A Related work

Our work builds upon a series of advances in private SGD [51, 7, 6, 28, 60, 80, 41] to make advance in understanding the tradeoff of privacy and sample complexity for PCA. Such tradeoffs have been studied extensively in canonical statistical estimation problems of mean (and covariance) estimation and linear regression.

Private mean estimation. As one of the most fundamental problem in private data analysis, mean estimation is initially studied under the bounded support assumptions, and the optimal error rate is now well understood. More recently, [5] considered the private mean estimation problem for k-th moment bounded distributions where the support of the data is *unbounded* and provided minimax error bound in various settings. [56] studied private mean estimation from Gaussian sample, and obtained an optimal error rate. There has been a lot of recent interests on private mean estimation under various assumptions, including mean and covariance joint estimation [50, 8], heavy-tailed mean estimation [54], mean estimation for general distributions [30, 74], distribution adaptive mean estimation [12], estimation for unbounded distribution parameters [53], mean estimation under pure differential privacy [39], local differential privacy [18, 19, 32, 47], user-level differential privacy [26], Mahalanobis distance[10] and robust and differentially private mean estimation [61, 59, 62].

Private linear regression The goal of private linear regression is to learn a linear predictor of response variable y from a set of examples $\{x_i, y_i\}_{i=1}^n$ while guarantee the privacy of the examples. Again, the work on private linear regression can be divided into two categories: deterministic and randomized. In the deterministic setting where the data is deterministically given without any probabilistic assumptions, significant advances in DP linear regression has been made [77, 57, 68, 16, 7, 83, 31, 67, 82, 71]. In the randomized settings where each example $\{\mathbf{x}_i, y_i\}$ is drawn i.i.d. from a distribution [66], [20] proposes an exponential time algorithm that achieves asymptotic consistency. [13] provides an efficient and minimax optimal algorithm under sub-Gaussian design and nearly identity covariance assumptions. Very recently, [62] for the first time gives an exponential time algorithm that achieves minimax risk for general covariance matrix under sub-Gaussian and hypercontractive assumptions. [75] gives the first computationally efficient algorithm to achieve nearly optimal risk using DP-SGD with adaptive clipping.

Private PCA without spectral gap. There is a long line of work in Private PCA [37, 38, 36, 9, 14, 55, 23, 4]. We explain the closely related ones in Section 2.3, with analysis when the covariance matrix has a spectral gap.

When there is no spectral gap, one can still learn a principal component. However, since the principal component is not unique, the error is typically measured in how much of the variance is captured in the estimated direction: $1 - \hat{v}^\top \Sigma \hat{v}/\|\Sigma\|$. [14] introduces an exponential mechanism (from [64]) which samples an estimate from a distribution $f_{\widehat{\Sigma}}(\hat{v}) = (1/C)\exp\{((\varepsilon n)/c^2)\hat{v}^\top \widehat{\Sigma}\hat{v}\}$, where C is a normalization constant to ensure that the pdf integrates to one. This achieves a stronger pure DP, i.e., $(\varepsilon, 0)$-DP, but is computationally expensive; [14] does not provide a tractable implementation and [55] provides a polynomial time implementation with time complexity at least cubic in d. This achieves error rate $1 - \hat{v}^\top \Sigma \hat{v}/\|\Sigma\| = \tilde{O}(d^2/(\varepsilon n))$ in [14, Theorem 7], which, when there is a spectral gap, translates into

$$\sin(\hat{v}, v_1)^2 \quad = \quad \tilde{O}\Big(\frac{\kappa d^2}{\varepsilon n}\Big)\,, \tag{16}$$

with high probability. Closest to our setting is the analyses in [62, Corollary 6.5] that proposed an exponential mechanism that achieves $1 - \hat{v}^\top \Sigma \hat{v}/\|\Sigma\| = \tilde{O}(\sqrt{d/n} + (d + \log(1/\delta))/(\varepsilon n))$ with high probability under (ε, δ)-DP and Gaussian samples, but this algorithm is computationally intractable. This is shown to be tight when there is no spectral gap. When there is a spectral gap, this translates into

$$\sin(\hat{v}, v_1)^2 \quad = \quad \tilde{O}\Big(\kappa\Big(\sqrt{\frac{d}{n}} + \frac{d + \log(1/\delta)}{\varepsilon n}\Big)\Big)\,. \tag{17}$$

Distributed PCA. In distributed PCA, the dataset is stored across different local servers [43, 44, 81, 33]. [43, 44, 81] consider differentially private distributed PCA under the assumption that the examples are deterministic and have norms bounded by a fixed and known constant. The algorithms

appeared in [43, 44, 81] are based on the Gaussian mechanism [23] on local server and an aggregator in the central server. The resulting utility guarantees are the same as those from [23], which are discussed in Section 2.3.

B Preliminary on differential privacy

Lemma B.1 (Stability-based histogram [56, Lemma 2.3]). *For every $K \in \mathbb{N} \cup \infty$, domain Ω, for every collection of disjoint bins $B_1, \ldots, B_K$ defined on Ω, $n \in \mathbb{N}$, $\varepsilon \geq 0, \delta \in (0, 1/n)$, $\beta > 0$ and $\alpha \in (0, 1)$ there exists an (ε, δ)-differentially private algorithm $M : \Omega^n \to \mathbb{R}^K$ such that for any set of data $X_1, \ldots, X_n \in \Omega^n$*

1. $\hat{p}_k = \frac{1}{n} \sum_{X_i \in B_k} 1$
2. $(\tilde{p}_1, \ldots, \tilde{p}_K) \leftarrow M(X_1, \ldots, X_n)$, *and*
3. $$n \geq \min \left\{ \frac{8}{\varepsilon\beta} \log(2K/\alpha), \frac{8}{\varepsilon\beta} \log(4/\alpha\delta) \right\}$$

then,

$$\mathbb{P}(|\tilde{p}_k - \hat{p}_k| \leq \beta) \geq 1 - \alpha$$

Since we focus on one-pass algorithms where a data point is only accessed once, we need a basic parallel composition of DP.

Lemma B.2 (Parallel composition [65]). *Consider a sequence of interactive queries $\{q_k\}_{k=1}^K$ each operating on a subset S_k of the database and each satisfying (ε, δ)-DP. If S_k's are disjoint then the composition $(q_1(S_1), q_2(S_2), \ldots, q_K(S_K))$ is (ε, δ)-DP.*

We also utilize the following serial composition theorem.

Lemma B.3 (Serial composition [22]). *If a database is accessed with an $(\varepsilon_1, \delta_1)$-DP mechanism and then with an $(\varepsilon_2, \delta_2)$-DP mechanism, then the end-to-end privacy guarantee is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP.*

When we apply private histogram learner to each coordinate, we require more advanced composition theorem from [49].

Lemma B.4 (Advanced composition [49]). *For $\varepsilon \leq 0.9$, an end-to-end guarantee of (ε, δ)-differential privacy is satisfied if a database is accessed k times, each with a $(\varepsilon/(2\sqrt{2k \log(2/\delta)}), \delta/(2k))$-differential private mechanism.*

C Converse results

When privacy is not required, we know from Theorem 2.2 that under Assumptions A.1-A.3, we can achieve an error rate of $\tilde{O}(\kappa\sqrt{V/n})$. In the regime of $V = O(d)$ and $\kappa = O(1)$, $n = O(d)$ samples are enough to achieve an arbitrarily small error. The next lower bounds shows that we need $n = O(d^2)$ samples when $(\varepsilon = O(1), 0)$-DP is required, showing that private PCA is significantly more challenging than a non-private PCA, when assuming only the support and moment bounds of assumptions A.1-A.3. We provide a proof in Appendix C.3.

Theorem C.1 (Lower bound without Assumption A.4). *Let $\mathcal{M}_\varepsilon$ be 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 distributions satisfying Assumptions A.1–A.3 with $M = O(d \log n)$ and $V = O(d)$ is denoted by $\tilde{\mathcal{P}}_{(\lambda_1, \lambda_2)}$. There exists a universal constant $C > 0$ such that*

$$\inf_{\hat{v} \in \mathcal{M}_\varepsilon} \sup_{P \in \tilde{\mathcal{P}}_{(\lambda_1, \lambda_2)}} \mathbb{E}_{S \sim P^n} \left[\sin(\hat{v}(S), v_1) \right] \geq C \min \left(\frac{\kappa d^2}{\varepsilon n} \sqrt{\frac{\lambda_2}{\lambda_1}}, \sqrt{\frac{\lambda_2}{\lambda_1}} \right) . \tag{18}$$

We next provide the proofs of all the lower bounds.

C.1 Proof of Theorem 5.3 on the lower bound for Gaussian case

Our proof is based on following differentially private Fano's method [3, Corollary 4].

Theorem C.2 (DP Fano's method [3, Corollary 4])**.** *Let $\mathcal{P}$ denote family of distributions of interest and $\theta : \mathcal{P} \to \Theta$ denote the population parameter. Our goal is to estimate θ from i.i.d. samples $x_1, x_2, \ldots, x_n \sim P \in \mathcal{P}$. Let $\hat{\theta}_\varepsilon$ be an $(\varepsilon, 0)$-DP estimator. Let $\rho : \Theta \times \Theta \to \mathbb{R}^+$ be a pseudo-metric on parameter space Θ. Let $\mathcal{V}$ be an index set with finite cardinality. Define $\mathcal{P}_\mathcal{V} = \{P_v, v \in \mathcal{V}\} \subset \mathcal{P}$ be an indexed family of probability measures on measurable set $(\mathcal{X}, \mathcal{A})$. If for any $v \neq v' \in \mathcal{V}$,*

1. *$\rho(\theta(P_v), \theta(P_{v'})) \geq \tau$,*
2. *$D_{\mathrm{KL}}(P_v, P_{v'}) \leq \beta$,*
3. *$D_{\mathrm{TV}}(P_v, P_{v'}) \leq \phi$,*

then

$$\inf_{\hat{\theta}_\varepsilon} \max_{P \in \mathcal{P}} \mathbb{E}_{S \sim P^n} \left[\rho(\hat{\theta}_\varepsilon(S), \theta(P)) \right] \geq \max \left(\frac{\tau}{2} \left(1 - \frac{n\beta + \log(2)}{\log(|\mathcal{V}|)} \right), 0.4\tau \min \left(1, \frac{|\mathcal{V}|}{e^{10n\phi\varepsilon}} \right) \right) . \tag{19}$$

For our problem, we are interested in Gaussian $\mathcal{P}_\Sigma$ and metric $\rho(u, v) = \sin(u, v)$. Using Theorem C.2, it suffices to construct such indexed set $\mathcal{V}$ and the indexed distribution family $\mathcal{P}_\mathcal{V}$. We use the same construction as in [78, Theorem 2.1] introduced to prove a lower bound for the (non-private) sparse PCA problem. The construction is given by the following lemma.

Lemma C.3 ([78, Lemma 3.1.2])**.** *Let $d > 10$. For $\alpha \in (0, 1]$, there exists $\mathcal{V}_\alpha \subset \mathbb{S}_2^{d-1}$ and an absolute constant $c_1 > 0.0233$ such that for every $v \neq v' \in \mathcal{V}_\alpha$, $\alpha/\sqrt{2} \leq \|v - v'\|_2 \leq \sqrt{2}\alpha$ and $\log(|\mathcal{V}_\alpha|) \geq c_1 d$.*

Fix $\alpha \in (0, 1]$. For each $v \in \mathcal{V}_\alpha$, we define $\Sigma_v = (\lambda_1 - \lambda_2) v v^\top + \lambda_2 \mathbf{I}_d$ and $P_v = \mathcal{N}(0, \Sigma_v)$. It is easy to see that Σ_v has eigenvalues $\lambda_1 > \lambda_2 = \cdots = \lambda_n$. The top eigenvector of Σ_v is v. Using Lemma F.4, we know for any $v \neq v' \in \mathcal{V}$,

$$\frac{\alpha}{2} \leq \frac{1}{\sqrt{2}} \|v - v'\| \leq \rho(v, v') = \sqrt{1 - \langle v, v' \rangle^2} \leq \|v - v'\| \leq \sqrt{2}\alpha . \tag{20}$$

Using [78, Lemma 3.1.3], we know

$$D_{\mathrm{KL}}(P_v, P_{v'}) = \frac{(\lambda_1 - \lambda_2)^2}{\lambda_1 \lambda_2} (1 - \langle v, v' \rangle^2) \leq \frac{(\lambda_1 - \lambda_2)^2 \alpha^2}{\lambda_1 \lambda_2} . \tag{21}$$

Using Pinsker's inequality, we have

$$D_{\mathrm{TV}}(P_v, P_{v'}) \leq \sqrt{\frac{D_{\mathrm{KL}}(P_v, P_{v'})}{2}} \leq \alpha \sqrt{\frac{(\lambda_1 - \lambda_2)^2}{2\lambda_1 \lambda_2}} . \tag{22}$$

Now we set

$$\alpha := \min \left(1, \max \left(\sqrt{\frac{d c_1 \lambda_1 \lambda_2}{2n(\lambda_1 - \lambda_2)^2}}, \frac{c_1 d}{10 n \varepsilon} \sqrt{\frac{2\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)^2}} \right) \right) \tag{23}$$

Combining all cases, it follows from Theorem C.2 and $d > 10$ that there exists a constant C such that

$$\inf_{\hat{v}} \sup_{P \in \mathcal{P}_\Sigma} \mathbb{E}_{S \sim P^n} [\sin(\hat{v}(S), v_1(\Sigma))] \geq C \min \left(\left(\sqrt{\frac{d}{n}} + \frac{d}{\varepsilon n} \right) \sqrt{\frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)^2}}, 1 \right) . \tag{24}$$

C.1.1 Proof of Lemma C.3

We first point out that Lemma C.3 is a special case of [78, Lemma 3.1.2]. Here is the original statement from [78].

Lemma C.4 ([78, Lemma 3.1.2]). *Define* $\mathbb{B}_q^p(R_q) = \left\{\theta \in \mathbb{R}^p : \sum_{j=1}^p |\theta_j|^q \le R_q\right\}$. *Let* $\bar{R}_q = R_q - 1 \ge 1$ *and* $p \ge 5$. *There exists a finite subset* $\Theta_\epsilon \subset \mathbb{S}_2^{p-1} \cap \mathbb{B}_q^p(R_q)$ *and an absolute constant* $c > 0$ *such that every distinct pair* $\theta_1, \theta_2 \in \Theta_\epsilon$ *satisfies*

$$\epsilon/\sqrt{2} < \|\theta_1 - \theta_2\|_2 \le \sqrt{2}\epsilon$$

and

$$\log|\Theta_\epsilon| \ge c\left(\frac{\bar{R}_q}{\epsilon^q}\right)^{\frac{2}{2-q}}\left[\log(p-1) - \log\left(\frac{\bar{R}_q}{\epsilon^q}\right)^{\frac{2}{2-q}}\right]$$

for all $q \in [0,1]$ *and* $\epsilon \in (0,1]$.

Assume $d \ge 10$ and set $q = 0$ and $R_q = \frac{d}{8} + 1$. Lemma C.4 implies that there exists a finite subset $\mathcal{V}_\alpha \subset \mathbb{S}_2^{d-1} \cap \mathbb{B}_q^d\left(\frac{d}{8}+1\right)$ and an absolute constant c such that for $v \ne v' \in \mathcal{V}_\alpha$ satisfies

$$\frac{\alpha}{\sqrt{2}} \le \|v - v'\| \le \sqrt{2}\alpha \tag{25}$$

and

$$\log(|\mathcal{V}_\alpha|) \ge c\frac{d}{8}\left(\log(d-1) - \log(\frac{d}{8})\right) = \frac{cd}{8}\log\left(8\left(1 - \frac{1}{d}\right)\right) \ge \frac{cd}{8}\log(6.3)\ . \tag{26}$$

For completeness, we also provide a direct proof of Lemma C.3, following the proof strategy of Lemma C.4. The following lemma is a variant of classic Varshamov-Gilbert bounds that appeared in [63, Lemma 4.10]. A similar lemma can be also found in [3, Lemma 6].

Lemma C.5 ([63, Lemma 4.10]). *Let* l *be a positive integer that is at most* $k/4$. *Then there exists a subset* $\Theta \subset \{0,1\}^k$ *and absolute constant* $c' > 0.233$ *such that*

1. *For any* $w \in \Theta$, $\|w\|_0 = l$,
2. *For any* $w \ne w' \in \Theta$, $\|w - w'\|_0 \ge l/2$,
3. $\log(|\Theta|) \ge c' l \log(k/l)$.

For $d \ge 10$, let $k = d-1$ and l be an integer between 1 and $(d-1)/4$. We will choose l later. Let Θ be such a set that satisfies the conditions in Lemma C.5. Now for $\alpha \in (0,1]$, we construct $\mathcal{V}_\alpha$. Define $f : \{0,1\}^{d-1} \to \mathbb{R}^d$ as follows.

$$f(w) = \left(\sqrt{1-\alpha^2}, \frac{w\alpha}{\sqrt{l}}\right) \in \mathbb{R}^d. \tag{27}$$

Let

$$\mathcal{V}_\alpha := \{f(w) : w \in \Theta\}\ . \tag{28}$$

It is easy to see that

$$\|f(w)\| = \sqrt{1 - \alpha^2 + \|w\|^2\alpha^2/l} = 1\ . \tag{29}$$

For any $v \ne v' \in \mathcal{V}_\alpha$, if $v = f(w)$ and $v' = f(w')$, we know

$$\frac{\alpha}{\sqrt{2}} \le \|v - v'\| = \sqrt{\frac{\|w - w'\|^2\alpha^2}{l}} \le \sqrt{2}\alpha \tag{30}$$

where the last inequality follows from the fact that $\|w - w'\|_0 \le 2l$.

Note that above inequalities hold for any l between 1 and $(d-1)/4$. Let $l = (d-1)/8$. Then we have

$$\log(|\mathcal{V}_\alpha|) = \log(|\Theta|) \ge c'((d-1)/8)\log\left(\frac{d-1}{(d-1)/8}\right) \ge \frac{c'd}{10} \tag{31}$$

for any $d \ge 2$.

C.2 Proof of Theorem 5.4

We first construct an indexed set $\mathcal{V}$ and indexed distribution family $\mathcal{P}_{\mathcal{V}}$ such that $x_i x_i^\top$ satisfies A.1, A.2 and A.3 in Assumption 1. Our construction is defined as follows.

By [3, Lemma 6] , there exists a finite set $\mathcal{V} \subset \mathbb{S}_2^{d-1}$, with cardinality $|\mathcal{V}| \geq 2^d$, such that for any $v \neq v' \in \mathcal{V}$, $\|v - v'\| \geq 1/2$.

Let $f_{(0,\mathbf{I}_d)}$ denotes the density function of $\mathcal{N}(0, \mathbf{I}_d)$. Let Q_v be a uniform distribution on two point masses $\{\pm\alpha^{-\frac{1}{4}} v\}$. Let Q_0 be Gaussian distribution $\mathcal{N}(0, \mathbf{I}_d)$. For $\alpha \in (0, 1]$, we construct $P_v := (1-\alpha)Q_0 + \alpha Q_v$. It is easy to see that P_v is a distribution over $\mathbb{R}^d$ with the following density function.

$$P_v(x) = \begin{cases} \frac{\alpha}{2}, & \text{if } x = -\alpha^{-\frac{1}{4}} v \ , \\ \frac{\alpha}{2}, & \text{if } x = \alpha^{-\frac{1}{4}} v \ , \\ (1-\alpha) f_{(0,\mathbf{I}_d)}(x) & \text{otherwise} \end{cases} . \tag{32}$$

The mean of P_v is 0. The covariance of P_v is $\Sigma_v = (1-\alpha)\mathbf{I}_d + \sqrt{\alpha} v v^\top$. The top eigenvalue is $\lambda_1 = 1 - \alpha + \sqrt{\alpha}$, the top eigenvector is v, and the second eigenvalue is $\lambda_2 = 1 - \alpha$. And $\kappa = O(\alpha^{-1/2})$.

If $x = \alpha^{-1/4} v$, then $\|xx^\top - \Sigma_v\|_2 = O(\alpha^{-1/2})$. If $x \sim \mathcal{N}(0, \mathbf{I}_d)$, we know $\|xx^\top - \Sigma_v\|_2 = O(d)$. This implies P_v satisfies A.2 in Assumption 1 with $M = O((d + \alpha^{-1/2}) \log(n))$ for n i.i.d. samples.

It is easy to see that $\|\mathbb{E}[(xx^\top - \Sigma_v)(xx^\top - \Sigma_v)^\top]\|_2 = O(d)$. This means P_v satisfies A.3 in Assumption 1 with $V = O(d)$.

By the fact that $\mathbb{E}[\langle x, u\rangle^2] = O(1)$ and $\mathbb{E}[\langle x, u\rangle^4] = O(1)$ for any unit vector u, we have $\gamma^2 = \|\mathbb{E}[(xx^\top - \Sigma_v) u u^\top (xx^\top - \Sigma_v)^\top]\|_2 = O(1)$ for any unit vector u.

Our proof technique is based on following lemma.

Lemma C.6 ([5, Theorem 3])**.** *Fix $\alpha \in (0, 1]$. Define $P_v = (1-\alpha)Q_0 + \alpha Q_v$ for $v \in \mathcal{V}$ such that such that $\rho(\theta(P_v), \theta(P_{v'})) \geq 2t$. Let $\hat{\theta}$ be a (ε, δ) differentially private estimator. Then,*

$$\frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} P_v \left(\rho\left(\widehat{\theta}, \theta(P_v)\right) \geq t \right) \geq \frac{(|\mathcal{V}| - 1) \cdot \left(\frac{1}{2} e^{-\varepsilon \lceil n\alpha \rceil} - \delta \frac{1 - e^{-\varepsilon \lceil n\alpha \rceil}}{1 - e^{-\varepsilon}} \right)}{1 + (|\mathcal{V}| - 1) \cdot e^{-\varepsilon \lceil n\alpha \rceil}} . \tag{33}$$

Set $\rho(\theta(P_v), \theta(P_{v'})) = \sin(v, v')/\kappa$. By Lemma F.4, $\rho(\theta(P_v), \theta(P_{v'})) \geq \|v - v'\|/\kappa = \Omega(\sqrt{\alpha})$.

Lemma C.6 implies

$$\sup_{P \in \tilde{\mathcal{P}}} \mathbb{E}_{S \sim P^n}[\sin(\hat{v}(S), v_1(\Sigma))] \geq \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \mathbb{E}_{S \sim P_v^n}[\sin(\hat{v}(S), v_1(\Sigma_v))] \tag{34}$$

$$= \kappa t \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} P_v \left(\frac{\sin(\hat{v}(S), v_1(\Sigma_v))}{\kappa} \geq t \right) \tag{35}$$

$$\gtrsim \kappa t \frac{(2^d - 1) \cdot \left(\frac{1}{2} e^{-\varepsilon \lceil n\alpha \rceil} - \frac{\delta}{1 - e^{-\varepsilon}} \right)}{1 + (2^d - 1) e^{-\varepsilon \lceil n\alpha \rceil}} , \tag{36}$$

For $d \geq 2$, we know $2^d - 1 \geq e^{d/2}$. We choose

$$\alpha = \min \left\{ \frac{1}{n\varepsilon} \left(\frac{d}{2} - \varepsilon \right), \frac{1}{n\varepsilon} \log \left(\frac{1 - e^{-\varepsilon}}{4\delta e^{\varepsilon}} \right), 1 \right\} . \tag{37}$$

This implies

$$\frac{1}{2} e^{-\varepsilon \lceil n\alpha \rceil} - \frac{\delta}{1 - e^{-\varepsilon}} \geq \frac{1}{4} e^{-\varepsilon(n\alpha + 1)} > 0 . \tag{38}$$

So we have there exists a constant C such that

$$\inf_{\hat{v}} \sup_{P\in\tilde{\mathcal{P}}} \mathbb{E}_{S\sim P^n}\left[\sin(\hat{v}(S), v_1(\Sigma))\right] \geq C\kappa\sqrt{\alpha}\frac{\frac{1}{4}e^{d/2}e^{-\varepsilon(n\alpha+1)}}{1+e^{d/2}e^{-\varepsilon(n\alpha+1)}} \tag{39}$$

$$\gtrsim \kappa \min\left(1, \sqrt{\frac{d\wedge\log\left(\left(1-e^{-\varepsilon}\right)/\delta\right)}{n\varepsilon}}\right) . \tag{40}$$

C.3 Proof of Theorem C.1

Similar to the proof of Theorem 5.3, we use DP Fano's method in Theorem C.2. It suffices to construct an indexed set $\mathcal{V}$ and indexed distribution family $\mathcal{P}_{\mathcal{V}}$ such that $x_i x_i^\top$ satisfies A.1, A.2 and A.3 in Assumption 1. Our construction is defined as follows.

Let $\lambda_1 > \lambda_2 > 0$. By Lemma C.3, there exists a finite set $\mathcal{V}_\alpha \subset \mathbb{S}_2^{d-1}$, with cardinality $|\mathcal{V}_\alpha| = 2^{\Omega(d)}$, such that for any $v \neq v' \in \mathcal{V}_\alpha$, $\alpha/\sqrt{2} \leq \|v - v'\| \leq \sqrt{2}$, where $\alpha := \sqrt{\lambda_2/\lambda_1}$.

Let $f_{(0,S)}$ denotes the density function of $\mathcal{N}(0, S)$. We construct P_v over $\mathbb{R}^d$ for $v \in \mathcal{V}_\alpha$ with the following density function.

$$P_v(x) = \begin{cases} \frac{1-\lambda_2/\lambda_1}{2d}, & \text{if } x = -\sqrt{d\lambda_1}v\ , \\ \frac{1-\lambda_2/\lambda_1}{2d}, & \text{if } x = \sqrt{d\lambda_1}v\ , \\ 1 - \frac{1-\lambda_2/\lambda_1}{d}\ f_{(0, \frac{\lambda_2}{1-\frac{1-\lambda_2/\lambda_1}{d}}\mathbf{I}_d)}(x) & \text{otherwise} \end{cases} . \tag{41}$$

The mean of P_v is 0. The covariance of P_v is $\Sigma_v := (\lambda_1 - \lambda_2)vv^\top + \lambda_2\mathbf{I}_d$. It is easy to see that the top eigenvalue is λ_1, the top eigenvector is v, and the second eigenvalue is λ_2.

If $x = \sqrt{d\lambda_1}v$, then $\|xx^\top - \Sigma_v\|_2 = \|(d\lambda_1 - \lambda_1 + \lambda_2) - \lambda_2\mathbf{I}_d\|_2 = O(d\lambda_1)$. If $x \sim \mathcal{N}(0, \frac{\lambda_2}{1-\frac{1-\lambda_2/\lambda_1}{d}}\mathbf{I}_d)$, by the fact that $\frac{\lambda_2}{1-\frac{1-\lambda_2/\lambda_1}{d}} \leq \lambda_1$, we know $\|xx^\top - \Sigma_v\|_2 \leq O(d\lambda_1)$. This implies P_v satisfies A.2 in Assumption 1 with $M = O(d\log(n))$ for n i.i.d. samples.

Similarly, $\|\mathbb{E}[(xx^\top - \Sigma_v)(xx^\top - \Sigma_v)^\top]\|_2 \leq \|d(\lambda_1^2 - \lambda_1\lambda_2)vv^\top + d\lambda_2\lambda_1 + 3\Sigma_v\Sigma_v^\top\|_2 = O(d\lambda_1^2)$. This means P_v satisfies A.3 in Assumption 1 with $V = O(d)$.

For $v \neq v' \in \mathcal{V}_\alpha$, we have $D_{\mathrm{TV}}(P_v, P_{v'}) = (1 - \lambda_2/\lambda_1)/d$. By Lemma F.4, $\sin(v, v') \geq \|v - v'\|/\sqrt{2} \geq (\sqrt{\lambda_2/\lambda_1})/2$.

By Theorem C.2, there exists a constant C such that

$$\inf_{\hat{v}} \sup_{P\in\mathcal{P}_\Sigma} \mathbb{E}_{S\sim P^n}\left[\sin(\hat{v}(S), v_1(\Sigma))\right] \geq C\min\left(\sqrt{\frac{\lambda_2}{\lambda_1}}, \frac{d^2}{n\varepsilon}\sqrt{\frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_2)^2}}\right) . \tag{42}$$

D The analysis of Private Oja's Algorithm

We analyze Private Oja's Algorithm in Algorithm 2.

D.1 Proof of privacy in Lemma 3.1

We use following Theorem D.1 to prove our privacy guarantees.

Theorem D.1 (Privacy amplification by shuffling [29, Theorem 3.8])**.** *For any domain $\mathcal{D}$, let $\mathcal{R}^{(i)} : \mathcal{S}^{(1)} \times \cdots \times \mathcal{S}^{(i-1)} \times \mathcal{D} \to \mathcal{S}^{(i)}$ for $i \in [n]$ (where $\mathcal{S}^{(i)}$ is the range space of $\mathcal{R}^{(i)}$) be a sequence of algorithms such that $\mathcal{R}^{(i)}(z_{1:i-1}, \cdot)$ is an $(\varepsilon_0, \delta_0)$-DP local randomizer for all values of auxiliary inputs $z_{1:i-1} \in \mathcal{S}^{(1)} \times \cdots \times \mathcal{S}^{(i-1)}$. Let $\mathcal{A}_S : \mathcal{D}^n \to \mathcal{S}^{(1)} \times \cdots \times \mathcal{S}^{(n)}$ be the algorithm that given a dataset $x_{1:n} \in \mathcal{D}^n$, samples a uniform random permutation π over $[n]$, then sequentially computes $z_i = \mathcal{R}^{(i)}(z_{1:i-1}, x_{\pi(i)})$ for $i \in [n]$ and outputs $z_{1:n}$. Then for any $\delta \in [0, 1]$ such that $\varepsilon_0 \leq \log\left(\frac{n}{16\log(2/\delta)}\right)$, $\mathcal{A}_s$ is $(\varepsilon, \delta + O(e^\varepsilon\delta_0 n))$-DP, where*

$$\varepsilon = O\left((1 - e^{-\varepsilon_0})\left(\frac{\sqrt{e^{\varepsilon_0}\log(1/\delta)}}{\sqrt{n}} + \frac{e^{\varepsilon_0}}{n}\right)\right) . \tag{43}$$

Let $\mathcal{R}^{(t)}(w_{t-1}, A_{\pi(t)}) := w_t$. Let $\varepsilon_0 = \frac{\sqrt{2\log(1.25/\delta_0)}}{\alpha}$. We show $\mathcal{R}^{(t)}(w_{t-1}, \cdot)$ is an $(\varepsilon_0, \delta_0)$-DP local randomizer.

If there is no noise in each update step, the update rule is

$$w_t' \leftarrow w_{t-1} + \eta_t \mathrm{clip}_\beta\left(A_t w_{t-1}\right) \ , \tag{44}$$

$$w_t \leftarrow w_{t-1}/\|w_{t-1}\| \tag{45}$$

The sensitivity of w_t' is $2\beta\eta_t$ with respect to a difference in A_t. By Gaussian mechanism in Lemma 2.4 and post processing property of local differential privacy, we know w_t is $(\varepsilon_0, \delta_0)$-DP local randomizer.

Assume that $\varepsilon_0 = \frac{\sqrt{2\log(1.25/\delta_0)}}{\alpha} \le \frac{1}{2}$. By Theorem D.1, for $\hat{\delta} \in [0,1]$ such that $\varepsilon_0 \le \log\left(\frac{n}{16\log(2/\hat{\delta})}\right)$, Algorithm 2 is $(\hat{\varepsilon}, \hat{\delta} + O(e^{\hat{\varepsilon}}\delta_0 n))$-DP and for some constant $c_1 > 0$,

$$\hat{\varepsilon} \le c_1\left((1-e^{-\varepsilon_0})\left(\frac{\sqrt{e^{\varepsilon_0}\log(1/\hat{\delta})}}{\sqrt{n}} + \frac{e^{\varepsilon_0}}{n}\right)\right) \tag{46}$$

$$\le c_1\left((e^{0.5} - e^{-0.5\varepsilon_0})\frac{\sqrt{\log(1/\hat{\delta})}}{\sqrt{n}} + \frac{e^{\varepsilon_0}-1}{n}\right) \tag{47}$$

$$\le c_1\left(((1+\varepsilon_0) - (1-\varepsilon_0/2))\frac{\sqrt{\log(1/\hat{\delta})}}{\sqrt{n}} + \frac{1+2\varepsilon_0-1}{n}\right) \tag{48}$$

$$= c_1\varepsilon_0\left(\frac{1}{2}\sqrt{\frac{\log(1/\hat{\delta})}{n}} + \frac{2}{n}\right) \tag{49}$$

$$\le c_2\frac{\sqrt{\log(1/\delta_0)}}{\alpha}\sqrt{\frac{\log(1/\hat{\delta})}{n}} \ , \tag{50}$$

for some absolute constant $c_2 > 0$.

Set $\hat{\delta} = \delta/2$, $\delta_0 = c_3\delta/(e^{\hat{\varepsilon}}n)$ for some $c_3 > 0$ and $\alpha = C'\log(n/\delta)/(\varepsilon\sqrt{n})$. We have

$$\hat{\varepsilon} \le c_2\frac{\sqrt{\log(e^{\hat{\varepsilon}}n/(c_3\delta))}}{\alpha}\sqrt{\frac{\log(2/\delta)}{n}} \tag{51}$$

$$= \frac{\sqrt{\log(e^{\hat{\varepsilon}}n/(c_3\delta))\log(2/\delta)}}{C'\log(n/\delta)}\cdot\varepsilon. \tag{52}$$

For any $\varepsilon \le 1$, by Eq. (52), there exists some sufficiently large $C' > 0$ such that $\hat{\varepsilon} \le \varepsilon$.

Recall that we assume $\varepsilon_0 = \frac{\sqrt{2\log(1.25/\delta_0)}}{\alpha} \le \frac{1}{2}$. This means $\varepsilon = O(\sqrt{\frac{\log(n/\delta)}{n}})$.

D.2 Proof of clipping in Lemma 3.2

Let $z_t = A_t w_{t-1}$. Let $\mu_t := \mathbb{E}[z_t] = \Sigma w_{t-1}$. By Lemma 2.1, we know for any $\|v\| = 1$, with probability $1-\zeta$,

$$|v^\top(z_t - \mu_t)| \le K\gamma\lambda_1\log^a(2/\zeta) \ . \tag{53}$$

Applying union bound over all basis vectors $v \in \{e_1, \ldots, e_d\}$ and all samples, we know with probability $1-\zeta$, for all $j \in [d]$ and $t \in [n]$

$$|z_{t,j}| \le K\gamma\lambda_1\log^a(2nd/\zeta) + \lambda_1 \ . \tag{54}$$

This implies that with probability $1-\zeta$, for all $t \in [n]$, we have

$$\|z_t\| \le (K\gamma\log^a(2nd/\zeta) + 1)\lambda_1\sqrt{d} \ . \tag{55}$$

D.3 Proof of utility in Theorem 3.3

Lemma 3.2 implies that with probability $1 - O(\zeta)$, Algorithm 2 does not have any clipping. Under this event, the update rule becomes

$$w_t' \leftarrow w_{t-1} + \eta_t \left(A_t + 2\alpha\beta G_t\right) w_{t-1} \tag{56}$$

$$w_t \leftarrow w_t' / \|w_t'\| \,, \tag{57}$$

where $\beta = (K\gamma \log^a(nd/\zeta) + 1)\lambda_1\sqrt{d}$ and each entry in $G_t \in \mathbb{R}^{d\times d}$ is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. This follows form the fact that $\|w_{t-1}\| = 1$ and $G_t w_{t-1} \sim \mathcal{N}(0, \mathbf{I}_d)$.

Let $B_t = A_t + 2\alpha\beta G_t$. We show B_t satisfies the three conditions in Theorem 2.2 ([45, Theorem 4.12]). It is easy to see that $\mathbb{E}[B_t] = \Sigma$ from Assumption A.1. Next, we show upper bound of $\max\left\{\left\|\mathbb{E}\left[(B_t-\Sigma)(B_t-\Sigma)^\top\right]\right\|_2, \left\|\mathbb{E}\left[(B_t-\Sigma)^\top(B_t-\Sigma)\right]\right\|_2\right\}$. We have

$$\begin{aligned}
&\left\|\mathbb{E}\left[(B_t-\Sigma)(B_t-\Sigma)^\top\right]\right\|_2 \\
=\ &\left\|\mathbb{E}[(A_t + 2\alpha\beta G_t - \Sigma)(A_t + 2\alpha\beta G_t - \Sigma)^\top]\right\|_2 \\
\le\ &\left\|\mathbb{E}[(A_t-\Sigma)(A_t-\Sigma)^\top]\right\|_2 + 4\alpha^2\beta^2\|\mathbb{E}[G_t G_t^\top]\|_2 \\
\le\ &V\lambda_1^2 + 4\alpha^2\beta^2 C_2 d \,,
\end{aligned} \tag{58}$$

where the last inequality follows from Lemma F.3 and $C_2 > 0$ is an absolute constant. Let $\widetilde{V} := V\lambda_1^2 + 4\alpha^2\beta^2 C_2 d$. Similarly, we can show that $\left\|\mathbb{E}\left[(B_t-\Sigma)^\top(B_t-\Sigma)\right]\right\|_2 \le \widetilde{V}$.

By Lemma F.2, we know with probability $1-\zeta$, for all $t \in [T]$,

$$\begin{aligned}
&\|B_t - \Sigma\|_2 \\
=&\|A_t + 2\alpha\beta G_t - \Sigma\|_2 \\
\le&\|A_t - \Sigma\|_2 + 2\alpha\beta\|G_t\|_2 \\
\le&M\lambda_1 + 2C_3\alpha\beta\left(\sqrt{d} + \sqrt{\log(n/\zeta)}\right) \,.
\end{aligned}$$

Let $\widetilde{M} := M\lambda_1 + 2C_3\alpha\beta\left(\sqrt{d} + \sqrt{\log(n/\zeta)}\right)$.

Under the event that $\|B_t - \Sigma\|_2 \le \widetilde{M}$ for all $t \in [n]$, we apply Theorem 2.2 with a learning rate $\eta_t = \frac{h}{(\lambda_1-\lambda_2)(\xi+t)}$ where

$$\xi = 20\max\left(\frac{\widetilde{M}h}{(\lambda_1-\lambda_2)}, \frac{\left(\widetilde{V} + \lambda_1^2\right)h^2}{(\lambda_1-\lambda_2)^2\log(1+\frac{\zeta}{100})}\right) \,. \tag{59}$$

Then Theorem 2.2 implies that with probability $1-\zeta$,

$$\sin^2(w_n, v_1) \le \frac{C\log(1/\zeta)}{\zeta^2}\left(d\left(\frac{\xi}{n}\right)^{2h} + \frac{h^2\widetilde{V}}{(2h-1)(\lambda_1-\lambda_2)^2 n}\right) \,, \tag{60}$$

for some positive constant C.

Set $\alpha = \frac{C'\log(n/\delta)}{\varepsilon\sqrt{n}}$, the above bound implies

$$\sin^2(w_n, v_1) \le \frac{C\log(1/\zeta)}{\zeta^2}\left(\frac{h^2 V\lambda_1^2}{(2h-1)(\lambda_1-\lambda_2)^2 n} + \frac{(K\gamma\log^a(nd/\zeta)+1)^2\lambda_1^2\log^2(n/\delta)d^2h^2}{(2h-1)(\lambda_1-\lambda_2)^2\varepsilon^2 n^2} + d\left(\tilde{\xi}\right)^h\right) \,, \tag{61}$$

where $\tilde{\xi} = (\xi/n)^2$, and

$$\tilde{\xi} := \max\Bigg(\frac{M^2\lambda_1^2 h^2}{(\lambda_1-\lambda_2)^2 n^2} + \frac{(K\gamma\log^a(nd/\zeta)+1)^2\lambda_1^2\log^3(n/\delta)h^2d^2}{(\lambda_1-\lambda_2)^2\varepsilon^2 n^3}, \\ \frac{V^2\lambda_1^4 h^4}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})n^2} + \frac{(K\gamma\log^a(nd/\zeta)+1)^4\lambda_1^4\log^4(n/\delta)h^4d^4}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})\varepsilon^4 n^4} \\ + \frac{\lambda_1^4 h^4}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})n^2}\Bigg) . \tag{62}$$

For $\zeta = O(1)$ and $K = O(1)$, selecting $h = c\log n$, and assuming

$$n \geq C\Bigg(\frac{M\lambda_1\log(n)}{\lambda_1-\lambda_2} + \frac{(K\gamma\log^a(nd/\zeta)+1)^{2/3}\lambda_1^{2/3}\log(n/\delta)\log^{2/3}(n)d^{2/3}}{(\lambda_1-\lambda_2)^{2/3}\varepsilon^{2/3}} \\ + \frac{V\lambda_1^2(\log(n))^2}{(\lambda_1-\lambda_2)^2} + \frac{(K\gamma\log^a(nd/\zeta)+1)\lambda_1\log(n/\delta)\log(n)d}{(\lambda_1-\lambda_2)\varepsilon} + \frac{\lambda_1^2\log^2(n)}{(\lambda_1-\lambda_2)^2}\Bigg) , \tag{63}$$

with large enough positive constants c, and C, we have $\tilde{\xi} \leq 1$ and $d\tilde{\xi}^\alpha \leq 1/n^2$. Hence it is sufficient to have

$$n = \tilde{O}\Big(\frac{\lambda_1^2}{(\lambda_1-\lambda_2)^2} + \frac{M\lambda_1}{\lambda_1-\lambda_2} + \frac{V\lambda_1^2}{(\lambda_1-\lambda_2)^2} + \frac{d\,(\gamma+1)\lambda_1\,\log(1/\delta)}{(\lambda_1-\lambda_2)\varepsilon}\Big) ,$$

with a large enough constant.

E The analysis of DP-PCA

We provides the proofs for Theorem 5.1, Theorem 6.1, and Lemma 6.2 that guarantees the privacy and utility of DP-PCA.

E.1 Proof of Theorem 5.1 on the privacy and utility of DP-PCA

From Theorem 6.1 we know that Alg. 4 returns $\hat{\Lambda}$ satisfying $2\hat{\Lambda} \geq \lambda_1^2\|H_u\|_2$ with high probability. Then, from Lemma 6.2, we know that with high probability Alg 5 returns an unbiased estimate of the gradient mean with added Gaussian noise. Under this case, the update rule becomes

$$w_t' \leftarrow w_{t-1} + \eta_t\left(\frac{1}{B}\sum_{i=1}^{B} A_{B(t-1)+i} + \beta_t G_t\right) w_{t-1} \tag{64}$$

$$w_t \leftarrow w_t'/\|w_t'\| , \tag{65}$$

where $\beta_t = \frac{8K\sqrt{2\hat{\Lambda}_t}\log^a(Bd/\zeta)\sqrt{2d\log(2.5/\delta)}}{\varepsilon B}$, $\hat{\Lambda}_t$ denote the estimated eigenvalue of covariance of the gradients at t-th iteration, and each entry in $G_t \in \mathbb{R}^{d\times d}$ is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. This follows form the fact that $\|w_{t-1}\| = 1$ and $G_t w_{t-1} \sim \mathcal{N}(0, \mathbf{I}_d)$.

Let $\beta := \frac{16K\gamma\lambda_1\log^a(Bd/\zeta)\sqrt{2d\log(2.5/\delta)}}{\varepsilon B}$ such that $\beta \geq \beta_t$, which follows from the fact that $\hat{\Lambda} \leq \sqrt{2}\lambda_1^2\|H_u\|_2 \leq \sqrt{2}\lambda_1^2\gamma^2$ (Theorem 6.1 and Assumption A.4). Let $B_t = (1/B)\sum_{i=1}^{B} A_{B(t-1)+i} + \beta_t G_t$. We show B_t satisfies the three conditions in Theorem 2.2 ([45, Theorem 4.12]). It is easy to see that $\mathbb{E}[B_t] = \Sigma$ from Assumption A.1. Next, we show upper bound of

$\max\left\{\left\|\mathbb{E}\left[(B_t-\Sigma)(B_t-\Sigma)^\top\right]\right\|_2, \left\|\mathbb{E}\left[(B_t-\Sigma)^\top(B_t-\Sigma)\right]\right\|_2\right\}$. We have

$$\begin{aligned}
&\left\|\mathbb{E}\left[(B_t-\Sigma)(B_t-\Sigma)^\top\right]\right\|_2 \\
=&\left\|\mathbb{E}[(\frac{1}{B}\sum_{i=1}^{B}A_{B(t-1)+i}+\beta_t G_t-\Sigma)(\frac{1}{B}\sum_{i=1}^{B}A_{B(t-1)+i}+\beta_t G_t-\Sigma)^\top]\right\|_2 \\
\le&\left\|\mathbb{E}[(\frac{1}{B}\sum_{i=1}^{B}A_{B(t-1)+i}-\Sigma)(\frac{1}{B}\sum_{i=1}^{B}A_{B(t-1)+i}-\Sigma)^\top]\right\|_2+\beta^2\|\mathbb{E}[G_tG_t^\top]\|_2 \\
=&V\lambda_1^2/B+\beta^2\|\mathbb{E}[G_tG_t^\top]\|_2 \\
\le&V\lambda_1^2/B+\beta^2C_2d\,,
\end{aligned} \tag{66}$$

where the last inequality follows from Lemma F.3 and $C_2>0$ is an absolute constant. Let $\widetilde{V}:=V\lambda_1^2/B+\beta^2C_2d$. Similarly, we can show that $\left\|\mathbb{E}\left[(B_t-\Sigma)^\top(B_t-\Sigma)\right]\right\|_2\le\widetilde{V}$. By Lemma F.5 and Lemma F.2, we know with probability $1-\zeta$, for all $t\in[T]$,

$$\begin{aligned}
&\|B_t-\Sigma\|_2 \\
=&\left\|\frac{1}{B}\sum_{i=1}^{B}A_{B(t-1)+i}+\beta_tG_t-\Sigma\right\|_2 \\
\le&C_3\left(\frac{M\lambda_1\log(dT/\zeta)}{B}+\sqrt{\frac{V\lambda_1^2\log(dT/\zeta)}{B}}+\beta\left(\sqrt{d}+\sqrt{\log(T/\zeta)}\right)\right)\,.
\end{aligned}$$

Let $\widetilde{M}:=C_3\left(\frac{M\lambda_1\log(dT/\zeta)}{B}+\sqrt{\frac{V\lambda_1^2\log(dT/\zeta)}{B}}+\beta\left(\sqrt{d}+\sqrt{\log(T/\zeta)}\right)\right)$. Under the event that $\|B_t-\Sigma\|_2\le\widetilde{M}$ for all $t\in[T]$, we apply Theorem 2.2 with a learning rate $\eta_t=\frac{\alpha}{(\lambda_1-\lambda_2)(\xi+t)}$ where

$$\xi=20\max\left(\frac{\widetilde{M}\alpha}{(\lambda_1-\lambda_2)},\frac{\left(\widetilde{V}+\lambda_1^2\right)\alpha^2}{(\lambda_1-\lambda_2)^2\log(1+\frac{\zeta}{100})}\right)\,. \tag{67}$$

Then Theorem 2.2 implies that with probability $1-\zeta$,

$$\sin^2(w_T,v_1)\le\frac{C\log(1/\zeta)}{\zeta^2}\left(d\left(\frac{\xi}{T}\right)^{2\alpha}+\frac{\alpha^2\widetilde{V}}{(2\alpha-1)(\lambda_1-\lambda_2)^2T}\right)\,, \tag{68}$$

for some positive constant C. Using $n=BT$ and Eq. (66), the above bound implies

$$\sin^2(w_T,v_1)\le\frac{C\log(1/\zeta)}{\zeta^2}\left(\frac{\alpha^2V\lambda_1^2}{(2\alpha-1)(\lambda_1-\lambda_2)^2n}+\frac{K^2\gamma^2\lambda_1^2\log^{2a}(nd/(T\zeta))\log(1/\delta)d^2\alpha^2T}{(2\alpha-1)(\lambda_1-\lambda_2)^2\varepsilon^2n^2}+d\left(\tilde{\xi}\right)^{\alpha}\right)\,. \tag{69}$$

where $\tilde{\xi}=(\xi/T)^2$, and

$$\begin{aligned}
\tilde{\xi}:=\max\Bigg(&\frac{M^2\lambda_1^2\alpha^2\log^2(dT/\zeta)}{(\lambda_1-\lambda_2)^2n^2}+\frac{V\lambda_1^2\log(dT/\zeta)\alpha^2}{(\lambda_1-\lambda_2)^2nT}+\frac{K^2\gamma^2\lambda_1^2\log^{2a}(nd/(T\zeta))\log(1/\delta)\log(T/\zeta)\alpha^2d^2}{(\lambda_1-\lambda_2)^2\varepsilon^2n^2},\\
&\frac{V^2\lambda_1^4\alpha^4}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})n^2}+\frac{K^4\gamma^4\lambda_1^4\log^{4a}(nd/(T\zeta))\log^2(1/\delta)\alpha^4d^4T^2}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})\varepsilon^4n^4}\\
&+\frac{\lambda_1^4\alpha^4}{(\lambda_1-\lambda_2)^4\log^2(1+\frac{\zeta}{100})T^2}\Bigg)\,.
\end{aligned} \tag{70}$$

For $\zeta = O(1)$ and $K = O(1)$, selecting $\alpha = c \log n$, $T = c'(\log n)^2$, and assuming $\log n \geq \lambda_1^2/(\lambda_1 - \lambda_2)^2$ and

$$n \geq C \left(\frac{M\lambda_1 \log(n) \log(d \log(n))}{\lambda_1 - \lambda_2} + \frac{\sqrt{V\lambda_1^2 \log(dT)}}{(\lambda_1 - \lambda_2)} + \frac{\gamma \lambda_1 \log^a(nd/\log(n)) \sqrt{\log(1/\delta) \log(\log(n))} \log(n) d}{(\lambda_1 - \lambda_2)\varepsilon} \right.$$
$$\left. + \frac{V\lambda_1^2 (\log(n))^2}{(\lambda_1 - \lambda_2)^2} + \frac{\gamma \lambda_1 \log^a(nd/\log(n)) \sqrt{\log(1/\delta)} (\log(n))^2 d}{(\lambda_1 - \lambda_2)\varepsilon} \right) , \tag{71}$$

with large enough positive constants c, c', and C, we have $\tilde{\xi} \leq 1$ and $d\tilde{\xi}^\alpha \leq 1/n^2$. Hence it is sufficient to have

$$n = \tilde{O}\Big(\exp(\lambda_1^2/(\lambda_1 - \lambda_2)^2) + \frac{M\lambda_1}{\lambda_1 - \lambda_2} + \frac{V\lambda_1^2}{(\lambda_1 - \lambda_2)^2} + \frac{d\,\gamma\,\lambda_1 \sqrt{\log(1/\delta)}}{(\lambda_1 - \lambda_2)\varepsilon} \Big) ,$$

with a large enough constant.

E.2 Algorithm and proof of Theorem 6.1 on top eigenvalue estimation

Algorithm 4: Private Top Eigenvalue Estimation

Input: $S = \{g_i\}_{i=1}^B$, (ε, δ)-DP, failure probability ζ

Let $\tilde{g}_i \leftarrow g_{2i} - g_{2i-1}$ for $i \in 1, 2, \ldots, \lfloor B/2 \rfloor$. Let $\tilde{S} = \{\tilde{g}_i\}_{i=1}^{\lfloor B/2 \rfloor}$

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

Let $\lambda_1^{(j)}$ be the top eigenvalue of $(1/b) G_j G_j^\top$ for $\forall j \in [k]$

Partition $[0, \infty)$ into
$\Omega \leftarrow \left\{ \ldots, \left[2^{-2/4}, 2^{-1/4}\right), \left[2^{-1/4}, 1\right), \left[1, 2^{1/4}\right), \left[2^{1/4}, 2^{2/4}\right), \ldots \right\} \cup \{[0, 0]\}$

Run (ε, δ)-DP histogram learner of Lemma B.1 on $\{\lambda_1^{(j)}\}_{j=1}^k$ over Ω

if all the bins are empty **then** Return $\perp$

Let $[l, r]$ be a non-empty bin that contains the maximum number of points in the DP histogram

Return $\hat{\Lambda} = l$

Taking the difference ensures that $\tilde{g}_i$ is zero mean, such that we can directly use the top eigenvalue of $(1/b) G_j G_j^\top$ for $j \in [k]$. We compute a histogram over those k top eigenvalues. This histogram is privatized by adding noise only to the occupied bins and thresholding small entries of the histogram to be zero. The choice $k = \Omega(\log(1/\zeta)/\varepsilon)$ ensures that the most occupied bin does not change after adding the DP noise to the histograms, and $k = \Omega(\log(1/\delta)/\varepsilon)$ is necessary for handling unbounded number of bins. We emphasize that we do not require any upper and lower bounds on the eigenvalue, thanks to the private histogram learner from [11, 56] that gracefully handles unbounded number of bins.

The privacy guarantee follows from the privacy guarantee of the histogram learner provided in Lemma B.1.

For utility analysis, we follow the analysis of [53, Theorem 3.1]. The main difference is that we prove a smaller sample complexity sine we only need the top eigenvalue, and we analyze a more general distribution family. The random vector $\tilde{g}_i$ is zero mean with covariance $2\lambda_1^2 H_u \in \mathbb{R}^{d \times d}$, where $H_u = \mathbb{E}[(A_i - \Sigma) u u^\top (A_i - \Sigma)^\top]/\lambda_1^2$, and $\tilde{g}_i$ satisfies with probability $1 - \zeta$,

$$|\, \langle \tilde{g}_i, v \rangle \,| \;\leq\; 2K\lambda_1 \sqrt{\|H_u\|_2} \log^a(1/\zeta) \,, \tag{72}$$

which follows from Lemma 2.1. Applying union bound over all basis vectors $v \in \{e_1, \ldots, e_d\}$, we know with probability $1 - \zeta$,

$$\|\tilde{g}_i\| \;\leq\; 2K\lambda_1 \sqrt{d\|H_u\|_2} \log^a(d/\zeta) \,.$$

We next show that conditioned on event $\mathcal{E} = \{\|\tilde{g}_i\| \le 2K\lambda_1\sqrt{d\|H_u\|_2}\log^a(d/\zeta)\}$, the covariance $\mathbb{E}[\tilde{g}_i\tilde{g}_i^\top|\mathcal{E}]$ is close to the true covariance $\mathbb{E}[\tilde{g}_i\tilde{g}_i^\top] = 2\lambda_1^2 H_u$. Note that

$$\begin{aligned}\mathbb{E}[\tilde{g}_i\tilde{g}_i^\top|\mathcal{E}] &= \frac{\mathbb{E}[\tilde{g}_i\tilde{g}_i^\top\mathbb{I}\{\|\tilde{g}_i\| \le 2K\lambda_1\sqrt{d\|H_u\|_2}\log^a(d/\zeta)\}]}{\mathbb{P}(\mathcal{E})} \\ &\preceq \frac{\mathbb{E}[\tilde{g}_i\tilde{g}_i^\top]}{\mathbb{P}(\mathcal{E})} \preceq \frac{2\lambda_1^2 H_u}{1-\zeta}\,. \end{aligned} \tag{73}$$

We next show the empirical covariance $(1/b)\sum_{i=1}^b \tilde{g}_i\tilde{g}_i^\top$ concentrates around $2\lambda_1^2 H_u$. First of all, using union bound on Eq. (72), we have with probability $1-\zeta$, for all $i \in [b]$ and $j \in [d]$,

$$|\tilde{g}_{ij}| \le 2K\lambda_1\sqrt{\|H_u\|_2}\log^a(bd/\zeta)\,.$$

Under the event that $|\tilde{g}_{ij}| \le 2K\lambda_1\sqrt{\|H_u\|_2}\log^a(nd/\zeta)$ for all $i \in [b]$, $j \in [d]$, [79, Corrollary 6.20] together with Eq. (73) implies

$$\mathbb{P}\left(\left\|\frac{1}{b}\sum_{i=1}^b \tilde{g}_i\tilde{g}_i^\top - 2\lambda_1^2 H_u\right\|_2 \ge \alpha\right) \le 2d\exp\left(-\frac{b\alpha^2}{8K^2\lambda_1^2\|H_u\|_2\log^{2a}(\frac{bd}{\zeta})d(2\lambda_1^2\|H_u\|_2/(1-\zeta)+\alpha)}\right)\,.$$

The above bound implies that with probability $1-\zeta$,

$$\left\|\frac{1}{b}\sum_{i=1}^b \tilde{g}_i\tilde{g}_i^\top - \lambda_1^2 2H_u\right\|_2 = O\Big(K\lambda_1^2\|H_u\|_2\log^a(bd/\zeta)\sqrt{\frac{d\log(d/\zeta)}{b}} + K^2\lambda_1^2\|H_u\|_2\log^{2a}(bd/\zeta)\frac{d\log(d/\zeta)}{b}\Big)\,.$$

This means if $b = \Omega(K^2 d\log(dk/\zeta)\log^{2a}(bdk/\zeta))$, then with probability $1-\zeta$, for all $j \in [k]$, $(1-2^{1/8})\lambda_1^2\|H_u\|_2 \le \lambda_1^{(j)} \le (1+2^{1/8})\lambda_1^2\|H_u\|_2$. This means all of $\lambda_1^{(j)}$ must be within $2^{1/4}\lambda_1^2\|H_u\|_2$ interval. Thus, at most two consecutive buckets are filled with $\lambda_1^{(j)}$. By private histogram from Lemma B.1, if $k \ge \log(1/(\delta\zeta))/\varepsilon$, one of those two bins are released. The resulting total multiplicative error is bounded by $2^{1/2}$.

E.3 Algorithm and proof of Lemma 6.2 on DP mean estimation

Algorithm 5: Private Mean Estimation [56, 50]

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

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

for *j=1, 2, …, d* **do**

  Run $(\frac{\varepsilon}{4\sqrt{2d\log(4/\delta)}}, \frac{\delta}{4d})$-DP histogram learner of Lemma B.1 on $\{g_{ij}\}_{i\in[B]}$ over $\Omega = \{\cdots, (-2\tau, -\tau], (-\tau, 0], (0, \tau], (\tau, 2\tau], (2\tau, 3\tau]\cdots\}$.

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

  $\bar{g}_j \leftarrow l$

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

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

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

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

Return $\hat{\mu}$

The histogram learner is called d times, each with $(\varepsilon/(4\sqrt{2d\log(4/\delta)}), \delta/(4d))$-DP guarantee, and the end-to-end privacy guarantee is $(\varepsilon/2, \delta/2)$ from Lemma B.4 for $\varepsilon \in (0, 0.9)$. The sensitivity of the clipped mean estimate is $\Delta = \sqrt{d}6K\sqrt{\hat{\Lambda}}\log^a(Bd/\zeta)$. Gaussian mechanism with covariance $(2\Delta\sqrt{2\log(2.5/\delta)}/\varepsilon)^2\mathbf{I}_d$ satisfy $(\varepsilon/2, \delta/2)$-DP from Lemma 2.4 for $\varepsilon \in (0, 1)$. Putting these two together, with serial composition of Lemma B.3, we get the desired privacy guarantee.

The proof of utility follows similarly as [61, Lemma D.2]. Let $I_l = (l\sqrt{\hat{\Lambda}}, (l+1)\sqrt{\hat{\Lambda}}]$. Denote the population probability of j-th coordinate at I_l as $h_{j,l} = \mathbb{P}(g_{ij} \in I_l)$. Denote the empirical probability as $\hat{h}_{j,l} = \frac{1}{B}\sum_{i=1}^{B} \mathbb{I}(g_{ij} \in I_l)$. Denote the private empirical probability being released as $\tilde{h}_{j,l}$.

Fix $j \in [d]$. Let I_k be the bin that contains the μ_j. Then we know $[\mu_j - K\lambda_1\sqrt{\|H_u\|_2}\log^a(25), \mu_j + K\lambda_1\sqrt{\|H_u\|_2}\log^a(25)] \subseteq [\mu_j - \tau, \mu_j + \tau] \subset (I_{k-1} \cup I_k \cup I_{k+1})$. By Lemma 2.1, we know $\mathbb{P}(|g_{ij} - \mu_j| \geq \tau) \leq \mathbb{P}(|g_{ij} - \mu_j| \geq K\lambda_1\sqrt{\|H_u\|_2}\log^a(25)) \leq 0.04$. This means $h_{(k-1),j} + h_{k,j} + h_{(k+1),j} \geq 0.96$ and $\min(h_{(k-1),j}, h_{k,j}, h_{(k+1),j}) \geq 0.32$.

By Dvoretzky-Kiefer-Wolfowitz inequality and an union bound over $j \in [d]$, we have that with probability $1-\zeta$, $\max_{j,l}|h_{j,l} - \hat{h}_{j,l}| \leq \sqrt{\log(d/\zeta)/B}$. Using Lemma B.1, if $B = \Omega((\sqrt{d\log(1/\delta)}/\varepsilon)\log(d/(\zeta\delta)))$, with probability $1-\zeta$, we have $\max_{j,l}|\tilde{h}_{j,l} - \hat{h}_{j,l}| \leq 0.005$. Thus, with our assumption on B, we can make sure with probability $1-\zeta$, $\max_{j,l}|\tilde{h}_{j,l} - h_{j,l}| \leq 0.01$. Then we have $\min(h_{(k-1),j}, h_{k,j}, h_{(k+1),j}) - 0.01 \geq 0.31 \geq 0.04 + 0.01 \geq \max_{l\neq k-1,k,k_1} h_{j,l} + 0.01$. This implies with probability $1-\zeta$, the algorithm must pick one of the bins from I_{k-1}, I_k, I_{k+1}. This means $|\bar{g}_j - \mu_j| \leq 2\tau \leq 2^{1.5}K\lambda_1\sqrt{\|H_u\|_2}\log^a(25)$. By tail bound of Lemma 2.1, we know for all $j \in [d]$ and $i \in [B]$, $|g_{ij} - \bar{g}_j| \leq |g_{ij} - \mu_j| + |\bar{g}_j - \mu_j| \leq 3K\lambda_1\sqrt{\|H_u\|_2}\log^a(Bd/\zeta)$. This completes our proof.

F Technical lemmas

Lemma F.1. *Let $x \in \mathbb{R}^d \sim \mathcal{N}(0, \Sigma)$. Then there exists universal constant C such that with probability $1-\zeta$,*

$$\|x\|^2 \leq C\,\mathrm{Tr}(\Sigma)\log(1/\zeta)\ . \tag{74}$$

Proof. Let $\tilde{x} := \Sigma^{-1/2}x$. Then $\tilde{x}$ is also a Gaussian with $\tilde{x} \sim \mathcal{N}(0, \mathbf{I}_d)$. By Hanson-Wright inequality ([76, Theorem 6.2.1]), there exists universal constant $c > 0$ such that with probability $1-\zeta$,

$$\|x\|^2 = \tilde{x}^\top\Sigma\tilde{x} \leq \mathrm{Tr}(\Sigma) + c(\|\Sigma\|_{\mathbf{F}} + \|\Sigma\|_2)\log(2/\zeta) \leq C\,\mathrm{Tr}(\Sigma)\log(1/\zeta)\ . \tag{75}$$

□

Lemma F.2 ([76, Theorem 4.4.5])**.** *Let $G \in \mathbb{R}^{d\times d}$ be a random matrix where each entry G_{ij} is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. Then there exists universal constant $C > 0$ such that with probability $1 - 2e^{-t^2}$, $\|G\|_2 \leq C(\sqrt{d} + t)$ for $t > 0$.*

Lemma F.3. *Let $G \in \mathbb{R}^{d\times d}$ be a random matrix where each entry G_{ij} is i.i.d. sampled from standard Gaussian $\mathcal{N}(0,1)$. Then we have $\|\mathbb{E}[GG^\top]\|_2 \leq C_2 d$ and $\|\mathbb{E}[G^\top G]\|_2 \leq C_2 d$.*

Proof. By Lemma F.2, there exists universal constant $C_3 > 0$ such that

$$\mathbb{P}\left(\|G\| \geq C_1(\sqrt{d} + s)\right) \leq e^{-s^2}, \quad \forall s > 0\ . \tag{76}$$

Then

$$\begin{aligned}
\|\mathbb{E}[GG^\top]\|_2 &\leq \mathbb{E}[\|GG^\top\|_2] && (77)\\
&\leq \mathbb{E}[\|G\|_2^2] && (78)\\
&= \int_0^\infty 2r\mathbb{P}(\|G\|_2 \geq r)dr \leq C_1 d + C_3\int_{\sqrt{d}}^\infty 2re^{-\frac{(r-\sqrt{d})^2}{2}}d && (79)\\
&= C_1(d + \sqrt{2\pi d} + 2) \leq C_2 d\ , && (80)
\end{aligned}$$

where C_2 is an absolute constant. The proof for the second claim follows similarly. □

Lemma F.4. *Let $x, y \in \mathbb{S}_2^{d-1}$. Then*

$$1 - \langle x, y \rangle^2 \leq \|x - y\|^2 \ . \tag{81}$$

If $\|x - y\|^2 \leq 2$, then

$$1 - \langle x, y \rangle^2 \geq \frac{1}{2}\|x - y\|^2 \ . \tag{82}$$

The following lemma follows from matrix Bernstein inequality [73].

Lemma F.5. *Under A.1, A.2, and A.3, in Assumption 1, with probability $1 - \zeta$,*

$$\Big\| \frac{1}{B} \sum_{i \in [B]} A_i - \Sigma \Big\|_2 \ = \ O\Big(\sqrt{\frac{\lambda_1^2 V \log(d/\zeta)}{B}} + \frac{\lambda_1 M \log(d/\zeta)}{B} \Big) \ . \tag{83}$$