WARM-STARTED OJA MATCHES POOLED PCA TO FIRST ORDER

Anonymous authors
Paper under double-blind review

ABSTRACT

Suppose historical observations are no longer available, but their leading empirical principal component, sample count, and spectral calibration have been retained. Can one-pass continuation track the PCA estimates that would be obtained by pooling the old data with every incoming prefix? In a fixed-dimensional, stationary rank-one isotropic Gaussian spike with iid historical and streaming observations and comparable sample sizes, we answer yes to first order. We prove that the actual normalized Oja recursion with a sample-count-shifted harmonic effective step is, uniformly over every prefix and on the same realization, within $o(N^{-1})$ expected maximal squared projector distance of both projector-regularized PCA and pooled empirical PCA. All three share the same influence path and, uniformly in the prefix, attain risk $(d-1)(\sigma^2+\Delta)\sigma^2/\{\Delta^2(\nu+k)\}+o(N^{-1})$. At the final prefix this is the Gaussian local asymptotic minimax constant. Clipped old-sample plug-in calibration preserves the conclusion, and the historical sample size is the unique first-order pseudocount. The projector-risk comparison extends to $d+\log N=o(N)$ at the natural $o(d/N)$ scale. The Oja proof is entirely discrete. Empirical Riemannian flows restarted on the same prefixes track the Oja path at calibrated times; a separate fixed-data time profile identifies pooled-efficient early stopping.

1 INTRODUCTION

When observations are discarded after a principal component is fitted, what information does that component retain for future learning? A device or representation pipeline may store a loading and its sample count but not the historical covariance. New observations can be processed by warm-starting Oja's streaming update (Oja, 1982), whereas exact pooled PCA would require the unavailable covariance. The question is not only whether the streaming estimate is accurate, but whether it reproduces the estimates that the discarded observations would have produced.

We study this question through *matched estimator trajectories*. After each new observation, compare the actual normalized Oja iterate, PCA regularized toward the retained historical projector, and ordinary PCA refitted on the old observations plus the new prefix. All three use the same underlying realization. Their path index counts arriving observations, not eigensolver iterations; only Oja updates in one pass from its previous direction. In a stationary rank-one isotropic Gaussian spike, with comparable historical and incoming sample sizes, a sample-count-calibrated harmonic schedule makes these paths agree to first order, uniformly over all prefixes. Their expected maximal squared projector discrepancy is smaller than their statistical risk. Thus the result identifies which realized fluctuations a warm start preserves, rather than merely matching convergence rates or endpoint limiting distributions.

Neither path comparisons nor efficient streaming estimation are new in themselves. Existing work compares gradient-flow and regularization paths (Ali et al., 2019; 2020), proves exact Oja–GROUSE equivalence (Balzano, 2022), and gives time-uniform Oja guarantees (Pham et al., 2025). Historical-projector compression (Yang et al., 2018), efficient recursive estimation (Sharia, 2010; Tripuraneni et al., 2018), and same-data distributed/pooled PCA comparisons (Chen et al., 2022; Li et al., 2026b) are also established. Our focus is the specific uniform, below-risk-scale comparison of the unaveraged Oja, anchored, and pooled *incoming-prefix* paths when the historical empirical pilot

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.

retains a nonvanishing first-order fluctuation. Section 7 and Appendix G compare the closest results and their different notions of a path.

Contributions. Our main contributions are:

1. **A discrete stochastic trajectory comparison.** The actual normalized Oja path tracks anchored and pooled PCA with $o(N^{-1})$ expected maximal squared projector discrepancy, where N is the total historical and incoming sample size (Theorem 3.1).
2. **Statistical efficiency from compressed history.** The three paths share an explicit influence expansion and sharp first-order risk, attaining the Gaussian local asymptotic minimax constant at the endpoint (Equation (3.11)). An $O(d)$ retained state therefore matches the leading-projector risk of the $O(d^2)$ historical covariance summary, not its full information.
3. **An aligned empirical-flow interpretation.** Riemannian flows restarted from the historical direction on each new prefix track oracle Oja at calibrated times, with $O(N^{-2})$ expected maximal squared projector discrepancy. On one fixed dataset, flow follows an anchored-PCA regularization path to the same order, uniformly over each fixed bounded time interval, and admits a pooled-efficient stopping time (Corollaries 5.1 and 5.2). These are empirical, not population, flows; their evaluation is batch-based.

Supporting results make the calibration operational: clipped historical estimates preserve first-order equivalence, and the historical sample count uniquely minimizes the first-order pseudocount risk. The oracle projector comparison also extends to $d + \log N = o(N)$, whereas the flow results remain fixed-dimensional. Equal gaps delimit the positive theorem; the restricted unequal-gap obstruction is a scope explanation, not a separate contribution. The Oja proof is entirely discrete and requires no SDE.

1.1 Paper organization

Section 2 defines the experiment and three estimator paths. Section 3 gives trajectory equivalence, efficiency, and calibration; Section 4 explains the proof mechanism. Section 5 develops the flow interpretation, and Section 6 reports the numerical audit. Sections 7 and 8 discuss prior work and scope. Proofs appear in Appendices A to F; Appendix G gives detailed comparisons with the closest literature.

2 Warm-started streaming PCA

2.1 Model and historical state

Let $d \geq 2$, $v \in \mathbb{S}^{d-1}$, and

$$\Sigma_v = \sigma^2 I_d + \Delta v v^\top, \qquad \lambda_1 = \sigma^2 + \Delta, \tag{2.1}$$

where $\sigma^2, \Delta > 0$. We observe independent historical samples $Y_1, \dots, Y_\nu$ and new samples $X_1, \dots, X_n$, all distributed as $N(0, \Sigma_v)$. Put $N = \nu + n$, and assume

$$\nu/N \longrightarrow \rho \in (0, 1). \tag{2.2}$$

The clean theorem fixes d, while Corollary 3.3 allows it to grow. Vector norms are Euclidean, matrix norms without a subscript are operator norms, and $\|\cdot\|_F$ denotes the Frobenius norm. For a symmetric matrix, Top_1 selects a unit leading eigenvector by a fixed measurable convention in the event of a tie; all losses depend only on its projector.

Let u be the leading empirical eigenvector of the historical covariance $S_{\mathrm{old}} = \nu^{-1} \sum_{j=1}^{\nu} Y_j Y_j^\top$. Before the historical sample becomes unavailable, retain a unit representative u (equivalently, store $P_u = uu^\top$ in factored form), together with

$$\nu, \qquad \widehat{\lambda}_1 = \lambda_{\max}(S_{\mathrm{old}}), \qquad \widehat{\Delta} = \widehat{\lambda}_1 - \frac{\mathrm{tr}(S_{\mathrm{old}}) - \widehat{\lambda}_1}{d-1}. \tag{2.3}$$

This is the warm-start setup. Warm starting itself imposes no rank, Gaussianity, or equal-gap condition; those are assumptions under which we can characterize exactly what the retained state remembers.

Table 1: State and per-observation work after the historical stage. The comparison concerns recovery of the leading projector, not the entire covariance. Both covariance-based methods additionally require an eigensolve whenever their principal component is requested.

Method	retained numerical state	streaming work
Warm-started rank-one Oja	$O(d)$	$O(d)$
Anchored prefix PCA	$O(d^2)$	$O(d^2)$
Pooled prefix PCA	$O(d^2)$	$O(d^2)$

2.2 Three estimator paths indexed by arriving samples

At prefix k, compare three directions defined from the same underlying historical sample and new observations $X_1, \dots, X_k$, but with different access to historical information. The index k counts observations, not optimization steps on a fixed dataset.

One-pass Oja: update the previous direction. For the oracle presentation, define the effective and raw steps

$$c_k = \frac{1}{\Delta(\nu + k)}, \qquad \eta_k = \frac{c_k}{1 - c_k\lambda_1} = \frac{1}{\Delta(\nu + k) - \lambda_1}. \tag{2.4}$$

For all sufficiently large N, the denominator is positive. Starting at $w_0 = u$, normalized one-pass Oja is

$$w_k = \frac{(I + \eta_k X_k X_k^\top) w_{k-1}}{\left\|(I + \eta_k X_k X_k^\top) w_{k-1}\right\|}, \qquad 1 \le k \le n. \tag{2.5}$$

The plug-in algorithm replaces (Δ, λ_1) by clipped versions of $(\widehat{\Delta}, \widehat{\lambda}_1)$ from Equation (2.3). Each update uses only the previous direction, the arriving observation, and these scalars.

Anchored prefix PCA: refit using compressed history. Retain the historical direction and count, and form the new-prefix covariance:

$$S_k = \frac{1}{k}\sum_{i=1}^{k} X_i X_i^\top, \qquad q_k = \mathrm{Top}_1\left(S_k + \frac{\nu\Delta}{k} u u^\top\right). \tag{2.6}$$

The added projector rewards alignment with u, with strength decreasing as new data accumulate. Each q_k is the exact leading eigendirection of this matrix, not one iterative update. This is an existing regularized estimator used as a comparator, not a newly proposed algorithm.

Pooled prefix PCA: refit with full historical information. The benchmark is ordinary PCA on the old sample plus the first k arrivals:

$$\widetilde{q}_k = \mathrm{Top}_1\left(\frac{\sum_{j=1}^{\nu} Y_j Y_j^\top + \sum_{i=1}^{k} X_i X_i^\top}{\nu + k}\right). \tag{2.7}$$

Keeping S_{old} and ν would suffice to compute this path; retaining the raw old observations is unnecessary. Our compressed historical state does not generally determine S_{old}, so this path is an inaccessible benchmark. No particular eigensolver is part of its definition.

Set $q_0 = \widetilde{q}_0 = w_0 = u$. The three sequences are indexed by the same arrivals; only w_k is generated by the one-pass recursion above.

For a unit vector a, write $P_a = aa^\top$ and measure error by

$$L(P_a, P_v) = \frac{1}{2}\|P_a - P_v\|_F^2 = \sin^2\angle(a, v). \tag{2.8}$$

3 UNIFORM EQUIVALENCE OF THE THREE PREFIX PATHS

Choose an orthogonal complement $V_\perp$ to v and define the local projective chart

$$g_v(a) = \frac{V_\perp^\top a}{v^\top a}. \tag{3.1}$$

Assign the chart the value zero on the probability-zero boundary $v^\top a = 0$; all moment statements for the raw chart remain localized. Writing $X_i = a_i v + V_\perp y_i$, let

$$\xi_i = (v^\top X_i) V_\perp^\top X_i = a_i y_i, \qquad \Gamma = \mathbb{E}\,\|\xi_i\|^2 = (d-1)\lambda_1 \sigma^2. \tag{3.2}$$

The common influence path is

$$Z_k = \frac{\nu}{\nu + k} g_v(u) + \frac{1}{\Delta(\nu + k)} \sum_{i=1}^{k} \xi_i. \tag{3.3}$$

Theorem 3.1 (Uniform same-realization prefix-path equivalence)**.** *Suppose Equations (2.1) and (2.2) hold with fixed d, and let the parameters range over a compact set on which σ^2 and Δ are bounded above and away from zero. There are simultaneous events $\mathcal{G}_N$ with $\mathbb{P}(\mathcal{G}_N^c) = O(N^{-2})$ such that*

$$\mathbb{E}\left[\max_{k \le n} \|g_v(w_k) - Z_k\|^2 \mathbf{1}_{\mathcal{G}_N}\right] = O(N^{-2}), \tag{3.4}$$

$$\mathbb{E}\left[\max_{1 \le k \le n} \|g_v(q_k) - Z_k\|^2 \mathbf{1}_{\mathcal{G}_N}\right] = o(N^{-1}). \tag{3.5}$$

Consequently, without chart qualifications,

$$\mathbb{E} \max_{k \le n} \|P_{w_k} - P_{q_k}\|_F^2 = o(N^{-1}), \tag{3.6}$$

$$\mathbb{E} \max_{k \le n} \|P_{w_k} - P_{\widetilde{q}_k}\|_F^2 = o(N^{-1}), \tag{3.7}$$

$$\mathbb{E} \max_{k \le n} \|P_{q_k} - P_{\widetilde{q}_k}\|_F^2 = o(N^{-1}). \tag{3.8}$$

Moreover, uniformly over prefixes and for each of the three estimators,

$$\sup_{k \le n} \left| \mathbb{E} L(P_{w_k}, P_v) - \frac{(d-1)\lambda_1 \sigma^2}{\Delta^2(\nu + k)} \right| = o(N^{-1}), \tag{3.9}$$

where w_k may be replaced by q_k or $\widetilde{q}_k$. At the final prefix,

$$\sqrt{N}(P_{w_n} - P_v) \Rightarrow v G^\top + G v^\top, \quad G \sim N_{v^\perp}\left(0, \frac{\lambda_1 \sigma^2}{\Delta^2} P_{v^\perp}\right). \tag{3.10}$$

The distinction between Equation (3.4) and Equations (3.6) and (3.7) is essential. Gaussian projective coordinates have infinite unconditional second moments because a random denominator can approach zero. The tangent expansion is localized; the projector conclusions are unconditional because Equation (2.8) is bounded.

3.1 STATISTICAL PAYOFF

The constant in Equation (3.9) agrees with classical Gaussian PCA asymptotics (Anderson, 1963) and is the local asymptotic minimax constant for estimating P_v in the Gaussian experiment (van der Vaart, 1998). In particular,

$$N \mathbb{E} L(P_{w_n}, P_v) \longrightarrow \frac{(d-1)\lambda_1 \sigma^2}{\Delta^2}. \tag{3.11}$$

Using only the new sample gives the same numerator divided by n, while freezing the historical component gives it divided by ν. Thus, when $\nu = n$, either naive choice has twice the first-order risk of calibrated warm-started Oja. The gain is not faster optimization: it is statistically correct fusion of historical and new information.

The retained projector is not exactly sufficient for the full covariance. The precise statement is local and parameter-specific: after a bounded-chart truncation, the reduced central sequence measurable from $(P_u, \nu, X_{1:n})$ differs by $o_{L^2}(1)$ from the full-data central sequence for perturbations of v. Nuisance eigenvalues require the calibration scalars in Equation (2.3).

3.2 PLUG-IN CALIBRATION AND THE PSEUDOCOUNT ORACLE

Corollary 3.2 (Implementable warm-started Oja). *Clip $\widehat{\Delta}$ and $\widehat{\lambda}_1$ from Equation (2.3) to a fixed compact parameter envelope containing the true compact parameter set in its interior, with $\widehat{\Delta} \geq d_0 > 0$ and $\widehat{\lambda}_1 \leq L_0 < \infty$, and use*

$$\widehat{\eta}_i = \frac{1}{\widehat{\Delta}(\nu + i) - \widehat{\lambda}_1}, \tag{3.12}$$

with a positive finite-N denominator clip. This extra clip is inactive for every sufficiently large N. Also define

$$\widehat{q}_k = \mathrm{Top}_1\left(S_k + \frac{\nu\widehat{\Delta}}{k} uu^\top\right).$$

Set $\widehat{q}_0 = u$. If $\widehat{w}_k$ is the resulting Oja path, then the unconditional projector comparisons, uniform risk expansion, and endpoint limit in Theorem 3.1 remain valid with (w_k, q_k) replaced by $(\widehat{w}_k, \widehat{q}_k)$.

The count ν is not an arbitrary optimization offset. Replacing it by a candidate pseudocount a, use

$$c_i(a) = \frac{1}{\Delta(a + i)}, \qquad \eta_i(a) = \frac{c_i(a)}{1 - \lambda_1 c_i(a)},$$

and denote the resulting endpoint by $w_n(a)$. For fixed $0 < c_- < 1 < c_+ < \infty$, the nonlinear analysis is uniform over $a/\nu \in [c_-, c_+]$ and gives

$$\sup_{a/\nu \in [c_-, c_+]} \left|\mathbb{E}L(P_{w_n(a)}, P_v) - \mathcal{R}_n(a)\right| = o(N^{-1}), \tag{3.13}$$

where the first-order endpoint risk is

$$\mathcal{R}_n(a) = \frac{\Gamma}{\Delta^2}\frac{a^2/\nu + n}{(a+n)^2} = \frac{\Gamma}{\Delta^2(\nu+n)} + \frac{\Gamma}{\Delta^2}\frac{n(a-\nu)^2}{\nu(a+n)^2(\nu+n)}. \tag{3.14}$$

The same expansion holds for $\mathrm{Top}_1(S_n + a\Delta uu^\top/n)$. By Equation (3.13), every sequence of exact risk minimizers over this compact ratio class obeys $a/\nu \to 1$. This is first-order uniqueness, not exact finite-sample optimality.

Corollary 3.3 (Slowly growing dimension). *If $d = d_N$, $d + \log N = o(N)$, balance holds, and the spectral parameters remain in a compact positive set, then the oracle-calibrated paths satisfy*

$$\mathbb{E}\max_{k \leq n} \|P_{w_k} - P_{\widetilde{q}_k}\|_F^2 \leq C\frac{d(d + \log N)}{N^2} = o(d/N), \tag{3.15}$$

with analogous comparisons to q_k, and the risk expansion in Equation (3.9) holds with remainder $o(d/N)$. This does not assert a growing-dimensional CLT or LAN theorem.

4 WHY THE HARMONIC SCHEDULE WORKS

Normalization in Equation (2.5) cancels exactly in projective coordinates. Let

$$\Xi_i = X_i X_i^\top - \Sigma_v, \quad \alpha_i = a_i^2 - \lambda_1, \quad \beta_i = a_i y_i, \quad C_i = y_i y_i^\top - \sigma^2 I,$$

and put $m_i = \nu + i$, $r_i = 1 - m_i^{-1}$, and $h_i(g) = \alpha_i + \beta_i^\top g$. With $g_i = g_v(w_i)$, the exact recursion is

$$g_i = \frac{r_i g_{i-1} + c_i\{\beta_i + C_i g_{i-1}\}}{1 + c_i h_i(g_{i-1})}. \tag{4.1}$$

Its linear part is

$$Z_i = r_i Z_{i-1} + c_i \beta_i, \qquad Z_0 = g_v(u), \tag{4.2}$$

which telescopes exactly to Equation (3.3). The factor $\nu/(\nu + k)$ gives the old estimator its sample-count weight, while every new score receives weight $1/\{\Delta(\nu + k)\}$.

While the iterates remain in the chart, the nonlinear error $D_k = g_k - Z_k$ satisfies the whole-prefix identity

$$D_k = \frac{1}{m_k}\left\{\frac{1}{\Delta}\sum_{i=1}^{k} U_i(g_{i-1}) + \sum_{i=1}^{k} m_i \mathfrak{q}_i(g_{i-1})\right\}, \tag{4.3}$$

where

$$U_i(g) = C_i g - r_i g h_i(g),$$
$$\mathfrak{q}_i(g) = -\frac{c_i^2 h_i(g)\{U_i(g) + \beta_i\}}{1 + c_i h_i(g)}.$$

The first sum in Equation (4.3) is a martingale after predictable localization; bounding it as a sum of absolute errors would destroy the N^{-2} scale. The second sum is already quadratic in the steps. A bootstrap gives $\mathbb{E}\left[\max_{k\le n} \|D_k\|^2 \mathbf{1}_{\mathcal{G}_N}\right] = O(N^{-2})$ on a common chart event, while the exit probability is $O(N^{-2})$.

The anchored estimator and pooled PCA share the same path because their eigenvector perturbations have the same first-order off-diagonal block. If $B_0 = \Sigma_v + \mu vv^\top$, $B = B_0 + H$, and $q = \mathrm{Top}_1(B)$, assume $H = H^\top$, $\Delta + \mu > 0$, and $\|H\|/(\Delta + \mu) \le 1/8$. A local block eigen-equation gives

$$\left\|g_v(q) - (\Delta+\mu)^{-1} V_\perp^\top H v\right\| \le 8\frac{\|H\|}{\Delta+\mu}\frac{\left\|V_\perp^\top H v\right\|}{\Delta+\mu}. \tag{4.4}$$

Substituting $\mu_k = \nu\Delta/k$ yields the first-order path in Equation (3.3), up to the perturbation and cubic pilot remainders; the pooled covariance does the same after linearizing u from the old sample. The appendix supplies the localization, moment, perturbation, and risk-transfer details.

5 Interpreting the Streaming Trajectory through Empirical Flow

The flow interprets the discrete theorem, rather than proposing another streaming method. The statistical experiment is unchanged: use the same historical pilot u and incoming observations as Theorem 3.1. Distinguish the data prefix k from optimization time t. For each fixed $k \ge 1$, let $F_k(t)$ ascend $\frac{1}{2}w^\top S_k w$ on the sphere:

$$\dot{F}_k(t) = (I - F_k(t)F_k(t)^\top)S_k F_k(t), \qquad F_k(0) = u, \qquad F_k(t) = \frac{e^{tS_k}u}{\|e^{tS_k}u\|}. \tag{5.1}$$

Each flow is deterministic conditional on the empirical covariance and pilot, but random under sampling. It is neither population flow nor an SDE.

5.1 Alignment at Each Data Prefix

Corollary 5.1 (Uniform prefix-flow alignment)**.** *Under the fixed-dimensional assumptions of Theorem 3.1, set*

$$t_k = \Delta^{-1}\log(1 + k/\nu), \qquad f_k = F_k(t_k), \qquad f_0 = u.$$

Then, for the oracle Oja path w_k,

$$\mathbb{E}\max_{k\le n}\|P_{f_k} - P_{w_k}\|_F^2 = O(N^{-2}). \tag{5.2}$$

Consequently, f_k inherits the pooled/anchored comparisons, sharp risk, and endpoint limit in Theorem 3.1. Replacing Δ in t_k by the clipped historical estimate preserves these conclusions, including Equation (5.2) against oracle Oja. Comparison to plug-in Oja is asserted only at the first-order $o(N^{-1})$ scale.

The interpretation is a common clock for historical memory. In the first-order recursion Equation (4.2), Oja retains weight $\prod_{i=1}^{k}(1 - \Delta c_i) = \nu/(\nu + k)$ on the historical error; flow retains $e^{-\Delta t}$. The time t_k makes these weights identical. Its new-score coefficient is also identical: $(1 - e^{-\Delta t_k})/(\Delta k) = 1/\{\Delta(\nu + k)\}$. Thus both reproduce the whole influence path Z_k, not

merely its variance. Corollary 5.1 controls the nonlinear discrepancy uniformly across prefixes, turning this interpretation into a quantitative statement about the actual Oja trajectory.

The flow family is a statistical representation, not a claimed one-pass implementation: F_k restarts from u for each S_k, while Oja continues from its previous iterate. Evaluating the family requires new-data covariance products; it is not one streaming ODE and does not inherit Oja's linear streaming-storage guarantee.

5.2 Optimization time on a fixed dataset

Now freeze the final prefix and write $f(t) = F_n(t)$. Varying t gives a separate early-stopping profile, not additional data arrivals.

Corollary 5.2 (Flow and anchored-PCA profile)**.** *Under the fixed-dimensional assumptions of Theorem 3.1, fix $T < \infty$, independently of N. For $t > 0$, define*

$$\mu(t) = \frac{\Delta}{e^{\Delta t} - 1}, \qquad q^{\mathrm{F}}(t) = \mathrm{Top}_1(S_n + \mu(t) P_u), \tag{5.3}$$

and put $q^{\mathrm{F}}(0) = u$. Then

$$\mathbb{E} \sup_{0 \le t \le T} \left\| P_{f(t)} - P_{q^{\mathrm{F}}(t)} \right\|_F^2 = O(N^{-2}). \tag{5.4}$$

For both $h = f$ and $h = q^{\mathrm{F}}$,

$$\sup_{0 \le t \le T} \left| \mathbb{E} L(P_{h(t)}, P_v) - \frac{\Gamma}{\Delta^2} \left\{ \frac{e^{-2\Delta t}}{\nu} + \frac{(1 - e^{-\Delta t})^2}{n} \right\} \right| = o(N^{-1}). \tag{5.5}$$

If T contains

$$t_* = \frac{1}{\Delta} \log(1 + n/\nu), \qquad \mu(t_*) = \frac{\nu \Delta}{n}, \tag{5.6}$$

eventually, t_* uniquely minimizes the displayed first-order risk and attains $\Gamma/(\Delta^2 N)$. At t_*, both estimators share the pooled endpoint influence and limit law. This is not exact finite-sample risk optimality.

The exponential calibration in Equation (5.3) matches the old weight $e^{-\Delta t} = \mu/(\Delta + \mu)$; it is only asymptotic to $1/t$ near zero. This is an equal-gap statistical comparison, not a general-spectrum gradient-flow/ridge equivalence.

Appendix E proves both comparisons directly, without interchanging statistical and discretization limits. An exact sample-by-sample flow interpretation of Oja's endpoints is given separately in Appendix E.4; that identity is not an additional statistical result.

6 Numerical audit

We test the sharp constants rather than use simulation as evidence for the proof. Samples follow $N(0, I_8 + 2vv^\top)$, old and new sample sizes are equal, and results are averaged over 1000 independent trials. The sharp constant is $A = (d-1)(\sigma^2+\Delta)\sigma^2/\Delta^2 = 5.25$, so every entry in Table 2 is $N\mathbb{E}L/A$. The plug-in methods use Equation (2.3) with a fixed truth-independent clip. The half- and double-count columns replace ν by $\nu/2$ and 2ν.

Across a separate 2000-trial audit, $N^2 \mathbb{E} \|P_{w_n} - P_{\widetilde{q}_n}\|_F^2$ remains approximately constant over $\nu = 100, \ldots, 1600$, while the oracle Oja, projector-regularized, pooled, and stopped full-gradient normalized risks approach one. These endpoint tests are consistent with the risk constant and lower-order coupling scale; they do not check the maximum over prefixes or prove an N^{-2} rate. The plug-in results and the half/double-count columns also match the corresponding first-order predictions. Code and fixed seeds accompany the submission.

Table 2: Normalized projector risk; the target is one. Half or double pseudocount has theoretical limit $10/9$. Monte Carlo standard errors are between 0.0172 and 0.0234; estimator-specific values accompany the saved CSV.

$\nu = n$	oracle Oja	plug-in Oja	plug-in anchor	pooled	half count	double count
100	1.032	1.017	1.003	0.997	1.099	1.176
200	1.009	1.004	1.002	0.999	1.128	1.106
400	0.984	0.982	0.980	0.978	1.085	1.102
800	1.005	1.005	1.003	1.002	1.113	1.115

7 Related work

Oja rates, linearization, and inference. Finite-sample Oja theory includes near-optimal rates (Jain et al., 2016; Li et al., 2018), harmonic warm-start analysis for top-r PCA (Huang et al., 2021; Liang, 2023), and sparse support selection with Oja refitting (Kumar & Sarkar, 2024). Liang (2023) already decomposes the warm graph-chart trajectory into contraction, martingale, and drift terms; neither this technique nor discrete analysis is new here. Lunde et al. (2021); Kumar et al. (2025); Xu & Li (2026) develop Gaussian approximations and bootstrap or entrywise inference, including general-rank, sub-Gaussian results. Their principal inference targets concern endpoint distributions, not recovery of the same realized pooled-prefix path from a retained empirical pilot. Our sharper comparison target comes with substantially narrower distributional and spectral assumptions.

Uniform bounds and functional limits. Pham et al. (2025) give all-time population-angle bounds for warm-started bounded-data Oja with harmonic steps. This is stronger in time horizon than our balanced finite-horizon statement, but controls error to the population direction, not a below-risk-scale difference between empirical estimators. Functional limits for unaveraged harmonic stochastic-gradient trajectories also predate our work (Arnaudon et al., 2012; Flamand & Brunel, 2026); Oja diffusion approximations are developed by Li et al. (2017); Liu & Liu (2023). A weak process limit is distinct from coupling specified estimators on the original observations in expected maximal squared distance.

Recursive efficiency and same-noise expansions. Calibrated efficient recursion and averaging are classical (Amari, 1998; Polyak & Juditsky, 1992; Zhou & Said, 2019). Sharia (2010) proves recursive asymptotic linearity and discusses same-data equivalence to maximum likelihood; Tripuraneni et al. (2018) obtain efficient averaged Riemannian SGD, including general-gap PCA. Pilot-based efficient estimator processes appear in Kutoyants & Motrunich (2016). Li et al. (2026a) give a uniform same-noise representation for an averaged SA path, under step assumptions excluding the harmonic schedule. These are substantive precedents, not only endpoint rate bounds. Here the nonnegligible historical fluctuation, unaveraged normalized update, and unconditional expected prefix maximum must be handled together.

Historical summaries, transfer, and pooled comparators. Eigenmodel merging without original data is longstanding (Hall et al., 1999). History PCA repeatedly combines a count-weighted previous projector with a new block and includes a harmonic Oja specialization (Yang et al., 2018). TL-PCA supplies the projector-regularized eigensystem used here (Hendy & Dar, 2024); multi-study transfer uses shared subspace aggregation (Li et al., 2024). Distributed PCA can match centralized rates (Fan et al., 2019), and Chen et al. (2022); Li et al. (2026b) already compare iterates with the *same pooled empirical covariance*. Those algorithms revisit retained local covariance operators as communication rounds progress. Our index instead counts arrivals, changes the empirical objective, and permits Oja to discard the historical covariance. We claim neither a new regularizer nor the first empirical pooled comparison.

Optimization and regularization paths. Balzano (2022) proves exact Oja–GROUSE iterate equivalence after state-dependent step conversion. For least squares, Ali et al. (2019) compare gradient-flow and ridge risk along the full path and identify an exact generalized-penalty representation; Ali et al. (2020) also bound same-data stochastic-flow/ridge estimator distance. AgFlow applies a regression-path construction to approximate PCA scores (Jiang et al., 2021). Our flow result is a nonlinear

empirical Rayleigh-flow comparison with an anchored eigendirection: its calibrated prefix evaluations interpret the discrete Oja path, while its fixed-data time profile gives a separate early-stopping comparison. Appendix G records theorem locations, version qualifications, and the precise boundaries of these precedents.

8 Limitations and extensions

The result is intentionally sharp and correspondingly structured. It assumes a stationary rank-one isotropic spike, balanced old and new sample sizes, and Gaussian observations. The growing-dimensional risk comparison in Corollary 3.3 is not a high-dimensional CLT. The one-pass prefix theorem is also not a fixed-objective early-stopping theorem: the empirical objective changes as observations arrive. The separate flow result has fixed dimension and a bounded time interval; it is not asserted in the growing-dimensional regime. Its batch covariance operations do not provide Oja's one-pass storage guarantee.

Equal tangent gaps are built into the rank-one isotropic spike. The linearized obstruction in Appendix F explains why the fixed scalar harmonic family cannot generally retain the pooled weights across unequal gaps under its stated pilot assumptions. This is a caveat, not a general impossibility theorem for scalar schedules or for Oja.

Gaussianity appears to be the least expensive assumption to relax. The exact recursion and telescoping are distribution-free, suggesting a route under sub-Gaussian tails. Uniform localization and identification of the resulting influence covariance remain to be proved; the Gaussian LAN constant does not transfer automatically.

An equal-gap top-r extension is structurally plausible because every Grassmann tangent mode still contracts at the same rate. It requires a matrix graph chart, uniform inverse-block localization, and matrix-martingale control; we do not claim it here. General spectra are qualitatively different by Proposition F.1. Strongly unbalanced horizons introduce explicit $\log(1+n/\nu)$ factors, and covariance drift changes the relevant oracle away from pooled PCA.

Historical samples being unavailable is a storage or access model. It is not machine unlearning, differential privacy, or a claim that the retained component reveals no information about individuals.

9 Conclusion

A historical principal component is more than a convenient initialization. When paired with its provenance and a sample-count-calibrated harmonic schedule, it has the same leading influence that pooled PCA assigns to the unavailable historical sample. Warm-started normalized Oja then fuses this influence with each new observation in one pass, using linear state and work; its expected maximal squared-projector discrepancy from pooled PCA is $o(N^{-1})$. Restarted empirical Riemannian flows align with this same prefix trajectory; on a fixed dataset, flow has an anchored profile and a pooled-efficient stopping time. These statistical comparisons are specific to the stated isotropic-spike and balanced-history experiment; general spectra remain outside the positive theorem.

Reproducibility statement. Appendices A to F record the proofs and localization qualifications. The supplement includes the script, the full-precision CSV for Table 2, the historical rounded record of the separate endpoint audit, and fixed-seed reproduction instructions.

AI use statement. OpenAI's GPT-5.6 Sol and GPT-6 Astra were used to assist with literature discovery, proof development and auditing, experiment implementation, and manuscript drafting. The main results were obtained through discussions between the author and generative AI agents. The author is responsible for reviewing and approving the mathematical claims, citations, numerical results, and final disclosure before release.

References

Alnur Ali, J. Zico Kolter, and Ryan J. Tibshirani. A continuous-time view of early stopping for least squares regression. In *Proceedings of the Twenty-Second International Conference on*

Artificial Intelligence and Statistics, volume 89 of *Proceedings of Machine Learning Research*, pp. 1370–1378, 2019. URL https://proceedings.mlr.press/v89/ali19a.html.

Alnur Ali, Edgar Dobriban, and Ryan J. Tibshirani. The implicit regularization of stochastic gradient flow for least squares. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 233–244, 2020. URL https://proceedings.mlr.press/v119/ali20a.html.

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746.

T. W. Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 1963. doi: 10.1214/aoms/1177704248.

Marc Arnaudon, Clément Dombry, Anthony Phan, and Le Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455, 2012. doi: 10.1016/j.spa.2011.12.011.

Laura Balzano. On the equivalence of Oja's algorithm and GROUSE. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 7014–7030, 2022. URL https://proceedings.mlr.press/v151/balzano22a.html.

Xi Chen, Jason D. Lee, He Li, and Yun Yang. Distributed estimation for principal component analysis: An enlarged eigenspace analysis. *Journal of the American Statistical Association*, 117 (540):1775–1786, 2022. doi: 10.1080/01621459.2021.1886937. URL https://arxiv.org/abs/2004.02336v3.

Jianqing Fan, Dong Wang, Kaizheng Wang, and Ziwei Zhu. Distributed estimation of principal eigenspaces. *The Annals of Statistics*, 47(6):3009–3031, 2019. doi: 10.1214/18-AOS1713.

Kessang Flamand and Victor-Emmanuel Brunel. Functional central limit theorem for stochastic gradient descent. *arXiv preprint arXiv:2602.15538*, 2026. doi: 10.48550/arXiv.2602.15538. URL https://arxiv.org/abs/2602.15538v1.

Peter Hall, David Marshall, and Ralph Martin. Adding and subtracting eigenspaces. In *Proceedings of the British Machine Vision Conference*, pp. 453–462, 1999. URL https://users.cs.cf.ac.uk/Dave.Marshall/Papers/BMVC99Addingeigen.pdf.

Sharon Hendy and Yehuda Dar. TL-PCA: Transfer learning of principal component analysis. *arXiv preprint arXiv:2410.10805*, 2024. URL https://arxiv.org/abs/2410.10805.

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. URL https://proceedings.mlr.press/v134/huang21a.html.

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 *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 1147–1164, 2016. URL https://proceedings.mlr.press/v49/jain16.html.

Haiyan Jiang, Haoyi Xiong, Dongrui Wu, Ji Liu, and Dejing Dou. AgFlow: Fast model selection of penalized PCA via implicit regularization effects of gradient flow. *Machine Learning*, 110(8): 2131–2150, 2021. doi: 10.1007/s10994-021-06025-3.

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

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, 2025. URL https://proceedings.mlr.press/v286/kumar25a.html.

Yury A. Kutoyants and Anastasia Motrunich. On multi-step MLE-process for Markov sequences. *Metrika*, 79(6):705–724, 2016. doi: 10.1007/s00184-015-0574-4. URL https://arxiv.org/abs/1601.08174.

Chris Junchi Li, Mengdi Wang, Han Liu, and Tong Zhang. Diffusion approximations for online principal component estimation and global convergence. In *Advances in Neural Information Processing Systems*, volume 30, pp. 645–655, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/13f3cf8c531952d72e5847c4183e6910-Abstract.html.

Chris Junchi 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.

Xiang Li, Jiadong Liang, and Zhihua Zhang. Online statistical inference for nonlinear stochastic approximation with Markovian data. *arXiv preprint arXiv:2302.07690*, 2026a. URL https://arxiv.org/abs/2302.07690v3. Version 3.

Zeyu Li, Kangxiang Qin, Yong He, Wang Zhou, and Xinsheng Zhang. Knowledge transfer across multiple principal component analysis studies. *arXiv preprint arXiv:2403.07431*, 2024. doi: 10.48550/arXiv.2403.07431. URL https://arxiv.org/abs/2403.07431.

Zeyu Li, Xinsheng Zhang, and Wang Zhou. Few-round distributed principal component analysis: Closing the statistical efficiency gap by consensus. *Journal of Computational and Graphical Statistics*, pp. 1–30, 2026b. doi: 10.1080/10618600.2026.2674240. URL https://arxiv.org/abs/2503.03123v3. Published online June 12, 2026. Theorem comparisons use arXiv:2503.03123v3.

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

Jian-Guo Liu and Zibu Liu. Diffusion approximations of Oja's online principal component analysis. *arXiv preprint arXiv:2301.01339*, 2023. URL https://arxiv.org/abs/2301.01339.

Robert Lunde, Purnamrita Sarkar, and Rachel Ward. Bootstrapping the error of Oja's algorithm. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6240–6252, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/3152e3b1e52e2cb123363787d5f76c95-Abstract.html.

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

Tuan Pham, Alessandro Rinaldo, and Purnamrita Sarkar. Time-uniform concentration bounds for iterative algorithms. *arXiv preprint arXiv:2511.18273*, 2025. URL https://arxiv.org/abs/2511.18273.

Boris T. Polyak and Anatoli B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992. doi: 10.1137/0330046. URL https://epubs.siam.org/doi/10.1137/0330046.

Teo Sharia. Recursive parameter estimation: Asymptotic expansion. *Annals of the Institute of Statistical Mathematics*, 62(2):343–362, 2010. doi: 10.1007/s10463-008-0179-z. URL https://arxiv.org/abs/0705.1783.

Nilesh Tripuraneni, Nicolas Flammarion, Francis Bach, and Michael I. Jordan. Averaging stochastic gradient descent on Riemannian manifolds. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 650–687, 2018. URL https://proceedings.mlr.press/v75/tripuraneni18a.html.

A. W. van der Vaart. *Asymptotic Statistics*. Number 3 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. doi: 10.1017/CBO9780511802256.

Haoshu Xu and Hongzhe Li. Inference and uncertainty quantification for streaming r-PCA. *arXiv preprint arXiv:2608.18374*, 2026. URL https://arxiv.org/abs/2608.18374.

Puyudi Yang, Cho-Jui Hsieh, and Jane-Ling Wang. History PCA: A new algorithm for streaming PCA. *arXiv preprint arXiv:1802.05447*, 2018. doi: 10.48550/arXiv.1802.05447. URL https://arxiv.org/abs/1802.05447.

Jialun Zhou and Salem Said. Fast, asymptotically efficient, recursive estimation in a Riemannian manifold. *Entropy*, 21(10):1021, 2019. doi: 10.3390/e21101021.

A PROOF DETAILS FOR THE MAIN THEOREM

This appendix gives the proof at the level needed to expose every stochastic and localization step. Constants are uniform over the compact parameter sets in Theorem 3.1.

A.1 A COUPLED HISTORICAL PILOT

Let

$$E_0 = \{\|S_{\text{old}} - \Sigma_v\| \le \Delta/8\}.$$

On E_0, orient the unique empirical top eigenvector toward v and put $\bar{g}_0 = g_v(u)$; on E_0^c, set $\bar{g}_0 = 0$ without evaluating the chart.

Lemma A.1 (Pilot expansion). *For every fixed $s \in \{1, 2, 3\}$,*

$$\mathbb{P}(E_0^c) \le N^{-5}, \qquad \mathbb{E}\,\|\bar{g}_0\|^{2s} \le C_s \nu^{-s}, \tag{A.1}$$

$$\bar{g}_0 = \frac{1}{\nu\Delta}\sum_{j=1}^{\nu}(v^\top Y_j)V_\perp^\top Y_j + r_{0,N}, \qquad \mathbb{E}\,\|r_{0,N}\|^2 \le C\nu^{-2}. \tag{A.2}$$

Proof. Fixed-dimensional Wishart concentration proves the probability bound. In the $(v, V_\perp)$ blocks, the empirical eigen-equation and the event E_0 give

$$\|\bar{g}_0\| \le C\Delta^{-1}\left\|V_\perp^\top(S_{\text{old}} - \Sigma_v)v\right\|$$

and a remainder bounded by

$$C\Delta^{-2}\|S_{\text{old}} - \Sigma_v\|\left\|V_\perp^\top(S_{\text{old}} - \Sigma_v)v\right\|.$$

Gaussian covariance moments through order six yield Equations (A.1) and (A.2). On E_0^c, the coupled pilot is zero; the corresponding remainder is controlled by the exponentially small Wishart tail and Gaussian score moments. □

Define the coupled influence path

$$\bar{Z}_k = \frac{\nu}{\nu + k}\bar{g}_0 + \frac{1}{\Delta(\nu + k)}\sum_{i=1}^{k}\beta_i. \tag{A.3}$$

It equals Z_k on E_0.

A.2 PREDICTABLE LOCALIZATION OF THE OJA RECURSION

Fix a constant chart radius $r > 0$, let $t_N = 6\log N$, and choose

$$B_N = \lambda_1\{d + 2\sqrt{dt_N} + 2t_N + 1\} = O(\log N)$$

so that, for $H_i = \{\|\Xi_i\| \le B_N\}$,

$$\mathbb{P}\left(\bigcup_{i\le n} H_i^c\right) \le N^{-5}. \tag{A.4}$$

On $H_i \cap \{\|g\| \le r\}$, large N guarantees $|c_i h_i(g)| \le 1/2$.

Let $\mathcal{F}_i = \sigma(Y_{1:\nu}, X_{1:i})$. Define the piecewise rational remainder $\mathfrak{q}_i^H(g) = \mathfrak{q}_i(g)$ on the safe set $H_i \cap \{\|g\| \le r\}$ and zero elsewhere. Starting with $\widetilde{D}_0 = 0$ and $\widetilde{g}_i = \bar{Z}_i + \widetilde{D}_i$, set

$$J_i = \mathbf{1}\left\{\|\bar{g}_0\| \le r/4, \max_{0\le j<i}\|\widetilde{g}_j\| \le r\right\}, \tag{A.5}$$

including J_{n+1}, and define

$$\widetilde{D}_i = r_i\widetilde{D}_{i-1} + J_i\{c_i U_i(\widetilde{g}_{i-1}) + \mathfrak{q}_i^H(\widetilde{g}_{i-1})\}. \tag{A.6}$$

The indicator J_i is $\mathcal{F}_{i-1}$-measurable. Crucially, we do not truncate U_i, so $J_i U_i$ remains a martingale difference.

For any $\mathcal{F}_{i-1}$-measurable G with $\|G\| \le r$, direct Gaussian calculation gives

$$\mathbb{E}[U_i(G) \mid \mathcal{F}_{i-1}] = 0, \tag{A.7}$$

$$\mathbb{E}[\|U_i(G)\|^2 \mid \mathcal{F}_{i-1}] \le K_2 \|G\|^2, \tag{A.8}$$

$$\left\|J_i \mathfrak{q}_i^H(\widetilde{g}_{i-1})\right\|_{L^2} \le 2c_i^2 Q_2, \tag{A.9}$$

where $K_2, Q_2 < \infty$ depend only on the compact parameter set and r. Unrolling Equation (A.6) recovers the harmonic identity

$$\widetilde{D}_k = \frac{1}{m_k}\left\{\frac{1}{\Delta}\sum_{i=1}^{k} J_i U_i + \sum_{i=1}^{k} m_i J_i \mathfrak{q}_i^H\right\}. \tag{A.10}$$

Put

$$A_N = \max_{0\le j\le n} m_j \mathbb{E}[J_{j+1}\|\widetilde{g}_j\|^2].$$

Martingale orthogonality, Minkowski's inequality, and balance yield

$$\mathbb{E}\left\|\widetilde{D}_k\right\|^2 \le \frac{C(A_N+1)}{N^2}. \tag{A.11}$$

Independence and Lemma A.1 give $\sup_k \mathbb{E}\left\|\bar{Z}_k\right\|^2 \le C/N$, hence

$$A_N \le C + C(A_N+1)/N.$$

Thus $A_N = O(1)$. Doob's inequality applied to $\sum_{i\le k} J_i U_i$ and Minkowski applied to the rational remainder give

$$\mathbb{E}\max_{k\le n}\left\|\widetilde{D}_k\right\|^2 = O(N^{-2}). \tag{A.12}$$

Fourth-moment martingale bounds similarly give

$$\mathbb{E}\max_{k\le n}\left\|\bar{Z}_k\right\|^4 = O(N^{-2}). \tag{A.13}$$

Define

$$\mathcal{G}_N^{\mathrm{Oja}} = E_0 \cap \{\|\bar{g}_0\| \le r/4\} \cap \bigcap_{i\le n} H_i \cap \left\{\max_{0\le j\le n}\|\widetilde{g}_j\| \le r\right\}. \tag{A.14}$$

Markov's inequality, Equations (A.12) and (A.13), the pilot fourth moment, and Equation (A.4) give $\mathbb{P}((\mathcal{G}_N^{\mathrm{Oja}})^c) = O(N^{-2})$. Before the first exit the stopped and actual recursions coincide. Hence

$$\mathbb{E}\left[\max_{k\le n}\left\|g_v(w_k) - \bar{Z}_k\right\|^2; \mathcal{G}_N^{\mathrm{Oja}}\right] = O(N^{-2}).$$

A.3 Projector-regularized and pooled eigendirection expansions

We use the following deterministic block perturbation lemma.

Lemma A.2 (Local eigendirection expansion)**.** *Let $B_0 = \Sigma_v + \mu vv^\top$, whose top eigengap is $D = \Delta + \mu > 0$, and let $B = B_0 + H$ with $H = H^\top$. If $\|H\|/D \le 1/8$, orient $q = \mathrm{Top}_1(B)$ toward v, put $b = V_\perp^\top H v$, and let $g = g_v(q)$. Then*

$$\left\|g - D^{-1}b\right\| \le 8\frac{\|H\|}{D}\frac{\|b\|}{D}. \tag{A.15}$$

Proof. The block eigen-equation is

$$Dg = b + (H_{\perp\perp} - H_{11}I)g - g(b^\top g).$$

Davis–Kahan puts g in a fixed chart. Absorb the linear terms to get $\|g\| \le 2\|b\|/D$, and substitute this bound once more. □

For Equation (2.6), put

$$T^Y = \sum_{j=1}^{\nu}(Y_jY_j^\top - \Sigma_v),$$
$$T_k^X = \sum_{i=1}^{k}(X_iX_i^\top - \Sigma_v), \qquad T_0^X = 0,$$
$$H_k = k^{-1}T_k^X + (\nu\Delta/k)(uu^\top - vv^\top),$$
$$\mathsf{D}_k = \Delta + \nu\Delta/k = \Delta(\nu+k)/k.$$

Uniform Wishart bounds normalized by D_k, rather than by the possibly unconcentrated short-prefix covariance, give an event $\mathcal{G}_N^{\mathrm{cov}}$ with $\mathbb{P}((\mathcal{G}_N^{\mathrm{cov}})^c) \le N^{-4}$ on which the following bounds hold simultaneously for $1 \le k \le n$ and $0 \le k \le n$, respectively:

$$\frac{\|H_k\|}{\mathsf{D}_k} \le C\left\{\sqrt{\frac{d+\log N}{\nu+k}} + \frac{d+\log N}{\nu+k}\right\}, \tag{A.16}$$
$$\frac{\|T^Y + T_k^X\|}{\Delta(\nu+k)} \le C\left\{\sqrt{\frac{d+\log N}{\nu+k}} + \frac{d+\log N}{\nu+k}\right\}. \tag{A.17}$$

Combining Lemma A.2 and Equation (A.2) with a vector-martingale maximal bound gives

$$\mathbb{E}\left[\max_{1\le k\le n}\left\|g_v(q_k) - \bar{Z}_k\right\|^2; \mathcal{G}_N^{\mathrm{Oja}} \cap \mathcal{G}_N^{\mathrm{cov}}\right] = o(N^{-1}). \tag{A.18}$$

Define the full-sample influence path

$$I_k = \frac{1}{\Delta(\nu+k)}\left\{\sum_{j=1}^{\nu}(v^\top Y_j)V_\perp^\top Y_j + \sum_{i=1}^{k}\xi_i\right\}. \tag{A.19}$$

Set

$$\mathcal{G}_N = \mathcal{G}_N^{\mathrm{Oja}} \cap \mathcal{G}_N^{\mathrm{cov}} \cap \left\{\max_{k\le n}\|I_k\| \le r/4\right\}. \tag{A.20}$$

Then $\bar{Z}_k = I_k + \nu(\nu+k)^{-1}r_{0,N}$. Applying Lemma A.2 with $\mu = 0$ to every pooled prefix and retaining the vector-martingale maximum gives

$$\mathbb{E}\left[\max_{k\le n}\|g_v(\widetilde{q}_k) - I_k\|^2; \mathcal{G}_N\right] = o(N^{-1}). \tag{A.21}$$

The fourth-moment bound in Equation (A.22) and the preceding event bounds give $\mathbb{P}(\mathcal{G}_N^c) = O(N^{-2})$. Since $\mathcal{G}_N \subseteq \mathcal{G}_N^{\mathrm{Oja}} \cap \mathcal{G}_N^{\mathrm{cov}}$, the Oja and anchored expansions above also hold on this final common event.

On bounded charts the chart-to-projector map is Lipschitz, so Equations (A.18), (A.21) and (3.4) prove all three pairwise comparisons on $\mathcal{G}_N$. Off the event, squared projector distance is at most two. This proves the unconditional Equations (3.6) to (3.8).

A.4 Exact risk constant and endpoint limit

The $\nu + k$ terms in Equation (A.19) are iid and centered, hence

$$\mathbb{E}\|I_k\|^2 = \frac{\Gamma}{\Delta^2(\nu+k)}, \qquad \mathbb{E}\max_{k\le n}\|I_k\|^4 = O(N^{-2}). \tag{A.22}$$

For $\ell(g) = \|g\|^2/(1+\|g\|^2)$, if $g = I_k + R_k$ on the good event, then

$$|\ell(g) - \|I_k\|^2| \le 2\|I_k\|\|R_k\| + \|R_k\|^2 + (\|I_k\| + \|R_k\|)^4.$$

On the common event, both the charts and I_k are bounded, so $\mathbb{E}[\max_k \|R_k\|^4 ; \mathcal{G}_N] \le C\mathbb{E}[\max_k \|R_k\|^2 ; \mathcal{G}_N] = o(N^{-1})$. Cauchy–Schwarz, Equation (A.22), the three remainder estimates, and the bounded loss off the good event give Equation (3.9). The multivariate CLT for Equation (A.19), the pilot expansion, and Slutsky's lemma give Equation (3.10).

For local alternatives $v_{N,h} = \frac{v+V_\perp h/\sqrt{N}}{\|v+V_\perp h/\sqrt{N}\|}$, the Gaussian score in tangent direction h is

$$s_v(x) = \frac{\Delta}{\lambda_1\sigma^2}(v^\top x)V_\perp^\top x,$$

so the per-observation tangent Fisher information is

$$\mathcal{I}_v = \frac{\Delta^2}{\lambda_1\sigma^2} I_{d-1}.$$

Concatenate the observations as $W_{1:N} = (Y_{1:\nu}, X_{1:n})$. The Gaussian experiment is locally asymptotically normal:

$$\log \frac{dP^N_{v_{N,h}}}{dP^N_v} = h^\top \frac{1}{\sqrt{N}} \sum_{m=1}^{N} s_v(W_m) - \frac{1}{2} h^\top \mathcal{I}_v h + o_{P_v}(1). \tag{A.23}$$

Consequently, the bounded-loss local asymptotic minimax theorem gives

$$\lim_{C\to\infty} \lim_{M\to\infty} \liminf_{N\to\infty} \inf_{\widehat{P}_N} \sup_{\|h\|\le M} \mathbb{E}_{v_{N,h}} \min\{NL(\widehat{P}_N, P_{v_{N,h}}), C\}$$
$$\ge \operatorname{tr}(\mathcal{I}_v^{-1}) = \frac{(d-1)\lambda_1\sigma^2}{\Delta^2}. \tag{A.24}$$

Rotational equivariance makes the preceding probability and moment bounds uniform over bounded h, so, for every fixed $M < \infty$,

$$\sup_{\|h\|\le M} \left| N\mathbb{E}_{v_{N,h}} L(P_{w_n}, P_{v_{N,h}}) - \operatorname{tr}(\mathcal{I}_v^{-1}) \right| \longrightarrow 0. \tag{A.25}$$

Thus Equation (3.11) is local-uniform attainment, not merely a pointwise risk calculation.

For the information-retention statement, choose the fixed chart radius r large enough to contain the historical pilot on E_0, and define the retained-summary-measurable truncation $\check{g}_0 = g_v(u)\mathbf{1}\{\|g_v(u)\| \le r\}$. It equals $\bar{g}_0$ on E_0. The pilot expansion gives

$$\frac{\nu}{\sqrt{N}} \mathcal{I}_v \check{g}_0 = \frac{1}{\sqrt{N}} \sum_{j=1}^{\nu} s_v(Y_j) + o_{L^2}(1). \tag{A.26}$$

Therefore the reduced central sequence

$$\Delta_N^{\mathrm{red}} = \frac{\nu}{\sqrt{N}} \mathcal{I}_v \check{g}_0 + \frac{1}{\sqrt{N}} \sum_{i=1}^{n} s_v(X_i) \tag{A.27}$$

is measurable from $(P_u, \nu, X_{1:n})$ at the base point and differs from the full-data central sequence in Equation (A.23) by $o_{L^2}(1)$. The standard LAN score-projection criterion (van der Vaart, 1998) therefore gives the reduced experiment the same tangent information. This is the precise first-order information-retention claim; it is not full sufficiency for nuisance parameters.

B Schedule stability and proof of the plug-in result

Let an $\mathcal{F}_0$-measurable schedule have effective steps $c_{i,N} = \eta_{i,N}/(1+\eta_{i,N}\lambda_1)$, and put

$$\epsilon_N = \max_{i\le n} |\Delta(\nu+i)c_{i,N} - 1|, \qquad L_N = \log(1+n/\nu).$$

If clipping ensures $c_{i,N} \asymp (\nu+i)^{-1}$, positive contractions, and $\|\epsilon_N L_N\|_{L^4} = o(1)$, comparison of the nonharmonic products with the telescoping oracle products gives

$$\mathbb{E} \max_{k\le n} \left\| P_{w_k^{\mathrm{plug}}} - P_{w_k} \right\|_F^2 = o(N^{-1}). \tag{B.1}$$

Indeed, each product ratio is

$$\prod_{\ell=i+1}^{k}\left(1-\frac{\delta_\ell}{\nu+\ell-1}\right)=1+O(\epsilon_N L_N),$$

where $\delta_i=\Delta(\nu+i)c_{i,N}-1$. Put

$$\widehat{r}_i=1-\Delta c_{i,N},\qquad \widehat{R}_i=\prod_{\ell=1}^{i}\widehat{r}_\ell,$$

and define the perturbed linear path

$$\widehat{Z}_k=\widehat{R}_k\bar{g}_0+\sum_{i=1}^{k}c_{i,N}\frac{\widehat{R}_k}{\widehat{R}_i}\beta_i. \tag{B.2}$$

The product comparison, conditional martingale maxima, and the pilot fourth moment give

$$\mathbb{E}\max_{k\le n}\left\|\widehat{Z}_k-\bar{Z}_k\right\|^2=o(N^{-1}).$$

The innovation born at time i has the additional coefficient ratio $c_{i,N}/\{1/[\Delta(\nu+i)]\}=1+\delta_i$. Balance makes $L_N\to\log(1/\rho)>0$, so this factor is absorbed by the same $O(\epsilon_N L_N)$ bound.

For the nonlinear term, set

$$\widehat{U}_i(g)=C_i g-\widehat{r}_i g h_i(g),\qquad \widehat{\mathfrak{q}}_i^H(g)=-\frac{c_{i,N}^2 h_i(g)\{\widehat{U}_i(g)+\beta_i\}}{1+c_{i,N}h_i(g)}$$

on $H_i\cap\{\|g\|\le r\}$, the safe set from Equation (A.4), and set $\widehat{\mathfrak{q}}_i^H=0$ elsewhere. Construct $\widehat{\widehat{g}}_i=\widehat{Z}_i+\widehat{D}_i$ and the predictable indicator $\widehat{J}_i$ as in Equation (A.5). The stopped error has the exact nonharmonic representation

$$\widehat{D}_k=\widehat{R}_k\sum_{i=1}^{k}\widehat{R}_i^{-1}\left\{c_{i,N}\widehat{J}_i\widehat{U}_i+\widehat{J}_i\widehat{\mathfrak{q}}_i^H\right\}. \tag{B.3}$$

Clipping and the product comparison imply $\widehat{R}_k/\widehat{R}_i\le C(\nu+i)/(\nu+k)$. Conditional on $\mathcal{F}_0$, $\widehat{J}_i\widehat{U}_i$ is a martingale difference, and the conditional versions of Equations (A.8) and (A.9) have uniform constants. Consequently, the same bootstrap and maximal argument gives a common event $\widehat{\mathcal{G}}_N$ such that

$$\mathbb{E}\left[\max_{k\le n}\left\|\widehat{D}_k\right\|^2;\widehat{\mathcal{G}}_N\right]=O(N^{-2}),\qquad \mathbb{P}(\widehat{\mathcal{G}}_N^c)=O(N^{-2}). \tag{B.4}$$

Combining the linear-path comparison and Equation (B.4) with bounded projector loss proves Equation (B.1).

For the estimates in Equation (2.3), Gaussian covariance concentration and clipping give

$$\left\|\widehat{\Delta}/\Delta-1\right\|_{L^4}=O(\nu^{-1/2}),\qquad \left\|\widehat{\lambda}_1-\lambda_1\right\|_{L^4}=O(\nu^{-1/2}).$$

For every sufficiently large N, when the auxiliary denominator clip is inactive, the true effective step induced by Equation (3.12) is

$$\widehat{c}_i=\frac{1}{\widehat{\Delta}(\nu+i)+\lambda_1-\widehat{\lambda}_1},$$

so Equation (B.1) applies. The plug-in anchor's old and new linear weights are

$$\frac{\nu\widehat{\Delta}}{\Delta k+\nu\widehat{\Delta}},\qquad \frac{1}{\Delta k+\nu\widehat{\Delta}};$$

they differ relatively from the oracle weights by $O(|\widehat{\Delta}/\Delta-1|)$. The eigendirection lemma controls its quadratic remainder. Projector-risk polarization and Slutsky's lemma complete Corollary 3.2.

For completeness, Equation (3.13) follows by replacing $m_i=\nu+i$ by $a+i$ throughout the stopped argument. Uniformly over $a/\nu\in[c_-,c_+]$, the harmonic sums, product bounds, and no-exit constants remain bounded in the required scales, while direct evaluation of the ideal linear path gives variance $\mathcal{R}_n(a)$.

C DIMENSION AUDIT

Let $M_N = d + \log N$. The pilot moments become

$$\mathbb{E}\,\|\bar{g}_0\|^{2s} \le C(d/\nu)^s, \qquad \mathbb{E}\,\|r_{0,N}\|^2 \le Cd^2/N^2,$$

and Gaussian covariance concentration gives $\mathbb{P}(E_0^c) \le N^{-5}$ under $M_N = o(N)$. Replace the fixed-dimensional observation cap by

$$B_N = \lambda_1\{d + 2\sqrt{6d\log N} + 12\log N + 1\} = O(M_N).$$

Then $\mathbb{P}(\max_i \|\Xi_i\| > B_N) \le N^{-5}$ and $\max_i c_i B_N\sqrt{1+r^2} = O(M_N/N) = o(1)$, preserving the safe denominators. In the stopped recursion, $K_2 = O(d)$, $Q_2^2 = O(d)$, and the dimension-explicit bootstrap is

$$\mathbb{E}\max_k \left\|\widetilde{D}_k\right\|^2 \le C\frac{dA_N + d}{N^2}, \qquad A_N \le Cd + C\frac{dA_N + d}{N}.$$

Thus $A_N = O(d)$ and the Oja discrepancy is $O(d^2/N^2)$. The eigendirection expansion and the relative prefix-covariance event contribute CdM_N/N^2. Bounded projector loss and the common-event failure bound

$$\mathbb{P}(\mathcal{G}_N^c) \le Cd^2/N^2 + N^{-4} = o(d/N)$$

prove Equation (3.15) and its risk counterpart.

D A FIXED-OBJECTIVE STOPPING COROLLARY

The one-pass theorem should not be mislabeled as early stopping because its empirical objective changes with k. A separate fixed-objective statement is nevertheless available. Let S_n be the covariance of the fixed new sample and run

$$w_0^{\mathrm{FG}} = u, \qquad w_\ell^{\mathrm{FG}} = \frac{(I + \eta_N S_n)w_{\ell-1}^{\mathrm{FG}}}{\left\|(I + \eta_N S_n)w_{\ell-1}^{\mathrm{FG}}\right\|}, \qquad 1 \le \ell \le K.$$

Here K is fixed as $N \to \infty$, and is chosen large enough, uniformly on the compact parameter set, that $c_N\lambda_1 < 1$, where

$$r_N^K = (1 - c_N\Delta)^K = \frac{\nu}{\nu + n}, \qquad c_N = \frac{1 - (\nu/(\nu+n))^{1/K}}{\Delta},$$

and $\eta_N = c_N/(1 - c_N\lambda_1)$.

For completeness, put

$$E_n = S_n - \Sigma_v, \qquad \bar{\beta} = V_\perp^\top E_n v = n^{-1}\sum_{i=1}^n \xi_i.$$

Linearizing the exact constant-matrix chart map gives

$$z_\ell = r_N z_{\ell-1} + c_N\bar{\beta}, \qquad z_K = r_N^K\bar{g}_0 + \frac{1 - r_N^K}{\Delta}\bar{\beta}.$$

For a sufficiently small fixed $\varepsilon > 0$, define

$$\mathcal{A}_N = E_0 \cap \{\|\bar{g}_0\| + \|E_n\|/\Delta \le \varepsilon\}.$$

Gaussian concentration and balance give $\mathbb{P}(\mathcal{A}_N^c) = o(N^{-1})$. On $\mathcal{A}_N$, the K-fold rational chart map is twice continuously differentiable in (g_0, E_n), with derivative bounds depending only on the compact parameter set and fixed K. Taylor's theorem yields

$$\left\|g_v(w_K^{\mathrm{FG}}) - z_K\right\| \le C_K\{\|\bar{g}_0\| + \|E_n\|/\Delta\}^2.$$

Let $B_N = \|\bar{g}_0\| + \|E_n\|/\Delta$. Since $\mathbb{E}[B_N^4; \mathcal{A}_N] = O(N^{-2})$, the quadratic Taylor bound yields the following expansion and remainder estimate:

$$g_v(w_K^{\mathrm{FG}}) = \frac{\nu}{\nu+n}g_v(u) + \frac{1}{\Delta(\nu+n)}\sum_{i=1}^n \xi_i + R_{K,N}, \qquad \mathbb{E}[\|R_{K,N}\|^2; \mathcal{A}_N] = O(N^{-2}),$$

on an event with $\mathbb{P}(\mathcal{A}_N^c) = o(N^{-1})$. Therefore

$$\mathbb{E}\left\|P_{w_K^{\mathrm{FG}}} - P_{\widetilde{q}_n}\right\|_F^2 = o(N^{-1}), \qquad \mathbb{E}L(P_{w_K^{\mathrm{FG}}}, P_v) = \frac{\Gamma}{\Delta^2 N} + o(N^{-1}).$$

No uniformity in $K \to \infty$ is claimed. In the iterated limit,

$$\lim_{K\to\infty}\lim_{N\to\infty} Kc_N = \Delta^{-1}\log(1/\rho) = \lim_{N\to\infty}\Delta^{-1}\log(1+n/\nu).$$

E RIEMANNIAN FLOW COMPARISONS

Throughout this appendix d is fixed. We reuse the coupled pilot and its moments from Lemma A.1; in particular, the actual chart equals $\bar{g}_0$ only on E_0. No unconditional chart moment is used.

E.1 A UNIFORM COMPLETE-EXPONENT EXPANSION

Put $P = vv^\top$. For a in a fixed compact interval of $[0,\infty)$, symmetric H, and local g, define

$$F(a,g,H) = \frac{V_\perp^\top e^{aP+H}(v+V_\perp g)}{v^\top e^{aP+H}(v+V_\perp g)}.$$

At $g = H = 0$, the denominator is e^a and the numerator is zero. The complete Fréchet derivative

$$De^{aP}[H] = \int_0^1 e^{(1-s)aP} H e^{saP}\, ds$$

does not require P and H to commute. The derivative of the denominator cancels from the first-order quotient, giving

$$F(a,g,H) = e^{-a}g + \psi(a)V_\perp^\top Hv + R(a,g,H), \quad \psi(a) = \frac{1-e^{-a}}{a}, \quad \psi(0) = 1. \tag{E.1}$$

The exact remainder is the integral of $(1-s)D^2_{(g,H)}F(a,sg,sH)[(g,H),(g,H)]$ over $s\in[0,1]$. On a sufficiently small fixed-radius neighborhood, the denominator is uniformly separated from zero and this entire Hessian is uniformly bounded. Consequently

$$\|R(a,g,H)\| \le C(\|g\| + \|H\|)^2.$$

This bound is applied after cancelling the scalar identity part of the exponent and differentiating the full quotient. It retains mixed pilot/covariance and numerator/denominator contributions.

E.2 FIXED-OBJECTIVE TIME PROFILE

Set $E = S_n - \Sigma_v$, $b = V_\perp^\top Ev$, and $y(t) = e^{-\Delta t}$. The scalar factor $e^{t\sigma^2}$ cancels exactly from normalization. Substitution of $a = \Delta t$, $H = tE$, and $g = \bar{g}_0$ in Equation (E.1) gives, on E_0, the leading path

$$z(t) = y(t)\bar{g}_0 + \frac{1-y(t)}{\Delta}b. \tag{E.2}$$

The actual flow has remainder bounded uniformly for $t \le T$ by $C(\|\bar{g}_0\| + \|E\|/\Delta)^2$ on a fixed-radius event.

At $t = 0$, the finite definition of the anchored matrix is

$$q^{\mathrm{F}}(t) = \mathrm{Top}_1\left(\frac{1-y(t)}{\Delta}S_n + y(t)P_u\right).$$

Its population baseline has gap one, including at zero. For $t > 0$, apply Equation (4.4) with $H^{\mathrm{A}} = E + \mu(t)(P_u - P)$ and $D = \Delta + \mu(t)$. The exact coefficients are

$$\frac{\mu(t)}{D} = y(t), \qquad \frac{1}{D} = \frac{1-y(t)}{\Delta}.$$

On E_0, the resulting off-diagonal term is

$$D^{-1}V_\perp^\top H^{\mathrm{A}} v = \frac{1-y(t)}{\Delta} b + y(t)\frac{\bar{g}_0}{1+\|\bar{g}_0\|^2}.$$

Thus the common first-order path cancels before estimating the difference; the remaining pilot correction is cubic. Moreover, $\left\|H^{\mathrm{A}}\right\|/D \le \|E\|/\Delta + \|P_u - P\|$, uniformly in time. On a sufficiently small fixed-radius event this is at most $1/8$, so Equation (4.4) yields a quadratic remainder plus that cubic term.

By Lemma A.1 and fixed-dimensional Gaussian covariance moments, the squared supremum of each remainder has expectation $O(N^{-2})$; the pilot cubic term uses its sixth moment. The common radius event, intersected with E_0, fails with probability $O(N^{-2})$ by fourth moments. Both charts are bounded there. The full squared difference has only remainder cross terms, which are controlled at the same N^{-2} scale. The projector map is Lipschitz on these charts, and its loss on the complement is bounded. This proves Equation (5.4).

For risk, replace the pilot in Equation (E.2) using Equation (A.2). The common efficient path is

$$I(t) = \frac{y(t)}{\Delta\nu}\sum_{j=1}^{\nu}\xi_j^Y + \frac{1-y(t)}{\Delta n}\sum_{i=1}^{n}\xi_i.$$

Independence and centering eliminate its old/new cross term, giving exactly the leading second moment in Equation (5.5). Its supremum has fourth moment $O(N^{-2})$. The complete loss identity

$$\frac{\|I+R\|^2}{1+\|I+R\|^2} - \|I\|^2 = 2\langle I, R\rangle + \|R\|^2 - \frac{\|I+R\|^4}{1+\|I+R\|^2}$$

retains the cross term and quartic correction. Bounded charts and the second-moment remainder bound give a localized fourth-moment remainder $O(N^{-2})$; the influence fourth moment also controls its missing energy off the event. The cross term is $O(N^{-3/2})$, and the bounded-loss complement costs $O(N^{-2})$, proving the uniform risk expansion. Finally, the intact quadratic identity

$$\frac{y^2}{\nu} + \frac{(1-y)^2}{n} = \frac{1}{N} + \frac{N}{\nu n}\left(y - \frac{\nu}{N}\right)^2$$

proves the first-order stopping claim. The endpoint influence agrees with pooled PCA, so its existing limit law transfers.

E.3 All prefixes and historical-gap plug-in

Let $T_k = \sum_{i\le k}(X_iX_i^\top - \Sigma_v)$, $a_k = \Delta t_k = \log(1+k/\nu)$, and $H_k = (t_k/k)T_k$, with $H_0 = 0$. The complete exponent is $t_kS_k = t_k\sigma^2 I + a_kP + H_k$. Although $S_k - \Sigma_v$ need not be small for short prefixes, the exact time factor satisfies $t_k/k \le 1/(\Delta\nu)$. After centering the intact matrix sum, the finite-dimensional martingale fourth-moment bound gives

$$\mathbb{E}\max_{k\le n}\|T_k\|^4 \le Cn^2, \qquad \mathbb{E}\max_{k\le n}\|H_k\|^4 = O(N^{-2}).$$

This can be obtained by vectorizing T_k, applying the same maximal moment argument as for the score sums, and using fixed dimension. On E_0, the linear term in Equation (E.1) is exactly

$$e^{-a_k}\bar{g}_0 + \psi(a_k)V_\perp^\top H_k v = \frac{\nu}{\nu+k}\bar{g}_0 + \frac{1}{\Delta(\nu+k)}\sum_{i\le k}\xi_i = \bar{Z}_k.$$

The radius event $E_0 \cap \{\|\bar{g}_0\| + \max_k\|H_k\| \le \epsilon\}$ fails with probability $O(N^{-2})$, and the localized expected squared maximal flow remainder is $O(N^{-2})$. Intersect with the Oja event in Theorem 3.1: the entire influence path cancels in the difference, and the remainder cross term preserves this order. Bounded projector loss on the complement proves Equation (5.2).

For plug-in times $\widehat{t}_k = \widehat{\Delta}^{-1}\log(1+k/\nu)$, put $r = \Delta/\widehat{\Delta}$, $\ell_k = \log(1+k/\nu)$, and $b_k = \sum_{i\le k}\xi_i$. The clipping and fourth-moment calibration bounds in Appendix B give r in a fixed positive compact interval and $\|r-1\|_{L^4} = O(N^{-1/2})$. The complete difference of the two linear flow paths is

$$(e^{-r\ell_k} - e^{-\ell_k})\left(\bar{g}_0 - \frac{b_k}{\Delta k}\right), \qquad k \ge 1.$$

Retain both its signs and the exact identity

$$e^{-r\ell_k} - e^{-\ell_k} = -(r-1)\ell_k \int_0^1 e^{-[1+s(r-1)]\ell_k}\, ds.$$

The factor $\ell_k/k \le 1/\nu$ removes the short-prefix singularity. The complete maximal difference is bounded by $C|r-1|(\|\bar{g}_0\| + \nu^{-1}\max_k \|b_k\|)$. The first product uses fourth moments, without independence between the old gap and pilot. For the second, the gap estimate is independent of the new score path. Its expected squared maximum is therefore $O(N^{-2})$. The same uniform-exponent localization applies under clipping, proving Equation (5.2) also for the plug-in flow against oracle Oja. First-order equivalence to plug-in Oja follows from Corollary 3.2. No growing-dimensional flow or unbounded-time conclusion is asserted.

E.4 Exact sample-wise flow interpretation

For one observation x, the exact Riemannian ascent flow of $\frac{1}{2}(w^\top x)^2$ is

$$\Phi_x^s(w) = \frac{e^{sxx^\top} w}{\|e^{sxx^\top} w\|}, \qquad e^{sxx^\top} = I + \frac{e^{s\|x\|^2} - 1}{\|x\|^2} xx^\top.$$

Hence a positive raw Oja step η_i is exactly the arrival endpoint of this flow at sample-dependent duration

$$s_i = \frac{\log(1+\eta_i\|X_i\|^2)}{\|X_i\|^2}, \qquad \Phi_{X_i}^{s_i}(w_{i-1}) = \frac{(I+\eta_i X_i X_i^\top)w_{i-1}}{\|(I+\eta_i X_i X_i^\top)w_{i-1}\|},$$

with $s_i = \eta_i$ at $X_i = 0$. The original theorem transfers verbatim to these endpoints, including plug-in calibration. This is an exact interpretation, not additional statistical novelty. It is not a fixed-objective flow, an SDE, or the geodesic exponential-map step at the same numerical step size.

F Scope of the equal-gap assumption

This appendix explains a restriction of the positive theorem. In Equation (2.1), a rank-one isotropic spike and equal tangent gaps are the same structural assumption. The restriction is natural for this model but limits the generality of the pooled-fusion conclusion. The statement below concerns only a linearized recursion with a fixed scalar harmonic schedule.

Proposition F.1 (Scalar harmonic fusion fails across unequal gaps)**.** *Fix $\gamma > 0$. Let a general covariance have eigenbasis $v_1 = v, v_2, \dots, v_d$, leading eigenvalue λ_1, and tangent gaps $\delta_j = \lambda_1 - \lambda_j > 0$. For $j \ge 2$, define the new and historical mode scores*

$$\xi_{j,i} = (v_1^\top X_i)(v_j^\top X_i), \qquad \xi_{j,t}^Y = (v_1^\top Y_t)(v_j^\top Y_t).$$

Consider the first-order Oja recursion in mode j,

$$z_{j,i} = \left(1 - \frac{\gamma\delta_j}{\nu+i}\right) z_{j,i-1} + \frac{\gamma}{\nu+i}\xi_{j,i}, \tag{F.1}$$

where $\nu/(\nu+n) \to \rho \in (0,1)$. Assume the pilot is measurable from the historical sample and is its coupled/truncated empirical-PCA chart (or its influence linearization), with

$$z_{j,0} = \frac{1}{\nu\delta_j}\sum_{t=1}^{\nu} \xi_{j,t}^Y + e_{j,N}, \qquad \|e_{j,N}\|_{L^2} = o(\nu^{-1/2}),$$

where both score arrays are centered and the historical and new observations are independent. In each compared mode assume

$$\liminf_N \nu \mathbb{E} z_{j,0}^2 > 0.$$

No unconditional chart moment of the raw empirical eigenvector is assumed. If, on the same realization, the endpoint differs from the pooled empirical-PCA linearization by $o_{L^2}(N^{-1/2})$ in two modes j and ℓ, then $\delta_j = \delta_\ell = 1/\gamma$. Hence no schedule $c_i = \gamma/(\nu+i)$ yields pooled first-order equivalence across unequal eigengaps.

Proof. The old-pilot coefficient in mode j is

$$A_{j,N} = \prod_{i=1}^{n} \left(1 - \frac{\gamma\delta_j}{\nu + i}\right).$$

Balance and a logarithmic expansion give

$$\log A_{j,N} = -\gamma\delta_j \sum_{i=1}^{n} (\nu + i)^{-1} + o(1) \longrightarrow \gamma\delta_j \log \rho,$$

so $A_{j,N} \to \rho^{\gamma\delta_j}$. Write $\rho_N = \nu/N$ and let $z_{j,n}^{\mathrm{pool}} = (\delta_j N)^{-1}(\sum_t \xi_{j,t}^{Y} + \sum_i \xi_{j,i})$. The exact endpoint difference is

$$z_{j,n} - z_{j,n}^{\mathrm{pool}} = (A_{j,N} - \rho_N) z_{j,0} + \rho_N e_{j,N} + \sum_{i=1}^{n} \left(b_{j,i,N} - \frac{1}{\delta_j N}\right) \xi_{j,i},$$

where $b_{j,i,N} = \gamma(\nu + i)^{-1} \prod_{s=i+1}^{n} (1 - \gamma\delta_j/(\nu + s))$. Conditioning on all historical observations eliminates the centered new sum. Conditional expectation is an L^2 contraction, so the assumed endpoint comparison and pilot remainder force $(A_{j,N} - \rho_N) z_{j,0} = o_{L^2}(N^{-1/2})$. Pilot nondegeneracy therefore gives $A_{j,N} - \rho_N \to 0$. Hence $\rho^{\gamma\delta_j} = \rho$, so $\gamma\delta_j = 1$. The same argument for ℓ proves the result. □

This obstruction does not say that Oja fails for a general covariance. It says that the particular first-order pooled-fusion phenomenon cannot hold across unequal gaps within the scalar harmonic family $c_i = \gamma/(\nu + i)$ under the stated pilot assumptions. General spectra call for either a weaker anisotropic bias–variance comparison or a preconditioned update approximating the inverse gap operator.

G Detailed comparison with trajectory and efficiency results

The relevant distinction is not whether a paper studies a "path," but which path, which data are retained, and which discrepancy is controlled. Table 3 separates several established comparison targets. Our discrete theorem uses the original observations to couple nonlinear estimators, retains the empirical historical error at first order, and controls an unconditional expectation of the maximum over arriving prefixes. This is different from either a family of marginal risk bounds or weak convergence of a rescaled process. Conversely, our balanced Gaussian-spike theorem does not subsume general-spectrum guarantees or infinite-horizon concentration results.

G.1 Closest discrete-Oja precedents

In arXiv:2104.00512v2, Liang (2023, Section 3.2, Lemmas 3.1 and 3.3) analyzes the whole warm-start graph-chart recursion through its linear contraction, martingale, and drift remainder. This is close methodological precedent for our proof, not merely a population-rate statement. Together with Huang et al. (2021); Li et al. (2018), it establishes that harmonic warm starts, localization, and discrete moment analysis are standard tools. The extra target here is the realized historical-pilot influence and its simultaneous agreement with two empirical refit paths in expected maximal projector loss.

Pham et al. (2025, Section 3.4, Proposition 4 and Theorem 3) treat bounded observations, a high-probability warm angle, and harmonic steps $2/\{\mathrm{gap}(t + L)\}$, with population-loss control for every $t \geq 0$. Neither harmonic steps nor time uniformity is therefore a novelty claim here. Sparse Oja (Kumar & Sarkar, 2024) also provides statistically sharp estimation through support selection and restricted refitting, but does not state the retained-pilot three-path coupling.

Distributional results further narrow the distinction. Lunde et al. (2021); Kumar et al. (2025) give squared-angle or coordinate-wise inference, and Xu & Li (2026, Section 3.1, Proposition 3) linearize normalized Oja for general-rank subspaces. The latter uses Haar initialization and horizon-dependent constant steps under sub-Gaussian sampling. Our result instead keeps the first-order error of an empirical historical pilot and uses a count-shifted harmonic schedule. We do not claim novelty for normalization-aware linearization or replace their broader inference theory.

Table 3: Different trajectory comparisons in the closest literature. The distinctions concern the stated target and access model, not a ranking of the results. Detailed locations and version qualifications appear below.

Work	Path and comparison	Distinction from our prefix theorem
Balzano (2022)	Same-stream Oja/GROUSE iterates; exact subspace identity after step conversion	Alternative online updates, not exact empirical PCA refits
Pham et al. (2025)	All-time Oja population-angle bounds	Population target and infinite horizon, not pairwise empirical-path error
Arnaudon et al. (2012); Flamand & Brunel (2026)	Weak limits of unaveraged harmonic stochastic-gradient trajectories	Process laws, not the stated same-data expected-maximal PCA coupling
Ali et al. (2019; 2020)	Fixed-data optimization-time flow and regularization paths	Least-squares coefficients; the data do not grow along the path
Chen et al. (2022); Li et al. (2026b)	Communication-round iterates compared with the pooled empirical eigenspace	Repeated covariance access on a fixed distributed dataset
Theorem 3.1	Oja, anchored, and pooled PCA on every incoming prefix	Compressed historical state; expected maximal squared projector discrepancy $o(N^{-1})$ in the stated model

G.2 Historical summaries and empirical pooled comparisons

Hall et al. (1999, Section 2) combine eigenmodels using retained means, eigenvectors, eigenvalues, and counts; its Section 2.3 leaves rigorous approximation-error analysis open. History PCA (Yang et al., 2018, Section 4.1, Algorithm 1; Section 4.4, Theorems 1–2) combines a count-weighted *previous* projector with an incoming block, and its one-inner-iteration specialization is harmonic Oja. That summary is updated repeatedly, whereas our anchored comparator keeps the original historical direction and uses the cumulative new-prefix covariance. The earlier algorithms establish the relevance of compression; our result quantifies its first-order effect on the pooled estimator path.

The distributed-PCA literature already contains direct empirical comparisons. In arXiv:2004.02336v3, Chen et al. (2022, Theorem 3.3, Corollary 3.4, and Theorem 3.10) controls optimization iterates relative to eigenspaces of the same pooled empirical covariance, with enlarged-eigenspace formulations when gaps are small. Algorithm 1 repeatedly evaluates retained local covariance operators. Its path index counts communication or optimization rounds on fixed data, not incoming observations after the historical covariance is discarded.

Li et al. (2026b) give an even closer first-order pooled comparison. In the inspected arXiv:2503.03123v3, the final display of Appendix A.2 proves $\|P^{(2)} - P_{\mathrm{pool}}\|_F^2 = O_p(K^{-2})$ in the Gaussian-spike regime of Theorem 4, where K is the number of machines. Appendix A.5 also compares iterates with pooled PCA and controls quantities uniformly over communication rounds. Thus neither same-data pooled agreement, below-statistical-scale accuracy, nor uniformity alone distinguishes our work. Our theorem combines an arrival-indexed, changing empirical objective with one-pass compressed-history access and an unconditional expected prefix maximum. The reference now has a journal publication; the theorem locations above refer to the explicitly identified preprint version, not unverified numbering in the journal version.

G.3 Efficient recursion and functional stochastic approximation

Sharia (2010, Section 2, Lemma 1) establishes asymptotic linearity of recursive estimators, and Tripuraneni et al. (2018, Theorems 1 and 3; Appendix C.2.1, Theorem 8) couple Riemannian stochastic gradients to a linear recursion to prove efficiency of their averages, including PCA with unequal gaps. These results explain why the common leading influence and endpoint efficiency are natural goals. They do not, by themselves, state the present normalized, unaveraged Oja/anchor/pooled comparison with a historical sample of nonnegligible relative size.

Kutoyants & Motrunich (2016) construct efficient estimator processes from a preliminary estimator based on a sublinear learning sample, with recursive implementation discussed separately; its Theorem 1 asserts efficiency at each fixed positive sample fraction, not our expected-maximal comparison. Our historical sample remains comparable to the new sample, so its realized fluctuation must persist in the matched path rather than disappear asymptotically. Li et al. (2026a, Theorem 1, Equation (19)) give a uniform same-noise martingale representation for the partial-sum/averaged SA path. Its condition $t\eta_t \to \infty$ excludes harmonic steps; the conclusion is in probability rather than the expected maximal projector statement here.

Unaveraged harmonic process limits also have precedent: Arnaudon et al. (2012, Theorem 2.6) study stochastic approximation for manifold means, and Flamand & Brunel (2026, Theorem 3) prove a functional central limit theorem for harmonic SGD in a convex setting. Such results capture temporal dependence, not just one endpoint. The latter theorem uses a fixed initialization and a gain strictly larger than the inverse smallest Hessian eigenvalue; its topology is uniform on compact positive time intervals. Flamand and Brunel also bound the expected supremum of the limiting diffusion in their Theorem 4, not a finite-sample same-data estimator discrepancy. These process results do not identify pooled and anchored PCA on the original observations. This distinction does not rule out adapting classical SA methods; the contribution claimed here is the quantitative PCA comparison, not a new general theory of efficient stochastic approximation.

G.4 Exact algorithm identities and regularization paths

Balzano (2022, Theorem 1, Equations (14)–(16)) makes Oja, GROUSE, and a projected-gradient update generate identical subspaces at each step through state- and sample-dependent step conversion. This is already an exact same-stream trajectory comparison. The comparators are alternative updates, however, not exact pooled or anchored empirical eigendirections.

For a fixed least-squares dataset, Ali et al. (2019, Section 3, Lemmas 3–4) represent discrete gradient descent and gradient flow by generalized quadratic penalties. The flow penalty has eigenvalues $s/(e^{ts} - 1)$, making it a direct conceptual precedent for the exponential gap calibration in Equation (5.3). Ali et al. (2020, Theorems 2–3) give risk and same-data coefficient-distance bounds for stochastic gradient flow and ridge. The coefficient bound is stated at each optimization time, rather than as an expectation of a trajectory supremum. Our fixed-data companion instead compares empirical Rayleigh flow with a nonlinear anchored eigendirection, uniformly over a bounded time interval in the stated statistical model.

AgFlow (Jiang et al., 2021, Sections 3.3–3.4) constructs approximate PCA scores and uses them as responses in a regression path; its Theorem 1 controls score quality, while subsequent comparisons use the least-squares flow/ridge theory. Our companion uses the Rayleigh objective itself. Most importantly, Corollary 5.1 also aligns restarted empirical flows with the incoming-prefix Oja path. This family is an interpretation requiring batch covariance products, not a one-pass flow algorithm.

Taken together, these comparisons support a narrowly stated contribution: the joint retained-pilot, same-realization, expected-maximal prefix theorem and its empirical-flow interpretation. We do not claim that path analysis, pooled equivalence, or calibrated efficient recursion is new in isolation.