Université Lille Nord de France

Ecole doctorale régionale Sciences Pour l'Ingénieur
Lille Nord-de-France - 072

Titre : ***Méthodes d'optimisation numérique pour l'intelligence artificielle***

Directeur de thèse : *Guillaume Dujardin*
E-mail : guillaume.dujardin@inria.fr

Co-directeur de thèse : *David Cohen*
E-mail : david.cohen@chalmers.se

Laboratoire : Laboratoire Paul Painlevé

Equipe : Analyse numérique et équations aux dérivées partielles

Descriptif

Le machine learning (ML) et le deep learning (DL) sont deux composantes importantes de l'intelligence artificielle (IA) [Good 16, Russ 10]**.** L'optimisation est sans doute l'un des fondements les plus importantes du ML. Dans ce contexte de ML, de nombreux problèmes conduisent à déterminer un minimum d'une fonction potentiellement non convexe [Bott 18]. C'est, par exemple, essentiellement ce qu'on cherche à faire lorsque l'on entraîne un réseau de neurones multi-couches : on cherche des paramètres minimisant une fonctionnelle de coût [High 18]. Qui plus est, plus le réseau est profond, plus la tache est numériquement difficile. En fonction du contexte applicatif, ces problèmes peuvent être déterministes comme stochastiques.

Le principal but de ce projet de thèse est d'exploiter les relations entre les systèmes dynamiques, potentiellement aléatoires et ces problématiques d'optimisation, dans l'objectif d'obtenir et d'analyser de nouvelles méthodes numériques d'optimisation pour le ML et le DL, fondées sur des outils d'intégration numérique d'équations différentielles ordinaires comme stochastiques.

Objectifs visés et résultats escomptés

Les objectifs visés sont l'introduction et l'analyse de nouvelles méthodes d'optimisation en ML, fondées sur des méthodes numériques issues de la simulation de dynamiques déterministes et stochastiques modélisées par des équations différentielles ordinaires ou stochastiques.

Concernant les méthodes d'optimisation déterministes.

Les principaux algorithmes numériques pour résoudre les problèmes de minimisation en IA sont par exemple la méthode de Newton, la méthode de descente de gradient ou la méthode de descente de gradient accélérée de Nesterov. Ces problèmes de minimisation ont souvent un lien étroit avec des dynamiques de type gradient modélisées par des équations différentielles déterministes ou stochastiques [High 18]. Par exemple, les auteurs de [Su 16] prouvent qu'une équation différentielle du second ordre peut être considérée comme la limite exacte de la méthode du gradient accéléré de Nesterov (en prenant de petites tailles de pas dans cet algorithme). Cette interprétation permet alors de mieux comprendre la méthode de Nesterov, et permet aux auteurs de développer toute une famille de schémas numériques avec le même taux de convergence que celui de Nesterov, puis de prouver la convergence d'une version de redémarrage de la méthode de Nesterov ([Su 16, Théorème 10]). Une démarche similaire est mise en œuvre dans [Wibi 16] avec une équation hamiltonienne de Bregman. L'utilisation de méthodes symplectiques pour ce type de problèmes de minimisation issus du ML devrait notamment permettre d'utiliser des pas de temps sensiblement plus grands, et ainsi d'être plus efficace, dans le calcul numérique de minimiseurs [Shi 19]. La direction est pointée dans [Shi 19], qui note que *"the literature has not yet provided a full exploration of the transition from continuous-time ODEs to discrete-time algorithms [in ML]"*.

Concernant les méthodes d'optimisation stochastiques

On sait que les méthodes traditionnelles de descente de type gradient sont très coûteuses dans le cas, fréquent, d'un grand nombre de paramètres et avec un grand nombre de points d'entraînement. Une alternative plus efficace dans ce contexte peut etre la méthode de descente dite de gradient stochastique (SGDM) [Jent 18]. Une généralisation naturelle du ce lien entre les équations différentielles et les méthodes d'optimisation déterministes décrit dans le paragraphe précédent est de connecter SGDM avec la discrétisation temporelle des équations différentielles stochastiques (EDS). Des études préliminaires ont déjà eu lieu, qui montrent par exemple que, en utilisant le lien entre les SDE et le SGDM, on permet de se ramener à des calculs plus efficaces. La connexion entre les EDS et les méthodes d'optimisation stochastique est loin d'être complète et offrent ainsi une piste de recherche prometteuse en ML et DL [High 18].

Bibliographie

[Bott 18] L. Bottou, F. E. Curtis, and J. Nocedal. "Optimization methods for large-scale machine learning".
SIAM Rev., Vol. 60, No. 2, pp. 223–311, 2018.

[Good 16] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016.
http://www.deeplearningbook.org.

[High 18] C. F. Higham and D. J. Higham. "Deep Learning: An Introduction for Applied Mathematicians".
SIAM Review, Vol. 61, No. 4, pp 860-891, 2019.

[Jent 18] A. Jentzen, B. Kuckuck, A. Neufeld, and P. von Wurstemberger. "Strong error analysis for stochastic gradient descent optimization algorithms".
IMA Journal of Numerical Analysis, Vol. 41, Issue 1, January 2021, pp 455–492.

[Russ 10] S. Russell and P. Norvig. Artificial Intelligence: A Modern Approach.
Pearson, pearson Ed., 2010.

[Shi 19] B. Shi, S. S. Du, W. J. Su, and M. I. Jordan. "Acceleration via Symplectic Discretization of High-Resolution Differential Equations".
Advances in Neural Information Processing Systems 32 (NeurIPS 2019).

[Su 16] W. Su, S. Boyd, and E. J. Candès. "A differential equation for modeling Nesterov's accelerated gradient method: theory and insights".
J. Mach. Learn. Res., Vol. 17, pp. Paper No. 153, 43, 2016.

[Wibi 16] A. Wibisono, A. C. Wilson, and M. I. Jordan. "A variational perspective on accelerated methods in optimization". Proc. Natl. Acad. Sci. USA, Vol. 113, No. 47, pp. E7351–E7358, 2016.