

Note de synthèse

HAMMOUDA Ilyes

Mon stage d'application de la deuxième année, c'est déroulé au sein du CREST, du 12 juin 2023 jusqu'au 08 septembre 2023. Pendant ce stage j'ai été sous la supervision de Monsieur Alexandre TSYBAKOV et Monsieur Mohamed NDAOUD.

Comme je l'ai mentionné dans mon rapport de Mi-parcours, dès le début du stage on m'a proposé de me concentrer sur l'étude théorique du problème d'optimisation à l'ordre zéro, plutôt que l'étude des modèles génératifs et en particulier l'étude des algorithmes de décompression basés sur des modèles génératifs par rapport aux méthodes classiques basées sur la sparsité, comme il a été mentionné dans la convention de stage.

L'optimisation à l'ordre zéro consiste en l'optimisation d'une fonction dont on ignore la forme analytique. Une autre difficulté réside dans le fait que les observations auxquels on a accès sont bruitées. Or il est connu que dans le cadre des statistiques en grande dimension il est impossible de séparer le bruit des observations. L'optimisation à l'ordre zéro s'utilise également dans les situations où on n'a pas d'informations sur les moments d'ordre un et deux de la fonction objective, où lorsque le calcul numérique de ces moments est coûteux. En d'autres termes cette branche trouve son utilité dans les situations où les démarches d'optimisation classiques sont inefficaces surtout dans le cadre des applications en grandes dimensions.

Ainsi mon principal travail était de passer en revue, dans un premier temps, les travaux de recherche qui s'intéressent à la problématique d'optimisation à l'ordre zéro. J'ai commencé par étudier un des premiers articles qui propose une solution à cette problématique en grande dimension [1].

L'étude théorique de cet article a montré que les résultats de celui-ci sont sous-optimaux et parfois les démonstrations sont ambiguës. L'implémentation et les tests empiriques de l'algorithme de sélection de composants successifs, ont confirmé cette remarque.

C'est pour cela j'ai étudié dans un second temps la démarche proposée dans [2]. En effet les résultats de celle-ci sont bien plus optimaux et mieux justifiée. Cependant cette démarche n'est pas adaptative. C'est-à-dire l'algorithme repose sur des paramètres auxquels on n'a pas accès dans les applications réelles, notamment la sparsité. Il est vrai qu'une méthode adaptative a également été proposée dans l'article [2] mais aucun résultat théorique n'a été donné. Cette démarche proposée est également complexe, ce qui rend difficile la démonstration de résultats équivalents à ceux proposés pour l'algorithme non adaptatif.

Dans la deuxième partie du stage j'ai essayé d'étudier une méthode adaptative plus simple et démontrer quelques résultats théoriques et les confirmer empiriquement. Mon étude s'est basée sur les résultats proposés dans [3] qui permettent de résoudre les problèmes liés à la dépendance à certains paramètres auxquels on ne peut pas avoir accès dans le cadre de données réelles. La démarche adaptative qu'on a mise en place a l'avantage d'être simple. Ceci nous a permis de montrer des résultats théoriques sur le nombre d'itérations nécessaires pour atteindre le minimum et certaines autres bornes. Ces résultats théoriques se rapprochent de ce qu'on trouve à l'état de l'art et les résultats empiriques confirment ce qu'on a pu démontrer.

Pendant ce stage on a travaillé exclusivement sur l'amélioration de la procédure de

l'estimation du gradient. Ainsi il serait envisageable d'améliorer notre proposition en s'intéressant à la procédure de l'optimisation en elle-même et utiliser des algorithmes plus avancés comme la méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS) ou L-BFGS. Finalement j'aimerais remercier encore une fois Monsieur Alexandre Tsybakov et Mohamed Ndaoud pour leur soutien et encadrement hebdomadaire, ça fut un réel plaisir de travailler avec vous.

Stage d'application 2A

ENSAE Paris

Summary Report

HAMMOUDA Ilyes

My second-year internship took place at the CREST from June 12, 2023, to September 8, 2023. During this internship, I was supervised by Mr. Alexandre TSYBAKOV and Mr. Mohamed NDAOUD.

As I mentioned in my midterm report, from the beginning of the internship, I was offered the opportunity to focus on the theoretical study of zeroth-order optimization problems rather than the study of generative models. In fact, the main focus was on the study of the iterative algorithms related to gradient estimation based on sparse models.

Zeroth-order optimization involves optimizing a function for which the analytical form is unknown. Another challenge is that the available observations are noisy. In fact, it is known that in high-dimensional statistics models, it is impossible to separate noise from observations. Zeroth-order optimization is also useful in situations where there is no information about the first and second moments of the objective function or when the numerical computation of these moments is expensive. In other words, this branch finds its utility in situations where classical optimization approaches are ineffective, especially in high-dimensional applications.

Therefore, my main task was to first review the research work related to zeroth-order optimization. I began by studying one of the first articles that proposed a solution to this high-dimensional optimization problem [1]. The theoretical study of this article revealed that its results are suboptimal, and sometimes the proofs are ambiguous. The implementation and empirical tests of the successive component selection algorithm confirmed these observations.

Thus for this reason, I studied the approach proposed in [2] in the second part of the internship. The results in this work are more optimal and better justified. However, this approach is not adaptive, meaning it relies on parameters that are not available in real applications, especially sparsity. Although an adaptive method was also proposed in [2], however, no theoretical results were provided. This proposed approach is also complex, making it challenging to prove results equivalent to those for the non-adaptive algorithm. In the second part of the internship, I attempted to study a simpler adaptive method, demonstrate some theoretical results, and empirically confirm them. My study was based on the results presented in [3], which address issues related to dependencies on parameters that are not accessible in real-world data. The adaptive approach we developed is advantageous for its simplicity. This allowed us to demonstrate theoretical results regarding the number of iterations required to reach the minimum and some other bounds.

These theoretical results are in line with the state of the art, and empirical results confirm what we were able to demonstrate.

Throughout the internship, we exclusively worked on improving the gradient estimation procedure. It would be possible to enhance our proposal by focusing on the optimization procedure itself and using more advanced algorithms like the Broyden-Fletcher-Goldfarb-Shanno (BFGS) or L-BFGS methods. Finally, I would like to express my gratitude once again to Mr. Alexandre Tsybakov and Mohamed Ndaoud for their weekly support and guidance. It has been a real pleasure working with you.

References

- [1] Yining Wang, Simon Du, Sivaraman Balakrishnan, and Aarti Singh. Stochastic zeroth-order optimization in high dimensions, 2018.
- [2] HanQin Cai, Daniel McKenzie, Wotao Yin, and Zhenliang Zhang. Zeroth-order regularized optimization (ZORO): Approximately sparse gradients and adaptive sampling. *SIAM Journal on Optimization*, 32(2):687–714, apr 2022.
- [3] Mohamed Ndaoud. Scaled minimax optimality in high-dimensional linear regression: A non-convex algorithmic regularization approach, 2020.