We thank the reviewers for their time and effort in providing feedback. We are encouraged by the universally positive scores, and that all the reviewers appreciated the paper for the following: (i) significant results (**R1**,**R3**,**R4**), (ii) technical contribution (**R1**,**R3**,**R4**), (iii) a unified view of heavy-tailed and robust mean estimation (**R3**), and (iv) clarity (**R1**, **R2**, **R3**). For completeness, we summarize the contributions of our paper below.

Summary: Our goal is to show that a host of recent *computationally efficient* algorithms achieve optimal (or near-optimal) statistical results for two important families of distributions. We achieve this by showing that the underlying deterministic structural condition, *stability*, holds with optimal (or near-optimal) rates. Thus, our work simultaneously improves the statistical error guarantee of these existing algorithms without designing a new algorithm.

We address the individual questions and comments by the reviewers below.

Reviewer 1 (R1): We thank the reviewer for the positive feedback. Regarding the tolerance of [DL19,LLVZ19] to adversarial contamination: The algorithms and analyses in these papers establish tolerance to additive contamination but not strong contamination. More broadly, there is a significant difference between additive and strong contamination.

Reviewer 2 (R2): We thank the reviewer for the positive feedback. The reviewer asked regarding the *polynomial complexity* and *practicality* of known stability-based algorithms. We would like to emphasize that the computational aspects of these algorithms ([DK19, CDG18, SCV18, DKK+17,DHL19, CDG20]) are well-studied. For concreteness, we specify the running time for two existing algorithms that achieve the rate in Proposition 1.6:

- Universal filter [DK19]: $\tilde{O}(\min(k,d)k^2 d)$
- Quantum entropy filter [DHL19]: $\tilde{O}(k^2 d)$

As shown in [DKK+17, DHL19], the filter algorithm (and its variants) is **scalable** and **practical**. In particular, these algorithms have been successfully implemented in practical applications (see [DHL19] and [DKK+17] for experiments). Combining our statistical results with the runtime of these filtering algorithms, we obtain fast algorithms for heavy-tailed robust mean estimation in the strong contamination model. Prior to our work, no polynomial-time algorithm (with provable guarantees) was known in the strong contamination model.

Reviewer 3 (R3): We thank the reviewer for a detailed and encouraging feedback. We agree that it is an important question, both conceptually and practically, if the stability-based algorithms achieve the optimal rate without pre-processing.

Comparison with prior work: Additional details for lines $140-143$: Some prior works state their guarantees in terms of sample complexity to get $O(\sqrt{\epsilon})$ error, either with constant probability or with probability $1-\tau$. In this terminology, our sample complexity is $n = \Omega((d\log d + \log(1/\tau))/\epsilon)$. We mention the rates from the prior work below, all of which are sub-optimal in one or more parameters:

- [DKK+17]: Guarantees are stated for large constant probability.
- Goodness [DHL19]: $\delta = \sqrt{(d\log d)/(n\tau)} + \sqrt{\epsilon}$.
- Resilience [SCV18]: Even for constant probability, the sample complexity is $\Omega(d^{3/2}/\epsilon + d/\epsilon^2)$.
- Generalized resilience [ZJS19]: They give two bounds: (i) $\delta = O(\sqrt{\epsilon} + \sqrt{\frac{d\log(d/\tau)}{n}})$, (ii) $\delta = \sqrt{\frac{\epsilon}{\tau}} + \frac{1}{\tau}\sqrt{d/n}$.
- [PBR19]: $\delta = O(\sqrt{d\log(d/\tau)/n})$.

Thanks for pointing out the typos, we will fix them.

Reviewer 4 (R4): We thank the reviewer for the detailed feedback.

Presentation: We would work on the provided suggestions. "*Some examples are in the additional comments section*" — it seems, unfortunately, that this field is missing from the review. We will be happy to address these points once the reviewer updates their review with these comments after the rebuttal phase.

Rank-deficient Σ: Why is the generalized covariance compared to identity?

We note that for rank-deficient Σ, the tightest bound that we can show is $\delta = O(\sqrt{\epsilon})$ (Theorem 1.4). This is precisely because for this choice of δ, the eigenvalue has trivial lower bound of $1 - \delta^2/\epsilon$, which is negative (see proof of Claim 2.1). As the reviewer points out, we need more information about Σ to obtain $o(\sqrt{\epsilon})$ error. For such cases, we assume the knowledge of Σ, and obtain tighter rates in Theorem 1.8.