Rebuttal

We thank all reviewers for your positive and insightful feedback. Below are our answers to your questions.

R1

"Your main result is given with high probability, in expectation. Could you elaborate with some discussion on what that guarantee means, and what it does not in terms of the actual convergence of the algorithm? In my opinion that's the only piece of discussion missing from an otherwise great paper."

Thanks for pointing this out: In short, $\mathcal{G}_t$ in Theorem 1 can be interpreted as a "success event" (the event that the algorithm's iterates up to time t all stay within the "basin of attraction"). Our analysis shows that $\mathcal{G}_t$ holds with high probability, and the expectation over Δ^t is restricted to this success event.

The same type of guarantee was used in [9] (Balsubramani et al).

Due to space limit, we have moved the definition of $\mathcal{G}_t$ to the Appendix before submission. We will add an explanation on this topic to our paper.

R2

"However, the significance of this comparison would be strengthened by considering runtime, not just epochs."

This is indeed a good point and will be very helpful for practitioners who would like to try out this method. We will add experiment that compares runtime.

R3

"(1) Why was the VR-PCA method not included in the simulation study?"

Our goal for the simulation study was to validate our theoretical result about Matrix Krasulina (exponential convergence rate on low-rank data), to see how its convergence rate is affected by both the intrinsic and the data dimension, and also to check whether the algorithm attains fast convergence (if not exactly exponential) when the data is effectively low-rank, to empirically test out our conjecture about the algorithm.

"(2) It would have been interesting to see some computation times."

Thanks for pointing this out. We'll add experiments to compare runtime as well.