Workshop KCL-UP *Mean-field Reinforcement learning*

14 and 15 October 2021
Online workshop

Program and abstracts

Schedule given in **London time (UTC +1)**

Thursday October, 14th

Morning session *(Chair: Huyên Pham)*

09h00 – 09h05 : **Welcome**

09h05 – 09h45 : **Lukasz Szpruch**, Edinburgh University and The Alan Turing Institute, UK
Gradient Flows for Regularized Stochastic Control Problems
09h50 – 10h30 : **Matthieu Geist**, Google Brain, France
Concave Utility Reinforcement Learning: the Mean-field Game viewpoint
10h30 – 10h50: Break and virtual rooms
10h50 – 11h30: **Justin Sirignano**, Oxford University, UK
Laws of Large Numbers for Neural Networks

Young researchers (Chair: Mathieu Laurière)

11h35 – 12h00: **Médéric Motte**, Université de Paris, France
Online click prediction learning algorithm for targeted advertising
12h05 – 12h30: **Haoyang Cao**, The Alan Turing Institute, UK
Identifiability in inverse reinforcement learning

12h30 – 14h00 : Break and virtual rooms

Afternoon session *(Chair: Jean-Francois Chassagneux)*

14h00 – 14h40 : **Olivier Guéant**, Université Paris Panthon Sorbonne, France
Market making algorithms on OTC markets: the next success of RL?
14h45 – 15h25 : **Christoph Reisinger**, Oxford University, UK
A fast iterative PDE-based algorithm for feedback controls of nonsmooth mean-field control problems
15h25 – 15h45 : Break and virtual rooms
15h45 – 16h25: **Mathieu Laurière**, Google Brain, France
Mean field MDP and mean field RL
16h25 – 18h00: Virtual discussions

Friday October, 15th

Morning session *(Chair: Roxana Dumitrescu)*

09h00 – 09h40 : **Nizar Touzi**, Ecole Polytechnique, CMAP, France
Mean field game of mutual holding
09h45 – 10h25 : **Samuel Cohen**, Oxford University and The Alan Turing Institute, UK
Optimal Control with Online Learning
10h25 – 10h45: Break and virtual rooms
10h45 – 11h25: **Peter Tankov**, ENSAE, France
A mean-field game of energy transition

Young researchers (Chair: Samuel Cohen)

11h30 – 11h55: **Xiaoli Wei**, Tsinghua Berkeley Shenzen Institute
Multi-agent reinforcement learning: a mean field perspective
12h – 12h25: **Sarah Perrin**, Université de Lille, France
Generalization in Mean Field Games by Learning Master Policies

12h25 – 14h00 : Break and virtual rooms

Afternoon session *(Chair: Peter Tankov)*

14h00 – 14h40 : **Clémence Alasseur**, EDF, France - *MFG model with a long-lived penalty at random jump times:application to demand side management for electricity contracts*
14h45 – 15h25 : **Davide Pigoli**, King's College London, UK - *Dynamic reconstruction of growth curves in forensic entomology.*

Abstracts

Clémence Alasseur (EDF, France)

Title: *MFG model with a long-lived penalty at random jump times:application to demand side management for electricity contracts*

Abstract: We consider an energy system with n consumers who are linked by a Demand Side Management (DSM) contract, i.e. they agreed to diminish, at random times, their aggregated power consumption by a predefined volume during a predefined duration. Their failure to deliver the service is penalised via the difference between the sum of the n power consumptions and the contracted target. We are led to analyse a non-zero sum stochastic game with n players, where the interaction takes place through a cost which involves a delay induced by the duration included in the DSM contract. When $n \to \infty$, we obtain a Mean-Field Game (MFG) with random jump time penalty and interaction on the control. We prove a stochastic maximum principle in this context, which allows to compare the MFG solution to the optimal strategy of a central planner. In a linear quadratic setting we obtain an semi-explicit solution through a system of decoupled forward-backward stochastic differential equations with jumps, involving a Riccati Backward SDE with jumps. We show that it provides an approximate Nash equilibrium for the original n-player game for n large. Finally, we propose a numerical algorithm to compute the MFG equilibrium and present several numerical experiments. This is a joint work with Luciano Campi, Roxana Dumitrescu and Jia Zeng.

Haoyang Cao (The Alan Turing Institute, UK)

Title: *Identifiability in inverse reinforcement learning*

Abstract: In this talk, we will focus on a class of stochastic inverse optimal control problems with entropy regularization. We first characterize the set of solutions for the inverse control problem. This solution set exemplifies the issue of degeneracy in generic inverse control problems that there exist multiple reward or cost functions that can explain the displayed optimal policy. Then we establish one resolution for the degeneracy issue by providing one additional optimal policy under a different discount factor. This resolution does not depend on any prior knowledge of the solution set. Through a simple numerical experiment with deterministic transition kernel, we demonstrate the ability of accurately extracting the cost function through our proposed resolution. This is a joint with Sam Cohen (Oxford) and Lukasz Szpruch (Edinburgh).

Samuel Cohen (Oxford University and The Alan Turing Institute, UK)

Title: *Optimal Control with Online Learning*

Abstract: If one takes a Bayesian view, optimal control with model uncertainty can be theoretically reduced to classical optimal control. The key difficulty is that the state space for the control problem is typically very large, leading to numerically intractable problems. In this talk, we will see that this view is nevertheless productive, as one can then exploit asymptotic expansions for the control problem to yield a computationally efficient and flexible algorithm, which performs well in practice. We will consider applications of this approach to multiarmed bandit problems, which include controlled learning as a key part of the optimal control problem, and can be used as models

of interesting problems in finance.

Matthieu Geist (Google Brain, France)
Title: *Concave Utility Reinforcement Learning: the Mean-field Game viewpoint*
Abstract: Concave Utility Reinforcement Learning (CURL) extends RL from linear to concave utilities in the occupancy measure induced by the agent's policy. This encompasses not only RL but also imitation learning and exploration, among others. Yet, this more general paradigm invalidates the classical Bellman equations, and calls for new algorithms. Mean-field Games (MFGs) are a continuous approximation of many-agent RL. They consider the limit case of a continuous distribution of identical agents, anonymous with symmetric interests, and reduce the problem to the study of a single representative agent in interaction with the full population. Our core contribution consists in showing that CURL is a subclass of MFGs. We think this important to bridge together both communities. It also allows to shed light on aspects of both fields: we show the equivalence between concavity in CURL and monotonicity in the associated MFG, between optimality conditions in CURL and Nash equilibrium in MFG, or that Fictitious Play (FP) for this class of MFGs is simply Frank-Wolfe, bringing the first convergence rate for discrete-time FP for MFGs. We also experimentally demonstrate that, using algorithms recently introduced for solving MFGs, we can address the CURL problem more efficiently.

Olivier Guéant (Université Paris Panthon Sorbonne, France)
Title: *Market making algorithms on OTC markets: the next success of RL?*
Abstract: In this talk, I will present the problem faced by market makers on OTC markets and the models that have been proposed to tackle this problem. In particular, I will discuss how the stochastic optimal control models introduced in the literature can be adapted to allow for the use of reinforcement learning techniques. The talk will be based on a series a paper by the authors on market making issues. In particular, we shall discuss "Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality" (2019) and "Reinforcement Learning Methods in Algorithmic Trading" (2021).

Mathieu Laurière (Google Brain, France)
Title: *Mean field MDP and mean field RL*
Abstract: Multi-agent reinforcement learning has attracted a lot of interest in the past decades. However, most existing methods do not scale well with the number of agents. In this talk, we study a limiting case in which there is an infinite population of cooperative agents. A mean field approach allows us to reduce the complexity of the problem and propose efficient learning methods. The problem is first phrased as a discrete time mean field control (MFC) problem. The model includes not only individual noise and individual action randomization at the agent level, but also common noise and common randomization at the population level. We relate this MFC problem to a lifted Mean Field Markov Decision Process (MFMDP), in which the state is the population distribution and for which we prove a dynamic programming principle. This allows us to connect

closed-loop and open-loop controls for the original MFC problem. Building on this MFMDP, we propose two reinforcement learning (RL) methods: one based on tabular Q-learning, for which convergence can be proved, and one based on deep RL. Several numerical examples are provided, in discrete and continuous spaces. This is joint work with Rene Carmona and Zongjun Tan.

Médéric Motte (Université de Paris, France)

Title: *Online click prediction learning algorithm for targeted advertising*

Abstract: We introduce and study an online click prediction learning algorithm for targeted advertising. The algorithm is based on a polynomial classifier with soft or hard margin, and, to learn, only requires to observe clicks on displayed ads. We show that all the ads that would lead to a click will be displayed, and that the expected number of non-clicked displayed ads is logarithmic in the total number of ads. In classification terms, this is to say that our algorithm makes no false negative and only makes a logarithmic amount of false positives in the number of stages. We conclude by discussing the boundedness of the average memory usage and computational complexity.

Sarah Perrin (Université de Lille, France)

Title: *Generalization in Mean Field Games by Learning Master Policies*

Abstract: Mean Field Games (MFGs) can potentially scale multi-agent systems to extremely large populations of agents. Yet, most of the literature assumes a single initial distribution for the agents, which limits the practical applications of MFGs. Machine Learning has the potential to solve a wider diversity of MFG problems thanks to generalizations capacities. We study how to leverage these generalization properties to learn policies enabling a typical agent to behave optimally against any population distribution. In reference to the Master equation in MFGs, we coin the term "Master policies" to describe them and we prove that a single Master policy provides a Nash equilibrium, whatever the initial distribution. We propose a method to learn such Master policies. Our approach relies on three ingredients: adding the current population distribution as part of the observation, approximating Master policies with neural networks, and training via Reinforcement Learning and Fictitious Play. We illustrate on numerical examples not only the efficiency of the learned Master policy but also its generalization capabilities beyond the distributions used for training.

Davide Pigoli (King's College London, UK)

Title: *Dynamic reconstruction of growth curves in forensic entomology*

Abstract: Larvae (or maggots) collected at crime scenes contribute important pieces of information to police investigations. Their hatching time provides a lower bound for the post-mortem interval, i.e. the interval between death and the discovery of the body. A functional data analysis approach is described here to model the local growth rate from the experimental data on larval development, where larvae have been exposed to a small number of constant temperature profiles. This allows us to reconstruct varying temperature growth profiles and use them to estimate the hatching time for a sample of larvae from the crime scene.

Christoph Reisinger (Oxford University, UK)
Title: *A fast iterative PDE-based algorithm for feedback controls of nonsmooth mean-field control problems*
Abstract: A PDE-based accelerated gradient algorithm is proposed to seek optimal feedback controls of McKean-Vlasov dynamics subject to nonsmooth costs, whose coefficients involve mean-field interactions both on the state and action. It exploits a forward-backward splitting approach and iteratively refines the approximate controls based on the gradients of smooth costs, the proximal maps of nonsmooth costs, and dynamically updated momentum parameters. At each step, the state dynamics is realized via a particle approximation, and the required gradient is evaluated through a coupled system of nonlocal linear PDEs. The latter is solved by finite difference approximation or neural network-based residual approximation, depending on the state dimension. Exhaustive numerical experiments for low and high-dimensional mean-field control problems, including sparse stabilization of stochastic Cucker-Smale models, are presented, which reveal that our algorithm captures important structures of the optimal feedback control, and achieves a robust performance with respect to parameter perturbation.

Justin Sirignano (Oxford University, UK)
Title: *Laws of Large Numbers for Neural Networks*
Abstract: The asymptotics of neural networks can be studied as the number of hidden units become large. Two different limits can be established, depending upon the choice of normalization for the network. The "mean-field limit" satisfies a PDE while the "kernel limit" satisfies an ODE. The differences between these two types of limits will be discussed. Then, we will prove a kernel limit for the Q-learning algorithm in reinforcement learning where the function approximator is a neural network. This is commonly called a "Q-network". The limit satisfies a nonlinear ODE whose stationary solution is the optimal policy. Under a certain condition, we can also prove convergence to the optimal policy.

Lukasz Szpruch (Edinburgh University and The Alan Turing Institute, UK)
Title: *Gradient Flows for Regularized Stochastic Control Problems*
Abstract: We study stochastic control problems regularized by the relative entropy, where the action space is the space of measures. This setting includes relaxed control problems, problems of finding Markovian controls with the control function replaced by an idealized infinitely wide neural network and can be extended to the search for causal optimal transport maps. By exploiting the Pontryagin optimality principle, we identify suitable metric space on which we construct gradient flow for the measure-valued control process along which the cost functional is guaranteed to decrease. It is shown that under appropriate conditions, this gradient flow has an invariant measure which is the optimal control for the regularized stochastic control problem. If the problem we work with is sufficiently convex, the gradient flow converges exponentially fast. Furthermore, the optimal measured valued control admits Bayesian interpretation which means that one can incor-

porate prior knowledge when solving stochastic control problem. This work is motivated by a desire to extend the theoretical underpinning for the convergence of stochastic gradient type algorithms widely used in the reinforcement learning community to solve control problems.

Peter Tankov (ENSAE, France)
Title: *A mean-field game of energy transition*
Abstract: We develop a model for the industry dynamics in the electricity market, based on mean-field games of optimal stopping. In our model, there are several types of agents representing various electricity production technologies. The renewable producers choose the optimal moment to build new renewable plants, and the conventional producers can both build new plants and exit the market. The agents interact through the market price, determined by matching the aggregate supply of all producers with an exogenous demand function. Using a relaxed formulation of optimal stopping mean-field games, we prove the existence of a Nash equilibrium and the uniqueness of the equilibrium price process. An empirical example, inspired by the German electricity market is presented. Based on joint works with René Aid and Roxana Dumitrescu.

Nizar Touzi (Ecole Polytechnique, CMAP, France)
Title: *Mean field game of mutual holding*
Abstract: We introduce a mean field model for optimal holding of a representative agent of her peers as a natural expected scaling limit from the corresponding $N-$agent model. The induced mean field dynamics appear naturally in a form which is not covered by standard McKean-Vlasov stochastic differential equations. We study the corresponding mean field game of mutual holding in the absence of common noise. Our main result provides existence of an explicit equilibrium of this mean field game, defined by a bang-bang control consisting in holding those competitors with positive drift coefficient of their dynamic value. Our analysis requires to prove an existence result for our new class of mean field SDE with the additional feature that the diffusion coefficient is irregular.

Xiaoli Wei (Tsinghua Berkeley Shenzen Institute)
Title: *Multi-agent reinforcement learning: a mean field perspective*
Abstract: Multi-agent reinforcement learning (MARL), despite its popularity and empirical success, suffers from the curse of dimensionality. This paper builds the mathematical framework to approximate cooperative MARL by a mean-field control (MFC) approach, and shows that the approximation error is of $O(1/\sqrt{N})$. By establishing an appropriate form of the dynamic programming principle for both the value function and the Q function, it proposes a model-free kernel-based Q-learning algorithm (MFC-K-Q), which is shown to have a linear convergence rate for the MFC problem, the first of its kind in the MARL literature. It further establishes that the convergence rate and the sample complexity of MFC-K-Q are independent of the number of agents N, which provides an $O(1/\sqrt{N})$ approximation to the MARL problem with N agents in the learning environment. If time allows, we will also discuss decentralized MARL with the network of states, with homogeneous (a.k.a. mean-field type) agents, frequently used for modeling self-driving vehicles,

ride-sharing, and data and traffic routing. The key idea is to utilize the homogeneity of agents and regroup them according to their states, thus the formulation of a networked Markov decision process with teams of agents, enabling the update of the Q-function in a localized fashion. In order to design an efficient and scalable reinforcement learning algorithm under such a framework, we adopt the actor-critic approach with over-parameterized neural networks, and establish the convergence and sample complexity for our algorithm, shown to be scalable with respect to the size of both agents and states.