, Markov Decision Process, Non-Stationary Markov Decision Process, Model-Based Reinforcement Learning

