DIFFERENTIALLY PRIVATE PRECONDITIONED POWER ITERATION METHOD FOR PRINCIPAL COMPONENT ANALYSIS THAT ADAPTS TO THE COHERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

In Principal Component Analysis (PCA), we are given as input a matrix $A \in \mathbb{R}^{n \times d}$ and an integer $k \leq d$, and the goal is to find the top k right-singular vectors of A. We consider the setting where each row of A is the private data of an individual, and we design an (ϵ, δ)-differentially private algorithm whose utility guarantee improves with the coherence of the input matrix. Prior works were either restricted to the case $k = 1$ or ensured significantly weaker privacy protections. Our algorithm is based on a preconditioned variant of the power iteration method that is designed to significantly reduce the noise needed to ensure privacy.

1 INTRODUCTION

Principal component analysis (PCA) is a fundamental tool for dimensionality reduction, data visualization, and low-rank approximation (Blum et al., 2020). Given a data matrix $A \in \mathbb{R}^{n \times d}$, PCA finds a low-dimensional subspace that captures as much of the variation in the data as possible. If $A = U\Sigma V^\top$ is the singular value decomposition of A, the optimal rank-k subspace is spanned by the top k right singular vectors $V_k = [v_1 \ \cdots \ v_k]$. It is also the rank k subspace that maximizes the total squared lengths of the projections of the data.

When the rows of A contain information about individuals, releasing a subspace computed from A can reveal information about a particular row. This motivates differentially private algorithms for PCA. We consider row-level privacy, where each row represents the data of one individual. We assume that every row has ℓ_2 norm at most 1, and two matrices are adjacent if one row is replaced. Normalizing the lengths of the rows is natural because we would like to measure the correlation with the entire dataset and no single data point should have an outsize contribution. The goal is to estimate V_k while protecting the privacy of every individual row. In this model, Dwork et al. (2014) give a private rank-k PCA algorithm whose subspace error scales with $\widetilde{O}\left(\sqrt{d}\right)$. Thus, the error can be large when the number of dimensions d is large. This motivates utility guarantees that exploit additional structure of the input matrix. One such structural parameter is coherence, which is related to the largest absolute entries of the left singular vectors U. Intuitively, coherence captures the difficulty in privatizing the singular vectors because it measures the maximum correlation between the singular vectors and individual data rows. The more the singular vectors reveal about one row, the more perturbation we need to protect privacy. On the other hand, for many natural datasets, such as when the data is close to i.i.d. data generated from a generative model, the coherence can be shown to be small. Previous work has shown that coherence-dependent guarantees can substantially improve on worst-case bounds on low-coherence data (Hardt and Roth, 2012; 2013; Hardt and Price, 2014; Balcan et al., 2016; Nicolas et al., 2025; d'Orsi and Novikov, 2026; Nguyen et al., 2026).

However, most earlier coherence-dependent guarantees were obtained under privacy models different from the row-level ℓ_2 model considered in our paper. For example, Hardt and Roth (2013) consider entry-level adjacency, under which neighboring matrices differ in a single entry by at most 1. Subsequent rank-k results use structured adjacency models that, when restricted to a one-row change, impose a ℓ_1-norm restriction on the changed row (Nicolas et al., 2025; d'Orsi and Novikov, 2026). These models do not cover all neighboring datasets in our setting, since a row a satisfying $\|a\|_2 \leq 1$ can have $\|a\|_1$ as large as $\sqrt{d}$. Nguyen et al. (2026) recently gave the first coherence-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Comparison of deterministic rank-k PCA guarantees. The error is $\|(I - \widehat{X}\widehat{X}^\top)V_k\|_2^2 = \sin^2(\theta_{\max})$ where $\theta_{\max}$ is the largest principal angle between the column spaces of $\widehat{X}$ and V_k. Poly-logarithmic factors in n, d, $\frac{1}{\epsilon}$, $\frac{1}{\delta}$ are suppressed. We let $\Delta_k := \sigma_k^2 - \sigma_{k+1}^2$.

	Algorithm	Privacy model	$\sin^2$-Error	Condition
Rank k	Hardt and Price (2014)	Entry	$\widetilde{O}\left(\frac{k^3\sigma_k^4 \max\{n\|U\|_\infty^2, d\|V\|_\infty^2\}}{\epsilon^2\Delta_k^3}\right)^2$	None
	d'Orsi and Novikov (2026)	See note[1]	$\widetilde{O}\left(\frac{\Delta_{dn}^2\sigma_k^2 k \max\{n\|U\|_\infty^2, d\|V\|_\infty^2\}}{\epsilon^2\Delta_k^2}\right)^2$	None
	Dwork et al. (2014)	Row-ℓ_2	$\widetilde{O}\left(\frac{d}{\epsilon^2\Delta_k^2}\right)$	$\Delta_k = \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon}\right)$
	Our algorithm	Row-ℓ_2	$\widetilde{O}\left(\frac{dk\sigma_k^2\|U\|_\infty^2}{\epsilon^2\Delta_k^2}\right)$	$\Delta_k = \widetilde{\Omega}\Big(\frac{\sqrt{d}}{\epsilon} + \left(\frac{\sigma_k\sqrt{k}}{\epsilon}\right)^{2/3} + \left(\frac{k\sigma_k^2\|U\|_\infty}{\epsilon}\right)^{2/3}\Big)$
Rank 1	Nguyen et al. (2026)	Row-ℓ_2	$\widetilde{O}\left(\frac{\left(\min\{4n, d\sigma_1^2\} + \frac{\sigma_1^4}{\Delta_1} + d\Delta_1\right)\|U\|_\infty^2}{\epsilon^2\Delta_1^2}\right)^3$	$\Delta_1 = \widetilde{\Omega}\Big(\left(\frac{\sigma_1^2\|U\|_\infty}{\epsilon}\right)^{2/3} + \left(\frac{\sigma_1}{\epsilon}\right)^{2/3}\Big)^3$
	Our algorithm	Row-ℓ_2	$\widetilde{O}\left(\frac{d\sigma_1^2\|U\|_\infty^2}{\epsilon^2\Delta_1^2}\right)$	$\Delta_1 = \widetilde{\Omega}\Big(\frac{\sqrt{d}}{\epsilon} + \left(\frac{\sigma_1^2\|U\|_\infty}{\epsilon}\right)^{2/3} + \left(\frac{\sigma_1}{\epsilon}\right)^{2/3}\Big)$

adaptive guarantee under row-level ℓ_2 privacy. However, their algorithm and analysis are specific to estimating a single leading singular vector, leaving the analogous rank-k problem open.

In this work, we give a coherence-adaptive algorithm for estimating the top-k right singular subspace under row-level ℓ_2 differential privacy. Our result extends the coherence-adaptive utility guarantee of Nguyen et al. (2026) from the leading singular vector to the top-k singular subspace while retaining the same privacy model. Our algorithm is based on a preconditioned and clipped variant of block power iteration. It does not require the coherence or the ratio $\sigma_{k+1}^2/\sigma_k^2$ as input, and its utility improves when the input has low coherence. The prior work is also suboptimal when the ratio $\sigma_{k+1}^2/\sigma_k^2$ is very close to 1 whereas our algorithm remains near optimal for all regimes of the ratio.

1.1 Our contributions

Let $\Delta_k := \sigma_k^2 - \sigma_{k+1}^2$ and let $\left\|(I - \widehat{X}\widehat{X}^\top)V_k\right\|_2$ denote the error of an output $\widehat{X} \in \mathbb{R}^{d\times k}$ with orthonormal columns. Let $a_i^\top$ denote the ith row of A, and recall that $\|a_i\|_2 \le 1$. We write $U_k =$

[1]Let $E := \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix} - \begin{pmatrix} 0 & A' \\ A'^\top & 0 \end{pmatrix}$. The matrices A and A' are considered Δ_{dN}-adjacent when $\sqrt{\sum_{i,j}\left|(EE^\top)_{ij}\right|} \le \Delta_{\mathrm{dN}}$. For a replacement of row a by row a', $\sqrt{\|EE^\top\|_1} = \sqrt{\|a-a'\|_2^2 + \|a-a'\|_1^2}$

[2]We express the relevant coherence parameters using $\max\{n\|U\|_\infty^2, d\|V\|_\infty^2\}$. In particular, the rank-k coherence of d'Orsi and Novikov (2026) is $\mu_k(A) = \max\left\{\frac{n}{k}\|U_kU_k^\top\|_{\max}, \frac{d}{k}\|V_kV_k^\top\|_{\max}\right\}$, where $\|M\|_{\max} := \max_{i,j}|M_{ij}|$. Since $\|U_kU_k^\top\|_{\max} \le k\|U\|_\infty^2$ and $\|V_kV_k^\top\|_{\max} \le k\|V\|_\infty^2$, we have $\mu_k(A) \le \max\left\{n\|U\|_\infty^2, d\|V\|_\infty^2\right\}$.

[3]The bounds depend only on the maximum entry of the first singular vector in U, not $\|U\|_\infty$

$U_{:,1:k}$, $\Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$, $\|U\|_\infty = \max_{i,j} |U_{ij}|$. For the ith standard basis vector e_i, we have $\left\|e_i^\top U_k\right\|_2 \le \sqrt{k}\,\|U\|_\infty$. Our main result is the following:

Theorem 1. *Let $\epsilon, \delta, \beta \in (0,1)$, and suppose that every row of A has ℓ_2 norm at most 1. There is an (ϵ, δ)-differentially private algorithm that, under the condition*

$$\Delta_k \ge \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon} + \left(\frac{\sigma_k\sqrt{k}}{\epsilon}\right)^{2/3} + \left(\frac{k\sigma_k^2\,\|U\|_\infty}{\epsilon}\right)^{2/3}\right),$$

outputs $\widehat{X} \in \mathbb{R}^{d\times k}$ satisfying, with probability at least $1-\beta$,

$$\left\|(I - \widehat{X}\widehat{X}^\top)V_k\right\|_2^2 \le \left(\frac{\beta\delta\epsilon}{n}\right)^2 + \widetilde{O}\left(\frac{dk\sigma_k^2\,\|U\|_\infty^2 + 1}{\epsilon^2\Delta_k^2}\right).$$

The algorithm does not require $\|U\|_\infty$, Δ_k, or $\sigma_{k+1}^2/\sigma_k^2$ as input.

1.2 Comparison with prior works

We additionally write $\|V\|_\infty := \max_{i,j} |V_{ij}|$. We summarize our results in comparison with prior work in Table 1.

Comparison with the row-ℓ_2 baseline. The most direct general rank-k comparison is the Analyze Gauss algorithm of Dwork et al. (2014), since it provides privacy under the same row-ℓ_2 model. Our result replaces the worst-case dependence on d by a dependence on the coherence of the input. Subject to the Δ_k condition in Theorem 1, our bound improves on Analyze Gauss when $k\sigma_k^2\|U\|_\infty^2$ is small. Analyze Gauss does not require low coherence and can give the stronger guarantee outside this regime.

Comparison under other privacy models. The results of Hardt and Price (2014); d'Orsi and Novikov (2026) provide coherence-dependent guarantees for general rank k, but they use privacy models different from ours. Hardt and Price (2014) consider entry-level adjacency. Any replacement between two rows of ℓ_2 norm at most 1 can be implemented using at most $2d$ entry changes by passing through the zero row. Group privacy therefore converts their guarantee to our row-level model by setting $\epsilon_{\text{ent}} = \Theta(\epsilon/d)$ and adjusting δ_{ent} accordingly. So Hardt and Price (2014) have a converted $\sin^2$ bound of:

$$\widetilde{O}\left(\frac{k^3\sigma_k^4 \max\left\{n\|U\|_\infty^2, d\|V\|_\infty^2\right\}}{(\epsilon/d)^2\,\Delta_k^3}\right) = \widetilde{O}\left(\frac{d^2k^3\sigma_k^4 \max\left\{n\|U\|_\infty^2, d\|V\|_\infty^2\right\}}{\epsilon^2\Delta_k^3}\right)$$

For d'Orsi and Novikov (2026) model, the replacement of row a by a' will result in: $\sqrt{\|EE^\top\|_1} = \sqrt{\|a-a'\|_2^2 + \|a-a'\|_1^2} = O\left(\sqrt{d\|a-a'\|_2^2}\right) = O\left(\sqrt{d}\right)$. Thus it suffices to set $\Delta_{dn} = O\left(\sqrt{d}\right)$, leading to $\sin^2$ error of:

$$\widetilde{O}\left(\frac{dk\sigma_k^2 \max\left\{n\|U\|_\infty^2, d\|V\|_\infty^2\right\}}{\epsilon^2\Delta_k^2}\right)$$

The above bound is a factor n larger than ours. Note that while this bound does not require any condition, it is only meaningful if the fraction inside $\widetilde{O}$ is less than 1, implying that the gap $\Delta_k \ge \Omega\left(\sqrt{dk\sigma_k^2 \max\left\{n\|U\|_\infty^2, d\|V\|_\infty^2\right\}}/\epsilon\right)$. This can be significant e.g. when $k=1$, the incoherence $n\,\|U\|_\infty^2 = \widetilde{O}(1)$ (the minimum possible value and what we obtain with i.i.d. data), we still need $\Delta_k \ge \sigma_1\sqrt{d}/\epsilon$. In contrast, the constraint for ours in the same setting is $\Delta_k = \widetilde{\Omega}\Big(\frac{\sqrt{d}}{\epsilon} + \big(\frac{\sigma_1}{\epsilon}\big)^{2/3} + \frac{\sigma_1^{4/3}}{n^{1/3}\epsilon^{2/3}}\Big)$, which is smaller.

Comparison for $k=1$. The result of Nguyen et al. (2026) uses the same row-ℓ_2 privacy model as ours and is the most relevant comparison when $k=1$. Specializing our result to $k=1$ gives a comparable coherence-dependent guarantee. Our requirement is more stringent as we need $\Delta_1 = \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon}\right)$ but our bound is better whenever the gap is small $\Delta_1 < \frac{\sigma_1^2}{d}$.

1.3 Our approach and techniques

Let $a_i^\top$ denote the ith row of A, and recall that $\|a_i\|_2 \le 1$. We write $U_k = U_{:,1:k}$, $\Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$, $\|U\|_\infty = \max_{i,j} |U_{ij}|$. For the ith standard basis vector e_i, we have $\left\|e_i^\top U_k\right\|_2 \le \sqrt{k}\,\|U\|_\infty$. Thus, when $\|U\|_\infty$ is small, every row of U_k has small norm.

A standard noisy power method (Hardt and Price, 2014) maintains a matrix $X^{(t)} \in \mathbb{R}^{d\times k}$ with orthonormal columns and performs an update of the form

$$\begin{aligned} Z^{(t+1)} &= A^\top A X^{(t)} + \text{Gaussian noise}, \\ X^{(t+1)} R^{(t+1)} &\text{be QR-decomposition of } Z^{(t+1)} \end{aligned}$$

Since $A^\top A X^{(t)} = \sum_{i=1}^n a_i \left(a_i^\top X^{(t)}\right)$, the contribution of row a_i to the update has Frobenius norm $\left\|a_i \left(a_i^\top X^{(t)}\right)\right\|_F = \|a_i\|_2 \left\|a_i^\top X^{(t)}\right\|_2$.

A direct implementation under row-level privacy would therefore calibrate the noise using bounds on these row contributions. The difficulty is that these contributions are not controlled by coherence alone. To see the difficulty, suppose that $X^{(t)}$ has already reached the target V_k, so that $X^{(t)} = V_k$. The contribution of row a_i then has norm

$$\left\|a_i \left(a_i^\top V_k\right)\right\|_F = \left\|a_i \left(e_i^\top A V_k\right)\right\|_F = \left\|a_i e_i^\top U_k \Sigma_k\right\|_F = \|a_i\|_2 \left\|e_i^\top U_k \Sigma_k\right\|_2$$

Low coherence controls $\left\|e_i^\top U_k\right\|_2 \le \sqrt{k}\,\|U\|_\infty$, but the direct update contains

$$\left\|e_i^\top U_k \Sigma_k\right\|_2 = \left(\sum_{j=1}^k \sigma_j^2 U_{ij}^2\right)^{1/2} \le \sigma_1 \left\|e_i^\top U_k\right\|_2 \le \sigma_1 \sqrt{k}\,\|U\|_\infty$$

This gives $\left\|a_i \left(a_i^\top V_k\right)\right\|_F \le \sigma_1 \sqrt{k} \|U\|_\infty$. Thus, the Gaussian noise adjusted to the above upper bound inherits a dependence on σ_1. On the other hand, in the normal power method, the subspace error contracts at rate $\sigma_{k+1}^2/\sigma_k^2$. This creates the main difficulty in analyzing the above power method: the bound on Gaussian noise can depend on σ_1, whereas the recurrence for the subspace error depends on σ_k and σ_{k+1}. This is not an issue for the prior work for $k = 1$ but it is a major issue for our setting.

Preconditioner. Our main idea is to remove the Σ_k term on the contribution of row a_i, using the preconditioner method. Our new update has this form:

$$\begin{aligned} H^{(t)} &= X^{(t)\top} A^\top A X^{(t)} + \text{symmetric Gaussian noise}, \\ \widetilde{H}^{(t)} &= H^{(t)} \text{ with every eigenvalue below } \gamma \text{ replaced by } \gamma, \\ B^{(t)} &= \left(\widetilde{H}^{(t)}\right)^{-1/2}, \\ Z^{(t+1)} &= A^\top A X^{(t)} B^{(t)} + \text{Gaussian noise}, \\ X^{(t+1)} R^{(t+1)} &\text{be QR-decomposition of } Z^{(t+1)} \end{aligned}$$

where the second step is to ensure that the inverse of $\widetilde{H}^{(t)}$ exists and $B^{(t)}$ is defined. The parameter γ is set roughly the same as the magnitude of the symmetric Gaussian noise so that we remove the artifact from the noise and the components smaller than the noise.

To explain the role of the preconditioner $B^{(t)}$, first suppose that the Gaussian noise is absent and $\widetilde{H}^{(t)} = H^{(t)} = X^{(t)\top} A^\top A X^{(t)}$. Provided that $X^{(t)\top} A^\top A X^{(t)}$ is positive definite, we have $B^{(t)} = \left(X^{(t)\top} A^\top A X^{(t)}\right)^{-1/2}$, and the columns of $A X^{(t)} B^{(t)}$ are orthonormal. Moreover, since $B^{(t)}$ is invertible, $\operatorname{range}\left(A^\top A X^{(t)} B^{(t)}\right) = \operatorname{range}\left(A^\top A X^{(t)}\right)$. Thus, in this idealized iteration, the preconditioner does not change the next subspace produced by the block power method. Instead, it changes the basis in which $a_i^\top X^{(t)} B^{(t)}$ is measured.

At $X^{(t)} = V_k$, we have $B^{(t)} = \left(V_k^\top A^\top A V_k\right)^{-1/2} = \Sigma_k^{-1}$ and therefore $a_i^\top V_k B^{(t)} = e_i^\top U_k$. The contribution of row a_i consequently has norm

$$\left\|a_i \left(a_i^\top V_k B^{(t)}\right)\right\|_F = \|a_i\|_2 \left\|e_i^\top U_k\right\|_2$$

which is controlled by the coherence and no longer contains $\sigma_1, \ldots, \sigma_k$. When the iterate is close to V_k and $B^{(t)}$ is computed from the noisy $k \times k$ matrix, this identity does not hold exactly. Our analysis instead shows that the part of $a_i^\top X^{(t)} B^{(t)}$ from V_k remains bounded by a constant multiple of a coherence-related quantity. The remaining part of $a_i^\top X^{(t)} B^{(t)}$ from $V_{>k}$ is controlled through a separate row-wise analysis. This allows the final error bound to depend on σ_k, rather than on all of the leading singular values.

Clipping. The preceding bounds hold with high probability when the iterates remain close to V_k. Differential privacy, however, requires a sensitivity bound that holds for every input. For a row vector $y \in \mathbb{R}^{1\times k}$ and a radius $R > 0$, define $\mathrm{Clip}_R(y) = y \min\left\{1, \frac{R}{\|y\|_2}\right\}$,with $\mathrm{Clip}_R(0) = 0$. Our update has the form:

$$Z^{(t+1)} = \sum_{i=1}^{n} a_i \mathrm{Clip}_R\left(a_i^\top X^{(t)} B^{(t)}\right) + \text{Gaussian noise}.$$

Since $\|a_i\|_2 \le 1$, $\left\|a_i \mathrm{Clip}_R\left(a_i^\top X^{(t)} B^{(t)}\right)\right\|_F \le R$. Under replacement adjacency, the sensitivity of the sum is therefore at most $2R$, and the Gaussian noise can be adjust to this sensitivity.

Clipping supplies a worst-case sensitivity bound but is never active with high probability. In particular, under the assumptions of our main result, our analysis gives a radius $R = \widetilde{O}\left(\sqrt{k}\,\|U\|_\infty\right)$ for which $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \le R$ for every row and every iteration, with high probability.

Private initialization. The preconditioner and row-wise bounds described above apply only when the current iterate has sufficient overlap with the target subspace. We obtain such an initialization using the Gaussian mechanism of Dwork et al. (2014). In particular, we add symmetric Gaussian noise to $A^\top A$ and let $X^{(0)}$ contain the top k eigenvectors of the resulting matrix. Under an initial condition, this initialization satisfies $\left\|\left(I - X^{(0)} X^{(0)\top}\right) V_k\right\|_2 \le 1/4$.

Private selection of R and T. The appropriate choices of R and T depend on private properties of the input: R depends on the coherence, while T depends on $\sigma_{k+1}^2/\sigma_k^2$. We consider public powers-of-two grids for R and T and use the private selection method of Papernot and Steinke (2022) to choose among the resulting outputs. The construction and analysis of this procedure are given in the algorithm section. Thus, our algorithm does not require the coherence or $\sigma_{k+1}^2/\sigma_k^2$ as input.

Analysis technique. The quality of the solution $X^{(t)}$ is evaluated via the ratio between its projection on the top k subspace $C = V_k^\top X^{(t)}$ and its projection on the bottom $d-k$ subspace $D = V_{>k}^\top X^{(t)}$. Let $E^{(t)} = D^{(t)}\left(C^{(t)}\right)^{-1}$. We track $\left\|E^{(t)}\right\|_2$ over the iterations, ideally showing that it decreases exponentially. If no clipping happens, we prove the recurrence $E^{(t+1)} = \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + noise\right)(I + noise)^{-1}$ where the noise terms come from the Gaussian noise in the update for $Z^{(t+1)}$. Without noise, we can see that $\left\|\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}\right\|_2 \le \frac{\sigma_{k+1}^2}{\sigma_k^2}\left\|E^{(t)}\right\|_2 = (1-\kappa_k)\left\|E^{(t)}\right\|_2$, indeed decreasing exponentially with iterations. Naively bounding the noises per iteration would create an error term linear in the number of iterations, which is too large. In different iterations, we add independent noises so intuitively, these noises are nearly orthogonal to each others and have a lot of cancellation. To exploit the cancellation, we need to perform a global analysis involving all iterations together. However, the dependence on the noise in each iteration is complex with a term in the numerator and a term in the denominator. Using $(I+F)^{-1} = I - F(I+F)^{-1}$ for any noise matrix F, the noise contribution is split into two parts: one part being added up linearly across all iterations, which is amenable to a martingale analysis, and one complicated part, which is bounded separately per iteration. The second part is small so even though the bound might be suboptimal, it is sufficiently small to not matter. This analysis is possible only if we never clip. First we show that clipping does not happen for the initial solution using Gaussian concentration. Then we use induction to show that clipping does not happen for the rest of the iterations.

2 Preliminaries

Let $A \in \mathbb{R}^{n\times d}$ be the input data matrix whose rows a_i are datapoints in $\mathbb{R}^d$. We assume that $n \ge d$ and each row a_i of A has length $\|a\|_2 \le 1$. As noted by Dwork et al. (2014), the latter assumption

Algorithm 1 Algorithm for constructing an initial solution

1: **Input**: Matrix $A \in \mathbb{R}^{n \times d}$, privacy parameters $\epsilon_{\text{init}}, \delta_{\text{init}}$
2: Let $\Delta_{\text{init}} = 2\sqrt{2\log(2/\delta_{\text{init}})}/\epsilon_{\text{init}}$.
3: Let $\widehat{M} = A^\top A + \Delta_{\text{init}} \cdot \text{sym}\left(G^{(\text{init})}\right)$ where $G^{(\text{init})} \sim \mathcal{N}(0,1)^{d \times d}$ and $\text{sym}(G) = \frac{1}{2}\left(G + G^\top\right)$
4: Let $X^{(0)}$ be a $d \times k$ matrix whose columns are the top k eigenvectors of $\widehat{M}$
5: Return $X^{(0)}$

is necessary for achieving non-trivial utility guarantees and it is a natural restriction for PCA, since otherwise the solution could be dominated by a single datapoint with very large norm and it could fail to capture the underlying structure shared by most of the datapoints.

Notation. Let $A = U\Sigma V^\top$ be the thin singular value decomposition (SVD) of A, where $U \in \mathbb{R}^{n \times d}$ is the matrix whose columns are the left-singular vectors $u_1, \ldots, u_d$, $V \in \mathbb{R}^{d \times d}$ is the matrix whose columns are the right-singular vectors $v_1, \ldots, v_d$, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_d) \in \mathbb{R}^{d \times d}$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$ are the singular values. Let $U_k \in \mathbb{R}^{n \times k}$ and $V_k \in \mathbb{R}^{d \times k}$ be the matrices whose columns are the top-k left-singular and right-singular vectors, respectively. Let $\Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k)$. We assume that there is a positive eigengap $\Delta_k := \sigma_k^2 - \sigma_{k+1}^2 > 0$.

We let $\|U\|_\infty := \max_{i,j} |U_{ij}|$, and we refer to $\|U\|_\infty$ as the **coherence** of the input matrix A. Note that $U_{ij} = \langle a_i, v_j \rangle / \sigma_j$ and thus $\|U\|_\infty$ measures the maximum correlation between the datapoints (the rows of A) and the right-singular vectors. The coherence value is not known to the algorithm.

We use the $\widetilde{O}/\widetilde{\Omega}$ notation to suppress factors that are poly-logarithmic in n, d, $1/\epsilon$, $1/\delta$, and $1/\beta$, where n and d are the input dimensions, ϵ and δ are the privacy parameters, and β is the failure probability.

Clipping. For a row vector $y \in \mathbb{R}^{1 \times k}$ and a radius $R > 0$, we define the clipping operation:

$$\text{Clip}_R(y) := y \cdot \min\left\{1, \frac{R}{\|y\|_2}\right\}$$

with the convention that $\text{Clip}_R(0) = 0$.

Differential privacy. We consider the privacy model where each row of the input matrix A is the private data of an user. More precisely, two input matrices A and A' are neighboring inputs if A and A' differ in exactly one row. A randomized algorithm $\mathcal{M}$ is (ϵ, δ)-differentially private ((ϵ, δ)-DP) if, for every pair of neighboring inputs A, A' and every measurable subset of possible outcomes $\mathcal{O}$, we have

$$\Pr[\mathcal{M}(A) \in \mathcal{O}] \leq e^\epsilon \Pr[\mathcal{M}(A') \in \mathcal{O}] + \delta.$$

Our algorithm and analysis use the Gaussian mechanism and advanced composition, and we give the definitions and theorems in Section A.

3 Our Algorithm

In this section, we give an overview of our algorithm and analysis. Our main algorithm is shown in Algorithm 3. The algorithm uses the algorithm of Dwork et al. (2014) to construct an initial solution (Algorithm 1). Our main optimization algorithm is shown in Algorithm 2. This algorithm takes as input the initial solution constructed by Algorithm 1 together with a clipping radius R and an iteration count T, and it iteratively constructs a solution using a private variant of a preconditioned power method as discussed in the introduction. In Section C in the Appendix, we show how to privately select the hyperparameters R and T using the private hyperparameter-selection method of Papernot and Steinke Papernot and Steinke (2022).

We now give an overview of our analysis. Due to space constraints, we defer the complete analysis to Section B in the Appendix.

Privacy analysis. Our algorithm uses the Gaussian mechanism to privatize the updates, and we show that the algorithm is (ϵ, δ)-DP via standard applications of the Gaussian mechanism and the advanced composition theorem.

Algorithm 2 Private Preconditioned Power Method

Input: Matrix $A \in \mathbb{R}^{n\times d}$ with rows a_i, initial solution $X^{(0)}$,number of iterations T, $k \le d$, per-iteration privacy parameters $\epsilon_{\text{iter}}, \delta_{\text{iter}}$, failure probability β, clipping radius R
Let $\Delta = 2\sqrt{2\log(2/\delta_{\text{iter}})}/\epsilon_{\text{iter}}$, $\gamma = 4\Delta\left(2\sqrt{k} + \sqrt{2\log(8T/\beta)}\right)$
for $t = 0 \dots T-1$
    Let $H^{(t)} = X^{(t)\top} A^\top A X^{(t)} + \Delta \cdot \text{sym}\left(G^{(t)}\right)$ where $G^{(t)} \sim \mathcal{N}(0,1)^{k\times k}$ and $\text{sym}(G) = \frac{1}{2}\left(G + G^\top\right)$
    Let $H^{(t)} = Q^{(t)}\Lambda^{(t)}Q^{(t)\top}$, $\widetilde{H}^{(t)} = Q^{(t)}\text{diag}\left(\max\left\{\Lambda^{(t)}_{11}, \gamma\right\}, \dots, \max\left\{\Lambda^{(t)}_{kk}, \gamma\right\}\right) Q^{(t)\top}$, $B^{(t)} = \left(\widetilde{H}^{(t)}\right)^{-1/2}$
    Let $Z^{(t+1)} = \sum_{i=1}^n a_i \,\text{Clip}_R\left(a_i^\top X^{(t)} B^{(t)}\right) + \Delta R \cdot \widetilde{G}^{(t)}$ where $\widetilde{G}^{(t)} \sim \mathcal{N}(0,1)^{d\times k}$
    Let $X^{(t+1)}R^{(t+1)} = Z^{(t+1)}$ be the QR-decomposition of $Z^{(t+1)}$
end for
return $X^{(T)}$

Algorithm 3 Main Algorithm

Input: Matrix $A \in \mathbb{R}^{n\times d}$, privacy parameters ϵ, δ, failure probability β
Set $\epsilon_{\text{init}} = \frac{\epsilon}{2}$ and $\delta_{\text{init}} = \frac{\delta}{2}$
Let $X^{(0)}$ be the solution returned by Algorithm 1 when run with $\epsilon_{\text{init}} = \frac{\epsilon}{2}$ and $\delta_{\text{init}} = \frac{\delta}{2}$
Set ϵ_{iter} and δ_{iter} so that Algorithm 2 is $\left(\frac{\epsilon}{2}, \frac{\delta}{2}\right)$-DP
Let $X^{(T)}$ be the solution returned by Algorithm 2 when run with ϵ_{iter} and δ_{iter}, and failure probability $\frac{\beta}{2}$
Return $X^{(T)}$

Utility analysis. We measure the utility of our algorithm via the error $\left\|\left(I - XX^\top\right)V_k\right\|_2$, where X is a solution with orthonormal columns and V_k is the matrix whose columns are the top-k right-singular vectors. We recall that this error is the sine of the maximum principal angle between the column spaces of X and V_k. We let $s_t := \left\|\left(I - X^{(t)}X^{(t)\top}\right)V_k\right\|_2$ be the error of solution $X^{(t)}$ constructed in iteration $t-1$ of our algorithm.

We analyze the error s_t via the following quantity. Let $C^{(t)} := V_k^\top X^{(t)}$, $D^{(t)} := V_{>k}^\top X^{(t)}$, and $E^{(t)} := D^{(t)}\left(C^{(t)}\right)^{-1}$ provided that $C^{(t)}$ is invertible. The initial solution $X^{(0)}$ constructed by Algorithm 1 satisfies $s_0 \le 1/4$, which guarantees that $C^{(0)}$ is invertible. Given this property, we show by induction that the algorithm maintains the invariant that $s_t < 1$, which ensures that $C^{(t)}$ is invertible (Lemma 13). We analyze the error s_t via the spectral norm $m_t := \left\|E^{(t)}\right\|_2$ of $E^{(t)}$, which upper bounds the error (i.e., we have $s_t \le m_t$).

In an ideal update with no clipping or privacy noise, we have $E^{(t+1)} = \Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}$ and thus

$$\left\|E^{(t+1)}\right\|_2 \le \left\|\Sigma_{>k}^2\right\|_2 \left\|E^{(t)}\right\|_2 \left\|\Sigma_k^{-2}\right\|_2 = \frac{\sigma_{k+1}^2}{\sigma_k^2}\left\|E^{(t)}\right\|_2 = (1-\kappa_k)\left\|E^{(t)}\right\|_2,$$

where $\kappa_k = \frac{\sigma_k^2 - \sigma_{k+1}^2}{\sigma_k^2}$. Thus, in this ideal case, the error decreases exponentially with the eigengap. For our private update, we show that, if the clipping is inactive in iteration t (i.e., $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \le R$ for all $i \in [n]$), then the following modified recurrence holds:

$$E^{(t+1)} = \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + L^{(t)}\right)\left(I + F^{(t)}\right)^{-1}, \tag{1}$$

where $L^{(t)}$ and $F^{(t)}$ are matrices arising from the Gaussian noise added, defined as follows: $F^{(t)} := \Delta R V_k^\top \widetilde{G}^{(t)}\left(K^{(t)}\right)^{-1}$ and $L^{(t)} := \Delta R V_{>k}^\top \widetilde{G}^{(t)}\left(K^{(t)}\right)^{-1}$, where $K^{(t)} := \Sigma_k^2 C^{(t)} B^{(t)}$. The main progress term $\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}$ is the same as in the ideal update, and its spectral norm decreases exponentially. Thus if we can show that the clipping remains inactive throughout the algorithm

and the contribution of the error terms is sufficiently small, we can show that we approximately preserve the exponential decrease in the error.

We show that the error terms can be bounded with high probability using Gaussian concentration and the properties of the preconditioner $B^{(t)}$ used by the main update. We have

$$\left\|F^{(t)}\right\|_2 \leq \Delta R \left\|V_k^\top \widetilde{G}^{(t)}\right\|_2 \left\|\left(K^{(t)}\right)^{-1}\right\|_2 \quad \left\|L^{(t)}\right\|_2 \leq \Delta R \left\|V_{>k}^\top \widetilde{G}^{(t)}\right\|_2 \left\|\left(K^{(t)}\right)^{-1}\right\|_2.$$

Since the Gaussian distribution is rotationally invariant, $V_k^\top \widetilde{G}^{(t)}$ is a $k \times k$ Gaussian matrix and $V_{>k}^\top \widetilde{G}^{(t)}$ is a $(d-k) \times k$ Gaussian matrix. Using Gaussian concentration results (Lemma 5), we can show that we have $\left\|V_k^\top \widetilde{G}^{(t)}\right\|_2 \leq \widetilde{O}\left(\sqrt{k}\right)$ and $\left\|V_{>k}^\top \widetilde{G}^{(t)}\right\|_2 \leq \widetilde{O}\left(\sqrt{d-k}+\sqrt{k}\right)$ with high probability. We can also show that $\left\|\left(K^{(t)}\right)^{-1}\right\|_2 \leq O\left(\frac{1}{\sigma_k}\right)$ (Lemma 18) as follows. We have $K^{(t)} := \Sigma_k^2 C^{(t)} B^{(t)}$. We can show that the preconditioner $B^{(t)}$ roughly whitens $\Sigma_k V_k^\top X^{(t)} = \Sigma_k C^{(t)}$, i.e., the singular values of $\Sigma_k C^{(t)} B^{(t)}$ are bounded from above and below by absolute constants (Lemma 17). This allows us to bound $\left\|\left(K^{(t)}\right)^{-1}\right\|_2 \leq \left\|\left(\Sigma_k C^{(t)} B^{(t)}\right)^{-1}\right\|_2 \left\|\Sigma_k^{-1}\right\|_2 = O(1/\sigma_k)$ since $\left\|\left(\Sigma_k C^{(t)} B^{(t)}\right)^{-1}\right\|_2 = O(1)$.

If the clipping is inactive, we can use the recurrence above together with the upper bounds on the noise terms described above to analyze the error progress. In order to obtain a stronger bound, we use a martingale argument to show that the contribution of some of the error terms is smaller. To simplify notation, for a matrix M, let us define $\mathcal{T}(M) := \Sigma_{>k}^2 M \Sigma_k^{-2}$. Thus the main recurrence (1) becomes:

$$E^{(t+1)} = \left(\mathcal{T}\left(E^{(t)}\right) + L^{(t)}\right)\left(I + F^{(t)}\right)^{-1}.$$

Using the identity $(I+F)^{-1} = I - F(I+F)^{-1}$, we can rewrite the recurrence as follows:

$$E^{(t+1)} = \mathcal{T}\left(E^{(t)}\right) + L^{(t)} \underbrace{- \left(\mathcal{T}\left(E^{(t)}\right) + L^{(t)}\right) F^{(t)} \left(I + F^{(t)}\right)^{-1}}_{:=R^{(t)}} = \mathcal{T}\left(E^{(t)}\right) + L^{(t)} + R^{(t)}.$$

The reason this is beneficial is that it separates the noise terms depending on $\widetilde{G}^{(t)}$ into the term $L^{(t)}$ that is linear in $\widetilde{G}^{(t)}$ and a more complex term $R^{(t)}$. The simple linear structure of the terms $L^{(t)}$ allows us to bound their contribution via a martingale argument. By unrolling the recurrence, we obtain

$$E^{(t)} = \mathcal{T}^t\left(E^{(0)}\right) + \sum_{j=0}^{t-1} \mathcal{T}^{t-1-j}\left(L^{(j)}\right) + \sum_{j=0}^{t-1} \mathcal{T}^{t-1-j}\left(R^{(j)}\right).$$

We upper bound the spectral norm of the middle term $\sum_{j=0}^{t-1} \mathcal{T}^{t-1-j}\left(L^{(j)}\right)$ via a martingale concentration argument (Lemma (30)). For the last term, we upper bound its spectral norm term by term, and we bound $\left\|\mathcal{T}^r\left(R^{(j)}\right)\right\|_2 = \left\|\Sigma_{>k}^{2r} R^{(j)} \Sigma_k^{-2r}\right\|_2$ using a similar approach as discussed above.

The analysis discussed above relied on the clipping being inactive in every iteration. A significant part of our analysis is devoted to showing that this event holds with high probability. Note that the clipping is inactive in iteration t if $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \leq R$ for all $i \in [n]$, and thus we need to upper bound the lengths $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2$. We can write $AX^{(t)}B^{(t)} = \left(U_k + U_{>k} S^{(t)}\right) P^{(t)}$ where $S^{(t)} := \Sigma_{>k} E^{(t)} \Sigma_k^{-1}$ and $P^{(t)} := \Sigma_k C^{(t)} B^{(t)}$. Thus, for each row i, we have $a_i^\top X^{(t)} B^{(t)} = e_i^\top \left(U_k + U_{>k} S^{(t)}\right) P^{(t)}$ and thus

$$\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \leq \left(\left\|e_i^\top U_k\right\|_2 + \left\|e_i^\top U_{>k} S^{(t)}\right\|_2\right)\left\|P^{(t)}\right\|_2.$$

We have $\left\|e_i^\top U_k\right\|_2 \leq \sqrt{k}\left\|U\right\|_\infty$ and $\left\|P^{(t)}\right\|_2 \leq O(1)$ with high probability as noted earlier, and thus it only remains to upper bound $\left\|e_i^\top U_{>k} S^{(t)}\right\|_2$. To this end, we introduce the following potential for every $0 \leq r, t \leq T$:

$$\Phi_{r,t} := \max_{i \in [n]} \left\|e_i^\top U_{>k} \Sigma_{>k}^{2r} S^{(t)}\right\|_2.$$

The potential values with $r = 0$ are precisely the terms that we need to upper bound, and the values with larger r allows us to analyze these terms. We first use a concentration argument to show that the initialization allows us to upper bound $\Phi_{r,0}$ for all r with high probability (Lemma 21). We then use an inductive argument to show that we can upper bound the potentials $\Phi_{0,t}$ using the initial potentials $\Phi_{r,0}$. Here we assume inductively that the clipping was inactive in all prior iterations and we use the recurrence for $E^{(t+1)}$ discussed above, and show that the clipping is inactive in the current iteration as well (Lemma 28). Overall, we obtain that, a radius $R \geq \widetilde{O}\left(\sqrt{k}\left\|U\right\|_\infty\right)$ suffices to ensure that clipping is inactive in all iterations with high probability.

4 Related work

Privacy models on coherence-sensitive PCA. We study row-level adjacency, under which two matrices differ in one row and the ℓ_2 norm of the difference between those rows is at most 1. The guarantees of Dwork et al. (2014) and Nguyen et al. (2026) apply under this model. By contrast, most earlier coherence-sensitive guarantees were established under different notions of adjacency. Earlier power-method results consider entry-level adjacency on the input matrix (Hardt and Roth, 2013; Hardt and Price, 2014; Balcan et al., 2016). Nicolas et al. (2025) instead allow difference between neighboring input matrices after symmetrization E satisfying $\sqrt{\sum_{i=1}^n \left\|E_{i,:}\right\|_1^2} \leq 1$, while d'Orsi and Novikov (2026) consider the same metric E satisfying $\sqrt{\sum_{i,j}\left|(EE^\top)_{ij}\right|} \leq \Delta$. Since these adjacency notions do not coincide with row-level ℓ_2 adjacency, Table 1 reports the bounds obtained after converting them to our model.

Worst-case bounds. Dwork et al. (2014) give a rank-k guarantee without any assumption on coherence. It provides a benchmark for coherence-independent k-PCA. Their guarantee gives the principal-angle error bound $\widetilde{O}\left(\frac{\sqrt{d}}{\epsilon\left(\sigma_k^2 - \sigma_{k+1}^2\right)}\right)$. Tran et al. (2025) study the spectral-norm error $\left\|\widetilde{M}_k - M_k\right\|_2$, where M_k and $\widetilde{M}_k$ are the best rank-k approximations before and after the Gaussian perturbation. Under their gap condition, converting this spectral-norm guarantee to principal-angle error recovers the bound of Dwork et al. (2014) up to logarithmic factors.

Beyond-worst-case bounds. Hardt and Roth (2013) is the first paper to study coherence-sensitive guarantees for private singular-vector computation. They give a private rank-1 algorithm and use repeated deflation to obtain a rank-k approximation. The resulting guarantee deteriorates when σ_1/σ_k is large. Subsequent works analyze block power methods, which update k directions simultaneously and avoid the dependence on σ_1 introduced by repeated deflation (Hardt and Price, 2014; Balcan et al., 2016; Nicolas et al., 2025). A different approach is taken by d'Orsi and Novikov (2026), who privately estimate the spectral gap and coherence and use these estimates as input to a Gaussian mechanism for estimating the top-k subspace.

Private block power method. Hardt and Price (2014) give a general analysis of the noisy block power method and apply it to private rank-k PCA. Their algorithm uses oversampling, maintaining $p \geq k$ vectors to approximate the subspace spanned by the top k singular vectors. Its coherence-sensitive principal-angle guarantee depends on the gap $\sigma_k - \sigma_{k+1}$. They conjecture that, with oversampling, this dependence can be replaced by dependence on the larger gap $\sigma_k - \sigma_{p+1}$. Balcan et al. (2016) partially address this conjecture by replacing the dependence on $\sigma_k - \sigma_{k+1}$ with dependence on $\sigma_k - \sigma_{q+1}$ for an intermediate rank $q \in [k, p/2]$. Building on this framework, Nicolas et al. (2025) sharpen the sensitivity bound from $\sqrt{p}\left\|X_t\right\|_{\max}$ to $\max_i \left\|\left(X_t\right)_{i,:}\right\|_2$. By contrast, our algorithm does not use oversampling and maintains exactly k vectors throughout the iterations.

Statistical / i.i.d private PCA. This literature obtains stronger statistical guarantees by assuming that the data are i.i.d samples from a structured distribution. Liu et al. (2022) estimates the top singular vector with sub-Gaussian distributed data. Düngler and Sanyal (2025) generalizes this work for top k singular vectors.

5 AI USE STATEMENT

In this work, we used generative AI tools for retrieving and summarizing related literature; discussing research directions; reviewing mathematical arguments and suggesting possible proofs; comparing and interpreting theoretical guarantees; and drafting, restructuring, and editing portions of the manuscript. The authors checked the cited literature against the original sources and reviewed all AI-assisted text and mathematical content. We did not use generative AI tools to generate or analyze data, conduct experiments, or implement the algorithm; these tasks are not applicable to this theoretical work. We take responsibility for the final content of this work, including all text and mathematical claims produced with the aid of generative AI.

REFERENCES

Maria-Florina Balcan, Simon Shaolei Du, Yining Wang, and Adams Wei Yu. An improved gap-dependency analysis of the noisy power method. In *COLT*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 284–309. JMLR.org, 2016.

Avrim Blum, John Hopcroft, and Ravindran Kannan. *Foundations of data science*. Cambridge University Press, 2020.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Kenneth R. Davidson and Stanisław J. Szarek. Local operator theory, random matrices and banach spaces. In *Handbook of the Geometry of Banach Spaces*, volume 1, pages 317–366. Elsevier, 2001. doi: 10.1016/S1874-5849(01)80010-3.

Tommaso d'Orsi and Gleb Novikov. Tight differentially private PCA via matrix coherence. In Kasper Green Larsen and Barna Saha, editors, *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2026, Vancouver, BC, Canada, January 11-14, 2026*, pages 10–51. SIAM, 2026. doi: 10.1137/1.9781611978971.2. URL https://epubs.siam.org/doi/10.1137/1.9781611978971.2.

John C. Duchi, Peter L. Bartlett, and Martin J. Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012. doi: 10.1137/110831659.

Johanna Düngler and Amartya Sanyal. An iterative algorithm for differentially private k-PCA with adaptive noise. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=SWsDJbIloF.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, volume 4004 of *Lecture Notes in Computer Science*, pages 486–503. Springer, 2006.

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze Gauss: optimal bounds for privacy-preserving principal component analysis. In *STOC*, pages 11–20. ACM, 2014.

Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. In *NIPS*, pages 2861–2869, 2014.

Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. In *STOC*, pages 1255–1268. ACM, 2012.

Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *STOC*, pages 331–340. ACM, 2013.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020. doi: 10.1214/18-PS321.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *IEEE Trans. Inf. Theory*, 63(6):4037–4049, 2017.

Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. DP-PCA: statistically optimal and differentially private PCA. In *NeurIPS*, 2022.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, 2017. doi: 10.1109/CSF.2017.11.

Ta Duy Nguyen, Alina Ene, and Huy Le Nguyen. Adaptive power iteration method for differentially private PCA. *CoRR*, abs/2602.11454, 2026.

Julien Nicolas, César Sabater, Mohamed Maouche, Sonia Ben Mokhtar, and Mark Coates. Differentially private and decentralized randomized power method, 2025. URL https://arxiv.org/abs/2411.01931.

Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with Rényi differential privacy. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=-70L8lpp9DF.

G. W. Stewart and Ji guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990. ISBN 978-0-12-670230-9.

Phuc Tran, Van Vu, and Nisheeth K. Vishnoi. Spectral perturbation bounds for low-rank approximation with applications to privacy. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=F0JzotXYgC.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

A MAIN TOOLS

In this section, we state the main theorems from prior work that we will use in our utility analysis shown in Section B.2. Our algorithm and analysis use the following tools from differential privacy.

Gaussian mechanism. Let f be a function that takes as input a matrix A and outputs a vector. We say f has ℓ_2-sensitivity Δ if on any neighboring inputs A and A', we have $\|f(A) - f(A')\|_2 \leq \Delta$. The following result was shown by Dwork et al. (2006).

Theorem 2 (Gaussian mechanism, Dwork et al. (2006))**.** *Assume $0 < \epsilon \leq 1$ and $0 < \delta < 1$. Given an input matrix A and a function f with ℓ_2-sensitivity Δ, the algorithm $\mathcal{M}$, which outputs $f(A) + G$ where G is sampled from the Gaussian distribution $\mathcal{N}(0, \sigma^2 I)$ with $\sigma \geq \frac{\Delta}{\epsilon}\sqrt{2\log\frac{1.25}{\delta}}$, is (ϵ, δ)-DP.*

Theorem 3 (Advanced composition Kairouz et al. (2017))**.** *For all ϵ, δ, the adaptive composition of k algorithms, each of which is (ϵ, δ)-DP, is (ϵ', δ')-private where*

$$\epsilon' = \epsilon\sqrt{2\ln\left(\frac{1}{\delta}\right)k} + \epsilon\frac{e^\epsilon - 1}{e^\epsilon + 1}k \quad \textit{and} \quad \delta' = \delta(k+1)$$

Gaussian concentration results. We also use the following concentration results for Gaussian random variables.

Theorem 4 (Boucheron et al. (2013, Theorem 5.6, Tsirelson–Ibragimov–Sudakov inequality))**.** *If $G \sim \mathcal{N}(0,1)^{d\times d}$ and $f : \mathbb{R}^{d\times d} \to \mathbb{R}$ is L-Lipschitz in Euclidean norm, then, for every $x \geq 0$,*

$$\Pr\left[f(G) - \mathbb{E}\left[f(G)\right] \geq x\right] \leq \exp\left(-\frac{x^2}{2L^2}\right), \qquad \Pr\left[f(G) - \mathbb{E}\left[f(G)\right] \leq -x\right] \leq \exp\left(-\frac{x^2}{2L^2}\right).$$

Lemma 5 (Davidson and Szarek (2001, Theorem II.13))**.** *Let $G \sim \mathcal{N}(0,1)^{q\times k}$. For every $x \geq 0$,* $\Pr\left[\|G\|_2 > \sqrt{q} + \sqrt{k} + x\right] \leq e^{-x^2/2}$.

Proof. The first assertion follows from Davidson-Szarek bound Davidson and Szarek (2001, Theorem II.13). We must rescale our matrix to match their normalization. Assume first that $k \leq q$, and define $\Gamma := \frac{1}{\sqrt{q}}G$. Then $\Gamma \in \mathbb{R}^{q\times k}$ has independent $\mathcal{N}(0, 1/q)$ entries. Apply the Davidson-Szarek Theorem with parameters $n = q, m = k, \beta = \frac{k}{q}, t = \frac{x}{\sqrt{q}}$,and using the identity $\|G\|_2 = \sqrt{q}\, s_1(\Gamma)$, we obtain:

$$\begin{aligned}\Pr\left[\|G\|_2 \geq \sqrt{q} + \sqrt{k} + x\right] &= \Pr\left[s_1(\Gamma) \geq 1 + \sqrt{\frac{k}{q}} + \frac{x}{\sqrt{q}}\right] \\ &\leq \exp\left(-\frac{q}{2}\left(\frac{x}{\sqrt{q}}\right)^2\right) = e^{-x^2/2}.\end{aligned}$$

If $k > q$, we apply the preceding argument to $G^\top$, since $\left\|G^\top\right\|_2 = \|G\|_2$.

□

Lemma 6. *Let $M \in \mathbb{R}^{q\times k}$ and $R \in \mathbb{R}^{k\times k}$ be fixed matrices, and let $G \sim \mathcal{N}(0,1)^{q\times k}$. For every $x \geq 0$,* $\Pr\left[\left\|M^\top GR\right\|_2 > \|M\|_2 \|R\|_2 \left(2\sqrt{k} + x\right)\right] \leq e^{-x^2/2}$.

Proof. Let $r = \operatorname{rank}(M) \leq k$, and write a compact singular value decomposition $M = P\Lambda Q^\top$, where $P \in \mathbb{R}^{q\times r}$ has orthonormal columns. By rotational invariance, $P^\top G \sim \mathcal{N}(0,1)^{r\times k}$. Hence Lemma 5 gives

$$\Pr\left[\left\|P^\top G\right\|_2 \leq \sqrt{r} + \sqrt{k} + x \leq 2\sqrt{k} + x\right] \geq 1 - e^{-x^2/2}$$

But $\left\|M^\top GR\right\|_2 = \left\|Q\Lambda P^\top GR\right\|_2 \leq \|M\|_2 \left\|P^\top G\right\|_2 \|R\|_2$, which implies that:

$$\Pr\left[\left\|M^\top GR\right\|_2 > \|M\|_2 \|R\|_2 \left(2\sqrt{k} + x\right)\right] \leq \Pr\left[\|M\|_2 \left\|P^\top G\right\|_2 \|R\|_2 > \|M\|_2 \|R\|_2 \left(2\sqrt{k} + x\right)\right]$$

$$= \Pr\left[\left\|P^\top G\right\|_2 > 2\sqrt{k} + x\right]$$
$$\leq e^{-x^2/2}$$

□

Spectral analysis tools. Our analysis uses Weyl's inequality and the Davis-Kahan $\sin\Theta$ theorem.

Theorem 7 (Weyl's inequality)**.** *Let M and W be two symmetric matrices and for a matrix A, let $\lambda_i(A)$ denote the i-th largest eigenvalue of A. For every i, we have*

$$|\lambda_i(M+W) - \lambda_i(M)| \leq \|W\|_2.$$

Let us also recall the definition of principal angles between two subspaces. Let $V, \widehat{V} \in \mathbb{R}^{d\times k}$ have orthonormal columns. Let $\sigma_1, \ldots, \sigma_k$ be the singular values of $\widehat{V}^\top V$. The principal angles between the column spaces of V and $\widehat{V}$ are $\theta_i := \arccos(\sigma_i)$ for $1 \leq i \leq k$. We let $\Theta\left(V, \widehat{V}\right) = \operatorname{diag}(\theta_1, \ldots, \theta_k)$ and $\sin\Theta\left(V, \widehat{V}\right) = \operatorname{diag}(\sin(\theta_1), \ldots, \sin(\theta_k))$.

Theorem 8 (Davis-Kahan $\sin\Theta$ theorem, Stewart and guang Sun (1990, Theorem V.3.6), Yu et al. (2015, Theorem 1))**.** *Let $M, \widehat{M} \in \mathbb{R}^{d\times d}$ be symmetric matrices with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$ and $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_d$, respectively. Fix $1 \leq r \leq s \leq d$, let $k = s - r + 1$ and let $V = (v_r, v_{r+1}, \ldots, v_s) \in \mathbb{R}^{d\times k}$ and $\widehat{V} = (\widehat{v}_r, \widehat{v}_{r+1}, \ldots, \widehat{v}_s) \in \mathbb{R}^{d\times k}$ have orthonormal columns satisfying $Mv_i = \lambda_i v_i$ and $\widehat{M}\widehat{v}_i = \widehat{\lambda}_i \widehat{v}_i$ for all $r \leq i \leq s$. Let*

$$\delta := \inf\left\{\left|\lambda - \widehat{\lambda}\right| : \lambda \in [\lambda_s, \lambda_r], \widehat{\lambda} \in (-\infty, \widehat{\lambda}_{s+1}] \cup [\widehat{\lambda}_{r-1}, +\infty)\right\}$$

where $\widehat{\lambda}_0 := +\infty$ and $\widehat{\lambda}_{d+1} = -\infty$.

If $\delta > 0$, we have

$$\left\|\left(I - \widehat{V}\widehat{V}^\top\right)V\right\|_F = \left\|\sin\Theta\left(V, \widehat{V}\right)\right\|_F \leq \frac{\left\|M - \widehat{M}\right\|_F}{\delta},$$

and

$$\left\|\left(I - \widehat{V}\widehat{V}^\top\right)V\right\|_2 = \left\|\sin\Theta\left(V, \widehat{V}\right)\right\|_2 \leq \frac{\left\|M - \widehat{M}\right\|_2}{\delta}.$$

Weyl's inequality (Theorem 7) and the Davis-Kahan theorem (Theorem 8) imply the following lemma. A similar argument appears in Dwork et al. (2014).

Lemma 9. *Let $\widehat{M} = A^\top A + W$ be a symmetric matrix, and let $\Delta_k = \lambda_k\left(A^\top A\right) - \lambda_{k+1}\left(A^\top A\right) > 0$ denote the spectral gap. If $\|W\|_2 < \Delta_k$, the top-k eigenspace $\widehat{V}_k$ of $\widehat{M}$ satisfies*

$$\left\|\left(I - \widehat{V}_k\widehat{V}_k^\top\right)V_k\right\|_2 \leq \frac{\|W\|_2}{\Delta_k - \|W\|_2}.$$

Proof. We apply the Davis–Kahan $\sin\Theta$ theorem (Theorem 8) with $M = A^\top A$ and $\widehat{M} = A^\top A + W$, and $r = 1$ and $s = k$. The spectral separation appearing in the theorem is $\delta := \lambda_k\left(A^\top A\right) - \lambda_{k+1}\left(A^\top A + W\right)$. We obtain

$$\left\|\left(I - \widehat{V}_k\widehat{V}_k^\top\right)V_k\right\|_2 = \left\|\sin\Theta\left(\widehat{V}_k, V_k\right)\right\|_2 \leq \frac{\|W\|_2}{\delta}.$$

We can upper bound $\lambda_{k+1}\left(A^\top A + W\right)$ using Weyl's inequality (Theorem 7):

$$\lambda_{k+1}\left(A^\top A + W\right) \leq \lambda_{k+1}\left(A^\top A\right) + \|W\|_2,$$

and thus

$$\delta \geq \lambda_k\left(A^\top A\right) - \lambda_{k+1}\left(A^\top A\right) - \|W\|_2 = \Delta_k - \|W\|_2.$$

□

Lemma 10. *Let M be a matrix. If $\|M\|_2 < 1$, $I + M$ is invertible and $\left\|(I+M)^{-1}\right\|_2 \leq \frac{1}{1-\|M\|_2}$.*

Proof. We can show that $I + M$ is invertible as follows. Suppose we have $(I + M)\, z = 0$ for some vector z. Then $z = -Mz$ and thus

$$\|z\|_2 = \|Mz\|_2 \leq \|M\|_2 \|z\|_2$$

Since $\|M\|_2 < 1$, the above inequality can only hold if $\|z\|_2 = 0$. Thus $z = 0$ and hence $I + M$ is invertible.

Next, we show the inequality in the lemma statement. For any vector z, we have

$$\|(I + M)\, z\|_2 \geq \|z\|_2 - \|Mz\|_2 \geq (1 - \|M\|_2)\, \|z\|_2$$

Thus

$$\sigma_{\min}\,(I + M) \geq 1 - \|M\|_2$$

Combining with $\|M\|_2 < 1$ gives

$$\left\|(I+M)^{-1}\right\|_2 = \frac{1}{\sigma_{\min}\,(1+M)} \leq \frac{1}{1-\|M\|_2},$$

as needed. □

B Algorithm Analysis

B.1 Privacy Analysis

Lemma 11. *Assume every row a_i of input matrix A satisfies $\|a_i\|_2 \leq 1$, and let $\epsilon_{\text{init}}, \delta_{\text{init}} \in (0, 1)$. Algorithm 1 is $(\epsilon_{\text{init}}, \delta_{\text{init}})$-differentially private.*

Proof. Under replacement adjacency, two neighboring datasets A and A' differ by exactly one row, meaning a row a is replaced by a'. Therefore:

$$A^\top A - A'^\top A' = aa^\top - a'a'^\top$$

and using triangle inequality, we have:

$$\left\|A^\top A - A'^\top A'\right\|_F \leq \left\|aa^\top\right\|_F + \left\|a'a'^\top\right\|_F = \|a\|_2^2 + \|a'\|_2^2 \leq 2$$

Thus, $\Delta \left\|A^\top A\right\|_F \leq 2$.

In the orthonormal basis $\{E_{ii} : 1 \leq i \leq d\} \cup \left\{\frac{E_{ij}+E_{ji}}{\sqrt{2}} : 1 \leq i < j \leq d\right\}$ of the symmetric-matrix space, the coordinates of the matrix $\operatorname{sym}\left(G^{(\text{init})}\right)$ are independent standard Gaussian random variables. Hence $\widehat{M} = A^\top A + \Delta_{\text{init}} \operatorname{sym}\left(G^{(\text{init})}\right)$ is the Gaussian mechanism with sensitivity at most 2, and is therefore $(\epsilon_{\text{init}}, \delta_{\text{init}})$-differentially private. The subsequent step of computing $X^{(0)}$ from $\widehat{M}$ is a deterministic post-processing step. □

Lemma 12. *Fix an initial matrix $X^{(0)}$ with orthonormal columns, a number of iterations $T \geq 1$, and a clipping radius R. Let $\epsilon_{\text{pow}}, \delta_{\text{pow}} \in (0, 1)$ and set*

$$\epsilon_{\text{iter}} = \frac{\epsilon_{\text{pow}}}{4\sqrt{2T \log\left((T+1)/\delta_{\text{pow}}\right)}},$$
$$\delta_{\text{iter}} = \frac{\delta_{\text{pow}}}{2\,(T+1)}.$$

Then Algorithm 2 is $(\epsilon_{\text{pow}}, \delta_{\text{pow}})$-differentially private.

Proof. Since the algorithm is adaptive, we condition on the transcript preceding each release. Previously released quantities may therefore be treated as fixed when analyzing the conditional privacy guarantee.

First we analyze the release of $H^{(t)}$. For a single row a, its contribution to $X^{(t)\top}A^\top AX^{(t)}$ is:

$$X^{(t)\top}aa^\top X^{(t)} = \left(X^{(t)\top}a\right)\left(X^{(t)\top}a\right)^\top$$

Because $X^{(t)}$ is a matrix with orthonormal columns, $\left\|X^{(t)\top}a\right\|_2 \le \|a\|_2$, and thus:

$$\left\|X^{(t)\top}aa^\top X^{(t)}\right\|_F = \left\|\left(X^{(t)\top}a\right)\left(X^{(t)\top}a\right)^\top\right\|_F = \left\|X^{(t)\top}a\right\|_2^2 \le \|a\|_2^2 \le 1$$

When comparing neighboring datasets where row a is replaced by a', the total change is:

$$\left\|X^{(t)\top}aa^\top X^{(t)} - X^{(t)\top}a'a'^\top X^{(t)}\right\|_F \le \left\|X^{(t)\top}aa^\top X^{(t)}\right\|_F + \left\|X^{(t)\top}a'a'^\top X^{(t)}\right\|_F \le 2$$

Therefore, $\Delta\left\|X^{(t)\top}A^\top AX^{(t)}\right\|_F \le 2$. Since $G^{(t)} \sim \mathcal{N}(0,1)^{k\times k}$, adding $\Delta\,\mathrm{sym}\left(G^{(t)}\right)$ applies the symmetric Gaussian mechanism. Because $\Delta = 2\sqrt{2\log(2/\delta_{\mathrm{iter}})}/\epsilon_{\mathrm{iter}}$, the release $H^{(t)} = X^{(t)\top}A^\top AX^{(t)} + \Delta\,\mathrm{sym}\left(G^{(t)}\right)$ is $(\epsilon_{\mathrm{iter}}, \delta_{\mathrm{iter}})$-differentially private. Construction of $B^{(t)}$ are post-processing steps.

Second, we analyze the release of $Z^{(t+1)}$. The contribution of one row a to $Z^{(t+1)}$ is $f_t(a) := a\,\mathrm{Clip}_R\left(a^\top X^{(t)}B^{(t)}\right)$. Since the norm of an outer product is the product of the two vector norms,

$$\|f_t(a_i)\|_F = \|a_i\|_2\left\|\mathrm{Clip}_R\left(a_i^\top X^{(t)}B^{(t)}\right)\right\|_2 \le R$$

Thus $\Delta\left(\sum_{i=1}^n a_i\,\mathrm{Clip}_R\left(a_i^\top X^{(t)}B^{(t)}\right)\right) \le 2R$. Consequently, $Z^{(t+1)} = \sum_{i=1}^n f_t(a_i) + \Delta R\widetilde{G}^{(t)}$ is another $(\epsilon_{\mathrm{iter}}, \delta_{\mathrm{iter}})$-differentially private Gaussian mechanism. The QR factorization is post-processing.

By sequential composition, a single iteration is $(2\epsilon_{\mathrm{iter}}, 2\delta_{\mathrm{iter}})$-differentially private. Advanced composition over the T iterations states that the whole algorithm is $(\epsilon_{\mathrm{total}}, \delta_{\mathrm{total}})$-differentially private, where:

$$\begin{aligned}
\delta_{\mathrm{total}} &= T(2\delta_{\mathrm{iter}}) + \frac{\delta_{\mathrm{pow}}}{T+1} = \delta_{\mathrm{pow}} \\
\epsilon_{\mathrm{total}} &\le 2\epsilon_{\mathrm{iter}}\sqrt{2T\log\left(\frac{T+1}{\delta_{\mathrm{pow}}}\right)} + T(2\epsilon_{\mathrm{iter}})\left(e^{2\epsilon_{\mathrm{iter}}} - 1\right)
\end{aligned}$$

We now show that $\epsilon_{\mathrm{total}} \le \epsilon_{\mathrm{pow}}$. We have $2\epsilon_{\mathrm{iter}} < 1$ so $e^{2\epsilon_{\mathrm{iter}}} - 1 \le 2(2\epsilon_{\mathrm{iter}})$. Our choice of ϵ_{iter} gives

$$\begin{aligned}
2\epsilon_{\mathrm{iter}}\sqrt{2T\log\left(\frac{T+1}{\delta_{\mathrm{pow}}}\right)} &= \frac{\epsilon_{\mathrm{pow}}}{2} \\
T(2\epsilon_{\mathrm{iter}})\left(e^{2\epsilon_{\mathrm{iter}}} - 1\right) &\le 2T(2\epsilon_{\mathrm{iter}})^2 \\
&= 2T\left(\frac{\epsilon_{\mathrm{pow}}}{2\sqrt{2T\log\left(\frac{T+1}{\delta_{\mathrm{pow}}}\right)}}\right)^2 \\
&= \frac{\epsilon_{\mathrm{pow}}^2}{4\log\left(\frac{T+1}{\delta_{\mathrm{pow}}}\right)} \le \frac{\epsilon_{\mathrm{pow}}}{2}
\end{aligned}$$

where the last inequality is true since $\epsilon_{\mathrm{pow}} < 1 \le 2\log\left(\frac{T+1}{\delta_{\mathrm{pow}}}\right)$.

Hence $\epsilon_{\mathrm{total}} \le \epsilon_{\mathrm{pow}}$, and Algorithm 2 is $(\epsilon_{\mathrm{pow}}, \delta_{\mathrm{pow}})$-differentially private. □

Lemma 11, Lemma 12, and sequential composition imply that Algorithm 3 is (ϵ, δ)-differentially private.

B.2 Utility Analysis

We start by introducing some notation that we will use throughout the analysis. We define

$$\kappa_k = \frac{\sigma_k^2 - \sigma_{k+1}^2}{\sigma_k^2} = \frac{\Delta_k}{\sigma_k^2}, \qquad \|U\|_\infty = \max_{i,j} |U_{ij}|.$$

For every iteration $t \geq 0$ of Algorithm 2, we define

$$C^{(t)} := V_k^\top X^{(t)}, D^{(t)} := V_{>k}^\top X^{(t)}, s_t := \left\| \left(I - X^{(t)} X^{(t)\top} \right) V_k \right\|_2.$$

For every iteration t in which $C^{(t)}$ is invertible, we define

$$E^{(t)} := D^{(t)} \left(C^{(t)} \right)^{-1}, m_t := \left\| E^{(t)} \right\|_2.$$

Our goal is to analyze m_t and show that it decreases appropriately with high probability. Our analysis will show that the matrices $C^{(t)}$ are invertible with high probability and thus the matrices $E^{(t)}$ are well-defined. As shown in the following lemma, having $s_t < 1$ suffices to ensure that $C^{(t)}$ is invertible.

Lemma 13. *Let $X \in \mathbb{R}^{d \times k}$ be any matrix with orthonormal columns. Let $s = \left\| \left(I - XX^\top \right) V_k \right\|_2$ and $C = V_k^\top X$. We have $s^2 = 1 - \sigma_{\min}^2 (C)$, where $\sigma_{\min}(C)$ is the smallest singular value of C. Thus, if $s < 1$, then $\sigma_{\min}(C) > 0$ and thus C is invertible.*

Proof. Since X has orthonormal columns, $P = I - XX^\top$ is an orthogonal projector and thus $P^\top P = P^2 = P$. Thus we have

$$\begin{aligned} s^2 &= \left\| \left(I - XX^\top \right) V_k \right\|_2^2 \\ &= \left\| V_k^\top \left(I - XX^\top \right) V_k \right\|_2 \\ &= \left\| I - CC^\top \right\|_2 \\ &= 1 - \sigma_{\min}^2 (C) \end{aligned}$$

as needed. □

We will use the following standard identity relating s_t and m_t.

Lemma 14. *Let $X \in \mathbb{R}^{d \times k}$ be any matrix with orthonormal columns. Let $s = \left\| \left(I - XX^\top \right) V_k \right\|_2$. Let $C = V_k^\top X$ and $D = V_{>k}^\top X$. Suppose that C is invertible and let $m = \left\| DC^{-1} \right\|_2$. We have $s = \frac{m}{\sqrt{1+m^2}}$.*

Proof. Let $E = DC^{-1}$. Since $V = [V_k \; V_{>k}]$ is orthogonal, we can write

$$X = V_k C + V_{>k} D = (V_k + V_{>k} E) C$$

Since C is invertible, we have

$$\operatorname{range}(X) = \operatorname{range}(V_k + V_{>k} E)$$

Note that we have

$$(V_k + V_{>k} E)^\top (V_k + V_{>k} E) = I + E^\top E$$

Thus $Y = (V_k + V_{>k} E) \left(I + E^\top E \right)^{-1/2}$ is an orthonormal basis for $\operatorname{range}(X)$, and therefore $YY^\top = XX^\top$. Thus

$$\left\| \left(I - XX^\top \right) V_k \right\|_2^2 = \left\| \left(I - YY^\top \right) V_k \right\|_2^2 = \left\| V_k^\top \left(I - YY^\top \right)^\top \left(I - YY^\top \right) V_k \right\|_2$$

Since Y has orthonormal columns, $YY^\top$ is an orthogonal projector. Thus $P = I - YY^\top$ is also an orthogonal projector, and thus $P^\top P = P^2 = P$. Thus

$$\left\| \left(I - X^{(t)} X^{(t)\top} \right) V_k \right\|_2^2 = \left\| V_k^\top \left(I - YY^\top \right) V_k \right\|_2 = \left\| I - \left(Y^\top V_k \right)^\top Y^\top V_k \right\|_2$$

Recall that $Y = (V_k + V_{>k}E)\left(I + E^\top E\right)^{-1/2}$. Thus we have

$$Y^\top V_k = \left(I + E^\top E\right)^{-1/2}$$

Thus

$$\left\|\left(I - XX^\top\right)V_k\right\|_2^2 = \left\|I - \left(I + E^\top E\right)^{-1}\right\|_2$$

If E has singular values s_i, $I - \left(I + E^\top E\right)^{-1}$ has eigenvalues $1 - \frac{1}{1+s_i^2} = \frac{s_i^2}{1+s_i^2}$. Thus

$$\left\|\left(I - XX^\top\right)V_k\right\|_2^2 = \frac{\|E\|_2^2}{1 + \|E\|_2^2}$$

as needed. □

We now proceed with the main analysis that bounds m_t. For every $t \geq 0$, we also define

$$\begin{aligned}
K^{(t)} &:= \Sigma_k^2 C^{(t)} B^{(t)} \\
S^{(t)} &:= \Sigma_{>k} E^{(t)} \Sigma_k^{-1} \\
P^{(t)} &:= \Sigma_k C^{(t)} B^{(t)} \\
Y^{(t)} &:= \Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} \\
F^{(t)} &:= \Delta R V_k^\top \widetilde{G}^{(t)} \left(K^{(t)}\right)^{-1} \\
L^{(t)} &:= \Delta R V_{>k}^\top \widetilde{G}^{(t)} \left(K^{(t)}\right)^{-1}
\end{aligned}$$

We also define the following event.

Definition 15. Define

$$\begin{aligned}
q_k &:= 2\sqrt{k} + \sqrt{2\log(32T/\beta)}, \\
q_{d-k} &:= \sqrt{d-k} + \sqrt{k} + \sqrt{2\log(32T/\beta)}, \\
q_{\text{row}} &:= \sqrt{k} + \sqrt{2\log(16nT(T+1)/\beta)}.
\end{aligned}$$

Let $\mathcal{G}_{\text{clip}}$ be the event where, simultaneously for all $0 \leq t < T$ and $0 \leq r \leq T$, we have

$$\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \leq \frac{\gamma}{4}, \left\|V_k^\top \widetilde{G}^{(t)}\right\|_2 \leq q_k, \left\|V_{>k}^\top \widetilde{G}^{(t)}\right\|_2 \leq q_{d-k}, \tag{2}$$

$$\max_i \left\|e_i^\top U_{>k} \Sigma_{>k}^{2r+1} V_{>k}^\top \widetilde{G}^{(t)}\right\|_2 \leq \sigma_{k+1}^{2r} q_{\text{row}} \tag{3}$$

Lemma 16. *We have* $\Pr\left(\mathcal{G}_{\text{clip}}\right) \geq 1 - \frac{\beta}{4}$.

Proof. Note that $\left\|\operatorname{sym}(G^{(t)})\right\|_2 \leq \left\|G^{(t)}\right\|_2$. By Lemma 5, for a fixed t, we have $\left\|\Delta \operatorname{sym}(G^{(t)})\right\|_2 \leq \gamma/4$ with probability $1 - \frac{\beta}{8T}$. Union bound over all t gives the total failure probability $\beta/8$.

By rotational invariance of the normal distribution, $V^\top \widetilde{G}^{(t)}$ is identically distributed as $\widetilde{G}^{(t)}$. Consequently, $V_k^\top \widetilde{G}^{(t)} \sim \mathcal{N}(0,1)^{k \times k}$ and $V_{>k}^\top \widetilde{G}^{(t)} \sim \mathcal{N}(0,1)^{(d-k)\times k}$. The two bounds in (2) then follow from Lemma 5 with failure probability $\frac{\beta}{32T}$ for each bound and each t. Union bound over all t and two bounds gives the total failure probability of $\beta/16$.

For (3), the ith row $e_i^\top U_{>k} \Sigma_{>k}^{2r+1} V_{>k}^\top \widetilde{G}^{(t)}$ is a k-dimensional Gaussian vector whose coordinate variance is at most σ_{k+1}^{4r}. Indeed, each coordinate is identically distributed as a standard Gaussian scaled by

$$\left\|e_i^\top U_{>k} \Sigma_{>k}^{2r+1}\right\|_2 \leq \sigma_{k+1}^{2r} \left\|e_i^\top U_{>k} \Sigma_{>k}\right\|_2 \leq \sigma_{k+1}^{2r} \|a_i\|_2$$

Applying Theorem 4 and a union bound over i, t, r give the stated row-wise bound with total failure probability $\beta/16$.

The total failure probability for all conditions is $\beta/8 + \beta/16 + \beta/16 \leq \beta/4$. □

Next, we show that, under certain assumptions, the preconditioner $B^{(t)}$ roughly whitens $\Sigma_k V_k^\top X^{(t)} = \Sigma_k C^{(t)}$, i.e., the singular values of $\Sigma_k C^{(t)} B^{(t)}$ are bounded from above and below by absolute constants. To this end, we assume that $\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \leq \frac{\gamma}{4}$, $s_t \leq \frac{1}{2}$, and $\gamma \leq \frac{\sigma_k^2}{4}$. As shown in the proof of Lemma 16, the first condition holds with high probability. Later in this section, we will show that the second condition also holds with high probability and the third condition is implied by a sufficiently large eigengap Δ_k. Recall that $P^{(t)} := \Sigma_k C^{(t)} B^{(t)}$.

Lemma 17. *Suppose that* $\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \leq \frac{\gamma}{4}$, $s_t \leq \frac{1}{2}$, *and* $\gamma \leq \frac{\sigma_k^2}{4}$. *We have* $H^{(t)} = \widetilde{H}^{(t)}$, $B^{(t)} = \left(H^{(t)}\right)^{-1/2}$, *and* $\sigma_{\min}\left(P^{(t)}\right) \geq \frac{2}{3}$ *and* $\left\|P^{(t)}\right\|_2 \leq \sqrt{2}$.

Proof. We first show that $H^{(t)} = \widetilde{H}^{(t)}$, i.e., we have $H^{(t)} \succeq \gamma I_k$. We have

$$\begin{aligned} H^{(t)} &= X^{(t)\top} A^\top A X^{(t)} + \Delta \operatorname{sym}\left(G^{(t)}\right) \\ &= X^{(t)\top}\left(V_k \Sigma_k^2 V_k^\top + V_{>k} \Sigma_{>k}^2 V_{>k}^\top\right) X^{(t)} + \Delta \operatorname{sym}\left(G^{(t)}\right) \\ &= C^{(t)\top} \Sigma_k^2 C^{(t)} + D^{(t)\top} \Sigma_{>k}^2 D^{(t)} + \Delta \operatorname{sym}\left(G^{(t)}\right). \end{aligned}$$

We lower bound each of the terms in turn. Consider $C^{(t)\top} \Sigma_k^2 C^{(t)}$. Since $\left\|D^{(t)}\right\|_2 = s_t \leq \frac{1}{2}$, we have $D^{(t)\top} D^{(t)} \preceq \frac{1}{4} I_k$. Since $C^{(t)\top} C^{(t)} + D^{(t)\top} D^{(t)} = I_k$, it follows that $C^{(t)\top} C^{(t)} \succeq \frac{3}{4} I_k$ and thus $C^{(t)\top} \Sigma_k^2 C^{(t)} \succeq \frac{3}{4} \sigma_k^2 I_k$. Next, we have $D^{(t)\top} \Sigma_{>k}^2 D^{(t)} \succeq 0$. We have $\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \leq \frac{\gamma}{4} \leq \frac{\sigma_k^2}{16}$ and thus $\Delta \operatorname{sym}\left(G^{(t)}\right) \succeq -\frac{1}{16} \sigma_k^2 I_k$. Putting everything together, we obtain

$$H^{(t)} \succeq \frac{3}{4} \sigma_k^2 I_k - \frac{1}{16} \sigma_k^2 I_k = \frac{11}{16} \sigma_k^2 I_k \geq \gamma I_k.$$

Thus we have $H^{(t)} = \widetilde{H}^{(t)}$ and $B^{(t)} = \left(H^{(t)}\right)^{-1/2}$.

Finally, consider $P^{(t)} := \Sigma_k C^{(t)} B^{(t)}$. We will lower and upper bound $P^{(t)\top} P^{(t)}$. We have

$$P^{(t)\top} P^{(t)} = B^{(t)\top}\left(C^{(t)\top} \Sigma_k^2 C^{(t)}\right) B^{(t)}$$

Let $M := X^{(t)\top} A^\top A X^{(t)} = C^{(t)\top} \Sigma_k^2 C^{(t)} + D^{(t)\top} \Sigma_{>k}^2 D^{(t)}$ and $N := \Delta \operatorname{sym}\left(G^{(t)}\right)$, and thus $H^{(t)} = M + N$. As shown above, we have $M \succeq \frac{3}{4} \sigma_k^2 I_k$ and $N \preceq \frac{1}{16} \sigma_k^2 I_k$, and thus

$$\begin{aligned} -\frac{1}{12} M \preceq N \preceq \frac{1}{12} M, \\ \frac{11}{12} M \preceq H^{(t)} \preceq \frac{13}{12} M. \end{aligned}$$

Thus

$$\frac{11}{12} B^{(t)\top} M B^{(t)} \preceq B^{(t)\top} H^{(t)} B^{(t)} \preceq \frac{13}{12} B^{(t)\top} M B^{(t)}.$$

Since $B^{(t)} = \left(H^{(t)}\right)^{-1/2}$, we have $B^{(t)\top} H^{(t)} B^{(t)} = I_k$. Thus

$$\frac{12}{13} I_k \preceq B^{(t)\top} M B^{(t)} \preceq \frac{12}{11} I_k.$$

Since $D^{(t)\top} \Sigma_{>k}^2 D^{(t)} \succeq 0$, we have $M \succeq C^{(t)\top} \Sigma_k^2 C^{(t)}$ and thus

$$B^{(t)\top}\left(C^{(t)\top} \Sigma_k^2 C^{(t)}\right) B^{(t)} \preceq B^{(t)\top} M B^{(t)} \preceq \frac{12}{11} I_k.$$

Since $C^{(t)\top} \Sigma_k^2 C^{(t)} \succeq \frac{3}{4} \sigma_k^2 I_k$ and $D^{(t)\top} \Sigma_{>k}^2 D^{(t)} \preceq \frac{1}{4} \sigma_k^2 I_k$, we have $M \preceq \frac{4}{3}\left(C^{(t)\top} \Sigma_k^2 C^{(t)}\right)$ and thus

$$B^{(t)\top}\left(C^{(t)\top} \Sigma_k^2 C^{(t)}\right) B^{(t)} \succeq \frac{3}{4} B^{(t)\top} M B^{(t)} \succeq \frac{9}{13} I_k.$$

Putting everything together, we have

$$\frac{9}{13} I_k \preceq P^{(t)\top} P^{(t)} \preceq \frac{12}{11} I_k$$

and thus

$$\sigma_{\min}\left(P^{(t)}\right) \geq \sqrt{\frac{9}{13}} \geq \frac{2}{3},$$
$$\left\|P^{(t)}\right\|_2 \leq \sqrt{\frac{12}{11}} \leq \sqrt{2}.$$

□

Using Lemma 17, we can show the following.

Lemma 18. *Suppose that* $\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \leq \frac{\gamma}{4}$, $s_t \leq \frac{1}{2}$, *and* $\gamma \leq \frac{\sigma_k^2}{4}$. *We have* $\left\|\left(K^{(t)}\right)^{-1}\right\|_2 \leq \frac{3}{2}\frac{1}{\sigma_k}$.

Proof. We have $K^{(t)} = \Sigma_k^2 C^{(t)} B^{(t)} = \Sigma_k P^{(t)}$. By Lemma 17, we have $\sigma_{\min}\left(P^{(t)}\right) \geq \frac{2}{3}$. Thus $P^{(t)}$ is invertible and

$$\begin{aligned}\left\|\left(K^{(t)}\right)^{-1}\right\|_2 &= \left\|\left(\Sigma_k P^{(t)}\right)^{-1}\right\|_2 = \left\|\left(P^{(t)}\right)^{-1}\Sigma_k^{-1}\right\|_2 \leq \left\|\left(P^{(t)}\right)^{-1}\right\|_2 \left\|\Sigma_k^{-1}\right\|_2 \\ &= \frac{1}{\sigma_{\min}\left(P^{(t)}\right)}\frac{1}{\sigma_k} \leq \frac{3}{2}\frac{1}{\sigma_k}.\end{aligned}$$

□

Main recurrence. In the following, we say that clipping is inactive in iteration t if $\operatorname{Clip}_R\left(a_i^\top X^{(t)} B^{(t)}\right) = a_i^\top X^{(t)} B^{(t)}$ for all $i \in [n]$, i.e., we have $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \leq R$ for all $i \in [n]$. Note that, in this case, the main update in Algorithm 2 is $Z^{(t+1)} = A^\top A X^{(t)} B^{(t)} + \Delta R \widetilde{G}^{(t)}$. First, we will show a recurrence for $E^{(t+1)}$ assuming that clipping is inactive in iteration t (Lemma 19). This recurrence expresses $E^{(t+1)}$ in terms of the ideal progress term $\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}$ (note that $\left\|\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}\right\|_2 \leq \frac{\sigma_{k+1}^2}{\sigma_k^2}\left\|E^{(t)}\right\|_2 = (1-\kappa_k)\left\|E^{(t)}\right\|_2$), and additional error terms that we will bound in Theorem 31.

Lemma 19. *Suppose* $C^{(t)}$ *is invertible and clipping is inactive in iteration* t. *If* $\left\|F^{(t)}\right\|_2 < 1$, *then* $C^{(t+1)}$ *is invertible and*

$$E^{(t+1)} = \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + L^{(t)}\right)\left(I + F^{(t)}\right)^{-1}.$$

Proof. Let us start by observing that

$$\begin{aligned}C^{(t+1)} R^{(t+1)} &= V_k^\top X^{(t+1)} R^{(t+1)} \\ &= V_k^\top Z^{(t+1)} \\ &= V_k^\top \left(A^\top A X^{(t)} B^{(t)} + \Delta R \widetilde{G}^{(t)}\right) \\ &= V_k^\top A^\top A X^{(t)} B^{(t)} + \Delta R V_k^\top \widetilde{G}^{(t)} \\ &= \Sigma_k^2 V_k^\top X^{(t)} B^{(t)} + \Delta R V_k^\top \widetilde{G}^{(t)} \\ &= \Sigma_k^2 C^{(t)} B^{(t)} + \Delta R V_k^\top \widetilde{G}^{(t)} \\ &= K^{(t)} + \Delta R V_k^\top \widetilde{G}^{(t)}.\end{aligned}$$

Since $C^{(t)}$, Σ_k^2, and $B^{(t)}$ are invertible, $K^{(t)} = \Sigma_k^2 C^{(t)} B^{(t)}$ is invertible. Then $F^{(t)} := \Delta R V_k^\top \widetilde{G}^{(t)}\left(K^{(t)}\right)^{-1}$ is well-defined.

Multiplying the above identity by $\left(K^{(t)}\right)^{-1}$ then gives

$$C^{(t+1)} R^{(t+1)}\left(K^{(t)}\right)^{-1} = I + \left(\Delta R V_k^\top \widetilde{G}^{(t)}\right)\left(K^{(t)}\right)^{-1} = I + F^{(t)}$$

Since $\left\|F^{(t)}\right\|_2 < 1$, Lemma 10 implies that $I + F^{(t)}$ is invertible. Thus $C^{(t+1)}R^{(t+1)}$ is invertible and therefore $C^{(t+1)}$ and $R^{(t+1)}$ are invertible.

This identity also gives

$$C^{(t+1)} = \left(I + F^{(t)}\right) K^{(t)} \left(R^{(t+1)}\right)^{-1}.$$

Now we establish the recurrence. Using the above identity, we obtain

$$\begin{aligned}
E^{(t+1)} &= D^{(t+1)} \left(C^{(t+1)}\right)^{-1} \\
&= V_{>k}^\top X^{(t+1)} \left(C^{(t+1)}\right)^{-1} \\
&= V_{>k}^\top Z^{(t+1)} \left(R^{(t+1)}\right)^{-1} \left(C^{(t+1)}\right)^{-1} \\
&= V_{>k}^\top Z^{(t+1)} \left(C^{(t+1)} R^{(t+1)}\right)^{-1} \\
&= V_{>k}^\top Z^{(t+1)} \left(K^{(t)}\right)^{-1} \left(I + F^{(t)}\right)^{-1}.
\end{aligned}$$

Note that

$$\begin{aligned}
V_{>k}^\top Z^{(t+1)} &= V_{>k}^\top \left(A^\top A X^{(t)} B^{(t)} + \Delta R \widetilde{G}^{(t)}\right) \\
&= \Sigma_{>k}^2 V_{>k}^\top X^{(t)} B^{(t)} + \Delta R V_{>k}^\top \widetilde{G}^{(t)} \\
&= \Sigma_{>k}^2 D^{(t)} B^{(t)} + \Delta R V_{>k}^\top \widetilde{G}^{(t)} \\
&= \Sigma_{>k}^2 E^{(t)} C^{(t)} B^{(t)} + \Delta R V_{>k}^\top \widetilde{G}^{(t)} \\
&= \Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} K^{(t)} + \Delta R V_{>k}^\top \widetilde{G}^{(t)} \\
&= \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + \Delta R V_{>k}^\top \widetilde{G}^{(t)} \left(K^{(t)}\right)^{-1}\right) K^{(t)} \\
&= \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + L^{(t)}\right) K^{(t)}.
\end{aligned}$$

Plugging this into the previous identity for $E^{(t+1)}$, we obtain the claimed recurrence. □

Showing that clipping is inactive. Using Lemma 19, we will show that, for an appropriate choice of R, clipping is inactive in every iteration with high probability. Recall that the clipping is inactive in iteration t if $\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \le R$ for all $i \in [n]$. In the following, we analyze $a_i^\top X^{(t)} B^{(t)}$ for each $i \in [n]$. Recall that $S^{(t)} := \Sigma_{>k} E^{(t)} \Sigma_k^{-1}$.

Lemma 20. *Suppose that* $\left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 \le \frac{\gamma}{4}$, $s_t \le \frac{1}{2}$, *and* $\gamma \le \frac{\sigma_k^2}{4}$. *For each* $1 \le i \le n$, *we have*

$$\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \le \sqrt{2}\left(\sqrt{k}\left\|U\right\|_\infty + \left\|e_i^\top U_{>k} S^{(t)}\right\|_2\right).$$

Proof. We have

$$\begin{aligned}
A X^{(t)} B^{(t)} &= \left(U_k \Sigma_k V_k^\top + U_{>k} \Sigma_{>k} V_{>k}^\top\right) X^{(t)} B^{(t)} \\
&= U_k \underbrace{\Sigma_k V_k^\top X^{(t)} B^{(t)}}_{P^{(t)}} + U_{>k} \Sigma_{>k} \underbrace{V_{>k}^\top X^{(t)}}_{D^{(t)} = E^{(t)} C^{(t)}} B^{(t)} \\
&= \left(U_k + U_{>k} S^{(t)}\right) P^{(t)}.
\end{aligned}$$

Thus, for each row $i \in [n]$, we have

$$a_i^\top X^{(t)} B^{(t)} = e_i^\top A X^{(t)} B^{(t)} = e_i^\top \left(U_k + U_{>k} S^{(t)}\right) P^{(t)}$$

and thus

$$\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 \le \left(\left\|e_i^\top U_k\right\|_2 + \left\|e_i^\top U_{>k} S^{(t)}\right\|_2\right) \left\|P^{(t)}\right\|_2.$$

By Lemma 17, we have $\left\|P^{(t)}\right\|_2 \le \sqrt{2}$. We also have $\left\|e_i^\top U_k\right\|_2 \le \sqrt{k}\left\|U\right\|_\infty$. □

Next, we upper bound $\max_{i\in[n]} \left\|e_i^\top U_{>k} S^{(t)}\right\|_2$. To this end, we introduce the following potential. For every $0 \le r, t \le T$, let

$$\Phi_{r,t} := \max_{i\in[n]} \left\|e_i^\top U_{>k} \Sigma_{>k}^{2r} S^{(t)}\right\|_2 .$$

Our goal is to upper bound $\Phi_{0,t} = \max_{i\in[n]} \left\|e_i^\top U_{>k} S^{(t)}\right\|_2$, and we will use $\Phi_{r,\cdot}$ for larger r to do so. First, we show that the initialization allows us to upper bound $\Phi_{r,0}$ for all $0 \le r \le T$ as follows.

Lemma 21. *Suppose that* $\Delta_k \ge 8\Delta_{\text{init}} \left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}}\right)$. *With probability at least* $1 - \frac{\beta}{4}$, *simultaneously for all* $0 \le r \le T$, *we have*

$$\Phi_{r,0} \le \sigma_{k+1}^{2r} \omega_{\text{init}} \quad \omega_{\text{init}} := 3\|U\|_\infty \left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T+1)}{\beta}\right)}\right) \tag{4}$$

Let $\mathcal{G}_{\text{init}}$ *be this event.*

Proof. Recall that we have $\Phi_{r,0} = \max_{i\in[n]} \left\|e_i^\top U_{>k} \Sigma_{>k}^{2r+1} E^{(0)} \Sigma_k^{-1}\right\|_2$. To analyze the norm, it is useful to view $E^{(0)}$ as a function of the Gaussian noise matrix added by Algorithm 1. More precisely, for a Gaussian matrix $G \sim \mathcal{N}(0,1)^{d\times d}$, let $X\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)$ be the top-k eigenvectors of the matrix $A^\top A + \Delta_{\text{init}} \operatorname{sym}(G)$, and let

$$\begin{aligned}
C\left(\Delta_{\text{init}} \operatorname{sym}(G)\right) &:= V_k^\top X\left(\Delta_{\text{init}} \operatorname{sym}(G)\right) \\
D\left(\Delta_{\text{init}} \operatorname{sym}(G)\right) &:= V_{>k}^\top X\left(\Delta_{\text{init}} \operatorname{sym}(G)\right) \\
E\left(\Delta_{\text{init}} \operatorname{sym}(G)\right) &:= D\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)\left(C\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)\right)^{-1}
\end{aligned}$$

where $E\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)$ is defined whenever $C\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)$ is invertible. If the function $M \to E(M)$ was Lipschitz under the Frobenius norm, we could use Theorem 4 to analyze $\left\|e_i^\top U_{>k} \Sigma_{>k}^{2r+1} E(M) \Sigma_k^{-1}\right\|_2$. Unfortunately, the function is not globally Lipschitz, but it is Lipchitz if we restrict it to the following ball:

$$\mathcal{B} := \left\{M \colon M = M^\top, \|M\|_2 \le \frac{\Delta_k}{8}\right\}.$$

We show in the following claim that the function $M \to E(M)$ is Lipschitz on $\mathcal{B}$. It is convenient to construct a Lipschitz extension of this function that is defined everywhere as follows. Let $\Pi_{\mathcal{B}}$ be the Frobenius norm projection onto $\mathcal{B}$, i.e., for every matrix M, we let $\Pi_{\mathcal{B}}(M) = \arg\min_{M'\in\mathcal{B}} \|M - M'\|_F$. Let

$$\overline{E}(G) = E\left(\Pi_{\mathcal{B}}\left(\Delta_{\text{init}} \operatorname{sym}(G)\right)\right).$$

Next, we show that $M \to E(M)$ is Lipschitz on $\mathcal{B}$ under the Frobenius norm, which will then imply that $G \to \overline{E}(G)$ is globally Lipschitz under the Frobenius norm.

Claim 22. For any two matrices $M, M' \in \mathcal{B}$ for which $E(M)$ and $E(M')$ are well-defined (i.e., $C(M)$ and $C(M')$ are invertible), we have

$$\|E(M) - E(M')\|_F \le \frac{6}{\Delta_k} \|M - M'\|_F .$$

Proof. We start by writing $E(M)$ and $E(M')$ in a more convenient form. For $W \in \{M, M'\}$, let $P(W) := X(W) X(W)^\top$. We can write

$$V^\top P(W) V = \begin{pmatrix} P_{11}(W) & P_{12}(W) \\ P_{21}(W) & P_{22}(W) \end{pmatrix},$$

where

$$\begin{aligned}
P_{11}(W) &= V_k^\top P(W) V_k = C(W) C(W)^\top \in \mathbb{R}^{k\times k} \\
P_{12}(W) &= V_k^\top P(W) V_{>k} = C(W) D(W)^\top \in \mathbb{R}^{k\times(d-k)} \\
P_{21}(W) &= V_{>k}^\top P(W) V_k = D(W) C(W)^\top \in \mathbb{R}^{(d-k)\times k}
\end{aligned}$$

$$P_{22}(W) = V_{>k}^\top P(W) V_{>k} = D(W) D(W)^\top \in \mathbb{R}^{(d-k)\times(d-k)}.$$

We have

$$E(W) = D(W) C(W)^{-1} = P_{21}(W) P_{11}(W)^{-1}.$$

Using the above identity, we obtain

$$\begin{aligned}
&\|E(M) - E(M')\|_F \\
&= \left\|P_{21}(M) P_{11}(M)^{-1} - P_{21}(M') P_{11}(M')^{-1}\right\|_F \\
&= \left\|(P_{21}(M) - P_{21}(M')) P_{11}(M)^{-1} + P_{21}(M') \left(P_{11}(M)^{-1} - P_{11}(M')^{-1}\right)\right\|_F \\
&\overset{(a)}{=} \left\|(P_{21}(M) - P_{21}(M')) P_{11}(M)^{-1} + P_{21}(M') P_{11}(M)^{-1} (P_{11}(M') - P_{11}(M)) P_{11}(M')^{-1}\right\|_F \\
&\overset{(b)}{\leq} \|P_{21}(M) - P_{21}(M')\|_F \left\|P_{11}(M)^{-1}\right\|_2 \\
&\quad + \|P_{21}(M')\|_2 \left\|P_{11}(M)^{-1}\right\|_2 \left\|P_{11}(M')^{-1}\right\|_2 \|P(M) - P(M')\|_F \\
&\overset{(c)}{\leq} \left(1 + \|P_{21}(M')\|_2 \left\|P_{11}(M')^{-1}\right\|_2\right) \left\|P_{11}(M)^{-1}\right\|_2 \|P(M) - P(M')\|_F
\end{aligned}$$

In (a), we used the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. In (b), we used the inequalities $\|A + B\|_F \leq \|A\|_F + \|B\|_F$, $\|AB\|_F \leq \|A\|_F \|B\|_2$, and $\|AB\|_2 \leq \|A\|_2 \|B\|_2$. In (c), we used that $\|P_{21}(M) - P_{21}(M')\|_F \leq \|P(M) - P(M')\|_F$, which we can show as follows:

$$\begin{aligned}
\|P_{21}(M) - P_{21}(M')\|_F &= \left\|V_{>k}^\top (P(M) - P(M')) V_k\right\|_F \\
&\leq \left\|V_{>k}^\top\right\|_2 \|P(M) - P(M')\|_F \|V_k\|_2 \\
&= \|P(M) - P(M')\|_F .
\end{aligned}$$

where the last equality follows from the fact that V_k and $V_{>k}$ have orthonormal columns.

Next, we bound the terms in the above inequality. Consider $\|P_{21}(M')\|_2$. We have

$$\begin{aligned}
\|P_{21}(M')\|_2 &= \left\|V_{>k}^\top X(M') X(M')^\top V_k\right\|_2 \\
&\leq \left\|V_{>k}^\top\right\|_2 \|X(M')\|_2^2 \|V_k\|^2 \leq 1
\end{aligned}$$

where the last inequality follows from the fact that V_k, $V_{>k}$, and $X(M')$ have orthonormal columns.

Consider $\left\|P_{11}(M)^{-1}\right\|_2$ and $\left\|P_{11}(M')^{-1}\right\|_2$. For $W \in \{M, M'\}$, we have

$$\|I_k - P_{11}(W)\|_2 = \left\|V_k^\top (I - P(W)) V_k\right\|_2 = \|(I - P(W)) V_k\|_2^2$$

where we used that $P_{11}(W) = V_k^\top P(W) V_k$ and $Q := I - P(W)$ is an orthogonal projector and thus $Q^\top Q = Q$.

Lemma 9 together with $\|W\|_2 \leq \frac{\Delta_k}{8}$ give

$$\|(I - P(W)) V_k\|_2 \leq \frac{\|W\|_2}{\Delta_k - \|W\|_2} \leq \frac{2\|W\|_2}{\Delta_k} \leq \frac{1}{4}.$$

Thus

$$\|I_k - P_{11}(W)\|_2 \leq \frac{1}{16}$$

and therefore $P_{11}(W) \succeq \frac{15}{16} I_k$ and $\left\|P_{11}(W)^{-1}\right\|_2 \leq \frac{16}{15}$.

Finally, consider $\|P(M) - P(M')\|_F$. We have

$$\begin{aligned}
\|P(M) - P(M')\|_F &= \left\|X(M) X(M)^\top - X(M') X(M')^\top\right\|_F \\
&= \sqrt{2} \left\|\left(I - X(M') X(M')^\top\right) X(M)\right\|_F
\end{aligned}$$

We can upper bound the last term using the Davis-Kahan theorem (Theorem 8) and Weyl's inequality (Theorem 7) as follows. Let

$$\delta = \max\left\{\lambda_k\left(A^\top A + M\right) - \lambda_{k+1}\left(A^\top A + M'\right), \lambda_k\left(A^\top A + M'\right) - \lambda_{k+1}\left(A^\top A + M\right)\right\}.$$

We apply the Davis-Kahan theorem with matrices $A^\top A + M$ and $A^\top A + M'$, and $s = 1$ and $r = k$. We obtain

$$\left\|\left(I - X\left(M'\right) X\left(M'\right)^\top\right) X\left(M\right)\right\|_F \leq \frac{\left\|M - M'\right\|_F}{\delta}.$$

To lower bound δ, we use Weyl's inequality and the fact that $\left\|M\right\|_2, \left\|M'\right\|_2 \leq \frac{\Delta_k}{8}$. By Weil's inequality, for $W \in \{M, M'\}$ and $j \in \{k, k+1\}$, we have

$$\left|\lambda_j\left(A^\top A + W\right) - \lambda_j\left(A^\top A\right)\right| \leq \left\|W\right\|_2.$$

Thus

$$\begin{aligned}
\lambda_k\left(A^\top A + M\right) - \lambda_{k+1}\left(A^\top A + M'\right) &\geq \left(\lambda_k\left(A^\top A\right) - \left\|M\right\|_2\right) - \left(\lambda_{k+1}\left(A^\top A\right) + \left\|M'\right\|_2\right) \\
&= \sigma_k^2 - \sigma_{k+1}^2 - \left\|M\right\|_2 - \left\|M'\right\|_2 \\
&= \Delta_k - \left\|M\right\|_2 - \left\|M'\right\|_2 \\
&\geq \frac{3}{4}\Delta_k
\end{aligned}$$

where in the last inequality we used that $\left\|M\right\|_2, \left\|M'\right\|_2 \leq \frac{\Delta_k}{8}$.

Similarly, we have

$$\lambda_k\left(A^\top A + M'\right) - \lambda_{k+1}\left(A^\top A + M\right) \geq \frac{3}{4}\Delta_k$$

and thus $\delta \geq \frac{3}{4}\Delta_k$.

Thus we obtain

$$\left\|\left(I - X\left(M'\right) X\left(M'\right)^\top\right) X\left(M\right)\right\|_F \leq \frac{\left\|M - M'\right\|_F}{\delta} \leq \frac{4}{3\Delta_k}\left\|M - M'\right\|_F.$$

and

$$\left\|P\left(M\right) - P\left(M'\right)\right\|_F = \sqrt{2}\left\|\left(I - X\left(M'\right) X\left(M'\right)^\top\right) X\left(M\right)\right\|_F \leq \frac{8}{3\Delta_k}\left\|M - M'\right\|_F.$$

Putting everything together, we have

$$\left\|E\left(M\right) - E\left(M'\right)\right\|_F \leq \left(1 + \frac{16}{15}\right)\frac{16}{15}\frac{8}{3\Delta_k}\left\|M - M'\right\|_F \leq \frac{6}{\Delta_k}\left\|M - M'\right\|_F,$$

as needed. □

Using the above claim and the fact that $\Pi_{\mathcal{B}}\left(\cdot\right)$ and $\mathrm{sym}\left(\cdot\right)$ are non-expansive in Frobenius norm, we can show that $G \to \overline{E}\left(G\right)$ is Lipschitz.

Claim 23. For any two matrices G, G' for which $\overline{E}\left(G\right)$ and $\overline{E}\left(G'\right)$ are well-defined, we have

$$\left\|\overline{E}\left(G\right) - \overline{E}\left(G'\right)\right\|_F \leq \frac{6\Delta_{\mathrm{init}}}{\Delta_k}\left\|G - G'\right\|_F.$$

where c_1 is an absolute constant.

Proof. We have

$$\begin{aligned}
\left\|\overline{E}\left(G\right) - \overline{E}\left(G'\right)\right\|_F &= \left\|E\left(\Pi_{\mathcal{B}}\left(\Delta_{\mathrm{init}}\,\mathrm{sym}\left(G\right)\right)\right) - E\left(\Pi_{\mathcal{B}}\left(\Delta_{\mathrm{init}}\,\mathrm{sym}\left(G'\right)\right)\right)\right\|_F \\
&\leq \frac{6}{\Delta_k}\left\|\Pi_{\mathcal{B}}\left(\Delta_{\mathrm{init}}\,\mathrm{sym}\left(G\right)\right) - \Pi_{\mathcal{B}}\left(\Delta_{\mathrm{init}}\,\mathrm{sym}\left(G'\right)\right)\right\|_F \\
&\leq \frac{6}{\Delta_k}\left\|\Delta_{\mathrm{init}}\,\mathrm{sym}\left(G\right) - \Delta_{\mathrm{init}}\,\mathrm{sym}\left(G'\right)\right\|_F \\
&\leq \frac{6\Delta_{\mathrm{init}}}{\Delta_k}\left\|G - G'\right\|_F
\end{aligned}$$

where we used the previous claim and the fact that $\Pi_{\mathcal{B}}\left(\cdot\right)$ and $\mathrm{sym}\left(\cdot\right)$ are non-expansive in Frobenius norm. □

Next, we use Gaussian concentration to upper bound $\left\|e_i^\top U_{>k}\Sigma_{>k}^{2r+1}\overline{E}\left(G^{(\text{init})}\right)\Sigma_k^{-1}\right\|_2$ for every row $1 \le i \le n$ and every $0 \le r \le T$. To this end, for a fixed row i, exponent r, and an arbitrary unit vector $z \in \mathbb{R}^k$, we consider the following scalar-valued function:

$$f_{i,r,z}(G) := e_i^\top U_{>k}\Sigma_{>k}^{2r+1}\overline{E}(G)\,\Sigma_k^{-1}z$$

Since $\overline{E}(G)$ is Lipschitz, we can show that $f_{i,r,z}$ is Lipschitz.

Claim 24. For any two matrices G, G' for which $\overline{E}(G)$ and $\overline{E}(G')$ are well-defined, we have

$$|f_{i,r,z}(G) - f_{i,r,z}(G')| \le L_{i,r}\|G - G'\|_F,$$

where $L_{i,r} := \frac{6\Delta_{\text{init}}}{\Delta_k}\frac{\sigma_{k+1}^{2r+1}}{\sigma_k}\left\|e_i^\top U_{>k}\right\|_2$.

Proof. We have

$$\begin{aligned}
|f_{i,r,z}(G) - f_{i,r,z}(G')| &= \left|e_i^\top U_{>k}\Sigma_{>k}^{2r+1}\left(\overline{E}(G) - \overline{E}(G')\right)\Sigma_k^{-1}z\right| \\
&\le \left\|e_i^\top U_{>k}\right\|_2\left\|\Sigma_{>k}^{2r+1}\right\|_2\left\|\overline{E}(G) - \overline{E}(G')\right\|_F\left\|\Sigma_k^{-1}\right\|_2\|z\|_2 \\
&= \left\|e_i^\top U_{>k}\right\|_2\sigma_{k+1}^{2r+1}\left\|\overline{E}(G) - \overline{E}(G')\right\|_F\frac{1}{\sigma_k} \\
&\le \left\|e_i^\top U_{>k}\right\|_2\sigma_{k+1}^{2r+1}\frac{6\Delta_{\text{init}}}{\Delta_k}\|G - G'\|_F\frac{1}{\sigma_k}
\end{aligned}$$

where in the last inequality we used the fact that $\overline{E}(G)$ is $\frac{6\Delta_{\text{init}}}{\Delta_k}$-Lipschitz. □

Next, we show that $\mathbb{E}[f_{i,r,z}(G)] = 0$.

Claim 25. We have $\mathbb{E}[f_{i,r,z}(G)] = 0$.

Proof. Consider a Gaussian matrix $G \sim \mathcal{N}(0,1)^{d\times d}$. Let

$$J = V\begin{pmatrix} -I_k & 0 \\ 0 & I_{d-k}\end{pmatrix}V^\top.$$

The matrix J is orthogonal and symmetric. Let $G' = JGJ^\top$. Since the Gaussian distribution is rotationally invariant, G' has the same distribution as G. Next, we will show that $f_{i,r,z}(G') = -f_{i,r,z}(G)$ which will imply that $\mathbb{E}[f_{i,r,z}(G)] = 0$. Let $M := \Pi_{\mathcal{B}}(\Delta_{\text{init}}\,\text{sym}(G))$ and $M' := \Pi_{\mathcal{B}}(\Delta_{\text{init}}\,\text{sym}(G'))$. Note that we have $M' = JMJ^\top$ and $J\left(A^\top A\right)J^\top = A^\top A$, and thus

$$A^\top A + M' = J\left(A^\top A + M\right)J^\top.$$

As in the proof of Claim 22, for $W \in \{M, M'\}$, we let $P(W) = X(W)X(W)^\top$ and we write it as follows:

$$V^\top P(W)V = \begin{pmatrix} P_{11}(W) & P_{12}(W) \\ P_{21}(W) & P_{22}(W)\end{pmatrix}$$

where $P_{11}(W) = V_k^\top P(W)V_k$, $P_{12}(W) = V_k^\top P(W)V_{>k}$, $P_{21}(W) = V_{>k}^\top P(W)V_k$, and $P_{22}(W) = V_{>k}^\top P(W)V_{>k}$.

Since $A^\top A + M' = J\left(A^\top A + M\right)J^\top$, $JV_k = -V_k$, and $JV_{>k} = V_{>k}$, we have $P(M') = JP(M)J^\top$ and

$$V^\top P(M')V = \begin{pmatrix} P_{11}(M) & -P_{12}(M) \\ -P_{21}(M) & P_{22}(M)\end{pmatrix}$$

Since $E(M) = P_{21}(M)P_{11}(M)^{-1}$, we obtain that $\overline{E}(G') = -\overline{E}(G)$. Therefore $f_{i,r,z}(G') = -f_{i,r,z}(G)$ and

$$\mathbb{E}[f_{i,r,z}(G)] = \mathbb{E}[f_{i,r,z}(G')] = -\mathbb{E}[f_{i,r,z}(G)].$$

The expectation exists since $f_{i,r,z}$ is globally Lipschitz, and thus $\mathbb{E}[f_{i,r,z}(G)] = 0$. □

By applying Theorem 4 to $f_{i,r,z}$ and using a net argument, we obtain the following guarantee.

Claim 26. With probability at least $1 - \frac{\beta}{8}$, simultaneously for all $1 \le i \le n$ and $0 \le r \le T$, we have

$$\left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1} \right\|_2 \le 2L_{i,r}\left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)}\right).$$

Proof. We discretize the set of all possible unit vectors $z \in \mathbb{R}^k$ (i.e., the unit sphere in $\mathbb{R}^k$) using a $1/2$-net $\mathcal{N}$, i.e., $\mathcal{N}$ is a set of unit vectors with the property that, for every unit vector $z \in \mathbb{R}^k$, there is a vector $z' \in \mathcal{N}$ such that $\|z - z'\|_2 \le 1/2$. A standard volume argument shows that there exists a net of size $|\mathcal{N}| \le 5^k$. In the following, we let $\mathcal{N}$ be a $1/2$-net of size $|\mathcal{N}| \le 5^k$.

Fix a row index i and exponent r. Let $w := e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1}$. Note that $w \in \mathbb{R}^{1\times k}$ is a row vector. In the following, we upper bound $\|w\|_2$. We may assume that $w \neq 0$, since otherwise $\|w\|_2 = 0$ and we are done. Let $\overline{w} = w^\top / \|w\|_2$ and thus $\|w\|_2 = w\overline{w}$. Since $\mathcal{N}$ is a $1/2$-net, there is a vector $z' \in \mathcal{N}$ such that $\|\overline{w} - z'\|_2 \le 1/2$. We have

$$\begin{aligned}
\|w\|_2 = w\overline{w} &= wz' + w\left(\overline{w} - z'\right) \\
&\le |wz'| + \|w\|_2 \|\overline{w} - z'\|_2 \\
&\le \max_{z\in\mathcal{N}} |wz| + \frac{1}{2}\|w\|_2 .
\end{aligned}$$

Rearranging, we obtain

$$\|w\|_2 \le 2\max_{z\in\mathcal{N}} |wz| .$$

Since $w = e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1}$ and $f_{i,r,z}\left(G^{(\text{init})}\right) = wz$, we have

$$\left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1} \right\|_2 \le 2\max_{z\in\mathcal{N}} \left| f_{i,r,z}\left(G^{(\text{init})}\right)\right| .$$

Since $f_{i,r,z}$ is $L_{i,r}$-Lipschitz and has expectation 0, Theorem 4 gives that, for every $x \ge 0$,

$$\Pr\left[\left| f_{i,r,z}\left(G^{(\text{init})}\right)\right| \ge L_{i,r}x\right] \le 2\exp\left(-\frac{x^2}{2}\right).$$

There are $n\,(T+1)\,|\mathcal{N}| \le n\,(T+1)\,5^k$ triples (i,r,z) where $1 \le i \le n$ is a row index, $0 \le r \le T$ is an exponent, and z is a vector in $\mathcal{N}$. Setting $x = \sqrt{2\log\left(\frac{16n(T+1)5^k}{\beta}\right)}$ in the above inequality and using an union bound over all triples (i,r,z), we obtain that, with probability at least $1 - \frac{\beta}{8}$, simultaneously for all i and r, we have

$$\begin{aligned}
\left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1} \right\|_2 &\le 2\max_{z\in\mathcal{N}} \left| f_{i,r,z}\left(G^{(\text{init})}\right)\right| \\
&\le 2L_{i,r}\sqrt{2\log\left(\frac{16n\,(T+1)\,5^k}{\beta}\right)} \\
&\le 2L_{i,r}\left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)}\right)
\end{aligned}$$

as needed. □

Finally, we show that the assumption on Δ_k implies that, with probability at least $1 - \frac{\beta}{8}$, $\Delta_{\text{init}}\,\text{sym}\left(G^{(\text{init})}\right) \in \mathcal{B}$ and thus $\overline{E}\left(G^{(\text{init})}\right) = E^{(0)}$.

Claim 27. Suppose that $\Delta_k \ge 8\Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}}\right)$. With probability at least $1 - \frac{\beta}{8}$, $\left\|\Delta_{\text{init}}\,\text{sym}\left(G^{(\text{init})}\right)\right\|_2 \le \frac{1}{8}\Delta_k$ and thus $\Delta_{\text{init}}\,\text{sym}\left(G^{(\text{init})}\right) \in \mathcal{B}$ and $\overline{E}\left(G^{(\text{init})}\right) = E^{(0)}$.

Proof. By Lemma 5, we have

$$\Pr\left[\left\|\text{sym}\left(G^{(\text{init})}\right)\right\|_2 > 2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}}\right] \le \exp\left(-\frac{1}{2}\left(\sqrt{2\log\frac{8}{\beta}}\right)^2\right) = \frac{\beta}{8}.$$

Thus, with probability $1 - \beta/8$, we have

$$\Delta_{\text{init}} \left\| \text{sym}\left(G^{(\text{init})}\right) \right\|_2 \leq \Delta_{\text{init}} \left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}} \right) \leq \frac{1}{8}\Delta_k.$$

□

Thus, by combining the last two claims, we obtain that, with probability at least $1 - \frac{\beta}{4}$, simultaneously for all $0 \leq r \leq T$, we have

$$\begin{aligned}
\Phi_{r,0} &:= \max_{i\in[n]} \left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} E^{(0)} \Sigma_k^{-1} \right\|_2 \\
&= \max_{i\in[n]} \left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} \overline{E}\left(G^{(\text{init})}\right) \Sigma_k^{-1} \right\|_2 \\
&\leq \max_{i\in[n]} 2L_{i,r} \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right) \\
&= \max_{i\in[n]} \frac{12\Delta_{\text{init}}}{\Delta_k} \frac{\sigma_{k+1}^{2r+1}}{\sigma_k} \left\| e_i^\top U_{>k} \right\|_2 \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right).
\end{aligned}$$

Using that $\left\| e_i^\top U_{>k} \right\|_2 \leq \sqrt{d-k}\,\|U\|_\infty$ and the assumption $\Delta_k \geq 8\Delta_{\text{init}} \left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}} \right)$, we can further upper bound

$$\begin{aligned}
\Phi_{r,0} &\leq \frac{12\Delta_{\text{init}}}{\Delta_k} \frac{\sigma_{k+1}^{2r+1}}{\sigma_k} \sqrt{d-k}\,\|U\|_\infty \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right) \\
&\leq \frac{12}{8\left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}} \right)} \frac{\sigma_{k+1}^{2r+1}}{\sigma_k} \sqrt{d-k}\,\|U\|_\infty \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right) \\
&\leq \frac{3}{2} \frac{\sigma_{k+1}^{2r+1}}{\sigma_k} \|U\|_\infty \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right) \\
&\leq \frac{3}{2} \sigma_{k+1}^{2r} \|U\|_\infty \left(2\sqrt{k} + \sqrt{2\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right) \\
&\leq 3\sigma_{k+1}^{2r} \|U\|_\infty \left(\sqrt{k} + \sqrt{\log\left(\frac{16n\,(T+1)}{\beta}\right)} \right)
\end{aligned}$$

as needed. □

Using Lemma 21, we will show by induction that the clipping is inactive in all iterations with high probability. We will use the following lemma for the inductive step. Recall that $\mathcal{G}_{\text{clip}}$ is the event defined in Definition 15 and $\mathcal{G}_{\text{init}}$ is the event in Lemma 21.

Lemma 28. *Suppose that* $\Delta_k \geq \max\left\{ 8\Delta_{\text{init}} \left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}} \right), 8\Delta q_{\text{row}} \right\}$, $\gamma \leq \frac{\sigma_k^2}{4}$, $p_R := \frac{3\Delta R q_k}{2\sigma_k} \leq \frac{\kappa_k}{16}$, *and* $R \geq 15\,\|U\|_\infty \left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T+1)}{\beta}\right)} \right)$. *Suppose that* $\mathcal{G}_{\text{clip}}$ *and* $\mathcal{G}_{\text{init}}$ *hold. If* $s_j \leq \frac{1}{2}$ *for all* $0 \leq j \leq t$ *and clipping was inactive in all iterations* $0 \leq j \leq t-1$, *then clipping is also inactive in iteration* t.

Proof. First, let us verify that the conditions of Lemma 19 are satisfied for all iterations $j \leq t-1$. By assumption, clipping is inactive in each of those iterations, and thus it only remains to check that we have $\left\| F^{(j)} \right\|_2 < 1$ for each $j \leq t-1$. We have

$$\left\| F^{(j)} \right\|_2 = \left\| \Delta R V_k^\top \widetilde{G}^{(j)} \left(K^{(j)} \right)^{-1} \right\|_2$$

$$\leq \Delta R \left\| V_k^\top \widetilde{G}^{(j)} \right\|_2 \left\| \left(K^{(j)} \right)^{-1} \right\|_2$$
$$\leq \Delta R q_k \frac{3}{2} \frac{1}{\sigma_k} = p_R.$$

In the last inequality, we used that $\mathcal{G}_{\text{clip}}$ holds (and thus $\left\| V_k^\top \widetilde{G}^{(j)} \right\|_2 \leq q_k$) and Lemma 18 (note that the conditions of the lemma are satisfied by assumption). By assumption, we have $p_R \leq \frac{\kappa_k}{16} \leq \frac{1}{16}$, and thus $\left\| F^{(j)} \right\|_2 < 1$ as needed.

Thus, by Lemma 19, for all iterations $j \leq t-1$, we have that $C^{(j+1)}$ is invertible and

$$E^{(j+1)} = \left(\Sigma_{>k}^2 E^{(j)} \Sigma_k^{-2} + L^{(j)} \right) \left(I + F^{(j)} \right)^{-1}.$$

Recall that $S^{(j)} := \Sigma_{>k} E^{(j)} \Sigma_k^{-1}$ for all j. Multiplying the above recurrence with $\Sigma_{>k}$ on the left and Σ_k^{-1} on the right, we obtain

$$S^{(j+1)} = \left(\Sigma_{>k}^2 S^{(j)} \Sigma_k^{-1} + \Sigma_{>k} L^{(j)} \right) \left(I + F^{(j)} \right)^{-1} \Sigma_k^{-1}.$$

Using the above recurrence, we can obtain a recurrence for the potential $\Phi_{r,t}$. Recall that $L^{(j)} = \Delta R V_{>k}^\top \widetilde{G}^{(j)} \left(K^{(j)} \right)^{-1}$. For every $r \geq 0$ and every $j \leq t-1$, we have

$$\begin{aligned}
\Phi_{r,j+1} &= \max_{i \in [n]} \left\| e_i^\top U_{>k} \Sigma_{>k}^{2r} S^{(j+1)} \right\|_2 \\
&= \max_{i \in [n]} \left\| \left(e_i^\top U_{>k} \Sigma_{>k}^{2(r+1)} S^{(j)} \Sigma_k^{-1} + e_i^\top U_{>k} \Sigma_{>k}^{2r} \Sigma_{>k} L^{(j)} \right) \left(I + F^{(j)} \right)^{-1} \Sigma_k^{-1} \right\|_2 \\
&\leq \max_{i \in [n]} \left(\left\| e_i^\top U_{>k} \Sigma_{>k}^{2(r+1)} S^{(j)} \right\|_2 \left\| \Sigma_k^{-1} \right\|_2 + \Delta R \left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} V_{>k}^\top \widetilde{G}^{(j)} \right\|_2 \left\| \left(K^{(j)} \right)^{-1} \right\|_2 \right) \\
&\qquad \cdot \left\| \left(I + F^{(j)} \right)^{-1} \right\|_2 \left\| \Sigma_k^{-1} \right\|_2 \\
&\leq \left(\Phi_{r+1,j} \frac{1}{\sigma_k} + \Delta R \sigma_{k+1}^{2r} q_{\text{row}} \frac{3}{2} \frac{1}{\sigma_k} \right) \frac{1}{1 - p_R} \frac{1}{\sigma_k} \\
&= \frac{\Phi_{r+1,j}}{(1 - p_R)\, \sigma_k^2} + \frac{3 \Delta R q_{\text{row}} \sigma_{k+1}^{2r}}{2\,(1 - p_R)\, \sigma_k^2}
\end{aligned}$$

In the last inequality, we used the following inequalities: $\left\| e_i^\top U_{>k} \Sigma_{>k}^{2r+1} V_{>k}^\top \widetilde{G}^{(j)} \right\|_2 \leq \sigma_{k+1}^{2r} q_{\text{row}}$ since $\mathcal{G}_{\text{clip}}$ holds, $\left\| \left(K^{(j)} \right)^{-1} \right\|_2 \leq \frac{3}{2} \frac{1}{\sigma_k}$ by Lemma 18, and $\left\| \left(I + F^{(j)} \right)^{-1} \right\|_2 \leq \frac{1}{1 - p_R}$ which follows from Lemma 10 since $\left\| F^{(j)} \right\|_2 \leq p_R < 1$.

Unrolling the recurrence, we obtain

$$\begin{aligned}
\Phi_{0,t} &\leq \frac{\Phi_{t,0}}{\sigma_k^{2t} \left(1 - p_R\right)^t} + \frac{3 \Delta R q_{\text{row}}}{2 \sigma_k^2 \left(1 - p_R\right)} \sum_{j=0}^{t-1} \left(\frac{1 - \kappa_k}{1 - p_R} \right)^j \\
&= \frac{\Phi_{t,0}}{\sigma_k^{2t} \left(1 - p_R\right)^t} + \frac{3 \Delta R q_{\text{row}}}{2 \sigma_k^2 \left(\kappa_k - p_R\right)} \left(1 - \left(\frac{1 - \kappa_k}{1 - p_R} \right)^t \right)
\end{aligned}$$

Since $\mathcal{G}_{\text{init}}$ holds, we have $\Phi_{t,0} \leq \omega_{\text{init}} \sigma_{k+1}^{2t}$. Using this bound together with the assumptions $\Delta q_{\text{row}} \leq \frac{1}{8} \Delta_k$ and $p_R \leq \frac{\kappa_k}{16}$, we obtain

$$\begin{aligned}
\Phi_{0,t} &\leq \frac{\omega_{\text{init}} \sigma_{k+1}^{2t}}{\sigma_k^{2t} \left(1 - p_R\right)^t} + \frac{3 \Delta R q_{\text{row}}}{2 \sigma_k^2 \left(\kappa_k - p_R\right)} \left(1 - \left(\frac{1 - \kappa_k}{1 - p_R} \right)^t \right) \\
&= \omega_{\text{init}} \left(\frac{1 - \kappa_k}{1 - p_R} \right)^t + \frac{3 \Delta R q_{\text{row}}}{2 \sigma_k^2 \left(\kappa_k - p_R\right)} \left(1 - \left(\frac{1 - \kappa_k}{1 - p_R} \right)^t \right)
\end{aligned}$$

$$\leq \omega_{\text{init}} + \frac{3\Delta R q_{\text{row}}}{2\sigma_k^2 \frac{15}{16}\kappa_k} = \omega_{\text{init}} + \frac{8\Delta R q_{\text{row}}}{5\Delta_k} \leq \omega_{\text{init}} + \frac{1}{5}R$$

Combining with Lemma 20 and using that $R \geq \max\left\{15\sqrt{k}\left\|U\right\|_\infty, 5\omega_{\text{init}}\right\}$, we obtain

$$\begin{aligned}\max_{i\in[n]}\left\|a_i^\top X^{(t)} B^{(t)}\right\|_2 &\leq \sqrt{2}\left(\sqrt{k}\left\|U\right\|_\infty + \Phi_{0,t}\right)\\ &\leq \sqrt{2}\left(\sqrt{k}\left\|U\right\|_\infty + \omega_{\text{init}} + \frac{1}{5}R\right)\\ &\leq \sqrt{2}\left(\frac{1}{15}R + \frac{1}{5}R + \frac{1}{5}R\right) \leq R\end{aligned}$$

and thus clipping is inactive in iteration t as well. □

Convergence analysis. Next, using Lemmas 19 and 28, we analyze s_t and m_t for all iterations $t \geq 0$. We proceed by induction and show that the clipping is inactive and thus we can apply the main recurrence shown in Lemma 19. Recall that the recurrence states that $E^{(t+1)} = \left(\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2} + L^{(t)}\right)\left(I + F^{(t)}\right)^{-1}$. The term $\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}$ captures the main progress, and we have $\left\|\Sigma_{>k}^2 E^{(t)} \Sigma_k^{-2}\right\|_2 \leq \frac{\sigma_{k+1}^2}{\sigma_k^2}\left\|E^{(t)}\right\|_2 = (1-\kappa_k)\, m_t$. Thus in an ideal case where $L^{(t)}$ and $F^{(t)}$ are zero, m_t decreases geometrically in every iteration. In our setting, $L^{(t)}$ and $F^{(t)}$ are non-zero and they depend on the Gaussian noise matrix $\widetilde{G}^{(t)}$, and our analysis will upper bound their contribution. We start by showing that the initial solution satisfies $s_0 \leq \frac{1}{4}$ and thus $m_0 \leq \frac{1}{\sqrt{15}}$, which will provide the base case for the induction.

Lemma 29. *Suppose that* $\Delta_k \geq 8\Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log\frac{4}{\beta}}\right)$. *With probability at least* $1-\frac{\beta}{4}$, $s_0 \leq \frac{1}{4}$.

Proof. Let $W = \Delta_{\text{init}} \operatorname{sym}\left(G^{(\text{init})}\right)$, and thus the columns of the initial solution $X^{(0)}$ are the top-k eigenvectors $X^{(0)}$ of $\widehat{M} = A^\top A + W$. By Lemma 5, we have

$$\Pr\left[\left\|\operatorname{sym}\left(G^{(\text{init})}\right)\right\|_2 > 2\sqrt{d} + \sqrt{2\log\frac{4}{\beta}}\right] \leq \exp\left(-\frac{1}{2}\left(\sqrt{2\log\frac{4}{\beta}}\right)^2\right) = \frac{\beta}{4}.$$

In the following, we condition on the event $\left\|\operatorname{sym}\left(G^{(\text{init})}\right)\right\|_2 \leq 2\sqrt{d} + \sqrt{2\log\frac{4}{\beta}}$. We have

$$\left\|W\right\|_2 = \Delta_{\text{init}}\left\|\operatorname{sym}\left(G^{(\text{init})}\right)\right\|_2 \leq \Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log\frac{4}{\beta}}\right) \leq \frac{1}{8}\Delta_k.$$

Thus Lemma 9 gives

$$s_0 = \left\|\left(I - X^{(0)}X^{(0)\top}\right)V_k\right\|_2 \leq \frac{\left\|W\right\|_2}{\Delta_k - \left\|W\right\|_2} \leq \frac{2\left\|W\right\|_2}{\Delta_k} \leq \frac{1}{4}.$$

□

Before giving our inductive analysis, we first show the following lemma which will be used to upper bound one of the error terms in the main recurrence. The lemma follows from a martingale argument. We first introduce some notation. We define a filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ as follows. For $t < T$, $\mathcal{F}_t$ includes $G^{(\text{init})}$, $G^{(j)}$ and $\widetilde{G}^{(j)}$ for all iterations $0 \leq j < t$ prior to t, and $G^{(t)}$; $\widetilde{G}^{(t)}$ is not included in $\mathcal{F}_t$ and is independent of $\mathcal{F}_t$. We let $\mathcal{F}_T$ include the full transcript.

Let τ be the following stopping time:

$$\tau := \inf\left\{0 \leq t < T\colon C^{(t)} \text{ is not invertible}, s_t > \frac{1}{2}, \text{ or } \left\|\Delta \operatorname{sym}\left(G^{(t)}\right)\right\|_2 > \frac{\gamma}{4}\right\} \tag{5}$$

with the convention that $\inf\emptyset = T$.

We define

$$W^{(t)} := \begin{cases} \left(K^{(t)}\right)^{-1} & t < \tau \\ 0 & t \geq \tau \end{cases} \qquad \widehat{L}^{(t)} := \Delta R V_{>k}^\top \widetilde{G}^{(t)} W^{(t)}.$$

Note that $\widehat{L}^{(t)} = L^{(t)}$ for $t < \tau$ and $\widehat{L}^{(t)} = 0$ for $t \geq \tau$.

Finally, for a matrix M, we let

$$\mathcal{T}(M) := \Sigma_{>k}^2 M \Sigma_k^{-2}.$$

Lemma 30. *Suppose that $\gamma \leq \frac{\sigma_k^2}{4}$. Let*

$$\nu_R := \frac{3\Delta R}{\sigma_k} \sqrt{\frac{2d \log 9 + 2 \log(8T/\beta)}{\kappa_k (2 - \kappa_k)}}.$$

With probability at least $1 - \frac{\beta}{4}$, simultaneously for all $1 \leq t \leq T$, we have

$$\left\| \sum_{j=0}^{t-1} \mathcal{T}^{t-1-j} \left(\widehat{L}^{(j)} \right) \right\|_2 \leq \nu_R. \tag{6}$$

Let $\mathcal{G}_{\mathrm{mart}}$ be this event.

Proof. Fix an iteration $1 \leq t \leq T$ and two unit vectors $u \in \mathbb{R}^{d-k}$ and $v \in \mathbb{R}^k$. For every $0 \leq j < t$, let

$$\xi_j := u^\top \mathcal{T}^{t-1-j} \left(\widehat{L}^{(j)} \right) v = u^\top \Sigma_{>k}^{2(t-1-j)} \widehat{L}^{(j)} \Sigma_k^{-2(t-1-j)} v.$$

Recall that $\widehat{L}^{(j)} := \Delta R V_{>k}^\top \widetilde{G}^{(j)} W^{(j)}$ and $\mathcal{F}_j$ includes all random matrices in the iterations prior to j and $G^{(j)}$, but not $\widetilde{G}^{(j)}$. Thus, conditioned on $\mathcal{F}_j$, ξ_j is a Gaussian random variable with expectation 0 and variance

$$\begin{aligned}
\mathbb{E}\left[\xi_j^2 | \mathcal{F}_j\right] &= (\Delta R)^2 \left\| \Sigma_{>k}^{2(t-1-j)} u \right\|_2^2 \left\| W^{(j)} \Sigma_k^{-2(t-1-j)} v \right\|_2^2 \\
&\leq (\Delta R)^2 \left\| \Sigma_{>k}^{2(t-1-j)} \right\|_2^2 \|u\|_2^2 \left\| W^{(j)} \right\|_2^2 \left\| \Sigma_k^{-2(t-1-j)} \right\|_2^2 \|v\|_2^2 \\
&= (\Delta R)^2 \left(\frac{\sigma_{k+1}^2}{\sigma_k^2} \right)^{2(t-1-j)} \left\| W^{(j)} \right\|_2^2 \\
&\leq (\Delta R)^2 \left(\frac{\sigma_{k+1}^2}{\sigma_k^2} \right)^{2(t-1-j)} \left(\frac{3}{2\sigma_k} \right)^2 \\
&= \left(\frac{3\Delta R}{2\sigma_k} \right)^2 (1 - \kappa_k)^{2(t-1-j)}.
\end{aligned}$$

The last inequality follows from Lemma 18: if $j < \tau$, we have $W^{(j)} = \left(K^{(j)}\right)^{-1}$ and the conditions of Lemma 18 are satisfied; otherwise, $W^{(j)} = 0$.

We can use the subgaussian version of Azuma's inequality to bound $\sum_j \xi_j$ (see e.g. Duchi et al. (2012, Lemma F.4) or Howard et al. (2020, Theorem 1(b)) with $\psi(\lambda) = \lambda^2/2$). For every $x \geq 0$, we have

$$\begin{aligned}
\Pr\left[\left| \sum_{j=0}^{t-1} \xi_j \right| \geq x \right] &\leq 2 \exp\left(- \frac{x^2}{2 \sum_{j=0}^{t-1} \left(\frac{3\Delta R}{2\sigma_k} \right)^2 (1 - \kappa_k)^{2(t-1-j)}} \right) \\
&\leq 2 \exp\left(- \frac{x^2 \kappa_k (2 - \kappa_k)}{2 \left(\frac{3\Delta R}{2\sigma_k} \right)^2} \right).
\end{aligned}$$

We can use the above guarantee to upper bound $\left\|\sum_{j=0}^{t-1}\mathcal{T}^{t-1-j}\left(\widehat{L}^{(j)}\right)\right\|_2$ using a net argument, similarly to the proof of Claim 28 and 26. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be $\frac{1}{4}$-nets of the two unit spheres in $\mathbb{R}^{d-k}$ and $\mathbb{R}^k$ respectively, of sizes $|\mathcal{N}_1| \le 9^{d-k}$ and $|\mathcal{N}_2| \le 9^k$. For any $(d-k)\times k$ matrix M, we have $\|M\|_2 \le 2\max_{u\in\mathcal{N}_1, v\in\mathcal{N}_2}\left|u^\top M v\right|$. Indeed, let u^*, v^* be unit left and right top singular vectors of M and leet u, v the closest vectors in the nets to u^*, v^*. We have

$$\|M\|_2 = u^{*\top} M v^* = u^{*\top} M(v^* - v) + (u^{*\top} - u^\top) M v + u^\top M v \le \frac{1}{4}\|M\|_2 + \frac{1}{4}\|M\|_2 + u^\top M v.$$

By rearranging, we obtain the claimed inequality.

There are $|\mathcal{N}_1|\,|\mathcal{N}_2|\,T = 9^d T$ triples (u, v, t) with $u \in \mathcal{N}_1$, $v \in \mathcal{N}_2$, and $1 \le t \le T$. Setting $x = \frac{3\Delta R}{2\sigma_k}\sqrt{\frac{2\log\left(\frac{8\cdot 9^d T}{\beta}\right)}{\kappa_k(2-\kappa_k)}}$ in the guarantee above and taking an union bound over all $9^d T$ triples, we obtain the following guarantee. With probability at least $1 - \frac{\beta}{4}$, simultaneously for all $1 \le t \le T$,

$$\begin{aligned}\left\|\sum_{j=0}^{t-1}\mathcal{T}^{t-1-j}\left(\widehat{L}^{(j)}\right)\right\|_2 &\le 2\max_{u\in\mathcal{N}_1, v\in\mathcal{N}_2}\left|u^\top M v\right| \\ &\le \frac{3\Delta R}{\sigma_k}\sqrt{\frac{2\log\left(\frac{8\cdot 9^d T}{\beta}\right)}{\kappa_k\,(2-\kappa_k)}} = \nu_R,\end{aligned}$$

as needed. □

We now give the main inductive argument. Recall that $\mathcal{G}_{\text{clip}}$ is the event defined in Definition 15, and $\mathcal{G}_{\text{init}}$ and $\mathcal{G}_{\text{mart}}$ are the events defined in Lemma 21 and Lemma 30 respectively. Recall also that $p_R := \frac{3\Delta R q_k}{2\sigma_k}$ (see Lemma 28) and $\nu_R := \frac{3\Delta R}{\sigma_k}\sqrt{\frac{2d\log 9 + 2\log(8T/\beta)}{\kappa_k(2-\kappa_k)}}$ (see Lemma 30).

Theorem 31. *Suppose that we have* $\Delta_k \ge \max\left\{8\Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}}\right), 8\Delta q_{\text{row}}\right\}$, $\gamma \le \frac{\sigma_k^2}{4}$, $p_R \le \frac{\kappa_k}{16}$, $R \ge 15\,\|U\|_\infty\left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T+1)}{\beta}\right)}\right)$, *and* $\nu_R \le \frac{1}{8}$. *Suppose that* $\mathcal{G}_{\text{clip}}$, $\mathcal{G}_{\text{init}}$, $\mathcal{G}_{\text{mart}}$, *and the event in Lemma 21 and 29 hold. Then, in every iteration,* $C^{(t)}$ *is invertible, clipping is inactive, and*

$$s_t \le m_t \le \left(\frac{1-\kappa_k}{1-p_R}\right)^t m_0 + 2\nu_R \le \left(1 - \frac{\kappa_k}{2}\right)^t m_0 + 2\nu_R. \tag{7}$$

Proof. In the following, we will use the definition of the stopping time τ (eqn. (5)) and the identity $s_t = \frac{m_t}{\sqrt{1+m_t^2}}$ (Lemma 14). We will prove by induction on $t \ge 0$ that $C^{(j)}$ is invertible and $m_j \le \frac{1}{\sqrt{3}}$ (and thus $s_j \le \frac{1}{2}$) for all $j \le t$, clipping is inactive in all iterations $j < t$, $\tau > t$ (if $t = T$, $\tau = T$), and (7) holds for all iterations $j \le t$. Consider the base case $t = 0$. Since the event in Lemma 21 and 29 holds, we have $s_0 \le \frac{1}{4}$ and thus $m_0 \le \frac{1}{\sqrt{15}}$. Since $s_0 < 1$, Lemma 13 implies that $C^{(0)}$ is invertible. Additionally, $\tau > 0$ and (7) holds for $t = 0$.

Next, we consider the inductive step. Consider $t \ge 0$ and suppose that $C^{(j)}$ is invertible and $m_j \le \frac{1}{\sqrt{3}}$ (and thus $s_j \le \frac{1}{2}$) for all $j \le t$, and clipping is inactive in all iterations $j < t$, and $\tau \ge t$. These assumptions and the assumptions in the theorem statement imply that Lemma 28 applies and thus clipping is inactive in iteration t as well.

For each $j \le t$, we can show that $F^{(j)} := \Delta R V_k^\top \widetilde{G}^{(j)}\left(K^{(j)}\right)^{-1}$ satisfies $\left\|F^{(j)}\right\|_2 \le p_R$ as follows. We have

$$\begin{aligned}\left\|F^{(j)}\right\|_2 &= \left\|\Delta R V_k^\top \widetilde{G}^{(j)}\left(K^{(j)}\right)^{-1}\right\|_2 \\ &\le \Delta R\left\|V_k^\top \widetilde{G}^{(j)}\right\|_2\left\|\left(K^{(j)}\right)^{-1}\right\|_2\end{aligned}$$

$$\leq \Delta R q_k \frac{3}{2}\frac{1}{\sigma_k} = p_R.$$

In the last inequality, we used that $\mathcal{G}_{\text{clip}}$ holds (and thus $\left\|V_k^\top \widetilde{G}^{(j)}\right\|_2 \leq q_k$) and Lemma 18. Note that the conditions of Lemma 18 hold: we have $\left\|\Delta \operatorname{sym}\left(G^{(j)}\right)\right\|_2 \leq \frac{\gamma}{4}$ since $\mathcal{G}_{\text{clip}}$ holds, $\gamma \leq \frac{\sigma_k^2}{4}$, and $s_j \leq \frac{1}{2}$ for all $j \leq t$.

For all $0 \leq j \leq t$, we have $\left\|F^{(j)}\right\|_2 \leq p_R \leq \frac{\kappa_k}{16} < 1$, $C^{(j)}$ is invertible, and clipping is inactive in iteration j. Thus Lemma 19 gives that $C^{(t+1)}$ is invertible and the following recurrence holds for all $0 \leq j \leq t$:

$$E^{(j+1)} = \left(\Sigma_{>k}^2 E^{(j)} \Sigma_k^{-2} + L^{(j)}\right)\left(I + F^{(j)}\right)^{-1} = \left(\mathcal{T}\left(E^{(j)}\right) + L^{(j)}\right)\left(I + F^{(j)}\right)^{-1} \tag{8}$$

where we used the notation $\mathcal{T}(M) = \Sigma_{>k}^2 M \Sigma_k^{-2}$ defined earlier. Using the above recurrence, we obtain the following.

Claim 32. Let

$$b_R := \Delta R q_{d-k} \frac{3}{2\sigma_k} \quad \eta_R := \nu_R + \frac{b_R p_R}{\kappa_k (1 - p_R)} \quad h_R := \frac{(1-\kappa_k) p_R}{1 - p_R}.$$

For every $1 \leq \ell \leq t+1$, we have

$$m_\ell \leq (1-\kappa_k)^\ell m_0 + \eta_R + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} m_j. \tag{9}$$

Proof. Using the identity $(I+F)^{-1} = I - F(I+F)^{-1}$, we can write the recurrence (8) as follows:

$$E^{(j+1)} = \mathcal{T}\left(E^{(j)}\right) + L^{(j)} \underbrace{- \left(\mathcal{T}\left(E^{(j)}\right) + L^{(j)}\right) F^{(j)} \left(I + F^{(j)}\right)^{-1}}_{:= R^{(j)}} = \mathcal{T}\left(E^{(j)}\right) + L^{(j)} + R^{(j)} \tag{10}$$

We rearranged the recurrence in order to separate the noise terms depending on $\widetilde{G}^{(j)}$ into the term $L^{(j)}$ that is linear in $\widetilde{G}^{(j)}$ and a term $R^{(j)}$ which has a more complex dependency on $\widetilde{G}^{(j)}$. This separation allows us to bound the contribution of the linear terms $L^{(j)}$ via the martingale argument in Lemma 30. The terms $R^{(j)}$ are more complex, and we will bound their contribution by bounding the norm of each $R^{(j)}$ separately.

For each $1 \leq \ell \leq t+1$, by unrolling the recurrence (10) and using that $L^{(j)} = \widehat{L}^{(j)}$ since $\tau > t$, we obtain

$$\begin{aligned} E^{(\ell)} &= \mathcal{T}^\ell\left(E^{(0)}\right) + \sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(L^{(j)}\right) + \sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(R^{(j)}\right) \\ &= \mathcal{T}^\ell\left(E^{(0)}\right) + \sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(\widehat{L}^{(j)}\right) + \sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(R^{(j)}\right). \end{aligned}$$

Thus, for each $1 \leq \ell \leq t+1$, we have

$$m_\ell = \left\|E^{(\ell)}\right\|_2 \leq \left\|\mathcal{T}^\ell\left(E^{(0)}\right)\right\|_2 + \left\|\sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(\widehat{L}^{(j)}\right)\right\|_2 + \sum_{j=0}^{\ell-1} \left\|\mathcal{T}^{\ell-1-j}\left(R^{(j)}\right)\right\|_2.$$

We have

$$\begin{aligned} \left\|\mathcal{T}^\ell\left(E^{(0)}\right)\right\|_2 &= \left\|\Sigma_{>k}^{2\ell} E^{(0)} \Sigma_k^{-2\ell}\right\|_2 \\ &\leq \left\|\Sigma_{>k}^{2\ell}\right\|_2 \left\|E^{(0)}\right\|_2 \left\|\Sigma_k^{-2\ell}\right\|_2 \end{aligned}$$

$$= \frac{\sigma_{k+1}^{2\ell}}{\sigma_k^{2\ell}} m_0 = (1-\kappa_k)^{\ell} m_0.$$

Since $\mathcal{G}_{\mathrm{mart}}$ holds, for all $1 \le \ell \le t+1$, we have

$$\left\| \sum_{j=0}^{\ell-1} \mathcal{T}^{\ell-1-j}\left(\widehat{L}^{(j)}\right) \right\|_2 \le \nu_R.$$

We have

$$\begin{aligned}
\left\|R^{(j)}\right\|_2 &= \left\| -\left(\mathcal{T}\left(E^{(j)}\right) + L^{(j)}\right) F^{(j)} \left(I + F^{(j)}\right)^{-1} \right\|_2 \\
&\le \left(\left\|\mathcal{T}\left(E^{(j)}\right)\right\|_2 + \left\|L^{(j)}\right\|_2 \right) \left\|F^{(j)}\right\|_2 \left\| \left(I + F^{(j)}\right)^{-1} \right\|_2
\end{aligned}$$

Additionally,

$$\begin{aligned}
\left\|\mathcal{T}\left(E^{(j)}\right)\right\|_2 &= \left\|\Sigma_{>k}^2 E^{(j)} \Sigma_k^{-2}\right\|_2 \le \left\|\Sigma_{>k}^2\right\|_2 \left\|E^{(j)}\right\|_2 \left\|\Sigma_k^{-2}\right\|_2 \\
&= \frac{\sigma_{k+1}^2}{\sigma_k^2} m_j = (1-\kappa_k)\, m_j.
\end{aligned}$$

Using that $\left\|V_{>k}^{\top} \widetilde{G}^{(t)}\right\|_2 \le q_{d-k}$ (since $\mathcal{G}_{\mathrm{clip}}$ holds) and Lemma 18, we obtain

$$\begin{aligned}
\left\|L^{(j)}\right\|_2 &= \Delta R \left\| V_{>k}^{\top} \widetilde{G}^{(j)} \left(K^{(j)}\right)^{-1} \right\|_2 \\
&\le \Delta R \left\|V_{>k}^{\top} \widetilde{G}^{(j)}\right\|_2 \left\| \left(K^{(j)}\right)^{-1} \right\|_2 \\
&\le \underbrace{\Delta R q_{d-k} \frac{3}{2\sigma_k}}_{:= b_R}
\end{aligned}$$

As shown earlier, we have $\left\|F^{(j)}\right\|_2 \le p_R$. Since $\left\|F^{(j)}\right\|_2 \le p_R \le \frac{\kappa_k}{16} < 1$, Lemma 10 gives $\left\| \left(I + F^{(j)}\right)^{-1} \right\|_2 \le \frac{1}{1-p_R}$.

Putting everything together, we obtain

$$\left\|R^{(j)}\right\|_2 \le \left((1-\kappa_k)\, m_j + b_R \right) \frac{p_R}{1-p_R} =: h_R m_j + f_R,$$

where we let $h_R := \frac{(1-\kappa_k) p_R}{1-p_R}$ and $f_R := \frac{b_R p_R}{1-p_R}$.

Thus

$$\begin{aligned}
\left\|\mathcal{T}^{\ell-1-j}\left(R^{(j)}\right)\right\|_2 &= \left\|\Sigma_{>k}^{2(\ell-1-j)} R^{(j)} \Sigma_k^{-2(\ell-1-j)}\right\|_2 \\
&\le \left\|\Sigma_{>k}^{2(\ell-1-j)}\right\|_2 \left\|R^{(j)}\right\|_2 \left\|\Sigma_k^{-2(\ell-1-j)}\right\|_2 \\
&\le \frac{\sigma_{k+1}^{2(\ell-1-j)}}{\sigma_k^{2(\ell-1-j)}} \left\|R^{(j)}\right\|_2 \\
&= (1-\kappa_k)^{\ell-1-j} \left\|R^{(j)}\right\|_2 \\
&\le (1-\kappa_k)^{\ell-1-j} \left(h_R m_j + f_R\right)
\end{aligned}$$

It follows that

$$m_\ell \le (1-\kappa_k)^{\ell} m_0 + \nu_R + \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} \left(h_R m_j + f_R\right)$$

$$\leq (1-\kappa_k)^{\ell} m_0 + \underbrace{\nu_R + \frac{f_R}{\kappa_k}}_{:=\eta_R} + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} m_j$$

as needed. □

We can prove by induction that 9 implies the following.

Claim 33. For every $1 \leq \ell \leq t+1$, we have

$$m_\ell \leq \left(\frac{1-\kappa_k}{1-p_R}\right)^{\ell} m_0 + \frac{\kappa_k}{\kappa_k - h_R}\eta_R. \tag{11}$$

Proof. We prove by induction on ℓ that

$$m_\ell \leq (1-\kappa_k+h_R)^{\ell} m_0 + \eta_R \left(1 + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k+h_R)^j\right).$$

The base case $\ell = 1$ follows from (9). In the inductive step, we have

$$\begin{aligned} m_\ell &\leq (1-\kappa_k)^{\ell} m_0 + \eta_R + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} m_j \\ &\leq (1-\kappa_k)^{\ell} m_0 + \eta_R + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} \left((1-\kappa_k+h_R)^j m_0 + \eta_R \left(1 + h_R \sum_{r=0}^{j-1} (1-\kappa_k+h_R)^r\right)\right) \end{aligned}$$

First, let us consider the coefficient of m_0. Since we have $a^\ell = b^\ell + (a-b)\left(\sum_{j=0}^{\ell-1} a^j b^{\ell-1-j}\right)$ for all a, b, we can add the coefficients of m_0 and obtain

$$(1-\kappa_k)^{\ell} + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} (1-\kappa_k+h_R)^j = (1-\kappa_k+h_R)^{\ell}.$$

Next, we consider the coefficient of η_R. We have

$$\begin{aligned} &1 + h_R \sum_{j=0}^{\ell-1} (1-\kappa_k)^{\ell-1-j} \left(1 + h_R \sum_{r=0}^{j-1} (1-\kappa_k+h_R)^r\right) \\ &= 1 + h_R \sum_{i=0}^{\ell-1} \left((1-\kappa_k)^i + h_R \sum_{j=0}^{i-1} (1-\kappa_k)^j (1-\kappa_k+h_R)^{i-1-j}\right) \\ &= 1 + h_R \sum_{i=0}^{\ell-1} (1-\kappa_k+h_R)^i \end{aligned}$$

which completes the induction.

Finally, using that $h_R := \frac{(1-\kappa_k)p_R}{1-p_R}$ and the geometric sum formula, we obtain the claim. □

By (11), for every $1 \leq \ell \leq t+1$, we have

$$m_\ell \leq \left(\frac{1-\kappa_k}{1-p_R}\right)^{\ell} m_0 + \frac{\kappa_k}{\kappa_k - h_R}\eta_R.$$

Note that $b_R \leq \nu_R$, since $q_{d-k} \leq 2\sqrt{d + \log(32T/\beta)}$. The assumption $p_R \leq \frac{\kappa_k}{16}$ implies $h_R \leq p_R \leq \frac{\kappa_k}{16}$, and $\frac{p_R}{\kappa_k}\frac{b_R}{1-p_R} \leq \frac{b_R}{15} \leq \frac{\nu_R}{15}$.

$$\eta_R := \nu_R + \frac{b_R p_R}{\kappa_k (1-p_R)} \leq \frac{16}{15}\nu_R$$

and

$$\frac{\kappa_k}{\kappa_k - h_R}\eta_R \le \left(\frac{16}{15}\right)^2 \nu_R.$$

Thus we obtain

$$m_\ell \le \left(\frac{1-\kappa_k}{1-p_R}\right)^\ell m_0 + \left(\frac{16}{15}\right)^2 \nu_R \le \left(1 - \frac{\kappa_k}{2}\right)^\ell m_0 + \left(\frac{16}{15}\right)^2 \nu_R,$$

which proves that (7) holds for all iterations $\ell \le t+1$.

In particular, we have

$$m_{t+1} \le m_0 + 2\nu_R \le \frac{1}{\sqrt{15}} + \frac{1}{4} < \frac{1}{\sqrt{3}}.$$

This concludes the main induction, including showing that $\tau > t+1$ (if $t+1 = T$, $\tau = T$).

□

Finally, we show the following corollary of the main theorem for ideal choices of T and R that depend on the data. In Section C, we show how to privately obtain an analogous guarantee via a private hyperparameter search.

Corollary 34. *Suppose that* $R \ge 15\|U\|_\infty \left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T+1)}{\beta}\right)}\right)$, $T = \left\lceil \frac{2}{\kappa_k} \log\left(\frac{n}{\beta\delta\epsilon}\right)\right\rceil$, *and*

$$\Delta_k \ge \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon} + \left(\frac{\sigma_k\sqrt{k}}{\epsilon}\right)^{2/3} + \left(\frac{R\sigma_k^2\sqrt{k}}{\epsilon}\right)^{2/3}\right),$$

where the $\widetilde{\Omega}(\cdot)$ *hides sufficiently large poly-logarithmic factors. With probability at least* $1-\beta$, *the solution* $X^{(T)}$ *returned by Algorithm 3 satisfies*

$$\left\|\left(I - X^{(T)}X^{(T)\top}\right)V_k\right\|_2^2 \le \left(\frac{\beta\delta\epsilon}{n}\right)^2 + \widetilde{O}\left(\frac{d\sigma_k^2 R^2}{\epsilon^2\Delta_k^2}\right).$$

In particular, if $R = \widetilde{\Theta}\left(\sqrt{k}\|U\|_\infty\right)$, *we have*

$$\left\|\left(I - X^{(T)}X^{(T)\top}\right)V_k\right\|_2^2 \le \left(\frac{\beta\delta\epsilon}{n}\right)^2 + \widetilde{O}\left(\frac{dk\sigma_k^2\|U\|_\infty^2}{\epsilon^2\Delta_k^2}\right).$$

Proof. Let us verify that the assumption imply the conditions in Theorem 31. The choice of T gives

$$\Delta = \widetilde{O}\left(\frac{1}{\epsilon\sqrt{\kappa_k}}\right) = \widetilde{O}\left(\frac{\sigma_k}{\epsilon\sqrt{\Delta_k}}\right).$$

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\left(\frac{\sigma_k\sqrt{k}}{\epsilon}\right)^{2/3}\right)$ implies $\Delta q_{\text{row}} = \widetilde{O}\left(\frac{\sigma_k\sqrt{k}}{\epsilon\sqrt{\Delta_k}}\right) = O(\Delta_k)$ and also implies $\gamma = \widetilde{O}\left(\Delta\sqrt{k}\right) = O(\Delta_k) < \frac{\sigma_k^2}{4}$.

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\left(\frac{R\sigma_k^2\sqrt{k}}{\epsilon}\right)^{2/3}\right)$ implies

$$p_R = \frac{3\Delta R q_k}{2\sigma_k} = \widetilde{O}\left(\frac{R\sqrt{k}}{\epsilon\sqrt{\Delta_k}}\right) \le \frac{\Delta_k}{16\sigma_k^2} = \frac{\kappa_k}{16}.$$

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon}\right)$ ensures that $\Delta_k \ge 8\Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta}}\right)$.

By Lemmas 16, 21, 30, and 29, the events $\mathcal{G}_{\text{clip}}$, $\mathcal{G}_{\text{init}}$, $\mathcal{G}_{\text{mart}}$, and the event $s_0 \le \frac{1}{4}$ hold jointly with probability at least $1-\beta$.

If $\nu_R \le \frac{1}{8}$, all of the conditions of Theorem 31 are satisfied and the theorem implies the desired guarantee. Otherwise, we have $s_T^2 \le 1 \le 64\nu_R^2$, and we have the same asymptotic guarantee.

□

C PRIVATE SELECTION OF THE CLIPPING RADIUS R AND ITERATION COUNT T

Both the clipping radius R and the iteration count T in Corollary 34 depend on the private data. We can try all of possible values for them from powers of two in an appropriate range and return the approximately best solution using the exponential mechanism, incurring an overhead logarithmic in the ranges (because we only consider powers of two). To be more economical, we select them jointly using the private hyperparameter-selection method of Papernot and Steinke Papernot and Steinke (2022). A candidate consists of a pair (R, T) and the subspace produced by that run. Each run receives the same total Renyi privacy budget, so runs with fewer iterations use smaller per-iteration noise. The guarantee compares the returned subspace with the best pair satisfying the utility hypotheses in the following public grid.

Fix failure probability $\beta \in (0, 1)$ and public grids of hyperparameters

$$\begin{aligned} \mathcal{R} &= \{R_j = 2^j R_{\min} : 0 \le j \le J\}, & m_R &= J + 1, \\ \mathcal{T}_{\text{grid}} &= \{T_\ell = 2^\ell T_{\min} : 0 \le \ell \le H\}, & m_T &= H + 1, \\ T_{\max} &= 2^H T_{\min}, R_{\max} = 2^J R_{\min} & m &= m_R m_T, \end{aligned} \tag{12}$$

All endpoints are fixed independently of the data. On any dataset satisfying the warm-start gap condition, $\Delta_k \ge 16\Delta_{\text{init}}\sqrt{d}$, whereas $\sigma_k^2 \le \operatorname{tr}(A^\top A) \le n$. Consequently $\kappa_k \ge \Delta_{\text{init}}\sqrt{d}/n$ in the regime with utility guarantees. Thus we can set $T_{\min} = 1, T_{\max} = \frac{2n}{\Delta_{\text{init}}\sqrt{d}} \log\left(\frac{n}{\beta\delta\epsilon}\right)$.

Since $\sqrt{1/n} \le \|U\|_\infty \le 1$, we can set $R_{\min} = \sqrt{k/n}$. Because $\|U\|_\infty \le 1$, we can set $R_{\max} = 15\left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T+1)}{\beta}\right)}\right)$.

We run the initialization algorithm (Algorithm (1)) only once with failure probability $\beta_0 = \beta/2$. We use this initial solutions for all of the runs of the optimization algorithm (Algorithm 2). The optimization runs using different choices for radius and iteration count will use the following parameters. For public $\Delta_{\text{init}}, \chi, \zeta > 0$, a run with t iterations uses

$$\Delta(t) = \chi\sqrt{t}, \qquad \gamma(t) = 4\Delta(t)\left(2\sqrt{k} + \sqrt{2\log(4t/\beta_0)}\right) \tag{13}$$

The noise parameters will be adjusted and slightly different from the even allocation between initialization and optimization in Algorithm 3. Let $q_k(t), q_{\text{row}}(t), p_R(t), \nu_R(t)$ be the quantities in the analysis as functions of the iteration count t, noise scale $\Delta(t)$, and failure probability β_0. In particular,

$$\begin{aligned} p_R(t) &= \frac{3\Delta(t)Rq_k(t)}{2\sigma_k}, \\ \nu_R(t) &= \frac{3\Delta(t)R}{\sigma_k}\sqrt{\frac{2\big(d\log 9 + \log(32t/\beta_0)\big)}{\kappa_k(2-\kappa_k)}}. \end{aligned}$$

C.1 A PRIVATE SCORE FOR COMPARING CANDIDATE SUBSPACES

In this section, we describe the score to compare the candidate subspaces. Ideally we would like to use $\|(I - XX^\top)V_k V_k^\top\|_2$ as the score but a more robust version of this is required to reduce the sensitivity of the score when the inputs do not have a large eigenvalue gap.

Recall that the initialization algorithm (Algorithm 1) produces the matrix

$$\widehat{M} = A^\top A + \Delta_{\text{init}} \operatorname{sym}(G^{(\text{init})}), \qquad e_{\text{w}} := \Delta_{\text{init}}\left(2\sqrt{d} + \sqrt{2\log(8/\beta)}\right).$$

Write its eigenvalues as $\widehat{\lambda}_1 \ge \cdots \ge \widehat{\lambda}_d$ and let $X^{(0)}$ be its top-k eigenbasis. Set

$$a = \widehat{\lambda}_{k+1} + e_{\text{w}}, \qquad b = \widehat{\lambda}_k - e_{\text{w}}, \qquad h = b - a. \tag{14}$$

If $h \le 0$, return $(R_J, 0, X^{(0)})$ immediately. Intuitively this happens when the eigenvalue gap is too small compared with the noise. If $h > 0$, consider the following scalar function

$$f_{a,b}(x) = \begin{cases} 0, & x \le a, \\ (x-a)/h, & a < x < b, \\ 1, & x \ge b, \end{cases}$$

Let $f_{a,b}(M)$ be the function that keeps all the singular vectors of M but applies the scalar function $f_{a,b}(x)$ to all the singular values (this is called the matrix functional calculus of $f_{a,b}$). For a candidate solution $X \in \mathbb{R}^{d\times k}$ with orthonormal columns, its score is

$$\mathsf{s}_{a,b}(A, X) := \|(I - XX^\top) f_{a,b}(A^\top A)\|_2 \in [0, 1]. \tag{15}$$

The following lemma bounds the sensitivity of $f_{a,b}(\cdot)$ for all inputs regardless of the eigenvalue gap and proves that for inputs with large singular value gaps, $f_{a,b}(\cdot)$ is identical to the top singular value subspace.

Lemma 35 (Sensitivity and accuracy of the comparison score)**.** *For each fixed initialization with $h > 0$, and every fixed orthonormal X, the replacement sensitivity of $\mathsf{s}_{a,b}(A, X)$ is at most $2/h$. If $\|\widehat{M} - A^\top A\|_2 \le e_{\mathrm{w}}, \Delta_k \ge 8e_{\mathrm{w}}$ then $h \ge \Delta_k/2$, we have $f_{a,b}(A^\top A) = V_k V_k^\top$ and consequently, simultaneously for all X, $\mathsf{s}_{a,b}(A, X) = \|(I - XX^\top)V_k\|_2$.*

Proof. The scalar function $f_{a,b}$ is $1/h$-Lipschitz. Consider symmetric matrices M, M' with orthonormal eigenbases u_i, v_j and eigenvalues λ_i, λ'_j respectively. We will show that

$$\sum_{i,j} (\lambda_i - \lambda'_j)^2 |u_i^\top v_j|^2 = \|M - M'\|_F^2 \tag{16}$$

Because the Frobenius norm does not change under multiplication with an orthonormal matrix, $\|M - M'\|_F^2 = \left\|U^\top (M - M') V\right\|_F^2 = \sum_{ij} (u_i^\top (M - M') v_j)^2$. By the definition of eigenvalues, $u_i^\top M = \lambda_i u_i^\top$ and $M' v_j = \lambda'_j v_j$. Thus, $u_i^\top (M - M') v_j = (\lambda_i - \lambda'_j) u_i^\top v_j$ and this proves (16). Applying 16, we have

$$\begin{aligned} \|f_{a,b}(M) - f_{a,b}(M')\|_F^2 &= \sum_{i,j} \big(f_{a,b}(\lambda_i) - f_{a,b}(\lambda'_j)\big)^2 |u_i^\top v_j|^2 \\ &\le h^{-2} \sum_{i,j} (\lambda_i - \lambda'_j)^2 |u_i^\top v_j|^2 = h^{-2} \|M - M'\|_F^2. \end{aligned}$$

Consequently, for neighboring datasets A, A',

$$|\mathsf{s}_{a,b}(A, X) - \mathsf{s}_{a,b}(A', X)| \le \|f_{a,b}(A^\top A) - f_{a,b}(A'^\top A')\|_F \le \frac{\|A^\top A - A'^\top A'\|_F}{h} \le \frac{2}{h}.$$

For the accuracy claim, Weyl's inequality (Theorem 7) gives $\sigma_{k+1}^2 \le a$ and $b \le \sigma_k^2$. It also gives $h \ge \Delta_k - 4e_{\mathrm{w}} \ge \Delta_k/2$. Thus $f_{a,b}$ is one on the top k eigenvalues of $A^\top A$ and zero on all remaining eigenvalues. Finally, $\|(I - XX^\top)V_k V_k^\top\|_2 = \|(I - XX^\top)V_k\|_2$. □

C.2 Joint random search and its privacy cost

We now apply the method of Papernot and Steinke (2022) to our problem with the above score function. Conditioned on the shared initialization $\widehat{M}$ with $h > 0$, define one base trial Q as follows. Sample (j, ℓ) uniformly from $\{0, \dots, J\} \times \{0, \dots, H\}$, and run Algorithm 2 from $X^{(0)}$ for T_ℓ iterations, with radius R_j, noise parameter $\Delta(T_\ell)$, and floor $\gamma(T_\ell)$. Denote its output by X. With a fresh $Z \sim \mathcal{N}(0, 1)$, form the noisy score

$$\widetilde{s} = \mathsf{s}_{a,b}(A, X) + \frac{2\zeta}{h} Z. \tag{17}$$

The base output is the tuple $(R_j, T_\ell, X, \widetilde{s})$, ordered by smaller $\widetilde{s}$ being better, with a fixed total ordering to break ties.

Independently sample a number of trials N with

$$\Pr[N = r] = \theta(1-\theta)^{r-1}, \quad r \geq 1, \qquad \theta = \frac{\beta}{8m}, \qquad \mathbb{E}[N] = \frac{1}{\theta}. \tag{18}$$

Run the base trial Q independently N times with the same initialization. Release the triple $(\widehat{R}, \widehat{T}, \widehat{X})$ with the smallest noisy score.

Theorem 36. *Define*

$$\varrho_{\text{init}} = \frac{2}{\Delta_{\text{init}}^2}, \qquad \varrho_{\text{b}} = \frac{4}{\chi^2} + \frac{1}{2\zeta^2}, \qquad \varrho_{\text{tot}} = \varrho_{\text{init}} + \varrho_{\text{b}} \tag{19}$$

The selection procedure satisfies $(\lambda, \varepsilon_{\text{RDP}}(\lambda))$-Rényi differential privacy for every $\lambda > 1$, with

$$\varepsilon_{\text{RDP}}(\lambda) \leq \lambda\varrho_{\text{tot}} + 4\sqrt{\varrho_{\text{b}} \log(1/\theta)} + \frac{\log(1/\theta)}{\lambda - 1}. \tag{20}$$

In particular it is (ϵ, δ)-DP whenever

$$\varrho_{\text{tot}} + 2\sqrt{\varrho_{\text{tot}} \log(1/(\theta\delta))} + 4\sqrt{\varrho_{\text{b}} \log(1/\theta)} \leq \epsilon. \tag{21}$$

For $\epsilon \in (0, 1)$, an explicit setting of parameters is

$$v = \frac{\epsilon^2}{100 \log(1/(\theta\delta))}, \qquad \Delta_{\text{init}}^2 = \frac{6}{v}, \qquad \chi^2 = \frac{12}{v}, \qquad \zeta^2 = \frac{3}{2v}. \tag{22}$$

Thus $\Delta(t) = \frac{\sqrt{1200t \log(1/(\theta\delta))}}{\epsilon}$ for each candidate iteration count t.

Proof. A Gaussian mechanism with Euclidean sensitivity s and noise standard deviation σ has Rényi cost $\lambda s^2/(2\sigma^2)$ at order λ; these costs add under adaptive composition Mironov (2017). The initialization has $s = 2$ and $\sigma = \Delta_{\text{init}}$ so it has cost $\lambda\varrho_{\text{init}}$.

Fix any initialization output with $h > 0$. Conditional on previous noisy outputs, the preconditioner and clipped-update queries have sensitivities 2 and $2R_j$, with noise scales $\Delta(T_\ell)$ and $\Delta(T_\ell)R_j$, respectively. Their cost over T_ℓ steps is

$$\frac{4T_\ell\lambda}{\Delta(T_\ell)^2} = \frac{4\lambda}{\chi^2},$$

uniformly for any choice of j and ℓ. By the sensitivity of the score shown in Lemma 35, the noisy score additionally costs $\lambda/(2\zeta^2)$. Thus the entire base tuple has cost $\lambda\varrho_{\text{b}}$. Each trial starts from the same private initialization and uses independent iteration noises so we do not count the initialization cost toward the cost of each trial.

Apply Theorem 2 of Papernot and Steinke (2022) with negative-binomial shape parameter 1, corresponding to the geometric distribution (18). For any auxiliary order $\widehat{\lambda} > 1$, the cost of returning the best base tuple is at most

$$\lambda\varrho_{\text{b}} + 2(\widehat{\lambda} - 1)\varrho_{\text{b}} + \frac{2\log(1/\theta)}{\widehat{\lambda}} + \frac{\log(1/\theta)}{\lambda - 1}.$$

Taking $\widehat{\lambda} = 1 + \sqrt{\log(1/\theta)/\varrho_{\text{b}}}$ bounds the middle two terms by $4\sqrt{\varrho_{\text{b}} \log(1/\theta)}$. This is a uniform bound for every initialization output because in the $h \leq 0$ case, we only perform postprocessing with no privacy loss. The bound (20) then follows from adaptive composition with the initialization.

To obtain an (ϵ, δ)-DP bound, we convert the Rényi-DP bound, which adds a $\log(1/\delta)/(\lambda - 1)$ term to ϵ Mironov (2017). Minimizing over $\lambda > 1$ gives (21). Under (22), $\varrho_{\text{tot}} = v$ and $\varrho_{\text{b}} = 2v/3 \leq v$. Writing $D = \log(1/(\theta\delta)) > \log 8$, the left-hand side of (21) is bounded by $\epsilon^2/(100D) + \epsilon/5 + 2\epsilon/5 < \epsilon$. □

C.3 Utility guarantee from hyperparameter search

In this section we show that the algorithm recovers a good solution if the input data satisfies the assumptions for the utility proof without prior knowledge of $\kappa_k, \|U\|_\infty$. Let $T_\star = \left\lceil \frac{2}{\kappa_k} \log \frac{n}{\beta\delta\epsilon} \right\rceil$ and $R_\star = 15\,\|U\|_\infty \left(\sqrt{k} + \sqrt{\log\left(\frac{16n(T_{\max}+1)}{\beta_0}\right)} \right)$. As argued before, if the input satisfies the assumptions for the utility guarantee then $T_{\min} \le T_\star \le T_{\max}$ and $R_{\min} \le R_\star \le R_{\max}$. Consider the smallest grid duration $T^\dagger \ge T_\star$ and grid radius $R^\dagger \ge R_\star$. We have

$$T_\star \le T^\dagger \le 2T_\star, \qquad R_\star \le R^\dagger \le 2R_\star.$$

Using a similar proof to that of Corollary 34, we can show the following guarantee.

Theorem 37. *Let $\epsilon, \delta, \beta \in (0,1)$ and up to logarithmic factors*

$$\Delta_k \ge \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon} + \left(\frac{\sigma_k \sqrt{k}}{\epsilon} \right)^{2/3} + \left(\frac{R_\star \sigma_k^2 \sqrt{k}}{\epsilon} \right)^{2/3} \right)$$

With probability at least $1-\beta$, the selected output satisfies

$$\|(I - \widehat{X}\widehat{X}^\top) V_k\|_2^2 \le \left(\frac{\beta\delta\epsilon}{n} \right)^2 + \widetilde{O}\left(\frac{dk\sigma_k^2 \|U\|_\infty^2 + 1}{\epsilon^2 \Delta_k^2} \right). \tag{23}$$

Proof. Let $T = T^\dagger$, $R = R^\dagger$, and $\Delta = \Delta(T)$. Let us verify that the assumptions imply the conditions in Theorem 31, with failure parameter $\beta_0 = \beta/2$. The calibration in (22) and the choice $T_\star \le T \le 2T_\star$ give

$$\Delta_{\text{init}} = \widetilde{O}\left(\frac{1}{\epsilon}\right), \qquad \Delta = \widetilde{O}\left(\frac{1}{\epsilon\sqrt{\kappa_k}}\right) = \widetilde{O}\left(\frac{\sigma_k}{\epsilon\sqrt{\Delta_k}}\right), \qquad \zeta = \widetilde{O}\left(\frac{1}{\epsilon}\right).$$

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\left(\frac{\sigma_k\sqrt{k}}{\epsilon} \right)^{2/3} \right)$ implies $\Delta q_{\text{row}}(T) = \widetilde{O}\left(\frac{\sigma_k\sqrt{k}}{\epsilon\sqrt{\Delta_k}} \right) \le \Delta_k/8$ and also implies $\gamma(T) = \widetilde{O}\left(\Delta\sqrt{k}\right) \le \sigma_k^2/4$, when the hidden factors in the gap condition are sufficiently large.

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\left(\frac{R_\star \sigma_k^2 \sqrt{k}}{\epsilon} \right)^{2/3} \right)$ and $R \le 2R_\star$ imply

$$p_R(T) = \frac{3\Delta R q_k(T)}{2\sigma_k} = \widetilde{O}\left(\frac{R_\star\sqrt{k}}{\epsilon\sqrt{\Delta_k}} \right) \le \frac{\Delta_k}{16\sigma_k^2} = \frac{\kappa_k}{16}.$$

The gap condition $\Delta_k \ge \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon} \right)$ ensures that

$$\Delta_k \ge 8\Delta_{\text{init}} \left(2\sqrt{d} + \sqrt{2\log\frac{8}{\beta_0}} \right) \ge 8e_{\text{w}}.$$

The choice $R \ge R_\star$ and $T \le T_{\max}$ also ensure the lower bound on R in Theorem 31.

By Lemmas 21 and 5, the initialization event

$$\mathcal{G}_{\text{init}} \cap \left\{ \left\| \widehat{M} - A^\top A \right\|_2 \le e_{\text{w}} \right\}$$

holds with probability at least $1 - \beta_0/4 - \beta/8 = 1 - \beta/4$, where $\mathcal{G}_{\text{init}}$ is defined with iteration count T and failure parameter β_0. On this event, Lemma 9 gives $s_0 \le 1/4$, and Lemma 35 gives $h \ge \Delta_k/2$ and identifies every comparison score with the true subspace error. In the following, condition on any initialization in this event.

By the proofs of Lemmas 16 and 30, the events $\mathcal{G}_{\text{clip}}$ and $\mathcal{G}_{\text{mart}}$ for a trial with parameters (R,T) hold jointly with conditional probability at least $1 - 3\beta_0/8 - \beta_0/4 \ge 1/2$. Here the floor in (13)

gives a failure probability of at most $3\beta_0/8$ for $\mathcal{G}_{\mathrm{clip}}$, and we use the martingale event from Lemma 30 with failure parameter β_0; its threshold is no larger than $\nu_R(T)$.

If $\nu_R(T) \le 1/8$, all of the conditions of Theorem 31 are satisfied on these events, and the theorem gives

$$s_T \le \left(1 - \frac{\kappa_k}{2}\right)^T m_0 + 2\nu_R(T) \le \frac{\beta\delta\epsilon}{n\sqrt{15}} + 2\nu_R(T),$$

where we used $m_0 \le 1/\sqrt{15}$ and $T \ge T_\star$. Otherwise, we have $s_T \le 1 \le 8\nu_R(T)$, and we have the same asymptotic guarantee. Thus, in both cases, the trial produces a candidate with true error at most

$$e_\dagger := \frac{\beta\delta\epsilon}{n\sqrt{15}} + 8\nu_R(T)$$

with conditional probability at least $1/2$. Moreover, since $R \le 2R_\star$ and $R_\star = \widetilde{O}\left(\sqrt{k}\,\|U\|_\infty\right)$,

$$\nu_R^2(T) = \widetilde{O}\left(\frac{\Delta^2 R^2 d}{\sigma_k^2 \kappa_k}\right) = \widetilde{O}\left(\frac{dk\sigma_k^2\,\|U\|_\infty^2}{\epsilon^2\Delta_k^2}\right).$$

It remains to account for the hyperparameter search and the noisy comparison scores. Each independent base trial samples (R, T) with probability $1/m$, so it produces a candidate of error at most $e_\dagger$ with probability $p \ge 1/(2m)$. The probability that no such candidate occurs before the geometric stopping time is

$$\mathbb{E}(1-p)^N = \frac{\theta(1-p)}{p + \theta(1-p)} \le \frac{\theta}{p} \le \frac{\beta}{4}.$$

Let

$$t_{\mathrm{val}} = \frac{2\zeta}{h}\sqrt{2\log\frac{8}{\beta\theta}}.$$

Each noisy score differs from its true score by more than t_{val} with probability at most $\beta\theta/4$. Conditioning on N and taking a union bound over the trials, then using $\mathbb{E}N = 1/\theta$, shows that all score errors are at most t_{val} with probability at least $1 - \beta/4$. If a candidate of error at most $e_\dagger$ occurs and all score errors satisfy this bound, the selected candidate has true error at most $e_\dagger + 2t_{\mathrm{val}}$.

Since $h \ge \Delta_k/2$ and $\zeta = \widetilde{O}(1/\epsilon)$, we have $t_{\mathrm{val}}^2 = \widetilde{O}\left(1/(\epsilon^2\Delta_k^2)\right)$. Consequently,

$$\begin{aligned}\left\|(I - \widehat{X}\widehat{X}^\top)V_k\right\|_2^2 &\le (e_\dagger + 2t_{\mathrm{val}})^2 \\ &\le \frac{2}{15}\left(\frac{\beta\delta\epsilon}{n}\right)^2 + O\left(\nu_R^2(T) + t_{\mathrm{val}}^2\right) \\ &\le \left(\frac{\beta\delta\epsilon}{n}\right)^2 + \widetilde{O}\left(\frac{dk\sigma_k^2\,\|U\|_\infty^2 + 1}{\epsilon^2\Delta_k^2}\right).\end{aligned}$$

The initialization, search, and score-error events fail with total probability at most $\beta/4 + \beta/4 + \beta/4 \le \beta$, proving (23).

□