We are glad the reviewers appreciate the theoretical contributions and the claims of the empirical experiments. We remark that these theoretical improvements are the primary contributions of the paper; the experiments were meant primarily to corroborate. We believe that the design of squared / asymmetric linear system solver based on variance reduction is intrinsically interesting. That it provides a new powerful primitive (and alternative to ridge regression) that enables nearly-linear runtimes for solving more complex problems (e.g. PCP / PCR) further demonstrates its utility and is the main result of this paper which the empirical experiments merely corroborate.

Reviewer 2 Thank you for the review; we are glad you found our result interesting. Below, we address in detail the concern you raised. With this concern addressed, we hope this elevates your view of the paper.

Proof of statement: $(M + M^T)/2 \succeq \mu I$ *if and only if* $x^\top M x \geq \mu x^T x$

Proof. Since $x^\top M x$ is a scalar by linearity we have

$$x^\top M x = x^\top M^\top x = \frac{1}{2}(x^\top M x + x^\top M^\top x) = x^\top \left(\frac{M^\top + M}{2}\right) x \,.$$

The claim follows as $(M + M^\top)/2 \succeq \mu I$ if and only if $x^\top((M + M^\top)/2)x \geq \mu x^\top x$ for all x by definition of $\succeq$. $\square$

We will be sure to add the proof of this fact in the full version.

Reviewer 4 Thank you for the kind review and for recognizing the importance of the problem and the novelty of our technique. We do hope both the rational approximation idea and the squared / asymmetric system solver we develop would inspire and assist further development in principle component analysis and linear system solving. We will correct all the typographical problems you mentioned in the final version of the paper.

Some additional discussion of rLanczos and sLanczos be added to main text if possible.

Thanks for this kind suggestion! We will add a few sentences giving a high level description of both algorithms. We were unable to demonstrate them in greater detail in the main paper due to the page limit. The particular details of *rLanczos* and *sLanczos* including theoretical runtime and implementation are included in Appendix F.1 and F.2; both methods rely on the key idea of solving squared system as subroutines discussed in Section 3.

Reviewer 5 Thank you for the kind review; we are pleased that you found the technique interesting. We do view this method for solving certain asymmetric systems as a variant of SVRG, and hope it will inspire and assist future improvement in obtaining faster algorithms on general linear system solving and other optimization tasks. We will address all the minor comments you pointed out in the final version of the paper.

The experiments look well executed, ..., would be nice to see a real-world application.

We believe the major contribution of the paper lies in its theoretical side, and we designed experiments on synthetic data primarily to give a clearer view of when and by how much our algorithm works better, confirming the theoretical results. We choose synthetic data because it gives more flexibility in controlling the patterns (eigenvalue distribution, eigen-gap, etc.). We agree that further empirical experiments on real data would be beneficial - it is a very interesting direction that we wish to explore in future work.

"had superlinear running times" should be sub-linear?

This is a good question. The previous methods can be superlinear because the dominating term in their runtime are either $k \cdot \mathrm{nnz}(A)$ or $\mathrm{nnz}(A)/\gamma$ where k is the number of eigenvalues above the threshold λ and γ is the eigen-gap around λ. If k gets larger or γ gets smaller as problem size ($\mathrm{nnz}(A)$) gets bigger then both methods can be superlinear. We will revise the introduction to clarify this in our final version of the paper.