Proceedings of Machine Learning Research vol 196:1–2, 2024 37th Annual Conference on Learning Theory

Online Newton Method for Bandit Convex Optimisation

Hidde Fokkema H.J.FOKKEMA@UVA.NL
Korteweg-de Vries Institute for Mathematics, University of Amsterdam

Dirk van der Hoeven DIRK@DIRKVANDERHOEVEN.COM
Mathematical Institute, Leiden University

Tor Lattimore LATTIMORE@GOOGLE.COM
Google DeepMind

Jack J. Mayo JACKJAMESMAYO@GMAIL.COM
Korteweg-de Vries Institute for Mathematics, University of Amsterdam

Editors: Shipra Agrawal and Aaron Roth

Abstract

We introduce a computationally efficient algorithm for zeroth-order bandit convex optimisation[1] and prove the following two results:

Theorem 1 *There exists an algorithm such that with probability at least* $1 - \delta$*,*

$$\mathrm{Reg}_n \leq \begin{cases} d^{3.5}\sqrt{n}\,\mathrm{polylog}(n, d, 1/\delta) & \textit{Adversarial}; \\ d^{1.5}\sqrt{n}\,\mathrm{polylog}(n, d, 1/\delta) & \textit{Stochastic}, K \textit{ in John's position or symmetric isotropic}; \\ d^{1.75}\sqrt{n}\,\mathrm{polylog}(n, d, 1/\delta) & \textit{Stochastic}, K \textit{ isotropic}. \end{cases}$$

Bandit convex optimisation is framed as a game between a learner and an adversary where the learner plays actions in a convex, compact set $K \subset \mathbb{R}^d$ with nonempty interior. The adversary secretely chooses a sequence of convex functions $\ell_1, \dots, \ell_n \to [0, 1]$. In round t, the learner chooses an action $A_t \in K$, suffers loss $\ell_t(A_t)$ and observes $\ell_t(A_t) + \varepsilon_t$ with ε_t a conditionally subgaussian random variable. The goal in bandit convex optimisation is to control the regret:

$$\mathrm{Reg}_n = \sup_{x \in K} \sum_{t=1}^{n} (\ell_t(A_t) - \ell_t(x)).$$

In the stochastic setting it is assumed that $\ell_t = \ell$ for all $t = 1, \dots, n$.

Our algorithm is based on the bandit version of online Newton step for the *stochastic unconstrained* bandit convex optimisation by Lattimore and György [2023], which estimates the gradients and Hessian through a Gaussian smoothed version of the losses. Some crucial new ingredients are added to handle the constraint set and the adversarial setting. We provide a reduction from unconstrained to constrained bandit convex optimisation based on a generalisation of the convex extension proposed by Mhammedi [2022] for the full information setting. The convex extension is not nicely behaved outside the constraint set, but we show that due to the design of the extension our meta-algorithm mostly plays on or close to the constraint set where the extension is well behaved. Finally, in the adversarial setting we employ an improved version of the restarting condition used by Suggala et al. [2021].

1. Extended abstract. Full version appears as [arXiv:2406.06706v1]

Acknowledgments

This work was partially done while DvdH and JJM were at the University of Amsterdam supported by Netherlands Organization for Scientific Research (NWO), grant number VI.Vidi.192.095. Thank you also to our colleagues Tim van Erven, Wouter Koolen and András György for many useful discussions.

References

T. Lattimore and A. György. A second-order method for stochastic bandit convex optimisation. In *Conference on Learning Theory*, 2023.

Z. Mhammedi. Efficient projection-free online convex optimization with membership oracle. In *Conference on Learning Theory*, pages 5314–5390, 2022.

A. Suggala, P. Ravikumar, and P. Netrapalli. Efficient bandit convex optimization: Beyond linear losses. In *Conference on Learning Theory*, pages 4008–4067, 2021.