

Методы безградиентной оптимизации в случае гладких задач

В. С. Ерошин, А. В. Гасников, А. В. Лобанов

Московский физико-технический институт (национальный исследовательский университет)

Задачи “Black-Box” оптимизации могут возникать в случае, когда вычисление градиента невозможно, либо является дорогой операцией. Решения задач оптимизации в постановке “Black-Box” вызывают интерес в таких областях, как обучении с подкреплением, федеративном обучении, распределенном обучении и перепараметризации. Основным способ решения таких задач - применение методов безградиентной оптимизации.

Для исследования методов безградиентной оптимизации вводится понятие оракула нулевого порядка, или “безградиентного” оракула [1]. Можно выделить несколько интересных нас концепций оракулов нулевого порядка: с детерминированным враждебным шумом и со стохастическим враждебным шумом.

В настоящее время можно выделить два основных подхода к решению задач безградиентной оптимизации: эволюционные алгоритмы и алгоритмы, основанные на методах оптимизации с использованием значений градиентов. Работа строится вокруг второго подхода, более конкретно - применения ускоренного стохастического градиентного спуска с использованием kernel-based аппроксимации градиента [2].

В работе рассмотрено обобщение результатов сходимости ускоренного градиентного спуска [3] на оракула, предоставляющего значение градиента со смещением [4]. Исследованы оценки на смещение и второй момент аппроксимации градиента в концепции стохастического шума [1]. Были получены соответствующие оценки для оракула с детерминированным шумом.

Рассмотрены оценки сходимости алгоритма и ограничений на величину шума в концепции оракула нулевого порядка со стохастическим шумом [1]. Получены оценки сходимости алгоритма и ограничения на величину шума в концепции оракула нулевого порядка с детерминированным шумом и сформулирована теорема, связывающая их.

Источники:

[1] Aleksandr Lobanov, Nail Bashirov, and Alexander Gasnikov. “The Black-Box Optimization Problem: Zero-Order Accelerated Stochastic Method via Kernel Approximation”. 2023. arXiv: 2310.02371.

[2] B. T. Polyak and A. B. Tsybakov. “Optimal orders of accuracy for search algorithms of stochastic optimization”. In: Problemy Peredachi Informatsii 26.2 (1990), pp. 45–53. issn: 0555-2923.

[3] Sharan Vaswani, Francis Bach, and Mark Schmidt. “Fast and Faster Convergence of SGD for Over-Parameterized Models and an Accelerated Perceptron”. In: Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Ed. by Kamalika Chaudhuri and Masashi Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, 16–18 Apr 2019, pp. 1195–1204.

[4] Ahmad Ajalloeian and Sebastian U. Stich. “Analysis of SGD with Biased Gradient Estimators”. In: CoRR abs/2008.00051 (2020). arXiv: 2008.00051.