We thank the reviewers for the valuable comments and suggested improvements.

Reviewer 1: Theorems in Section 2.1 stated without reference to motivation. In the final version, we will interleave the theorem statements from Section 2.1 and Section 4 with the corresponding discussion in Section 1.2. This will hopefully aid the flow and add value to the remarks and discussion in Section 1.2. The reference to the preliminaries will be replaced with one to Section 1.1 (Problem setup and motivation).

Reviewer 2: Technical challenges compared to [6]. The works [6] and [7] developed bounds for how the top-k eigenspace is perturbed under sampling and aggregation. In our algorithm, we need to analyze how the *low rank approximation* is perturbed. While the techniques are inspired by those works, we extended (and simplified) their results for our purposes. We will highlight the differences further in the final version. Finally, the analysis for imprecise k (Theorem 9) is entirely novel; indeed the setting itself has not been studied in prior works. Not knowing the precise location of an eigenvalue-gap is quite natural in applications, and thus we view this as an important setting to study.

Reviewer 5: Intuition about cancellations in Theorem 4. We thank the reviewer for pointing this out. The explanation for the linear (dominant in magnitude) terms cancelling currently appears in the proof at the end of page 5. We will include the main idea in the proof overview at the start of the section.

Reviewer 5: Comparison to work on sketching for EVD/SVD. The works based on sketching have been crucial to our understanding of distributed SVD. We will include a discussion on these techniques in the final version. A key difference between the setting we consider and many of those works is that we focus on finding the precise eigenvectors/values, as opposed to minimizing the low rank error. The two notions are related in many realistic settings, and it is an interesting direction to compare them. At a high level, averaging methods (such as ours) perform really well as the number of points per machine (parameter n) grows. In contrast, sketching methods perform better as the sketch size grows (while not being too dependent on n).

Reviewer 6: Contribution and tightness of the bounds. Our results may indeed be viewed as correcting the estimators from prior work. Indeed, the algorithm itself changes only slightly. However, this change leads to significant improvements both theoretically and empirically for the fundamental problem of estimating the SVD. Further, the analysis now involves more work, as described in the response to review #2 (please see above).

The bounds in the theorems are tight as far as the dependence on the parameters m, n and the gaps (Δ_k terms) go. It is possible that dependence on the trace can be improved. We will include a detailed discussion (and relevant open questions) in the final version. We note that optimality in terms of n (i.e., the first term in Theorem 1) was shown in [7], and the dependence on $\sqrt{mn}$ follows from the Sin-theta theorem and known optimality results for matrix concentration.