The Price of Adaptivity in Stochastic Convex Optimization*

Yair Carmon YCARMON@TAUEX.TAU.AC.IL
Oliver Hinder OHINDER@PITT.EDU

Editors: Shipra Agrawal and Aaron Roth

Stochastic optimization methods in modern machine learning often require carefully tuning algorithmic parameters at significant cost in time, computation, and expertise. This reality has led to sustained interest in developing adaptive (or parameter-free) algorithms that require minimal or no tuning [1, 2, 4–8, 10–15, 17–20]. However, these adaptive methods typically have worse suboptimality bounds than their non-adaptive counterparts. Are existing methods "as adaptive as possible," or is there room for improvement? Put differently, is there a fundamental price to be paid (in terms of rate of convergence) for not knowing the problem parameters in advance?

To answer these questions, we draw inspiration from the "price of anarchy" in algorithmic game theory [16] and introduce the "price of adaptivity" (PoA). Roughly speaking, the PoA measures the multiplicative increase in suboptimality due to uncertainty in problem parameters. We show the following PoA lower bounds for non-smooth stochastic convex optimization:

1. When the initial distance to the optimum is unknown we show that the PoA is at least logarithmic for expected suboptimality. When the gradient norm is known, the upper bound of McMahan and Orabona [9] matches our lower bound. Combined with Carmon and Hinder [1], our lower bound also implies that in-expectation parameter-free guarantees are fundamentally *worse* than high-probability guarantees.

2. When the initial distance to the optimum is unknown we also show that the PoA is at least doubly-logarithmic for any suboptimality quantile. This establishes the near-optimality of Carmon and Hinder [1].

3. When there is uncertainty in both distance and gradient norm, we show that the PoA must be polynomial in the level of uncertainty. This lower bound also nearly matches a combination of upper bounds from the literature [1, 3].

4. Heavy-tailed noise (specifically, stochastic gradients with only bounded second moment) significantly increases the previous PoA bound, precluding meaningful adaptivity when both the distance and gradient second moment are unknown. In contrast, when all problem parameters are known, heavy-tailed noise does not affect the optimal rate of convergence.

* Extended abstract. Full version appears as [arXiv:2402.10898, v3].

Acknowledgements

We thank Amit Attia, Chi Jin, Ahmed Khaled, and Tomer Koren for helpful discussions. This work was supported by the NSF-BSF program, under NSF grant #2239527 and BSF grant #2022663. OH acknowledges support from Pitt Momentum Funds, and AFOSR grant #FA9550-23-1-0242. YC acknowledges support from the Israeli Science Foundation (ISF) grant no. 2486/21, the Alon Fellowship, and the Adelis Foundation.

References

[1] Yair Carmon and Oliver Hinder. Making SGD parameter-free. In *Conference on Learning Theory (COLT)*, 2022.

[2] Keyi Chen, John Langford, and Francesco Orabona. Better parameter-free stochastic optimization with ODE updates for coin-betting. In *AAAI Conference on Artificial Intelligence*, 2022.

[3] Ashok Cutkosky. Artificial constraints and hints for unbounded online learning. In *Conference on Learning Theory (COLT)*, 2019.

[4] Ashok Cutkosky and Francesco Orabona. Black-box reductions for parameter-free online learning in Banach spaces. In *Conference on Learning Theory (COLT)*, 2018.

[5] Maor Ivgi, Oliver Hinder, and Yair Carmon. DoG is SGD's best friend: A parameter-free dynamic step size schedule. In *International Conference on Machine Learning (ICML)*, 2023.

[6] Andrew Jacobsen and Ashok Cutkosky. Unconstrained online learning with unbounded losses. In *International Conference on Machine Learning (ICML)*, 2023.

[7] Ali Kavis, Kfir Y Levy, Francis Bach, and Volkan Cevher. UniXGrad: A universal, adaptive algorithm with optimal guarantees for constrained optimization. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[8] H Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18(1):3117–3166, 2017.

[9] H Brendan McMahan and Francesco Orabona. Unconstrained online linear learning in Hilbert spaces: Minimax algorithms and normal approximations. In *Conference on Learning Theory (COLT)*, 2014.

[10] Zakaria Mhammedi, Wouter M Koolen, and Tim Van Erven. Lipschitz adaptivity with multiple learning rates in online learning. In *Conference on Learning Theory*, pages 2490–2511, 2019.

[11] Francesco Orabona. Simultaneous model selection and optimization through parameter-free stochastic learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[12] Francesco Orabona. A modern introduction to online learning. *arXiv:1912.13213*, 2021.

[13] Francesco Orabona and Dávid Pál. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[14] Francesco Orabona and Dávid Pál. Parameter-free stochastic optimization of variationally coherent functions. *arXiv:2102.00236*, 2021.

[15] Courtney Paquette and Katya Scheinberg. A stochastic line search method with expected complexity analysis. *SIAM Journal on Optimization*, 30(1):349–376, 2020.

[16] Tim Roughgarden. *Selfish routing and the price of anarchy*. MIT press, 2005.

[17] Louis Sharrock and Christopher Nemeth. Coin sampling: Gradient-based bayesian inference without learning rates. In *International Conference on Machine Learning (ICML)*, 2023.

[18] Sharan Vaswani, Aaron Mishkin, Issam Laradji, Mark Schmidt, Gauthier Gidel, and Simon Lacoste-Julien. Painless stochastic gradient: Interpolation, line-search, and convergence rates. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[19] Jiujia Zhang and Ashok Cutkosky. Parameter-free regret in high probability with heavy tails. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[20] Zhiyu Zhang, Ashok Cutkosky, and Ioannis Paschalidis. PDE-based optimal strategy for unconstrained online learning. In *International Conference on Machine Learning (ICML)*, 2022.