We thank the reviewers for their careful consideration of our paper and their positive feedback. In the following paragraphs, we address some comments and questions asked by the reviewers.

Reviewer 1

Thank you very much for your review and suggestions.

Question: *Is the sample complexity upper bound of $1/\gamma^2\epsilon^2$ optimal especially under the restricted approximation factor considered in Theorem 3.2 and Theorem 3.3?*

Answer: Theorem 3.2 considers the agnostic version of the problem (what we call 1-agnostic in our paper). This is the most stringent notion of approximation and it is known that the sample complexity of this problem is $\Omega(1/(\epsilon^2\gamma^2))$ (information-theoretic lower bound). See, e.g., [BS00] or [SSS09] for an explicit reference.

Regarding weaker notions of approximation (like the ones addressed in our Theorems 3.1, 3.3 and our upper bounds in the supplementary material) we note the following: For a constant factor approximation ratio α, one can show that indeed the $\Omega(1/(\epsilon^2\gamma^2))$ bound applies as well. Since we could not find an explicit reference of this fact, we remark the following: Even for the basic case that the data is *linearly separable with margin* γ (i.e., $\mathrm{OPT}_\gamma^{\mathcal{D}} = 0$), the sample complexity of learning is $\Omega(1/(\epsilon\gamma^2))$. This is a known fact and can be found, e.g., in Cristianini Shawe-Taylor's book. Therefore, even for weaker notions of approximation (i.e. $\alpha = \infty$) our sample complexity is optimal as a function of γ and within a quadratic of optimal as a function of ϵ. Note that previous algorithms (with similar approximation guarantees) had sample complexity exponential in $1/\gamma$.

Question: *Line 254, by the standard arguments, it seems that the ϵ term should contain a multiplicative factor of $1/\gamma$.*

Answer: We are not sure exactly what the reviewer means here. Fact 2.4 (lines 251-254) is a standard generalization bound that we quoted from the literature, saying that after $\Omega(1/(\epsilon^2\gamma^2))$ samples the empirical distribution is accurate with high constant probability. In particular, we believe that line 254 is correct as is.

We will fix the other typos in the revised version of the paper.

Reviewer 2

Thank you very much for your review and suggestions.

Reviewer 3

Thank you very much for your review and suggestions.

Question: *In section 1.4, Chow parameters are mentioned, but they are not subsequently used in the presented analysis, they are only used in the supplementary material. It is not clear how the algorithm and analysis in the supplementary relate to the ones in the paper.*

Answer: The reviewer is correct that the Chow parameters do not appear again in the main body of the paper. We use the notion of Chow parameters in our second algorithm for large values of α (Theorem 1.2), and its proof is entirely in the supplementary material. This algorithm is based on a somewhat different approach than our main algorithmic result (Theorem 2.1). We will make sure to highlight this more clearly in the revised version.

A brief high-level description of our algorithm for large values of α can be found at the end of page 3. To reiterate here, it is well-known that the Chow parameters of a halfspace uniquely determine the function. With a margin condition, one can show a robust version of this statement. Our algorithm attempts to approximate the true Chow parameters. In the realizable setting, this would amount to simply using the empirical Chow parameters. However, in the agnostic setting we consider, the corruptions can only produce a large change in the empirical Chow parameters if the error points in question are very large in the direction of the change. Our algorithm works by guessing a small number of points and guessing a correction for the empirical Chow parameters in the subspace spanned by them. Once we have robustly learned the Chow parameters, we can use them to efficiently find the weights of an approximate linear separator by leveraging known techniques.

We will fix the other typos pointed out in the revised version of our paper.