

---

# An Iterative Algorithm for Differentially Private $k$ -PCA with Adaptive Noise

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1        Given  $n$  i.i.d. random matrices  $A_i \in \mathbb{R}^{d \times d}$  that share common expectation  $\Sigma$ , the  
2        objective of *Differentially Private Stochastic PCA* is to identify a subspace of  
3        dimension  $k$  that captures the largest variance directions of  $\Sigma$ , while preserving  
4        differential privacy (DP) of each individual  $A_i$ . Existing methods either (i) require  
5        the sample size  $n$  to scale super-linearly with dimension  $d$ , even under Gaussian  
6        assumptions on the  $A_i$ , or (ii) introduce excessive noise for DP even when the  
7        intrinsic randomness within  $A_i$  is small. Liu et al. [2022a] addressed these issues  
8        for sub-Gaussian data but only for estimating the top eigenvector ( $k = 1$ ) using  
9        their algorithm DP-PCA. We propose the first algorithm capable of estimating  
10       the top  $k$  eigenvectors for arbitrary  $k \leq d$ , whilst overcoming both limitations  
11       above. For  $k = 1$ , our algorithm matches the utility guarantees of DP-PCA,  
12       achieving near-optimal statistical error even when  $n = \tilde{O}(d)$ . We further provide  
13       a lower bound for general  $k > 1$ , matching our upper bound up to a factor of  $k$ ,  
14       and experimentally demonstrate the advantages of our algorithm over comparable  
15       baselines.

## 16    1 Introduction

17    Principal Component Analysis (PCA) is a foundational statistical method widely utilized for dimen-  
18    sionality reduction, data visualisation, and noise filtering. Given  $n$  data points  $\{x_i\}_{i=1}^n$ , classical  
19    PCA computes the top eigenvectors of the empirical covariance matrix  $X := \sum_{i=1}^n x_i x_i^\top \in \mathbb{R}^{d \times d}$ .  
20    This problem of extracting the top  $k$  eigenvectors is commonly known as  $k$ -PCA. In this work,  
21    we consider the problem of Stochastic  $k$ -PCA, which differs from the standard setting as follows:  
22    instead of inputting a single matrix, we input a stream of matrices  $A_1, \dots, A_n$ , that are sampled  
23    independently from distributions that share the same expectation  $\Sigma$ . Given this input, the goal of a  
24    Stochastic  $k$ -PCA algorithm is to approximate the dominant  $k$  eigenvectors of  $\Sigma$ .

25    Differential privacy (DP) [Dwork et al., 2006] provides rigorous, quantifiable guarantees of individual  
26    data privacy and has been widely adopted in sensitive data contexts, such as census reporting [Abowd  
27    et al., 2020] and large-scale commercial analytics [Team et al., 2017]. Despite extensive study of  
28    differentially private PCA [Blum et al., 2005, Chaudhuri et al., 2013, Hardt and Roth, 2013, Dwork  
29    et al., 2014b], existing methods in the stochastic setting suffer from sample complexity super-linear  
30    in  $d$  or inject noise at a scale that ignores the underlying stochasticity in the data. When applied to the  
31    stochastic setting, these works generally yield suboptimal error rates of  $O(\sqrt{dk/n} + d^{3/2}k/(\varepsilon n))$   
32    where  $\varepsilon$  is the DP parameter.

33    **Example 1** (Spiked Covariance). *In the spiked covariance model, we observe i.i.d. matrices*  
34     *$A_i \in \mathbb{R}^{d \times d}$  that contain both a deterministic (low-rank) signal and random noise, causing the*  
35     *$A_i$  to be full-rank. As a concrete illustration, consider data points  $x_i = s_i + n_i$ , composed of a signal*  
36     *$s_i \sim \text{Unif}(\{v, -v\})$  with  $v$  a unit vector and  $n_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ . Therefore  $A_i := x_i x_i^\top$  consists of a*

37 *deterministic part  $vv^\top$  and noise terms that scale with  $\sigma^2$ . One would hope that the privacy noise*  
 38 *that is needed, shrinks as the noise variance  $\sigma^2$  decreases. Instead, most differentially private PCA*  
 39 *methods employ non-adaptive clipping thresholds, so their added privacy noise scales only with that*  
 40 *threshold, resulting in unnecessarily large privacy noise for many distributions.*

41 Recent advances by Liu et al. [2022a] address these limitations for sub-Gaussian distributions, but only  
 42 for the top eigenvector case ( $k = 1$ ). Cai et al. [2024] achieve optimal performance specifically for the  
 43  $k$ -dimensional spiked covariance model, yet their privacy guarantees only apply under distributional  
 44 assumptions on the data.

45 **Our Contributions.** *In this work, we propose  $k$ -DP-PCA, the first DP algorithm for stochastic*  
 46 *PCA that simultaneously (1) achieves sample complexity  $n = \tilde{O}(d)$  under similar assumptions*  
 47 *as Liu et al. [2021], (2) adapts its privacy noise to the data’s inherent randomness, (3) generalizes*  
 48 *seamlessly to any target dimension  $k \leq d$ , and (4) is simple to implement.*

49 For  $k = 1$ ,  $k$ -DP-PCA matches the risk of Liu et al. [2022a] under sub-Gaussian assumptions. For  
 50 general  $k$ , we prove a nearly matching lower bound up to a linear factor in  $k$ , precisely characterising  
 51 the cost of privacy in this general setting. Technically, we employ the *deflation* framework: iteratively  
 52 estimate the top eigenvector, project it out, and repeat. We extend the recent deflation analysis of  
 53 Jambulapati et al. [2024] to the stochastic setting via a novel *stochastic  $e$ -PCA oracle* (Definition 5),  
 54 which may be of independent interest. We then adapt DP subroutines from Liu et al. [2022a]  
 55 based on Oja’s algorithm and finally, through a novel utility analysis of non-private Oja’s algorithm,  
 56 demonstrate that the adapted subroutines satisfy the oracle’s requirements, yielding a simple to  
 57 implement, memory-efficient method.

58 The remainder of this paper is structured as follows. We formally define our setting in Section 2,  
 59 state main results in Section 3, present technical analyses in Section 4 and empirical evaluations  
 60 demonstrating the effectiveness of our approach in Section 5. Finally, we end with a discussion and  
 61 open questions in Section 6 and conclusion in Section 7.

## 62 2 Problem formulation

63 Let  $A_1, \dots, A_n \in \mathbb{R}^{d \times d}$  be independent random matrices with common expectation  $\Sigma = \mathbb{E}[A_i]$ . We  
 64 assume  $\Sigma$  is symmetric positive semi-definite (PSD) with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$ .  
 65 For a given  $k < d$ , we assume the eigengap  $\Delta_k = \lambda_k - \lambda_{k+1} > 0$ . The goal of *Stochastic PCA* is  
 66 to produce a  $U \in \mathbb{R}^{d \times k}$  whose orthonormal columns approximate the top- $k$  eigenspace of  $\Sigma$ . We  
 67 measure the utility of  $U$  by comparing it to  $V_k$ , the matrix containing the true top  $k$  eigenvectors of  
 68  $\Sigma$  as columns. Throughout,  $\|\cdot\|_2$  denotes the operator norm and  $\langle \cdot, \cdot \rangle$  the Frobenius inner product:  
 69  $\langle A, B \rangle = \text{Tr}(A^\top B)$ .

70 **Definition 1** ( $\zeta$ -approximate Utility). We say  $U \in \mathbb{R}^{d \times k}$  is  $\zeta$ -approximate if  $U$  has orthonormal  
 71 columns and

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

72 Although several utility measures exist for PCA, our choice is motivated by the error measure used  
 73 in Jambulapati et al. [2024]. This is a natural measure of usefulness, as  $\langle UU^\top, \Sigma \rangle$  quantifies how  
 74 much of the original “energy” of  $\Sigma$  is retained when projecting onto the lower-dimensional subspace  
 75 spanned by  $U$ , and by the Eckart-Young Theorem we know  $V_k$  is the optimal rank- $k$  approximation  
 76 of  $\Sigma$ .

77 Further, we use the add/remove model of differential privacy, namely

**Definition 2** (Differential Privacy ([Dwork et al., 2006])). Given two multi-sets  $S$  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  $(\epsilon, \delta)$ -differential privacy for some  $\epsilon > 0$  and  $\delta \in (0, 1)$  if

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

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

79 Before discussing the main results of our work, we first formalize the assumptions on the data  
 80 in Assumption A. Note that Assumption A is only required for our utility guarantee and is not  
 81 necessary for the privacy guarantee.

82 **Assumption A**  $((\Sigma, \{\lambda_i\}_{i=1}^d, M, V, K, \kappa, a, \gamma^2)$ -model). Let  $A_1, \dots, A_n \in \mathbb{R}^{d \times d}$  be sampled inde-  
83 *pendently from distributions satisfying:*

84 **A.1**  $\mathbb{E}[A_i] = \Sigma$ , where  $\Sigma$  is PSD with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_d \geq 0$ , corresponding eigenvectors  
85  $v_1, \dots, v_d$ ,  $0 < \Delta = \min_{i \in [k]} \Delta_k$  and  $\kappa' := \frac{\lambda_1}{\Delta}$ .

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

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

88 **A.4** For all unit vectors  $u, v$  and projection matrices  $P$ ,

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

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

Assumptions A.1 to A.3 are standard for matrix concentration (e.g., under the matrix Bernstein inequality [Tropp, 2012]) and thus also required for the utility guarantees of Oja’s algorithm even in the non-private setting. Assumption A.4 guarantees that for any unit vectors  $u, v$ , and projection  $P$

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

90 with probability  $1 - \vartheta$ , for some sufficiently large constant  $K$ . This bound, which controls the size  
91 of the bilinear form, can be seen as a Gaussian-like tail bound, which tells us that the magnitude of  
92 the projection of the  $A_i$  along any direction is bounded with high probability. It is an extension of  
93 the assumptions in [Liu et al., 2022a] to the higher dimensional case. Distributions that fulfill this  
94 assumption include bounded matrices and (sub-)gaussian outer product matrices:

95 **Example 2** (Gaussian Data, Remark 3.4 in Liu et al. [2022a]). Let  $A_i = x_i x_i^\top$  with  $x_i \sim \mathcal{N}(0, \Sigma)$ ,  
96 then comparing to Assumption A we have that  $M = O(d \log(n))$ ,  $V = O(d)$ ,  $K = 4$ ,  $a = 1$ , and  
97  $\gamma^2 = O(1)$

98 Distributions that violate assumption 4 include heavy-tailed outer products, for example  $r \sim$   
99  $\text{Pareto}(\alpha)$ ,  $x = ru$ ,  $A_i = xx^\top$ , or mixtures with rare but huge spikes:

100 **Example 3.** Let  $A_i = x_i x_i^\top$ , with  $x_i$  be sampled as follows:

$$x_i = \begin{cases} x \sim \mathcal{N}(0, \mathbf{I}_d) & \text{w.p. } 1 - \alpha \\ x \sim \text{Unif}\{\alpha^{-1/4}v, -\alpha^{-1/4}v\} & \text{w.p. } \alpha \end{cases}$$

101 where  $v$  is a unit vector and  $0 < \alpha < 1$ . Then the mean of this distribution is 0 and its covariance is  
102  $\Sigma = (1 - \alpha)\mathbf{I}_d + \sqrt{\alpha}vv^\top$ . So for  $u = v$  and  $P = \mathbb{I}_d$ , if  $x = \pm\alpha^{-1/4}v$

$$\begin{aligned} u^\top (A_i - \Sigma)u &= v^\top (x_i x_i^\top - \Sigma)v = (v^\top x_i)^2 - v^\top \Sigma v \\ &= \alpha^{-1/2} - v^\top \Sigma v = \alpha^{-1/2} - (1 - \alpha) + \sqrt{\alpha} \simeq \alpha^{-1/2} \end{aligned}$$

103 and for  $\alpha \rightarrow 0$  this term blows up, so for any fixed  $K, \lambda_1, \gamma$  the overall expectation will exceed 1, and  
104 hence violate Assumption A.4.

### 105 3 Main Results

106 In this section, we first discuss our main proposed algorithm in Section 3.1. In Section 3.2 we then  
107 discuss our main upper bounds and complement that with lower bounds in Section 3.3

#### 108 3.1 Our Algorithm

109 Our first proposed algorithm k-DP-PCA, defined in Algorithm 1, follows a classical deflation [Jam-  
110 bulapati et al., 2024] approach. The algorithm proceeds in  $k$  rounds and in each of the  $k$  rounds it  
111 invokes the sub-routine MODIFIEDDP-PCA (Line 3), to identify the current top eigenvector. Then,  
112 the algorithm removes its contribution by projecting out the direction of the eigenvector from the  
113 remaining data (Line 4), on which it carries out the next round.

---

**Algorithm 1**  $k$ -DP-PCA

---

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

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

---

114 The MODIFIEDDP-PCA subroutine (Algorithm 2) itself is based on Oja’s streaming Algorithm  
115 [Jain et al., 2016], but importantly replaces the vanilla gradient update in Oja’s algorithm  $\omega_T \leftarrow$   
116  $\omega_{t-1} + \eta_t A_{t-1} \omega_{t-1}$ , with a two-stage algorithm: first, Line 3 privately estimates the range of  
117 a batch of  $\{A_i \omega_{t-1}\}$ , then Line 4 leverages that range to calibrate the added noise to privately  
118 compute the batch’s mean. By tailoring the noise scale to the empirical spread of the data, we inject  
119 significantly less (privacy) noise whenever the batch concentrates tightly around its mean. Thanks to  
120 those additional steps the algorithm enjoys certain statistical benefits as discussed in the paragraph  
121 below Corollary 2.

122 Nevertheless, it is possible to replace the MODIFIEDDP-PCA subroutine with other simpler subrou-  
123 tines that can privately estimate the top eigenvector. We present one such algorithm in Algorithm 3.  
124 In Section 5, we present simulations with both of these algorithms highlighting their respective  
125 advantages.

---

**Algorithm 2** ModifiedDP-PCA

---

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

- 1: Choose  $\omega'_0$  uniformly at random from the unit sphere,  $\omega_0 \leftarrow P\omega'_0 / \|P\omega'_0\|$
- 2: **for**  $t = 1, 2, \dots, T = \lfloor m/B \rfloor$  **do**
- 3:      $\hat{\Lambda} \leftarrow \text{PRIVRANGE}(\{PA_{B(t-1)+i}P\omega_{t-1}\}_{i=1}^{\lfloor B/2 \rfloor}, (\varepsilon/2, \delta/2), \tau/(2T))$  (Algorithm 6)
- 4:      $\hat{g}_t \leftarrow \text{PRIVMEAN}(\{PA_{B(t-1)+i}P\omega_{t-1}\}_{i=1}^{\lfloor B/2 \rfloor}, \hat{\Lambda}, (\varepsilon/2, \delta/2), \tau/(2T))$  (Algorithm 7)
- 5:      $\omega'_t \leftarrow \omega_{t-1} + \eta_t P \hat{g}_t$
- 6:      $\omega_t \leftarrow P\omega'_t / \|P\omega'_t\|$
- 7: **end for**
- 8: **return**  $\omega_T$

---

### 126 3.2 Upper Bound

127 We now state the main privacy and utility guarantees of  $k$ -DP-PCA (Algorithm 1).

128 **Theorem 1** (Main Theorem). *Let  $\varepsilon, \delta \in (0, 0.9)$  and  $1 \leq k < d$ . Then  $k$ -DP-PCA satisfies the*  
129 *following:*

130 **Privacy:** *For any input sequence  $\{A_i \in \mathbb{R}^{d \times d}\}$ , the algorithm is  $(\varepsilon, \delta)$ -differentially private.*

131 **Utility:** *Suppose  $A_1, \dots, A_n$  are i.i.d. satisfying Assumption A with parameters*  
132  *$(\Sigma, M, V, K, \kappa', a, \gamma^2)$ . If*

$$n \geq C \max \begin{cases} e^{\kappa'^2} + \frac{d \kappa' \gamma \sqrt{\ln(1/\delta)}}{\varepsilon} + \kappa' M + \kappa'^2 V + \frac{\sqrt{d} (\ln(1/\delta))^{3/2}}{\varepsilon}, \\ \lambda_1^2 \kappa'^2 k^3 V, \\ \frac{\kappa'^2 \gamma k^2 d \sqrt{\ln(1/\delta)}}{\varepsilon} \end{cases}, \quad (1)$$

133 *for a sufficiently large constant  $C$ , then with probability at least 0.99, the output  $U \in \mathbb{R}^{d \times k}$  is*  
134  *$\zeta$ -approximate with*

$$\zeta = \tilde{O} \left( \kappa' \left( \sqrt{\frac{Vk}{n}} + \frac{\gamma dk \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right), \quad (2)$$

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

**Remark.** The proof of our main Theorem can be found in Appendix E. For  $k = 1$ , Theorem 1 recovers the bound of Liu et al. [2022a] for DP-PCA. Moreover, the linear dependence on  $d$  in  $\zeta$  matches the lower bound in Liu et al. [2022a]. On the other hand, the additional linear factor in  $k$  may be an artifact of our analysis: if one could reuse samples across deflation steps, this factor could potentially be improved. Further, in  $\zeta$ , the first term  $\sqrt{Vk/n}$  is the non-private statistical error of PCA, while the second term  $(\gamma dk \sqrt{\ln(1/\delta)})/(\varepsilon n)$  is the cost of privacy. Lastly, the sample-size condition (1) arises because (i) each batch must be large enough to accurately estimate the range in PRIVRANGE in Algorithm 2, and (ii) errors accumulate across the  $k$  deflation steps (Line 4).

As a direct consequence of applying Theorem 1 to Examples 1 and 2, we obtain the following Corollaries:

**Corollary 1** (Upper bound, Gaussian distribution). *Under the same setting as Theorem 1, let  $A_i = x_i x_i^\top$  with  $x_i \sim \mathcal{N}(0, \Sigma)$ . Then with high probability the output is  $\zeta$ -approximate with*

$$\zeta = \tilde{O} \left( \kappa' \left( \sqrt{\frac{dk}{n}} + \frac{dk \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right)$$

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

**Corollary 2** (Upper bound, Spiked Covariance). *If  $A_i$  follows the spiked covariance model from Example 1, then  $V = O(\sigma^2 d)$ ,  $\gamma^2 = \sigma^2$ , and  $K = 1$ . Hence, with high probability the output is  $\zeta$ -approximate with*

$$\zeta = \tilde{O} \left( \sigma \cdot \kappa' \left( \sqrt{\frac{dk}{n}} + \frac{dk \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right) \quad (3)$$

**Adaptive noise:** Our algorithm's advantage is most pronounced when  $\gamma$  and  $V$  grow with the data randomness, as in Corollary 2. Since for  $\zeta = \tilde{O}(\sigma \kappa'(\sqrt{dk/n} + (dk \sqrt{\ln(1/\delta)})/(\varepsilon n)))$ , the approximation error decreases as the noise standard deviation  $\sigma$  shrinks. Moreover, by comparison with Corollary 5, this bound is tight up to a factor of  $k$ .

### 3.3 Lower Bounds

In this section, we derive an information-theoretic lower bound for differentially private PCA under our setting. Formal proofs can be found in Appendix F.1. Recall that our utility metric  $\zeta$  defined in Definition 1 measures the *relative* loss in captured variance compared to the optimal top- $k$  subspace of  $\Sigma$ . By contrast, most classical lower bounds for PCA (e.g., Cai et al. [2024], Liu et al. [2022a]) quantify error in terms of the squared Frobenius norm  $\|\tilde{U}\tilde{U}^\top - V_k V_k^\top\|_F^2$ . These two measures are fundamentally different: the ratio of captured variance directly reflects variance explained in  $\Sigma$ , whereas the Frobenius-norm loss measures subspace distance without respecting the eigenvalue gaps in  $\Sigma$ . To connect them, we first establish:

**Lemma 1** (Reduction to Frobenius norm). *Let  $\Sigma$  be a PSD  $d \times d$  matrix with top- $k$  eigenvectors  $V_k \in \mathbb{R}^{d \times k}$  and eigenvalues  $\lambda_1 \geq \dots \geq \lambda_d$ . Any  $U \in \mathbb{R}^{d \times k}$  that satisfies  $\|UU^\top - V_k V_k^\top\|_F^2 \geq \gamma$ , must incur*

$$\zeta^2 \geq \frac{\gamma \Delta_k}{2 \sum_{i=1}^k \lambda_i}$$

where  $\Delta_k := \lambda_k - \lambda_{k+1}$ .

Note that if all eigenvalues of  $\Sigma$  are equal, every subspace captures the same variance so  $\zeta = 0$  for any estimate, yet two such subspaces can be far apart in Frobenius norm. This gap in sensitivity to eigengaps is precisely why our reduction from Frobenius error to  $\zeta$  incurs a factor of  $\Delta_k$ . With this reduction in hand, we prove the spiked-covariance lower bound by invoking standard Frobenius-norm minimax rates [Cai et al., 2024] for differentially private PCA in the spiked covariance model.

174 **Corollary 3** (Lower bound, Spiked Covariance). *Let the  $d \times n$  data matrix  $X$  have i.i.d. columns sam-*  
 175 *ples from a distribution  $P = \mathcal{N}(0, U^\top \Lambda U^\top + \sigma^2 \mathbf{I}_d) \in \mathcal{P}(\lambda, \sigma^2)$  where  $\mathcal{P}(\lambda, \sigma^2) = \{\mathcal{N}(0, \Sigma), \Sigma =$*   
 176  *$U \Lambda U^\top + \sigma^2 \mathbf{I}_d, c\lambda \leq \lambda_k \leq \dots \leq \lambda_1 \leq C\lambda\}$ . Suppose  $\lambda \leq c'_0 \exp\{e\varepsilon - c_0(\varepsilon\sqrt{ndk} + dk)\}$  for*  
 177 *some small constants  $c_0, c'_0 > 0$ . Then, there exists an absolute constant  $c_1 > 0$  such that*

$$\inf_{\tilde{U} \in \mathcal{U}_{\varepsilon, \delta}} \sup_{P \in \mathcal{P}(\lambda, \sigma^2)} \mathbb{E}[\zeta] \geq c_1 \left( \left( \frac{\sigma\sqrt{\lambda_1 + \sigma^2}}{\sum_{i=1}^k (\lambda_i + \sigma^2)} \right) \left( \sqrt{\frac{dk}{n}} + \frac{dk}{n\varepsilon} \right) \wedge 1 \right).$$

178 Comparing to our upper bound (Corollary 2), we see matching dependence on  $\sigma$ ,  $d$ ,  $n$ , and  $\varepsilon$ , up  
 179 to a multiplicative factor of  $k$ ,  $\sqrt{\lambda_1 + \sigma^2}$ , and  $\sqrt{\log(1/\delta)}$ . The gap in  $k$  arises from our sequential  
 180 deflation approach, which currently requires independent batches at each step. Reusing samples  
 181 across rounds could remove this up to a  $\sqrt{k}$  factor<sup>1</sup>.

182 **Special case  $k = 1$ .** When  $k = 1$ , k-DP-PCA reduces exactly to MODIFIEDDP-PCA. Theorem 10  
 183 guarantees that the sine of the angle between the privately estimated eigenvector of MODIFIEDDP-  
 184 PCA and the true top eigenvector is small, which is equivalent to being close in the Frobenius norm.  
 185 This matches the upper bound of Liu et al. [2022a] and thus also the lower bound up to a factor of  
 186  $\log(1/\delta)$  (restated in Theorem 12 in the Appendix).

## 187 4 Technical Results

188 We now sketch the proof of Theorem 1 by first proving a more general “meta-theorem” that applies  
 189 to any *stochastic ePCA oracle* (defined below in Definition 5). At a high level, k-DP-PCA uses  
 190 the classical deflation strategy: 1. Extract the top eigenvector of the current residual using a 1-PCA  
 191 subroutine. 2. Project this vector out of the data. 3. Repeat until  $k$  components are obtained.  
 192 In Theorem 1 we implement the 1-PCA step with MODIFIEDDP-PCA, but the same proof carries  
 193 through for any algorithm satisfying the following guarantee.

194 **Definition 3** (stochastic ePCA oracle). An algorithm  $O_{\text{ePCA}}$  is a  $\zeta$ -approximate 1-ePCA oracle if  
 195 the following holds. On independent inputs  $A_1, \dots, A_n \in \mathbb{R}^{d \times d}$  with  $\mathbb{E}[A_i] = \Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$  for all  $i$   
 196 and any orthogonal projector  $P \in \mathbb{R}^{d \times d}$ ,  $O_{\text{ePCA}}$  returns a unit vector  $u \in \text{Im}(P)$  such that, with  
 197 high probability,

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

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

199 This notion was inspired by Jambulapati et al. [2024], who analyzed deflation in the non-stochastic  
 200 setting. Their results do not extend the stochastic setting that we explore here.

201 **Theorem 2** (Meta Theorem). *Let  $\Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$  and  $A_1, \dots, A_n$  be  $n$  i.i.d. samples with  $\mathbb{E}[A_i] = \Sigma$ .*  
 202 *Suppose we replace each 1-PCA step in Line 3 of Algorithm 1 by a  $\zeta$ -approximate stochastic ePCA*  
 203 *oracle  $O_{\text{1PCA}}$ . Then the deflation algorithm outputs  $U \in \mathbb{R}^{d \times k}$  satisfying*

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

204 *Further, for any  $\varepsilon > 0$ ,  $\delta \in (0, 1)$ , if  $O_{\text{1PCA}}$  is  $\varepsilon, \delta$ -DP then the entire algorithm remains  $(\varepsilon, \delta)$ -DP.*

205 **Remark.** This Theorem is a consequence of the stochastic deflation method we prove in Appendix C  
 206 and Parallel Composition (Lemma 16).

207 One important thing we would like to highlight in this section is that this proof strategy is not unique  
 208 to MODIFIEDDP-PCA. In fact, our novel analysis of non-private Oja’s algorithm (Theorem 8) shows  
 209 that Algorithm 3 is also a stochastic ePCA oracle. We highlight the two results below.

210 **Theorem 3.** *Given  $A_1, \dots, A_n$  are i.i.d. and satisfy Assumption A, MODIFIEDDP-PCA and k-DP-*  
 211 *Ojas as defined Algorithms 2 and 3 are stochastic ePCA with  $\zeta = \tilde{O} \left( \kappa' \left( \sqrt{\frac{V}{n}} + \frac{\gamma d \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right)$*

212 *and  $\zeta = \tilde{O} \left( \kappa' \left( \sqrt{\frac{V}{n}} + \frac{(\gamma+1)d \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right)$  respectively.*

<sup>1</sup>Reusing will allow us to use all  $n$  samples every round (instead of  $n/k$ ), however we will incur an additional  
 $\sqrt{k}$  factor due to privacy composition, which is why it will only lead to a total improvement of  $\sqrt{k}$  and not  $k$ .

---

**Algorithm 3** DP-Ojas

---

**Input:**  $\{A_1, \dots, A_m\}$ , a projection  $P$ , privacy parameters  $(\varepsilon, \delta)$ , learning rates  $\{\eta_t\}_{t=1}^{\lfloor m \rfloor}$

```

1: Set DP noise multiplier:  $\alpha \leftarrow C' \log(n/\delta)/(\varepsilon\sqrt{n})$ 
2: Set clipping threshold:  $\beta \leftarrow C\lambda_1\sqrt{d}(K\gamma\log^a(nd/\zeta) + 1)$ 
3: Choose  $\omega'_0$  uniformly at random from the unit sphere,  $\omega_0 \leftarrow P\omega'_0/\|P\omega'_0\|$ 
4: for  $t = 1, 2, \dots, m$  do
5:   Sample  $z_t \sim \mathcal{N}(0, \mathbf{I}_d)$ 
6:    $\omega'_t \leftarrow \omega_{t-1} + \eta_t P (\text{clip}_\beta(PA_t P\omega_{t-1}) + 2\beta\alpha z_t)$ 
7:    $\omega_t \leftarrow P\omega'_t/\|P\omega'_t\|$ 
8: end for
9: return  $\omega_T$ 

```

where  $\text{clip}_\beta(x) = x \cdot \min\{1, \frac{\beta}{\|x\|_2}\}$

---

213 *Remark.* In Appendix E, we establish that both MODIFIEDDP-PCA and k-DP-Ojas are valid ePCA  
214 oracles, with each result stated and proved as a separate theorem.

215 Note that we cannot plug in the DP-PCA algorithm of Liu et al. [2022a] in Theorem 2, since it only  
216 guarantees relative error on  $\mathbb{E}[P]\Sigma\mathbb{E}[P]$ :

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

217 rather than on  $P\Sigma P$ , and  $\mathbb{E}[P]$  need not be a projection matrix.

218 The proof of Theorem 9 follows directly from the utility proof of MODIFIEDDP-PCA (Theorem 10)  
219 and of DP-Ojas (Theorem 11). Combining this with Theorem 2 immediately gives us Theorem 1 and  
220 the following Corollary 4.

221 To prove the utility of MODIFIEDDP-PCA we proceed in three steps: 1. Prove non-private Oja's  
222 algorithm is a stochastic ePCA oracle via a Novel analysis in Appendix D 2. Show that with high  
223 probability, the update step (Line 5 in Algorithm 2) can be reduced to an update step of non-private  
224 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  
225 matrix. 3. Bound the accumulated projection error across deflation steps (Lemma 27). Importantly, a  
226 similar argument also shows that DP-Ojas Algorithm 3 satisfies the same property with a slightly  
227 differently  $\zeta$ .

228 **Corollary 4** (k-DP-Ojas). *Under Assumption A, if  $n$  is sufficiently large then using Algorithm 3 in*  
229 *each 1-PCA step returns  $U \in \mathbb{R}^{d \times k}$  that is  $\zeta$ -approximate with*

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

230 *hiding poly-logarithmic factors in  $n, d, 1/\varepsilon, \ln(1/\delta)$  and polynomial factors in  $K$ .*

231 *Remark.* This Corollary follows directly from Theorem 2 together with Theorem 9.

232 When comparing the utility bounds of MODIFIEDDP-PCA and k-DP-Ojas the difference is partic-  
233 ularly apparent when considering Example 1, as for k-DP-Ojas when  $\sigma \rightarrow 0$  the bound becomes  
234  $\tilde{O} \left( \frac{dk \log(1/\delta)}{\varepsilon n} \right)$ , as due to the second term of the utility bound containing the multiplicative factor of  
235  $(\gamma + 1)$  (as opposed  $\gamma$  as in MODIFIEDDP-PCA) it does not vanish. Therefore in the low-noise cases  
236 MODIFIEDDP-PCA will outperform k-DP-Ojas. However, for other cases such as (sub-)Gaussian  
237 data we expect them to perform similarly. In those cases it can be preferential to use k-DP-Ojas as  
238 due to its simplicity it requires less hyperparameters to be set and is more stable to changes in learning  
239 rates.

## 240 5 Experiments

241 In our experiments, we compare k-DP-PCA and k-DP-Ojas against two modified versions of the  
242 DP-Gauss algorithms of Dwork et al. [2014b]. Their work operates in a deterministic setting with

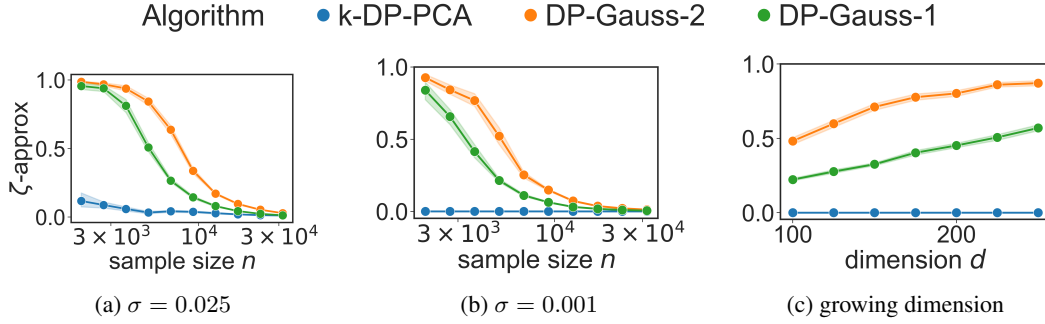


Figure 1: Comparison of k-DP-PCA vs DP-Gauss-1 (input perturbation) and DP-Gauss-2 (output perturbation) on the spiked covariance model. We plot the mean over 50 trials, with shaded regions representing 95% confidence intervals. We set  $k = 2$ ,  $d = 200$ ,  $\lambda_1 = 10$ ,  $\varepsilon = 1$ , and  $\delta = 0.01$ .

each row of the data matrix  $X \in \mathbb{R}^{n \times d}$  bounded in  $\ell_2$ -norm by 1 and the estimate the top eigenvectors of  $X^\top X$ . By contrast, our setting is stochastic: we draw independent matrices  $A_i$  without any norm constraint and we estimate the top eigenvectors of  $\mathbb{E}[A_i] = \Sigma$ . Thus, we first adapt their algorithm to also work in the stochastic setting. Note that if we draw observations  $x_i$  from a distribution with mean zero and covariance  $\Sigma$ , then  $X^\top X = \sum_{i=1}^n x_i x_i^\top$  serves as an unbiased estimate of  $n\Sigma$ . A naive way to enforce the bounded norm requirement of Dwork et al. [2014b], is to define  $\tilde{x}_i = x_i / \max\{\|x_i\|_2\}$ . However, this non-private pre-processing step [Hu et al., 2024] will violate privacy: modifying a single  $x_i$  can potentially change the maximum norm and thus affect all of the  $\tilde{x}_i$ . A natural next attempt is to scale each vector exactly to unit norm, i.e.,  $\tilde{x}_i = x_i / \|x_i\|_2$ . However, this will result in a biased estimator as  $\mathbb{E}[x x^\top / \|x\|^2] \neq \Sigma$  and thus does not enjoy meaningful utility guarantees. Instead, we clip each  $x_i$  at  $\beta$  so that with probability at least  $1 - \vartheta$ ,  $\|x_i\|_2 \leq \beta$ . Then scaling the Gaussian noise in the DP-Gauss mechanisms by  $\beta$  maintains  $(\varepsilon, \delta)$ -DP guarantee. For the spiked covariance model this would mean  $\beta = C\sqrt{\lambda_1} + \sigma\sqrt{d \log(n/\vartheta)}$ . Using this strategy we modify Algorithm 1 and 2 in Dwork et al. [2014b] and refer to them as DP-Gauss-1 and DP-Gauss-2 respectively. DP-Gauss-1 first clips each  $x_i$ , adds appropriately scaled Gaussian noise to the sum  $\sum_i \tilde{x}_i \tilde{x}_i^\top$ , and then performs standard (non-private) PCA. DP-Gauss-2, on the other hand, begins by privately estimating the eigengap of the clipped covariance matrix, runs non-private PCA on the clipped data, and finally perturbs the resulting top- $k$  eigenvectors with noise that scales with that that privately computed eigengap. In the rest of this section, Figure 1 compares k-DP-PCA with DP-Gauss-1 and DP-Gauss-2 across various noise levels  $\sigma$  and dimensions  $d$ . Figure 2 also incorporates the much simpler-to-implement k-DP-Ojas algorithm and shows that a simpler, more scalable algorithm can match or even outperform k-DP-PCA in practice, despite its slightly weaker theoretical guarantee.

**Experimental Results using Spiked Covariance Data** We evaluate all methods on the spiked-covariance model (see Example 1). Figures 1a and 1b show utility as a function of sample size for large and small noise levels, respectively. Our results show that across both regimes, k-DP-PCA consistently outperforms the two DP-Gauss baselines, with the gap widening when the noise level is significantly smaller than the signal strength ( $\sigma \ll \lambda_1$ ). Figure 1c examines the effect of increasing ambient dimension  $d$  at fixed  $n$ . As  $d$  grows, the DP-Gauss methods' utility degrades faster than k-DP-PCA's, reflecting the fact that their theoretical utility scales like  $O(d^{3/2}/n)$ , whereas our guarantee only incurs a linear dependence on  $d$ .

In Figure 2a, we plot the utility against the eigengap ( $\lambda_k - \lambda_{k+1}$ ) for different algorithms. DP-Gauss-2, which is designed with large eigen-gaps in mind steadily improves in utility as the gap grows and nearly matches the utility of k-DP-PCA for very large eigengap. By contrast, DP-Gauss-1 which offers better scalability with dimension  $d$  but is insensitive to the eigen-gap, maintains a nearly flat utility as the eigen-gap grows. Throughout, k-DP-PCA consistently outperforms both DP-Gauss algorithms.

Next, in Figure 2, we compare k-DP-PCA against the much simpler k-DP-Ojas algorithm. As predicted by Corollaries 2 and 4, k-DP-PCA clearly outperforms k-DP-Ojas in the low-noise regime ( $\sigma \ll \lambda_1$ ). Conversely, at larger noise levels k-DP-Ojas often matches or even exceeds k-DP-PCA in practice, owing to its fewer hyperparameters and greater robustness to learning-rate choices (see Figure 2b and appendix G). Although both algorithms require knowledge of the eigenvalues of  $\Sigma$  to set optimal step sizes, these can be obtained privately via the Gaussian mechanism. Nevertheless, it is interesting to note that k-DP-Ojas remains effective even when its step size is chosen without any explicit eigenvalue estimates (see Appendix G).



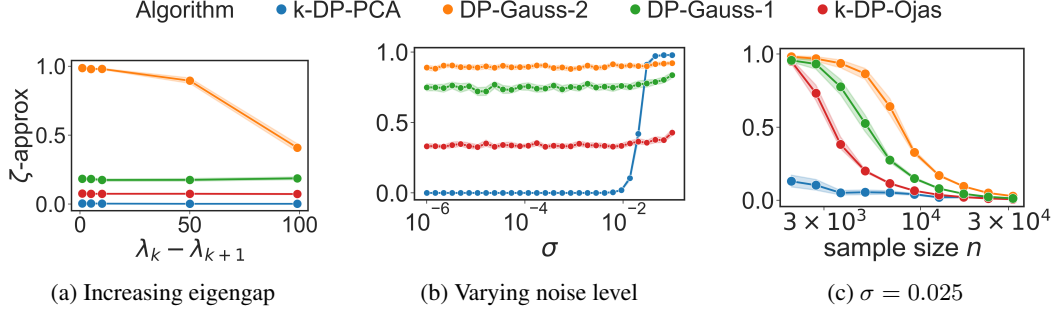


Figure 2: Comparison of k-DP-PCA and k-DP-Ojas in a higher noise regime (also including DP-Gauss-1 (input perturbation) and DP-Gauss-2 (object perturbation)) on the spiked covariance model. We plot the mean over 50 trials, with shaded regions representing 95% confidence intervals. We set  $k = 2$ ,  $d = 200$ ,  $\lambda_1 = 10$ ,  $\varepsilon = 1$ , and  $\delta = 0.01$ .

## 6 Related Work and Open problems

**Related Work** Differentially private PCA has been studied extensively [Blum et al., 2005, Chaudhuri et al., 2013, Hardt and Roth, 2013, Dwork et al., 2014b]. However, when applied to the stochastic setting, these methods typically suffer from sample complexity that scales super-linearly in  $d$  or inject noise at a scale that ignores the underlying stochasticity in the data, resulting in suboptimal error rates of  $O(\sqrt{dk/n} + d^{3/2}k/(\varepsilon n))$ . The first to address these limitations were [Liu et al., 2022b, Cai et al., 2024]; however the results by [Liu et al., 2022b] only apply for  $k = 1$  and Cai et al. [2024] provide an algorithm whose privacy guarantee is conditional on distributional assumptions on the data. In contrast, our algorithm applies to all  $k \leq d$ , is private for all inputs, provides an error rate that scales linearly with  $d$ , and the injected noise scales with the inherent stochasticity in the data.

A complimentary line of work, [Singhal and Steinke, 2021, Tsfadia, 2024] obtains sample complexity that scales independently of the dimension  $d$  but requires a strong multiplicative eigengap  $(\lambda_k/\lambda_{k+1}) = O(\sqrt{d})$ , which is a strictly stronger assumption than ours.

**Open Problems** Despite being a mild concentration requirement also seen in prior work [Liu et al., 2022a], Assumption A.4 is perhaps the most non-standard assumption in Assumption A. As observed by Liu et al. [2022a], this can be relaxed to a bounded  $k$ -th moment condition, at which point the second term in (14) grows to  $O(d(\log(1/\delta)/\varepsilon n)^{1-1/k})$ . Further, empirical improvements may also be possible from applying private robust mean estimation [Liu et al., 2021, Hopkins et al., 2022], as opposed to clipping around the mean of the gradients. We leave these to future work.

The sample size condition in Equation (1) includes an exponential dependence on the spectral gap:  $n \geq \exp(\kappa')$ . While this is relatively harmless as there is no such exponential dependence in the utility guarantee Equation (2), we show in Appendix E.2 how to get rid of this exponential dependence by incurring an additional  $\tilde{O}(\gamma d^2 \log(1/\delta)/(\varepsilon n))$  term in the utility guarantee.

As already mentioned in Section 3.3, our upper bounds are loose in their dependence in  $k$  and  $\delta$ . We incur this additional  $k$  factor, because each deflation step must use a fresh batch of samples, so that the projection matrices  $P$  remain independent of the data matrices in Line 4 of Algorithm 1. If one could safely reuse the same  $A_i$ 's across rounds, this could potentially be improved to  $O(\sqrt{k})$ . Finally, although inspired by the streaming analysis of Oja's method [Jain et al., 2016, Huang et al., 2021], our subroutines (MODIFIEDDP-PCA, PRIVRANGE, PRIVMEAN) are not directly streaming-compatible. Adapting them to the online setting is an interesting avenue for future work.

## 7 Conclusion

We have presented the first algorithm for stochastic  $k$ -PCA that is both differentially private and computationally efficient, supports any  $k \leq d$ , and achieves near-optimal error. Our analysis critically relies on our adaptation of the DP-PCA algorithm [Liu et al., 2022a], a stochastic deflation framework inspired by [Jambulapati et al., 2024], and our novel analysis of non-private Oja's algorithm [Jain et al., 2016]. Along with our novel results in the *Stochastic k-PCA* problem, we believe the above mentioned theoretical results are of independent interest, and may inspire the development of new algorithms for this and related problems.

## References

- John M Abowd, Gary L Benedetto, Simson L Garfinkel, Scot A Dahl, Aref N Dajani, Matthew Graham, Michael B Hawes, Vishesh Karwa, Daniel Kifer, Hang Kim, et al. The modernization of statistical disclosure limitation at the us census bureau. *URL: bit.ly/DPcensus20*, 2020.
- Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.
- Zeyuan Allen-Zhu and Yuanzhi Li. Lazysvd: Even faster svd decomposition yet without agonizing pain. *Advances in neural information processing systems*, 29, 2016.
- Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pages 464–473. IEEE, 2014.
- Raef Bassily, Vitaly Feldman, Kunal Talwar, and Abhradeep Guha Thakurta. Private stochastic convex optimization with optimal rates. *Advances in neural information processing systems*, 32, 2019.
- Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. *Advances in Neural Information Processing Systems*, 33:14475–14485, 2020.
- 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.
- Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. *Advances in Neural Information Processing Systems*, 32, 2019.
- T Tony Cai, Dong Xia, and Mengyue Zha. Optimal differentially private pca and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*, 2024.
- 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.
- 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.
- Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014a.
- 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, 2014b.
- Vitaly Feldman and Thomas Steinke. Calibrating noise to variance in adaptive data analysis. In *Conference On Learning Theory*, pages 535–544. PMLR, 2018.
- Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: optimal rates in linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 439–449, 2020.
- 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.
- Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1406–1417, 2022.
- Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

373 Lijie Hu, Shuo Ni, Hanshen Xiao, and Di Wang. High dimensional differentially private stochastic  
 374 optimization with heavy-tailed data. In *Proceedings of the 41st ACM SIGMOD-SIGACT-SIGAI*  
 375 *Symposium on Principles of Database Systems*, pages 227–236, 2022.

376 Yaxi Hu, Amartya Sanyal, and Bernhard Schölkopf. Provable privacy with non-private pre-processing.  
 377 In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Pro-*  
 378 *ceedings of Machine Learning Research*, pages 19402–19437. PMLR, 2024.

379 De Huang, Jonathan Niles-Weed, and Rachel Ward. Streaming k-pca: Efficient guarantees for oja’s  
 380 algorithm, beyond rank-one updates. In *Conference on Learning Theory*, pages 2463–2498. PMLR,  
 381 2021.

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

385 Arun Jambulapati, Syamantak Kumar, Jerry Li, Shourya Pandey, Ankit Pensia, and Kevin Tian.  
 386 Black-box  $k$ -to-1-pca reductions: Theory and applications. *arXiv preprint arXiv:2403.03905*,  
 387 2024.

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

392 Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional  
 393 distributions. In *Conference on Learning Theory*, pages 1853–1902. PMLR, 2019.

394 Gautam Kamath, Xingtu Liu, and Huanyu Zhang. Improved rates for differentially private stochastic  
 395 convex optimization with heavy-tailed data. In *International Conference on Machine Learning*,  
 396 pages 10633–10660. PMLR, 2022.

397 Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *Proceedings*  
 398 *of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1395–1414.  
 399 SIAM, 2013.

400 Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv*  
 401 *preprint arXiv:1711.03908*, 2017.

402 Pravesh Kothari, Pasin Manurangsi, and Ameya Velingker. Private robust estimation by stabilizing  
 403 convex relaxations. In *Conference on Learning Theory*, pages 723–777. PMLR, 2022.

404 Janardhan Kulkarni, Yin Tat Lee, and Daogao Liu. Private non-smooth empirical risk minimization  
 405 and stochastic convex optimization in subquadratic steps. *arXiv preprint arXiv:2103.15352*, 2021.

406 Xiyang Liu, Weihao Kong, Sham Kakade, and Sewoong Oh. Robust and differentially private mean  
 407 estimation. *Advances in neural information processing systems*, 34:3887–3901, 2021.

408 Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. Dp-pca: Statistically optimal and  
 409 differentially private pca. *Advances in neural information processing systems*, 35:29929–29943,  
 410 2022a.

411 Xiyang Liu, Weihao Kong, and Sewoong Oh. Differential privacy and robust statistics in high  
 412 dimensions. In *Conference on Learning Theory*, pages 1167–1246. PMLR, 2022b.

413 Lester Mackey. Deflation methods for sparse pca. *Advances in neural information processing systems*,  
 414 21, 2008.

415 Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical*  
 416 *biology*, 15:267–273, 1982.

417 Alain Pajor. Metric entropy of the grassmann manifold. *Convex Geometric Analysis*, 34(181-188):  
 418 0942–46013, 1998.

419 Vikrant Singhal and Thomas Steinke. Privately learning subspaces. *Advances in neural information*  
420 *processing systems*, 34:1312–1324, 2021.

421 Apple Differential Privacy Team et al. Learning with privacy at scale. *Apple Mach. Learn. J*, 1(8):  
422 1–25, 2017.

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

425 Eliad Tsfadia. On differentially private subspace estimation in a distribution-free setting. *arXiv*  
426 *preprint arXiv:2402.06465*, 2024.

427 Christos Tzamos, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Ilias Zadik. Optimal private  
428 median estimation under minimal distributional assumptions. *Advances in Neural Information*  
429 *Processing Systems*, 33:3301–3311, 2020.

430 Di Wang, Hanshen Xiao, Srinivas Devadas, and Jinhui Xu. On differentially private stochastic convex  
431 optimization with heavy-tailed data. In *International Conference on Machine Learning*, pages  
432 10081–10091. PMLR, 2020.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: Yes. Our main contributions are also detailed in Section 3 and Appendix E contains the relevant mathematical proofs.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Yes, see Section 6 for limitations. We also comment on the limitations of the different algorithms in Section 5.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Yes, please see Appendix E for a detailed proof of the Main Theorem, and Appendix C, Appendix D for the more general novel results we developed in order to proof the Main Theorem. Lastly in Appendix F we proof the lower bound.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We describe our algorithm in Detail in Section 3 and state all the hyperparameters used for the plots in Appendix G.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We will release the code publically after we have cleaned it.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: A detailed discussion can be found in Appendix G

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We ran a minimum of 50 trials for each experiment and included the variance of results in the plots.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: All experiments were run locally on a MacBook M3 Pro.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We followed the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

## 10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: This work is mainly a theory result. The numerical experiments were run on synthetic data and are therefore not related to any private or personal data, and there's no explicit negative social impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.



- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We do not foresee any high risk for misuse of this work.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We have not released any new assets as part of this work.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758

## 16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

## Appendix

The appendix is structured as follows. In Appendix A, we provide a more detailed overview of related work to complement the discussion in the main text. Appendix B introduces mathematical and privacy-related preliminaries that lay the groundwork for our analysis. In Appendices C and D, we present novel technical contributions: in C we extend the recent deflation analysis of Jambulapati et al. [2024] to the stochastic setting and also prove Theorem 2 in the main text, and in D we provide a new utility analysis of the non-private Oja’s algorithm. These results are then used to prove our main theorem and establish the utility and privacy guarantees for our second proposed algorithm (Corollary 4),  $k$ -DP-Ojas, in Appendix E. In Appendix F, we prove our lower bound result from Section 3.3. We then provide additional experimental details in Appendix G, and conclude by restating the subroutines from Liu et al. [2022b], which are used in MODIFIEDDP-PCA, in Appendix H.

## A Related Work

The problem of private  $k$ -PCA has been the subject of extensive research, with many works exploring it under various constraints. Several works address  $k$ -PCA in the standard setting, while assuming an additive eigengap [Blum et al., 2005, Chaudhuri et al., 2013, Hardt and Roth, 2013, Dwork et al., 2014b]. When applied to the stochastic setting, these works generally yield suboptimal error rates of

$$O(\sqrt{dk/n} + d^{3/2}k/(\epsilon n))$$

More recent work has considered the multiplicative eigengap setting [Tsfadia, 2024, Singhal and Steinke, 2021], however this is a strictly stronger assumption. Finally there are a group of results without spectral gap assumptions [Chaudhuri et al., 2013, Kapralov and Talwar, 2013, Liu et al., 2022b]. However, these works either do not provide a tractable implementation or give utility bounds that are super-linear in their dependence on  $d$ .

A widely used strategy to mitigate the complexity of designing algorithms for  $k$ -PCA is to reduce the  $k$ -dimensional problem to a series of 1-dimensional problems, using a technique known as the deflation method [Mackey, 2008, Allen-Zhu and Li, 2016]. Jambulapati et al. [2024] proved significantly sharper bounds on the approximation parameter degradation of deflation methods for  $k$ -PCA. While their analysis was conducted in the standard (non-stochastic) setting, assuming access to the true covariance matrix  $\Sigma$ , their results serve as a conceptual foundation for our own work, where we extend similar arguments to the stochastic setting, where only access to sample matrices  $A_i$  with shared expectation  $\mathbb{E}[A_i] = \Sigma$  is available.

Our 1-PCA method builds upon Oja’s algorithm [Oja, 1982], (see Algorithm 5), one of the oldest and most popular algorithms for streaming PCA. The first formal utility guarantees for Oja’s algorithm in the  $k = 1$  case were established by Jain et al. [2016], whose analysis inspired our proofs in Appendix D. Subsequent extensions to the  $k > 1$  case were provided by Huang et al. [2021].

Lastly, our ePCA oracle MODIFIEDDP-PCA is largely inspired by the DP-PCA algorithm of Liu et al. [2022b]. Their result builds upon a series of advances in private SGD [Kamath et al., 2022, Bassily et al., 2014, 2019, Feldman et al., 2020, Kulkarni et al., 2021, Wang et al., 2020, Hu et al., 2022], and private mean estimation [Bun and Steinke, 2019, Karwa and Vadhan, 2017, Kamath et al., 2019, Biswas et al., 2020, Feldman and Steinke, 2018, Tzamos et al., 2020]. In this work we use we use some of the techniques proposed by Liu et al. [2022b] specifically their PRIVMEAN and PRIVRANGE algorithms. Replacing them with robust and private mean estimation [Liu et al., 2021, Kothari et al., 2022] one could relax Assumption A.4, but at the cost of sub-optimal sample complexity.

## B Preliminaries

In this section we list some mathematical and privacy preliminaries. A familiar reader is welcome to skip this section.

### B.1 Mathematics Preliminaries

**Lemma 2.**  $C \preceq D \implies ACA^\top \preceq ADA^\top$

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

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

806 so for any  $x$

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

807 which proofs our claim.  $\square$

808 **Theorem 4** (Woodbury matrix identity). *For  $A$  an  $n \times n$ ,  $C$  a  $k \times k$ ,  $U$  and  $n \times k$  and  $V$  a  $k \times n$*   
 809 *matrix. We have*

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

810 **Theorem 5** (Pinsker's Inequality). *For  $P$  and  $Q$  two probability distributions on a measurable space*  
 811 *then*

$$TV(P, Q) \leq \sqrt{\frac{1}{2}KL(P||Q)}$$

812 **Lemma 3** (Lemma F.2 in [Liu et al., 2022a]). *Let  $G \in \mathbb{R}^{d \times d}$  be a random matrix where each entry*  
 813  *$G_{ij}$  is i.i.d. sampled from standard Gaussian  $\mathcal{N}(0, 1)$ . Then there exists a universal constant  $C > 0$*   
 814 *such that with probability  $1 - 2e^{-t^2}$*

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

815 for  $t > 0$ .

816 **Lemma 4** (Lemma F.5 in [Liu et al., 2022a]). *Under Assumption 1.-3. with probability  $1 - \tau$*

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

817 **Lemma 5** (Adapted Version of Lemma F.3 in [Liu et al., 2022a]). *lemma-version-f3-liu Let  $G \in \mathbb{R}^{d \times d}$*   
 818 *be a random matrix where each entry  $G_{ij}$  is i.i.d. sampled from standard Gaussian  $\mathcal{N}(0, 1)$ . Then*  
 819 *we have*

$$\mathbb{E}[\|GG^\top\|_2] \leq C_2 d \quad (4)$$

820 *Proof.* By Lemma 3 the exists a universal constant  $C_3 > 0$  such that

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

821 then

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

822  $\square$

823 **Lemma 6** (Weyl's inequality [Horn and Johnson, 2012]). *Let  $G_1$  and  $G_2$  be two matrices with*  
 824 *eigenvalues  $\mu_1 \geq \dots \geq \mu_d$  and  $\nu_1 \geq \dots \geq \nu_d$  respectively, then*

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

825 **Lemma 7** (Conditional Markov Inequality). *Let  $\mathcal{F}$  be a conditioning event (or a sigma-algebra), let*  
 826  *$X$  be a non negative random variable, and  $a > 0$ , then*

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

827 *Proof.* As a first step we define

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

828 then by definition of the indicator function we have

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

829 which implies

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

830 by taking conditional expectation on both sides. And finally

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

831 gives us the wished result □

832 **Lemma 8** (Conditional Chebyshev's Inequality). *Let  $\mathcal{F}$  be a conditioning event (or a sigma-algebra),*  
833 *then for  $a > 0$*

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

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

*Proof.*

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

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

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

837 □

838 **Lemma 9** (Distributional Equivalence). *Let  $z \sim \mathcal{N}(0, \Sigma)$ ,  $P$  a projection matrix, and  $\omega \in \text{Im}(P)$  a*  
839 *unit vector, then there exists a random matrix  $G$  so that*

$$Pz \stackrel{d}{=} PGP\omega$$

840 and

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

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

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

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

843 Likewise we have

$$\text{Cov}(PGP\omega) = P\text{Cov}(GP\omega)P = P\text{Cov}(G\omega)P$$

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

$$\text{Cov}(G\omega) = \Sigma$$

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

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

846 we are done. Using rotation invariance of the spherical Gaussian random vectors and the fact that  
847  $\|w\|_2 = 1$ , 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  
848  $G'w \sim \mathcal{N}(0, \mathbf{I}_d)$ , which finishes our proof. □

849 **Lemma 10.** Assume we have a matrix  $A \in \mathbb{R}^{d \times d}$  and a projection matrix  $P$  then

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

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

852 **Lemma 11.** Let  $A \in \mathbb{R}^{d \times d}$  be a random matrix and  $P$  a random projection matrix independent of  $A$   
853 then

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

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

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

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

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

856 as  $x$  was an arbitrary unit vector, this implies:

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

857 If we now take expectations on both sides we get

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

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

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

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

861 **Lemma 12.** For  $A$  and  $B$  independent random matrices

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

862 *Proof.* By independence we have  $\mathbb{E}[ABA^\top] = \mathbb{E}[A\mathbb{E}[B]A^\top]$ . Then by using  $\mathbb{E}[B] \preceq \|\mathbb{E}[B]\|_2 \mathbf{I}_d$   
863 and Lemma 2 we obtain the wished inequality.  $\square$

864 **Lemma 13.** 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

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

865 then

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

866 where  $\gamma$  and  $\lambda_1$  are defined as in Assumption A

*Proof.*

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

867 and further

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

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

869 **Definition 4.**  $\mathbb{O}_{d,k}$  denotes the set of  $d \times k$  matrices satisfying  $U^\top U = \mathbf{I}_k$ .

870 **Lemma 14** (Proposition 8, Pajor [1998]). *For any  $q \in [1, \infty]$ , there exists an absolute constant*  
871  *$c' > 0$  and a subset  $\mathcal{S}_q^{(d-k)} \subset \mathbb{O}_{d-k,k}$  s.t. for any  $V_i \neq V_j \in \mathcal{S}_q^{(d-k)}$ ,*

$$\|V_i V_i^\top - V_j V_j^\top\| \geq c' r^{1/q}$$

872 *, where  $\|\cdot\|_q$  denotes the Schatten- $q$  norm and the cardinality of  $\mathcal{S}_q^{(d-k)}$  is at least  $2^{k(d-k)}$*

873 *Remark.* The Frobenius norm is equal to the Schatten-2 norm.

874 **Lemma 15** (Lemma 3 in [Jambulapati et al., 2024]). *Let  $\Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$ ,  $k \in [d]$ . If  $P \in \mathbb{R}^{d \times d}$  is a*  
875 *rank- $(d-k)$  orthogonal projection matrix, then  $\|P\Sigma P\|_{op} \geq \lambda_{k+1}(\Sigma)$ .*

## 876 B.2 Differential Privacy Preliminaries

877 **Lemma 16** (Parallel composition, [Dwork et al., 2014a]). *Consider a sequence of interactive queries*  
878  *$\{q_k\}_{k=1}^K$  each operating on a subset  $S_k$  of the database and each satisfying  $(\varepsilon, \delta)$ -DP. If  $S_k$ 's are*  
879 *disjoint then the composition  $(q_1(S_1), q_2(S_2), \dots, q_K(S_K))$  is  $(\varepsilon, \delta)$ -DP.*

880 **Lemma 17** (Advanced Composition, [Kairouz et al., 2015]). *For  $\varepsilon \leq 0.9$ , an end-to-end guar-*  
881 *antee of  $(\varepsilon, \delta)$ -differential privacy is satisfied if a database is accessed  $k$  times, each with a*  
882  *$(\varepsilon/(2\sqrt{2k \log(2/\delta)}), \delta/(2k))$ -differential private mechanism.*

883 **Lemma 18** (DP-constrained Fano's Lemma, [Acharya et al., 2021]). *Let  $\mathcal{P} := \{P : P = \mu^{(1)} \times \dots \times$*   
884  *$\mu^{(n)}\}$  be a family of product measures indexed by a parameter from a pseudo-metric space  $(\Theta, \rho)$ .*  
885 *Denote  $\theta(P) \in \Theta$  the parameter associated with the distribution  $P$ . Let  $\mathcal{Q} = \{P_1, \dots, P_N\} \subset \mathcal{P}$*   
886 *contain  $N$  probability measures and there exist constants  $\rho_0, l_0, t_0 > 0$  such that for all  $i \neq i' \in [N]$ ,*

$$\rho(\theta(P_i), \theta(P_{i'})) \geq \rho_0, \quad KL(P_i \| P_{i'}) \leq l_0,$$

887 *and*

$$\sum_{k \in [n]} TV(\mu_i^{(k)}, \mu_{i'}^{(k)}) \leq t_0,$$

888 *where  $P_i = \mu_i^{(1)} \times \dots \times \mu_i^{(n)}$  and  $P_{i'} = \mu_{i'}^{(1)} \times \dots \times \mu_{i'}^{(n)}$ . Then,*

$$\inf_{A \in \mathcal{A}_{\varepsilon, \delta}(\mathcal{P})} \sup_{P \in \mathcal{P}} \mathbb{E}_A \rho(A, \theta(P)) \geq \max \left\{ \frac{\rho_0}{2} \left( 1 - \frac{l_0 + \log 2}{\log N} \right), \frac{\rho_0}{4} \left( 1 \wedge \frac{N-1}{\exp(4\varepsilon t_0)} \right) \left( 1 - \frac{2\delta e^{4\varepsilon t_0}}{e^\varepsilon - 1} \right) \right\} \quad (5)$$

889 *where the infimum is taken over all  $(\varepsilon, \delta)$ -DP randomized algorithm defined by  $\mathcal{A}_{\varepsilon, \delta}(\mathcal{P}) := \{A :$   
890  *$X \rightarrow \Theta$  and  $A$  is  $(\varepsilon, \delta)$ -differentially private for all  $X \sim P \in \mathcal{P}\}$ .**

## 891 C Meta Algorithm for stochastic $k$ -PCA

892 In this section we prove that any stochastic ePCA oracle when passed to Appendix C will give us a  
893  $k$ -PCA algorithm. This is the basis for Theorem 2 as it holds for all randomized stochastic ePCA  
894 oracles and is not specific to privacy. We obtain this result by extending the work of Jambulapati et al.  
895 [2024] to the stochastic setting. We obtain the same utility results as them even when approximating  
896 the top eigenvector of the expectation of a stream of matrices.

897 **Definition 5** (stochastic ePCA oracle). An algorithm  $O_{\text{ePCA}}$  is a  $\zeta$ -approximate 1-ePCA oracle if  
898 the following holds. On independent inputs  $A_1, \dots, A_n \in \mathbb{R}^{d \times d}$  with  $\mathbb{E}[A_i] = \Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$  for all  $i$   
899 and any orthogonal projector  $P \in \mathbb{R}^{d \times d}$ ,  $O_{\text{ePCA}}$  returns a unit vector  $u \in \text{Im}(P)$  such that, with  
900 high probability,

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

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

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

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

903 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:**  $\{A_1, \dots, A_n\}$  i.i.d matrices sampled from a distribution with expectation  $\mathbb{E}[A_i] = \Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$ ,  $k \in [d]$ ,  $O_{1PCA}$  an algorithm which takes as input matrices  $A_1, \dots, A_n$  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
     $u_i \leftarrow O_{1PCA}(A_{B*(i-1)+1}, \dots, A_{B*i}, P_{i-1})$ 
     $P_i \leftarrow P_{i-1} - u_i u_i^\top$ 
end for
return  $U \leftarrow \{u_i\}_{i \in [k]}$ 

```

---

904 We will now show that for this type of approximation algorithm we can obtain a utility guarantee and  
 905 that it would be optimal for the spiked covariance setting. Jambulapati et al. [2024] define two types  
 906 of approximation notions for PCA. Our type of utility bound is equivalent to their first notion:

907 **Definition 6** (energy  $k$ -PCA, [Jambulapati et al., 2024]).  $U \in \mathbb{R}^{d \times k}$  is an  $\zeta$ -approximation energy  
 908  $k$ -PCA of  $M \in \mathbb{S}_{\geq 0}^{d \times d}$  if

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

909 where

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

910 and

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

911 is the Frobenius inner product.

912 **Lemma 19.** For  $v, w \in \mathbb{R}^d$  unit vectors,  $\theta$  the angle between the two, and  $\Sigma$  a psd matrix with  $v$  it's  
 913 top eigenvector we have

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

*Proof.*

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

914 Now as  $v$  is the top eigenvector of  $\Sigma$  we know

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

915 where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  denote the eigenvalues of  $\Sigma$  and  $v, v_2, \dots, v_d$  the corresponding  
 916 eigenvectors. Therefore

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

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

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

---

**Algorithm 5** OjasAlgorithm( $\{A_i\}_{i=1}^n$ )

---

Choose  $\omega_0$  uniformly at random from the unit sphere

**for**  $t = 1, \dots, n$  **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**  $w_n$

---

918 and by definition

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

919 so we have

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

920 which in turn means

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

921

□

922 **Theorem 6** ( $k$ -to-1-ePCA reduction). *Let  $A_1, \dots, A_n$  be i.i.d. matrices with expectation  $\Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$ ,  $\varepsilon \in (0, 1)$ , and  $O_{1PCA}$  a stochastic ePCA oracle. Then, Algorithm 4 returns  $U \in \mathbb{R}^{d \times k}$  so that  $\Sigma$ .*

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

924 *Proof.* We will proof this by induction, where the  $k = 1$  case follows from Lemma 19. For  $i + 1$  we  
 925 note  $P_i = \mathbb{I}_d - U_i U_i^\top$  then

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

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

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

927 as  $u_{i+1}$  can be seen as the approximation the oracle returns for the top eigenvalue of  $P_i \Sigma P_i$  and  
 928 therefore it must fulfill this equality by assumption on our oracle. By Lemma 3 in [Jambulapati et al.,  
 929 2024] (restated in Lemma 15) we know

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

930 and this in turn gives us

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

931 which proofs our Claim. □

932 **Theorem 2** (Meta Theorem). *Let  $\Sigma \in \mathbb{S}_{\geq 0}^{d \times d}$  and  $A_1, \dots, A_n$  be  $n$  i.i.d. samples with  $\mathbb{E}[A_i] = \Sigma$ .  
 933 Suppose we replace each 1-PCA step in Line 3 of Algorithm 1 by a  $\zeta$ -approximate stochastic ePCA  
 934 oracle  $O_{1PCA}$ . Then the deflation algorithm outputs  $U \in \mathbb{R}^{d \times k}$  satisfying*

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

935 *Further, for any  $\varepsilon > 0, \delta \in (0, 1)$ , if  $O_{1PCA}$  is  $\varepsilon, \delta$ -DP then the entire algorithm remains  $(\varepsilon, \delta)$ -DP.*

936 *Proof of theorem 2.* The proof follows directly by Theorem 6 and Lemma 16. □

## D Novel Analysis of non private Oja's Algorithm

Given i.i.d. random matrices  $A_1, \dots, A_n$  with  $\mathbb{E}[A_i] = \Sigma$ , let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  denote the eigenvalues of  $\Sigma$ , and let  $v_1, \dots, v_d$  be the corresponding eigenvectors. Let  $P$  be a projection matrix that is independent of the  $A_i$ . Our goal is to compute an  $\varepsilon$ -approximation to the top eigenvector of the projected matrix  $P\Sigma P$ .

When  $P$  is deterministic, the analysis of Jain et al. [2016] shows that Oja's algorithm produces an output vector that is close to the top eigenvector of  $P\Sigma P$ . However, in our setting, the projection matrix is itself *random*, defined as  $P = I - \sum_i u_i u_i^\top$ , where the  $u_i$  are random vectors estimated from an earlier, independent sample of the  $A_i$ . This dependence introduces a key difficulty: applying the result of Jain et al. [2016] directly would only guarantee closeness to the top eigenvector of  $\mathbb{E}[P]\Sigma\mathbb{E}[P]$ , but  $\mathbb{E}[P]$  is generally not a projection matrix and may not preserve the spectral structure of interest.

To address this, we provide a new analysis of Oja's algorithm tailored to inputs of the form  $\{PA_iP\}$ , where  $P$  is a random projection matrix. Our main result shows that, under suitable conditions, the algorithm still yields an accurate approximation to the top eigenvector of  $P\Sigma P$ , even when  $P$  is random and data-dependent.

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

We 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

$$B_n := (\mathbf{I} + \eta_n P A_n P)(\mathbf{I} + \eta_{n-1} P A_{n-1} P) \cdots (\mathbf{I} + \eta_1 P A_1 P) \quad (6)$$

$$w_n := \frac{B_n w_0}{\|B_n w_0\|_2} \quad (7)$$

$$\bar{\mathcal{V}} := \mathcal{V} + \tilde{\lambda}_1^2 \quad (8)$$

and note that  $w_n$  is the result of Oja's Algorithm after  $n$  steps given  $\{PA_iP\}$  as input. The proof of Theorem 8 will require the following result we will proof below

**Theorem 7.** *Given  $A_1, \dots, A_n$  that fulfill Assumption A.1-A.3 with parameters  $\Sigma, M, V, \kappa$ , a projection matrix  $P$  independent of the  $A_i$ ,  $\tilde{v}$  the top eigenvector of  $P\Sigma P$ , and  $B_n$  described as in Equation (6), the output  $\omega_n$  resulting from non-private Oja's Algorithm on inputs  $PA_1P, \dots, PA_nP$  fulfills*

$$\sin\left(\tilde{v}, \frac{B_n \omega_n}{\|B_n \omega_n\|_2}\right) \leq \frac{1}{Q} \exp \sum_{j \in [t]} \eta_j^2 5\bar{\mathcal{V}} \left( d \exp(-2(\tilde{\lambda}_1 - \tilde{\lambda}_2) \sum_{j \in [t]} \eta_j) \right) \quad (9)$$

where  $Q = \frac{\delta}{C \log(1/\delta)} \left(1 - \frac{1}{\sqrt{\delta}} \sqrt{\exp(\sum_{i \in [n]} 18\eta_i^2 \bar{\mathcal{V}}) - 1}\right)$

**Theorem 8.** (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  $\omega_n$  of Algorithm 5 satisfies:

$$1 - (\omega_n^\top \tilde{v})^2 \leq \frac{C \log(1/\delta)}{\delta^2} \left( d \left(\frac{\beta}{n}\right)^{2\alpha} + \frac{\alpha^2 \mathcal{V}}{(2\alpha - 1) \cdot (\tilde{\lambda}_1 - \tilde{\lambda}_2)^2} \cdot \frac{1}{n} \right), \quad (10)$$

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

971 *Proof.* Analogously to the proof of Theorem 4.1 in [Jain et al., 2016] by replacing their Theorem 3.1  
 972 with our Theorem 7.  $\square$

973 We state and proof several Lemmas that will allow us to proof Theorem 7 which in turn will directly  
 974 proof Theorem 8.

975 **Lemma 20** (One Step Power Method [Jain et al., 2016]). . Let  $B \in d \times d$ , let  $v \in d$  be a unit vector,  
 976 and let  $V_\perp$  be a matrix whose columns form an orthonormal basis of the subspace orthogonal to  $v$ . If  
 977  $w \in \mathbb{R}^d$  is chosen uniformly at random from the surface of the unit sphere then with probability at  
 978 least  $1 - \delta$

$$\sin^2(v, \frac{Bw}{\|Bw\|_2}) = 1 - (v^\top Bw)^2 \leq C \frac{\log(1/\delta)}{\delta} \frac{\text{Tr}(V_\perp^\top B B^\top V_\perp)}{v^\top B B^\top v} \quad (11)$$

979 where  $C$  is an absolute constant.

980 Based on the above Lemma we see that to show Oja's algorithm succeeds we simply need to show that  
 981 with high probability  $\text{Tr}(\tilde{V}_\perp^\top B_n B_n^\top \tilde{V}_\perp)$  is relatively large and  $\tilde{v}^\top B_n B_n^\top \tilde{v}$  is relatively small. Note  
 982 so long as we pick  $\eta_i$  sufficiently small, i.e.  $\eta_i = O(1/\max M, \tilde{\lambda}_1)$  then  $\mathbf{I} + \eta_i P A_i P$  is invertible,  
 983 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  
 984 finite. In order to explicitly bound the RHS we will utilize conditional Chebychev's and Markov's,  
 985 where the conditioning will serve to fix  $P$ .

986 **Lemma 21.**  $\|\mathbb{E}[B_t B_t^\top | P]\|_2 \leq \exp(\sum_{i \in [t]} 2\eta_i \tilde{\lambda}_1 + \eta_i^2 (\tilde{\lambda}_1^2 + \mathcal{V}))$

987 *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  
 988 Lemma 12

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

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

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

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

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

992 As  $\alpha_0$  and  $1 + x \leq e^x$  this gives us

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

993  $\square$

994 **Lemma 22.**  $\mathbb{E}[\tilde{v}^\top B_t B_t^\top \tilde{v} | P] \geq \exp(\sum_{i \in [t]} (2\eta_i \tilde{\lambda}_1 - 4\eta_i^2 \tilde{\lambda}_1^2))$

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

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

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

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

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

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

999 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

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

1000

□

1001 **Lemma 23.**  $\mathbb{E}[(\tilde{v}^\top B_t B_t \tilde{v})^2 | P] \leq \exp(\sum 4\eta_i \tilde{\lambda}_1 + 10\eta_i^2 \tilde{\nu})$

1002 *Proof.* We define  $\gamma_s := \mathbb{E}[(\tilde{v}^\top W_{t,s} W_{t,s}^\top \tilde{v})^2 | P]$  where  $W_{t,s} := (\mathbf{I} - \eta_t P A_i P) \cdot \dots (\mathbf{I} +$   
 1003  $\eta_{t-s+1} P A_{t-s+1} P)$ . So by this definition we see  $W_{t,t} = B_t$  and  $\gamma_t = \mathbb{E}[(\tilde{v}^\top B_t B_t^\top \tilde{v})^2 | P]$ . As  
 1004 the trace of a scalar is the scalar itself, we can exploit the cyclic permutation properties of the trace:

$$\begin{aligned}\gamma_t &= \text{Tr}(\mathbb{E}[W_{t,t}^\top \tilde{v} \tilde{v}^\top W_{t,t} W_{t,t}^\top \tilde{v} \tilde{v}^\top W_{t,t} | P]) \\ &= \text{Tr}(\mathbb{E}[(\mathbf{I} + \eta_1 A_1^\top) G_{t-1} (\mathbf{I} + \eta_1 A_1) (\mathbf{I} + \eta_1 A_1^\top) G_{t-1} (\mathbf{I} + \eta_1 A_1) | P])\end{aligned}$$

1005 where  $G_{t-1} := W_{t,t-1}^\top v_1 v_1^\top W_{t,t-1}$ . We first bound for an arbitrary  $G_{t-1} = G$ , and then take the  
 1006 expectation over only  $A_1$  and finally over  $G_{t-1}$ .

$$\begin{aligned}&\text{Tr}(\mathbb{E}[(\mathbf{I} + \eta_1 P A_1^\top P) G (\mathbf{I} + \eta_1 P A_1 P) (\mathbf{I} + \eta_1 P A_1^\top P) G (\mathbf{I} + \eta_1 P A_1 P) | P]) \\ &= \text{Tr}(\mathbb{E}[(G + \eta_1 P A_1^\top P G + \eta_1 G P A_1 P + \eta_1^2 P A_1^\top P G P A_1 P)^2 | P]) \\ &= \text{Tr}(G^2) + 4\eta_1 \text{Tr}(P \Sigma P G^2) + 2\eta_1^2 \text{Tr}(\mathbb{E}[P A_1 P A_1^\top P | P] G^2) \\ &\quad + \eta_1^2 \text{Tr}(\mathbb{E}[P A_1^\top P G P A_1 P G | P]) + \eta_1^2 \text{Tr}(\mathbb{E}[P A_1^\top P G P A_1^\top P G | P]) \\ &\quad + \eta_1^2 \text{Tr}(\mathbb{E}[G P A_1 P G P A_1 P | P]) + \eta_1^2 \text{Tr}(\mathbb{E}[G P A_1^\top P G P A_1 P | P]) \\ &\quad + 2\eta_1^3 \text{Tr}(\mathbb{E}[P A_1^\top P G P A_1^\top P G P A_1 P | P]) \\ &\quad + \eta_1^4 \text{Tr}(\mathbb{E}[P A_1^\top P G P A_1 P A_1^\top P G P A_1 P | P])\end{aligned}$$

1007 Let's begin with the first order terms:

$$\begin{aligned}\text{Tr}(P \Sigma P G^2) &\leq \|P \Sigma P\|_2 \text{Tr}(G^2) = \tilde{\lambda}_1 \text{Tr}(G^2) \\ \text{Tr}(\mathbb{E}[P A_1 P A_1^\top P | P] G^2) &\leq (\|\mathbb{E}[P(A_1 - \Sigma)P(A_1^\top - \Sigma)P] \|_2 + \|P \Sigma P\|_2) \text{Tr}(G^2) \leq (\mathcal{V} + \tilde{\lambda}_1^2) \text{Tr}(G^2)\end{aligned}$$

1008 where the last inequality follows by Lemma 11. Next we have 4 second order terms:

$$\begin{aligned}&\text{Tr}(\mathbb{E}[P A_1^\top P G P A_1 P G | P]) = \text{Tr}(\mathbb{E}[P A_1^\top P G P A_1^\top P G | P]) \\ &= \text{Tr}(\mathbb{E}[G P A_1 P G P A_1 P | P]) = \text{Tr}(\mathbb{E}[G P A_1^\top P G P A_1 P | P]) \\ &\leq \frac{1}{2} \mathbb{E}[\|P A_1^\top P G\|_F^2 + \|P A_1 P G\|_F^2 | P] \\ &= \frac{1}{2} \text{Tr}(G \mathbb{E}[P A_1 P A_1^\top P | P] G + G \mathbb{E}[P A_1 P A_1^\top P | P] G) \leq (\mathcal{V} + \tilde{\lambda}_1^2) \text{Tr}(G^2)\end{aligned}$$

1009 Third order terms we can bound as follows:

$$\begin{aligned}\text{Tr}(\mathbb{E}[P A_1^\top P G P A_1^\top P G P A_1 P | P]) &\leq \|P A_1^\top P\| \text{Tr}(\mathbb{E}[P A_1^\top P G G P A_1 P | P]) \\ &\leq (\|P(A_1 - \Sigma)P\|_2 + \|P \Sigma P\|_2) \text{Tr}(G \mathbb{E}[P A_1 P A_1^\top P | P] G) \\ &\leq (\mathcal{M} + \tilde{\lambda}_1)(\mathcal{V} + \tilde{\lambda}_1) \text{Tr}(G^2)\end{aligned}$$

1010 Finally the fourth order terms

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

1011 all of this together gives us

$$\begin{aligned} &\text{Tr}(\mathbb{E}[(\mathbf{I} + \eta_1 PA_1^\top P)G(\mathbf{I} + \eta_1 PA_1 P)(\mathbf{I} + \eta_1 PA_1^\top P)G(\mathbf{I} + \eta_1 PA_1 P)|P]) \\ &\leq \text{Tr}(G^2) + 4\eta_1 \tilde{\lambda}_1 \text{Tr}(G^2) + 5\eta_1^2 \bar{\mathcal{V}} \text{Tr}(G^2) + 4\eta_1^3 (\mathcal{M} + \tilde{\lambda}_1) \bar{\mathcal{V}} \text{Tr}(G^2) + \eta_1^4 (\mathcal{M} + \tilde{\lambda}_1)^2 \bar{\mathcal{V}} \text{Tr}(G^2) \\ &= (1 + 4\eta_1 \tilde{\lambda}_1 + 5\eta_1^2 \bar{\mathcal{V}} + 4\eta_1^3 (\mathcal{M} + \tilde{\lambda}_1) \bar{\mathcal{V}} + \eta_1^4 (\mathcal{M} + \tilde{\lambda}_1)^2 \bar{\mathcal{V}}) \text{Tr}(G^2) \\ &\leq (1 + 4\eta_1 \tilde{\lambda}_1 + 10\eta_1^2 \bar{\mathcal{V}}) \text{Tr}(G^2) \\ &\leq \exp(4\eta_1 \tilde{\lambda}_1 + 10\eta_1^2 \bar{\mathcal{V}}) \text{Tr}(G^2) \end{aligned}$$

1012 where we used  $\eta_i \leq \frac{1}{4 \max\{\tilde{\lambda}_1, \mathcal{M}\}}$  and  $1 + x \leq \exp(x)$ . All of this give us

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

1013 then using  $\gamma_0 = 1$  gives us the wished result.  $\square$

**Lemma 24.**

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

1014 *Proof.* Let  $\alpha_t := \mathbb{E}[\text{Tr}(\tilde{V}_\perp^\top B_t B_t^\top \tilde{V}_\perp)|P]$ . Then using the cyclic property of the trace and the fact  
1015 that  $\tilde{V}_\perp$  is not random in  $\mathbb{E}[\cdot|P]$ , we have

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

1016 the RHS of this matrix inner product equates to

$$\begin{aligned} &\tilde{V}_\perp \tilde{V}_\perp^\top + \eta_t P \Sigma P \tilde{V}_\perp \tilde{V}_\perp^\top + \eta_t \tilde{V}_\perp \tilde{V}_\perp^\top P \Sigma P + \eta_t^2 \mathbb{E}[P A_t P \tilde{V}_\perp \tilde{V}_\perp^\top P A_t P |P] \\ &\preceq \tilde{V}_\perp \tilde{V}_\perp^\top + 2\eta_t \tilde{\lambda}_2 \tilde{V}_\perp \tilde{V}_\perp^\top + \eta_t^2 \tilde{\lambda}_1^2 \tilde{V}_\perp \tilde{V}_\perp^\top + \mathbb{E}[P(A_t - \Sigma)P(A_t - \Sigma)P |P] \\ &\preceq (1 + 2\eta_t \tilde{\lambda}_2 + \eta_t^2 \bar{\mathcal{V}}) \tilde{V}_\perp \tilde{V}_\perp^\top + \eta_t^2 \mathcal{V} v_1 v_1^\top \end{aligned}$$

1017 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  
1018 orthogonal matrix. So plugging this into the inner product we get

$$\alpha_t \leq (1 + 2\eta_t \tilde{\lambda}_2 + \eta_t^2 \bar{\mathcal{V}}) \langle \mathbb{E}[B_{t-1} B_{t-1}^\top |P], \tilde{V}_\perp \tilde{V}_\perp^\top \rangle + \eta_t^2 \mathcal{V} \langle \mathbb{E}[B_{t-1} B_{t-1}^\top |P], v_1 v_1^\top \rangle$$

1019 using  $1 + x \leq \exp(x)$  we get

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

1020 using Lemma 21. Then by recursion we get

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

1021 the result follows by  $\alpha_0 = d - 1 \leq d$   $\square$

1022 *Proof of Theorem 7.* Using conditional Chebychev's (Lemma 8) we have

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

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  $1 - \delta$

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

with

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

where we use Lemma 22 and Lemma 23. So putting this together we get

$$\tilde{v}^\top B_n B_n^\top \tilde{v} \geq \exp\left(\sum_{i \in [n]} (2\eta_i \tilde{\lambda}_1 - 4\eta_i^2 \tilde{\lambda}_1^2)\right) \left(1 - \frac{1}{\sqrt{\delta}} \sqrt{\exp(\sum_{i \in [n]} 18\eta_i^2 \tilde{\mathcal{V}}) - 1}\right) \quad (12)$$

this lower bounds the denominator in Lemma 20. 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 24 we get

$$\text{Tr}(\tilde{V}_\perp^\top B_t B_t^\top \tilde{V}_\perp) \leq \frac{1}{\delta} \exp\left(\sum_{j \in [t]} 2\eta_j \tilde{\lambda}_2 + \eta_j^2 \tilde{\mathcal{V}}\right) \left(d + \mathcal{V} \sum_{i=1}^t \eta_i^2 \exp\left(\sum_{j \in [i]} 2\eta_j (\tilde{\lambda}_1 - \tilde{\lambda}_2)\right)\right) \quad (13)$$

so plugging Equation 12 and 13 into Lemma 20 we get that with probability at least  $1 - 2\delta$

$$\begin{aligned}&\sin^2(\tilde{v}, \frac{B_n w_n}{\|B_n w_n\|_2}) \\ &\leq \frac{C \log(1/\delta)}{\delta} \frac{1}{Q'} \exp(\sum_{j \in [t]} 2\eta_j (\tilde{\lambda}_2 - \tilde{\lambda}_1) + \eta_j^2 (\tilde{\mathcal{V}} + 4\tilde{\lambda}_1^2)) (d + \mathcal{V} \sum_{i=1}^t \eta_i^2 \exp(\sum_{j \in [i]} 2\eta_j (\tilde{\lambda}_1 - \tilde{\lambda}_2)))\end{aligned}$$

where  $Q' = \left(1 - \frac{1}{\sqrt{\delta}} \sqrt{\exp(\sum_{i \in [n]} 18\eta_i^2 \tilde{\mathcal{V}}) - 1}\right)$ . By  $\tilde{\mathcal{V}} + 4\tilde{\lambda}_1^2 \leq 5\tilde{\mathcal{V}}$  the result follows.  $\square$

## E Proof of Main Theorem

The privacy proof of Algorithm 1 follows straight using Advanced Composition (Lemma 17) together with the privacy of MODIFIEDDP-PCA, which in turn follows by [Liu et al., 2022a]. 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 likelihood we can reduce the update step to an update step of non private Oja's Algorithm with matrices  $PC_t P$ . We then use a novel result we proved (Appendix D), which shows that non private Oja's Algorithm given input  $\{PC_t P\}_t$  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.

1044 **Theorem 9.** Given  $A_1, \dots, A_n$  are i.i.d. and satisfy Assumption A, MODIFIEDDP-PCA and  $k$ -DP-  
 1045 Ojas as defined Algorithms 2 and 3 are stochastic ePCA with  $\zeta = \tilde{O}\left(\kappa' \left(\sqrt{\frac{V}{n}} + \frac{\gamma d \sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$   
 1046 and  $\zeta = \tilde{O}\left(\kappa' \left(\sqrt{\frac{V}{n}} + \frac{(\gamma+1)d \sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right)$  respectively.

1047 *Proof.* The proof follows by the utility proofs of MODIFIEDDP-PCA and DP-Ojas (Theorem 10  
 1048 and Theorem 11) and Lemma 19.  $\square$

1049 **Theorem 10** (Utility of MODIFIEDDP-PCA). For  $\varepsilon \in (0, 0.9)$  and  $0 < k < d$ , ModifiedDP-PCA  
 1050 fulfills  $(\varepsilon, \delta)$ -DP for all inputs  $\{A_i\}, B, \zeta$  and  $\delta$  and any projection matrix  $P$  (that we assume to  
 1051 already be private). Given  $n$  i.i.d. samples  $\{A_i \in \mathbb{R}^{d \times d}\}_{i=1}^n$  and  $P$  satisfying Assumptions Assump-  
 1052 tion A.1–Assumption A.4 with parameters  $(\Sigma, M, V, K, \kappa', a, \gamma^2)$ , if

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

1053 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

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

1054 then there exists a learning rate  $\eta_t$  that depends on  $(t, M, V, K, a, \lambda_1(\Sigma), \lambda_1(P\Sigma P) -$   
 1055  $\lambda_2(P\Sigma P), n, d\varepsilon, \delta)$  such that  $T = \lfloor n/B \rfloor$  steps of ModifiedDP-PCA with choices of  $\tau = 0.01$   
 1056 and  $B = c_1 n / (\log n)^3$  outputs  $\omega_T$  such that with probability 0.99

$$\sin(\omega_t, \tilde{v}) \leq \tilde{O}\left(\kappa' \left(\sqrt{\frac{V}{n}} + \frac{\gamma d \sqrt{\log(1/\delta)}}{\varepsilon n}\right)\right) \quad (14)$$

1057 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  
 1058  $\log(1/\delta)$  and polynomial factors in  $K$ .

1059 *Remark.* For readability we omitted the advanced composition details in the above proof. If we  
 1060 choose  $T = O(\log 2n)$ , we can simply set  $(\varepsilon', \delta') = (\varepsilon / (2\sqrt{2\log^2(n)\log(2/\delta)}), \delta / (2\log^2(n)))$  in  
 1061 every step and then by advanced composition we get. And in our utility guarantee we would only  
 1062 occur additional  $\log^2(n)$  factors which we omit.

1063 *Proof.* We choose the batch size  $B = \Theta(n/\log^3 n)$  such that we access the dataset only  $T =$   
 1064  $\Theta(\log^3 n)$  times. Hence we do not need to rely on amplification by shuffling. To add Gaussian noise  
 1065 that scales as the standard deviation of the gradients in each minibatch (as opposed to potentially  
 1066 excessively large mean of the gradients), DP-PCA first gets a private and accurate estimate of the  
 1067 range. Using this estimate,  $\Lambda$ , PRIVMEAN returns an unbiased estimate of the empirical mean of the  
 1068 gradients, as long as no truncation has been applied. As we choose the truncation threshold so that  
 1069 with high probability there will be no truncation the update step will look as follows:

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

1070 where  $z_t \sim \mathcal{N}(0, \mathbf{I})$  and  $\beta_t = \frac{8K\sqrt{2\Lambda_t} \log^a(Bd/\tau) \sqrt{2d\log(2.5/\delta)}}{\varepsilon B}$ . The privacy follows by the privacy  
 1071 of the subroutines private eigenvalue and private mean estimation [Liu et al., 2022a]. So all that is  
 1072 left to do is show the utility guarantee. We will do that by showing we can reduce it the accuracy of  
 1073 the non private case. First we note that  $P^2 = P$  so we get

$$\omega'_t = \omega_{t-1} + \eta_t \left( \frac{1}{B} \sum_{i \in [B]} P A_i P \omega_{t-1} + \beta_t P z_t \right)$$

1074 Using rotation invariance of the spherical Gaussian random vectors and the fact that  $\|\omega_{t-1}\| = 1$  and  
 1075  $\omega_{t-1} \in \text{Im}(P)$  (for details see Lemma 9), we can reformulate it as

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



1076 we can further pull out the projection matrices to obtain

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

1077 Where  $G$  is a matrix whose entries are i.i.d.  $\mathcal{N}(0, 1)$  distributed. So we have a matrix

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

1078 and we will now proof that  $C_t$  fulfills all requirements for Theorem 8 (our version of the non private  
1079 Oja's Algorithm utility guarantee), which will directly give us the wished utility guarantee. It is easy  
1080 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  
1081 upper bound of  $\max\{\|\mathbb{E}[(C_t - \Sigma)(C_t - \Sigma)^\top]\|_2, \|\mathbb{E}[(C_t - \Sigma)^\top(C_t - \Sigma)]\|_2\}$

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

1082 where the first inequality holds due to  $G_t$  being independent to  $A_i$ , and  $\mathbb{E}[G_t] = 0$ . The second  
1083 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.  
1084 And the last inequality holds with high probability due to  $G_t$  having i.i.d. Gaussian entries, and by  
1085 choosing

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

1086 we have  $\beta \geq \beta_t$  for all  $t$  as by Theorem 6.1 in [Liu et al., 2022a] and Assumption 4

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

1087 Lastly let us consider  $\|C_t - \Sigma\|_2$ . By Lemma 3 and Lemma 4 we know with probabability  $1 - \tau$  for  
1088 all  $t \in [T]$

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

1089 so by Theorem 8 with stepsize  $\eta_t := \frac{\alpha}{(\lambda_1 - \lambda_2)(\xi + t)}$  after  $T$  steps with

$$T \geq 20 \max \left( \frac{\tilde{M} \alpha}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)}, \frac{(\tilde{V} + \lambda_1^2) \alpha^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 \log(1 + \frac{\zeta}{100})} \right) := \xi \quad (15)$$

1090 with probability  $1 - \zeta$

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

1091 so if we fill in  $\tilde{M}$ ,  $\tilde{V}$ , and  $\beta$  into  $\xi$  and use  $n = BT$  we get

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

1092 in order for Theorem 8 to hold we need to force  $\xi/T \leq 1$ . Noting  $\tau = O(1)$ ,  $K = O(1)$  and  
1093 selecting  $\alpha = c \log n$ ,  $T = c'(\log n)^3$  we get that

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

1094 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  
1095 need

$$\frac{\lambda_1^2 (\log n)^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 T} \leq 1/3 \quad (16)$$

1096 as  $T = c'(\log(n))^3$  this means

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

1097 for the remaining terms we need

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

1098 We note that to obtain  $n/\log(n) \geq a$ ,  $n \simeq a \log(a) + a \log \log(a)$ . So

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

1099 with large enough constant suffices (where  $\gtrsim$  is hiding log terms) to obtain  $\xi/T \leq 1$  and  $d(\xi/T)^{2\alpha} \leq$   
1100  $1/n^2$ . And we get

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

1101 (where  $\lesssim$  is hiding log terms), so plugging this in our bound for  $\sin(\omega_T, \tilde{v})$  we get

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

1102 which finishes the proof  $\square$

1103 The above utility result depends on the eigenvalues of the input. After the first step of  $k$ -DP-PCA our  
1104 input is of the form  $PA_1P, \dots, PA_nP$ , so our utility bound depends on the eigenvalues of  $P\Sigma P$ . In  
1105 general  $\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)$  can be arbitrarily much smaller than the actual eigengap of  $\Sigma$ , and  
1106 therefore it is not a sufficient utility bound as is. However, as we iteratively apply projection matrices  
1107 of the form

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

1108 where  $u$  is a unit vector, and further  $u$  is  $\varepsilon$ -close to the top eigenvector of the matrix we apply it to,  
1109 we can actually relate the eigengap of  $P\Sigma P$  to the one of  $\Sigma$  using Weyl's Theorem.

1110 **Lemma 25.** Given  $\sin^2(\theta) \leq \xi$ , where  $\theta$  refers to the angle between  $v_1$  and  $u$  we have

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

1111 where  $\Delta = 8\lambda_1\sqrt{\xi}(1 + \sqrt{\xi})$

1112 *Proof.* We will use Weyl's Theorem to proof this, by defining

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

1113 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  
1114 we can use this as follows:

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

1115 where the last inequality follows by Weyl's Theorem. Next we will bound  $\|G_1 - G_2\|$

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

1116 where the last step follows as  $(uu^\top \Sigma uu^\top - v_1 v_1^\top \Sigma v_1 v_1^\top) = (uu^\top - v_1 v_1^\top) \Sigma uu^\top + v_1 v_1^\top \Sigma (uu^\top -$   
1117  $v_1 v_1^\top)$  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  
1118  $\sin^2(v_1, u) \leq \xi$ : First we note that as  $u$  and  $v_1$  are unit vectors we can write

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

1119 so this means

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

1120 and also gives us

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

1121 □

1122 so all in all this tells us

1123 **Lemma 26.** For  $\Sigma \in \mathbb{R}^{d \times d}$  a matrix with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ ,  $P = I - uu^\top$ , with  
1124  $u \in \text{Im}(\Sigma)$ , and  $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \dots \geq \tilde{\lambda}_{d-1}$  the eigenvalues of  $P\Sigma P$

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

1125 where  $\Delta = 8\lambda_1\sqrt{\xi}(1 + \sqrt{\xi})$  and  $\xi \geq \sin^2(\theta)$  with  $\theta$  the angle between  $u$  and  $v_1$ , the top eigenvector  
1126 of  $\Sigma$ .

1127 Lemma 26 together with Theorem 10 this give us a utility guarantee independent of eigenvalues of  
1128  $P\Sigma P$  for  $k \leq 2$ .

1129 We can see that, because first of all for  $k = 2$  we pass  $P = \mathbf{I}$  and  $P = I - u_1 u_1^\top$ . And secondly,  
1130 because the utility bound of MODIFIEDDP-PCA depends on several constants originating from  
1131 constraints on the data:

1132 1.  $\kappa = \frac{\lambda_1}{\lambda_1 - \lambda_2}$

- 1133 2.  $M$  so that  $\|A_i - \Sigma\|_2 \leq \lambda_1 M$  almost surely  
 1134 3.  $V$  so that  $\max\{\|\mathbb{E}[(A_i - \Sigma)(A_i - \Sigma)^\top]\|_2, \|\mathbb{E}[(A_i - \Sigma)^\top(A_i - \Sigma)]\|_2\} \leq \lambda_1^2 V$   
 1135 4.  $\gamma^2 := \max_{\|u\|=1} \|H_u\|_2$   
 1136 5.  $K$  so that  $\max_{\|u\|=1, \|v\|=1} \mathbb{E} \left[ \exp \left( \left( \frac{|u^\top (A_i^\top - \Sigma)v|^2}{K^2 \lambda_1^2 \|H_u\|_2} \right)^{1/(2a)} \right) \right] \leq 1$

1137 now if we replace the  $\{A_i\}$  with  $\{PA_iP\}$  where  $P$  is a projection matrix, the constants  $M, V, \lambda_1^2 \gamma^2$   
 1138 and  $K$  will still remain upper bounds (proved in Lemma 10, Lemma 11, Lemma 13). Therefore, if we  
 1139 just swapped  $\kappa$  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  
 1140 as a utility bound for  $\{PA_iP\}$  as input to our modified DP-PCA algorithm

$$\xi = \kappa B_n \quad (17)$$

1141 where

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

1142 so for  $k = 2$  Lemma 26 will give us a utility bound independent of  $P$ . However, we want to obtain a  
 1143 utility guarantee for arbitrary  $k < d$ . From now on we will denote

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

1144 and the goal is to upper bound  $\kappa_i$  with something independent of  $P$ . If we iteratively applying Lemma  
 1145 26 we get

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

1146 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  $\Delta_j$  still  
 1147 depends on previous projections and it's not even clear in general if  $\xi_j > \xi_{j+1}$  or the other way  
 1148 around. Ultimately we want to have an upper bound for all  $\xi_j$ , to get a utility bound for  $U = \{u_i\}$ . A  
 1149 natural approach is to try and choose  $n$  big enough so that

$$\begin{aligned} \lambda_1(P_i\Sigma P_i) &\leq \lambda_1 \\ \lambda_1(P_i\Sigma P_i) - \lambda_2(P_i\Sigma P_i) &\geq \delta \end{aligned}$$

1150 for some  $\delta > 0$ . If we achieve this we are done, as this will guarantee that

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

1151

1152 **Lemma 27.** *If for  $k$  fixed,  $0 < \Delta = \min_{i \in [k]} \lambda_i - \lambda_{i+1}$ ,  $0 < \delta < \Delta$  and a sufficiently large constant*  
 1153  *$C > 1$ , we are given  $\{A_i\}_{i=1}^n$  fulfilling Assumption A with  $n$  is big enough so that*

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

1154 then

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

1155 where  $\xi_i$  refers to the utility  $\xi_i$  of the vector  $u_i$  returned at iteration  $i \in [k]$  of Algorithm 1.

1156 *Proof.* We will proof that at every step:

$$\lambda_1(P_i\Sigma P_i) \leq \lambda_1 \quad (18)$$

$$\lambda_1(P_i\Sigma P_i) - \lambda_2(P_i\Sigma P_i) \geq \delta \quad (19)$$

1157 is fulfilled, which directly implies what we are trying to proof. We will proof these two statements by  
 1158 induction. For  $k = 1$  we have  $P_0 = \mathbf{I}$  which straightaway gives us equation 18. And as  $\delta$  is smaller  
 1159 then the minium eigengap equation 19, directly follows as well. For  $k + 1$  we start with showing  
 1160 equation 18. By Lemma 26

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

1161 first let's upper bound  $\sum_{j=1}^k \Delta_j$  by induction assumption:

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

1162 so equation 18 will be implied by

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

1163 which is surely fulfilled as by assumption

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

1164 To show equation 19, we see

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

1165 where the first inequality follows by Lemma 26 and the second by definition of  $\Delta$ . Using the upper  
 1166 bound on  $\sum_{j=1}^k \Delta_j$  we established

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

1167 will imply equation 19. □

1168 We will now combine all of this to proof our main theorem:

1169 **Theorem 1** (Main Theorem). *Let  $\varepsilon, \delta \in (0, 0.9)$  and  $1 \leq k < d$ . Then  $k$ -DP-PCA satisfies the*  
 1170 *following:*

1171 **Privacy:** *For any input sequence  $\{A_i \in \mathbb{R}^{d \times d}\}$ , the algorithm is  $(\varepsilon, \delta)$ -differentially private.*

1172 **Utility:** *Suppose  $A_1, \dots, A_n$  are i.i.d. satisfying Assumption A with parameters*  
 1173  *$(\Sigma, M, V, K, \kappa', a, \gamma^2)$ . If*

$$n \geq C \max \left\{ \begin{aligned} &e^{\kappa'^2} + \frac{d \kappa' \gamma \sqrt{\ln(1/\delta)}}{\varepsilon} + \kappa' M + \kappa'^2 V + \frac{\sqrt{d} (\ln(1/\delta))^{3/2}}{\varepsilon}, \\ &\lambda_1^2 \kappa'^2 k^3 V, \\ &\frac{\kappa'^2 \gamma k^2 d \sqrt{\ln(1/\delta)}}{\varepsilon} \end{aligned} \right\}, \quad (1)$$

1174 *for a sufficiently large constant  $C$ , then with probability at least 0.99, the output  $U \in \mathbb{R}^{d \times k}$  is*  
 1175  *$\zeta$ -approximate with*

$$\zeta = \tilde{O} \left( \kappa' \left( \sqrt{\frac{Vk}{n}} + \frac{\gamma dk \sqrt{\log(1/\delta)}}{\varepsilon n} \right) \right), \quad (2)$$

1176 where  $\tilde{O}(\cdot)$  hides factors polylogarithmic in  $n, d, 1/\varepsilon, \ln(1/\delta)$  and polynomial in  $K$ .

1177 *Proof of Theorem 1.* By Theorem 10 we know that when passing  $m = n/k$  matrices  $A_i$  at every step  
1178 of our deflation method we obtain a vector  $u_i$  fulfilling

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

1179 where  $v_i$  is the top eigenvector of  $P_{i-1}\Sigma P_{i-1}$ . Which by Lemma 19 give us

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

1180 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  
1181 Lemma 27 that

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

1182 where we used that  $(\Delta - \delta)\delta$  is maximized by  $\delta = \Delta/2$ . So finally Theorem 6 gives us that

$$\langle UU^\top, \Sigma \rangle \geq (1 - \zeta^2) \langle V_k V_k^\top, \Sigma \rangle \quad (20)$$

1183 where  $V_k$  is the matrix obtained by non private  $k$ -PCA.  $\square$

## 1184 E.1 Proof of Utility of DP-Ojas

1185 **Theorem 11. Privacy:** If  $\varepsilon = O(\sqrt{\log(n/\delta)/n})$  and the noise multiplier is chosen to be  $\alpha =$   
1186  $\Omega(\log(n/\delta)/(\varepsilon\sqrt{n}))$ , then Algorithm 3 is  $(\varepsilon, \delta)$ -DP.

1187 **Utility:** Given  $n$  i.i.d. samples  $\{A_i \in \mathbb{R}^{d \times d}\}_{i=1}^n$  satisfying Assumption A with parameters  
1188  $(\Sigma, M, V, K, \kappa', a, \gamma^2)$  if

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

1189 with large enough constant, then there exists a positive universal constant  $c_1$  and choice of learning  
1190 rate  $\eta_t$  such that Algorithm 3 with a choice of  $\zeta = 0.01$  outputs  $w_n$  that with probability 0.00,

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

1191 where  $\kappa' = \frac{\lambda_1(\Sigma)}{\lambda_1(P\Sigma P) - \lambda_2(P\Sigma P)}$  and  $\tilde{O}(\cdot)$  hides poly-logarithmic factors in  $n, d, 1/\varepsilon$ , and  $\log(1/\delta)$   
1192 and polynomial factors in  $K$ .

1193 *Proof. Privacy:* The privacy proof follows by Lemma 3.1 in [Liu et al., 2022b].

1194 **Utility:** By Assumption A.4 it follows analogously to Lemma 3.2 in [Liu et al., 2022b] that with  
1195 probability  $1 - O(\zeta)$  Algorithm 3 does not have any clipping. Under this event, the update rule  
1196 becomes

$$\begin{aligned} w'_t &\leftarrow w_{t-1} + \eta_t P(PA_t Pw_{t-1} + 2\beta\alpha z_t) \\ w_t &\leftarrow Pw'_t / \|Pw'_t\| \end{aligned}$$

1197 where  $\beta = C\lambda_1\sqrt{d}(K\gamma\log^2(nd/\zeta) + 1)$  and  $z_t \sim \mathcal{N}(0, \mathbf{I})$ . Just like in the proof of Theorem 10 we  
1198 use that  $P^2 = P$  and Lemma 9 to rewrite this as

$$w'_t \leftarrow w_{t-1} + \eta_t P(A_t + 2\beta\alpha G_t) Pw_{t-1}$$

1199 where  $G$  is a matrix whose entries are i.i.d.  $\mathcal{N}(0, 1)$  distributed. So if we define

$$\tilde{A}_t := A_t + 2\beta\alpha G_t$$

1200 this becomes

$$w'_t \leftarrow w_{t-1} + \eta_t P \tilde{A}_t P w_{t-1}$$

1201 so if we can show the  $\tilde{A}_t$ 's fulfill all requirements for Theorem 8, we will directly obtain the wished  
1202 utility guarantee. Equivalently to the proof of Theorem 10 we can show

$$\begin{aligned} \|\mathbb{E}[(\tilde{A}_t - \Sigma)(\tilde{A}_t - \Sigma)^\top]\|_2 &\leq V\lambda_1^2 + 4\alpha^2\beta^2 C_2 d =: \tilde{V} \\ \|\tilde{A}_t - \Sigma\|_2 &\leq M\lambda_1 + 2C_3\alpha\beta(\sqrt{d} + \sqrt{\log(n/\zeta)}) =: \tilde{V} \end{aligned}$$

1203 Under the event that  $\|\tilde{A}_t - \Sigma\|_2 \leq \tilde{M}$  for all  $t \in [n]$ , we apply Theorem 8 with a learning rate  
1204  $\eta_t = \frac{h}{(\lambda_1 - \lambda_2)(\xi + t)}$  where

$$\xi = 20 \max \left( \frac{\tilde{M}h}{(\lambda_1 - \lambda_2)}, \frac{(\tilde{V} + \lambda_1)^2 h^2}{(\lambda_1 - \lambda_2)^2 \log(1 + \frac{\zeta}{100})} \right)$$

1205 which tells us that with probability  $1 - \zeta$ , for  $n > \xi$

$$\sin^2(w_n, v_1) \leq \frac{C \log(1/\zeta)}{\zeta^2} \left( d \left( \frac{\xi}{n} \right)^{2h} + \frac{h^2 \tilde{V}}{(2h-1)(\lambda_1 - \lambda_2)^2 n} \right)$$

1206 for some positive constant  $C$ . If we plug in  $\alpha = \frac{C' \log(n/\delta)}{\varepsilon \sqrt{n}}$  (as defined in Algorithm 3), set  $\zeta = O(1)$ ,  
1207  $K = O(1)$ , select  $h = c \log(n)$  and assume

$$n \geq C \left( \frac{M\lambda_1 \log(n)}{\lambda_1 - \lambda_2} + \frac{V\lambda_1^2 (\log(n))^2}{(\lambda_1 - \lambda_2)^2} \frac{(K\gamma \log^2(nd/\zeta) + 1)\lambda_1 \log(n/\delta) \log(n)d}{(\lambda_1 - \lambda_2)\varepsilon} + \frac{\lambda_1^2 \log^2(n)}{(\lambda_1 - \lambda_2)^2} \right)$$

1208 we are guaranteed  $n \geq \xi$  and  $d(\xi/n)^{2\alpha} \leq 1/n^2$ , so we will obtain the wished bound.  $\square$

## 1209 E.2 Sample Size requirements

1210 The sample size condition in Theorem 1:

$$n \geq C \max \left\{ \begin{aligned} &e^{\kappa'^2} + \frac{d\kappa'\gamma\sqrt{\ln(1/\delta)}}{\varepsilon} + \kappa'M + \kappa'^2 V + \frac{\sqrt{d}(\ln(1/\delta))^{3/2}}{\varepsilon}, \\ &\lambda_1^2 \kappa'^2 k^3 V, \\ &\frac{\kappa'^2 \gamma k^2 d \sqrt{\ln(1/\delta)}}{\varepsilon} \end{aligned} \right\},$$

1211 includes an exponential dependence on the spectral gap:  $n \geq \exp(\kappa')$ . While this is relatively  
1212 harmless as there is no such exponential dependence in the utility guarantee of the Theorem, we  
1213 are able to get rid of this exponential dependency in exchange for an additional term in the utility  
1214 guarantee. When looking at the utility proof of MODIFIEDDP-PCA (Theorem 10) we see this term  
1215 arises as we choose  $T$  and  $n$  so that  $(\xi/T) < 1$ , as this is one of the requirements of Theorem 8. The  
1216 specific inequality that arose from bounding  $(\xi/T)$  and that lead to this exponential dependency is

$$\frac{\lambda_1^2 (\log n)^2}{(\tilde{\lambda}_1 - \tilde{\lambda}_2)^2 T} \leq 1/3 \quad (21)$$

1217 (see Equation (16)). As we selected  $T = c'(\log n)^3$ , we required  $\log(n) \geq \lambda_1/(\lambda_1 - \lambda_2)$ . By  
1218 selecting a slightly larger  $T = c\kappa \log^3 n$ , we would get rid of this exponential dependence, however  
1219 at the cost of getting an extra term of  $\tilde{O}(\kappa^r \gamma^2 d^2 \log(1/\delta)/(\varepsilon n)^2)$  in the utility guarantee.

## 1220 F Proof of Lower Bound

1221 **Lemma 28.** *Under certain assumptions for an orthonormal matrix  $U \in \mathbb{R}^{d \times k}$  and a psd matrix*  
 1222  *$X \in \mathbb{R}^{d \times d}$  with eigengap  $\Delta_k = \lambda_k - \lambda_{k+1}$  and top  $k$  eigenvectors  $V \in \mathbb{R}^{d \times k}$ , we have*

$$\frac{\|UU^\top - VV^\top\|_F^2 \Delta_k}{2} \leq \text{Tr}(VV^\top X) - \text{Tr}(UU^\top X)$$

1223 *Proof.* We will proof this by proving the following two (in)equalities:

1224 1.  $\Delta_k \|\sin \Theta(U, V)\|_F^2 \leq \text{Tr}(VV^\top X) - \text{Tr}(UU^\top X)$

1225 2.  $\|UU^\top - VV^\top\|_F = \sqrt{2} \|\sin \Theta(U, V)\|_F$

1226 Inequality 1: We first note that

$$\text{Tr}(VV^\top X) - \text{Tr}(UU^\top X) = \text{Tr}((VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d))$$

1227 as

$$\text{Tr}((VV^\top - UU^\top)\lambda_{k+1}) = \lambda_{k+1} (\text{Tr}(VV^\top) - \text{Tr}(UU^\top)) = 0$$

1228 where the last equality follows as  $\text{Tr}(UU^\top) = k = \text{Tr}(VV^\top)$ . Now

$$\begin{aligned} \text{Tr}((VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d)) &= \text{Tr}((VV^\top + (\mathbf{I}_d - VV^\top))(VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d)) \\ &= \text{Tr}(VV^\top(VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d)) + \text{Tr}((\mathbf{I} - VV^\top)(VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d)) \\ &\geq \text{Tr}(VV^\top(VV^\top - UU^\top)(X - \lambda_{k+1}\mathbf{I}_d)) \\ &\geq \Delta_k \text{Tr}((V_k V_k^\top - UU^\top)_+) \end{aligned}$$

1229 where  $(A)_+$  is obtained by replacing each eigenvalue of the matrix  $A$  with  $\max\{\mu_i, 0\}$ . Now we  
 1230 note that

$$\text{Tr}((V_k V_k^\top - UU^\top)_+) \geq \|\sin \Theta(U, V)\|_F^2$$

1231 Hence, since the  $\sin \theta_i$  are nonnegative (as the principal angles  $\theta_i$  lie in  $[0, \pi/2]$ ) we have  $\text{Tr}((V_k V_k^\top -$   
 1232  $UU^\top)_+) = \sum_{i=1}^k \sin \theta_i$ . Further, by definition we have

$$\|\sin \Theta(U, V)\|_F^2 = \sum_{i=1}^k \sin^2 \theta_i.$$

1233 So by noticing that for any angle  $\theta \in [0, \pi/2]$ ,  $\sin \theta \geq \sin^2 \theta$  we have proved the first inequality.

1234 Inequality 2:  $\|UU^\top - VV^\top\|_F^2 = \text{Tr}((UU^\top - VV^\top)^2)$ . By expanding  $(UU^\top - VV^\top)^2$  we see

$$(UU^\top - VV^\top)^2 = UU^\top - UU^\top VV^\top - VV^\top UU^\top + VV^\top$$

1235 which gives us

$$\begin{aligned} \text{Tr}((UU^\top - VV^\top)^2) &= 2k - 2\text{Tr}(UU^\top VV^\top) \\ &= 2k - \text{Tr}(V^\top UU^\top V) \\ &= 2k - \|U^\top V\|_F^2 \end{aligned}$$

1236 Lastly, utilizing

$$\|U^\top V\|_F^2 = 2 \sum_{i=1}^k \cos^2 \theta_i = 2 \sum_{i=1}^k (1 - \sin^2 \theta_i)$$

1237 the proof follows. □

1238 **Lemma 29** (Reduction to Frobenius norm). *Let  $\Sigma$  be a PSD  $d \times d$  matrix with top- $k$  eigenvectors*  
 1239  *$V_k \in \mathbb{R}^{d \times k}$  and eigenvalues  $\lambda_1 \geq \dots \geq \lambda_d$ . Any  $U \in \mathbb{R}^{d \times k}$  that satisfies  $\|UU^\top - V_k V_k^\top\|_F^2 \geq \gamma$ ,*  
 1240 *must incur*

$$\zeta^2 \geq \frac{\gamma \Delta_k}{2 \sum_{i=1}^k \lambda_i}$$

1241 where  $\Delta_k := \lambda_k - \lambda_{k+1}$ .



1242 *Proof.* As

$$\begin{aligned}\langle UU^\top, X \rangle &= \frac{\langle UU^\top, X \rangle}{\langle V_k V_k^\top, X \rangle} \langle V_k V_k^\top, X \rangle \\ &= \frac{\text{Tr}(UU^\top X)}{\text{Tr}(V_k V_k^\top X)} \langle V_k V_k^\top, X \rangle\end{aligned}$$

1243 this implies that

$$\frac{\text{Tr}(UU^\top X)}{\text{Tr}(V_k V_k^\top X)} \geq 1 - \zeta^2. \quad (22)$$

1244 So any upper bound on  $\frac{\text{Tr}(UU^\top X)}{\text{Tr}(V_k V_k^\top X)}$  will give us a lower bound on  $\zeta^2$ . By Lemma 28 we know

$$\frac{\|UU^\top - VV^\top\|_F^2 \Delta_k}{2} \leq \text{Tr}(VV^\top X) - \text{Tr}(UU^\top X)$$

1245 which gives us that

$$\frac{\text{Tr}(UU^\top X)}{\text{Tr}(V_k V_k^\top X)} \leq 1 - \frac{\|UU^\top - V_k V_k^\top\|_F^2 \Delta_k}{2\text{Tr}(V_k V_k^\top X)}.$$

1246 By equation 22 this gives us

$$\frac{\|UU^\top - V_k V_k^\top\|_F^2 \Delta_k}{2 \sum_{i=1}^k \lambda_i} \leq \zeta^2.$$

1247 □

1248 **Corollary 5** (Lower bound, Spiked Covariance). *Let the  $d \times n$  data matrix  $X$  have i.i.d. columns sam-*  
 1249 *ples from a distribution  $P = \mathcal{N}(0, U^\top \Lambda U^\top + \sigma^2 \mathbf{I}_d) \in \mathcal{P}(\lambda, \sigma^2)$  where  $\mathcal{P}(\lambda, \sigma^2) = \{\mathcal{N}(0, \Sigma), \Sigma =$   
 1250  $U \Lambda U^\top + \sigma^2 \mathbf{I}_d, c\lambda \leq \lambda_k \leq \dots \leq \lambda_1 \leq C\lambda\}$ . Suppose  $\lambda \leq c'_0 \exp\{e\varepsilon - c_0(\varepsilon\sqrt{ndk} + dk)\}$  for  
 1251 some small constants  $c_0, c'_0 > 0$ . Then, there exists an absolute constant  $c_1 > 0$  such that*

$$\inf_{\tilde{U} \in \mathcal{U}_{\varepsilon, \delta}} \sup_{P \in \mathcal{P}(\lambda, \sigma^2)} \mathbb{E}[\zeta] \geq c_1 \left( \left( \frac{\sigma\sqrt{\lambda_1 + \sigma^2}}{\sum_{i=1}^k (\lambda_i + \sigma^2)} \right) \left( \sqrt{\frac{dk}{n}} + \frac{dk}{n\varepsilon} \right) \wedge 1 \right).$$

1252 *Proof.* Combining Lemma 29 with Theorem 14 we obtain the lower bound in Corollary 5. □

## 1253 F.1 Existing Lower Bounds

1254 **Theorem 12** (Lower bound, Gaussian distribution, Theorem 5.3 in Liu et al. [2022a]). *Let  $\mathcal{M}_\varepsilon$  be*  
 1255 *a class of  $(\varepsilon, 0)$ -DP estimators that map  $n$  i.i.d. samples to an estimate  $\hat{v} \in \mathbb{R}^d$ . A set of Gaussian*  
 1256 *distributions with  $(\lambda_1, \lambda_2)$  as the first and second eigenvalues of the covariance matrix is denoted by*  
 1257  $\mathcal{P}_{(\lambda_1, \lambda_2)}$ . *There exists a universal constant  $C > 0$  such that*

$$\inf_{\hat{v} \in \mathcal{M}_\varepsilon} \sup_{P \in \mathcal{P}_{(\lambda_1, \lambda_2)}} \mathbb{E}_{S \sim P^n} [\sin(\hat{v}(S), v_1)] \geq C \min \left( \kappa \left( \sqrt{\frac{d}{n}} + \frac{d}{\varepsilon n} \right) \sqrt{\frac{\lambda_2}{\lambda_1}}, 1 \right)$$

1258 **Theorem 13** (Lower bound without Assumption 4, Theorem 5.4 in Liu et al. [2022a]). *Let  $\mathcal{M}_\varepsilon$  be a*  
 1259 *class of  $(\varepsilon, \delta)$ -DP estimators that map  $n$  i.i.d. samples to an estimate  $\hat{v} \in \mathbb{R}^d$ . A set of distributions*  
 1260 *satisfying 1.-3. of Assumption A with  $M = \tilde{O}(d + \sqrt{n\varepsilon/d})$ ,  $V = O(d)$  and  $\gamma = O(1)$  is denoted by*  
 1261  $\tilde{\mathcal{P}}$ . *For  $d \geq 2$ , there exists a universal constant  $C > 0$  such that*

$$\inf_{\hat{v} \in \mathcal{M}_\varepsilon} \sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_{S \sim P^n} [\sin(\hat{v}(S), v_1)] \geq C \kappa \min \left( \sqrt{\frac{d \wedge \log((1 - e^{-\varepsilon})/\delta)}{\varepsilon n}}, 1 \right)$$

1262 **Theorem 14** (Theorem 4.2 in Cai et al. [2024]). *Let the  $d \times n$  data matrix  $X$  have i.i.d. columns*  
 1263 *samples from a distribution  $P = \mathcal{N}(0, U^\top \Lambda U^\top + \sigma^2 \mathbf{I}_d) \in \mathcal{P}(\lambda, \sigma^2)$ . Suppose  $\lambda \leq c'_0 \exp\{e\varepsilon -$*

1264  $c_0(\varepsilon\sqrt{ndk} + dk)\}$  for some small constants  $c_0, c'_0 > 0$ . Then, there exists an absolute constant  
 1265  $c_1 > 0$  such that

$$\inf_{\tilde{U} \in \mathcal{U}_{\varepsilon, \delta}} \sup_{P \in \mathcal{P}(\lambda, \sigma^2)} \frac{\mathbb{E} \|\tilde{U}\tilde{U}^\top - UU^\top\|_F}{\sqrt{k}} \geq c_1 \left( \left( \frac{\sigma\sqrt{\lambda + \sigma^2}}{\lambda} \right) \left( \sqrt{\frac{d}{n}} + \frac{d\sqrt{k}}{n\varepsilon} \right) \wedge 1 \right)$$

1266 where the infimum is taken over all the possible  $(\varepsilon, \delta)$ -DP algorithms, denoted by  $\mathcal{U}_{\varepsilon, \delta}$  and the  
 1267 expectation is taken with respect to both  $\tilde{U}$  and  $P$  and

$$\mathcal{P}(\lambda, \sigma^2) := \{\mathcal{N}(0, \Sigma) : \Sigma = U\Lambda U^\top + \sigma^2 \mathbf{I}_d, \text{ where } U \in \mathbb{O}_{d, k}, \Lambda = \text{diag}(\lambda_1, \dots, \lambda_k), c_0\lambda \leq \lambda_k \leq \lambda_1 \leq C_0\lambda\}$$

## 1268 G Experiments

1269 In Section 5 we compare the performance of k-DP-PCA and k-DP-Ojas to two modified versions  
 1270 of the DP-Gauss algorithms of Dwork et al. [2014b], we refer to as DP-Gauss-1 and DP-Gauss-2  
 1271 respectively.

1272 Given a stream of matrices  $\{A_i\}$  and a clipping threshold  $\beta$  (that is set based on the distribu-  
 1273 tion of the input data), DP-Gauss-1 first clips each matrix to have trace at most  $\beta^2$ :  $\tilde{A}_i =$   
 1274  $A_i \cdot \min\{1, \beta^2/\text{Tr}(A_i)\}$ . In a second step it computes the sum of the  $\tilde{A}_i$ :  $X = \sum_i \tilde{A}_i$  and then per-  
 1275 forms the gaussian mechanism:  $X' = X + E$ , where  $E$  is a symmetric matrix with their upper triangle  
 1276 values (including its diagonal) i.i.d. sampled from  $\mathcal{N}(0, \Delta_1^2 \mathbf{I}_d)$  and  $\Delta_1 = \beta^2 \sqrt{2 \log(1.25/\delta)}/\varepsilon$ .  
 1277 Lastly, it performs an eigenvalue decomposition on  $X'$ , and releases the top  $k$  eigenvectors.

1278 DP-Gauss-2 just like DP-Gauss-1 clips the matrices and sums them up to obtain  $X$ . Next it  
 1279 extracts  $V_k$  the top  $k$  eigenvectors of  $X$  via an eigenvalue decomposition and privatizes its eigengap:  
 1280  $g_k = \lambda_k - \lambda_{k+1} + z$ , where  $z \sim \text{Lap}(2/\varepsilon)$ . It then applies the Gaussian mechanism to  $V_k$ :  
 1281  $V'_k = V_k + E$ , where  $E$  is a symmetric matrix with their upper triangle values (including its diagonal)  
 1282 i.i.d. sampled from  $\mathcal{N}(0, \Delta_2^2 \mathbf{I}_d)$  and

$$\Delta_2 = \frac{\beta^2(1 + \sqrt{2 \log(1/\delta)})/\varepsilon}{|g_k - 2(1 + \log(1/\delta))/\varepsilon|}.$$

1283 Finally it applies another eigenvalue decomposition on  $V'_k$  (as the noised matrix may no longer  
 1284 contain orthogonal columns) and releases the resulting  $k$  top eigenvectors. We want to note here  
 1285 that if  $g_k$  is not positive this is not private, even though we follow the algorithm stated in the paper  
 1286 (see Algorithm 2 [Dwork et al., 2014b]). A fully correct implementation (also mentioned in the  
 1287 paper) would be to use PTR at the cost of more privacy. For the sake of simplicity and to give more  
 1288 flexibility, we simply sample fresh noise if  $g_k \leq 0$ .

### 1289 G.1 Synthetic Data

1290 We sample data from the spiked model meaning, the  $A_i \in \mathbb{R}^{d \times d}$  consist of a rank  $k$  component that  
 1291 is deterministic, plus random noise that leads to the  $A_i$  being full rank. For  $k = 1$  we implement  
 1292 this by sampling  $x_i = s_i + n_i$ , where  $s_i \sim \text{Unif}(\{\lambda_1 v, -\lambda_1 v\})$  with  $v$  a unit vector,  $\lambda_1$  a scalar, and  
 1293  $n_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ . And finally defining  $A_i = x_i x_i^\top$ . We note that  $\lambda_1$  and  $\sigma$  are inputs to the sampling  
 1294 function, whereas  $v$  is obtained by generating a random vector of dimension  $d$  and normalizing it.  
 1295 For  $k > 1$  we sample a random  $d \times k$  matrix, perform a Gram Schmidt to obtain  $V_k$  a  $d \times k$  matrix  
 1296 with  $k$  orthogonal columns, and define  $A_i = V\Lambda V^\top + z_i z_i^\top$  where  $z_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ . Where  $\Lambda$  is a  
 1297  $k \times k$  diagonal matrix, whose diagonal entries are inputs to the sampling function.

1298 We set  $\beta = C\sqrt{\lambda_1} + \sigma\sqrt{d \log(n/\zeta)}$  for DP-Gauss-1 and DP-Gauss-2, where  $n$  is the number  
 1299 of samples,  $1 - \zeta$  is the probability of not clipping. We set  $\zeta = 0.01$  for both our algorithms  
 1300 (MODIFIEDDP-PCA and k-DP-Ojas) as well as for both Gauss algorithms. For k-DP-PCA and  
 1301 k-DP-Ojas we also need to give the parameters  $K$  and  $a$  (see Assumption A) as inputs. For data as  
 1302 described above we have  $a = 1$  and  $K = O(1)$ , so we set  $a = 1 = K$ . Further, for k-DP-PCA we set  
 1303 a batchsize  $B$ , which is used in the PRIVMEAN Algorithm. The theory suggests for the optimal batchsize  
 1304 should be  $n/\log^3(n)$ , where  $n$  is the sample size, however, we found empirically that  $B = \sqrt{n}$  gave  
 1305 better results. Lastly, we need to set a learning rate for k-DP-PCA and k-DP-Ojas. For k-DP-PCA we  
 1306 set the learning rates to be

$$\eta_t^i = 1/(20\sigma\lambda_i + (\lambda_i - \lambda_{i+1}) \cdot t/\log(n))$$

where  $t$  refers to the  $t$ th update step inside of MODIFIEDDP-PCA ( $t \in T = \lfloor n/B \rfloor$ ) and  $i$  to the  $i$ th iteration of  $k$ -DP-PCA. For  $k$ -DP-Ojas we empirically found that simply choosing a decreasing learning rate (independent of eigenvalues) resulted in good performance, so we set the learning rate to be

$$\eta_j = 1/(1+j)$$

for  $j \in [n]$  for all  $k$  iterations of  $k$ -DP-Ojas.

## G.2 Gaussian Data

For more general data distributions—that is, those that do not follow a clean signal-plus-noise structure—Corollary 1 suggests that  $k$ -DP-PCA can still outperform existing state-of-the-art algorithms, primarily due to its more favorable dependence on the ambient dimension  $d$ . However, our second algorithm  $k$ -DP-Ojas gives similar utility guarantees for those cases (Corollary 4). Given its simplicity — no hyperparameter tuning and stable behavior under learning rate variations — we recommend it in these scenarios. As shown in Figure 3,  $k$ -DP-Ojas consistently outperforms other state-of-the-art methods when applied to data of the form  $A_i = x_i x_i^\top$  with  $x_i \sim \mathcal{N}(0, \Sigma)$ .

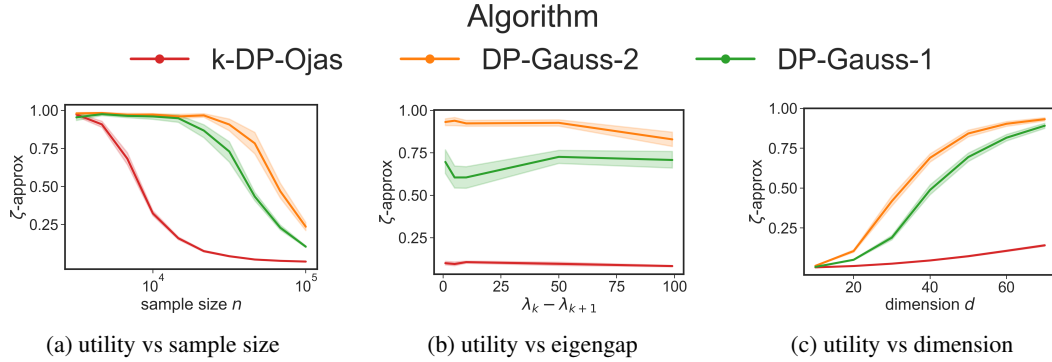


Figure 3:  $k$ -DP-Ojas

## G.3 Further comments

Lastly we would like to comment on the algorithm DP-Gauss-non-private. In order to compare it to  $k$ -DP-PCA, we would need to privately estimate  $\max\{\|x_i\|_2^2\} = \max\{\text{Tr}(x_i x_i^\top)\}$ , this is comparable to privately estimating the truncation threshold for the clipping in the private mean estimation (Algorithm 7) of  $k$ -DP-PCA. To simplify, we non privately scale the  $A_i = x_i x_i^\top$  by  $\max\{\|x_i\|_2^2\}$  and fix our truncation threshold to  $C(\lambda_1 + \sigma^2)\sigma \log(Bd/\zeta)$ . We see in Figure 4 how DP-Gauss-non-private performs when compared in this manner. Fixing the truncation threshold avoids the need to privately estimate it from data, which in turn eliminates the substantial sample complexity required to perform this estimation reliably under differential privacy. This explains the lower bound on sample size in  $k$ -DP-PCA: a minimum number of samples is needed to ensure that the truncation threshold can be estimated in a meaningful and privacy-preserving way. Interestingly, this also highlights why the algorithm may perform better in practice than the theoretical upper bounds

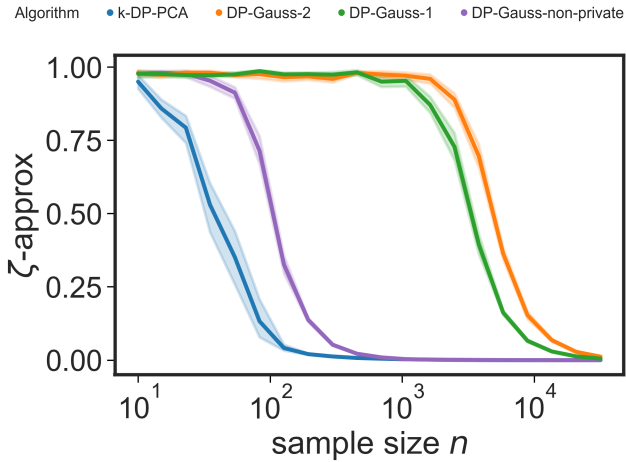


Figure 4: Top: comparing to semi-private algorithm ( $k = 2, d = 100, \sigma = 1e-4, \lambda_1 = 10, \lambda_2 = 5$ ). Bottom: [your second figure’s description here].

1346 suggest. In particular, if we use fewer samples to estimate the truncation threshold—below the level  
 1347 required for utility guarantees—k-DP-PCA might still perform well empirically, although we would  
 1348 no longer be able to guarantee its utility in a formal sense.

## 1349 H Algorithms used in Modified DP-PCA

1350 Below we describe the two subroutines that estimate the eigenvalue and mean of the gradients.

---

**Algorithm 6** Top-Eigenvalue-Estimation, Algorithm 4 in [Liu et al., 2022a]

---

**Input:**  $S = \{g_i\}_{i=1}^B$ , privacy parameters  $(\varepsilon, \delta)$ , failure probability  $\tau \in (0, 1)$

- 1:  $\tilde{g}_i \leftarrow g_{2i} - g_{2i-1}$  for  $i \in 1, 2, \dots, \lfloor B/2 \rfloor$
- 2:  $\tilde{S} = \{\tilde{g}_i\}_{i=1}^{\lfloor B/2 \rfloor}$
- 3: Partition  $\tilde{S}$  into  $k = C_1 \log(1/(\delta\tau))/\varepsilon$  subsets and denote each dataset as  $G_j \in \mathbb{R}^{d \times b}$  (where  $b = \lfloor B/2k \rfloor$  is the size of the dataset)
- 4:  $\lambda_1^{(j)} \leftarrow$  top eigenvalue of  $(1/b)G_j G_j^\top$  for all  $j \in [k]$
- 5:  $\Omega \leftarrow \{\dots, [2^{-2/4}, 2^{-1/4}), [1, 2^{1/4}), \dots\}$
- 6: run  $(\varepsilon, \delta)$ -DP histogram learner on  $\Omega$
- 7: **if** all bins are empty **then**
- 8:     **return**  $\perp$
- 9: **else**
- 10:    for  $[l, r]$  the bin with the maximum number of points
- 11:     **return**  $\hat{\Lambda} = l$
- 12: **end if**

---



---

**Algorithm 7** Private-Mean-Estimation, Algorithm 5 in [Liu et al., 2022a]

---

**Input:**  $S = \{g_i\}_{i=1}^B$ , privacy parameters  $(\varepsilon, \delta)$ , target error  $\alpha$ , failure probability  $\tau \in (0, 1)$ , approximate top eigenvalue  $\hat{\Lambda}$

- 1: let  $\tau = 2^{1/4} K \sqrt{\hat{\Lambda}} \log^2(25)$
- 2: **for**  $j = 1, 2, \dots, d$  **do**
- 3:    Run  $(\frac{\varepsilon}{4\sqrt{2d \log(4/\delta)}}, \frac{\delta}{4d})$ -DP histogram learner of Lemma on  $\{g_{ij}\}_{i \in [B]}$
- 4:    Let  $[l, h]$  be the bucket that contains maximum number of points in the private histogram
- 5:     $\bar{g}_j \leftarrow l$
- 6:    Truncate the  $j$ -th coordinate of gradient  $\{g_i\}_{i \in [B]}$  by  $[\bar{g}_j - 3K\sqrt{\hat{\Lambda}} \log^a(BD/\tau), \bar{g}_j + 3K\sqrt{\hat{\Lambda}} \log^a(BD/\tau)]$ .
- 7:    Let  $\tilde{g}_i$  be the truncated version of  $g_i$
- 8: **end for**
- 9: Compute empirical mean of truncated gradients  $\tilde{\mu} = (1/B) \sum_{i=1}^B \tilde{g}_i$  and add Gaussian noise:

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

10: **return**  $\hat{\mu}$

---