



Smoothed Online Optimization with Unreliable Predictions

Daan Rutten

drutten@gatech.edu

Georgia Institute of Technology

Atlanta, Georgia, United States

Debankur Mukherjee

debankur.mukherjee@isye.gatech.edu

Georgia Institute of Technology

Atlanta, Georgia, United States

Nicolas Christianson

nchristianson@caltech.edu

California Institute of Technology

Pasadena, California, United States

Adam Wierman

adamw@caltech.edu

California Institute of Technology

Pasadena, California, United States

ACM Reference Format:

Daan Rutten, Nicolas Christianson, Debankur Mukherjee, and Adam Wierman. 2023. Smoothed Online Optimization with Unreliable Predictions. In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '23 Abstracts)*, June 19–23, 2023, Orlando, FL, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3578338.3593570>

1 INTRODUCTION

We consider online optimization with switching costs in a normed vector space $(X, \|\cdot\|)$ wherein, at each time t , a decision maker observes a non-convex *hitting cost* function $f_t : X \rightarrow [0, \infty]$ and must decide upon some $x_t \in X$, paying $f_t(x_t) + \|x_t - x_{t-1}\|$, where $\|\cdot\|$ characterizes the *switching cost*. Throughout, we assume that f_t is globally α -polyhedral, i.e., f_t has a unique minimizer $v_t \in X$, and, for all $x \in X$, $f_t(x) \geq f_t(v_t) + \alpha \cdot \|x - v_t\|$. Moreover, we assume that the decision maker has access to an *untrusted prediction* $\tilde{x}_t$ of the optimal decision during each round, such as the decision suggested by a black-box AI tool.

The bulk of the literature on online optimization with switching costs has sought to design *competitive* algorithms for the task, i.e., algorithms with finite competitive ratios [2, 3]. Although competitive analysis yields strong performance guarantees, it has often been criticized as being unduly pessimistic, since algorithms are characterized by their *worst-case* performance, while worst-case conditions may never occur in practice. On the other hand, many real-world applications have access to vast amounts of historical data which could be leveraged by modern black-box AI tools to achieve significantly improved performance in the typical case. Making use of modern black-box AI tools is potentially transformational for online optimization; however, such machine-learned algorithms fail to provide any uncertainty quantification and thus do not have formal guarantees on their worst-case performance. As such, while their performance may improve upon competitive algorithms in typical cases, they may perform arbitrarily worse in scenarios where the training examples are not representative of the real world workloads due to, e.g., distribution shift.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '23 Abstracts, June 19–23, 2023, Orlando, FL, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0074-3/23/06.

<https://doi.org/10.1145/3578338.3593570>

This work aims to provide an algorithm that can achieve the best of both worlds – making use of black-box AI tools to provide good performance in the typical case while integrating competitive algorithms to ensure formal worst-case guarantees. We formalize this goal using the notions of *consistency* and *robustness* introduced in [4]. The untrusted predictions are fully arbitrary, i.e., we impose no statistical assumptions on the predictions, and the decision maker has no *a priori* knowledge of the predictions' accuracy. The goal of the decision maker is to be both *consistent* – to achieve performance comparable to that of the predictions if they are accurate, while remaining *robust* – to have a cost that is never much worse than the hindsight optimal, even if predictions are completely inaccurate. Thus, an algorithm that is consistent and robust is able to match the performance of the black-box AI tool when the predictions are accurate while also ensuring a worst-case performance bound.

2 MAIN CONTRIBUTIONS

We make five main contributions. *First*, we identify a fundamental trade-off between consistency and robustness for any deterministic algorithm. If an algorithm is $(1 + \delta)$ -consistent, then this implies a lower bound on its robustness guarantee. In fact, we identify a region of (α, δ) , which we term the *infeasible region*, for which no algorithm can be simultaneously $(1 + \delta)$ -consistent and ζ -robust for any $\zeta < \infty$. Outside of the infeasible region, we prove that there is an exponential trade-off between consistency and robustness.

THEOREM 2.1. *Let $\mathcal{A}$ be any deterministic online algorithm. If $\mathcal{A}$ is $(1 + \delta)$ -consistent, then $\mathcal{A}$ is at least L -robust, where*

$$L = \begin{cases} \frac{\alpha\delta}{4} \left(\frac{2}{\alpha + \delta(1 + \alpha)} \right)^{\frac{2 - \alpha(1 - \delta^2)}{\alpha\delta(1 + \delta)}} - O(1) & \text{if } (\alpha, \delta) \text{ is not infeasible,} \\ \infty & \text{if } (\alpha, \delta) \text{ is infeasible.} \end{cases}$$

This threshold-type behavior is unprecedented in the literature on learning-augmented algorithms and reveals the hardness of the problem brought about by the non-convexity of hitting costs.

Second, we introduce a new algorithm for online optimization in a normed vector space with untrusted predictions, *Adaptive Online Switching* (AOS), and provide bounds on its consistency and robustness. Our analysis shows that AOS can be used in combination with a black-box AI tool to match the performance of the black-box AI tool while also ensuring provable worst-case guarantees.

The AOS algorithm works as follows. At each time t , AOS either follows the predictions $\tilde{x}_t$ or the minimizers v_t of f_t , and adaptively switches between these two. The challenge in the design

Algorithm 1 Adaptive Online Switching (AOS)

```

1: Let  $p_t := \arg \min_{p \in X} (f_t(p) + \|p - p_{t-1}\| + \|p - \tilde{x}_t\|)$ ,  $\text{ROB}(t) :=$ 
    $f_t(v_t) + \|v_t - v_{t-1}\|$ , and  $\text{Adv}(t) := f_t(p_t) + \|p_t - p_{t-1}\|$ .
2:  $T_1 \leftarrow 1$ ,  $t \leftarrow 1$ 
3: for  $k = 1, 2, \dots$  do
4:   while  $\sum_{i=T_k}^{t-1} \text{Adv}(i) + \text{ROB}(t) + \|p_{t-1} - v_{t-1}\| + \|v_t - p_t\| \geq$ 
      $(1 + \delta) \sum_{i=T_k}^t \text{Adv}(i)$  do
5:      $x_t \leftarrow p_t$ ,  $t \leftarrow t + 1$  and stop if  $t = T + 1$ 
6:   end while
7:    $M_k \leftarrow t$   $\triangleright$  Start following the minimizers
8:    $x_t \leftarrow v_t$ ,  $t \leftarrow t + 1$  and stop if  $t = T + 1$ 
9:   while  $\sum_{i=M_k+1}^t \text{ROB}(i) + \|v_t - p_t\| - \|v_{M_k} - p_{M_k}\| \leq (1 +$ 
      $\delta) \sum_{i=M_k+1}^t \text{Adv}(i) + \gamma \sum_{i=T_k}^t \text{Adv}(i)$  do
10:     $x_t \leftarrow v_t$ ,  $t \leftarrow t + 1$  and stop if  $t = T + 1$ 
11:   end while
12:    $T_{k+1} \leftarrow t$   $\triangleright$  Start following the predictions
13:    $x_t \leftarrow p_t$ ,  $t \leftarrow t + 1$  and stop if  $t = T + 1$ 
14: end for

```

of AOS is that, on the one hand, switching must be infrequent in order to limit the switching cost, but, on the other hand, switching must be frequent enough to ensure that the algorithm does not get stuck following a suboptimal sequence of decisions from either the predictions or the minimizers.

If (α, δ) belongs to the *feasible region*, we prove that AOS is robust. In either case, we prove that AOS is consistent.

THEOREM 2.2. *Let $CR(\eta)$ be the competitive ratio of AOS. Then,*

$$CR(\eta) \leq (1 + \delta + \gamma)(1 + 2\eta). \quad (1)$$

Moreover, if (α, δ) is feasible, then,

$$CR(\eta) \leq \frac{12 + o(1)}{\gamma} \left(\frac{2}{\alpha + \delta(1 + \alpha)} \right)^{2/(\alpha\delta)}. \quad (2)$$

Here, η is an appropriate measure of the accuracy of the predictions and relates to the distance between the prediction $\tilde{x}_t$ and the optimal decision, and δ and γ are hyperparameters of the algorithm. Note that, even though the competitive ratio is a function of the accuracy, the algorithm is oblivious to it. If the predictions are accurate, i.e., if $\eta = 0$, then the competitive ratio of AOS is $1 + \delta + \gamma$. In other words, AOS is $(1 + \delta + \gamma)$ -consistent and almost reproduces the hindsight optimal sequence of decisions if the predictions are accurate. Moreover, even if predictions are completely inaccurate, i.e., if $\eta = \infty$, the competitive ratio of AOS is uniformly bounded.

The trade-off between consistency and robustness is characterized by the confidence hyperparameters δ and γ , where the robustness bound depends exponentially on both δ and α . In light of our lower bound, this means that AOS reproduces the order optimal trade-off between robustness and consistency in the feasible region. As a proof technique, the conventional potential function approach fails due to the non-convexity of the problem and the incorporation of predictions. Therefore, significant novelty in the technique is required for the proof of Theorem 2.2.

Third, we extend the above results to the case when only an approximate non-convex solver is available. As we do not make any assumptions on the convexity of f_t , it may be computationally

difficult or simply impossible to obtain the exact minimizer of f_t . We therefore extend AOS to work with any non-exact, approximate minimizer of f_t . AOS is oblivious to the accuracy of the solver and we prove that the competitive ratio decays smoothly in the accuracy of the solver. Moreover, we provide bounds for the case when predictions cannot be used in every time step due to the computational expense associated with the non-convex functions. Our bounds characterize how the consistency-robustness tradeoff is impacted by this computational constraint. In fact, the impact on the competitive ratio is linear in the number of time steps between available predictions.

Fourth, we characterize the importance of memory for algorithms seeking to use untrusted predictions. Interestingly, AOS requires full memory of all predictions. This is a stark contrast with well-known memoryless algorithms for online optimization with switching costs, which do not make use of any information about previous hitting costs or actions. We prove that any memoryless algorithm cannot have simultaneous non-trivial robustness and consistency bounds. Thus, *memory is necessary to benefit from untrusted predictions*.

Fifth, we consider an important special case when the vector space is $X = \mathbb{R}$ and each function is convex and show that it is possible to provide an improved trade-off between robustness and consistency using a memoryless algorithm in this special case. In this context, we introduce a new algorithm, called Adaptive Online Balanced Descent (AOBD), which is inspired by Online Balanced Descent [1].

THEOREM 2.3. *Let $CR(\eta)$ be the competitive ratio of AOBD. Then,*

$$CR(\eta) \leq \min \{ (1 + \delta)(1 + 4\eta), 1 + 3/\delta + 2/\delta^2 \}. \quad (3)$$

The competitive ratio of AOBD has a similar structure to AOS, but improves the robustness bound significantly by taking advantage of the additional structure available compared to the general non-convex case. The result is complemented by a lower bound.

THEOREM 2.4. *Let $\mathcal{A}$ be any deterministic algorithm for the convex, one-dimensional optimization problem. If $\mathcal{A}$ is $(1 + \delta)$ -consistent, then $\mathcal{A}$ is at least $1/(2\delta)$ -robust.*

ACKNOWLEDGMENTS

The work was partially supported by the NSF grants CIF-2113027, CNS-2146814, CPS-2136197, CNS-2106403, and NGSDI-2105648, an NSF Graduate Research Fellowship (DGE-1745301), and Amazon AWS.

REFERENCES

- [1] Nianjun Chen, Anish Agarwal, Adam Wierman, Siddharth Barman, and Lachlan LH Andrew. 2015. Online convex optimization using predictions. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*. 191–204.
- [2] Nianjun Chen, Gautam Goel, and Adam Wierman. 2018. Smoothed online convex optimization in high dimensions via online balanced descent. In *Conference On Learning Theory*. PMLR, 1574–1594.
- [3] Yiheng Lin, Gautam Goel, and Adam Wierman. 2020. Online optimization with predictions and non-convex losses. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4, 1 (2020), 1–32.
- [4] Thodoris Lykouris and Sergei Vassilvitskii. 2018. Competitive caching with machine learned advice. In *International Conference on Machine Learning*. PMLR, 3296–3305.