

1 We thank the reviewers for their insightful comments. We are humbled that the reviewers view this work as novel (R1),
2 that it considers an interesting problem that is closely relevant to the NeurIPS community (R2), and that our algorithmic
3 approach and methodology are “solid” (R4). We spent considerable effort to make the writing and structure of the paper
4 easy to read and follow, and so are grateful that this effort was reflected in the referees feedback on clarity (R1, R2,
5 R4). To recap, we *develop algorithms* to solve a *learning problem* in *online advertising auctions*. Our optimization
6 algorithms are capable of producing models that perform better than those produced by existing techniques.

7 **Contribution to ML.** R1 expressed concerns about the contribution of the work "from an ML standpoint". While
8 we do not deal directly with, e.g., statistical properties of our learned models, our work is firmly grounded in an ML
9 application. NeurIPS has historically made a home for such papers that broadly define ML. To pick just one example, an
10 Honorable Mention for the Outstanding Paper Award at NeurIPS 2019 (*Fast and Accurate Least-Mean-Squares Solvers*
11 by Maalouf, Jubran, and Feldman) is a “pure algorithms” paper on a crucial subroutine in ML training algorithms.

12 **Venue fit.** R1 stated that the paper "may be more suited for a more specialized venue for MIP or for auctions, like
13 WWW". We agree that this work is likely be of interest to specialized communities in auctions and discrete optimization.
14 However, both audiences are well-represented in the NeurIPS community, and indeed R2 states that "The topic [of this
15 paper] is closely relevant to the NeurIPS community".

16 **Connection with Mohri and Medina (2016).** R4 expressed a desire for a more detailed explanation of the connection
17 between our proposed method and the work of Mohri and Medina (2016); referred to MM to follow. R4 asks: "Are both
18 papers considering linear models"?, and the answer is: **Yes**. Both papers consider training over an identical linear model
19 with discontinuous loss. MM presents an algorithm for the training problem by smoothing the loss using a surrogate
20 function which thus does not guarantee global optimality to the original problem. Our novel contribution is a family of
21 algorithms which, as the computations show, can close the optimality gap of the MM method *to ultimately produce*
22 *better performing models*.

23 **Second-price auctions have limited target audience.** R4 expresses concern that the paper may have a "limited target
24 audience as the method is meant to be used in second-price auctions". We respectfully disagree: generalized second-
25 price (Vickrey) auctions are commonly used by online ad platforms, and have been extensively studied, including at
26 past NeurIPS. By our count, 5 papers at NeurIPS 2019 included "auction" in the title, and two of these directly consider
27 second-price auctions or generalizations thereof. We also highlight that, as noted in the introduction, our approach can
28 be slightly modified to handle other auction variants, such as first price auctions or pure price-setting problems.

29 **Why does LP do poorly on synthetic data?** R1 states that "it is interesting that the LP does not really do better than
30 constant prices for synthetic data", and then asks why this might be the case. The explanation is: The models handle
31 constant offsets in slightly different ways. The constant policy (CP) does not enforce bounds on the magnitude on
32 the offset term, whereas other models do. Therefore, the CP model was not actually attainable by LP. This does not
33 matter in high dimensions (eBay), but can make a difference on our low dimensional synthetic data. We have now
34 implemented models that tune over the magnitude bounds for the offset, and can verify that i) LP does substantially
35 better on the synthetic data (beating CP), and ii) the results on the eBay data set remain unchanged. We are currently
36 repeating the synthetic data experiments on higher dimensional data to mitigate the spurious impact of the constant
37 offset. We thank R1 for their great observation, which lead to a substantial improvement of the performance of LP.

38 **Data leakage.** We thank R4 for pointing out the potential for temporal data leakage. To provide more detail, we are not
39 doing validation on “future” data: our validation set is constructed by randomly holding out a portion of the training
40 data. Ultimately, we agree with R4 that this likely does not have a strong effect on the results.

41 **Why not approximate r with a smooth differentiable function?** R4 suggests, as an alternative approach, to "approx-
42 imate r with a smooth differentiable function." This is an interesting and natural idea! We are considering this approach
43 in ongoing work, and indeed MM studied a variant of this smoothing technique. One thing to keep in mind is that the
44 problem remains nonconvex even after smoothing, and so is still NP-hard. Practically, the natural way to solve this
45 problem would be using (stochastic) gradient ascent. However, we believe that the performance would not be materially
46 different from running gradient ascent directly on the discontinuous problem, which can be observed to have very poor
47 performance in Tables 1 and 2. In a sentence, we believe this because i) the gradients only differ substantially in a small
48 neighborhood around discontinuities, and ii) even after smoothing the problem is very likely still have many (bad) local
49 minimums (see Figure 1).

50 **Presentation of important claims.** R2 stated that "it would be better if important claims...were presented." We may be
51 misunderstanding this comment, but we interpret it to mean that R2 is asking for proofs of the claims. In the submission,
52 we rigorously proved each of the claims in the supplementary materials. To make this connection clear, we have added
53 proof "stubs" to follow each formal result, pointing to the specific subsection in the appendix containing the proof.