



**Laboratoire de Mathématiques et Informatique pour la Complexité et les
Systèmes
MICS**

Présente

L'AVIS DE SOUTENANCE

De Monsieur Vincent Ragel

Laboratoire MICS, CentraleSupélec, Université Paris Saclay, soutiendra
publiquement ses travaux de thèse de doctorat intitulés :

“Reinforcement Learning for systematic market making strategies”

Sous la Direction de Monsieur Damien Challet.

Le vendredi 6 décembre 2024 à 15h00

À l'école CentraleSupélec, dans la **salle E.068** - Bâtiment Bouygues.

Membres du jury :

Olivier GUEANT, Rapporteur & Examineur, Université Paris 1 Panthéon-Sorbonne

Eyal NEUMANN, Rapporteur & Examineur, Imperial College London

Caroline HILLAIRET, Examinatrice, ENSAE Paris

Mathieu GARCIN, Examineur, ESILV

Résumé :

Les teneurs de marché proposent simultanément des prix d'achat et de vente pour certains actifs. Les stratégies les plus simples, souvent fondées sur des règles *ad hoc*, sont aujourd'hui remises en question par les avancées dans le domaine du contrôle stochastique, qui repose sur des hypothèses fortes susceptibles de ne pas refléter la complexité des marchés financiers réels, ce qui peut entraîner des stratégies sous-optimales et des pertes financières. Dans le secteur hautement compétitif des teneurs de marché, il existe une forte incitation à intégrer des méthodes issues de l'intelligence artificielle (IA). L'apprentissage par renforcement (RL) est

particulièrement efficace pour développer des stratégies optimales grâce à une interaction répétée d'un agent avec son environnement, mais son entraînement en situation réelle est coûteux et risqué. Par conséquent, les agents RL sont souvent entraînés à l'aide de données historiques ou synthétiques.

Cette thèse aborde certains des défis rencontrés lors de l'entraînement d'un agent RL teneur de marché. Elle propose d'abord d'utiliser des nouvelles mesures d'état du carnet d'ordre couplées à de l'apprentissage profond pour améliorer la prédiction de la probabilité d'exécution des ordres limites, ce qui conduit mécaniquement à des gains plus importants pour un teneur de marché. Une deuxième partie propose un moyen d'utiliser les données historiques de manière plus cohérente lors de l'entraînement d'un agent RL, appelé voyage temporel dans les données. Cette méthode permet de simuler de trajectoires réalistes, qui tiennent mieux en compte l'impact des actions de l'agent sur la dynamique du marché. Enfin, nous proposons un moyen d'inclure les dépendances à plus long terme dans les réseaux neuronaux récurrents (RNN) en intégrant au moins deux échelles de temps dans un seul état caché, tant pour les processus neuronaux de Hawkes que les réseaux de neurones de type Long Short-Term Memory (LSTM).

Abstract :

Market makers simultaneously offer buy and sell quotes for a variety of assets. The simplest strategies, often based on simple *ad hoc* rules, are now being challenged by advances in stochastic control, which relies on strong assumptions that may not reflect the complexities of real financial markets, potentially resulting in suboptimal strategies and financial losses. In the highly competitive market-making field, there is a pressure to use methods from artificial intelligence (AI). Reinforcement learning (RL) is particularly effective in developing optimal strategies through environmental interaction but is expensive and risky to train in real life. Therefore, RL agents are often trained with either historical or synthetic data.

This thesis tackles some of the challenges encountered when training a RL agent for market-making. It first leverages new features and deep learning to improve the prediction of the order execution probability, which mechanically yields larger gains for a market maker. A second part proposes a way to use historical data in a more consistent way when training a RL agent, called data time travel, which enables realistic trajectory simulations that are consistent with the impact of the agent's actions on the market dynamics. Finally, we propose to a way to include longer term dependencies in recurrent neural networks by including at least two time scales in a single hidden state, both for Neural Hawkes Processes and vanilla Long Short-term Memory neural networks.