Proposition de stage LMI - INSA Rouen Normandie Développement et étude d'une version streaming d'un algorithme de Newton stochastique universel

Lieu du stage : Laboratoire de Mathématiques de l'INSA (LMI, UR3026), INSA Rouen Normandie. La personne recrutée pourra également être amenée à travailler au Laboratoire de Probabilités, Statistique et Modélisation à Sorbonne Université, les frais de déplacement étant pris en charge par le LMI.

Niveau d'études : Master 2, Ecole d'ingénieur.

Titre du sujet : Développement et étude d'une version streaming d'un algorithme de Newton stochastique universel

Date de début : A partir de Mars 2024.

Durée du contrat : 5 à 6 mois.

Gratification : Environ 650 euros par mois, conformément à la grille officielle en vigueur.

Projet à plus long terme : Une demande d'allocation de thèse a été déposée auprès de la région Normandie. L'objectif serait que l'étudiant recruté pour ce stage poursuive en thèse.

Profil du candidat : Ecole d'ingénieur ou Master en mathématiques appliquées, avec de bonnes compétences en statistiques et en informatique.

Encadrement : Le stage sera co-encadré par Bruno Portier (INSA Rouen Normandie, LMI) et Antoine Godichon-Baggioni (Sorbonne Université, LPSM).

Candidature : Envoyer un message électronique accompagné d'un CV à Bruno Portier `bruno.portier@insa-rouen.fr`.

Sujet : Avec l'avènement du Big-data, de nouvelles problématiques dans le traitement statistique des données ont émergé. Parmi celles-ci, on trouve le traitement statistique en temps réel des données lorsqu'elles arrivent en flux continu et qu'elles ne sont pas stockées, ou bien acquises séquentiellement à partir de fichiers trop volumineux pour être chargés en mémoire, empêchant ainsi les procédures de calcul habituelles. Des méthodes récursives sont alors requises pour le traitement statistique de ces données, parmi lesquelles on trouve les algorithmes stochastiques. Les algorithmes stochastiques, introduits dans les années 50 (Robbins and Monro, 1951), sont en partie utilisés comme des outils d'optimisation stochastique pour minimiser une fonction inconnue mais s'écrivant comme l'espérance mathématique d'une fonction connue dépendant d'un vecteur aléatoire. Plus précisément, ils sont utilisés pour estimer le minimiseur d'une fonction $G : \mathbb{R}^d \longrightarrow \mathbb{R}$ de la forme

$$G(h) := \mathbb{E}[g(X, h)].$$

Ce type de problème est très fréquent et est rencontré (entre autres) :

(i) pour l'apprentissage supervisé (Bach, 2014; Bercu et al., 2020),

(ii) en apprentissage profond (Refinetti et al., 2023),

(ii) en statistique robuste (Cardot et al., 2013; Godichon-Baggioni and Lu, 2023).

Pour estimer la solution de ce problème de minimisation, les algorithmes de premier ordre type algorithmes de gradient stochastiques sont devenus, dans le cadre des données massives, incontournables ces dernières années. Il sont en effet très peu coûteux en terme de temps de calculs, ils permettent de faire des taches d'apprentissage machine sur des grands jeux de données et permettent de traiter ces dernière de manière séquentielle. Ces algorithmes ont connu du point de vue théorique un premier essors dans les années 90 (Polyak and Juditsky, 1992; Duflo, 1996; Pelletier, 1998, 2000) et ont atteint leur apogée ces dernières années (Bach and Moulines, 2013; Godichon-Baggioni, 2019; Gadat and Panloup, 2017). Cependant, ce type de méthodes peut rencontrer un certain nombre de problèmes. En particulier, les algorithmes du premier ordre ne prennent en compte que les informations données par le gradient de la fonction que l'on cherche à minimiser et peuvent par exemple être très sensible au cas où le spectre de la Hessienne de la fonction que l'on cherche à minimiser est large, dans le sens où les valeurs propres sont à des échelles très différentes (voir l'exemple de la Section 5.2 dans Bercu et al. (2020) pour s'en convaincre).

Une solution pour pallier les problèmes rencontrés par les algorithmes de gradient stochastique consiste à mettre en place des méthodes de second-ordre, i.e prenant en compte les informations données par la matrice Hessienne de la fonction que l'on cherche à minimiser. Plus précisément on va s'intéresser à des algorithmes de Newton stochastiques. La première difficulté rencontrée par cette méthode est que l'on ne connaît pas forcément la matrice Hessienne et il faut donc construire un estimateur récursif de celle-ci pour pouvoir conserver le caractère séquentiel de l'algorithme stochastique. La deuxième difficulté est que c'est davantage son inverse qui va nous intéresser, et dans l'esprit des algorithmes en ligne, il faut donc être capable de mettre à jour l'inverse de l'estimateur de la Hessienne, et ce avec un coût, en terme de temps de calculs, le plus réduit possible.

Récemment (Bercu et al., 2020), un algorithme de Newton stochastique a été proposé, pour estimer les paramètres d'un modèle de régression logistique,. L'estimateur récursif de l'inverse de la matrice Hessienne est mis à jours à l'aide d'une formule de Ricatti (aussi appelée formule de Sherman-Morrison) très peu coûteuse en temps de calculs. Les auteurs ont démontré que les estimateurs ainsi obtenus ont un comportement asymptotiquement optimal, et une version "streaming" a récemment été proposé pour obtenir des temps de calculs réduits. Cependant, ces méthodes ne peuvent s'appliquer que pour certains cas particulier (régressions logistiques et linéaire, par exemple).

Dans le même temps (Godichon-Baggioni et al., 2024), une nouvelle approche basé sur l'algorithme de Robbins-Monro (Robbins and Monro, 1951) a été proposée pour estimer l'inverse de la Hessienne. Cela a permis d'obtenir des estimateurs de type Newton stochastiques universels, i.e qui peuvent être appliqués à toute fonction strictement convexe et deux fois continûment différentiable. Bien que cette approche permet de réduire considérablement les temps de calcul, ces derniers restent beaucoup plus importants que ceux nécessaires pour les algorithmes de gradient stochastiques.

L'objectif de ce stage serait donc de coupler l'approche de type streaming ainsi que les algorithmes de Newton universels, pour construire un nouvel estimateur de Newton, là encore universel, mais nécessitant un temps de calculs de l'ordre de $O(nd)$ opérations (i.e du même ordre que les algorithmes de gradient stochastiques), où n est la taille d'échantillon et d la dimension. Des expérimentations numériques compléterons l'étude théorique menée sur les algorithmes proposés afin de prouver et d'éprouver leurs performances sur des données simulées et réelles. Les algorithmes développés seront codés en R ou en Python.

Références

Bach, F. (2014). Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *The Journal of Machine Learning Research*, 15(1) :595–627.

Bach, F. and Moulines, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate o (1/n). In *Advances in neural information processing systems*, pages 773–781.

Bercu, B., Godichon, A., and Portier, B. (2020). An efficient stochastic newton algorithm for parameter estimation in logistic regressions. *SIAM Journal on Control and Optimization*, 58(1) :348–367.

Cardot, H., Cénac, P., and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19(1) :18–43.

Duflo, M. (1996). *Algorithmes stochastiques*. Springer Berlin.

Gadat, S. and Panloup, F. (2017). Optimal non-asymptotic bound of the ruppert-polyak averaging without strong convexity. *arXiv preprint arXiv :1709.03342*.

Godichon-Baggioni, A. (2019). Lp and almost sure rates of convergence of averaged stochastic gradient algorithms : locally strongly convex objective. *ESAIM : Probability and Statistics*, 23 :841–873.

Godichon-Baggioni, A. and Lu, W. (2023). Online stochastic newton methods for estimating the geometric median and applications. *arXiv preprint arXiv :2304.00770*.

Godichon-Baggioni, A., Lu, W., and Portier, B. (2024). Online estimation of the inverse of the hessian for stochastic optimization with application to universal stochastic newton algorithms. *arXiv preprint arXiv :2401.10923*.

Pelletier, M. (1998). On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic processes and their applications*, 78(2) :217–244.

Pelletier, M. (2000). Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1) :49–72.

Polyak, B. and Juditsky, A. (1992). Acceleration of stochastic approximation. *SIAM J. Control and Optimization*, 30 :838–855.

Refinetti, M., Ingrosso, A., and Goldt, S. (2023). Neural networks trained with sgd learn distributions of increasing complexity. In *International Conference on Machine Learning*, pages 28843–28863. PMLR.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407.