GAP-FREE STREAMING PCA BEYOND RANK-ONE UPDATES: NEAR-OPTIMAL RATES AND APPLICATIONS TO DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

Streaming principal component analysis (PCA) seeks to recover a leading spectral subspace in a single pass over a data stream. We give a new analysis of the ubiquitous *Oja's algorithm* (Oja, 1982) for the most general, *gap-free* variant of this problem, where no eigengap assumptions are made on the underlying mean matrix, complemented by a nearly-matching lower bound. Prior works achieving near-optimal rates for streaming PCA either required gap assumptions (Jain et al., 2016; Huang et al., 2021), or were limited to rank-one updates (Allen-Zhu & Li, 2017; Liang, 2023). Our proof only uses a second moment bound on the individual stochastic updates, bypassing the almost sure bounds needed by prior near-optimal analyses, and the analogous offline matrix Bernstein bound. We also extend our result to a Rayleigh quotient notion of approximate PCA, addressing an open question of Jain et al. (2016). As our main application, we give gap-free differentially private PCA guarantees for sub-Gaussian data, settling Conjecture 1.1 of Brown (2026) up to logarithmic factors.

1 INTRODUCTION

Let $\mathbf{A}_1, \mathbf{A}_2, \ldots \in \mathbb{R}^{d\times d}$ be i.i.d. stochastic matrices with common mean $\mathbf{\Sigma} \succeq \mathbf{0}_{d\times d}$. In *streaming principal component analysis* (PCA), the goal is to recover a unit vector close to the largest eigenvector of $\mathbf{\Sigma}$, while processing each update only once, ideally with small space overhead.

The classical Oja's algorithm (Oja, 1982) is perhaps the simplest method for this task: starting from a (randomly initialized) unit vector $\mathbf{w}_0$, it repeatedly iterates

$$\mathbf{w}_t \leftarrow \frac{(\mathbf{I}_d + \eta_t \mathbf{A}_t)\mathbf{w}_{t-1}}{\|(\mathbf{I}_d + \eta_t \mathbf{A}_t)\mathbf{w}_{t-1}\|_2}.$$

This update can be performed using $O(d)$ auxiliary space, i.e., without storing a matrix explicitly. Oja's algorithm is extremely well-studied (Balsubramani et al., 2013; Shamir, 2016; Jain et al., 2016; Allen-Zhu & Li, 2017; Huang et al., 2021; Liang, 2023), and is known to achieve near-optimal rates of convergence in various settings, under standard regularity assumptions on the sequence $\{\mathbf{A}_t\}_{t\geq 1}$, such as a second moment bound and almost sure bound (Model 3). Notably, these are the same assumptions required by the *matrix Bernstein* concentration inequality (cf. Proposition 1), which solves the same stochastic eigenvector estimation problem near-optimally, albeit in an offline setting and using $O(d^2)$ space.

We study the most general formulation of streaming PCA, where no gap assumptions are placed on $\mathbf{\Sigma}$'s spectrum. PCA objectives become ill-conditioned when the leading eigenvalues are equal or close to equal, e.g., if $\lambda_1(\mathbf{\Sigma}) = \lambda_2(\mathbf{\Sigma})$, then recovering the leading eigenvector is not even well-defined. A common alternative in such *gap-free* settings, popularized by (Garber & Hazan, 2015; Allen-Zhu & Li, 2016; 2017), is to ask for a unit vector with little mass on eigenvectors whose eigenvalues are below $(1-\gamma)\lambda_1(\mathbf{\Sigma})$, for a parameter $\gamma \in (0, 1)$. We formalize this correlation PCA (cPCA) objective in Definition 1.

Perhaps surprisingly, all prior near-optimal rates for streaming PCA, via Oja's algorithm or otherwise, either required an eigengap assumption on $\mathbf{\Sigma}$ (Jain et al., 2016; Huang et al., 2021), or were

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

limited to the setting where every $\mathbf{A}_t$ is rank-one (Allen-Zhu & Li, 2017; Liang, 2023). This motivates our work's central question.

Does Oja's algorithm achieve near-optimal convergence for streaming PCA, without eigengap assumptions on $\mathbf{\Sigma}$ or rank restrictions on the $\{\mathbf{A}_t\}_{t\geq 1}$?

1.1 Our results

Our main result (Theorem 1) answers this question affirmatively. In fact, its convergence guarantee holds under qualitatively weaker regularity assumptions (Model 1) than used by prior work. Assuming a bound on the standard matrix variance parameter,

$$V := \max\left\{\left\|\mathbb{E}[(\mathbf{A}_t - \mathbf{\Sigma})(\mathbf{A}_t - \mathbf{\Sigma})^\top]\right\|_{\text{op}}, \left\|\mathbb{E}[(\mathbf{A}_t - \mathbf{\Sigma})^\top(\mathbf{A}_t - \mathbf{\Sigma})]\right\|_{\text{op}}\right\},$$

Theorem 1 shows that with constant probability, Oja's algorithm returns a (γ, Δ)-cPCA (i.e., has squared correlation at most Δ with the eigenspace below $(1-\gamma)\lambda_1(\mathbf{\Sigma})$), using[1]

$$\widetilde{O}\left(\frac{V}{\lambda_1^2\gamma^2\Delta} + \frac{1}{\gamma}\right)$$

online samples $\mathbf{A}_t$. The first term in the above rate is complemented with a nearly-matching lower bound in Theorem 4, Appendix D, and the second term is a consequence of the standard convergence rate of the (offline) power method, in the special deterministic setting where all $\mathbf{A}_t = \mathbf{\Sigma}$.

Interestingly, Theorem 1 holds under weaker requirements than earlier convergence analyses of Oja's algorithm. In particular, it only posits a matrix variance bound V (Model 1), and circumvents the almost sure bound (Model 3) typically used by prior works on streaming PCA, as well as the matrix Bernstein inequality. As a tradeoff, it only offers a constant success probability (more generally, Theorem 1's sample complexity scales inverse-polynomially in the failure probability ζ). In Theorem 5, we give an alternative result that leverages geometric aggregation to achieve a $\text{polylog}(\frac{1}{\zeta})$ sample complexity overhead. This result analyzes an extension of Oja's algorithm to block matrices (Algorithm 3), and requires $d \cdot \text{polylog}(\frac{1}{\zeta})$ auxiliary space.

In Appendix F, we also consider the energy objective $\mathbf{w}^\top\mathbf{\Sigma}\mathbf{w} \geq (1-\alpha)\lambda_1$ (ePCA, Definition 4). While a black-box cPCA-to-ePCA conversion (Lemma 8, Jambulapati et al. (2024)) exists, its combination with Theorem 1 leads to a suboptimal sample complexity by a factor of $\frac{1}{\alpha}$. Instead, we give a multiscale reduction-based analysis in Theorem 6 that shows Oja's algorithm returns an α-ePCA using

$$\widetilde{O}\left(\frac{V}{\lambda_1^2\alpha^2} + \frac{1}{\alpha}\right)$$

samples. Here also, the first term is complemented with a nearly matching lower bound (Corollary 2). This result addresses an open question posed by Section 6 of Jain et al. (2016).

Finally, as our main application, we consider the setting of *differentially private* PCA, i.e., where the goal is to solve PCA subject to (ε, δ)-DP (Definition 2). A prior work by Liu et al. (2022a) achieved a near-optimal convergence rate for this problem under an eigengap. We give an analogous sample bound in the gap-free setting: for publicly known ν, λ_1, Theorem 2 returns an (ε, δ)-DP estimator that is a (γ, Δ)-cPCA with high probability, using

$$\widetilde{O}\left(\frac{d\nu^4}{\gamma^2\lambda_1^2\Delta} + \frac{d\nu^2}{\varepsilon\gamma\lambda_1\sqrt{\Delta}}\right)$$

samples. Our result is stated directly under sub-Gaussianity (Definition 3). For Gaussian data, $\nu^2 = \lambda_1$, and the polynomial dependence matches Corollary 5.2 of Liu et al. (2022a), with spectral resolution γ replacing the relative eigengap. We obtain a slightly better γ dependence than Liu et al. (2022a) by avoiding minibatches, instead taking full passes to obtain an improved sensitivity tradeoff. Its analysis uses Rényi differential privacy to compose the Gaussian queries and control

[1] We use the notation $\widetilde{O}$ to suppress polylogarithmic factors in problem parameters in informal rate summaries. All of our formal theorem statements specify all parameter dependences explicitly.

Work	General updates	Gap-free	Near-optimal rate
Shamir (2016), Corollary 1	✓	✓	–
Jain et al. (2016), Theorem 3	✓	–	✓
Allen-Zhu & Li (2017), Theorem 2	–	✓	✓
Huang et al. (2021), Theorem 3	✓	–	✓
Liang (2023), Theorem 3.3	–	✓	✓
This work, Theorem 1	✓	✓	✓

Table 1: Representative streaming PCA guarantees. "Near-optimal" means the rate matches the lower bound (Theorem 4) up to logarithmic factors and low-order terms. "General updates" means no rank restrictions are placed, and only a statistical assumption (e.g., Models 1 or 3) is used.

adaptive clipping. Further, applying the ePCA analysis to the same algorithm gives, under the same ν-sub-Gaussian model, an α-ePCA with sample complexity (Theorem 3),

$$\widetilde{O}\left(\frac{d\nu^4}{\lambda_1^2\alpha^2}+\frac{d\nu^2}{\varepsilon\lambda_1\alpha}\right).$$

In particular, for Gaussian data with a publicly known λ_1, a setting where $\nu^2 = O(\lambda_1)$, our new sample complexity bound above matches the rate conjectured by Brown (2026) up to logarithmic factors. The main outstanding questions left by Theorem 3 are to remove the remaining polylogarithmic overhead, and to privately estimate λ_1 from samples.

1.2 Problem Statement

We now introduce the main problem in this paper; deferring discussion on related work and notation to Appendices A and B, respectively.

To state our main problem, we recall the following helpful definition from Jambulapati et al. (2024), which has emerged as a useful gap-free notion of PCA in the literature (Garber & Hazan, 2015; Allen-Zhu & Li, 2016; 2017).

Definition 1 (cPCA)**.** *Let $(\gamma,\Delta)\in(0,1)^2$, and let $\mathbf{\Sigma}\in\mathbb{S}^{d\times d}$. We say that a unit vector $\mathbf{v}\in\mathbb{R}^d$ is a (γ,Δ)-*cPCA *(correlation PCA) of $\mathbf{\Sigma}$ if, letting orthonormal $\mathbf{S}\in\mathbb{R}^{d\times r}$ have the same column span as the eigenspace of $\mathbf{\Sigma}$ corresponding to eigenvalues $<(1-\gamma)\lambda_1(\mathbf{\Sigma})$, $\|\mathbf{S}^\top\mathbf{v}\|_{\mathrm{F}}^2\le\Delta$.*

We now state the main statistical model we consider in this paper.

Model 1. *Fix $\lambda_1>0$ and $V>0$. Let $\{\mathbf{A}_t\in\mathbb{R}^{d\times d}\}_{t\in[n]}$ be i.i.d. with $\mathbb{E}\mathbf{A}_t=\mathbf{\Sigma}\in\mathbb{S}^{d\times d}_{\succeq\mathbf{0}}$, and*

$$\|\mathbf{\Sigma}\|_{\mathrm{op}}=\lambda_1,\quad \max\left\{\left\|\mathbb{E}\left[(\mathbf{\Sigma}-\mathbf{A}_t)(\mathbf{\Sigma}-\mathbf{A}_t)^\top\right]\right\|_{\mathrm{op}},\left\|\mathbb{E}\left[(\mathbf{\Sigma}-\mathbf{A}_t)^\top(\mathbf{\Sigma}-\mathbf{A}_t)\right]\right\|_{\mathrm{op}}\right\}\le V.$$

The main problem this paper focuses on is computing a cPCA of $\mathbf{\Sigma}$, given access to $\{\mathbf{A}_t\}_{t\in[n]}$ arising from Model 1. Our algorithms' sample complexities will depend on five parameters: (V,λ_1) from Model 1, (γ,Δ) from Definition 1, and the failure probability, denoted $\zeta\in(0,1)$.

We consider this problem in two settings: the batch setting where one can arbitrarily manipulate the $\{\mathbf{A}_t\}_{t\in[n]}$, and the streaming setting, our main focus. In the streaming setting, the $\{\mathbf{A}_t\}_{t\in[n]}$ are given in a stream, and once we receive $\mathbf{A}_t$ we can perform an update and then it is discarded from memory. The goal is to solve the cPCA problem with low external memory, ideally $O(d)$.

In Proposition 1 in Appendix B.2 we recall a baseline in the batch setting via matrix Bernstein. The main algorithm we consider for streaming PCA is Oja's algorithm (Algorithm 1).

Note that Algorithm 1 is clearly a streaming algorithm. We introduce some helpful notation:

$$\begin{aligned}\mathbf{B}_t&\coloneqq(\mathbf{I}_d+\eta_t\mathbf{A}_t)\cdots(\mathbf{I}_d+\eta_1\mathbf{A}_1),\\ \mathbf{C}_t&\coloneqq(\mathbf{I}_d+\eta_t\mathbf{\Sigma})\cdots(\mathbf{I}_d+\eta_1\mathbf{\Sigma}).\end{aligned}\tag{1}$$

With this notation, the updates in Algorithm 1 are equivalent to $\mathbf{w}_t\leftarrow\frac{\mathbf{B}_t\mathbf{g}}{\|\mathbf{B}_t\mathbf{g}\|_2}$. Further, $\mathbf{C}_t$ helps track the unnormalized iterates for the corresponding updates using the average matrix $\mathbf{\Sigma}$.

Algorithm 1: Oja$(\{\mathbf{A}_t, \eta_t\}_{t\in[n]})$

Input: $\{\mathbf{A}_t \in \mathbb{R}^{d\times d}, \eta_t > 0\}_{t\in[n]}$
$\mathbf{w}_0 \leftarrow \mathbf{g}/\|\mathbf{g}\|_2$, where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$
for $t \in [n]$ **do**
  $\mathbf{w}_t \leftarrow (\mathbf{I}_d + \eta_t \mathbf{A}_t)\mathbf{w}_{t-1}$
  $\mathbf{w}_t \leftarrow \mathbf{w}_t/\|\mathbf{w}_t\|_2$
end
return $\mathbf{w}_n$

2 Oja's Algorithm

In this section we give a new analysis of Oja's algorithm (Algorithm 1), yielding cPCA guarantees in the general gap-free setting of Model 1; we defer some proofs to Appendix C. To simplify notation, we let orthonormal $\mathbf{S}$ span the eigenspace of $\mathbf{\Sigma}$ corresponding to eigenvalues $< (1-\gamma)\lambda_1(\mathbf{\Sigma})$ (in line with Definition 1), and denote the associated orthogonal projector by $\mathbf{P} \coloneqq \mathbf{S}\mathbf{S}^\top$.

We first analyze a one-step power method, analogously to Lemma 3.1, Jain et al. (2016).

Lemma 1. *Let $\zeta \in (0, \frac{1}{3})$, let $\mathbf{B} \in \mathbb{R}^{d\times d}$ not be the all-zeroes matrix, and let $\mathbf{V} \in \mathbb{R}^{d\times r}$ be orthonormal. If $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$, then with probability $\geq 1-\zeta$,*

$$\frac{\left\|\mathbf{V}^\top \mathbf{B}\mathbf{g}\right\|_2^2}{\left\|\mathbf{B}\mathbf{g}\right\|_2^2} \leq \frac{90\log\left(\frac{1}{\zeta}\right)}{\zeta^2} \cdot \frac{\left\|\mathbf{V}\mathbf{V}^\top \mathbf{B}\right\|_{\mathrm{F}}^2}{\left\|\mathbf{B}\right\|_{\mathrm{F}}^2}.$$

In Lemmas 2 and 3, we derive bounds on the ratio in Lemma 1 for $\mathbf{V} \leftarrow \mathbf{S}$, as $\mathbf{B} \leftarrow \mathbf{B}_t$ undergoes the updates of Algorithm 1. We begin by tracking a trace-based potential.

Lemma 2. *Under Model 1 and notation (1), the iterates of Algorithm 1 satisfy, for all $t \in [n]$,*

$$\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{B}_t\mathbf{B}_t^\top\right)\right] \leq \exp\left(V\sum_{s\in[t]}\eta_s^2\right)\mathrm{Tr}\left(\mathbf{C}_t\mathbf{C}_t^\top\right).$$

Proof. Let $\lambda_j \coloneqq \lambda_j(\mathbf{\Sigma})$ for shorthand. We prove inductively that, for every $s \geq 0$ and $k \in [d]$,

$$\sum_{j\in[k]}\lambda_j\left(\mathbb{E}[\mathbf{B}_s\mathbf{B}_s^\top]\right) \leq \left(\prod_{r\in[s]}(1+V\eta_r^2)\right)\sum_{j\in[k]}\prod_{r\in[s]}(1+\eta_r\lambda_j)^2. \tag{2}$$

Then, taking $s \leftarrow t$, $k \leftarrow d$, and using $1+x \leq \exp(x)$ proves the claim, since all $\mathbf{I}_d+\eta_s\mathbf{\Sigma}$ commute.

Clearly (2) holds for $s = 0$ (where we take empty products as 1). For the inductive step, suppose that (2) holds for $s-1$ and all $k \in [d]$. Upon expanding, we have

$$\mathbb{E}[\mathbf{B}_s\mathbf{B}_s^\top] = (\mathbf{I}_d+\eta_s\mathbf{\Sigma})\mathbb{E}[\mathbf{B}_{s-1}\mathbf{B}_{s-1}^\top](\mathbf{I}_d+\eta_s\mathbf{\Sigma}) + \eta_s^2\mathbb{E}\left[(\mathbf{A}_s-\mathbf{\Sigma})\mathbb{E}[\mathbf{B}_{s-1}\mathbf{B}_{s-1}^\top](\mathbf{A}_s-\mathbf{\Sigma})^\top\right]. \tag{3}$$

We bound the two terms separately in order to apply (2). For the first term, for any PSD $\mathbf{C}, \mathbf{H}$, letting $\mathbf{Q}$ be the projector onto any top-k eigenspace of $\mathbf{CHC}$, and using $\mathbf{CQC} \preceq \mathbf{C}^2$,

$$\begin{aligned}\sum_{j\in[k]}\lambda_j(\mathbf{CHC}) &= \langle\mathbf{Q}, \mathbf{CHC}\rangle = \mathrm{Tr}(\mathbf{HCQC}) \\ &\leq \sum_{j\in[k]}\lambda_j(\mathbf{H})\lambda_j(\mathbf{CQC}) \leq \sum_{j\in[k]}\lambda_j(\mathbf{C})^2\lambda_j(\mathbf{H}).\end{aligned}$$

For the second term, observe that for every rank-k orthogonal projector $\mathbf{Q}$,

$$\mathbf{0}_{d\times d} \preceq \mathbb{E}[(\mathbf{A}_s-\mathbf{\Sigma})^\top\mathbf{Q}(\mathbf{A}_s-\mathbf{\Sigma})] \preceq V\mathbf{I}_d, \quad \mathrm{Tr}\left(\mathbb{E}\left[(\mathbf{A}_s-\mathbf{\Sigma})^\top\mathbf{Q}(\mathbf{A}_s-\mathbf{\Sigma})\right]\right) \leq kV.$$

Then by the von Neumann trace inequality,

$$\mathrm{Tr}\left(\mathbf{Q}\mathbb{E}[(\mathbf{A}_s-\mathbf{\Sigma})\mathbf{H}(\mathbf{A}_s-\mathbf{\Sigma})^\top]\right)=\mathrm{Tr}\left(\mathbf{H}\mathbb{E}[(\mathbf{A}_s-\mathbf{\Sigma})^\top\mathbf{Q}(\mathbf{A}_s-\mathbf{\Sigma})]\right)\le V\sum_{j=1}^{k}\lambda_j(\mathbf{H}).$$

By supremizing this over rank-k projectors $\mathbf{Q}$, for every $\mathbf{H}\in\mathbb{S}^{d\times d}_{\succeq\mathbf{0}}$ and $k\in[d]$,

$$\sum_{j\in[k]}\lambda_j\left(\mathbb{E}[(\mathbf{A}_s-\mathbf{\Sigma})\mathbf{H}(\mathbf{A}_s-\mathbf{\Sigma})^\top]\right)\le V\sum_{j\in[k]}\lambda_j(\mathbf{H}),$$

which bounds the second term. Combining the above displays into (3), and using the triangle inequality of the Ky Fan norm,

$$\begin{aligned}\sum_{j\in[k]}\lambda_j(\mathbb{E}[\mathbf{B}_s\mathbf{B}_s^\top])&\le\sum_{j\in[k]}\left((1+\eta_s\lambda_j)^2+V\eta_s^2\right)\lambda_j(\mathbb{E}[\mathbf{B}_{s-1}\mathbf{B}_{s-1}^\top])\\&\le(1+V\eta_s^2)\sum_{j\in[k]}(1+\eta_s\lambda_j)^2\lambda_j(\mathbb{E}[\mathbf{B}_{s-1}\mathbf{B}_{s-1}^\top])\\&\le\left(\prod_{r\in[s]}(1+V\eta_r^2)\right)\sum_{j\in[k]}\prod_{r\in[s]}(1+\eta_r\lambda_j)^2\end{aligned}$$

where the third line applies Fact 1 with

$$\mathbf{c}_j\leftarrow(1+\eta_s\lambda_j)^2,\quad\mathbf{x}_j\leftarrow\lambda_j(\mathbb{E}[\mathbf{B}_{s-1}\mathbf{B}_{s-1}^\top]),\quad\mathbf{y}_j\leftarrow\left(\prod_{r\in[s-1]}(1+V\eta_r^2)\right)\prod_{r\in[s-1]}(1+\eta_r\lambda_j)^2,$$

where $\mathbf{y}$ weakly majorizes $\mathbf{x}$ is the inductive hypothesis. Thus (2) holds as desired. □

Using Markov's inequality and centering around the progress made by the population-level (noiseless) variant of Oja's algorithm, Appendix C proves the following consequence of Lemma 2.

Lemma 3. *Under Model 1 and notation (1), let $\zeta\in(0,\frac{1}{3})$, let $t\in[n]$, and suppose $\eta_s\le\frac{1}{\lambda_1}$ for all $s\in[t]$. If we let $q_t\coloneqq V\sum_{s\in[t]}\eta_s^2$, and $q_t\le\frac{\zeta}{2}$, then with probability $\ge1-\zeta$,*

$$\frac{\|\mathbf{P}\mathbf{B}_t\|_{\mathrm{F}}^2}{\|\mathbf{B}_t\|_{\mathrm{F}}^2}\le2d\exp\left(-\gamma\lambda_1\sum_{s\in[t]}\eta_s\right)+\frac{8\left(\exp(q_t)-1\right)}{\zeta}.$$

At this point, we are ready to give our main bound in this section.

Theorem 1. *Let $\zeta\in(0,\frac{1}{3})$ and $(\gamma,\Delta)\in(0,1)^2$. Under Model 1, if*

$$n=\Omega\left(\frac{V}{\lambda_1^2\gamma^2\Delta}\cdot\frac{\log^2\left(\frac{d}{\Delta\zeta}\right)\log\left(\frac{1}{\zeta}\right)}{\zeta^3}+\frac{\log\left(\frac{d}{\Delta\zeta}\right)}{\gamma}\right),$$

for an appropriate constant, then there exists $\beta\in\mathbb{R}_{>0}$ such that taking $\eta_t=\frac{\log(\frac{1440d}{\Delta\zeta^3})}{\gamma\lambda_1(\beta+t)}$ for $t\in[n]$, the output of Algorithm 1 is a (γ,Δ)-cPCA of $\mathbf{\Sigma}$ with probability $\ge1-\zeta$.

Proof. For shorthand, denote $L\coloneqq\log(\frac{1440d}{\Delta\zeta^3})$, and for a large enough constant C, let

$$\beta\coloneqq\max\left\{\frac{8L}{\gamma},\frac{CVL^2}{\gamma^2\lambda_1^2\Delta}\cdot\frac{\log\left(\frac{1}{\zeta}\right)}{\zeta^3}\right\},$$

and $n\ge4\beta$, so that all $\eta_s\le\frac{1}{\lambda_1}$. Moreover, $\sum_{s\in[n]}\frac{1}{(\beta+s)^2}\le\int_\beta^\infty\frac{\mathrm{d}x}{x^2}=\frac{1}{\beta}$, so for large enough C,

$$q_n=V\sum_{s\in[n]}\eta_s^2\le\frac{VL^2}{\gamma^2\lambda_1^2\beta}\le\frac{\Delta}{2}\cdot\frac{\zeta^3}{32\cdot360\cdot\log\left(\frac{2}{\zeta}\right)}.$$

Thus, $q_n \le \frac{\zeta}{4}$, so Lemmas 1 and 3 both apply at failure probability $\frac{\zeta}{2}$, and combining gives

$$\frac{\left\|\mathbf{S}^\top \mathbf{B}_n \mathbf{g}\right\|_2^2}{\left\|\mathbf{B}_n \mathbf{g}\right\|_2^2} \le \frac{360 \log\left(\frac{2}{\zeta}\right)}{\zeta^2}\left(2d \exp\left(-\gamma\lambda_1 \sum_{s\in[n]} \eta_s\right) + \frac{16(\exp(q_n)-1)}{\zeta}\right),$$

with probability $\ge 1-\zeta$ over the randomness of $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$ and Model 1. Condition on this event henceforth. For the first term above, since $n \ge 4\beta$, an integral comparison gives

$$\gamma\lambda_1 \sum_{s\in[n]} \eta_s = L \sum_{s\in[n]} \frac{1}{\beta+s} \ge L \int_1^{n+1} \frac{\mathrm{d}x}{\beta+x} \ge L \log\left(\frac{\beta+n+1}{\beta+1}\right) \ge L.$$

Combining the above three displays concludes the proof, upon simplifying using $\exp(q_n)-1 \le 2q_n$, and $\log(\frac{2}{\zeta}) \le \frac{1}{\zeta}$, in the relevant parameter regimes. □

We give a high-probability variant of Theorem 1 in Theorem 5 (with polylog dependence on $\frac{1}{\zeta}$), and demonstrate the tightness of its various parameter dependences in Theorem 4.

3 Application to Differentially Private PCA

In this section, we give our application to differentially private PCA; we defer proofs to Appendix G. Given i.i.d. sub-Gaussian samples with covariance $\mathbf{\Sigma}$, we seek a cPCA (Definition 1) or ePCA (Definition 4) subject to the following privacy guarantee.

Definition 2 (Differential privacy). *We say that a randomized algorithm $\mathcal{A} : (\mathbb{R}^d)^n \to \Omega$ is (ε, δ)-differentially private if for all measurable $\mathcal{E} \subseteq \Omega$, and all $S, S' \in (\mathbb{R}^d)^n$ differing in one entry, $\Pr[\mathcal{A}(S) \in \mathcal{E}] \le \exp(\varepsilon)\Pr[\mathcal{A}(S') \in \mathcal{E}] + \delta$.*

Our utility analysis holds under the following assumption on the dataset.

Definition 3 (ν-sub-Gaussianity). *A distribution $\mathcal{D}$ on $\mathbb{R}^d$ is ν-sub-Gaussian if, for every $\mathbf{v} \in \mathbb{R}^d$,*

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}\left[\exp\left(\langle \mathbf{x} - \mathbb{E}\mathbf{x}, \mathbf{v}\rangle\right)\right] \le \exp\left(\frac{\nu^2 \left\|\mathbf{v}\right\|_2^2}{2}\right).$$

Model 2. *The samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ are drawn i.i.d. from a mean-zero, ν-sub-Gaussian distribution with covariance $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ and $\lambda_1 = \left\|\mathbf{\Sigma}\right\|_{\mathrm{op}} > 0$.*

Throughout this section, we treat n, ν, λ_1 as public parameters, fixed independently of the dataset. As is standard for statistical DP algorithms, our privacy guarantee will hold regardless of the input dataset, and our utility guarantee will hold assuming that the dataset follows Model 2.

We state our main algorithm in Algorithm 2, which is patterned off of the DP-PCA algorithm of Liu et al. (2022a). After clipping the input dataset, the algorithm simply adds an appropriate Gaussian perturbation to each iterate of an empirical power method, which we show can be cast as an instance of Model 1. One major difference between Algorithm 2 and the variant in Liu et al. (2022a) is that we do not subdivide our dataset into minibatches, and instead use full-batch iterations; this difference ends up shaving a roughly $\gamma^{-1/2}$ factor from our final sample complexity.

We now state our main private cPCA guarantee, which directly applies Theorem 5, after a coupling argument shows that the iterates of Algorithm 2 are induced by Model 1, with high probability.

Theorem 2. *Under Model 2, let $\varepsilon \in (0, 1]$, $(\delta, \zeta) \in (0, \frac{1}{3})^2$, and $(\gamma, \Delta) \in (0, 1)^2$. If*

$$n = \Omega\left(\frac{\nu^4\left(d + \log\left(\frac{1}{\zeta}\right)\right)}{\lambda_1^2\gamma^2\Delta} + \frac{d\nu^2 \log\left(\frac{1}{\zeta}\right)\log\left(\frac{2d}{\Delta}\log\left(\frac{1}{\zeta}\right)\right)\log\left(\frac{n}{\zeta}\right)\sqrt{\log\left(\frac{n}{\delta\zeta}\right)}}{\varepsilon\lambda_1\gamma\sqrt{\Delta}}\right) \tag{4}$$

for an appropriate constant, there is a choice of inputs to Algorithm 2 that gives an (ε, δ)-DP algorithm returning a (γ, Δ)-cPCA of $\mathbf{\Sigma}$ with probability $\ge 1-\zeta$.

Algorithm 2: PrivBoostedSketchOja$(\{\mathbf{x}_i\}_{i\in[n]}, \{\eta_t\}_{t\in[T]}, R, T, R_1, R_2, \tau, \sigma)$

Input: $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i\in[n]}$, $\{\eta_t > 0\}_{t\in[T]}$, $(R, T) \in \mathbb{N}^2$, $(R_1, R_2, \tau, \sigma) \in \mathbb{R}^4_{>0}$
$\mathbf{G} \leftarrow d \times R$ matrix with i.i.d. $\mathcal{N}(0, 1)$ entries
for $r \in [R]$ **do**
  $\mathbf{Y}_r \leftarrow \mathbf{G}$
end
for $t \in [T]$ **do**
  for $r \in [R]$ **do**
    $\mathbf{U}_r\mathbf{D}_r\mathbf{V}_r^\top \leftarrow$ compact SVD of $\mathbf{Y}_r$, with $s_r \coloneqq \mathrm{rank}(\mathbf{Y}_r)$
    $\mathbf{Q}_{r,t} \leftarrow \mathbf{0}_{d\times s_r}$
    for $i \in [n]$ **do**
      `// Use multiplier 1 if a denominator in either clipping step is zero.`
      $\mathbf{s}_i \leftarrow \mathbf{x}_i \min\{1, \sqrt{R_1}/\left\|\mathbf{x}_i\right\|_2\}$
      $\mathbf{z}_i \leftarrow \mathbf{s}_i \min\{1, \sqrt{R_2}/\left\|\mathbf{U}_r^\top \mathbf{s}_i\right\|_2\}$
      $\mathbf{Q}_{r,t} \leftarrow \mathbf{Q}_{r,t} + \frac{1}{n}\mathbf{z}_i(\mathbf{z}_i^\top \mathbf{U}_r)$
    end
    $\mathbf{H}_{r,t} \leftarrow d \times s_r$ matrix with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries
    $\mathbf{Y}_r \leftarrow \mathbf{Y}_r + \eta_t(\mathbf{Q}_{r,t} + \mathbf{H}_{r,t})\mathbf{D}_r\mathbf{V}_r^\top$
  end
end
for $r \in [R]$ **do**
  $\mathbf{Z}_r \leftarrow \mathbf{Y}_r/\left\|\mathbf{Y}_r\right\|_{\mathrm{F}}$, or a fixed unit-Frobenius-norm matrix if $\mathbf{Y}_r = \mathbf{0}_{d\times R}$
end
for $r \in [R]$ **do**
  $\mathcal{N}_r \leftarrow \{r' \in [R] : \left\|\mathbf{Z}_r - \mathbf{Z}_{r'}\right\|_{\mathrm{F}} \le 2\tau\}$
end
$i \leftarrow$ any index with $|\mathcal{N}_i| \ge R/2$, else $i \leftarrow 1$
$\mathbf{w}_n \leftarrow$ any top left singular vector of $\mathbf{Z}_i$
Return: $\mathbf{w}_n$

We conclude by giving an analogous private ePCA guarantee for Oja's algorithm; we provide this extension because this was the form of PCA asked about in Conjecture 1.1, Brown (2026).

Theorem 3. *Under Model 2, let $\alpha \in (0, \frac{1}{4})$, $\varepsilon \in (0, 1]$, and $\delta \in (0, \frac{1}{3})$. If*

$$n = \Omega\left(\frac{\nu^4 d}{\lambda_1^2\alpha^2} + \frac{d\nu^2 \log^2\left(\frac{nd}{\alpha}\right)\sqrt{\log\left(\frac{nd}{\alpha\delta}\right)}}{\varepsilon\lambda_1\alpha}\right) \tag{5}$$

for an appropriate constant, there are choices of the inputs to Algorithm 2 that gives an (ε, δ)-DP algorithm returning an α-ePCA of $\mathbf{\Sigma}$ with probability $\ge \frac{2}{3}$.

We remark that the success probability of Theorem 3 can be boosted using holdout samples, analogously to Corollary 1. For brevity, we omit this extension.

Remark 1 (Gaussian specialization)**.** *For Gaussian data, $\nu^2 = \lambda_1$, so Theorem 3 matches the sample complexity in Conjecture 1.1, Brown (2026) up to logarithmic factors when λ_1 is known.*

4 Experiments

We conclude by providing empirical evaluations of Algorithm 1 (Section 4.1) and Algorithm 2 (Section 4.2), to complement our theoretical results.

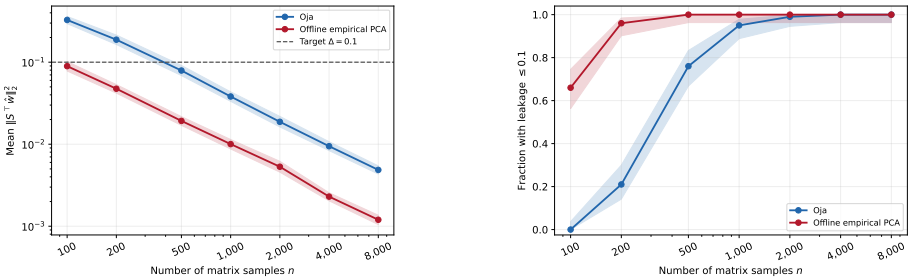


Figure 1: Oja's algorithm with $\eta_t = \frac{16}{30+t}$ and offline empirical PCA across 100 fresh paired trials. The left figure shows the cPCA error with 95% Student-t confidence intervals. The right figure shows the fraction of trials with cPCA error at most $\Delta = 0.1$ with 95% Wilson intervals.

4.1 Oja's algorithm

We first evaluate Oja's algorithm (Algorithm 1) on synthetic streams with nearly tied leading eigenvalues, by comparing it against the top principal component of the empirical covariance. Note that this empirical estimator is not applicable in streaming settings, and serves only as a baseline.

Our experiments study the performance of these two algorithms under the same sample size. In our experiments, we set $d = 50$ and set the population mean to $\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{Q}$ is a Haar-distributed orthonormal matrix. The first three eigenvalues in $\mathbf{\Lambda}$ are fixed at $(1, 0.99, 0.98)$, and the remaining eigenvalues are independently drawn from $\mathsf{Unif}(0, 0.95)$ and sorted in decreasing order. We generate our matrix stream as $\mathbf{A}_t := \mathbf{\Sigma} + \frac{1}{\sqrt{d}}\mathbf{G}_t$, with $[\mathbf{G}_t]_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. It is straightforward to check that this is an instance of Model 1 with $V = 1$.

We set $\gamma = 0.05$ and measure the cPCA success rate from Definition 1 with $\Delta = 0.1$. We implemented Algorithm 1 with $\eta_t = \frac{c}{\lambda_1(\beta+t)}$, where $\lambda_1 = 1$, and we performed a grid search for the pair of c and β that achieved the smallest mean cPCA error, over the choices $c \in \{0.5, 1, 2, 4, 8, 16, 32\}$ and $\beta \in \{0, 1, 3, 10, 30, 100\}$.

The baseline returns a principal eigenvector of $\widehat{\mathbf{\Sigma}} := \frac{1}{2n}\sum_{t\in[n]}(\mathbf{A}_t + \mathbf{A}_t^\top)$, i.e., the symmetrized empirical covariance. As shown in Figure 1, Oja's algorithm with tuned step sizes achieves results comparable to the baseline, but in the streaming setting.

4.2 Private PCA

We next evaluate Algorithm 2 for private PCA on synthetic Gaussian samples, comparing it against the AnalyzeGauss algorithm of Dwork et al. (2014), which clips samples and noises the empirical covariance matrix entrywise. Up to logarithmic factors, Theorem 2 and Theorem 6 of Dwork et al. (2014) show that the sample complexity of Algorithm 2 and AnalyzeGauss respectively scale as[2]

$$\approx \frac{d}{\gamma^2\Delta} + \frac{d}{\varepsilon\gamma\sqrt{\Delta}}, \quad \approx \frac{d}{\gamma^2\Delta} + \frac{d^{1.5}}{\varepsilon\gamma\sqrt{\Delta}}. \tag{6}$$

Observe that unless d is somewhat large or γ, Δ are somewhat small, the identical first term in each of the above expressions dominates. Thus, we expect our algorithm to have improved performance over AnalyzeGauss only in regimes with moderately large d and small γ, Δ.

In the following experiment, we set $d = 5000$ and vary $n \in \{10, 20, 30, 50, 100\} \times 10^6$. The population mean $\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ follows the exact same distribution as in Section 4.1, i.e., the first three eigenvalues are $(1, 0.99, 0.98)$, and the remaining eigenvalues are independently drawn from $\mathsf{Unif}(0, 0.95)$. At each sample size, we run 20 independent trials. We use the same cPCA error bound and success criterion of $\Delta = 0.1$ as before, and vary $\gamma \in \{\frac{1}{4}, \frac{1}{20}\}$ to measure the effect of this parameter. Finally, we set our DP parameters to $\varepsilon = 1$ and $\delta = 10^{-6}$, and as hyperparameters to

[2]To see this bound for Dwork et al. (2014) in the Gaussian setting, after adapting their notation and scaling convention to ours, plug in $n \leftarrow d$ and $\sigma_1^2 - \sigma_2^2 \leftarrow \frac{n\gamma}{d}$, and set the resulting $\sin^2$ error bound to Δ.

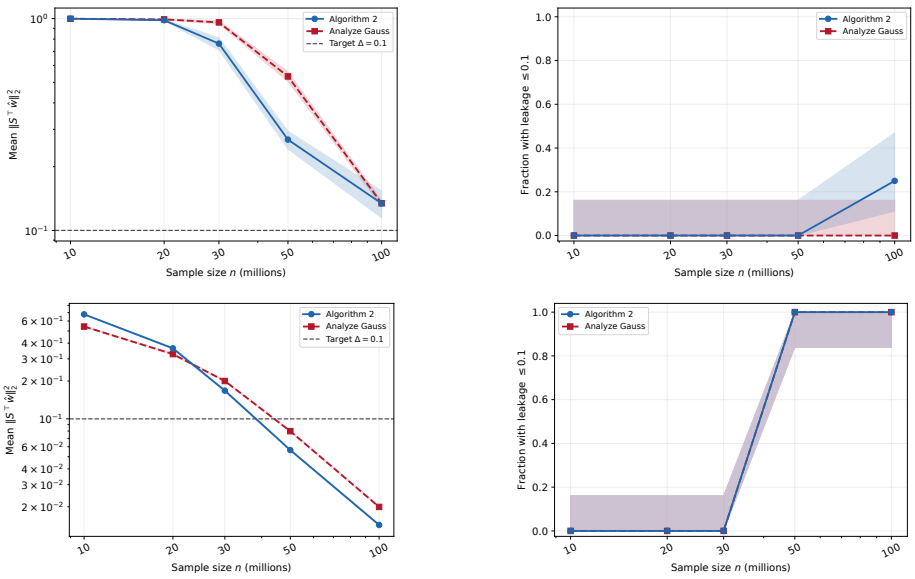


Figure 2: The left two figures show the cPCA error with 95% Student-t confidence intervals. The right two figures show the fraction of trials with cPCA error at most $\Delta = 0.1$ with 95% Wilson intervals. The top figures are for $\gamma = 0.05$ and bottom figures are for $\gamma = 0.25$.

Algorithm 2, we use $R = 3$, $T = 500$, $\eta_t = \frac{32}{100+t}$, and $\tau = \frac{1}{8}\sqrt{\frac{\Delta}{R}}$. The step sizes η_t were picked using another grid search, selected from the same choices as used in Section 4.1. We choose the clipping thresholds from (32) with $\nu = \lambda_1 = 1$ and $\zeta = 0.1$. The noise scale σ is selected according to Lemma 13, which guarantees DP.

We next briefly describe the AnalyzeGauss baseline from Dwork et al. (2014). We used the same norm clipping threshold R_1, i.e., we follow Line 11 of Algorithm 2 to produce clipped samples $\{\mathbf{s}_i\}_{i\in[n]}$. AnalyzeGauss then outputs a leading eigenvector of $\widehat{\mathbf{\Sigma}}_{\text{clip}} + \mathbf{H}$, where

$$\widehat{\mathbf{\Sigma}}_{\text{clip}} \coloneqq \frac{1}{n}\sum_{i\in[n]} \mathbf{s}_i\mathbf{s}_i^\top, \quad \sigma_{\text{AG}}^2 \coloneqq \frac{4R_1^2\log(1.25/\delta)}{n^2\varepsilon^2},$$

and $\mathbf{H}$ is a symmetric matrix with the upper triangle sampled i.i.d. from $\mathcal{N}(0, \sigma_{\text{AG}}^2)$.

In Figure 2, we show that Algorithm 2 achieves lower mean cPCA error than AnalyzeGauss under the given parameters. As expected from (6), Algorithm 2 performs better when γ is smaller. This improvement becomes less drastic when n is very large, because rearranging (6) shows that

$$\Delta \approx \max\left(\frac{d}{\gamma^2 n}, \left(\frac{d}{\varepsilon\gamma n}\right)^2\right)$$

is dominated by the first term for large n. In such regimes, our error decay matches AnalyzeGauss.

We also note that, consistently with our theory, this finding appears to require a moderately large dimension to emerge: for example, when $d = 3000$ and all other parameter settings remain fixed, AnalyzeGauss achieves lower error than our algorithm.

AI USE STATEMENT

We used ChatGPT 5.5 and 5.6 Pro to explore approaches for Theorem 1, primarily to aid with strategies for proving Lemma 2, but the final proof strategy was developed by the authors. The reduction in Proposition 2 was discovered in conversations with ChatGPT 5.6 Pro, allowing us to extend our cPCA results to ePCA. Additionally, we used ChatGPT 6 in the Codex environment to implement experiments. The manuscript was written solely by the authors, and we take responsibility for the final content of this work, including claims or artifacts produced with the aid of generative AI.

Reproducibility statement

We provide code for all experiments reported in this paper. Proofs of our theoretical results are provided in the main text and appendices.

References

Zeyuan Allen-Zhu and Yuanzhi Li. Even faster SVD decomposition yet without agonizing pain. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pp. 974–982, 2016.

Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: a global, gap-free, and near-optimal rate. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 487–492. IEEE, 2017.

Maria-Florina Balcan, Simon Shaolei Du, Yining Wang, and Adams Wei Yu. An improved gap-dependency analysis of the noisy power method. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 284–309. PMLR, 2016.

Akshay Balsubramani, Sanjoy Dasgupta, and Yoav Freund. The fast convergence of incremental PCA. In *Advances in Neural Information Processing Systems 26*, 2013.

Gavin Brown. Gap-free, computationally efficient private PCA. Open Problems, Workshop on the Intersections of Differential Privacy and Sublinear Algorithms, TTIC, 2026. Conjecture 1.1, p. 2; July 27–29, 2026.

T. Tony Cai, Dong Xia, and Mengyue Zha. Optimal differentially private PCA and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*, 2024. URL https://arxiv.org/abs/2401.03820.

Kamalika Chaudhuri, Anand D. Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14:2905–2943, 2013. URL https://jmlr.org/papers/v14/chaudhuri13a.html.

Sanjoy Dasgupta, Syamantak Kumar, Shourya Pandey, and Purnamrita Sarkar. Low precision streaming PCA. In *Advances in Neural Information Processing Systems 38*, pp. 157961–157996. Curran Associates, Inc., 2025. doi: 10.52202/085713-5279. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/e79f5871643773a83e41c06d42d81cb8-Paper-Conference.pdf.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pp. 11–20. ACM, 2014. doi: 10.1145/2591796.2591883.

Dan Garber and Elad Hazan. Fast and simple PCA via convex optimization. *CoRR*, abs/1509.05647, 2015.

Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. In *Advances in Neural Information Processing Systems 27*, pp. 2861–2869, 2014.

Samuel B. Hopkins, Gautam Kamath, Mahbod Majid, and Shyam Narayanan. Robustness implies privacy in statistical estimation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pp. 497–506. ACM, 2023. URL https://arxiv.org/abs/2212.05015.

Daniel Hsu, Sham M. Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(52):1–6, 2012.

De Huang, Jonathan Niles-Weed, and Rachel Ward. Streaming k-PCA: Efficient guarantees for Oja's algorithm, beyond rank-one updates. In *Proceedings of the Thirty-Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 2463–2498, 2021.

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

Arun Jambulapati, Syamantak Kumar, Jerry Li, Shourya Pandey, Ankit Pensia, and Kevin Tian. Black-box k-to-1-pca reductions: Theory and applications. In *The Thirty Seventh Annual Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 2564–2607. PMLR, 2024.

Minwoo Kim and Sungkyu Jung. Robust and differentially private principal component analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 18(6):e70053, 2025. doi: 10.1002/sam.70053.

Syamantak Kumar and Purnamrita Sarkar. Streaming PCA for Markovian data. In *Advances in Neural Information Processing Systems 36*, pp. 64650–64662. Curran Associates, Inc., 2023. doi: 10.52202/075280-2821. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/cbb1fa8e7f515e796cda6621a703492f-Paper-Conference.pdf.

Syamantak Kumar and Purnamrita Sarkar. Oja's algorithm for streaming sparse PCA. In *Advances in Neural Information Processing Systems 37*, pp. 74528–74578. Curran Associates, Inc., 2024. doi: 10.52202/079017-2372. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/8829f586a1ac0e6c41143f5d57b63c4b-Paper-Conference.pdf.

Syamantak Kumar, Shourya Pandey, and Purnamrita Sarkar. Beyond sin-squared error: Linear time entrywise uncertainty quantification for streaming PCA. In *Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 2396–2430. PMLR, 2025. URL https://proceedings.mlr.press/v286/kumar25a.html.

Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

Chengtao Li, Mengdi Wang, Han Liu, and Tong Zhang. Near-optimal stochastic approximation for online principal component estimation. *Mathematical Programming*, 167(1):75–97, 2018. doi: 10.1007/s10107-017-1182-z.

Xin Liang. On the optimality of Oja's algorithm for online PCA. *Statistics and Computing*, 33(3): 62, 2023. doi: 10.1007/s11222-023-10229-z.

Xiyang Liu, Weihao Kong, Prateek Jain, and Sewoong Oh. DP-PCA: Statistically optimal and differentially private PCA. In *Advances in Neural Information Processing Systems 35*, 2022a.

Xiyang Liu, Weihao Kong, and Sewoong Oh. Differential privacy and robust statistics in high dimensions. In *Proceedings of the Thirty-Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 1167–1246. PMLR, 2022b. URL https://proceedings.mlr.press/v178/liu22b.html.

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

Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain. Memory limited, streaming PCA. In *Advances in Neural Information Processing Systems 26*, 2013.

Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982. doi: 10.1007/BF00275687.

Christopher De Sa, Kunle Olukotun, and Christopher Ré. Global convergence of stochastic gradient descent for some non-convex matrix problems. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2332–2341. PMLR, 2015.

Ohad Shamir. Convergence of stochastic gradient descent for PCA. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 257–265. PMLR, 2016.

Kevin Tian. CS395T: Continuous Algorithms, Part XI: Low-Rank Approximation. Lecture notes, University of Texas at Austin, 2026. URL https://kjtian.github.io/notes/CS%20395T%20(Spring%202026)/Part11_main.pdf.

Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and trends® in machine learning*, 8(1-2):1–230, 2015.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, 2009. doi: 10.1007/b13794.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

A Related Work

Streaming and gap-free PCA. Finite-sample analyses of streaming PCA include incremental PCA (Balsubramani et al., 2013), memory-optimal block methods (Mitliagkas et al., 2013), and stochastic power or matrix-factorization methods (Sa et al., 2015). Other variants address Markovian data (Kumar & Sarkar, 2023), sparse leading eigenvectors (Kumar & Sarkar, 2024), entrywise uncertainty quantification (Kumar et al., 2025), and low-precision computation (Dasgupta et al., 2025). Most closely related to our work, Jain et al. (2016) obtained the first near-optimal gapped rates for general, possibly nonsymmetric matrix updates, while Li et al. (2018) give near-optimal gapped guarantees for sub-Gaussian PCA. Relatedly, Shamir (2016) gives an early eigengap-free guarantee permitting general PSD stochastic matrices, but with a slower objective rate and low success probability from random initialization. Later, Allen-Zhu & Li (2017) established an efficient global near-optimal gap-free analysis for rank-one streaming k-PCA, and Liang (2023) obtained sharp gap-free rates for sub-Gaussian data, again requiring rank-one updates. In another direction, Huang et al. (2021) extends nearly offline-optimal streaming-PCA guarantees to arbitrary-rank updates under an eigengap.

We note that this work focuses on the 1-PCA problem, i.e., approximating the top eigenvector of a population average $\mathbf{\Sigma}$ from samples. We leave open the analogous question for k-PCA for $k > 1$, where a similar situation holds in the current literature: Allen-Zhu & Li (2017) gave a gap-free result for k-PCA under rank-one updates, and Huang et al. (2021) removed the rank restriction, but used an eigengap.

Noisy power methods. Under Model 1 and an eigengap assumption $\lambda_2(\mathbf{\Sigma}) \le (1-\gamma)\lambda_1(\mathbf{\Sigma})$, Hardt & Price (2014) gives a suboptimal sample complexity scaling as $\widetilde{O}(\frac{V}{\lambda_1^2\gamma^3\Delta} + \frac{1}{\gamma})$ for minibatched stochastic matrix-vector products, even before accounting for their additional projected-noise condition (see the statement of their Corollary 1.1). This incurs an extra factor of $\frac{1}{\gamma}$ in the leading term compared with Theorem 1. Later, Balcan et al. (2016) replaces a dependence on $\lambda_1 - \lambda_2$ by $\lambda_1 - \lambda_{q+1}$, but requires maintaining at least q directions. Notably, both results hold only in the gapped setting.

Differentially private PCA. For arbitrary row-bounded datasets, early approaches sample a direction using the exponential mechanism (Chaudhuri et al., 2013), while Analyze Gauss (Dwork et al., 2014) adds a symmetric Gaussian matrix to the empirical covariance and then extracts its leading eigenspace. Notably, when applying such results to i.i.d. sub-Gaussian data, the resulting sample complexity is at least $d^{1.5}$ up to logarithmic factors. Specializing to i.i.d. statistical models, Liu et al. (2022b) use robust one-dimensional scores within a propose-test-release framework to obtain nearly information-theoretically optimal private PCA under sub-Gaussian and hypercontractive

assumptions, although the resulting estimator is not computationally efficient. The black-box reduction of Hopkins et al. (2023) converts suitable robust estimators into private mean and covariance estimators, from which PCA can be obtained by post-processing when covariance error controls the desired subspace. Closest to our algorithm, Liu et al. (2022a) give a single-pass minibatched Oja method with nearly optimal rates for sub-Gaussian data under an eigengap. Subsequent specialized results obtain minimax rates for rank-r spiked covariance models (Cai et al., 2024) and robustness to heavy tails and contamination under elliptical models (Kim & Jung, 2025).

B Notation and Additional Results

B.1 Notation

We use $X \perp Y$ to denote that random variables X and Y are independent. We use $\mathbb{I}_{\mathcal{E}}$ to denote the 0-1 indicator random variable of an event $\mathcal{E}$. For two measures π, μ over the same sample space Ω, which we identify with corresponding distributions, $\mathrm{TV}(\pi, \mu) \coloneqq \frac{1}{2}\int_\Omega |\pi - \mu| \mathrm{d}\omega$ denotes their TV distance and $\mathrm{KL}(\pi \| \mu) \coloneqq \int \pi \log \frac{\pi}{\mu} \mathrm{d}\omega$ denotes their KL divergence.

Vectors are denoted in lowercase boldface and matrices in uppercase boldface. We use $\mathbf{0}_d$ and $\mathbf{1}_d$ to denote the all-zeroes and all-ones vectors in $\mathbb{R}^d$, $\mathbf{I}_d$ to denote the $d \times d$ identity matrix, and $\mathbf{0}_{m\times n}$ to denote the $m \times n$ all-zeroes matrix. We use $[d]$ to denote $\{i \in \mathbb{N} : 1 \le i \le d\}$. For $p \ge 1$ including $p = \infty$ we use $\|\cdot\|_p$ to denote the ℓ_p norm of a vector, and $\|\cdot\|_{S_p}$ to denote the Schatten-p norm of a matrix. The set $\mathbb{S}^{d\times d}$ denotes all $d \times d$ symmetric matrices, and $\mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ denotes the subset of positive semidefinite matrices. We use $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ to denote the multivariate Gaussian with mean $\mathbf{m} \in \mathbb{R}^d$ and covariance $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$. We use $\|\cdot\|_{\mathrm{op}}$ to denote the (ℓ_2 induced) operator norm of a matrix, and $\|\cdot\|_{\mathrm{F}}$ to denote its Frobenius norm, i.e., Schatten-2 norm. We use $\lambda_i(\cdot)$ to denote the i^{th} largest eigenvalue of a symmetric matrix, and $\mathrm{Tr}(\cdot)$ for the trace. We say a matrix is orthonormal if its columns $\{\mathbf{u}_i\}$ satisfy $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbb{I}_{i=j}$. For unit vectors $\mathbf{u}, \mathbf{v}$ we define

$$\mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{v}) \coloneqq \min\left\{\|\mathbf{u} - \mathbf{v}\|_2, \|\mathbf{u} + \mathbf{v}\|_2\right\}.$$

Lemma 4. m_{sign} *satisfies the triangle inequality.*

Proof. For unit $\mathbf{u}, \mathbf{v}, \mathbf{w}$, if $\mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{w}) = \|\mathbf{u} - \sigma\mathbf{w}\|_2$, $\mathrm{m}_{\text{sign}}(\mathbf{w}, \mathbf{v}) = \|\mathbf{w} - \tau\mathbf{v}\|_2$, for $(\sigma, \tau) \in \{\pm 1\}^2$,

$$\mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{v}) \le \|\mathbf{u} - \sigma\tau\mathbf{v}\|_2 \le \|\mathbf{u} - \sigma\mathbf{w}\|_2 + \|\sigma\mathbf{w} - \sigma\tau\mathbf{v}\|_2 = \mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{w}) + \mathrm{m}_{\text{sign}}(\mathbf{w}, \mathbf{v}).$$

□

B.2 Baseline via matrix Bernstein

As a baseline, we recall a folklore result that in the offline setting, any approximate cPCA of the empirical covariance (with appropriate parameters) also solves the statistical cPCA problem. This result is stated under a slight strengthening of Model 1 that imposes a probability 1 bound on each sample $\mathbf{A}_t$, but naturally yields a high-probability guarantee unlike Theorem 1.

Model 3. *Fix $\lambda_1 > 0$, $V > 0$, and $M > 0$. Let $\{\mathbf{A}_t\}_{t\in[n]}$ from Model 1 additionally satisfy*

$$\|\mathbf{\Sigma} - \mathbf{A}_t\|_{\mathrm{op}} \le M \text{ with probability } 1.$$

Proposition 1 (Gap-free offline PCA via matrix Bernstein)**.** *Under Model 3, let $\widehat{\mathbf{\Sigma}} \coloneqq \frac{1}{2n}\sum_{t\in[n]}(\mathbf{A}_t + \mathbf{A}_t^\top)$, and let $\mathbf{v}$ be any $(\frac{\gamma}{6}, \frac{\Delta}{4})$-cPCA for $\widehat{\mathbf{\Sigma}}$. Then for any $\zeta \in (0, \frac{1}{3})$, $\mathbf{v}$ is a (γ, Δ)-cPCA for $\mathbf{\Sigma}$ with probability $\ge 1 - \zeta$, if for an appropriate constant,*

$$n = \Omega\left(\left(\frac{V}{\lambda_1^2\gamma^2\Delta} + \frac{M}{\lambda_1\gamma\sqrt{\Delta}}\right)\log\left(\frac{d}{\zeta}\right)\right).$$

Proof. This is almost the statement of Proposition 1, Tian (2026), up to the assumptions on $\{\mathbf{A}_t\}_{t\in[n]}$. The proof of Proposition 1, Tian (2026) shows the result if with probability $\ge 1 - \zeta$,

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathrm{op}} \le \frac{\lambda_1\gamma\sqrt{\Delta}}{4}.$$

To show this, set $\mathbf{E}_t \coloneqq \mathbf{A}_t - \mathbf{\Sigma}$ and $\mathbf{E} \coloneqq \frac{1}{n}\sum_{t\in[n]} \mathbf{E}_t$. The $\mathbf{E}_t$ are independent and mean zero, so the bounds from Model 3 and the matrix Bernstein inequality (Theorem 6.6.1, Tropp (2015)) prove the above bound on $\|\mathbf{E}\|_{\mathrm{op}}$. The claim follows from Jensen's inequality and $\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} = \frac{1}{2}(\mathbf{E} + \mathbf{E}^\top)$. □

C Deferred Proofs from Section 2

Fact 1. *Let $\mathbf{x}, \mathbf{y}, \mathbf{c} \in \mathbb{R}^d_{\geq 0}$ have nonincreasing coordinates, and suppose $\mathbf{y}$ weakly majorizes $\mathbf{x}$. Then*

$$\sum_{j\in[k]} \mathbf{c}_j \mathbf{x}_j \leq \sum_{j\in[k]} \mathbf{c}_j \mathbf{y}_j \text{ for all } k \in [d].$$

Proof. For all $k \in [d]$ let $\mathbf{s}_k \coloneqq \sum_{j\in[k]} \mathbf{x}_j$ and $\mathbf{t}_k \coloneqq \sum_{j\in[k]} \mathbf{y}_j$. Then Abel's summation formula gives

$$\sum_{j\in[k]} \mathbf{c}_j(\mathbf{y}_j - \mathbf{x}_j) = \mathbf{c}_k(\mathbf{t}_k - \mathbf{s}_k) + \sum_{j\in[k-1]} (\mathbf{c}_j - \mathbf{c}_{j+1})(\mathbf{t}_j - \mathbf{s}_j) \geq 0.$$

□

C.1 Proof of Lemma 1

Proof. Define $\mathbf{H} \coloneqq \mathbf{B}^\top\mathbf{B}$ and $\mathbf{K} \coloneqq \mathbf{B}^\top\mathbf{V}\mathbf{V}^\top\mathbf{B}$. Then our goal is to bound

$$\frac{\|\mathbf{V}^\top\mathbf{B}\mathbf{g}\|_2^2}{\|\mathbf{B}\mathbf{g}\|_2^2} = \frac{\mathbf{g}^\top\mathbf{K}\mathbf{g}}{\mathbf{g}^\top\mathbf{H}\mathbf{g}}.$$

For the denominator, standard Gaussian anti-concentration (e.g., Lemma A.2.1, Kumar & Sarkar (2024)) shows

$$\mathbf{g}^\top\mathbf{H}\mathbf{g} \geq \frac{\zeta^2}{4e}\mathrm{Tr}(\mathbf{H})$$

with probability at least $1 - \frac{\zeta}{2}$. Similarly, for the numerator, standard χ^2 concentration bounds (e.g., Lemma 1, Laurent & Massart (2000)) along with $\mathrm{Tr}(\mathbf{K}^2) \leq \mathrm{Tr}(\mathbf{K})^2$, $\|\mathbf{K}\|_{\mathrm{op}} \leq \mathrm{Tr}(\mathbf{K})$, gives

$$\Pr\left(\mathbf{g}^\top\mathbf{K}\mathbf{g} > (1 + 2\sqrt{t} + 2t)\mathrm{Tr}(\mathbf{K})\right) \leq \exp(-t) \text{ for all } t > 0.$$

Plugging in $t = \log(\frac{2}{\zeta}) \leq 2\log(\frac{1}{\zeta})$ and combining the above three displays gives the result. □

C.2 Proof of Lemma 3

Proof. First, observe that since $\|\mathbf{C}_t\|_{\mathrm{F}}^2 \geq \|\mathbf{C}_t^2\|_{\mathrm{op}} = \prod_{s\in[t]}(1+\eta_s\lambda_1)^2$, $\mathbf{P}$ commutes with all of the $\mathbf{I}_d + \eta_s\mathbf{\Sigma}$, and the corresponding eigenvalues of $\mathbf{\Sigma}$ are $\leq (1-\gamma)\lambda_1$,

$$\frac{\|\mathbf{P}\mathbf{C}_t\|_{\mathrm{F}}^2}{\|\mathbf{C}_t\|_{\mathrm{F}}^2} \leq d\prod_{s\in[t]}\left(\frac{1+\eta_s(1-\gamma)\lambda_1}{1+\eta_s\lambda_1}\right)^2 \leq d\exp\left(-\gamma\lambda_1\sum_{s\in[t]}\eta_s\right). \tag{7}$$

Next, independence of the stream in Model 1 shows that $\mathbb{E}[\mathbf{B}_t] = \mathbf{C}_t$, so Lemma 2 gives

$$\mathbb{E}\,\|\mathbf{B}_t - \mathbf{C}_t\|_{\mathrm{F}}^2 = \mathbb{E}\,\|\mathbf{B}_t\|_{\mathrm{F}}^2 - \|\mathbf{C}_t\|_{\mathrm{F}}^2 \leq (\exp(q_t) - 1)\,\|\mathbf{C}_t\|_{\mathrm{F}}^2. \tag{8}$$

Thus, by Markov's inequality, we have with probability $\geq 1 - \zeta$ that

$$\frac{\|\mathbf{B}_t - \mathbf{C}_t\|_{\mathrm{F}}}{\|\mathbf{C}_t\|_{\mathrm{F}}} \leq \sqrt{\frac{\exp(q_t) - 1}{\zeta}} < 1, \tag{9}$$

where the last inequality used our assumption $q_t \le \frac{\zeta}{2}$. Under this event, we have $\mathbf{B}_t \neq \mathbf{0}_{d\times d}$ by the triangle inequality. Further, for non-zero $\mathbf{X}, \mathbf{Y}$,

$$\left\| \frac{\mathbf{X}}{\|\mathbf{X}\|_{\mathrm{F}}} - \frac{\mathbf{Y}}{\|\mathbf{Y}\|_{\mathrm{F}}} \right\|_{\mathrm{F}} \le \left\| \frac{\mathbf{X}-\mathbf{Y}}{\|\mathbf{Y}\|_{\mathrm{F}}} \right\|_{\mathrm{F}} + \left\| \mathbf{X} \cdot \left(\frac{1}{\|\mathbf{X}\|_{\mathrm{F}}} - \frac{1}{\|\mathbf{Y}\|_{\mathrm{F}}} \right) \right\|_{\mathrm{F}} \le 2\frac{\|\mathbf{X}-\mathbf{Y}\|_{\mathrm{F}}}{\|\mathbf{Y}\|_{\mathrm{F}}}. \tag{10}$$

Finally, applying (10) with $(\mathbf{X}, \mathbf{Y}) \leftarrow (\mathbf{B}_t, \mathbf{C}_t)$ implies

$$\begin{aligned}
\frac{\|\mathbf{P}\mathbf{B}_t\|_{\mathrm{F}}}{\|\mathbf{B}_t\|_{\mathrm{F}}} &\le \left\| \mathbf{P}\frac{\mathbf{C}_t}{\|\mathbf{C}_t\|_{\mathrm{F}}} \right\|_{\mathrm{F}} + \left\| \mathbf{P}\left(\frac{\mathbf{B}_t}{\|\mathbf{B}_t\|_{\mathrm{F}}} - \frac{\mathbf{C}_t}{\|\mathbf{C}_t\|_{\mathrm{F}}} \right) \right\|_{\mathrm{F}} \\
&\le \frac{\|\mathbf{P}\mathbf{C}_t\|_{\mathrm{F}}}{\|\mathbf{C}_t\|_{\mathrm{F}}} + \left\| \frac{\mathbf{B}_t}{\|\mathbf{B}_t\|_{\mathrm{F}}} - \frac{\mathbf{C}_t}{\|\mathbf{C}_t\|_{\mathrm{F}}} \right\|_{\mathrm{F}} \\
&\le \sqrt{d\exp\left(-\gamma\lambda_1 \sum_{s\in[t]} \eta_s \right)} + 2\sqrt{\frac{\exp(q_t)-1}{\zeta}},
\end{aligned}$$

where we used (7) and (9) in the last line. The conclusion follows from $(a+b)^2 \le 2(a^2+b^2)$. □

D PCA Lower Bound

We give an information-theoretic lower bound that shows the leading-order parameter dependence in Theorem 1 is qualitatively tight, for any choice of V, λ_1. We state our result under Model 1, but our hard instance is even more well-behaved: the matrices $\mathbf{A}_i$ are always PSD.

To begin, we require a standard formulation of Le Cam's two point method.

Lemma 5 (Theorem 2.2(i), Tsybakov (2009))**.** *Let P_0, P_1 be probability distributions on the same measurable space Ω, and let $\phi : \Omega \to \{0,1\}$ be a (possibly randomized) function. Then,*

$$\Pr_{\omega\sim P_0}[\phi(\omega) = 1] + \Pr_{\omega\sim P_1}[\phi(\omega) = 0] \ge 1 - \mathrm{TV}(P_0, P_1).$$

We can now state and prove our lower bound.

Theorem 4. *Fix any choice of $\lambda_1 > 0$, $V > 0$, $\gamma \in (0, \frac{1}{4})$, and $\Delta \in (0, \frac{1}{4})$. There is no algorithm $\mathcal{A}$ that takes as input $\{\mathbf{A}_i\}_{i\in[n]}$ from Model 1, and outputs a (γ, Δ)-cPCA of $\mathbf{\Sigma}$ with probability $\ge \frac{2}{3}$, even assuming that $\mathbf{A}_i \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ for all $i \in [n]$, unless for an appropriate constant,*

$$n = \Omega\left(\frac{V}{\lambda_1^2\gamma^2\Delta} \right).$$

Proof. We begin by defining matrices used in our construction. Let $\alpha := 1.5 \arcsin\sqrt{\Delta}$, and

$$\mathbf{u}_0 := \begin{pmatrix} \cos\alpha \\ -\sin\alpha \end{pmatrix}, \quad \mathbf{v}_0 := \begin{pmatrix} \sin\alpha \\ \cos\alpha \end{pmatrix}, \quad \mathbf{u}_1 := \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \quad \mathbf{v}_1 := \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}.$$

Also, let

$$\mathbf{\Sigma}_0 := \lambda_1\mathbf{u}_0\mathbf{u}_0^\top + (1-2\gamma)\lambda_1\mathbf{v}_0\mathbf{v}_0^\top, \quad \mathbf{\Sigma}_1 := \lambda_1\mathbf{u}_1\mathbf{u}_1^\top + (1-2\gamma)\lambda_1\mathbf{v}_1\mathbf{v}_1^\top.$$

If $\mathbf{u}$ is a (γ, Δ)-cPCA of $\mathbf{\Sigma}_0$, then $\mathrm{m}_{\mathrm{sign}}(\mathbf{u}, \mathbf{u}_0) \le 2\sin(\frac{\alpha}{3})$, and a similar bound holds for $\mathbf{\Sigma}_1$. We observe two helpful reformulations of $\mathbf{\Sigma}_0, \mathbf{\Sigma}_1$ used in our constructions. First,

$$\begin{aligned}
&\mathbf{\Sigma}_i = \mathbf{D} + (2i-1)\beta\mathbf{H} \text{ for } i \in \{0,1\}, \text{ where } \beta := 2\gamma\lambda_1\sin(\alpha)\cos(\alpha), \\
&\mathbf{D} := \lambda_1 \begin{pmatrix} \cos^2\alpha + (1-2\gamma)\sin^2\alpha & 0 \\ 0 & \sin^2\alpha + (1-2\gamma)\cos^2\alpha \end{pmatrix}, \quad \mathbf{H} := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\end{aligned} \tag{11}$$

Second, letting the diagonal elements of $\mathbf{D}$ be d_1 and d_2, and $s := d_1 + d_2 = 2(1-\gamma)\lambda_1$,

$$\begin{aligned}
&\mathbf{\Sigma}_i = s\left(\left(\frac{1}{2} + (2i-1)\eta \right)\mathbf{w}\mathbf{w}^\top + \left(\frac{1}{2} - (2i-1)\eta \right)\mathbf{z}\mathbf{z}^\top \right), \text{ where } \eta := \frac{\beta}{2\sqrt{d_1 d_2}}, \\
&\mathbf{w} := \frac{1}{\sqrt{s}}\begin{pmatrix} \sqrt{d_1} \\ \sqrt{d_2} \end{pmatrix}, \quad \mathbf{z} := \frac{1}{\sqrt{s}}\begin{pmatrix} \sqrt{d_1} \\ -\sqrt{d_2} \end{pmatrix}.
\end{aligned} \tag{12}$$

Now suppose there is an algorithm $\mathcal{A}$ as in the theorem statement, and define $\phi : (\mathbb{S}^{d\times d}_{\succeq \mathbf{0}})^n \to \{0,1\}$ as follows. Given matrices $\{\mathbf{A}_i\}_{i\in[n]}$, let $\mathbf{u} \coloneqq \mathcal{A}(\{\mathbf{A}_i\}_{i\in[n]})$ be the assumed algorithm's output. Then we let ϕ be the composition of $\mathcal{A}$ with the map $\mathbf{u} \to i \in \{0,1\}$, where $i = 0$ if $\mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{u}_0) \le \mathrm{m}_{\text{sign}}(\mathbf{u}, \mathbf{u}_1)$ and $i = 1$ otherwise. Because $\mathrm{m}_{\text{sign}}(\mathbf{u}_0, \mathbf{u}_1) = 2\sin\alpha > 4\sin(\frac{\alpha}{3})$, Lemma 4 implies that ϕ identifies $i \in \{0,1\}$ whenever $\mathbf{u}$ is a (γ, Δ)-cPCA of the corresponding $\mathbf{\Sigma}_i$.

We next define our distributions on the $\{\mathbf{A}_i\}_{i\in[n]}$. We split into two cases depending on V. In each case, we define single-sample laws p_0, p_1 with means $\mathbf{\Sigma}_0, \mathbf{\Sigma}_1$, and let $P_0 \coloneqq p_0^{\otimes n}$, $P_1 \coloneqq p_1^{\otimes n}$ be the n-fold product laws. We show that our laws satisfy Model 1, and bound $\mathrm{KL}(p_0 \| p_1)$. In this setting, existence of the stated algorithm $\mathcal{A}$ implies that

$$\Pr_{\omega\sim P_0}[\phi(\omega) = 1] \le \frac{1}{3}, \quad \Pr_{\omega\sim P_1}[\phi(\omega) = 0] \le \frac{1}{3}. \tag{13}$$

Case 1: $V \le (1-2\gamma)\lambda_1^2$. We follow the notation (11). Let $t \coloneqq \sqrt{V+\beta^2}$. We define the law of $\mathbf{A} \sim p_i$ for $i \in \{0,1\}$ as follows. First, we draw $\sigma \in \{\pm 1\}$ with $\mathbb{E}[\sigma] = \frac{(2i-1)\beta}{t}$, so that

$$\sigma = \begin{cases} 1 & \text{with probability } \frac{1}{2} + \frac{(2i-1)\beta}{2t} \\ -1 & \text{with probability } \frac{1}{2} - \frac{(2i-1)\beta}{2t} \end{cases},$$

We then set $\mathbf{A} \leftarrow \mathbf{D} + \sigma t \mathbf{H}$. Observe that $\mathbb{E}_{p_i}[\mathbf{A}] = \mathbf{D} + (2i-1)\beta\mathbf{H} = \mathbf{\Sigma}_i$ from (11),

$$\mathbb{E}_{p_i}\left[(\mathbf{\Sigma}_i - \mathbf{A})^2\right] = \left(t^2 - \beta^2\right)\mathbf{H}^2 = V\mathbf{I}_2,$$

and $\mathbf{A}$ has positive entries on the diagonal, with

$$\det(\mathbf{A}) = \det(\mathbf{D}) - t^2 = (1-2\gamma)\lambda_1^2 + \beta^2 - t^2 \ge 0.$$

Thus, draws from both P_0 and P_1 are valid instances of Model 1. We also have

$$\mathrm{KL}(p_0\|p_1) = \frac{\beta}{t}\log\left(\frac{1+\frac{\beta}{t}}{1-\frac{\beta}{t}}\right) \le \frac{2\beta^2}{t^2-\beta^2} \le \frac{18\gamma^2\lambda_1^2\Delta}{V},$$

where we computed the KL between two distributions on $\{\pm 1\}$ with probabilities $\frac{1}{2} \pm \frac{\beta}{2t}$, and used the bounds $\beta^2 \le 9\gamma^2\lambda_1^2\Delta$ and $x\log\frac{1+x}{1-x} \le \frac{2x^2}{1-x^2}$ valid for $x \in [0,1)$.

Case 2: $V > (1-2\gamma)\lambda_1^2$. We follow the notation (11), (12). Let $R \coloneqq \lambda_1 + \frac{V}{\lambda_1}$ and $q \coloneqq \frac{s}{R} < 1$. We define the law of $\mathbf{A} \sim p_i$ for $i \in \{0,1\}$ as follows. We set

$$\mathbf{A} = \begin{cases} R\mathbf{w}\mathbf{w}^\top & \text{with probability } q\left(\frac{1}{2} + (2i-1)\eta\right) \\ R\mathbf{z}\mathbf{z}^\top & \text{with probability } q\left(\frac{1}{2} - (2i-1)\eta\right) \\ \mathbf{0}_{2\times 2} & \text{with probability } 1-q \end{cases}.$$

Observe that $\mathbb{E}_{p_i}[\mathbf{A}] = \mathbf{\Sigma}_i$ from (12), and $\mathbf{A}$ is clearly always PSD. Further, because $R\lambda_1 - \lambda_1^2 = V$ and $R(1-2\gamma)\lambda_1 - (1-2\gamma)^2\lambda_1^2 = (1-2\gamma)(V + 2\gamma\lambda_1^2) \le V$, we have

$$\mathbb{E}_{p_i}\left[(\mathbf{\Sigma}_i - \mathbf{A})^2\right] = R\mathbf{\Sigma}_i - \mathbf{\Sigma}_i^2 \preceq V\mathbf{I}_2.$$

Thus, draws from both P_0 and P_1 are valid instances of Model 1. We can finally directly compute

$$\mathrm{KL}(p_0\|p_1) = (2\eta q)\log\frac{1+2\eta}{1-2\eta} \le \frac{8q\eta^2}{1-4\eta^2} = \frac{2q\beta^2}{d_1d_2 - \beta^2} \le \frac{8\beta^2}{V} \le \frac{72\gamma^2\lambda_1^2\Delta}{V},$$

where we used $d_1d_2 - \beta^2 = (1-2\gamma)\lambda_1^2$, $q \le \frac{2\lambda_1}{R} \le \frac{2\lambda_1^2}{V}$, and our earlier bound $\beta^2 \le 9\gamma^2\lambda_1^2\Delta$.

In summary, in all regimes of (V, λ_1), there are instances P_0, P_1 of Model 1 such that $P_0 = p_0^{\otimes n}$, $P_1 = p_1^{\otimes n}$, the mean of p_i is $\mathbf{\Sigma}_i$ for $i \in \{0,1\}$, and $\mathrm{KL}(p_0\|p_1) = O(\frac{\gamma^2\lambda_1^2\Delta}{V})$.

By tensorization of KL divergence, we have for sufficiently small $n = o(\frac{V}{\lambda_1^2\gamma^2\Delta})$ that $\mathrm{TV}(P_0, P_1) < \frac{1}{3}$. Combining with Lemma 5 and (13), this contradicts existence of $\mathcal{A}$. □

E HIGH-PROBABILITY GUARANTEES

Theorem 1 achieves a near-optimal rate of error as a function of the parameters in Model 1 and Definition 1, but only provides a low-confidence guarantee (i.e., with polynomial dependence on $\frac{1}{\zeta}$). We next give a confidence amplification procedure with only polylogarithmic overheads, without worsening the dependence on either of the cPCA parameters (γ, Δ).

Notably, standard confidence boosting procedures based on a direct geometric aggregation (e.g., Lemma 3.10, Kumar & Sarkar (2024)) do not work, since if the leading eigenvalue has multiplicity larger than one, even two exact cPCAs may be orthogonal. Instead, we aggregate sketches of the unnormalized Oja product, piggybacking off of closeness guarantees from Lemma 3. The main technical novelty in this section is that we require a polylogarithmic-dimension sketch (rather than the single Gaussian used in Algorithm 1), so that the population-level sketch concentrates with high probability.

We begin with a standard Gaussian trace estimate.

Lemma 6. *Let $\mathbf{H} \succeq \mathbf{0}$ be fixed and let $\mathbf{G} \in \mathbb{R}^{d\times s}$ have i.i.d. $\mathcal{N}(0,1)$ entries. Then,*

$$\frac{s}{2}\mathrm{Tr}(\mathbf{H}) \le \mathrm{Tr}(\mathbf{G}^\top \mathbf{H}\mathbf{G}) \le 2s\mathrm{Tr}(\mathbf{H}), \tag{14}$$

with probability $\ge 1 - \exp(-\frac{s}{16})$ for each inequality.

Proof. Let the columns of $\mathbf{G}$ be $\{\mathbf{g}_i\}_{i\in[s]}$. Then, the middle term in (14) is $X \coloneqq \sum_{j\in[s]} \mathbf{g}_j^\top \mathbf{H}\mathbf{g}_j$. The claim follows from χ^2 concentration. In particular, Lemma 1, Laurent & Massart (2000) gives

$$\Pr\left(X - s\mathrm{Tr}(\mathbf{H}) \ge 2\sqrt{s\mathrm{Tr}(\mathbf{H}^2)x} + 2\left\|\mathbf{H}\right\|_{\mathrm{op}} x\right) \le \exp(-x)$$
$$\Pr\left(s\mathrm{Tr}(\mathbf{H}) - X \ge 2\sqrt{s\mathrm{Tr}(\mathbf{H}^2)x}\right) \le \exp(-x).$$

Using $x \leftarrow \frac{s}{16}$, $\mathrm{Tr}(\mathbf{H}^2) \le \mathrm{Tr}(\mathbf{H})^2$, and $\left\|\mathbf{H}\right\|_{\mathrm{op}} \le \mathrm{Tr}(\mathbf{H})$ proves the claim. □

Next, we show that independent product sketches cluster around a common population sketch.

Lemma 7. *Let $\mathbf{G} \in \mathbb{R}^{d\times R}$ have i.i.d. $\mathcal{N}(0,1)$ entries. Under Model 1, if $\eta_s \le \frac{1}{\lambda_1}$ for all $s \in [t]$ and $q_t \coloneqq V\sum_{s\in[t]}\eta_s^2 \le \frac{1}{128}$, then with probability $\ge 1 - 3\exp(-\frac{R}{16})$ over $\mathbf{G}$, the following hold. First,*

$$\left\|\mathbf{P}\mathbf{Z}_{\star,t}\right\|_{\mathrm{F}}^2 \le 4d\exp\left(-\gamma\lambda_1\sum_{s\in[t]}\eta_s\right), \text{ for } \mathbf{Z}_{\star,t} \coloneqq \frac{\mathbf{C}_t\mathbf{G}}{\left\|\mathbf{C}_t\mathbf{G}\right\|_{\mathrm{F}}}. \tag{15}$$

Second, conditioning on $\mathbf{G}$, over the randomness of Model 1,

$$\Pr\left[\left\|\frac{\mathbf{B}_t\mathbf{G}}{\left\|\mathbf{B}_t\mathbf{G}\right\|_{\mathrm{F}}} - \mathbf{Z}_{\star,t}\right\|_{\mathrm{F}} \le 8\sqrt{2q_t}\right] \ge \frac{3}{4}. \tag{16}$$

Proof. Throughout the proof, let $\mathbf{H}_t \coloneqq \mathbb{E}[(\mathbf{B}_t - \mathbf{C}_t)^\top(\mathbf{B}_t - \mathbf{C}_t)]$. Then Lemma 6 implies

$$\left\|\mathbf{C}_t\mathbf{G}\right\|_{\mathrm{F}}^2 \ge \frac{R}{2}\left\|\mathbf{C}_t\right\|_{\mathrm{F}}^2, \quad \left\|\mathbf{P}\mathbf{C}_t\mathbf{G}\right\|_{\mathrm{F}}^2 \le 2R\left\|\mathbf{P}\mathbf{C}_t\right\|_{\mathrm{F}}^2, \quad \mathrm{Tr}\left(\mathbf{G}^\top\mathbf{H}_t\mathbf{G}\right) \le 2R\mathrm{Tr}\left(\mathbf{H}_t\right), \tag{17}$$

all hold with the requisite failure probability. Condition on these events henceforth. Combining the first two events in (17) with (7) then immediately implies our first claim, (15).

Next, recall from (8) and our range on q_t that

$$\mathrm{Tr}\left(\mathbf{H}_t\right) = \mathbb{E}\left\|\mathbf{B}_t - \mathbf{C}_t\right\|_{\mathrm{F}}^2 \le 2q_t\left\|\mathbf{C}_t\right\|_{\mathrm{F}}^2.$$

Combining with the first and third events in (17) then gives

$$\mathbb{E}\left\|\left(\mathbf{B}_t - \mathbf{C}_t\right)\mathbf{G}\right\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\mathbf{G}^\top\mathbf{H}_t\mathbf{G}\right) \le 2R\mathrm{Tr}(\mathbf{H}_t) \le 4q_tR\left\|\mathbf{C}_t\right\|_{\mathrm{F}}^2 \le 8q_t\left\|\mathbf{C}_t\mathbf{G}\right\|_{\mathrm{F}}^2.$$

Algorithm 3: BoostedSketchOja$(\{\mathbf{A}_t, \eta_t\}_{t\in[n]}, R, \tau)$

Input: $\{\mathbf{A}_t \in \mathbb{R}^{d\times d}, \eta_t > 0\}_{t\in[n]}$, $R \in \mathbb{N}$, $\tau > 0$
$\mathbf{G} \leftarrow d \times R$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries
for $r \in [R]$ **do**
  $\mathbf{Y}_r \leftarrow \mathbf{G}$
end
for $j \in [\lfloor \frac{n}{R} \rfloor]$ **do**
  for $r \in [R]$ **do**
    $\mathbf{Y}_r \leftarrow (\mathbf{I}_d + \eta_{(j-1)R+r}\mathbf{A}_{(j-1)R+r})\mathbf{Y}_r$
  end
end
for $r \in [R]$ **do**
  $\mathbf{Z}_r \leftarrow \mathbf{Y}_r / \|\mathbf{Y}_r\|_{\mathrm{F}}$, or a fixed unit-Frobenius-norm matrix if $\mathbf{Y}_r = \mathbf{0}_{d\times R}$
end
for $r \in [R]$ **do**
  $\mathcal{N}_r \leftarrow \{r' \in [R] : \|\mathbf{Z}_r - \mathbf{Z}_{r'}\|_{\mathrm{F}} \le 2\tau\}$
end
$i \leftarrow$ any index with $|\mathcal{N}_i| \ge \frac{R}{2}$, else $i \leftarrow 1$
$\mathbf{w}_n \leftarrow$ any top left singular vector of $\mathbf{Z}_i$
Return: $\mathbf{w}_n$

Thus, by Markov's inequality, with probability $\ge \frac{3}{4}$ over Model 1,

$$\frac{\|(\mathbf{B}_t - \mathbf{C}_t)\mathbf{G}\|_{\mathrm{F}}}{\|\mathbf{C}_t\mathbf{G}\|_{\mathrm{F}}} \le 4\sqrt{2q_t} \le \frac{1}{2}.$$

Thus we have $\mathbf{B}_t\mathbf{G} \neq \mathbf{0}_{d\times R}$ on this event, so again applying (10) (with $(\mathbf{X}, \mathbf{Y}) \leftarrow (\mathbf{B}_t\mathbf{G}, \mathbf{C}_t\mathbf{G})$), and using the definition of $\mathbf{Z}_{\star,t}$, concludes the proof of (16). □

We give our full high-probability method in Algorithm 3, which uses a shared Gaussian matrix across all sample blocks, making the population center common to the independent runs. This lets us apply geometric aggregation to boost the guarantee (16). We make two further observations: first, although it is written with a specified horizon n, Algorithm 3 is an online algorithm, as the iterations do not use knowledge of n. Second, it is implementable using $O(dR^2) = O(d\log^2(\frac{1}{\zeta}))$ space, for the eventual $R = O(\log(\frac{1}{\zeta}))$ in Theorem 5, by storing only the $\mathbf{Y}_r$.

We first give the guarantee for a common stepsize schedule across the R runs.

Lemma 8. *Under Model 1, let $R, T \in \mathbb{N}$ and run R independent Oja products for T steps with common step sizes η_t, using the shared Gaussian initialization, $\mathbf{G}$, and aggregation rule of Algorithm 3. Let $\gamma \in (0,1)$ and $0 < \tau < \frac{1}{4\sqrt{R}}$. If $0 < \eta_t \le \frac{1}{\lambda_1}$ and*

$$V\sum_{t=1}^{T}\eta_t^2 \le \frac{\tau^2}{128}, \qquad 4d\exp\left(-\gamma\lambda_1\sum_{t=1}^{T}\eta_t\right) \le \tau^2,$$

then the aggregated output is a $(\gamma, 16R\tau^2)$-cPCA of $\mathbf{\Sigma}$ with probability $\ge 1 - 4\exp(-\frac{R}{72})$.

Proof. Set $n = RT$ and index the update in round j of run r by $\mathbf{A}_{(j-1)R+r}$, for $j \in [T]$ and $r \in [R]$. Let $\mathbf{P}$ project onto eigenvectors of $\mathbf{\Sigma}$ with eigenvalues below $(1-\gamma)\lambda_1$. Define

$$\mathbf{C}_j := (\mathbf{I}_d + \eta_j\mathbf{\Sigma})\cdots(\mathbf{I}_d + \eta_1\mathbf{\Sigma}),$$
$$\mathbf{B}_j^{(r)} := (\mathbf{I}_d + \eta_j\mathbf{A}_{(j-1)R+r})\cdots(\mathbf{I}_d + \eta_1\mathbf{A}_r),$$

and

$$\mathbf{Z}_\star := \frac{\mathbf{C}_T\mathbf{G}}{\|\mathbf{C}_T\mathbf{G}\|_{\mathrm{F}}}, \qquad \mathbf{Z}_r := \frac{\mathbf{B}_T^{(r)}\mathbf{G}}{\left\|\mathbf{B}_T^{(r)}\mathbf{G}\right\|_{\mathrm{F}}},$$

with the algorithm's fallback definition when $\mathbf{B}_T^{(r)}\mathbf{G} = \mathbf{0}_{d\times R}$. By Lemma 7 with $t \leftarrow T$ and $q_T \leftarrow V\sum_{j\in[T]}\eta_j^2 \le \frac{\tau^2}{128}$, with probability at least $1 - 3\exp(-\frac{R}{16})$ over $\mathbf{G}$, $\|\mathbf{P}\mathbf{Z}_\star\|_{\mathrm{F}}^2 \le \tau^2$. Further, for each run,

$$\Pr(\|\mathbf{Z}_\star - \mathbf{Z}_r\|_{\mathrm{F}} \le \tau \mid \mathbf{G}) \ge \frac{3}{4}.$$

Conditioning on this event for $\mathbf{G}$, the runs are independent. By a Chernoff bound, the set

$$\mathcal{G} \coloneqq \{r \in [R] \mid \|\mathbf{Z}_r - \mathbf{Z}_\star\|_{\mathrm{F}} \le \tau\}$$

has size at least $\frac{2R}{3}$ with conditional probability $\ge 1 - \exp(-\frac{R}{72})$. Every such candidate has at least $\frac{2R}{3}$ neighbors within distance 2τ. Conversely, any candidate with at least $\frac{R}{2}$ such neighbors has a neighbor in $\mathcal{G}$, and is therefore within distance 3τ of $\mathbf{Z}_\star$ by the triangle inequality.

Let $\widehat{\mathbf{Z}} \coloneqq \mathbf{Z}_i$ be the matrix selected in Line 17. On the preceding events, which hold with probability $\ge 1 - 3\exp(-\frac{R}{16}) - \exp(-\frac{R}{72}) \ge 1 - 4\exp(-\frac{R}{72})$,

$$\left\|\mathbf{P}\widehat{\mathbf{Z}}\right\|_{\mathrm{F}} \le \|\mathbf{P}\mathbf{Z}_\star\|_{\mathrm{F}} + \left\|\widehat{\mathbf{Z}} - \mathbf{Z}_\star\right\|_{\mathrm{F}} \le 4\tau.$$

Let $\mathbf{w}_n$ be a top left singular vector of $\widehat{\mathbf{Z}}$. Since $\|\widehat{\mathbf{Z}}\|_{\mathrm{F}} = 1$ and $\mathrm{rank}(\widehat{\mathbf{Z}}) \le R$, its top singular value is at least $R^{-1/2}$, so $\|\mathbf{P}\widehat{\mathbf{Z}}\|_{\mathrm{F}}^2 \ge \frac{1}{R}\|\mathbf{P}\mathbf{w}_n\|_2^2$ by expanding the singular value decomposition. Thus,

$$\|\mathbf{P}\mathbf{w}_n\|_2^2 \le R\left\|\mathbf{P}\widehat{\mathbf{Z}}\right\|_{\mathrm{F}}^2 \le 16R\tau^2.$$

□

We conclude by proving our main high-probability result, Theorem 5.

Theorem 5. *Let $\zeta \in (0, \frac{1}{3})$ and $(\gamma, \Delta) \in (0,1)^2$. Under Model 1, if*

$$n = \Omega\left(\frac{V}{\lambda_1^2\gamma^2\Delta}\log^2\left(\frac{d}{\Delta\zeta}\right)\log^2\left(\frac{1}{\zeta}\right) + \frac{\log\left(\frac{d}{\Delta\zeta}\right)\log\left(\frac{1}{\zeta}\right)}{\gamma}\right)$$

for an appropriate constant, then there exist choices of the inputs $R, \tau, \{\eta_t\}_{t\ge1}$, such that the output of Algorithm 3 is a (γ, Δ)-cPCA of $\mathbf{\Sigma}$ with probability $\ge 1 - \zeta$.

Proof. Throughout the proof, for sufficiently large $C > 0$ and small $c > 0$, we take

$$R \coloneqq \left\lceil C\log\left(\frac{1}{\zeta}\right)\right\rceil, \quad L \coloneqq C\log\left(\frac{d}{\Delta\zeta}\right), \quad \tau \coloneqq c\sqrt{\frac{\Delta}{R}}.$$

For an iteration $t \in [n]$ such that $t = (j-1)R + r$ for some $r \in [R]$, we also choose

$$\eta_t \coloneqq \frac{L}{\gamma\lambda_1(\beta + j)}, \text{ where } \beta \coloneqq C\left(\frac{L}{\gamma} + \frac{VRL^2}{\gamma^2\lambda_1^2\Delta}\right).$$

Let $T \coloneqq \lfloor\frac{n}{R}\rfloor$ and write $\alpha_j \coloneqq \eta_{(j-1)R+r}$ for the common stepsize in round $j \in [T]$, which is independent of $r \in [R]$. By the bound on β and n, we ensure $T \ge 4\beta$ and $\alpha_j \le \frac{1}{\lambda_1}$. Then, the same integral comparison as in the proof of Theorem 1 yields

$$q_T = V\sum_{j\in[T]}\alpha_j^2 \le \frac{VL^2}{\gamma^2\lambda_1^2\beta} \le \frac{1}{128}, \quad \gamma\lambda_1\sum_{j\in[T]}\alpha_j \ge L,$$

if C is sufficiently large relative to c. These imply

$$4d\exp\left(-\gamma\lambda_1\sum_{j\in[T]}\alpha_j\right) \le 4de^{-L} \le \tau^2, \quad 8\sqrt{2q_T} \le c\sqrt{\frac{\Delta}{R}} = \tau.$$

Applying Lemma 8 to these R runs of T updates, with $\eta_j \leftarrow \alpha_j$ and radius τ, gives

$$\|\mathbf{P}\mathbf{w}_n\|_2^2 \le 16R\tau^2 = 16c^2\Delta \le \Delta,$$

for $c \le \frac{1}{4}$, with failure probability at most $4\exp(-\frac{R}{72}) \le \zeta$ when C is sufficiently large. □

F ENERGY PCA GUARANTEES

In this section, we show how to extend the approach of Section 2 to an alternative gap-free notion of PCA often considered in the literature, energy PCA, which asks for a direction capturing nearly the largest possible variance. We use the following definition from Jambulapati et al. (2024).

Definition 4 (ePCA)**.** *Let $\alpha \in (0,1)$, and let $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$. We say that a unit vector $\mathbf{w} \in \mathbb{R}^d$ is an* α-ePCA *(energy PCA) of $\mathbf{\Sigma}$ if $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \geq (1-\alpha)\lambda_1(\mathbf{\Sigma})$.*

Lemma 8 of Jambulapati et al. (2024) shows that a (γ, Δ)-cPCA is also a $(\gamma + \Delta)$-ePCA. Taking $\gamma = \Delta \leftarrow \frac{\alpha}{2}$ in Theorem 1 therefore gives an α-ePCA, but the resulting sample complexity scales as $\frac{V}{\lambda_1^2 \alpha^3}$, which is suboptimal in its dependence on $\frac{1}{\alpha}$. This conversion bounds the squared projection onto eigenvalues below a single threshold $(1-\gamma)\lambda_1$. To avoid the lossy conversion, we will use the following identity that expresses $1 - \frac{\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}}{\lambda_1}$ as an integral over thresholds $(1-u)\lambda_1$.

Proposition 2 (Multiscale cPCA-to-ePCA)**.** *Let $\mathbf{\Sigma} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ with $\lambda_1 \coloneqq \lambda_1(\mathbf{\Sigma}) > 0$, and let $\mathbf{w} \in \mathbb{R}^d$ be a unit vector. For every $u \in (0,1)$, let $\mathbf{P}_u \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ project onto the eigenvectors of $\mathbf{\Sigma}$ with eigenvalues $< (1-u)\lambda_1$, and define $\Delta_u \coloneqq \|\mathbf{P}_u \mathbf{w}\|_2^2$. Then*

$$\mathbf{I}_d - \frac{\mathbf{\Sigma}}{\lambda_1} = \int_0^1 \mathbf{P}_u \mathrm{d}u. \tag{18}$$

In particular,

$$1 - \frac{\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}}{\lambda_1} = \int_0^1 \Delta_u \mathrm{d}u. \tag{19}$$

Consequently, if $\alpha \in (0,1)$ and $\int_0^1 \Delta_u \mathrm{d}u \leq \alpha$, then $\mathbf{w}$ is an α-ePCA of $\mathbf{\Sigma}$.

Proof. Write $\mathbf{\Sigma} = \sum_{j\in[d]} \lambda_j \mathbf{u}_j \mathbf{u}_j^\top$. Then since

$$\mathbf{P}_u = \sum_{j\in[d]} \mathbb{I}_{\left\{u < 1 - \frac{\lambda_j}{\lambda_1}\right\}} \mathbf{u}_j \mathbf{u}_j^\top,$$

integrating each indicator gives the matrix identity in (18). Taking the quadratic form with $\mathbf{w}$ proves the integral identity (19) and the claim. □

We apply Proposition 2 to the output $\mathbf{w}_n$ of Algorithm 1, so henceforth $\Delta_u \coloneqq \|\mathbf{P}_u \mathbf{w}_n\|_2^2$. Suppose $\eta_t \leq \frac{1}{\lambda_1}$ for all $t \in [n]$ and $V \sum_{t\in[n]} \eta_t^2$ is a sufficiently small constant. For each fixed $u \in (0,1)$, combining Lemmas 1 and 3 with $\gamma \leftarrow u$ and fixed constant ζ gives, for a universal constant C,

$$\Delta_u \leq C \min\left\{1, d\exp\left(-u\lambda_1 \sum_{t\in[n]} \eta_t\right) + V \sum_{t\in[n]} \eta_t^2\right\}.$$

These bounds hold with constant probability for each fixed u, but need not hold simultaneously for all u. Even if the bound held simultaneously for every u, integrating it would only give

$$\int_0^1 \Delta_u \mathrm{d}u = O\left(\frac{\log(ed)}{\lambda_1 \sum_{t\in[n]} \eta_t} + V \sum_{t\in[n]} \eta_t^2\right). \tag{20}$$

For step sizes $\eta_t = \frac{L}{n+t}$, making both terms in the integral bound (20) at most α requires

$$L = \Omega\left(\frac{\log(ed)}{\lambda_1 \alpha}\right), \quad n = \Omega\left(\frac{VL^2}{\alpha}\right) = \Omega\left(\frac{V\log^2(ed)}{\lambda_1^2 \alpha^3}\right),$$

which is still suboptimal. The following lemma gives a sharper bound on $\int_0^1 \Delta_u \mathrm{d}u$.

Lemma 9. *Under Model 1 with the products $\mathbf{B}_t, \mathbf{C}_t$ defined in (1), suppose $0 < \eta_s \le \frac{1}{\lambda_1}$ for all $s \in [n]$. For $u \in (0,1)$, define $\Delta_u := \|\mathbf{P}_u \mathbf{w}_n\|_2^2$ for the projector $\mathbf{P}_u$ in Proposition 2. There are universal constants $c, C > 0$ such that, if $V \sum_{s\in[n]} \eta_s^2 \le c$, then, with probability $\ge \frac{3}{4}$,*

$$\int_0^1 \Delta_u \mathrm{d}u \le C \left(\frac{\log(ed)}{\lambda_1 \sum_{s\in[n]} \eta_s} + V \sum_{i\in[n]} \eta_i^2 \min\left\{1, \frac{1}{\lambda_1 \sum_{s=i+1}^n \eta_s}\right\} \right). \tag{21}$$

Here the minimum for $i = n$ is interpreted as 1.

The proof retains the dependence on u in the bound on $\mathbb{E}\,\|\mathbf{P}_u \mathbf{B}_n\|_{\mathrm{F}}^2$. After division by $\|\mathbf{C}_n\|_{\mathrm{F}}^2$, this bound contains, up to a universal constant,

$$V \sum_{i\in[n]} \eta_i^2 \exp\left(-\frac{u\lambda_1}{2} \sum_{s=i+1}^n \eta_s \right).$$

Integrating each summand over u gives the factor $\min\{1, 1/(\lambda_1 \sum_{s=i+1}^n \eta_s)\}$ in (21), up to a universal constant. In Appendix F.1, we provide a proof of Lemma 9 where we integrate the bounds on $\mathbb{E}\,\|\mathbf{P}_u \mathbf{B}_n\|_{\mathrm{F}}^2$ over u before applying Markov's inequality and controlling the normalization of $\mathbf{w}_n$. This avoids a simultaneous union bound over all thresholds $u \in (0,1)$.

We next evaluate the error bound for the schedule used in both this section and our ePCA application in Section 3.

Lemma 10. *Under Model 1, let $a \ge 1$, $n \ge 4a$, and $\eta_t = \frac{a}{\lambda_1(n/4+t)}$ for $t \in [n]$. There are universal constants $c, C > 0$ such that, if $V \sum_{t\in[n]} \eta_t^2 \le c$, then with probability at least $\frac{3}{4}$,*

$$1 - \frac{\mathbf{w}_n^\top \mathbf{\Sigma} \mathbf{w}_n}{\lambda_1} \le C \left(\frac{\log(ed)}{a} + \frac{Va\log(ea)}{\lambda_1^2 n} \right). \tag{22}$$

Proof. Our strategy is to bound the two terms in Lemma 9 for this schedule, then apply Proposition 2 to obtain the ePCA bound. Since $\eta_t \lambda_1 \le \frac{4a}{n} \le 1$, Lemma 9 applies. The schedule satisfies

$$\lambda_1 \sum_{t\in[n]} \eta_t \ge \frac{4a}{5}, \quad \lambda_1 \sum_{s=i+1}^n \eta_s \ge \frac{4a(n-i)}{5n}, \quad \eta_i^2 \le \frac{16a^2}{\lambda_1^2 n^2}.$$

The bound on $\lambda_1 \sum_{t\in[n]} \eta_t$ controls the first term in Lemma 9 by $O(\frac{\log(ed)}{a})$. To bound the remaining variance sum, we write $k = n - i$ and use the bounds on $\lambda_1 \sum_{s=i+1}^n \eta_s$ and η_i^2:

$$\sum_{i\in[n]} \eta_i^2 \min\left\{1, \frac{1}{\lambda_1 \sum_{s=i+1}^n \eta_s}\right\} \le \frac{16a^2}{\lambda_1^2 n^2} \left(1 + \sum_{k\in[n-1]} \min\left\{1, \frac{5n}{4ak}\right\} \right) = O\left(\frac{a\log(ea)}{\lambda_1^2 n} \right).$$

Finally, applying Lemma 9 with our choices of η_t, and Proposition 2 with $\mathbf{w} \leftarrow \mathbf{w}_n$, gives with probability at least $\frac{3}{4}$,

$$1 - \frac{\mathbf{w}_n^\top \mathbf{\Sigma} \mathbf{w}_n}{\lambda_1} = \int_0^1 \Delta_u \mathrm{d}u \le C \left(\frac{\log(ed)}{a} + \frac{Va\log(ea)}{\lambda_1^2 n} \right).$$

□

We therefore obtain the following ePCA guarantee.

Theorem 6. *Under Model 1, let $\alpha \in (0, \frac{1}{4})$. Then if*

$$n = \Omega\left(\frac{V \log^2\left(\frac{d}{\alpha}\right)}{\lambda_1^2 \alpha^2} + \frac{\log\left(\frac{d}{\alpha}\right)}{\alpha} \right)$$

for an appropriate constant, then there exists $a \in \mathbb{R}_{>0}$ such that taking $\eta_t = \frac{a}{\lambda_1(n/4+t)}$ for $t \in [n]$, the output of Algorithm 1 is an α-ePCA of $\mathbf{\Sigma}$ with probability $\ge \frac{3}{4}$.

Proof. Let $L := \log(\frac{d}{\alpha})$ and take $a = \frac{C_1 L}{\alpha}$ for a sufficiently large universal constant C_1. Then our assumptions ensure $n \geq 4a$ and

$$V \sum_{t \in [n]} \eta_t^2 \leq \frac{4Va^2}{\lambda_1^2 n} \leq c,$$

where c is the constant in Lemma 10. Applying Lemma 10 with $a \leftarrow \frac{C_1 L}{\alpha}$ and using $\log(ea) = O(L)$ gives, with probability at least $\frac{3}{4}$,

$$1 - \frac{\mathbf{w}_n^\top \mathbf{\Sigma} \mathbf{w}_n}{\lambda_1} = O\left(\frac{L}{a} + \frac{VaL}{\lambda_1^2 n}\right) = O\left(\frac{\alpha}{C_1} + \frac{C_1 \alpha}{C}\right).$$

Choosing C_1 and then C sufficiently large makes the error at most α. □

Although Theorem 6 is stated with a constant failure probability, it is straightforward to use holdout samples to reduce its failure probability.

Corollary 1. *Let $(\zeta, \alpha) \in (0, \frac{1}{4})^2$. Under Model 1, if*

$$n = \Omega\left(\frac{V \log^2\left(\frac{d}{\alpha}\right) \log(\frac{1}{\zeta})}{\lambda_1^2 \alpha^2} + \frac{\log\left(\frac{d}{\alpha}\right) \log(\frac{1}{\zeta})}{\alpha}\right)$$

for an appropriate constant, we can obtain an α-ePCA of $\mathbf{\Sigma}$ with probability $\geq 1 - \zeta$.

Proof. Fix some unit vector $\mathbf{w} \in \mathbb{R}^d$. We claim that we can estimate its quadratic form $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$ up to additive error $\frac{\alpha \lambda_1}{4}$ using $O(\frac{V}{\alpha^2 \lambda_1^2})$ holdout samples. To see this, under Model 1, a single sample quadratic form $\mathbf{w}^\top \mathbf{A}_t \mathbf{w}$ is unbiased for $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$, and has variance at most V:

$$\mathbb{E}\left[\left(\mathbf{w}^\top (\mathbf{A}_t - \mathbf{\Sigma}) \mathbf{w}\right)^2\right] \leq \mathbb{E} \|(\mathbf{A}_t - \mathbf{\Sigma}) \mathbf{w}\|_2^2 \leq V.$$

Thus, averaging independent estimates and applying Chebyshev's inequality gives the claim.

Now, calling Theorem 6 $O(\log \frac{1}{\zeta})$ times independently with $\alpha \leftarrow \frac{\alpha}{2}$ implies that with probability $\geq 1 - \frac{\zeta}{2}$, at least one of the outputs will be an $\frac{\alpha}{2}$-ePCA. Taking the median of $O(\log \frac{1}{\zeta})$ estimates of the quadratic form $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$ obtained by each output $\mathbf{w}$, using independent holdout samples, then yields the unit vector with largest quadratic form up to additive error $\frac{\alpha \lambda_1}{2}$, concluding the proof. Note that the same holdout samples are simultaneously accurate for each output by applying independence and taking a union bound, and do not dominate the stated sample complexity. □

Matching lower bound. We briefly conclude the section by showing a matching lower bound, up to logarithmic factors, by appealing to Theorem 4.

Corollary 2. *Fix $d \geq 2$ and any choice of $\lambda_1 > 0$, $V > 0$, $\alpha \in (0, \frac{1}{32})$. There is no algorithm $\mathcal{A}$ that takes as input $\{\mathbf{A}_i\}_{i \in [n]}$ from Model 1, and outputs an α-ePCA of $\mathbf{\Sigma}$ with probability $\geq \frac{2}{3}$, even assuming that $\mathbf{A}_i \in \mathbb{S}_{\succeq \mathbf{0}}^{d \times d}$ for all $i \in [n]$, unless for an appropriate constant,*

$$n = \Omega\left(\frac{V}{\lambda_1^2 \alpha^2}\right).$$

Proof. By Lemma 7, Jambulapati et al. (2024) with $k = 1$, an α-ePCA of $\mathbf{\Sigma}$ is also a $(\gamma, \frac{\alpha}{\gamma})$-cPCA for every $\gamma \in (\alpha, 1)$, in the notation of Definition 1. Then, taking $\gamma \leftarrow 8\alpha$ and $\Delta \leftarrow \frac{1}{8}$ in Theorem 4 gives

$$n = \Omega\left(\frac{V}{\lambda_1^2 (8\alpha)^2 \cdot \frac{1}{8}}\right) = \Omega\left(\frac{V}{\lambda_1^2 \alpha^2}\right).$$

□

F.1 Proof of Lemma 9

Now we prove Lemma 9, our multiscale cPCA guarantee on Algorithm 1.

Our strategy is to first bound $\mathbb{E}\,\|\mathbf{P}_u\mathbf{B}_n\|_{\mathrm{F}}^2$ for each threshold u, and then give a normalization argument needed to integrate this bound. Throughout, work under Model 1 with the products $\mathbf{B}_s, \mathbf{C}_s$ defined in (1) and $0 < \eta_s \le \frac{1}{\lambda_1}$. For convenience, we also define

$$q_n \coloneqq V \sum_{s\in[n]} \eta_s^2, \qquad Z_s \coloneqq \|\mathbf{C}_s\|_{\mathrm{F}}^2, \qquad N_s(u) \coloneqq \mathbb{E}\,\|\mathbf{P}_u\mathbf{B}_s\|_{\mathrm{F}}^2,$$

where $\mathbf{P}_u$ is the projector from Proposition 2.

Lemma 11. *For every $u \in (0,1)$,*

$$\frac{N_n(u)}{Z_n} \le \exp(q_n)\left(\min\left\{1, d(1+q_n)\exp\left(-\frac{u\lambda_1}{2}\sum_{s\in[n]}\eta_s\right)\right\} + V\sum_{i\in[n]}\eta_i^2\exp\left(-\frac{u\lambda_1}{2}\sum_{s=i+1}^{n}\eta_s\right)\right). \tag{23}$$

Proof. Our proof strategy is to bound the initial contribution and the variance introduced at each update separately. For this, we first expand the second-moment recurrence, then divide by Z_n and bound the two contributions. Finally, we combine these bounds with Lemma 2.

Define $r_s(u) \coloneqq 1 + \eta_s(1-u)\lambda_1$. Taking the trace against $\mathbf{P}_u$ in the second moment recurrence (3) and using the variance bound in Model 1 then gives

$$N_s(u) \le r_s(u)^2 N_{s-1}(u) + V\eta_s^2\,\mathbb{E}\,\|\mathbf{B}_{s-1}\|_{\mathrm{F}}^2.$$

Iterating from $N_0(u) \le d$ and using Lemma 2 with $t \leftarrow i-1$ for $i > 1$ (and $\mathbf{B}_0 = \mathbf{C}_0 = \mathbf{I}_d$ for $i = 1$), we obtain

$$N_n(u) \le \underbrace{d\prod_{s\in[n]} r_s(u)^2}_{\text{initial contribution}} + \underbrace{V\exp(q_n)\sum_{i\in[n]}\eta_i^2 Z_{i-1}\prod_{s=i+1}^{n} r_s(u)^2}_{\text{variance term}}. \tag{24}$$

The initial contribution comes from $N_0(u) \le d$, multiplied by the recurrence factors over all n updates. The variance term sums the contribution introduced at each update i, multiplied by the factors from the subsequent updates. We first bound the initial contribution after dividing by Z_n, by comparing with the product for λ_1. The assumption $\eta_s\lambda_1 \le 1$ gives

$$\left(\frac{1+\eta_s\mu}{1+\eta_s\mu'}\right)^2 \le \exp(-\eta_s(\mu'-\mu)) \qquad \text{for } 0 \le \mu \le \mu' \le \lambda_1. \tag{25}$$

Since $Z_n \ge \prod_{s\in[n]}(1+\eta_s\lambda_1)^2$, the first term in (24), divided by Z_n, is at most

$$d\exp\left(-u\lambda_1\sum_{s\in[n]}\eta_s\right).$$

It remains to bound the variance terms in (24) after division by Z_n. For the ith summand, expand

$$Z_{i-1} = \sum_{j\in[d]}\prod_{s<i}(1+\eta_s\lambda_j)^2$$

and separate the eigenvalues at $(1-\frac{u}{2})\lambda_1$. Above this threshold, we compare the remaining factors with those for the same eigenvalue in Z_n; below it, we compare all factors with those for λ_1. In both cases the compared eigenvalues differ by at least $\frac{u\lambda_1}{2}$. For $\lambda_j \ge (1-\frac{u}{2})\lambda_1$, (25) gives

$$\prod_{s<i}(1+\eta_s\lambda_j)^2\prod_{s=i+1}^{n} r_s(u)^2 \le \exp\left(-\frac{u\lambda_1}{2}\sum_{s=i+1}^{n}\eta_s\right)\prod_{s\in[n]}(1+\eta_s\lambda_j)^2.$$

Summing over j with $\lambda_j \geq (1-\frac{u}{2})\lambda_1$ bounds their total contribution by $Z_n \exp(-\frac{u\lambda_1}{2}\sum_{s=i+1}^{n}\eta_s)$. For $\lambda_j < (1-\frac{u}{2})\lambda_1$, the scalar ratio bound in (25) and $\prod_{s\in[n]}(1+\eta_s\lambda_1)^2 \leq Z_n$ give

$$\prod_{s<i}(1+\eta_s\lambda_j)^2 \prod_{s=i+1}^{n} r_s(u)^2 \leq \prod_{s\in[n]}\left(1+\eta_s\left(1-\frac{u}{2}\right)\lambda_1\right)^2 \leq Z_n \exp\left(-\frac{u\lambda_1}{2}\sum_{s\in[n]}\eta_s\right).$$

There are at most d such j, so the two ranges together give

$$\frac{Z_{i-1}\prod_{s=i+1}^{n} r_s(u)^2}{Z_n} \leq \exp\left(-\frac{u\lambda_1}{2}\sum_{s=i+1}^{n}\eta_s\right) + d\exp\left(-\frac{u\lambda_1}{2}\sum_{s\in[n]}\eta_s\right).$$

Substituting the bounds for both terms into (24) and using $V\sum_{i\in[n]}\eta_i^2 = q_n$ gives the claimed bound without the minimum with 1. To obtain that minimum, apply Lemma 2 with $t \leftarrow n$ and use $\|\mathbf{P}_u\mathbf{B}_n\|_{\mathrm{F}} \leq \|\mathbf{B}_n\|_{\mathrm{F}}$ to get $\frac{N_n(u)}{Z_n} \leq \exp(q_n)$. Combining this bound on $\frac{N_n(u)}{Z_n}$ with the bound obtained from (24), using $\min\{1, x+y\} \leq \min\{1,x\}+y$ for $x, y \geq 0$ proves the claim. □

Lemma 12. *There are universal constants $c, C > 0$ such that, if $q_n \leq c$, then for any fixed $\mathbf{D} \succeq \mathbf{0}_{d\times d}$ and independent $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$, with probability at least $\frac{3}{4}$, $\mathbf{B}_n\mathbf{g} \neq \mathbf{0}_d$ and*

$$\frac{\mathbf{g}^\top\mathbf{B}_n^\top\mathbf{D}\mathbf{B}_n\mathbf{g}}{\|\mathbf{B}_n\mathbf{g}\|_2^2} \leq C\frac{\mathbb{E}\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top)}{Z_n}.$$

Proof. Our proof strategy is to control the numerator and denominator of the normalized output using bounds on $\mathbf{B}_n$. We first lower bound $\|\mathbf{B}_n\|_{\mathrm{F}}^2$ and upper bound $\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top)$. We then condition on $\mathbf{B}_n$, apply the Gaussian quadratic-form bounds, and combine the three events.

Choose c so that $\exp(c) - 1 \leq \frac{1}{48}$. Since $\mathbb{E}\mathbf{B}_n = \mathbf{C}_n$, Lemma 2 with $t \leftarrow n$ gives

$$\mathbb{E}\|\mathbf{B}_n - \mathbf{C}_n\|_{\mathrm{F}}^2 = \mathbb{E}\|\mathbf{B}_n\|_{\mathrm{F}}^2 - Z_n \leq (\exp(q_n)-1)Z_n.$$

Markov's inequality gives events

$$\mathcal{E}_B := \left\{\|\mathbf{B}_n - \mathbf{C}_n\|_{\mathrm{F}}^2 \leq \frac{Z_n}{4}\right\}, \qquad \mathcal{E}_D := \left\{\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top) \leq 12\mathbb{E}\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top)\right\},$$

each with failure probability at most $\frac{1}{12}$. Fix $\mathbf{B}_n$ in $\mathcal{E}_B \cap \mathcal{E}_D$; then $\|\mathbf{B}_n\|_{\mathrm{F}}^2 \geq \frac{Z_n}{4}$, and $\mathbf{g}$ remains an independent standard Gaussian.

We have thus obtained the required bounds on $\mathbf{B}_n$; it remains to control the ratio of quadratic forms in $\mathbf{g}$. The proof of Lemma 1 applies with $\mathbf{H} = \mathbf{B}_n^\top\mathbf{B}_n$ and $\mathbf{K} = \mathbf{B}_n^\top\mathbf{D}\mathbf{B}_n$. Its Gaussian quadratic-form bounds require only $\mathbf{H}, \mathbf{K} \succeq \mathbf{0}_{d\times d}$ and $\mathbf{H} \neq \mathbf{0}_{d\times d}$. With $\zeta = \frac{1}{12}$, it gives an event

$$\mathcal{E}_g := \left\{\|\mathbf{B}_n\mathbf{g}\|_2^2 > 0, \quad \frac{\mathbf{g}^\top\mathbf{B}_n^\top\mathbf{D}\mathbf{B}_n\mathbf{g}}{\|\mathbf{B}_n\mathbf{g}\|_2^2} \leq C_0\frac{\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top)}{\|\mathbf{B}_n\|_{\mathrm{F}}^2}\right\},$$

with conditional failure probability at most $\frac{1}{12}$, for a universal constant C_0. On $\mathcal{E}_B \cap \mathcal{E}_D \cap \mathcal{E}_g$, the claimed bound holds with $C = 48C_0$. The total failure probability is at most $3 \cdot \frac{1}{12} = \frac{1}{4}$. □

Proof of Lemma 9. Our strategy is to bound the integral by an expected trace using Lemma 12, then estimate this trace using Lemma 11. For this, we first write the integral as a quadratic form.

Set $\mathbf{D} := \mathbf{I}_d - \frac{\mathbf{\Sigma}}{\lambda_1} \succeq \mathbf{0}_{d\times d}$. Proposition 2, applied to $\mathbf{\Sigma}$, gives $\mathbf{D} = \int_0^1 \mathbf{P}_u\,\mathrm{d}u$. On the event $\mathbf{B}_n\mathbf{g} \neq \mathbf{0}_d$, write $\mathbf{w}_n = \mathbf{B}_n\mathbf{g}/\|\mathbf{B}_n\mathbf{g}\|_2$ and $\Delta_u = \|\mathbf{P}_u\mathbf{w}_n\|_2^2$. Then

$$\int_0^1 \Delta_u\,\mathrm{d}u = \mathbf{w}_n^\top\mathbf{D}\mathbf{w}_n = \frac{\mathbf{g}^\top\mathbf{B}_n^\top\mathbf{D}\mathbf{B}_n\mathbf{g}}{\|\mathbf{B}_n\mathbf{g}\|_2^2}.$$

Applying Lemma 12 with $\mathbf{D} \leftarrow \mathbf{I}_d - \frac{\mathbf{\Sigma}}{\lambda_1}$ and using linearity of trace and expectation therefore gives, with probability at least $\frac{3}{4}$,

$$\int_0^1 \Delta_u \, \mathrm{d}u \le C \frac{\mathbb{E}\mathrm{Tr}(\mathbf{D}\mathbf{B}_n\mathbf{B}_n^\top)}{Z_n} = C \int_0^1 \frac{N_n(u)}{Z_n} \, \mathrm{d}u. \tag{26}$$

We have reduced the desired bound to the integral of $\frac{N_n(u)}{Z_n}$ in (26). It remains to apply Lemma 11 and integrate its two terms. For $A \ge 1$ and $x > 0$, direct integration gives

$$\int_0^1 \min\left\{1, A\exp\left(-\frac{ux}{2}\right)\right\} \mathrm{d}u \le \frac{2(1+\log A)}{x}, \qquad \int_0^1 \exp\left(-\frac{ux}{2}\right) \mathrm{d}u \le 2\min\left\{1, \frac{1}{x}\right\}. \tag{27}$$

To conclude, we apply Lemma 11 for each $u \in (0,1)$. For its first term, use the first integration bound in (27) with $A = d(1+q_n)$ and $x = \lambda_1 \sum_{s\in[n]} \eta_s$. For its i^{th} variance summand with $i < n$, use the second integration bound with $x = \lambda_1 \sum_{s=i+1}^n \eta_s$; for $i = n$, the exponential is identically 1. Taking $c \le 1$, we have $\exp(q_n) = O(1)$ and $1 + \log(d(1+q_n)) = O(\log(ed))$, so

$$\int_0^1 \frac{N_n(u)}{Z_n} \, \mathrm{d}u \le C' \left(\frac{\log(ed)}{\lambda_1 \sum_{s\in[n]} \eta_s} + V \sum_{i\in[n]} \eta_i^2 \min\left\{1, \frac{1}{\lambda_1 \sum_{s=i+1}^n \eta_s}\right\} \right),$$

where $C' > 0$ is universal and the minimum for $i = n$ is interpreted as 1. Substituting this bound on $\int_0^1 \frac{N_n(u)}{Z_n} \, \mathrm{d}u$ into (26) proves (21). □

G Deferred Proofs from Section 3

G.1 Proof of Theorem 2

We start with the proof of Theorem 2, deferring the privacy and utility proofs.

Proof. We run Algorithm 2 with

$$R = \left\lceil 72 \log\left(\frac{8}{\zeta}\right) \right\rceil, \quad L = \log\left(\frac{256dR}{\Delta}\right), \quad T = \left\lceil \frac{40L}{\gamma} \right\rceil, \quad \eta_t = \frac{8L}{\lambda_1(10L + \gamma t)}, \quad \tau = \frac{1}{8}\sqrt{\frac{\Delta}{R}}.$$

Choose R_1, R_2 as in (32), δ_* as in (33), and set

$$\sigma^2 = \frac{12RTR_1R_2}{n^2\varepsilon^2} \log\left(\frac{1}{\delta_*}\right). \tag{28}$$

Privacy. Lemma 13 with $\delta \leftarrow \delta_*$ gives (ε, δ_*)-DP, and hence, (ε, δ)-DP.

Utility. Our utility proof strategy is to obtain a cPCA of $\widehat{\mathbf{\Sigma}}$ and transfer it to $\mathbf{\Sigma}$. For this, we first bound $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\text{op}}$, then couple the private updates to Oja and apply Lemma 8. We combine these guarantees on their common success event. Since $\lambda_1 \le \nu^2$, the first term in the sample complexity (4) and Theorem 6.5 of Wainwright (2019) give, with probability $\ge 1 - \frac{\zeta}{4}$,

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\text{op}} \le \frac{\lambda_1\gamma\sqrt{\Delta}}{4}, \qquad \frac{3}{4}\lambda_1 \le \left\|\widehat{\mathbf{\Sigma}}\right\|_{\text{op}} \le \frac{5}{4}\lambda_1. \tag{29}$$

Let $\mathcal{E}_{\text{cov}}$ denote the event that the covariance bounds in (29) hold. As in the proof of Proposition 1, it suffices on this event to obtain a $(\frac{\gamma}{6}, \frac{\Delta}{4})$-cPCA of $\widehat{\mathbf{\Sigma}}$.

Apply Lemma 15 with the chosen R, T, R_1, R_2, σ and failure parameter ζ. Let $\mathcal{E}_{\text{cpl}}$ denote the event that the two sequences of iterates agree throughout; then $\Pr(\mathcal{E}_{\text{cpl}}^c) \le \frac{\zeta}{4}$. We analyze these Oja updates conditional on a dataset satisfying the covariance bounds in (29); they are independent with mean $\widehat{\mathbf{\Sigma}}$ and variance $V = d\sigma^2$.

To obtain the required cPCA of $\widehat{\mathbf{\Sigma}}$, we check the hypotheses of Lemma 8 for the R coupled runs of T updates with $(\mathbf{\Sigma}, \gamma, V) \leftarrow (\widehat{\mathbf{\Sigma}}, \frac{\gamma}{6}, d\sigma^2)$, the chosen schedule $\{\eta_t\}_{t\in[T]}$, and radius τ. Its step-size condition holds since $\eta_t \left\|\widehat{\mathbf{\Sigma}}\right\|_{\mathrm{op}} \le 1$. Moreover, our parameter choices give

$$4d \exp\left(-\frac{\gamma}{6}\left\|\widehat{\mathbf{\Sigma}}\right\|_{\mathrm{op}} \sum_{t\in[T]} \eta_t\right) \le \tau^2, \quad q_T \coloneqq d\sigma^2 \sum_{t\in[T]} \eta_t^2 \le \frac{\tau^2}{128}. \tag{30}$$

The exponential bound in (30) uses the choices of T and η_t. The bound on q_T follows by substituting R_1, R_2 from (32) and σ from (28), and using the second term in the sample bound (4); the sample bound (4) ensures $R, T = O(n)$ and hence $\log(\frac{1}{\delta_*}) = O(\log \frac{n}{\delta\zeta})$.

Having verified the hypotheses of Lemma 8, we apply it for every fixed S in $\mathcal{E}_{\mathrm{cov}}$ to get

$$\|\mathbf{P}\mathbf{w}_{RT}\|_2^2 \le 16R\tau^2 = \frac{\Delta}{4}$$

with failure probability at most $4\exp(-\frac{R}{72}) \le \frac{\zeta}{2}$, where $\mathbf{P}$ projects onto eigenvectors of $\widehat{\mathbf{\Sigma}}$ with eigenvalues below $(1-\frac{\gamma}{6})\|\widehat{\mathbf{\Sigma}}\|_{\mathrm{op}}$. Thus, if $\mathcal{E}_{\mathrm{Oja}}$ denotes the event that the coupled Oja output is a $(\frac{\gamma}{6}, \frac{\Delta}{4})$-cPCA of $\widehat{\mathbf{\Sigma}}$, then $\Pr(\mathcal{E}_{\mathrm{Oja}}^c \mid S) \le \frac{\zeta}{2}$.

We have thus obtained the required cPCA of $\widehat{\mathbf{\Sigma}}$. On $\mathcal{E}_{\mathrm{cov}} \cap \mathcal{E}_{\mathrm{cpl}} \cap \mathcal{E}_{\mathrm{Oja}}$, the private output agrees with this Oja output, and the covariance bounds in (29), via the proof of Proposition 1, make it a (γ, Δ)-cPCA of $\mathbf{\Sigma}$. The proof follows by noting that

$$\Pr(\mathcal{E}_{\mathrm{cov}}^c) + \Pr(\mathcal{E}_{\mathrm{cpl}}^c) + \Pr(\mathcal{E}_{\mathrm{Oja}}^c \cap \mathcal{E}_{\mathrm{cov}}) \le \frac{\zeta}{4} + \frac{\zeta}{4} + \frac{\zeta}{2} = \zeta.$$

□

G.1.1 Privacy

We prove that Algorithm 2 satisfies (ε, δ)-DP when σ in Line 15 is appropriately chosen. Our proof is standard, and proceeds via *Rényi DP*, an alternative privacy accounting strategy that is particularly well-suited to the *Gaussian mechanism*. For brevity, we defer background on the Gaussian mechanism to Appendix A of Dwork & Roth (2014), and background on Rényi DP to Mironov (2017).

Lemma 13. *For $\varepsilon \in (0, 1]$ and $\delta \in (0, \frac{1}{3})$, the output of Algorithm 2 is (ε, δ)-DP if*

$$\sigma^2 \ge \frac{12RTR_1R_2}{n^2\varepsilon^2} \log\left(\frac{1}{\delta}\right). \tag{31}$$

Moreover, every $\mathbf{U}_r$ on Line 8 is (ε, δ)-DP for all $r \in [R]$, at every iteration $t \in [T]$.

Proof. The algorithm only accesses the dataset $\{\mathbf{x}_i\}_{i\in[n]}$ in the nested for loops from Lines 6 to 18. Fix the initialization and preceding noisy answers at the beginning of one loop, indexed by $t \in [T]$ and $r \in [R]$, so that $\mathbf{Y}_r$ and $\mathbf{U}_r$ are fixed. Each clipped summand to $\mathbf{Q}_{r,t}$ satisfies

$$\left\|\mathbf{z}_i\mathbf{z}_i^\top \mathbf{U}_r\right\|_{\mathrm{F}} = \|\mathbf{z}_i\|_2 \left\|\mathbf{U}_r^\top \mathbf{z}_i\right\|_2 \le \sqrt{R_1R_2},$$

and $\mathbf{z}_i$ is a deterministic function of $\mathbf{U}_r$ and $\mathbf{x}_i$, so replacing one sample $\mathbf{x}_i$ changes $\mathbf{Q}_{r,t}$ in Frobenius norm by $\le s \coloneqq \frac{2}{n}\sqrt{R_1R_2}$. Thus, Proposition 7 and Corollary 3 of Mironov (2017) show that the noisy answer $\mathbf{Q}_{r,t} + \mathbf{H}_{r,t}$ is $(p, \frac{ps^2}{2\sigma^2})$-RDP, and $\mathbf{Y}_r$ is updated by a deterministic function of this answer and the preceding state. Now RDP composition (Proposition 1, Mironov (2017)) over RT iterations shows that the transcript of all $\mathbf{Y}_r$ is (p, ρ)-RDP, where

$$\rho = \frac{ps^2RT}{2\sigma^2}.$$

Finally, taking $p = 1 + \frac{2\log(\frac{1}{\delta})}{\varepsilon}$, and using the lower bound on σ^2 in (31), implies

$$\frac{\log\left(\frac{1}{\delta}\right)}{p-1} \le \frac{\varepsilon}{2}, \quad \rho \le \frac{ps^2n^2\varepsilon^2}{24R_1R_2\log\left(\frac{1}{\delta}\right)} = \frac{p\varepsilon^2}{6\log\left(\frac{1}{\delta}\right)} \le \frac{\varepsilon^2}{6\log\left(\frac{1}{\delta}\right)} + \frac{\varepsilon}{3} \le \frac{\varepsilon}{2}.$$

Proposition 3 of Mironov (2017) then shows that the transcript of all of the $\mathbf{Y}_r$ is (ε, δ)-DP. The privacy of the algorithm's output and all $\mathbf{U}_r$ then follows, as postprocessings of the transcript. □

G.1.2 Utility

We next give our utility analysis. For convenience, denote the dataset and empirical covariance by

$$S \coloneqq \{\mathbf{x}_i\}_{i\in[n]}, \quad \widehat{\mathbf{\Sigma}} \coloneqq \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top.$$

We first show how to couple iterates of Algorithm 2 to an instance of Model 1. To begin, we show that with high probability, the clipping events on Lines 11 and 12 never occur. This step requires using our earlier privacy guarantee to handle a dependence between $\mathbf{U}$ and the dataset S.

Lemma 14. *Under Model 2, let $\mathcal{A}$ be an (ε, δ)-DP mechanism whose output $\mathbf{U} = \mathcal{A}(S)$ is an orthonormal matrix with at most R columns. For every $i \in [n]$ and $u > 0$,*

$$\Pr\left(\left\|\mathbf{U}^\top \mathbf{x}_i\right\|_2^2 > 3\nu^2(R+u)\right) \le \exp(\varepsilon - u) + \delta.$$

Proof. Replace $\mathbf{x}_i$ by an independent copy to form $S^{(i)}$, and set $\mathbf{U}' = \mathcal{A}(S^{(i)})$. Conditional on $\mathbf{U}'$, the vector $\mathbf{x}_i$ remains mean-zero and ν-sub-Gaussian. Theorem 2.1 of Hsu et al. (2012), applied to $(\mathbf{U}')^\top\mathbf{x}_i$, gives $\Pr(\left\|(\mathbf{U}')^\top\mathbf{x}_i\right\|_2^2 > 3\nu^2(R+u)) \le \exp(-u)$. Applying Definition 2 to the neighboring datasets $S, S^{(i)}$ and averaging over the independent copy gives the claim. □

For a failure probability parameter $\zeta \in (0,1)$, set

$$R_1 \coloneqq 3\nu^2\left(d + \log\left(\frac{8n}{\zeta}\right)\right), \qquad R_2 \coloneqq 3\nu^2\left(R + 2\log\left(\frac{16nRT}{\zeta}\right)\right). \tag{32}$$

We use the smaller privacy failure parameter

$$\delta_* \coloneqq \min\left\{\delta, \frac{\zeta}{16nRT}\right\}, \tag{33}$$

so that the additive privacy errors can be summed over all sample projections.

Lemma 15. *Under Model 2, let $\varepsilon \in (0,1]$, $\delta \in (0, \frac{1}{3})$, and $\zeta \in (0,1)$. Choose R_1, R_2 as in (32), and suppose σ^2 satisfies the lower bound in (31) with $\delta \leftarrow \delta_*$ from (33). There exist matrices*

$$\mathbf{A}_{(t-1)R+r} \coloneqq \widehat{\mathbf{\Sigma}} + \mathbf{G}_{r,t},$$

where the $\mathbf{G}_{r,t} \in \mathbb{R}^{d\times d}$ have independent $\mathcal{N}(0,\sigma^2)$ entries, such that the updates in Algorithm 2 can be coupled to $\mathbf{Y}_r \leftarrow (\mathbf{I}_d + \eta_t\mathbf{A}_{(t-1)R+r})\mathbf{Y}_r$ with probability $\ge 1 - \frac{\zeta}{4}$. Conditional on the dataset, these matrices are i.i.d. with mean $\widehat{\mathbf{\Sigma}}$ and satisfy both variance bounds in Model 1 with $V = d\sigma^2$.

Proof. We first show that neither clipping step changes any sample, except with probability $\frac{\zeta}{4}$. We then choose the Gaussian matrices so that the private and Oja updates agree whenever no clipping occurs. By Theorem 2.1 of Hsu et al. (2012) and the choice of R_1,

$$\Pr\left(\exists i \in [n] : \|\mathbf{x}_i\|_2^2 > R_1\right) \le n\exp\left(-\log\left(\frac{8n}{\zeta}\right)\right) = \frac{\zeta}{8}.$$

By Lemma 13 with $\delta \leftarrow \delta_*$, each subspace $\mathbf{U}_r$ computed by the algorithm is (ε, δ_*)-DP. For Line 12, apply Lemma 14 with $u = 2\log(\frac{16nRT}{\zeta})$, so that $R_2 = 3\nu^2(R+u)$. Since $\varepsilon \le 1 \le \log(\frac{16nRT}{\zeta})$, for every $(i,r,t) \in [n]\times[R]\times[T]$,

$$\Pr\left(\left\|\mathbf{U}_r^\top\mathbf{x}_i\right\|_2^2 > R_2\right) \le \exp(\varepsilon - u) + \delta_* \le \frac{\zeta}{16nRT} + \frac{\zeta}{16nRT} = \frac{\zeta}{8nRT}.$$

These bounds apply to the subspaces computed by the algorithm, including when earlier samples were clipped. A union bound gives total failure probability at most $\frac{\zeta}{8} + nRT\cdot\frac{\zeta}{8nRT} = \frac{\zeta}{4}$. On the complementary event, $\mathbf{z}_i = \mathbf{x}_i$ for every sample in every update.

Now draw the $d \times d$ matrices $\mathbf{G}_{r,t}$ with independent $\mathcal{N}(0, \sigma^2)$ entries, independently across updates and independently of the data and initialization. Set $\mathbf{H}_{r,t} = \mathbf{G}_{r,t}\mathbf{U}_r$. Given the dataset and all preceding updates, $\mathbf{U}_r$ is fixed and orthonormal, so $\mathbf{H}_{r,t}$ has independent $\mathcal{N}(0, \sigma^2)$ entries, as required by the algorithm. Whenever no clipping occurs, $\mathbf{Q}_{r,t} = \widehat{\mathbf{\Sigma}}\mathbf{U}_r$, and the SVD identity gives

$$\mathbf{Y}_r + \eta_t(\widehat{\mathbf{\Sigma}}\mathbf{U}_r + \mathbf{G}_{r,t}\mathbf{U}_r)\mathbf{D}_r\mathbf{V}_r^\top = (\mathbf{I}_d + \eta_t(\widehat{\mathbf{\Sigma}} + \mathbf{G}_{r,t}))\mathbf{Y}_r.$$

Thus, starting from the same initialization, the private and Oja iterates agree with probability at least $1 - \frac{\zeta}{4}$. Conditional on the dataset, the matrices $\mathbf{A}_{(t-1)R+r}$ are i.i.d. with mean $\widehat{\mathbf{\Sigma}}$, and

$$\mathbb{E}[\mathbf{G}_{r,t}\mathbf{G}_{r,t}^\top] = \mathbb{E}[\mathbf{G}_{r,t}^\top\mathbf{G}_{r,t}] = d\sigma^2\mathbf{I}_d.$$

This verifies Model 1 with $V = d\sigma^2$, conditional only on the dataset. □

G.2 Proof of Theorem 3

Proof. We run Algorithm 2 with $R = \tau = 1$ and

$$a \leftarrow \frac{C_3 \log\left(\frac{d}{\alpha}\right)}{\alpha}, \quad T \leftarrow \lceil C_4 a \rceil, \quad \eta_t \leftarrow \frac{a}{\lambda_1(T/4 + t)},$$

for sufficiently large universal constants C_3, C_4. We choose R_1, R_2 from (32) with $\zeta = \frac{1}{12}$, set $\delta_* = \min\{\delta, \frac{1}{192nT}\}$ as in (33), and choose σ as in (31) with $\delta \leftarrow \delta_*$. The privacy proof is identical to Theorem 2, so we focus on the utility proof.

First, Lemma 15 with $(R, \zeta) \leftarrow (1, \frac{1}{12})$ gives coupled Oja updates with mean $\widehat{\mathbf{\Sigma}}$ and variance $V = d\sigma^2$. Let $\mathcal{E}_{\mathrm{cpl}}$ denote the event that Lemma 15 succeeds; then $\Pr(\mathcal{E}_{\mathrm{cpl}}^c) \le \frac{1}{48}$.

Next, let $\mathcal{E}_{\mathrm{cov}}$ denote the event $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathrm{op}} \le \frac{\alpha\lambda_1}{8}$. The first term in (5) and Theorem 6.5 of Wainwright (2019) give $\Pr(\mathcal{E}_{\mathrm{cov}}^c) \le \frac{1}{48}$. Fix a dataset S in $\mathcal{E}_{\mathrm{cov}}$ and analyze the Oja run conditional only on S.

To apply Lemma 10 to $\widehat{\mathbf{\Sigma}}$, we express the step sizes using its top eigenvalue. Writing $\widehat{\lambda}_1 := \|\widehat{\mathbf{\Sigma}}\|_{\mathrm{op}}$ and $a' := \frac{a\widehat{\lambda}_1}{\lambda_1}$, the schedule becomes $\eta_t = \frac{a'}{\widehat{\lambda}_1(T/4+t)}$, with $(1 - \frac{\alpha}{8})a \le a' \le (1 + \frac{\alpha}{8})a$ and $T \ge 4a'$.

The schedule now has the form required by Lemma 10; it remains to check its variance condition. Using $V = d\sigma^2$, σ from (31) with $(R, \delta) \leftarrow (1, \delta_*)$, and the sample bound (5), we obtain

$$V\sum_{t=1}^{T}\eta_t^2 \le \frac{4Va^2}{\lambda_1^2 T} = \frac{48dR_1R_2a^2\log\left(\frac{1}{\delta_*}\right)}{n^2\varepsilon^2\lambda_1^2} \le c,$$

for a sufficiently small universal constant $c > 0$. We may therefore apply Lemma 10 with $(\mathbf{\Sigma}, n, a) \leftarrow (\widehat{\mathbf{\Sigma}}, T, a')$. For every fixed S in $\mathcal{E}_{\mathrm{cov}}$, it gives, with probability at least $\frac{3}{4}$,

$$1 - \frac{\mathbf{w}_T^\top\widehat{\mathbf{\Sigma}}\mathbf{w}_T}{\widehat{\lambda}_1} \le C_0\left(\frac{\log(ed)}{a'} + \frac{Va'\log(ea')}{\widehat{\lambda}_1^2 T}\right) \le \frac{\alpha}{2}.$$

The error bound $\frac{\alpha}{2}$ uses $R_1R_2 = O(\nu^4 d\log^2(\frac{nd}{\alpha}))$, $\log(\frac{1}{\delta_*}) = O(\log(\frac{nd}{\alpha\delta}))$, and the sample bound (5), with C_3 and then C sufficiently large. Let $\mathcal{E}_{\mathrm{Oja}}$ denote this $\frac{\alpha}{2}$-ePCA guarantee for the coupled Oja output $\mathbf{w}_T$; then $\Pr(\mathcal{E}_{\mathrm{Oja}}^c \mid S) \le \frac{1}{4}$.

We have obtained an $\frac{\alpha}{2}$-ePCA of $\widehat{\mathbf{\Sigma}}$; it remains to transfer this guarantee to $\mathbf{\Sigma}$. On $\mathcal{E}_{\mathrm{cov}} \cap \mathcal{E}_{\mathrm{cpl}} \cap \mathcal{E}_{\mathrm{Oja}}$, coupling and the covariance bound give

$$\mathbf{w}_n^\top\mathbf{\Sigma}\mathbf{w}_n \ge \left(1 - \frac{\alpha}{2}\right)\widehat{\lambda}_1 - \frac{\alpha\lambda_1}{8} \ge (1 - \alpha)\lambda_1.$$

The private output $\mathbf{w}_n$ is an α-ePCA of $\mathbf{\Sigma}$ on these events. The failure probability follows from

$$\Pr(\mathcal{E}_{\mathrm{cov}}^c) + \Pr(\mathcal{E}_{\mathrm{cpl}}^c) + \Pr(\mathcal{E}_{\mathrm{Oja}}^c \cap \mathcal{E}_{\mathrm{cov}}) \le \frac{1}{48} + \frac{1}{48} + \frac{1}{4} < \frac{1}{3}.$$

□