First, we would like to thank the reviewers for the careful reading of the paper and for helpful/thoughtful remarks. All the recommendations made by the reviewers will be taken into account in the revised version.

Response to Reviewer # 1 Thank you very much for your very enthusiastic report. Let us briefly comment on the question related to the choice of the parameter λ_o and on its impact on the quality of estimation.

The basic intuition included in your report is correct: the fact that $\lambda_o\sqrt{n}$ tends to infinity when $n \to \infty$ implies that all the outliers for which the outlyingness $y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - \xi_i$ is smaller than $c\sigma\sqrt{\log n}$ will be concerned only by the squared loss. As a consequence, if we were aware that all the outliers satisfy the condition $|y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - \xi_i| \leq c\sigma\sqrt{\log n}$ for some $c > 0$, there would be no need of using the Huber loss; the standard (penalized) least-squares estimator would have the statistical precision described in Theorem 3. However, one can never really check whether this condition is satisfied or not, since, for instance, $\boldsymbol{\beta}^*$ is unknown.

Response to Reviewer # 2 Thank you for finding our proofs sound and the paper well written. We find, however, that some other claims of your review are not fair and would like to explain our point of view.

- *This work shows that ℓ_1-penalized Huber's M-estimator can tolerate (up to constant fraction) of outliers in the response variables (y).*

  We believe that this formulation does not reflect well the content of our paper, since the words "can tolerate" have no clear meaning in this setting. The more accurate wording would be "is minimax-rate-optimal".

- *This paper only deals with the setting where response variables (y), and has to assume that x are sub-Gaussian, and similar results were already established in (Bhatia et al., 2017).*

  Unfortunately, due to the space limits, we could not include in the paper a more detailed comparison with prior work. We will do our best for including this comparison in the revised version of the paper.

  It is not fair to compare our results to those in (Bhatia et al., 2017) for the following reasons:

  1. (Bhatia et al., 2017) consider that the number of outliers and the sparsity, or a good upper bound on it, are known, while our method does not need this information.
  2. The method in (Bhatia et al., 2017) needs the Euclidean norm of the vector $\boldsymbol{\theta}^*$, while our method does not need this information.
  3. (Bhatia et al., 2017) do not cover the case of fully adversarial contamination. They consider, for instance, that the corruption vector $\boldsymbol{\theta}^*$ is independent of the design matrix and of the noise.
  4. Finally, (Bhatia et al., 2017) do not cover the high dimensional case under the sparsity constraint. Their estimator is provably consistent only when p/n goes to zero, where p is the dimension of the covariates (d in their paper) and n is the sample size. In our result, only s/n needs to go to zero, where s is the number of non-zero entries of the vector $\boldsymbol{\beta}^*$. This allows us to handle the situation where p is larger than n.

- *However, (Liu et al., 2019) can handle corruptions in both x and y, although the bound is slightly worse.*

  The words "can handle coruptions" being somewhat abstract, let us emphasize some striking differences between the results in (Liu et al., 2019) and those in our paper.

  1. (Liu et al., 2019) consider that the fraction of outliers, ε, and the sparsity, s (k in their paper) are known. They also need the norm $\|\boldsymbol{\beta}^*\|_2$ for determining the parameter T. None of these parameters are used by our algorithm.
  2. The constrains on ε in (Liu et al., 2019, Corollary 4.1) is of the form $\varepsilon = \tilde{O}(1/\sqrt{s})$ while in our result it is of the form $\varepsilon = \tilde{O}(1)$.
  3. Most importantly, the rate of convergence in (Liu et al., 2019, Corollary 4.1), is much slower than the one in our result. Indeed, they obtain the rate $\tilde{O}(\varepsilon\sqrt{s} + \frac{\sqrt{s}}{\sqrt{n}})$ whereas our rate is $\tilde{O}(\varepsilon + \frac{\sqrt{s}}{\sqrt{n}})$.

  This being said, we agree that (Liu et al., 2019) consider a more general setting, which fully justifies the important differences mentioned above. But we feel that it is unfair to claim that their bound is "slightly worse".

We sincerely hope that all these explanations will convince the reviewer that our paper contains significant improvements of previous results. It is also paramount to stress that our results are obtained for a simple estimator that is already used by many practitioners since several decades.

Response to Reviewer # 3 Thank you very much for your very positive and encouraging report. We agree with all the remarks/comments/suggestions you made. Concerning your remark (6), the trace should be removed. In an initial version of the paper, Lemma 3 was stated in a more general case in which $\boldsymbol{b}$ was a matrix. This is why the trace operator was used. We apologize for this typo.