We appreciate the valuable comments from reviewers on paper presentation and typos. We will revise our work accordingly. In what follows, we first address some common concerns as follows.

(Intuition Behind the Model Design.) Our model is a mixture of two single index models with α as the mixing probability. The intuition behind the design is to interpolate between two classes of the link functions, which exploit the first- and second-order Stein's identities for the estimation of β^*, respectively. Such a model design allows us to control the magnitude of first-order Stein's identity by the mixing probability α and study the effect of the first-order Stein's identity to the recovery of β^*.
(Intuition Behind the Function Class $\mathcal{C}(\psi)$.) The function class $\mathcal{C}(\psi)$ includes all the link functions with nonzero second-order Stein's identity under the marginal transformation ψ. Our work studies the minimax separation rate, which corresponds to the minimum signal to noise ratio (SNR) needed for the best algorithm when applying to the hardest model. Here ψ is one of the parameters that specify the model. Hence, by introducing the function class $\mathcal{C}(\psi)$ with ψ being arbitrary (yet smooth), we investigate a large family of models when searching for the hardest model.

Reviewer 1

(1). (Intuition Behind Our Model.) Please refer to the intuition behind the model design in the common concerns.
(2). (Lower Bound For Parameter Estimation Implied By Testing.) Intuitively, an algorithm that estimates β^* with sufficient accuracy can also be used to test nonzero β^* if the SNR is sufficiently large. Hence, by investigating the lower bound in the testing of nonzero β^*, we can characterize the minimal possible statistical error in the estimation of β^*. We refer to §3.3 for a detailed investigation of such an intuition.

Reviewer 2 (1). (Specify ψ, f_1, and f_2?) We refer to common concerns for the intuition behind ψ. We highlight that our goals is to characterizes the minimax separation rate, which corresponds to the minimal SNR needed for the best algorithm in solving the hardest model. Fixing arbitrary f_1, f_2, and ψ does not suffice, as they may not corresponds to the hardest model. In specific, we derive an upper bound that matches the lower bound we obtained by our selection of the parameters f_1, f_2, and ψ. Our work does not follow [2]. In specific, we study the phase transition in the minimax separation rate corresponding to the first-order Stein's association, whereas [2] studies a misspecified phase retrieval algorithm that estimates β^* via solely exploiting the second-order Stein's association, and their technique does not apply to our work. Moreover, the effect of computational tractability on the sample complexity is not studied in [2], which is a major concern of our work.
2. (Normalizing β^*?) Our work investigates the minimax separation rate, which corresponds to the minimum SNR needed for the hardest recovery of β^*. In specific, our work fixes the scale of link functions and the noise, and characterize the hardness of the recovery of β^* by the SNR $\|\beta^*\|_2^2/\sigma^2$. Hence we cannot normalize the quantity $\|\beta^*\|_2^2$, as such a quantity characterizes the hardness of recovering β^*, which is a major concern of our work.
(3). (Similarity to [1]?) Our model is a single index model where the link function is unknown, whereas [1] study the mixed linear regression model. They also show that phase retrieval model with link function $f(u) = |u|$ can be reduced to mixed linear regression and thus obtain the lower bound by resorting to the hardness of mixed linear regression. Thus, their detour approach cannot be used for showing the computational barriers in general single index models. Instead, our results cover their phase retrieval model as a special case.
(4).(Similarity to [3] and [4]?) Our work studies the statistical-computational tradeoffs in the single index model, whereas [3] study mean detection problems only. Moreover, [4] study the binary classification problem with the corrupted label. The model we consider are more complicated than their models and thus their analysis does not apply.
(5).(Technical Challenges.) Compared with these related work, we consider a larger model class. Our lower bound requires careful analysis of the χ_2-divergence the interpolated model introduced in (B.8). Moreover, our upper bounds holds for general single index models are thus handling the unknown nonlinear function f.

Reviewer 4 We thank the reviewer for the numerous advice in the layout and content of our work. We will revise accordingly. We move the results on upper bounds to the appendix due to the page limit. We will briefly introduce the upper bounds in the revision.

Reviewer 5 (1). (Intuition Behind The Function ψ.) Please refer to intuition behind the function class $\mathcal{C}(\psi)$ in the common concerns. Our result applies to the function class $\mathcal{C}(\psi)$, which contains the link functions that correspond to a large family of phase retrieval problems.
(2). (Explanation Of The Statistical Oracle.) Under the statistical query model, the sample complexity n becomes a parameter of the statistical oracle, and the sub-Gaussian error in the response of a query function is a natural extension of n independent realizations of the query function.

[1] Fan et al. " Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval" (2018).
[2] Neykov et al. "Agnostic estimation for misspecified phase retrieval models" (2016).
[3] Wang et al. "Sharp computational-statistical phase transitions via oracle computational model" (2015).
[4] Yi et al. "More supervision, less computation: Statistical-computational tradeoffs in weakly supervised learning"