We thank all three reviewers for their time and comments. Their suggestions will help us clarify the contributions of our work as we incorporate them in the next revision of our paper.

In the following, we respond to several specific points raised by the reviewers.

Reviewer #1: You mentioned that the paper is at the end not self-contained because we didn't include a summary of how the quasi-concave optimization (RecConcave) works. We take it into our attention and in the body of the final version we will try to add a short summary of how it works, and to add more details in the appendix.

Reviewer #2: You wrote "it might be helpful if the authors were to offer some thoughts on whether a linear dependence on $\log^* X$ is possible in higher dimensions". We take it into our attention. In the final version we will mention this question, and write that one option for answering it is by finding a different 1-dimensional quasi-concave optimization that is linear (or almost linear) in $\log^* X$, since RecConcave, the optimization that we are using, requires exponential dependency in $\log^* X$. Indeed, a recent work of Kaplan, Ligett, Mansour, Naor, and Stemmer [COLT 2020] shows an (almost) linear dependency in $\log^* X$ for 1-dimensional thresholds, which is a special case of a quasi-concave optimization, and it still remains open whether this result can be extended to the quasi-concave optimization case.

Reviewer #2: You wrote that it would be nice to offer some possible uses or obstacles to using the generalized QuasiConcave paradigm. Indeed, as you mentioned, for the linear feasibility it was more involved to define the domain at each iterative step, and might be even more involved for other d-dimensional functions. The point is that this technical issue is inherent for privately optimizing such functions (at least if the optimization is done coordinate by coordinate), because we know that we must pay at least $\log^*$ of the domain size by any private algorithm. So we cannot get away from finding finite domains. But, if we can find such domains with some finite bound on their sizes, even if it is very large and not tight at all, it is usually should be enough since we are going to pay just $\log^*$ of these sizes in the sample complexity. We will try to emphasize this point in the paper.

Reviewer #3: You wrote "it will be more clear if the authors can define the problem of learning half spaces in the introduction or preliminary, ..." and "it is worth mentioning that the results in this paper can be easily generalized into the statistical setting". We take all your suggestions into our attention. We will try to address them in the final version.