# Résumé du stage : Optimisation distribuée pour la recharge de véhicules électriques

Xinyu HUANG
Master in Artificial intelligence, Sorbonne University

## 1 Mots clés

Frank&Wolfe algorithme, problème de recharge intelligente, optimisation convexe

## 2 Introduction

This project aims to reproduce and compare the simulations done in the paper *Prefrontal cortex as a meta-reinforcement learning system*, Below is a short summary of work done, including:

- The reading and learning process of extracurricular, but required theories and tools
- The implementation of the model and the simulations
- The analysis of the results obtained and the understanding of the limitations of the meta-model

## 3 Etat de l'art

## 4 Développement des algorithmes

Our work is based on the code of a previous student who build a model to test the Two Step Task. However, as TensorFlow has been updated from version 1 to version 2 and the two have significant differences, the code written in tf1 differs from the tutorials found online and therefore is hard to understand and modify. We had decided to split into two group: one adapted the model to another simulations, the other built their own model in tf2 to reproduce the Two Step Task.

Nevertheless, the two versions of TensorFlow do share the same principles of implementations as below:

## 5 Développement des algorithmes

## 6 Formulation des différents problèmes

## 7 Évaluations numériques

## 8 Conclusion

## 9 Reference

[1] J. X. Wang et al., "Learning to reinforcement learn," ArXiv, vol. abs/1611.05763, 2016.

[2] J. X. Wang et al., "Prefrontal cortex as a meta-reinforcement learning system," Nature Neuroscience, vol. 21, no. 6, pp. 860–868, Jun. 2018, doi: 10.1038/s41593-018-0147-8.

[3] K.-I. Tsutsui, F. Grabenhorst, S. Kobayashi, and W. Schultz, "A dynamic code for economic object valuation in prefrontal cortex neurons," Nature Communications, vol. 7, no. 1, p. 12554, Sep. 2016, doi: 10.1038/ncomms12554.

[4] C. M. Stopper, M. T. L. Tse, D. R. Montes, C. R. Wiedman, and S. B. Floresco, "Overriding Phasic Dopamine Signals Redirects Action Selection during Risk/Reward Decision Making," Neuron, vol. 84, no. 1, pp. 177–189, 2014, doi: https://doi.org/10.1016/j.neuron.2014.08.033.

[5] V. Mnih et al., "Asynchronous Methods for Deep Reinforcement Learning," in Proceedings of The 33rd International Conference on Machine Learning, New York, New York, USA, Jun. 2016, vol. 48, pp. 1928–1937. [Online]. Available: https://proceedings.mlr.press/v48/mniha16.html

[6] B. Lau and P. W. Glimcher, "Dynamic response-by-response models of matching behavior in rhesus monkeys," Journal of the experimental analysis of behavior, vol. 84, no. 3, pp. 555–579, 2005.

[7] E. Bromberg-Martin, M. Matsumoto, S. Hong, and O. Hikosaka, "A Pallidus-Habenula-Dopamine Pathway Signals Inferred Stimulus Values," Journal of neurophysiology, vol. 104, pp. 1068–76, Aug. 2010, doi: 10.1152/jn.00158.2010.