We thank all the reviewers for the helpful comments. Here, we address the main concerns raised by the reviewers.

To Reviewer # 4

- [Additional discussion related to τ] Any τ that is less than the ground truth one is sufficient; in parctice, one can start with a large τ and decay its value until the algorithm finds a parameter that fits τn samples. That being said, of course a better estimate of τ means ILTS needs fewer samples. Additionally, the fraction of corruptions should always be no larger than a constant fraction of the interested component. Otherwise, it is impossible to identify if a mixture component comes from an adversary or not.
- [Compare with [14]] The computational complexity of [14] is **not** cheaper than ours. More specifically, in the "computation" column, the n in each row represents the sample complexity of the corresponding algorithm. For [14], the n in term nd has an exponential dependency on m. We will make this clearer in the revised version.
- [Claim about nearly-optimal computation] We would like to point out that the computation used in the global step is not due to any trade-off between sample-efficiency and computation. Instead, it is related to the hardness of the problem in the general setting. Also, from a practical standpoint, our algorithm is easily parallelizable (while [14] is not).
- [About ϵ-net] The ϵ-net is applied to a $O(m)$-dimensional subspace, where m is the number of mixtures. The mixed linear regression problem gets much harder as m gets larger. In this general setting, ϵ-net may not be an overkill. One intuition (also mentioned in [14]) is that recovering m Gaussian mixture models in general requires at least exponential in m number of samples. We consider the general setting, and hope the local/global analysis in this paper can shed light on understanding the structure of the problem. That being said, if additional assumptions (e.g., all components are orthogonal to each other) are added to the problem, there may exist more efficient approaches.
- [Writing of technical part and notations] We will keep improving the presentation of the technical part according to the reviewer's suggestion. In the current version, we try to make our notation self-contained: for example, the $\star$ in the superscript is used to denote "oracle knowledge" – $\theta^\star_{(j)}$ is the ground truth parameter for the j-th mixture component, $S^\star_{(j)}$ is the ground truth index set of samples that belong to the j-th mixture, $R^\star$ is the underlying index set of adversary samples.

To Reviewer # 5

- [Additional informal statements of Theorems 7 and 11 to improve readability.] We thank the reviewer for the kind suggestion. We will give an informal statement of our main results in the introduction section to improve readability.
- [c_0 in Theorem 7] c_0 is a constant such that $\kappa_t < 1$, and such a c_0 corresponds to an upper bound on c_j, i.e., the local region. We will add more detailed description to explain this Theorem in our revised version.
- [Explicit dependency on failure probability δ] High probability in all our arguments means with probability at least $1-n^{-c}$ (i.e., $\delta = n^{-c}$) for any given constant c. We did not write out the dependency explicitly because it contributes an additive factor of $\log d$ and is not the dominant term in the final expression for n. More intuitively speaking, our sample complexity is in the form of $d + \log(1/\delta)$, instead of $d\log(1/\delta)$. That being said, adding the dependency on δ would make our results more sensible, and we will add it in our revised version.
- [Typos] We will fix the typos and correct the confusing sentences as pointed out by the reviewer. "centered sphere" in line 5 of Algorithm 2 means the sphere with origin as its center.

To Reviewer # 6

- [Optimality of $\gamma^\star$] Our main convergence theorem holds when the corruption ratio is a constant fraction of the size of the interested mixture component. This corruption level is order-wise optimal. That being said, it is interesting to study other algorithms that can tolerate larger corruption ratio than ILTS does, in the mixed linear regression setting.
- [Numerical evaluations] We agree with the reviewer that adding rigorous numerical evaluations for ILTS and compare it with others in terms of scalability can be very helpful. We would like to add this section in our future versions. For the current paper, we would like to focus on the theoretical results, which may provide insights to understanding the mixed linear regression problem and the ILTS algorithm.