We thank all of the reviewers for their time, careful reading, and valuable feedback.

Before discussing individual comments, we first want to address an issue raised by Reviewers 1 and 4 about **Assumption 2**. As we explain in the paper, Assumption 2 may not be necessary; pp. 7 – 9 in the supplement contain a preliminary convergence result in that direction. Also, it is *numerically verifiable*, as it need only hold for all t elapsed by subspace iteration. Indeed, we verify that the assumption holds for the datasets used in the experiments (see Section 1.3 of the supplement).

Reviewer 1. We understand that lines 141–146 contain a somewhat hand-wavy discussion, as the formally correct residual is given in eq. (9). However, this discussion is meant to provide practical alternatives for cases when we do not know all information (e.g., incoherence or the exact value of sep). This is meant to mirror standard stopping criteria; for example, criteria with respect to the spectral norm often assume that $\lambda_1 - \lambda_2 \approx \lambda_1$, and $\lambda_1 \approx \hat{\lambda}_1^t$.

The distinction between norms is another good point. The $\ell_{2\to\infty}$ norm is invariant to unitary transformations from the right, while the ℓ_∞ norm is not. Thus, the $\ell_{2\to\infty}$ norm allows us to interpret a subspace geometrically, as in the case of spectral clustering and this is a reason why the $\ell_{2\to\infty}$ norm is used in clustering analysis [3] (of course, when $r = 1$ the two distances coincide). Also, for $E \in \mathbb{R}^{d\times r}$, we have that $\|E\|_\infty \leq \|E\|_{2\to\infty} \leq \sqrt{r}\,\|E\|_\infty$, which means $\ell_{2\to\infty}$ is a good "proxy" for ℓ_∞. We can clarify these points when revising the paper.

Regarding coverage of prior work: as we mention in the manuscript, $\ell_{2\to\infty}$ norm convergence from a computational perspective is absent from the literature (and this is a contribution of our paper) and prior research mostly focuses on perturbation theory. We believe that we have done due diligence in citing such past work, but we are happy to include any additional references that the reviewers think are relevant.

A final clarification: in lines 109–112, we write that $\mathrm{dist}_{2\to\infty} \ll \mathrm{dist}_2$ in the "typical" case, as this is only false when the error is highly localized in just a few columns of the matrix; we will clarify this point in the paper.

Reviewer 2. One of the issues raised about the experiments may be due to a misunderstanding: t_{naive} checks $\|A\hat{V}_{:,j} - \hat{\lambda}_j \hat{V}_{:,j}\| \leq \varepsilon\hat{\lambda}_j$ for all j **up to** $j = r$ (we will fix this in the updated manuscript). Therefore, t_{naive} does **not** compute additional eigenvectors compared to t_{comp}. We apologize for the confusion caused by that omission. We also appreciate you bringing further issues to our attention, which will all be addressed in the updated manuscript, specifically:

- "computing $\|E\|_{2,2\to\infty}$" means computing $\|E\|_2$ and $\|E\|_{2\to\infty}$; it does not refer to a mixed norm.
- Figure 1 depicts how far apart the $\ell_2, \ell_{2\to\infty}$ distances and residuals are during an execution of subspace iteration, with ε representing different target accuracy levels, for a single experiment per configuration.
- In Figure 2, the quantities from Eq. (11) are hypothesized **rates** of convergence, rather than approximations to the distance (e.g. in the sense of Eq. (10)), so we would prefer to leave this caption as is.

Reviewer 3. Limiting our experiments to eigenvector-based methods was intended, as the focus of the paper is on theory and stopping criteria for eigensolvers and we want the experiments to highlight how our methods accelerate computations of these existing methods. There are certainly alternatives to eigenvector-based methods, and we will make sure to include this point when revising the paper. However, there are already several papers that discuss the relative merits of these methods (e.g., of graph neural networks and spectral clustering) on downstream tasks. Thus, we think it is more productive to cite this work, rather than duplicating the analysis. There are a few options for taking these ideas further; for example, one could attempt a similar analysis for Krylov methods, such as the Lanczos algorithm. On the other hand, more applications could benefit computationally from adopting appropriate stopping criteria (e.g., [4]) or analyzing algorithms from the $2 \to \infty$ perspective [1].

Reviewer 4. We understand that spectral methods are not used in some engineering settings. However, they are still the basis of fundamental algorithms in machine learning and data science (e.g., PCA, factor analysis, robust statistics [2]) and an active area of applied research in technology companies such as Google[1] and StitchFix.[2]

References

[1] M. Boedihardjo, S. Deng, and T. Strohmer. A Performance Guarantee for Spectral Clustering. *arXiv:2007.05627*, 2020.

[2] I. Diakonikolas and D. M. Kane. Recent advances in algorithmic high-dimensional robust statistics. *arXiv:1911.05911*, 2019.

[3] V. Lyzinski, D. Sussman, M. Tang, A. Athreya, and C. Priebe. Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding. *arXiv:1310.0532*, 2013.

[4] E. Nathan, G. Sanders, J. Fairbanks, V. E. Henson, and D. A. Bader. Graph ranking guarantees for numerical approximations to katz centrality. *Procedia Computer Science*, 2017. International Conference on Computational Science.

[1] https://ai.googleblog.com/2020/05/understanding-shape-of-large-scale-data.html

[2] https://www.wired.com/story/the-style-maven-astrophysicists-of-silicon-valley