

1 We thank the reviewers for the positive comments and useful feedback. We provide responses to the main comments.

2 **Connections to Cotter et al:** There are two main differences between our paper and Cotter et al. (2019a;b):

3 1. Cotter et al. consider constrained optimization problems where the objective and constraints are linear functions of
4 rates, i.e. are expected losses over a subset of the instance space. We handle a much broader class of problems where
5 the objective and constraints can be general *non-decomposable* functions of rates including e.g. F-measure, KLD, etc.

6 2. Like us, Cotter et al. propose a formulation where the θ -player optimizes the true Lagrangian and the λ -player
7 optimizes a surrogate-approximation to the Lagrangian. They consider two algorithms, one where both the θ - and
8 λ -player seek to minimize (external) regret, and the other where the θ -player optimizes alone minimizes external
9 regret, while the λ -player minimizes swap regret. They are however able to show convergence guarantees only for the
10 second algorithm, which arguably has a more complicated set of updates (requiring computing eigenvectors). They had
11 acknowledged this saying they had “no theoretical justification” for their external regret algorithm (p. 27 in [2]).

12 We address that *lack of theoretical justification* (and will change to that wording rather than referring to it as an *open*
13 *problem*, in accordance with Reviewer 1’s comment) by providing a theoretically justified external regret algorithm
14 in that we show that having both the θ - and λ -players optimize external regret does indeed lead to convergence, and
15 show this result for a more general setting than the one in Cotter et al. We will clarify this better in the main text and
16 elaborate on this in Appendix D.

17 **Two vs Three-Player Viewpoint:** We agree that our formulation can be also be phrased as a two player game where
18 the min-player plays uses different strategies for the ξ - and θ - portions of the objective. While we find the three-player
19 viewpoint to be a useful way to think about the problem algorithmically in that the three sets of parameters can use
20 different optimization algorithms (see Table 1), this viewpoint is not crucial to the main contribution of the paper. We
21 will clarify early on in the main text that our formulation can equivalently be regarded as a two-player game, and that
22 the three-player viewpoint is an algorithmic perspective.

23 **Code:** We will make Tensorflow code available.

24 **Motivation for Surrogates.** Generalized rate metrics such as the G-mean or F-measure are non-convex and non-
25 continuous functions of the model parameters, and hence directly optimizing them is hard in general. The traditional
26 approach to optimizing non-continuous metrics is to work with convex surrogates that upper bound the metrics. For
27 standard metrics such as the error rate, this is straight-forward and amounts to replacing the indicator function in the
28 metrics with e.g. the hinge loss. This simple approach can however be problematic for more general non-decomposable
29 metrics of the form $\psi(R_1(\theta), \dots, R_K(\theta))$, where ψ is often defined in a restricted domain (e.g. square-root or KL-
30 divergence), and replacing the rate terms with convex/concave relaxations may render ψ undefined. We explain this
31 with an example in Appendix C.

32 The proposed framework provides a cleaner solution for optimizing generalized rate metrics by using surrogates
33 only where necessary. In particular, we formulate a max-min Lagrangian optimization problem, and use surrogate
34 approximations for the Lagrangian when optimizing over the model parameters, and use the original rate values R_k ’s
35 for all other updates. We will include a discussion on surrogates in Section 2.

36 References

- 37 [1] Cotter, A., Heinrich J., and Sridharan, K. (2019). Two-Player Games for Efficient Non-Convex Constrained
38 Optimization. *Algorithmic Learning Theory*.
- 39 [2] Cotter, A., Jiang, H., Wang, S., Narayan, T., Gupta, M., You, S., and Sridharan, K. (2019). Optimization with
40 Non-Differentiable Constraints with Applications to Fairness, Recall, Churn, and Other Goals. ArXiv:1809.04198.
41 URL: <https://arxiv.org/abs/1809.04198>.