We thank all the reviewers for the positive comments.

R1: Correctness of sr(A)-1 < k < sr(A) Yes, this statement is correct. Our goal in Remark 1 is to illustrate a sharp transition in the approximation factor as k approaches the stable rank from below. Remark 1(2) is an existential result i.e. there exists an $\mathbf{A}$ for which the lower bound on the approximation factor holds.

R1: L92 “stable rank of A is 20" should be 40 Here, we are referring to $\mathrm{sr}_0(\mathbf{A})$, which is 20 and in Figure 1 it coincides with the first peak in the approximation factor.

R1: Minor typos/comments Thanks for the careful read. We will make the suggested changes in the final version.

R1: Fig 2. Is the min $\Phi_s(k)$ line taken over $s \in \{10, 15, 25\}$ or over all values of $s \in [0, \mathbf{rank}(\mathbf{A})]$ The minimum is taken over all values of s.

R1: Proof of Thm 3. Importance of orthogonality of simplices For our setup, because of multiple simplices, the symmetry that holds for Deshpande et al. [2006] breaks, and thus different subsets of the same size have different possible errors. Orthogonality, along with the choice of α_i, β_i, ensures that we can pinpoint the best possible set of size k and its corresponding error amongst the different choices (see proof of Lemma 4 in Appendix F).

R1: Other works on beyond worst case analysis We feel that any thorough discussion of works on “beyond worst-case analysis” would have to span a very wide range of topics, which is beyond the scope of this paper. In the final version, we will reference Roughgarden [2019] which provides a broader discussion on the area. If the reviewer has any other works in mind which are particularly relevant, we will definitely consider adding them.

R2: Suggested additional related works on matrix approximation The list of published papers on algorithms for matrix approximation is indeed extensive. We thank the reviewer for the additional pointers, as they are relevant to our work. In particular, we believe that extending our results to the ℓ_p low-rank approximation setting is an interesting direction for future work.

R3: Connections with overparametrization While we are not aware of a formal connection between the multiple-descent phenomenon in column subset selection and overparameterization in machine learning, both settings have to do with obtaining a stable approximation of the data, and this stability can often be captured by the spectral properties of the problem.

R4: Difference between k-DPP and greedy on enuite2001 We also found the different behaviors of the approximation factor for the two methods intriguing. One explanation for this is that the expected approximation factor for a k-DPP is fully determined by the eigenvalues, whereas the performance of the greedy method depends on other characteristics of the data. However, it is not clear to us why this is most pronounced on the eunite2001 dataset.

R4: Correcting the effect of noisy eigenvalues We have not considered this, but it is indeed an interesting question. One simple observation is that if one chooses the subset size k small enough so that the noisy tail of the spectrum consists of only the eigenvalues with index sufficiently larger than k, then the effect of noise on the approximation factor should be minimal.

References

A. Deshpande, L. Rademacher, S. Vempala, and G. Wang. Matrix approximation and projective clustering via volume sampling. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, pages 1117–1126, Miami, FL, USA, January 2006.

T. Roughgarden. Beyond worst-case analysis. *Communications of the ACM*, 62(3):88–96, 2019.