

Approximation par tenseurs de la matrice pseudo-hessienne dans les méthodes du second ordre pour l'optimisation en apprentissage automatique

Supervisor: Quang Huy TRAN (IFPEN)

Ce sujet s'inscrit dans les thèmes scientifiques « méthodes d'approximation par tenseurs » et « optimisation en grande dimension ». L'étudiant qui travaillera sur ce sujet est Abdoulaye KOROKO, actuellement en première année de thèse à IFPEN.

En optimisation pour machine learning, une tendance actuelle est de considérer les méthodes du second ordre afin d'accélérer la convergence de la méthode traditionnelle du gradient stochastique. Dans une telle méthode, chaque itération requiert la résolution d'un système linéaire dont la matrice coûte cher à former, puis à inverser. Cette matrice contient des informations de courbure sur la fonction à minimiser. Elle peut être la matrice hessienne (pour la méthode de Newton) ou une approximation de celle-ci (dans une méthode quasi-Newton). Les deux cas qui nous intéressent sont : la matrice de Gauss-Newton (pour la méthode éponyme) et la matrice de Fisher (pour la méthode de gradient naturel).

Une approximation pertinente [1] pour la matrice de Fisher est le K-FAC (Kronecker Factorized Approximation of Curvature), dans laquelle la matrice de départ est remplacée par une matrice diagonale par bloc. Chaque bloc correspond aux variables d'une couche du réseau et se trouve à nouveau approché par le produit de Kronecker de deux plus petites matrices. Cette forme de Kronecker rend la résolution du système « facile » sans avoir jamais à calculer et à stocker le bloc. Un travail plus récent [2] inspiré de la décomposition de domaines améliore la qualité de l'approximation en ajoutant un terme de couplage au niveau grossier entre les couches. Mais la structure globale de l'approximation ne change pas : c'est toujours une somme de produits plus élémentaires.

L'objectif du travail proposé est d'examiner la question de l'approximation de cette matrice (ou de son inverse) d'un point de vue plus systématique à l'aide d'une théorie plus générale d'approximation par des tenseurs de bas rang, dans le but d'accroître encore la précision de l'approximation et/ou son efficacité. Les méthodes retenues seront appliquées non seulement à la matrice de Fisher pour la méthode du gradient naturel, mais aussi à la matrice de Gauss-Newton qui dans une certaine mesure présente une structure analogue.

References

- [1] Roger Grosse and James Martens. A kronecker-factored approximate fisher matrix for convolution layers. In *International Conference on Machine Learning*, pages 573–582. PMLR, 2016.
- [2] Nikolaos Tselepidis, Jonas Kohler, and Antonio Orvieto. Two-level k-fac preconditioning for deep learning. *arXiv preprint arXiv:2011.00573*, 2020.