AN ITERATIVE ALGORITHM FOR DIFFERENTIALLY PRIVATE k-PCA WITH ADAPTIVE NOISE

Johanna Düngler* Amartya Sanyal*

ABSTRACT

Given n i.i.d. matrices $A_i \in \mathbb{R}^{d\times d}$ that share a common expectation Σ, the objective of Stochastic Principal Component Analysis (PCA) is to identify a subspace of dimension k that captures the maximum variance in Σ. Private PCA aims to find this subspace while ensuring the privacy of each individual instance A_i. However, even when estimating only the top eigenvector, most existing techniques either (i) require the number of samples n to scale super-linearly with d, even for Gaussian data, or (ii) suffer from excessive (privacy) noise when the randomness in each A_i is small. Liu et al. [2022] overcame both limitations for sub-Gaussian data when estimating the top eigenvector with their algorithm DP-PCA. We propose an extension of their algorithm that estimates the top k eigenvectors for arbitrary $k \leq d$, while still overcoming challenges (i) and (ii). Furthermore, for $k = 1$, we recover the utility of DP-PCA, which for sub-Gaussian data achieves nearly optimal statistical error rates even for $n = \tilde{O}(d)$.

1 Introduction

Principal Component Analysis (PCA) is an important statistical technique widely used for dimensionality reduction, data visualization, and noise filtering. Given n data points $\{x_i\}_{i=1}^n$, standard PCA inputs a single matrix $\sum_i x_i x_i^\top$ and computes its dominant eigenvectors. The number of eigenvectors PCA computes (i.e. the dimension of the subspace) is generally considered to be an input parameter to PCA, however, we often refer to k-PCA, to emphasize we are interested in the dominant k eigenvectors. In this work, we consider the problem of Stochastic (Streaming) k-PCA, which differs from the standard setting in two important aspects. First, instead of inputting a single matrix, we input a stream of matrices $A_1, \ldots, A_n$, processing each matrix sequentially and then discarding it and second, these matrices are sampled independently from distributions that share the same expectation Σ. Given this input, the goal of a Stochastic (Streaming) k-PCA algorithm is to approximate the dominant k eigenvectors of Σ.[2]

In the non-private setting, Oja's algorithm, discussed in Algorithm 1, is known to be nearly optimal for this problem. Jain et al. [2016] showed that Algorithm 1 recovers the top eigenvector with approximation error $\tilde{\Theta}(\sqrt{d/n})$ where d refers to the dimension of the input and n to the number of samples. Oja's Algorithm can be extended to $k > 1$ by simply initializing a matrix $Q_0 \in \mathbb{R}^{d\times k}$ instead of a vector ω_0, and performing the Gram-Schmidt process to obtain an orthogonal matrix after every ascent step instead of vector normalizing. Huang et al. [2021] extended the utility guarantee to this setting and achieved nearly optimal error of

$$\|\hat{V}_k \hat{V}_k^\top - V_k V_k^\top\| = \tilde{\Theta}(\sqrt{dk/n})$$

under certain boundedness assumptions (Assumptions A.1 to A.3) where each column of $\hat{V}_k \in \mathbb{R}^{d\times k}$ represents an approximate principal component and $V_k \in \mathbb{R}^{d\times k}$ is the matrix containing the true top k eigenvectors of Σ.

In the private setting, there are several existing DP-PCA algorithms. Though mostly designed for the non-stochastic setting, most of these works [Blum et al., 2005, Chaudhuri et al., 2013, Dwork et al., 2014, Hardt and Roth, 2013] achieve suboptimal error rates of $O(\sqrt{dk/n} + d^{3/2}k/(\varepsilon n))$ when extended to the stochastic setting. Another issue with these algorithms is that the added noise does not scale with the inherent randomness in each data point. To understand this, consider the following example.

*Department of Computer Science, University of Copenhagen

[2]For purpose of brevity, we will refer to it as stochastic k-PCA and ignore the adjective "streaming".

Algorithm 1 OjasAlgorithm($\{A_i\}_{i=1}^n$)

Choose ω_0 uniformly at random from the unit sphere
for $t = 1, \ldots, n$ **do**
  $w_i' \leftarrow w_{i-1} + \eta_i A_i w_{i-1}$
  $w_i \leftarrow w_i'/\|w_i'\|_2$
end for
return w_n

Example 1.1. *In the spiked covariance model, we sample i.i.d. matrices $A_i \in \mathbb{R}^{d\times d}$ that contain both a deterministic (low-rank) signal and random noise that leads the observed matrices A_i to be full-rank. The signal takes the form*

$$U\Lambda U^\top \text{ where } \Lambda = diag(\lambda_1, \ldots, \lambda_k)$$

and the noise comes from Gaussian vectors $\eta_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$. In the rank 1 setting one way to sample from the spiked covariance model would be to sample i.i.d. vectors $x_i = s_i + \eta_i$, with $s_i = v$ (a unit vector) with probability 1/2 and $-v$ otherwise, and $n_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$. We then define the matrix $A_i = x_i x_i^\top$ which captures both the deterministic signal component $vv^\top$ and noise terms that scale with σ. As the (data) noise level decreases (i.e. $\sigma \to 0$) the data becomes nearly deterministic and we have $\Sigma = \mathbb{E}[A_i] \approx A_i$ for all i. In such low-noise regimes, it would be natural to expect that less privacy noise is needed to preserve differential privacy. However, most algorithms for private PCA add (privacy) noise either based on static clipping or by assuming that any input has a maximum norm of at most one. Consequently, we need to add (privacy) noise that scales with this clipping threshold or with $\max_i \|A_i\|$, even if the data passed to the algorithm has no randomness at all (i.e. $\mathbb{E}[A_i] = A_i$).

The DP-PCA algorithm by Liu et al. [2022] simultaneously achieves both statistically optimal error rates for sub-Gaussian distributions, including the spiked covariance case, while only requiring $\tilde{O}(d)$ samples. However, their algorithm is limited to estimating the top eigenvector. Cai et al. [2024] developed an algorithm that is statistically optimal for the spiked covariance model, even when computing k principal components. However, it only fulfills differential privacy when the data is sampled from the spiked covariance model.

Our Contribution *Our main contribution is that we introduce a novel algorithm, k-DP-PCA, that is simple to implement, ensures privacy for any input data, requires only a linear number of samples in d, and adds (privacy) noise that scales with the randomness in the input data.*

From a technical perspective, our algorithm is based on the deflation method, arguably the most natural reduction-based approach to k-PCA. The deflation method repeatedly employs a subroutine to estimate the top eigenvector and then projects out the previously computed eigenvectors from the matrix. By repeating this approach k times, such methods can compute k eigenvectors. Our approach is inspired by Jambulapati et al. [2024], who proved significantly sharper utility bounds for deflation methods in k-PCA. However, their results can only be directly applied for non-stochastic PCA, meaning that we would need direct access to Σ.

Our technical contributions are three-fold: We first extend the result of Jambulapati et al. [2024] to the stochastic setting and in the process introduce the stochastic ePCA oracle in Definition 3. In Appendix B, we state and prove the full result for the stochastic deflation problem which we believe may be of independent interest. Second, we propose an adapted version of the DP-PCA [Liu et al., 2022] in Algorithm 3. And finally, through a novel analysis of non-private Oja's algorithm, we demonstrate that our adapted approach meets all the essential requirements for stochastic deflation. The full results for this are presented in Appendix C. Combining all of these, we present our main algorithm in Algorithm 2 and the main utility bound in Theorem 2.1 which recovers the utility guarantees of Liu et al. [2022] in the $k = 1$ case.

2 Problem formulation and Main Theorem

In this work, our algorithm inputs n matrices $A_1, \ldots, A_n$ in $\mathbb{R}^{d\times d}$ and outputs a matrix $U \in \mathbb{R}^{d\times k}$. The columns of U are pairwise orthogonal and of unit norm. Throughout this work, $\|\cdot\|_2$ refers to the operator norm and $\langle\cdot,\cdot\rangle$ refers to the Frobenius inner product. Specifically, for any matrices A, B, $\langle A, B\rangle = \mathrm{Tr}(A^\top B)$. Before discussing the main result of our work, we first formalise the problem setting including the assumptions on data in Assumption A and the utility metric in Definition 2.

Assumption A ($(\Sigma, \{\lambda_i\}_{i=1}^d, M, V, K, \kappa, a, \gamma^2)$-model [Liu et al., 2022])**.** Let $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ be a sequence of (not necessarily symmetric) matrices sampled independently from distributions that fulfill the following assumptions with a PSD matrix $\Sigma \in \mathbb{R}^{d\times d}$, matrices $H_u \in \mathbb{R}^{d\times d}$, and positive scalars $M, V, K, \kappa, a, \gamma^2$:

A.1 $\mathbb{E}[A_i] = \Sigma$, with Σ a PSD matrix having eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$, corresponding eigenvectors $v_1, \ldots, v_d$, $0 < \Delta = \min_{i \in [k]}(\lambda_i - \lambda_{i+1})$ and $\kappa' := \frac{\lambda_1}{\Delta}$.

A.2 $\|A_i - \Sigma\|_2 \leq \lambda_1 M$ almost surely.

A.3 $\max\{\|\mathbb{E}[(A_i - \Sigma)(A_i - \Sigma)^\top]\|_2, \|\mathbb{E}[(A_i - \Sigma)^\top (A_i - \Sigma)]\|_2\} \leq \lambda_1^2 V$.

A.4 $\max_{\|u\|=1, \|v\|=1} \mathbb{E}\left[\exp\left(\left(\frac{|u^\top P(A_i^\top - \Sigma)Pv|^2}{K^2 \lambda_1^2 \gamma^2}\right)^{1/(2a)}\right)\right] \leq 1$, where

$$H_u := \frac{1}{\lambda_1^2}\mathbb{E}[(A_i - \Sigma)uu^\top (A_i - \Sigma)^\top], \quad \text{and} \quad \gamma^2 := \max_{\|u\|=1} \|H_u\|_2.$$

The first three assumptions Assumptions A.1 to A.3 are quite mild (and standard), as they are needed for concentration of measure (under the matrix Bernstein inequality [Tropp, 2012]) and therefore also required for Oja's algorithm even when privacy is not required. Assumption A.4 guarantees that for any unit vectors u, v, and projection P with probability $1 - \vartheta$

$$|u^\top P(A_i - \Sigma)Pv|^2 \leq K^2 \lambda_1^2 \gamma^2 \log^{2a}(1/\vartheta)$$

for some sufficiently large constant K. This bound controlling the size of the bilinear form, can be seen as a Gaussian-like tail bound, which tells us that the magnitude of the projection of the A_i along any direction is bounded with high probability. Further, we use the add/remove model of differential privacy, namely

Definition 1. ([Dwork et al., 2006]) Given two multisets S and S', we say the pair (S, S') is neighboring if $|S \setminus S'| + |S' \setminus S| \leq 1$. We say a stochastic query q over a dataset S satisfies (ε, δ)-differential privacy for some $\varepsilon > 0$ and $\delta \in (0, 1)$ if

$$P(q(S) \in A) \leq e^\varepsilon P(q(S') \in A) + \delta$$

for all neighboring (S, S') and all subsets A of the range of q.

Our algorithm outputs a matrix $U \in \mathbb{R}^{d \times k}$, where each column represents an approximate principal component, and all columns are mutually orthogonal. We measure the utility of U by comparing it to V_k, the matrix containing the true top k eigenvectors of Σ as columns.

Definition 2. We say $U \in \mathbb{R}^{d \times k}$ is ζ-useful if U has orthonormal columns and

$$\langle UU^\top, \Sigma \rangle \geq (1 - \zeta^2)\langle V_k V_k^\top, \Sigma \rangle.$$

Although several utility measures exist for PCA, our choice is motivated by the error measure used in Oja's Algorithm. This is a natural measure of usefulness, as $\langle UU^\top, \Sigma \rangle$ quantifies how much of the original energy is retained when projecting Σ onto the lower-dimensional subspace spanned by U and the Eckart-Young theorem tells us V_k is the best rank-k approximation of Σ.

Theorem 2.1 (Main Theorem)**.** *For $\varepsilon \in (0, 0.9)$ and $0 < k < d$, k-DP-PCA satisfies (ε, δ)-DP for all inputs $\{A_i\}, B, \zeta$ and δ. Given n i.i.d. samples $\{A_i \in \mathbb{R}^{d \times d}\}_{i=1}^n$ satisfying Assumptions A.1 to A.4 with parameters $(\Sigma, M, V, K, \kappa', a, \gamma^2)$, if*

$$n = C \max\left\{ e^{\kappa'^2} + \frac{d\kappa'\gamma(\log(1/\delta))^{1/2}}{\varepsilon} + \kappa' M + \kappa'^2 V + \frac{d^{1/2}(\log(1/\delta))^{3/2}}{\varepsilon}, \lambda_1^2 \kappa'^2 k^3 V, \frac{\kappa'^2 \gamma k^2 d\sqrt{\log(1/\delta)}}{\varepsilon} \right\} \tag{1}$$

with a large enough constant (and ignoring log terms) and $\delta \leq 1/n$, then Algorithm 2 outputs $U \in \mathbb{R}^{d \times k}$, which with probability at least 0.99 is ζ-useful, with

$$\zeta = \tilde{O}\left(\frac{\lambda_1}{\Delta}\left(\sqrt{\frac{Vk}{n}} + \frac{\gamma dk\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$$

where $\tilde{O}(\cdot)$ hides poly-logarithmic factors in $n, d, 1/\varepsilon$, and $\log(1/\delta)$ and polynomial factors in K.

We note that for $k = 1$, this recovers the same utility guarantee as DP-PCA [Liu et al., 2022] and continues to be optimal for spiked covariance as shown below in Corollary 2.1.1. Additionally, we highlight that the dependence of the estimation error on the dimension d is optimal, as demonstrated by the lower bound established in prior work by Liu et al. However, the linear dependency on k might be an artifact of our analysis, specifically, if it were possible to

Algorithm 2 k-DP-PCA

Input: $\{A_1, \ldots, A_n\}$, $k \in [d]$, privacy parameters (ε, δ), $B \in \mathbb{Z}_+$, learning rates $\{\eta_t\}_{t=1}^{\lfloor n/B \rfloor}$, and $\tau \in (0, 1)$

1: $m \leftarrow n/k$, $P_0 \leftarrow \mathbb{I}_d$
2: **for** $i \in [k]$ **do**
3: $\quad u_i \leftarrow$ MODIFIEDDP-PCA$(\{A_{m*(i-1)+j}\}_{j=1}^{m}, P_{i-1}, (\varepsilon, \delta), B, \{\eta_t\}, \tau)$
4: $\quad P_i \leftarrow P_{i-1} - u_i u_i^\top$
5: **end for**
6: **return** $U \leftarrow \{u_i\}_{i \in [k]}$

demonstrate that the same samples A_i could be effectively reused across iterations, we could potentially reduce this dependency on k. Furthermore, the first term of the utility guarantee $O(\sqrt{Vk/n})$ represents the statistical error of PCA without privacy constraint and the second term is the cost of privacy. Lastly, we require a lower bound on the number of samples n in Equation (1), due to two key factors: first, each batch must be sufficiently large to guarantee the accuracy of the range estimation in Algorithm 3 with high probability. And second, we need to carefully control the cumulative error introduced during each deflation step, as errors from earlier steps directly propagate and amplify subsequent errors.

Corollary 2.1.1 (Upper bound, Spiked Covariance)**.** *Under the assumptions of Theorem 2.1 and* $\{A_i = x_i x_i^\top\}_{i=1}^n$ *with random vectors sampled as described in Example* **??***, k-DP-PCA outputs* $U \in \mathbb{R}^{d \times k}$ *that is* ζ*-useful, with*

$$\zeta = \tilde{O}\left(\frac{\sigma(\lambda_1 + \sigma^2)}{\Delta}\left(\sqrt{\frac{dk}{n}} + \frac{dk\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$$

where $\tilde{O}(\cdot)$ *hides poly-logarithmic factors in* $n, d, 1/\varepsilon$*, and* $\log(1/\delta)$ *and polynomial factors in* K*.*

3 Algorithm and Proof Sketch

We present our main algorithm in Algorithm 2. The algorithm proceeds by applying the deflation method (Line 4) to a 1-PCA algorithm (Line 3), defined in Algorithm 3. The privacy proof of Algorithm 2 follows straightforwardly from the composition of private algorithms (Algorithms 5 and 6) used in Lines 3 and 4 of Algorithm 3. However, the utility proof is more involved as the arguments for deflation in Jambulapati et al. [2024] do not directly extend to the stochastic setting. So, we first extend the argument in Jambulapati et al. [2024] to the stochastic deflation setting by defining a stochastic ePCA oracle in Definition 3. Full proofs are relegated to Appendix B.

Definition 3 (stochastic ePCA oracle)**.** We say an algorithm O_{ePCA} is a ζ-approximate 1-ePCA oracle if, on independent inputs $A_1, \ldots, A_n \in \mathbb{R}^{d \times d}$ with $\mathbb{E}[A_i] = \Sigma$ for all i, an orthogonal projection matrix $P \in \mathbb{R}^{d \times d}$, a, the algorithm O_{ePCA} returns $u \in \mathbb{R}^d$ satisfying $u \in \text{Im}(P)$ and, with high probability

$$\langle uu^\top, P\Sigma P \rangle \geq (1 - \zeta^2)\langle vv^\top, P\Sigma P \rangle$$

where v is the top eigenvector of $P\Sigma P$.

However, we cannot use DP-PCA [Liu et al., 2022] as an input to the deflation method as it only satisfies

$$\langle uu^\top, \mathbb{E}[P]\Sigma\mathbb{E}[P] \rangle \geq (1 - \zeta)\langle vv^\top, \mathbb{E}[P]\Sigma\mathbb{E}[P] \rangle$$

which is insufficient, as $\mathbb{E}[P]$ is likely not even a projection matrix. For this reason, we propose MODIFIEDDP-PCA (Algorithm 3), which we prove is a stochastic ePCA oracle in two steps. First we show that non-private Oja's algorithm is an ePCA oracle, via a novel analysis of the algorithm we relegated to Appendix C. And secondly, we show that with high probability we can reduce the update step (Line 5 in Algorithm 3) to an update step of non-private Oja's algorithm with matrices PC_tP, where $C_t := \frac{1}{B}\sum_{i \in [B]} A_i + \beta_t G_t$ and G_t is a scaled Gaussian matrix. Lastly we show in Theorem B.2 that the deflation method also yields the required utility guarantee with a stochastic e-PCA oracle. For the full proof, see Appendix D.

References

Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005. 1

Algorithm 3 ModifiedDP-PCA

Input: $\{A_1, \ldots, A_m\}$, a projection P, privacy parameters (ε, δ), learning rates $\{\eta_t\}_{t=1}^{\lfloor n/B \rfloor}$, $B \in \mathbb{Z}_+$ and $\tau \in (0, 1)$

Choose ω_0' uniformly at random from the unit sphere, $\omega_0 \leftarrow P\omega_0'/\|P\omega_0'\|$
for $t = 1, 2, \ldots, T = \lfloor m/B \rfloor$ **do**
    $\hat{\Lambda} \leftarrow \text{PrivTopEigenval}(\{PA_{B(t-1)+i}P\omega_{t-1}\}_{i=1}^{\lfloor B/2 \rfloor}, (\varepsilon/2, \delta/2), \tau/(2T))$ (Algorithm 5)
    $\hat{g}_t \leftarrow \text{PrivMean}(\{PA_{B(t-1)+i}P\omega_{t-1}\}_{i=1}^{\lfloor B/2 \rfloor}, \hat{\Lambda}, (\varepsilon/2, \delta/2), \tau/(2T))$ (Algorithm 6)
    $\omega_t' \leftarrow \omega_{t-1} + \eta_t P\hat{g}_t$
    $\omega_t \leftarrow P\omega_t'/\|P\omega_t'\|$
end for
return ω_T

T Tony Cai, Dong Xia, and Mengyue Zha. Optimal differentially private pca and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*, 2024. 1

Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *The Journal of Machine Learning Research*, 14(1):2905–2943, 2013. 1

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006. 1

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20, 2014. 1

Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 331–340, 2013. 1

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012. A.5

De Huang, Jonathan Niles-Weed, and Rachel Ward. Streaming k-pca: Efficient guarantees for oja's algorithm, beyond rank-one updates. In *Conference on Learning Theory*, pages 2463–2498. PMLR, 2021. 1

Prateek Jain, Chi Jin, Sham M Kakade, Praneeth Netrapalli, and Aaron Sidford. Streaming pca: Matching matrix bernstein and near-optimal finite sample guarantees for oja's algorithm. In *Conference on learning theory*, pages 1147–1164. PMLR, 2016. 1, C, C, C.3

Arun Jambulapati, Syamantak Kumar, Jerry Li, Shourya Pandey, Ankit Pensia, and Kevin Tian. Black-box k-to-1-pca reductions: Theory and applications. *arXiv preprint arXiv:2403.03905*, 2024. 1, 3, B, 4, 5, B

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1376–1385, Lille, France, 07–09 Jul 2015. PMLR. A.14

Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. Dp-pca: Statistically optimal and differentially private pca. *Advances in neural information processing systems*, 35:29929–29943, 2022. (document), 1, 1, A, 2, 3, A.2, A.3, B, D, D, D.4, 5, 6

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12: 389–434, 2012. 2

Appendix

A Mathematics Preliminaries

Lemma A.1. $C \preceq D \implies ACA^\top \preceq ADA^\top$

Proof. $C \preceq D \implies C - D \preceq 0$. So for any $x \in \mathbb{R}^d$, set $y = A^\top x$ we have

$$x^\top A(C - D)Ax = y^\top (C - D)y \leq 0$$

so for any x

$$x^\top ACA^\top x \leq x^\top ADA^\top x$$

which proofs our claim. □

Lemma A.2. *(Lemma F.2 in [Liu et al., 2022] Let $G \in \mathbb{R}^{d\times d}$ be a random matrix where each entry G_{ij} is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. Then there exists a universal constant $C > 0$ such that with probability $1 - 2e^{-t^2}$*

$$\|G\|_2 \leq C(\sqrt{d} + t)$$

for $t > 0$.

Lemma A.3. *(Lemma F.5 in [Liu et al., 2022]) Under Assumption 1.-3. with probability $1 - \tau$*

$$\|\frac{1}{B}\sum_{i\in[B]} A_i - \Sigma\|_2 = \mathcal{O}\left(\sqrt{\frac{\lambda_1^2 V \log(d/\tau}{B}} + \frac{\lambda_1 M \log(d/\tau}{B}\right)$$

Lemma A.4. *Let $G \in \mathbb{R}^{d\times d}$ be a random matrix where each entry G_{ij} is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. Then we have*

$$\mathbb{E}[\|GG^\top\|_2] \leq C_2 d \tag{2}$$

Proof. By Lemma A.2 the exists a universal constant $C_3 > 0$ such that

$$\mathbb{P}(\|G\| \geq C_1(\sqrt{d} + s)) \leq e^{-s^2}, \forall s > 0$$

then

$$\begin{aligned}\mathbb{E}[\|GG^\top\|_2] &\leq \mathbb{E}[\|GG\|_2^2] \\ &= \int_0^\infty 2r\mathbb{P}(\|G\|_2 > r)dr \leq C_1 d + C_2 \int_{\sqrt{d}}^\infty 2re^{-\frac{(r-\sqrt{d})^2}{2}} d \\ &= C_1(d + \sqrt{2\pi d} + 2) \leq C_2 d\end{aligned}$$

□

Lemma A.5 (Weyl's inequality [Horn and Johnson, 2012])**.** *Let G_1 and G_2 be two matrices with eigenvalues $\mu_1 \geq \cdots \geq \mu_d$ and $\nu_1 \geq \cdots \geq \nu_d$ respectively, then*

$$|\nu_i - \mu_i| \leq \|G_1 - G_2\|_2$$

Lemma A.6. *(Conditional Markov Inequality) Let $\mathcal{F}$ be a conditioning event (or a sigma-algebra), let X be a non negative random variable, and $a > 0$, then*

$$P(X \geq a|\mathcal{F}) \leq \frac{\mathbb{E}[X|\mathcal{F}]}{a}$$

Proof. As a first step we define

$$I_{\{X\geq a\}} = \begin{cases} 1, \text{ if } X \geq a \\ 0, \text{ otherwise} \end{cases}$$

then by definition of the indicator function we have

$$XI_{\{X\geq a\}} \geq aI_{\{X\geq a\}}$$

which implies

$$\mathbb{E}[XI_{\{X\geq a\}}|\mathcal{F}] \geq \mathbb{E}[aI_{\{X\geq a\}}|\mathcal{F}]$$

by taking conditional expectation on both sides. And finally

$$\mathbb{E}[I_{\{X\geq a\}}|\mathcal{F}] = P(X \geq a|\mathcal{F})$$

gives us the wished result □

Lemma A.7. *(Conditional Chebyshev's Inequality) Let $\mathcal{F}$ be a conditioning event (or a sigma-algebra), then for $a > 0$*

$$P(|X - \mathbb{E}[X|\mathcal{F}]| \geq a|\mathcal{F}) \leq \frac{Var[X|\mathcal{F}]}{a^2}$$

where $Var[X|\mathcal{F}] = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}]$.

Proof.

$$P(|X - \mathbb{E}[X|\mathcal{F}]| \geq a|\mathcal{F}) = P((X - \mathbb{E}[X|\mathcal{F}])^2 \geq a^2|\mathcal{F})$$

$(X - \mathbb{E}[X|\mathcal{F}])^2$ is a non non negative random variable, so we can use conditional Markov, which gives us

$$P((X - \mathbb{E}[X|\mathcal{F}])^2 \geq a^2|\mathcal{F}) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}]}{a^2}$$

□

Lemma A.8. *(Distributional Equivalence) Let $z \sim \mathcal{N}(0, \Sigma)$, P a projection matrix, and $\omega \in Im(P)$ a unit vector, then there exists a random matrix G so that*

$$Pz \stackrel{d}{=} PGPw$$

and

$$G = \Sigma^{1/2}Y$$

with Y is a random matrix where each entry is i.i.d. sampled from $\mathcal{N}(0, 1)$.

Proof. First note that $\text{Cov}(Pz) = P\text{Cov}(z)P^\top$, and as $PP^\top = P^2 = P$ we have

$$Pz \sim \mathcal{N}(0, P\Sigma P)$$

Likewise we have

$$\text{Cov}(PGPw) = P\text{Cov}(GPw)P = P\text{Cov}(Gw)P$$

where the last equality follows as $w \in \text{Im}(P)$. So we want

$$\text{Cov}(Gw) = \Sigma$$

If we define $G = \Sigma^{1/2}G'$, we see that if we can find G' so that

$$Gw \stackrel{d}{\sim} \mathcal{N}(0, \mathbf{I}_d)$$

we are done. Using rotation invariance of the spherical Gaussian random vectors and the fact that $\|w\|_2 = 1$, we get that defining $G' \in \mathbb{R}^{d\times d}$ with each entry i.i.d. sampled from $\mathcal{N}(0, 1)$, we get $G'w \sim N(0, \mathbf{I}_d)$, which finishes our proof. □

Lemma A.9. *Assume we have a matrix $A \in \mathbb{R}^{d\times d}$ and a projection matrix P then*

$$\|PAP\|_2 \leq \|A\|_2$$

Proof. $\|PAP\|_2 \leq \|P\|_2\|A\|_2\|P\|_2 \leq \|A\|_2$, where the last inequality follows as projection matrices have eigenvalues in $\{0, 1\}$. □

Lemma A.10. *Let $A \in \mathbb{R}^{d\times d}$ be a random matrix and P a random projection matrix independent of A then*

$$\|\mathbb{E}[PAPA^\top P]\|_2 \leq \|\mathbb{E}[AA^\top]\|_2$$

Proof. Let $x \in \mathbb{R}^d$ be a unit vector, then

$$\|PAPx\|_2 \leq \|APx\|_2$$

as P is a projection matrix. Squaring both sides we get

$$x^\top PA^\top PAPx \leq x^\top PA^\top APx$$

as x was an arbitrary unit vector, this implies:

$$PA^\top PAP \preceq PA^\top AP$$

If we now take expectations on both sides we get

$$\mathbb{E}[PA^\top PAP] \preceq \mathbb{E}[PA^\top AP] \preceq \mathbb{E}[PA^\top AP] = \mathbb{E}_P[P\mathbb{E}[A^\top A|P]P] = \mathbb{E}[P\mathbb{E}[A^\top A]P] = \mathbb{E}[P\mathbb{E}[A^\top A]P]$$

where we can drop the conditioning as A is independent of P. So when taking the 2-norm on either side we get

$$\|\mathbb{E}[PA^\top PAP]\| \leq \|\mathbb{E}[P\mathbb{E}[A^\top A]P]\|_2 \leq \mathbb{E}[\|P\|_2\|\mathbb{E}[A^\top A]\|_2\|P\|_2] \leq \|\mathbb{E}[A^\top A]\|_2$$

where the last inequality follows as $\|P\|_2 \leq 1$. □

Lemma A.11. *For A and B independent random matrices*

$$\mathbb{E}[ABA^\top] \preceq \|\mathbb{E}[B]\|_2 \mathbb{E}[AA^\top]$$

Proof. By independence we have $\mathbb{E}[ABA^\top] = \mathbb{E}[A\mathbb{E}[B]A^\top]$. Then by using $\mathbb{E}[B] \preceq \|\mathbb{E}[B]\|_2 \mathbf{I}_d$ and Lemma A.1 we obtain the wished inequality. □

Lemma A.12. *We define*

$$H_u^P := \frac{1}{\lambda_1^2(P\Sigma P)} \mathbb{E}[P(A_i - \Sigma)Puu^\top P(A_i - \Sigma)P]$$

and

$$\gamma_P^2 = \max_{\|u\|=1} \|H_u^P\|_2$$

then

$$\lambda_1^2(P\Sigma P)\gamma_P^2 \leq \lambda_1^2 \gamma^2$$

where γ and λ_1 are defined as in Assumption A

Proof.

$$\begin{aligned}
\|\mathbb{E}\left[P(A_i - \Sigma)Puu^\top P(A_i - \Sigma)P\right]\| &= \|\mathbb{E}_P\left[P\mathbb{E}[(A_i - \Sigma)Puu^\top P(A_i - \Sigma)|P]P\right]\| \\
&\leq \mathbb{E}_P\left[\|P\|\|\mathbb{E}[(A_i - \Sigma)Puu^\top P(A_i - \Sigma)|P]\|\|P\|\right] \\
&\leq \mathbb{E}_P\left[\|\mathbb{E}[(A_i - \Sigma)Puu^\top P(A_i - \Sigma)|P]\|\right]
\end{aligned}$$

and further

$$\max_{\|u\|=1} \|\mathbb{E}[(A_i - \Sigma)Puu^\top P(A_i - \Sigma)|P]\| \leq \max_{\|u\|=1} \|\mathbb{E}[(A_i - \Sigma)uu^\top (A_i - \Sigma)|P]\| = \lambda_1^2\gamma^2$$

as $Puu^\top P \preceq uu^\top$. So, all together this proves the Lemma. □

Differential Privacy

Lemma A.13. *(Parallel composition). Consider a sequence of interactive queries $\{q_k\}_{k=1}^K$ each operating on a subset S_k of the database and each satisfying (ε, δ)-DP. If S_k's are disjoint then the composition $(q_1(S_1), q_2(S_2), ..., q_K(S_K))$ is (ε, δ)-DP.*

Lemma A.14. *(Advanced Composition [Kairouz et al., 2015]) For $\varepsilon \leq 0.9$, an end-to-end guarantee of (ε, δ)-differential privacy is satisfied if a database is accessed k times, each with a $(\varepsilon/(2\sqrt{2k\log(2/\delta)}), \delta/(2k))$-differential private mechanism.*

B Stochastic Black Box PCA

In this section we extend the work of Jambulapati et al. [2024] to the stochastic setting and obtain the same utility results as them even when approximating the top eigenvector of the expectation of a stream of matrices.

Definition 4. (stochastic ePCA oracle) We say an algorithm O_{ePCA} is a ζ-approximate 1-ePCA oracle (or, ζ-1-ePCA oracle) if, on independent inputs $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ with $\mathbb{E}[A_i] = \Sigma$ for all i, and $P \in \mathbb{R}^{d\times d}$, an orthogonal projection matrix, the algorithm O_{ePCA} returns $u \in \mathbb{R}^d$ satisfying $u \in \text{Im}(P)$ and, with high probability

$$\langle uu^\top, P\Sigma P\rangle \geq (1 - \zeta^2)\langle vv^\top, P\Sigma P\rangle$$

where v is the top eigenvector of $P\Sigma P$.

Remark. DP-PCA [Liu et al., 2022] is not a stochastic 1-ePCA oracle as it will only fulfill

$$\langle uu^\top, \mathbb{E}[P]\Sigma\mathbb{E}[P]\rangle \geq (1 - \zeta^2)\langle vv^\top, \mathbb{E}[P]\Sigma\mathbb{E}[P]\rangle$$

and it's not clear how close $\mathbb{E}[P]\Sigma\mathbb{E}[P]$ is to $P\Sigma P$.

Algorithm 4 BlackBoxPCA($\{A_i\}$,k,O_{1PCA}) [Jambulapati et al., 2024]

Input: $\{A_1, \ldots, A_n\}$ i.i.d matrices sampled from a distribution with expectation $\mathbb{E}[A_i] = \Sigma \in \mathbb{S}^{d\times d}_{\succeq 0}$, $k \in [d]$, O_{1PCA} an algorithm which takes as input matrices $A_1, \ldots A_n$ and returns a unit vector in $\mathbb{R}^d$

$P_0 \leftarrow \mathbb{I}_d$
$B \leftarrow \lfloor n/k \rfloor$
for $i \in [k]$ **do**
  $u_i \leftarrow O_{1PCA}(A_{B*(i-1)+1}, \ldots, A_{B*i}, P_{i-1})$
  $P_i \leftarrow P_{i-1} - u_i u_i^\top$
end for
return $U \leftarrow \{u_i\}_{i\in[k]}$

We will now show that for this type of approximation algorithm we can obtain a utility guarantee and that it would be optimal for the spiked covariance setting. Jambulapati et al. define two types of approximation notions for PCA. Our type of utility bound is equivalent to their first notion:

Definition 5. (energy k-PCA. [Jambulapati et al., 2024]) $U \in \mathbb{R}^{d\times k}$ is an ζ-approximation energy k-PCA of $M \in \mathbb{S}^{d\times d}_{\succeq 0}$ if

$$\langle UU^\top \rangle \geq (1-\zeta^2)\|M\|_k$$

where

$$\|M\|_k := \max_{\text{orthonormal} V\in\mathbb{R}^{d\times k}} \langle VV^\top, M\rangle$$

and

$$\langle A, B\rangle = \mathrm{Tr}(A^\top B)$$

is the frobenius inner product.

Lemma B.1. *For $v, w \in \mathbb{R}^d$ unit vectors,θ the angle between the two and Σ a psd matrix with v it's top eigenvector we have*

$$\langle ww^\top, \Sigma\rangle \geq (1-\sin^2(\theta))\langle vv^\top, \Sigma\rangle$$

Remark. We note that $\sin^2(\theta)$ is a tight lower bound for ε, as $w = v$ it is achieved.

Proof.

$$\begin{aligned}\langle ww^\top, \Sigma\rangle &= \langle vv^\top, \Sigma\rangle - \langle vv^\top - ww^\top, \Sigma\rangle \\ &= (1 - \frac{\langle vv^\top - ww^\top, \Sigma\rangle}{\langle vv^\top, \Sigma\rangle})\langle vv^\top, \Sigma\rangle\end{aligned}$$

Now as v is the top eigenvector of Σ we know

$$\langle vv^\top, \Sigma\rangle = \mathrm{Tr}(vv^\top\Sigma) = v^\top\Sigma v = \lambda_1$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ denote the eigenvalues of Σ and $v, v_2, \ldots, v_d$ the corresponding eigenvectors. Therefore

$$\begin{aligned}\frac{\langle vv^\top - ww^\top, \Sigma\rangle}{\langle vv^\top, \Sigma\rangle} &= 1 - \frac{1}{\lambda_1} w^\top\Sigma w = 1 - (w^\top vv^\top w + \frac{1}{\lambda_1}\sum_{i=2}^{d}\lambda_i w^\top v_i v_i^\top w) \\ &= 1 - < v, w >^2 - \sum_{i=2}^{d}\frac{\lambda_i}{\lambda_1} < v_i, w >^2\end{aligned}$$

As Σ is psd we know $\lambda_i \geq 0$, which in turn gives us

$$\frac{\langle vv^\top - ww^\top, \Sigma\rangle}{\langle vv^\top, \Sigma\rangle} \leq 1 - < v, w >^2$$

and by definition

$$\sin(\theta) = \sqrt{1 - (\langle v, w \rangle)^2}$$

so we have

$$\frac{\langle vv^\top - ww^\top, \Sigma \rangle}{\langle vv^\top, \Sigma \rangle} \leq \sin^2(\theta)$$

which in turn means

$$\langle ww^\top, \Sigma \rangle \geq (1 - \sin^2(\theta)) \langle vv^\top, \Sigma \rangle$$

□

Theorem B.2. *k-to-1-ePCA reduction. Let $\varepsilon \in (0, 1)$, let $\Sigma \in \mathbb{S}^{d \times d}_{\succeq 0}$, let $A_1, \ldots, A_n$ be i.i.d. (not sure independent is needed) matrices with expectation Σ and let O_{1PCA} be a stochastic ePCA oracle. Then, Algorithm 4 returns $U \in \mathbb{R}^{d \times k}$ so that Σ.*

$$\langle UU^\top, \Sigma \rangle \geq (1 - \zeta^2) \|\Sigma\|_k$$

Proof. We will proof this by induction, where the $k = 1$ case follows from Lemma B.1. For $i + 1$ we note $P_i = \mathbb{I}_d - U_i U_i^\top$ then

$$\begin{aligned} \mathrm{Tr}(U_{i+1}^\top \Sigma U_{i+1}) &= \mathrm{Tr}(U_i^\top \Sigma U_i) + u_{i+1}^\top \Sigma u_{i+1} \\ &\geq (1 - \zeta^2) \|\Sigma\|_i + u_{i+1}^\top \Sigma u_{i+1} \end{aligned}$$

where the first step follows by linearity and the second step by induction assumption. Now we note

$$u_{i+1}^\top \Sigma u_{i+1} = \langle u_{i+1} u_{i+1}^\top, \Sigma \rangle \geq (1 - \zeta^2) \|P_i \Sigma P_i\|_2$$

as u_{i+1} can be seen as the approximation the oracle returns for the top eigenvalue of $P_i \Sigma P_i$ and therefore it must fulfill this equality by assumption on our oracle. By Lemma 3 in [Jambulapati et al., 2024] we know

$$\|P_i \Sigma P_i\|_2 \geq \lambda_{i+1}(\Sigma)$$

and this in turn gives us

$$\mathrm{Tr}(U_{i+1}^\top \Sigma U_{i+1}) \geq (1 - \zeta^2) \|\Sigma\|_{i+1}$$

which proofs our Claim. □

So if we can show ModifiedDP-PCA classifies as a stochastic $k = 1$ ePCA oracle, we will automatically get a utility guarantee for $k > 1$.

C Novel Analysis of non private Oja's Algorithm

Given $A_1, \ldots, A_n$ i.i.d with $\mathbb{E}[A_i] = \Sigma$, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ be the eigenvalues of Σ and $v_1, \ldots, v_n$ the corresponding eigenvectors. Further let P be a projection matrix independent of the A_i. Our goal is to compute an ε approximation of the top eigenvector of $P\Sigma P$. For P a deterministic matrix the analysis of Jain et al. [2016] of Oja's Algorithm shows that the output of the algorithm will be close to the top eigenvector of $P\Sigma P$. The problem we face is that in our case $P = I - \sum_i u_i u_i^\top$ where the u_i are random and obtained as an estimation using a previously sampled set of A_i's. So applying Jain et al.'s main theorem would only guarantee us that the output is close to the top eigenvector of $\mathbb{E}[P]\Sigma\mathbb{E}[P]$, but $\mathbb{E}[P]$ might not even be a projection matrix. Therefore, we give an alternative proof of Oja's Algorithm, showing that with inputs of the form $\{PA_iP\}$ we do get utility guarantees for being close to the top eigenvector of $P\Sigma P$ even when P is random.

From now on we will let $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_d$ refer to the eigenvalues of $P\Sigma P$ and we note that

$$\lambda_i \geq \tilde{\lambda}_i$$

as a natural consequence of applying a projection. We further assume that there are scalars $\mathcal{M}, \mathcal{V}$ such that

1. $\|A_i - \Sigma\|_2 \leq \mathcal{M}$ with probability 1
2. $\max\{\|\mathbb{E}[(A_i - \Sigma)(A_i - \Sigma)^\top]\|_2, \|\mathbb{E}[(A_i - \Sigma)^\top (A_i - \Sigma)]\|_2\} \leq \mathcal{V}$

Remark we on purpose use different symbols here than in Assumption A, as $\mathcal{M} = \lambda_1 M$ and similarly for $\mathcal{V}$, when compared to Assumption A. We then define

$$B_n := (\mathbf{I} + \eta_n P A_n P)(\mathbf{I} + \eta_{n-1} P A_{n-1} P) \cdot \cdots \cdot (\mathbf{I} + \eta_1 P A_1 P)$$
$$w_n := \frac{B_n w_0}{\|B_n w_0\|_2}$$
$$\bar{\mathcal{V}} := \mathcal{V} + \tilde{\lambda}_1^2$$

and note that w_n is the result of Oja's Algortihm after n steps given $\{PA_iP\}$ as input. The proof of accuracy of Algorithm 3 will require the following result we will proof below

Theorem C.1. *Given $A_1, \ldots, A_n$ fulfill Assumption 1. - 3. with parameters Σ, M, V, κ and a projection matrix P independent of the A_i, $\tilde{v}$ the top eigenvector of $P\Sigma P$ and B_n, ω_n the outputs resulting from non-private Oja's Algorithm given input $PA_1P, \ldots, PA_nP$*

$$\sin\left(\tilde{v}, \frac{B_n \omega_n}{\|B_n \omega_n\|_2}\right) \leq \frac{1}{Q} \exp \sum_{j \in [t]} \eta_j^2 5 \bar{\mathcal{V}} \left(d \exp(-2(\tilde{\lambda}_1 - \tilde{\lambda}_2) \sum_{j \in [t]} \eta_j \right) \tag{3}$$

Theorem C.2. *(Main theorem of this section) Fix any $\delta > 0$ and suppose the step sizes are set to $\eta_t = \frac{\alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)(\beta + t)}$ for $\alpha > \frac{1}{2}$ and*

$$\beta := 20 \max\left(\frac{\mathcal{M}\alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)}, \frac{\bar{\mathcal{V}}\alpha^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\delta}{100})} \right)$$

Suppose the number of iterations $n > \beta$. Then the output ω_n of Algorithm 1 satisfies:

$$1 - (\omega_n^\top \tilde{v})^2 \leq \frac{C \log(1/\delta)}{\delta^2} \left(d \left(\frac{\beta}{n}\right)^{2\alpha} + \frac{\alpha^2 \mathcal{V}}{(2\alpha - 1) \cdot (\tilde{\lambda}_1 - \tilde{\lambda}_2)^2} \cdot \frac{1}{n} \right), \tag{4}$$

with probability at least $1 - \delta$. Here C is an absolute numerical constant.

Proof. Analogously to the proof of Theorem 4.1 in [Jain et al., 2016] by replacing their Theorem 3.1 with our Theorem C.1. □

We state and proof several Lemmas that will allow us to proof Theorem C.1 which in turn will directly proof Theorem C.2.

Lemma C.3 (One Step Power Method [Jain et al., 2016])**.** *. Let $B \in d \times d$, let $v \in d$ be a unit vector, and let $V_\perp$ be a matrix whose columns form an orthonormal basis of the subspace orthogonal to v. If $w \in \mathbb{R}^d$ is chosen uniformly at random from the surface of the unit sphere then with probability at least $1 - \delta$*

$$\sin^2(v, \frac{Bw}{\|Bw\|_2}) = 1 - (v^\top Bw)^2 \leq C \frac{\log(1/\delta)}{\delta} \frac{\mathrm{Tr}(V_\perp^\top B B^\top V_\perp)}{v^\top B B^\top v} \tag{5}$$

where C is an absolute constant.

Based on the above Lemma we see that to show that Oja's algorithm succeeds we simply need to show that with high probability $\mathrm{Tr}(\tilde{V}_\perp^\top B_n B_n^\top \tilde{V}_\perp)$ is relatively large and $\tilde{v}^\top B_n B_n^\top \tilde{v}$ is relatively small. Note so long as we pick η_i sufficiently small, i.e. $\eta_i = O(1/maxM, \tilde{\lambda}_1)$ then $\mathbf{I} + \eta_i PAiP$ is invertible, so in turn $B_n B_n^\top$, which guarantees $\tilde{v}^\top B_n B_n^\top \tilde{v} > 0$, so the RHS of the inequality will always be finite. In order to explicitly bound the RHS we will utilize conditional Chebychev's and Markov's, where the conditioning will serve to fix P.

Lemma C.4. $\|\mathbb{E}[B_t B_t^\top | P]\|_2 \leq \exp(\sum_{i \in [t]} 2\eta_i \tilde{\lambda}_1 + \eta_i^2(\tilde{\lambda}_1^2 + \mathcal{V}))$

Proof. We will denote $\alpha_t = \|\mathbb{E}[B_t B_t^\top | P]\|_2$ in this proof. Note that $\mathbb{E}[B_t B_t^\top | P] \preceq \alpha_t \mathbf{I}$, so by Lemma A.11

$$\begin{aligned}
\mathbb{E}[B_t B_t^\top | P] &= \mathbb{E}[(\mathbf{I} + \eta_t P A_t P) B_{t-1} B_{t-1}^\top (\mathbf{I} + \eta_t P A_t P)^\top | P] \\
&\preceq \alpha_{t-1} \mathbb{E}[(\mathbf{I} + \eta_t P A_t P)(\mathbf{I} + \eta_t P A_t P)^\top | P] \\
&= \alpha_{t-1} \mathbb{E}[\mathbf{I} + \eta_t P A_t P + + \eta_t P A_t^\top P + \eta_t^2 P A_t P A_t^\top P | P] \\
&= \alpha_{t-1} (\mathbf{I} + 2\eta_t P \Sigma P + \eta_t^2 \mathbb{E}[P A_t P A_t^\top P | P]
\end{aligned}$$

we can easily bound $P\Sigma P$ via $P\Sigma P \preceq \tilde{\lambda}_1 \mathbf{I}$. Further

$$\begin{aligned}
\mathbb{E}[PA_tPA_t^\top P|P] &= P\Sigma P\Sigma P + \mathbb{E}[(P(A_t-\Sigma)P(A_t-\Sigma)^\top P|P] \\
&= P\Sigma P\Sigma P + P\mathbb{E}[(A_t-\Sigma)P(A_t-\Sigma)^\top|P]P \\
&\preceq \tilde{\lambda}_1^2\mathbf{I} + \mathbb{E}[(A_t-\Sigma)(A_t-\Sigma)^\top|P] \\
&= \tilde{\lambda}_1^2\mathbf{I} + \mathbb{E}[(A_t-\Sigma)(A_t-\Sigma)^\top] \\
&\preceq (\tilde{\lambda}_1^2 + \mathcal{V})\mathbf{I}
\end{aligned}$$

where the third step follows as $\|P\|_2 \leq 1$, the 4th as P is independent of A_t and the last step by assumption on the A_i. So in total this gives us

$$\alpha_t \leq \alpha_{t-1}(1 + 2\eta_t\tilde{\lambda}_1 + \eta_t^2(\tilde{\lambda}_1^2 + \mathcal{V}))$$

As α_0 and $1 + x \leq e^x$ this gives us

$$\alpha_t \leq \exp(\sum_{i\in[t]} 2\eta_i\tilde{\lambda}_1 + \eta_i^2(\tilde{\lambda}_1^2 + V))$$

□

Lemma C.5. $\mathbb{E}[\tilde{v}^\top B_tB_t\tilde{v}|P] \geq \exp(\sum_{i\in[t]}(2\eta_i\tilde{\lambda}_1 - 4\eta_i^2\tilde{\lambda}_1^2))$

Proof. We define $\beta_t := \mathbb{E}[\tilde{v}^\top B_tB_t^\top\tilde{v}|P]$, since $B_t = (\mathbf{I} + \eta_tPA_tP)B_{t-1}$ we have

$$\beta_t = \langle \mathbb{E}[B_{t-1}B_{t-1}^\top|P], \mathbb{E}[(\mathbf{I} + \eta_tPA_tP)\tilde{v}\tilde{v}^\top(\mathbf{I} + \eta_tPA_tP)^\top|P]\rangle$$

because $\langle A, B\rangle := \mathrm{Tr}(A^\top B)$ and the trace is invariant under cyclic permutations. The RHS of the matrix inner product we can lower bound as follows:

$$\begin{aligned}
\mathbb{E}[(\mathbf{I} + \eta_tPA_tP)\tilde{v}\tilde{v}^\top(\mathbf{I} + \eta_tPA_tP)^\top|P] &= \tilde{v}\tilde{v}^\top + \eta_tP\Sigma P\tilde{v}\tilde{v}^\top + \eta_t\tilde{v}\tilde{v}^\top P\Sigma P + \eta_t^2\mathbb{E}[PA_tP\tilde{v}\tilde{v}^\top PA_t^\top P|P] \\
&\geq \tilde{v}\tilde{v}^\top + \eta_tP\Sigma P\tilde{v}\tilde{v}^\top + \eta_t\tilde{v}\tilde{v}^\top P\Sigma P \\
&= \tilde{v}\tilde{v}^\top + 2\eta_t\tilde{\lambda}_1\tilde{v}\tilde{v}^\top
\end{aligned}$$

where the last step follows as $\tilde{v}$ is the top eigenvector of $P\Sigma P$ by assumption. So all together we get

$$\beta_t \geq \langle \mathbb{E}[B_{t-1}B_{t-1}^\top|P], (1 + 2\tilde{\lambda}_1\eta_t)\tilde{v}\tilde{v}^\top\rangle = (1 + 2\tilde{\lambda}_1\eta_t)\beta_{t-1}$$

as $B_0 = \mathbf{I}$, we have $\beta_0 = \|\tilde{v}\|_2^2 = 1$ and then by applying $1 + x \geq \exp(x - x^2)$ for all $x > 0$ we get

$$\beta_t \geq \exp(\sum_{i=1}^t 2\tilde{\lambda}_1\eta_i - \sum_{i=1}^t 4\tilde{\lambda}_1^2\eta_i^2)$$

□

Lemma C.6. $\mathbb{E}[(\tilde{v}^\top B_tB_t\tilde{v})^2|P] \leq \exp(\sum 4\eta_i\tilde{\lambda}_1 + 10\eta_i^2\bar{\nu})$

Proof. We define $\gamma_s := \mathbb{E}[(\tilde{v}^\top W_{t,s}W_{t,s}^\top\tilde{v})^2|P]$ where $W_{t,s} := (\mathbf{I} - \eta_tPA_iP)\cdot\ldots(\mathbf{I} + \eta_{t-s+1}PA_{t-s+1}P)$. So by this definition we see $W_{t,t} = B_t$ and $\gamma_t = \mathbb{E}[(\tilde{v}^\top B_tB_t^\top\tilde{v})^2|P]$. As the trace of a scalar is the scalar itself, we can exploit the cyclic permutation properties of the trace:

$$\begin{aligned}
\gamma_t &= \mathrm{Tr}(\mathbb{E}[W_{t,t}^\top\tilde{v}\tilde{v}^\top W_{t,t}W_{t,t}^\top\tilde{v}\tilde{v}^\top W_{t,t}|P]) \\
&= \mathrm{Tr}(\mathbb{E}[(\mathbf{I} + \eta_1A_1^\top)G_{t-1}(\mathbf{I} + \eta_1A_1)(\mathbf{I} + \eta_1A_1^\top)G_{t-1}(\mathbf{I} + \eta_1A_1)|P])
\end{aligned}$$

where $G_{t-1} := W_{t,t-1}^\top v_1v_1^\top W_{t,t-1}$. We first bound for an arbitrary $G_{t-1} = G$, and then take the expectation over only A_1 and finally over G_{t-1}.

$$\begin{aligned}
&\mathrm{Tr}(\mathbb{E}[(\mathbf{I} + \eta_1PA_1^\top P)G(\mathbf{I} + \eta_1PA_1P)(\mathbf{I} + \eta_1PA_1^\top P)G(\mathbf{I} + \eta_1PA_1P)|P]) \\
=&\mathrm{Tr}(\mathbb{E}[(G + \eta_1PA_1^\top PG + \eta_1GPA_1P + \eta_1^2PA_1^\top PGPA_1P)^2|P]) \\
=&\mathrm{Tr}(G^2) + 4\eta_1\mathrm{Tr}(P\Sigma PG^2) + 2\eta_1^2\mathrm{Tr}(\mathbb{E}[PA_1PA_1^\top P|P]G^2) \\
&+ \eta_1^2\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1PG|P]) + \eta_1^2\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1^\top PG|P]) \\
&+ \eta_1^2\mathrm{Tr}(\mathbb{E}[GPA_1PGPA_1P|P]) + \eta_1^2\mathrm{Tr}(\mathbb{E}[GPA_1^\top PGPA_1P|P]) \\
&+ 2\eta_1^3\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1^\top PGPA_1P|P]) \\
&+ \eta_1^4\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1PA_1^\top PGPA_1P|P]))
\end{aligned}$$

Let's begin with the first order terms:

$$\mathrm{Tr}(P\Sigma PG^2) \le \|P\Sigma P\|_2 Tr(G^2) = \tilde{\lambda}_1 \mathrm{Tr}(G^2)$$
$$\mathrm{Tr}(\mathbb{E}[PA_1PA_1^\top P|P]G^2) \le (\|\mathbb{E}[P(A_1-\Sigma)P(A_1^\top-\Sigma)P]\|_2 + \|P\Sigma P\Sigma P\|_2)\mathrm{Tr}(G^2) \le (\mathcal{V}+\tilde{\lambda}_1^2)\mathrm{Tr}(G^2)$$

where the last inequality follows by Lemma A.10. Next we have 4 second order terms:

$$\begin{aligned}
&\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1PG|P]) = \mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1^\top PG|P]) \\
=&\mathrm{Tr}(\mathbb{E}[GPA_1PGPA_1P|P]) = \mathrm{Tr}(\mathbb{E}[GPA_1^\top PGPA_1P|P]) \\
\le&\frac{1}{2}\mathbb{E}[\|PA_1^\top PG\|_F^2 + \|PA_1PG\|_F^2|P] \\
=&\frac{1}{2}\mathrm{Tr}(G\mathbb{E}[PA_1PA_1^\top P|P]G + G\mathbb{E}[PA_1PA_1^\top P|P]G)) \le (\mathcal{V}+\tilde{\lambda}_1^2)\mathrm{Tr}(G^2)
\end{aligned}$$

Third order terms we can bound as follows:

$$\begin{aligned}
\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1^\top PGPA_1P|P]) &\le \|PA_1^\top P\|\mathrm{Tr}(\mathbb{E}[PA_1^\top PGGPA_1P)|P] \\
&\le (\|P(A_1-\Sigma)P\|_2 + \|P\Sigma P\|_2)\mathrm{Tr}(G\mathbb{E}[PA_1PA_1^\top P|P]G) \\
&\le (\mathcal{M}+\tilde{\lambda}_1)(\mathcal{V}+\tilde{\lambda}_1)\mathrm{Tr}(G^2)
\end{aligned}$$

Finally the fourth order terms

$$\begin{aligned}
\mathrm{Tr}(\mathbb{E}[PA_1^\top PGPA_1PA_1^\top PGPA_1P|P])) &\le \|\mathbb{E}[PA_1PA_1^\top P]\|_2\mathrm{Tr}(G\mathbb{E}[PA_1PA_1^\top P|P]G) \\
&\le (\mathcal{M}+\tilde{\lambda}_1)^2(\mathcal{V}+\tilde{\lambda}_1)\mathrm{Tr}(G^2)
\end{aligned}$$

all of this together gives us

$$\begin{aligned}
&\mathrm{Tr}(\mathbb{E}[(\mathbf{I}+\eta_1PA_1^\top P)G(\mathbf{I}+\eta_1PA_1P)(\mathbf{I}+\eta_1PA_1^\top P)G(\mathbf{I}+\eta_1PA_1P)|P]) \\
\le&\mathrm{Tr}(G^2) + 4\eta_1\tilde{\lambda}_1\mathrm{Tr}(G^2) + 5\eta_1^2\bar{\mathcal{V}}\mathrm{Tr}(G^2) + 4\eta_1^3(\mathcal{M}+\tilde{\lambda}_1)\bar{\mathcal{V}}\mathrm{Tr}(G^2) + \eta_1^4(\mathcal{M}+\tilde{\lambda}_1)^2\bar{\mathcal{V}}\mathrm{Tr}(G^2) \\
=&(1+4\eta_1\tilde{\lambda}_1 + 5\eta_1^2\bar{\mathcal{V}} + 4\eta_1^3(\mathcal{M}+\tilde{\lambda}_1)\bar{\mathcal{V}} + \eta_1^4(\mathcal{M}+\tilde{\lambda}_1)^2\bar{\mathcal{V}})\mathrm{Tr}(G^2) \\
\le&(1+4\eta_1\tilde{\lambda}_1 + 10\eta_1^2\bar{\mathcal{V}})\mathrm{Tr}(G^2) \\
\le&\exp(4\eta_1\tilde{\lambda}_1 + 10\eta_1^2\bar{\mathcal{V}})\mathrm{Tr}(G^2)
\end{aligned}$$

where we used $\eta_i \le \frac{1}{4\max\{\lambda_1,\mathcal{M}\}}$ and $1+x \le \exp(x)$. All of this give us

$$\gamma_t \le \exp(4\eta_1\tilde{\lambda}_1 + 10\eta_1^2\bar{\mathcal{V}})\mathbb{E}[\mathrm{Tr}(G_{t-1}^2)|P] = \exp(4\eta_1\tilde{\lambda}_1 + 10\eta_1^2\bar{\mathcal{V}})\gamma_{t-1}$$

then using $\gamma_0 = 1$ gives us the wished result. □

Lemma C.7. $\mathbb{E}[\mathrm{Tr}(\tilde{V}_\perp^\top B_tB_t^\top\tilde{V}_\perp)|P] \le \exp(\sum_{j=1}^t 2\eta_j\tilde{\lambda}_2 + \eta_j^2\bar{\mathcal{V}})(d + \sum_{i\in[t]}\eta_i^2\mathcal{V}\exp(\sum_{j\in[i]}2\eta_j(\tilde{\lambda}_1-\tilde{\lambda}_2))$

Proof. Let $\alpha_t := \mathbb{E}[\mathrm{Tr}(\tilde{V}_\perp^\top B_tB_t^\top\tilde{V}_\perp)|P]$. Then using the cyclic property of the trace and the fact that $\tilde{V}_\perp$ is not random in $\mathbb{E}[\cdot|P]$, we have

$$\begin{aligned}
\alpha_t &= \langle\mathbb{E}[B_tB_t^\top|P], \tilde{V}_\perp\tilde{V}_\perp^\top\rangle \\
&= \langle\mathbb{E}[B_{t-1}B_{t-1}^\top|P], \mathbb{E}[(\mathbf{I}+\eta_tPA_tP)\tilde{V}_\perp\tilde{V}_\perp^\top(\mathbf{I}+\eta_tPA_tP)|P]\rangle
\end{aligned}$$

the RHS of this matrix inner product equates to

$$\begin{aligned}
&\tilde{V}_\perp\tilde{V}_\perp^\top + \eta_tP\Sigma P\tilde{V}_\perp\tilde{V}_\perp^\top + \eta_t\tilde{V}_\perp\tilde{V}_\perp^\top P\Sigma P + \eta_t^2\mathbb{E}[PA_tP\tilde{V}_\perp\tilde{V}_\perp^\top PA_tP|P] \\
\preceq&\tilde{V}_\perp\tilde{V}_\perp^\top + 2\eta_t\tilde{\lambda}_2\tilde{V}_\perp\tilde{V}_\perp^\top + \eta_t^2\tilde{\lambda}_1^2)\tilde{V}_\perp\tilde{V}_\perp^\top + \mathbb{E}[P(A_t-\Sigma)P(A_t-\Sigma)P|P] \\
\preceq&(1+2\eta_t\tilde{\lambda}_2 + \eta_t^2\bar{\mathcal{V}})\tilde{V}_\perp\tilde{V}_\perp^\top + \eta_t^2\mathcal{V}v_1v_1^\top
\end{aligned}$$

where we used that $\tilde{V}_\perp$ is orthogonal to the top eigenvector of $P\Sigma P$ and that $\tilde{V}_\perp\tilde{V}_\perp^\top \preceq \mathbf{I}$ as it is an orthogonal matrix. So plugging this into the inner product we get

$$\alpha_t \le (1+2\eta_t\tilde{\lambda}_2 + \eta_t^2\bar{\mathcal{V}})\langle\mathbb{E}[B_{t-1}B_{t-1}^\top|P], \tilde{V}_\perp\tilde{V}_\perp^\top\rangle + eta_t^2\mathcal{V}\langle\mathbb{E}[B_{t-1}B_{t-1}^\top|P], v_1v_1^\top\rangle$$

using $1+x \leq \exp(x)$ we get

$$\begin{aligned}\alpha_t &\leq \exp(2\eta_t\tilde{\lambda}_2 + \eta_t^2\bar{\mathcal{V}})\alpha_{t-1} + \eta_t^2\mathcal{V}\|\mathbb{E}[B_{t-1}B_{t-1}^\top|P]\|_2 \\ &\leq \exp(2\eta_t\tilde{\lambda}_2 + \eta_t^2\bar{\mathcal{V}})\alpha_{t-1} + \eta_t^2\mathcal{V}\exp(\sum_{i\in[t-1]} 2\eta_i\tilde{\lambda}_1 + \eta_i^2\bar{\mathcal{V}})\end{aligned}$$

using Lemma C.4. Then by recursion we get

$$\alpha_t \leq \exp(\sum_{j=1}^{t} 2\eta_j\tilde{\lambda}_2 + \eta_j^2\bar{\mathcal{V}})\alpha_0 + \sum_{i\in[t]}\eta_i^2\mathcal{V}\exp(\sum_{j\in[i]} 2\eta_j\tilde{\lambda}_1 + \eta_j^2\bar{\mathcal{V}})\exp(\sum_{j=i+1}^{t} 2\eta_j\tilde{\lambda}_2 + \eta_j^2\bar{\mathcal{V}})$$

the result follows by $\alpha_0 = d-1 \leq d$ □

Proof of Theorem C.1 Using conditional Chebychev's (Lemma A.7) we have

$$\mathbb{P}\left[|\tilde{v}^\top B_n B_n^\top \tilde{v} - \mathbb{E}[\tilde{v}^\top B_n B_n^\top \tilde{v}|P]| > \frac{1}{\sqrt{\delta}}\sqrt{\text{Var}[\tilde{v}B_nB_n^\top\tilde{v}|P]}\right] < \delta$$

so with probability $1-\delta$ given P is fixed $\tilde{v}^\top B_n B_n^\top \tilde{v}$ lies in the interval around it's expectation. So we know with probability at least $1-\delta$

$$\begin{aligned}\tilde{v}^\top B_n B_n^\top \tilde{v} &> \mathbb{E}[\tilde{v}^\top B_n B_n^\top \tilde{v}|P] - \frac{1}{\sqrt{\delta}}\sqrt{\text{Var}[\tilde{v}B_nB_n^\top\tilde{v}|P]} \\ &= \mathbb{E}[\tilde{v}^\top B_n B_n^\top \tilde{v}|P] - \frac{1}{\sqrt{\delta}}\sqrt{\mathbb{E}[(\tilde{v}B_nB_n^\top\tilde{v})^2|P] - \mathbb{E}[\tilde{v}B_nB_n^\top\tilde{v}|P]^2} \\ &= \mathbb{E}[\tilde{v}^\top B_n B_n^\top \tilde{v}|P](1 - \frac{1}{\sqrt{\delta}}\sqrt{\Delta - 1}\end{aligned}$$

with

$$\begin{aligned}\Delta = \frac{\mathbb{E}[(\tilde{v}B_nB_n^\top\tilde{v})^2|P]}{\mathbb{E}[\tilde{v}B_nB_n^\top\tilde{v}|P]^2} &\leq \frac{\mathbb{E}[(\tilde{v}B_nB_n^\top\tilde{v})^2|P]}{\exp(\sum_{i\in[t]} 2\eta_i\tilde{\lambda}_1 - 4\eta_i^2\tilde{\lambda}_1^2)^2} \\ &\leq \frac{\exp(\sum_{i\in[t]} 4\eta_i\tilde{\lambda}_1 + 10\eta_i^2\bar{\mathcal{V}})}{\exp(\sum_{i\in[t]} 2\eta_i\tilde{\lambda}_1 - 4\eta_i^2\tilde{\lambda}_1^2)^2} \\ &= \frac{\exp(\sum_{i\in[t]} 4\eta_i\tilde{\lambda}_1 + 10\eta_i^2\bar{\mathcal{V}})}{\exp(\sum_{i\in[t]} 4\eta_i\tilde{\lambda}_1 - 8\eta_i^2\tilde{\lambda}_1^2)} \leq \exp(\sum_{i\in[t]} 18\eta_i^2\bar{\mathcal{V}})\end{aligned}$$

where we use Lemma C.5 and Lemma C.6. So putting this together we get

$$\tilde{v}^\top B_n B_n^\top \tilde{v} \geq \exp\left(\sum_{i\in[n]}(2\eta_i\tilde{\lambda}_1 - 4\eta_i^2\tilde{\lambda}_1^2)\right)\left(1 - \frac{1}{\sqrt{\delta}}\sqrt{\exp(\sum_{i\in[n]} 18\eta_i^2\bar{\mathcal{V}}) - 1}\right) \tag{6}$$

this lower bounds the denominator in Lemma C.3. So next we will upper bound the nominator to complete the proof. Markov's inequality gives us

$$\text{Tr}(\tilde{V}_\perp^\top B_t B_t^\top \tilde{V}_\perp) \leq \mathbb{E}[\text{Tr}(\tilde{V}_\perp^\top B_t B_t^\top \tilde{V}_\perp)] \cdot \frac{1}{\delta}$$

holds with probability $1-\delta$. So by Lemma C.7 we get

$$\text{Tr}(\tilde{V}_\perp^\top B_t B_t^\top \tilde{V}_\perp) \leq \frac{1}{\delta}\exp\left(\sum_{j\in[t]} 2\eta_j\tilde{\lambda}_2 + \eta_j^2\bar{\mathcal{V}}\right)\left(d + \mathcal{V}\sum_{i=1}^{t}\eta_i^2\exp\left(\sum_{j\in[i]} 2\eta_j(\tilde{\lambda}_1 - \tilde{\lambda}_2)\right)\right) \tag{7}$$

so plugging Equation 6 and 7 into Lemma C.3 we get that with probability at least $1-2\delta$

$$\sin^2(\tilde{v}, \frac{B_n w_n}{\|B_n w_n\|_2})$$

$$\leq \frac{C\log(1/\delta)}{\delta}\frac{1}{(1-\frac{1}{\sqrt{\delta}}\sqrt{\exp(\sum_{i\in[n]} 18\eta_i^2\bar{\mathcal{V}}) - 1}}\exp(\sum_{j\in[t]} 2\eta_j(\tilde{\lambda}_2 - \tilde{\lambda}_1) + \eta_j^2(\bar{\mathcal{V}} + 4\lambda_1^2))(d + \mathcal{V}\sum_{i=1}^{t}\eta_i^2\exp(\sum_{j\in[i]} 2\eta_j(\tilde{\lambda}_1 - \tilde{\lambda}_2)))$$

By $\bar{\mathcal{V}} + 4\tilde{\lambda}_1^2 \leq 5\bar{\mathcal{V}}$ and the definition of Q the result follows.

D Proof of Main Theorem

The privacy proof of Algorithm 2 follows straight from the privacy of ModifiedDP-PCA, which in turn follows by [Liu et al., 2022], however the utility proof is more involved. We cannot apply DP-PCA straight away as this would only give us a guarantee that the vector $\tilde{v}$ we obtain is a good approximation of the top eigenvector of $\mathbb{E}[P]\Sigma\mathbb{E}[P]$. This is not sufficient for the deflation method, as we require $\tilde{v}$ to be a good approximation of $P\Sigma P$. We show that for ModifiedDP-PCA this is indeed the case. By first showing that with high likelyhood we can reduce the update step to an update step of non private Oja's Algorithm with matrices PC_tP. We then use a novel result we proved (which for readability we added to the appendix), which shows that non private Oja's Algorithm given input $\{PC_tP\}_t$ will return a good approximation of $P\mathbb{E}[C_t]P$ under some assumptions on C_t which we will show, our data fulfills.

Theorem D.1. *(Utility of ModifiedDP-PCA) For $\varepsilon \in (0, 0.9)$ and $0 < k < d$, ModifiedDP-PCA fulfills (ε, δ)-DP for all inputs $\{A_i\}, B, \zeta$ and δ and any projection matrix P (that we assume to already be private). Given n i.i.d. samples $\{A_i \in \mathbb{R}^{d\times d}\}_{i=1}^n$ satisfying Assumption 1. - 4. with parameters $(\Sigma, M, V, K, \kappa, a, \gamma^2)$, if*

$$n = \tilde{O}\left(e^{\kappa'^2} + \frac{d\kappa'\gamma(\log(1/\delta))^{1/2}}{\varepsilon} + \kappa' M + \kappa'^2 V + \frac{d^{1/2}(\log(1/\delta))^{3/2}}{\varepsilon}\right)$$

where $\kappa' = \frac{\lambda_1(\Sigma)}{\lambda_1(P\Sigma P)-\lambda_2(P\Sigma P)}$ with a large enough constant and $\delta \leq 1/n$. If further

$$0 < \lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)$$

then there exists a learning rate η_t that depends on $(t, M, V, K, a, \lambda_1(\Sigma), \lambda_1(P\Sigma P) - \lambda_2(P\Sigma P), n, d\varepsilon, \delta)$ such that $T = \lfloor n/B \rfloor$ steps of ModifiedDP-PCA with choices of $\tau = 0.01$ and $B = c_1 n/(\log n)^3$ outputs ω_T such that with probability 0.99

$$\sin(\omega_t, \tilde{v}) \leq \tilde{O}\left(\kappa'\left(\sqrt{\frac{V}{n}} + \frac{\gamma d\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$$

where $\tilde{v}$ is the top eigenvector of $P\Sigma P$ and $\tilde{O}(\cdot)$ hides poly-logarithmic factors in $n, d, 1/\varepsilon$, and $\log(1/\delta)$ and polynomial factors in K.

Remark. For readability we omitted the advanced composition details in the above proof. If we choose $T = O(log2n)$, we can simply choose $(\varepsilon', \delta') = (\varepsilon/(2\sqrt{2\log^2(n)log(2/\delta)})), \delta/(2\log^2(n)))$ in every step and then by advanced composition we get. And in our utility guarantee we would only occur additional $\log^2(n)$ factors which we omit.

Proof. We choose the batch size $B = \Theta(n/\log^3 n)$ such that we access the dataset only $T = \Theta(\log^3 n)$ times. Hence we do not need to rely on amplification by shuffling. To add Gaussian noise that scales as the standard deviation of the gradients in each minibatch (as opposed to potentially excessively large mean of the gradients), DP-PCA first gets a private and accurate estimate of the range. Using this estimate, Λ, Private Mean Estimation returns an unbiased estimate of the empirical mean of the gradients, as long as no truncation has been applied. As we choose the truncation threshold so that with high probability there will be no truncation the update step will look as follows:

$$\omega_t' \leftarrow \omega_{t-1} + \eta_t P(\frac{1}{B}\sum_{i\in[B]} PA_iP\omega_{t-1} + \beta_t z_t)$$

where $z_t \sim \mathcal{N}(0, \mathbf{I})$ and $\beta_t = \frac{8K\sqrt{2\hat{\Lambda}_t}\log^a(Bd/\tau)\sqrt{2d\log(2.5/\delta)}}{\varepsilon B}$. The privacy follows by the privacy of the subroutines private eigenvalue and private mean estimation [Liu et al., 2022]. So all that is left to do is show the utility guarantee. We will do that by showing we can reduce it the accuracy of the non private case. First we note that $P^2 = P$ so we get

$$\omega_t' = \omega_{t-1} + \eta_t(\frac{1}{B}\sum_{iin[B]} PA_iP\omega_{t-1} + \beta_t Pz_t)$$

Using rotation invariance of the spherical Gaussian random vectors and the fact that $\|\omega_{t-1}\| = 1$ and $\omega_{t-1} \in \text{Im}(P)$ (for details see Lemma A.8), we can reformulate it as

$$\omega_t' \leftarrow \omega_{t-1} + \eta_t\left(\frac{1}{B}\sum_{i\in[B]} PA_iP + \beta_t PG_tP\right)\omega_{t-1}$$

we can further pull out the projection matrices to obtain

$$\omega_t' \leftarrow \omega_{t-1} + \eta_t P \left(\frac{1}{B} \sum_{i \in [B]} A_i + \beta_t G_t \right) P \omega_{t-1}$$

Where G is a matrix whose entries are i.i.d. $\mathcal{N}(0,1)$ distributed. So we have a matrix

$$C_t := \frac{1}{B} \sum_{i \in [B]} A_i + \beta_t G_t$$

and we will now proof that C_t fulfills all requirements for Theorem C.2 (our version of the non private Oja's Algorithm utility guarantee), which will directly give us the wished utility guarantee. It is easy to see that $\mathbb{E}[C_t] = \Sigma$ as z is a zero mean random variable and hence so is G_t. Next we show the upper bound of $\max\{\|\mathbb{E}[(C_t - \Sigma)(C_t - \Sigma)^\top]\|_2, \|\mathbb{E}[(C_t - \Sigma)^\top (C_t - \Sigma)]\|_2\}$

$$\begin{aligned}
&\|\mathbb{E}[(C_t - \Sigma)(C_t - \Sigma)^\top]\|_2 \\
=&\|\mathbb{E}[(\frac{1}{B} \sum_{i \in [B]} A_i + \beta_t G_t - \Sigma)(\frac{1}{B} \sum_{i \in [B]} A_i + \beta_t G_t - \Sigma)^\top]\|_2 \\
\leq&\|\mathbb{E}[(\frac{1}{B} \sum_{i \in [B]} A_i - \Sigma)(\frac{1}{B} \sum_{i \in [B]} A_i - \Sigma)^\top]\|_2 + \beta_t^2 \|\mathbb{E}[G_t G_t^\top]\|_2 \\
\leq&V\lambda_1^2/B + \beta_t^2 \|\mathbb{E}[G_t G_t^\top]\|_2 \\
\leq&V\lambda_1^2/B + \beta^2 C_2 d =: \tilde{V}
\end{aligned}$$

where the first inequality holds due to G_t being independent to A_i, and $\mathbb{E}[G_t] = 0$. The second inequality follows due to having B elements of $\frac{1}{B^2}\|\mathbb{E}[(A_i - \Sigma)^\top (A_i - \Sigma)]\|_2$ and Assumption 3. And the last inequality holds with high probability due to G_t having i.i.d. Gaussian entries, and by choosing

$$\beta := \frac{16K\gamma\lambda_1 \log^a(Bd/\tau)\sqrt{2d\log(2.5/\delta)}}{\varepsilon B}$$

we have $\beta \geq \beta_t$ for all t as by Theorem 6.1 in [Liu et al., 2022] and Assumption 4

$$\hat{\Lambda} \leq \sqrt{2}\lambda_1^2 \|H_u\|_2 \leq \sqrt{2}\lambda_1^2 \gamma$$

Lastly let us consider $\|C_t - \Sigma\|_2$. By Lemma A.2 and Lemma A.3 we know with proabability $1 - \tau$ for all $t \in [T]$

$$\begin{aligned}
&\|C_t - \Sigma\|_2 \\
=&\|\frac{1}{B} \sum_{i \in [B]} A_i + \beta_t G_t - \Sigma\| \\
\leq&\left(\frac{M\lambda_1 \log(dT/\tau}{B} + \sqrt{\frac{V\lambda_1^2 \log(dT/\tau}{B}} + \beta(\sqrt{d} + \sqrt{\log(T/\tau)}) \right) =: \tilde{M}
\end{aligned}$$

so by Theorem C.2 for

$$T \geq 20 \max\left(\frac{\tilde{M}\alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)}, \frac{(\tilde{V} + \lambda_1^2)\alpha^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\zeta}{100})} \right) := \xi \tag{8}$$

with probability $1 - \zeta$

$$\sin^2(w_T, \tilde{v}) \leq \frac{C\log(1/\delta)}{\delta^2} \left(d\left(\frac{\xi}{T}\right)^{2\alpha} + \frac{\alpha^2 \tilde{V}}{(2\alpha - 1)(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 T} \right)$$

so if we fill in $\tilde{M}$, $\tilde{V}$, and β into ξ and use $n = BT$ we get

$$\frac{\xi}{T} := 20 \max\left\{ \begin{array}{l} \frac{\lambda_1 M \log(dT/\tau\alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)n} + \sqrt{\frac{V\log(dT/\tau}{nT}} \cdot \frac{\lambda_1 \alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)} + \frac{K\gamma\lambda_1 \log^a(nd/T\tau\sqrt{2\log(2.5/\delta)}\sqrt{log(T/\tau}d\alpha}{\varepsilon n(\tilde{\lambda}_1 - \tilde{\lambda}_2)}, \\ \frac{V\lambda_1^2\alpha^2}{n(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\zeta}{100})} + \frac{K^2\gamma^2\lambda_1^2 \log^{2a}(Bd/\tau d^2 \log(2.5/\delta)\alpha^2}{\varepsilon^2 n^2 (\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\zeta}{100})} + \frac{\lambda_1^2\alpha^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\zeta}{100})T} \end{array} \right.$$

in order for Theorem C.2 to hold we need to force $\xi/T \leq 1$. Noting $\tau = O(1)$, $K = O(1)$ and selecting $\alpha = c \log n$, $T = c'(\log n)^3$ we get that

$$\frac{\xi}{T} \leq 20C \max \left\{ \begin{array}{l} \frac{\lambda_1 M \log(d \log(n)) \log n}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)n} + \sqrt{\frac{V \log(d \log(n))}{n}} \cdot \frac{\lambda_1}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)} + \frac{\gamma \lambda_1 \log^2(nd/\log(n)) \sqrt{\log(1/\delta) \log(\log(n))} \log(n) d}{\varepsilon(\tilde{\lambda}_1 - \tilde{\lambda}_2)} \\ \frac{V \lambda_1^2 (\log n)^2}{n(\tilde{\lambda}_1 - \tilde{\lambda}_2)} + \frac{\gamma^2 \lambda_1^2 \log^{2a}(nd/\log(n)) \log(1/\delta) d^2 \alpha^2}{\varepsilon^2 n^2 (\tilde{\lambda}_1 - \tilde{\lambda}_2)^2} + \frac{\lambda_1^2 (\log n)^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 T} \end{array} \right.$$

so $\frac{\xi}{T} \leq 1$ will be trivially fulfilled if each of the summand is smaller than 1/3. For the last term we need

$$\frac{\lambda_1^2 (\log n)^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 T} \leq 1/3$$

as $T = c'(\log(n))^3$ this means

$$\log n \geq 3 \frac{\lambda_1}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2}$$

for the remaining terms we need

$$\begin{aligned} \frac{n}{\log^a(n/\log n)\log(n)} &\geq 3 \frac{\gamma \lambda_1 \sqrt{\log(1/\delta)} d}{\varepsilon(\tilde{\lambda}_1 - \tilde{\lambda}_2)} \\ \frac{n}{(\log(n))^2} &\geq 3 \frac{V \lambda_1^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2} \\ \frac{n}{\log(\log(n))} &\geq \sqrt{3} \sqrt{V \log(d)} \\ \frac{n}{\log(n)\log(\log(n))} &\geq 3 \frac{\lambda_1 M \log(d)}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)} \end{aligned}$$

We note that to obtain $n/log(n) \geq a$, $n \simeq a \log(a) + a \log\log(a)$. So

$$n \gtrsim C' \left(\exp(\lambda_1^2/(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2) + \frac{M \lambda_1}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)} + \frac{V \lambda_1^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2} + \frac{d \gamma \lambda_1 \sqrt{\log(1/\delta)}}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)\varepsilon} \right)$$

with large enough constant suffices (where $\gtrsim$ is hiding log terms) to obtain $\xi/T \leq 1$ and $d(\xi/T)^{2\alpha} \leq 1/n^2$. And we get

$$\frac{\tilde{V}}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)} \lesssim C'' \left(\frac{V \lambda_1^2}{n} + \frac{\gamma^2 \lambda_1^2 d^2 \log(1/\delta)}{\varepsilon n} \right)$$

(where $\lesssim$ is hiding log terms), so plugging this in our bound for $\sin(\omega_T, \tilde{v})$ we get

$$\sin(\omega_T, \tilde{v}) \leq \tilde{O} \left(\kappa' \left(\sqrt{\frac{V}{n}} + \frac{\gamma d \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right)$$

which finishes the proof □

The above utility result depends on the eigenvalues of the input. After the first step of k-DP-PCA our input is of the form $PA_1P, \ldots, PA_nP$, so our utility bound depends on the eigengap of $P\Sigma P$. Now in general $\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)$ can be arbitrarily much smaller than the actual eigengap of Σ, and therefore it is not a sufficient utility bound as is. However, as we iteratively apply projection matrices of the form

$$P = I - uu^\top$$

where u is a unit vector, and further u is ε-close to the top eigenvector of the matrix we apply it to, we can actually relate the eigengap of $P\Sigma P$ to the one of Σ using Weyl's Theorem.

Lemma D.2. *Given $\sin^2(\theta) \leq \xi$, where θ refers to the angle between v_1 and u we have*

$$\begin{aligned} \tilde{\lambda}_i &\geq \lambda_{i-1} - \Delta \\ \tilde{\lambda}_i &\leq \lambda_{i-1} + \Delta \end{aligned}$$

where $\Delta = 8\lambda_1 \sqrt{\xi}(1 + \sqrt{\xi})$

Proof. We will use Weyl's Theorem to proof this, by defining

$$G_1 = (\mathbf{I} - v_1 v_1^\top)\Sigma(\mathbf{I} - v_1 v_1^\top)$$
$$G_2 = (\mathbf{I} - uu^\top)\Sigma(\mathbf{I} - uu^\top)$$

then by our previous definitions we know $\lambda_2 = \mu_1, \lambda_3 = \mu_2, \dots$ and $\tilde{\lambda}_1 = \nu_1, \tilde{\lambda}_2 = \nu_2, \dots$. Now we can use this as follows:

$$\begin{aligned}\tilde{\lambda}_i &= \lambda_{i-1} + (\tilde{\lambda}_i - \lambda_{i-1}) \\ &\leq \lambda_{i-1} + |\tilde{\lambda}_i - \lambda_{i-1}| \\ &\leq \lambda_{i-1} + \|G_1 - G_2\|\end{aligned}$$

where the last inequality follows by Weyl's Theorem. Next we will bound $\|G_1 - G_2\|$

$$\begin{aligned}\|G_1 - G_2\| &= \|(v_1 v_1^\top \Sigma - uu^\top \Sigma) + (\Sigma v_1 v_1^\top - \Sigma uu^\top) + (uu^\top \Sigma uu^\top - v_1 v_1^\top \Sigma v_1 v_1^\top)\| \\ &= 4\|v_1 v_1^\top - uu^\top\|_2 \|\Sigma\|_2\end{aligned}$$

where the last step follows as $(uu^\top \Sigma uu^\top - v_1 v_1^\top \Sigma v_1 v_1^\top = (uu^\top - v_1 v_1^\top)\Sigma uu^\top + v_1 v_1^\top \Sigma(uu^\top - v_1 v_1^\top$ and $\|v_1 v_1^\top\|_2 = \|uu^\top\|_2 = 1$. Further it turns out that we can bound $\|v_1 v_1^\top - uu^\top\|_2$ using $\sin^2(v_1, u) \leq \xi$: First we note that as u and v_1 are unit vectors we can write

$$u = \cos\theta v_1 + \sin\theta v_1^\perp$$

so this means

$$uu^\top = \cos^2\theta v_1 v_1^\top + \cos\theta(v_1 v_1^{\perp\top} + v_1^\perp v_1^\top) + \sin^2\theta v_1^\perp v_1^{\perp\top}$$

and also gives us

$$\begin{aligned}\|uu^\top - v_1 v_1^\top\|_2 &= \|(\cos^2\theta - 1)v_1 v_1^\top + \cos\theta\sin\theta(v_1 v_1^{\perp\top} + v_1^\perp v_1^\top) + \sin^2\theta v_1^\perp v_1^{\perp\top}\|_2 \\ &= \| - \sin^2\theta v_1 v_1^\top + \cos\theta(v_1 v_1^{\perp\top} + v_1^\perp v_1^\top) + \sin^2\theta v_1^\perp v_1^{\perp\top}\|_2 \\ &\leq |\sin^2\theta|\|v_1 v_1^\top\| + |\cos\theta\sin\theta|\|v_1 v_1^{\perp\top} + v_1^\perp v_1^\top\|_2 + |\sin^2\theta|\|v_1^\perp v_1^{\perp\top}\|_2 \\ &\leq 2|\sin^2\theta| + 2|\sin\theta| \leq 2\sqrt{\xi}(1 + \sqrt{\xi})\end{aligned}$$

□

so all in all this tells us

Lemma D.3. *For $\Sigma \in \mathbb{R}^{d\times d}$ a matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, $P = I - uu^\top$, with $u \in$ Im(Σ), and $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_{d-1}$ the eigenvalues of $P\Sigma P$*

$$\tilde{\lambda}_1 - \tilde{\lambda}_2 \geq \lambda_2 - \lambda_3 - 2\Delta$$

where $\Delta = 8\lambda_1\sqrt{\xi}(1 + \sqrt{\xi})$ and $\xi \geq \sin^2(\theta)$ with θ the angle between u and v_1, the top eigenvector of Σ.

Together with Theorem D.4 this give us a utility guarantee independent of eigenvalues of $P\Sigma P$ for $k \leq 2$. As in the first step of our recursive algorithm $P = \mathbf{I}$ and in the second we will pass $P = I - u_1 u_1^\top$.

Theorem D.4. *Utility for $k = 1$ (Main Theorem in [Liu et al., 2022]). For $\varepsilon \in (0, 0.9)$, Algorithm 3 guarantees (ε, δ)-DP for all S, B, ζ, and δ. Given n i.i.d. samples $\{A_i \in \mathbb{R}^{d\times d}\}_{i=1}^n$ and $P = \mathbf{I}$ satisfying Assumption 1 with parameters $(\Sigma, M, V, K, \kappa, a, \gamma^2)$, if*

$$n = \tilde{O}\left(e^{\kappa^2} + \frac{d^{1/2}(\log(1/\delta))^{3/2}}{\varepsilon} + \kappa M + \kappa^2 V + \frac{d\kappa\gamma(\log(1/\delta))^{1/2}}{\varepsilon}\right) \tag{9}$$

with a large enough constant and $\delta \leq 1/n$, then there exists a positive universal constant c_1 and a choice of learning rate η_t that depends on $(t, M, V, K, a, \lambda_1, \lambda_1 - \lambda_2, n, d, \varepsilon, \delta)$ such that $T = \lfloor n/B \rfloor$ steps in Algorithm 3 with choices of $\tau = 0.01$ and $B = c_1 n/(\log n)^2$, outputs ω_T such that with probability 0.99,

$$\sin(w_T, v_1) = \tilde{O}\left(\kappa(\sqrt{\frac{V}{n}} + \frac{\gamma d\sqrt{\log(1/\delta)}}{\varepsilon n}\right) \tag{10}$$

where $\tilde{O}(\cdot)$ hides poly-logarithmic factors in n, d, $1/\varepsilon$, and $\log(1/\delta)$ and polynomial factors in K.

We can see this, because the utility bound of DP-PCA depends on several constants originating form constraints on the data:

1. $\kappa = \frac{\lambda_1}{\lambda_1 - \lambda_2}$
2. M so that $\|A_i - \Sigma\|_2 \leq \lambda_1 M$ almost surely
3. V so that $\max\{\|\mathbb{E}[(A_i - \Sigma)(A_i - \Sigma)^\top]\|_2, \|, \|\mathbb{E}[(A_i - \Sigma)^\top (A_i - \Sigma)]\|_2\} \leq \lambda_1^2 V$
4. $\gamma^2 := \max_{\|u\|=1} \|H_u\|_2$
5. K so that $\max_{\|u\|=1, \|v\|=1} \mathbb{E}\left[\exp\left((\frac{|u^\top (A_i^\top - \Sigma)v|^2}{K^2 \lambda_1^2 \|H_u\|_2})^{1/(2a)}\right)\right] \leq 1$

now if we replace the $\{A_i\}$ with $\{PA_iP\}$ where P is a projection matrix, the constants $M, V, \lambda_1^2\gamma^2$ and K will still remain upper bounds (see Lemma A.9, Lemma A.10, Lemma A.12). Therefore, if we just swapped κ to be $\lambda_1(P\Sigma P)/(\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P))$ in the bound below it would still qualify as a utility bound for $\{PA_iP\}$ as input to our modified DP-PCA algorithm

$$\xi = \kappa B_n \tag{11}$$

where

$$B_n = \tilde{O}\left(\sqrt{\frac{V}{n}} + \frac{\gamma d \sqrt{\log(1/\delta)}}{\varepsilon n}\right)$$

so for $k = 2$ Lemma D.3 will give us a utility bound independent of P. However, we want to obtain a utility guarantee for arbitrary $k < d$. From now on we will denote

$$\kappa_i := \frac{\lambda_1(P_{i-1}\Sigma P_{i-1})}{\lambda_1(P_{i-1}\Sigma P_{i-1}) - \lambda_2(P_{i-1}\Sigma P_{i-1})}$$
$$\xi_i := \kappa_i \cdot B_n \text{ (upper bound ont the utility of the vector returned at step i)}$$

and the goal is to upper bound κ_i with something independent of P. If we iteratively applying Lemma D.3 we get

$$\kappa_i \leq \frac{\lambda_i(\Sigma) + \sum_{j=1}^{i-1} \Delta_j}{\lambda_i(\Sigma) - \lambda_{i+1}(\Sigma) - 2\sum_{j=1}^{i-1} \Delta_j}$$

where $\Delta_j = c\lambda_1(P_{j-1}\Sigma P_{j-1})\xi_j$ ($\Delta_0 := 0$ for completeness). Now the problem is that Δ_j still depends on previous projections and it's not even clear in general if $\xi_j > \xi_{j+1}$ or the other way around. Ultimately we want to have an upper bound for all ξ_j, to get a utility bound for $U = \{u_i\}$. A natural approach is to try and choose n big enough so that

$$\lambda_1(P_i\Sigma P_i) \leq \lambda_1$$
$$\lambda_1(P_i\Sigma P_i) - \lambda_2(P_i\Sigma P_i) \geq \delta$$

for some $\delta > 0$ then we are done. As this will guarantee that

$$\xi_i \leq \frac{\lambda_1}{\delta} B_n$$

which scales with λ_1 which guarantees that in the spiked covariance model the noise will vanish for $\sigma \to 0$.

Lemma D.5. *If for k fixed, $0 < \Delta = \min_{i \in [k]} \lambda_i - \lambda_{i+1}$ and a $0 < \delta < \Delta$ and given $C > 1$, we are given $\{A_i\}_{i=1}^n$ fulfilling Assumption A and n is big enough so that*

$$B_{n/k} \leq \frac{(\Delta - \delta)\delta}{Ck\lambda_1^2}$$

then the utility ξ_i of the vector u_i returned at step $i \in [k]$ of Algorithm 2 fulfills

$$\xi_i \leq \frac{\lambda_1}{\delta} B_{n/k}$$

Proof. We will proof that at every step:

$$\lambda_1(P_i\Sigma P_i) \leq \lambda_1 \tag{12}$$
$$\lambda_1(P_i\Sigma P_i) - \lambda_2(P_i\Sigma P_i) \geq \delta \tag{13}$$

is fulfilled, which directly implies what we are trying to proof. We will proof these two statements by induction. For $k = 1$ we have $P_0 = \mathbf{I}$ which straightaway gives us equation 12. And as δ is smaller then the minium eigengap equation 13, directly follows as well. For $k+1$ we start with showing equation 12. By Lemma D.3

$$\lambda_1(P_k \Sigma P_k) \leq \lambda_{k+1}(\Sigma) + \sum_{j=1}^{k} \Delta_j$$

first let's upper bound $\sum_{j=1}^{k} \Delta_j$ by induction assumption:

$$\begin{aligned}\sum_{j=1}^{k} \Delta_j &= \sum_{j=1}^{k} c \frac{\lambda_1^2(P_{j-1}\Sigma P_{j-1})}{\lambda_1(P_{j-1}\Sigma P_{j-1}) - \lambda_2(P_{j-1}\Sigma P_{j-1})} \cdot B_n \\ &\leq c B_{n/k} \cdot \sum_{j=1}^{k} \frac{\lambda_1^2}{\delta}\end{aligned}$$

so equation 12 will be implied by

$$B_{n/k} \leq (\lambda_1 - \lambda_{k+1}) \cdot \frac{\delta}{ck\lambda_1^2}$$

which is surely fulfilled as by assumption

$$B_{n/k} \leq \frac{(\Delta - \delta)\delta}{ck\lambda_1^2}$$

To show equation 13, we see

$$\begin{aligned}\lambda_1(P_k \Sigma P_k) - \lambda_2(P_k \Sigma P_k) &\geq \lambda_{k+1}(\Sigma) - \lambda_{k+2}(\Sigma) - 2\sum_{j=1}^{k} \Delta_j \\ &\geq \Delta - 2\sum_{j=1}^{k} \Delta_j\end{aligned}$$

where the first inequality follows by Lemma D.3 and the second by definition of Δ. Using the upper bound on $\sum_{j=1}^{k} \Delta_j$ we established

$$B_{n/k} \leq \frac{(\Delta - \delta)\delta}{ck\lambda_1^2}$$

will imply equation 13. □

We will now combine all of this to proof our main theorem:

Proof of Theorem ?? By Theorem D.1 we know that when passing $m = n/k$ A_i at every step of our deflation method we obtain a vector u_i fulfilling

$$\sin(u_i, v_i) \leq \tilde{O}\left(\frac{\lambda_1(P\Sigma P)}{\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)}\left(\sqrt{\frac{Vk}{n}} + \frac{\gamma dk\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$$

where v_i is the top eigenvector of $P_{i-1}\Sigma P_{i-1}$. Which by Lemma B.1 give us

$$\langle u_i u_i^\top, P_{i-1}\Sigma P_{i-1}\rangle \geq (1 - \zeta_i^2)\langle v_i v_i^\top, P_{i-1}\Sigma P_{i-1}\rangle$$

with $\zeta_i = \tilde{O}\left(\frac{\lambda_1(P\Sigma P)}{\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)}\left(\sqrt{\frac{Vk}{n}} + \frac{\gamma dk\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$. By our choice of n we know by Lemma D.5 that

$$\zeta_i \leq \tilde{O}\left(\frac{\lambda_1}{\Delta}\left(\sqrt{\frac{Vk}{n}} + \frac{\gamma dk\sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$$

where we used that $(\Delta - \delta)\delta$ is maximized by $\delta = \Delta/2$. So finally Theorem B.2 gives us that

$$\langle UU^\top, \Sigma\rangle \geq (1 - \zeta^2)\langle V_k V_k^\top, \Sigma\rangle \tag{14}$$

where V_k is the matrix obtained by non private k-PCA.

E Algorithms used in Modified DP-PCA

Below we describe the two subroutines that estimate the eigenvalue and mean of the gradients.

Algorithm 5 Top-Eigenvalue-Estimation, Algorithm 4 in [Liu et al., 2022]

Input: $S = \{g_i\}_{=1}^{B}$, privacy parameters (ε, δ), failure probability $\tau \in (0, 1)$

$\tilde{g}_i \leftarrow g_{2i} - g_{2i-1}$ for $i \in 1, 2, \ldots, \lfloor B/2 \rfloor$

$\tilde{S} = \{\tilde{g}_i\}_{=1}^{\lfloor B/2 \rfloor}$

Partition $\tilde{S}$ into $k = C_1 \log(1/(\delta\tau)/\varepsilon$ subsets and denote each dataset as $G_j \in G_j \in \mathbb{R}^{d \times b}$ (where $b = \lfloor B/2k \rfloor$ is the size of the dataset)

$\lambda_1^{(j)} \leftarrow$ top eigenvalue of $(1/b)G_j G_j^\top$ for all $j \in [k]$

$\Omega \leftarrow \{\ldots, [2^{-2/4}, 2^{-1/4}), [1, 2^{1/4}), \ldots\}$

run (ε, δ)-DP histogram learner on Ω

if all bins are empty **then**

  return $\perp$

else

  for $[l, r]$ the bn with the maximum number of points

  return $\hat{\Lambda} = l$

end if

Algorithm 6 Private-Mean-Estimation, Algorithm 5 in [Liu et al., 2022]

Input: $S = \{g_i\}_{=1}^{B}$, privacy parameters (ε, δ), target error α, failure probability $\tau \in (0, 1)$, approximate top eigenvalue $\hat{\Lambda}$

let $\tau = 2^{1/4} K \sqrt{\hat{\Lambda}} \log^2(25)$

for $j = 1, 2, \ldots, d$ **do**

  Run $(\frac{\varepsilon}{4\sqrt{2d\log(4/\delta)}}, \frac{\delta}{4d})$-DP histogram learner of Lemma on $\{g_{ij}\}_{i \in [B]}$

  Let $[l, h]$ be the bucket that contains maximum number of points in the private histogram

  $\bar{g}_j \leftarrow l$

  Truncate the j-th coordinate of gradient $\{g_i\}_{i \in [B]}$ by $[\bar{g}_j - 3K\sqrt{\hat{\Lambda}}\log^a(BD/\tau), \bar{g}_j + 3K\sqrt{\hat{\Lambda}}\log^a(BD/\tau)]$.

  Let $\tilde{g}_i$ be the truncated version of g_i

end for

Compute empirical mean of truncated gradients $\tilde{\mu} = (1/B)\sum_{i=1}^{B} \tilde{g}_i$ and add Gaussian noise:

$$\hat{\mu} = \tilde{\mu} + \mathcal{N}\left(0, \left(\frac{12K\sqrt{\hat{\Lambda}}\log^a(BD/\tau)\sqrt{2d\log(2.5/\delta)}}{\varepsilon B}\right)^2 \mathbf{I}_d\right)$$

return $\hat{\mu}$