Overlap-Driven Subspace Drift in Rolling Covariance Windows

June 28, 2026

Abstract

We analyze adjacent increments of top-r spectral projectors computed from rolling sample covariance matrices. Adjacent windows of length W share $W-1$ observations, so their covariance difference is driven only by the entering and exiting boundary observations and has rank at most two. Under a static Gaussian bulk-plus-spikes null, conditioning on the shared core yields: conditional exchangeability of the two adjacent projectors; an exact one-pair sign-randomization calibration for oriented projector contrasts; a null fluctuation benchmark of order r/W^2; and, in the fixed-rank strongly spiked regime, the leading expansion

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{2}{\alpha_W^2}\frac{\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{W^2} + o(W^{-2}) = \frac{2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{(W-1)^2} + o(W^{-2}),$$

where $\alpha_W = (W-1)/W$ and $\mathfrak{d}_{\mathrm{sub}}$ is the spectral difficulty of Definition 4.1. Thus the common shorthand $2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)/W^2$ is asymptotically correct, while $2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)/(W-1)^2$ is the more accurate finite-W leading benchmark. Because the population subspace is fixed under the null, this constant is a noise floor for detecting genuine subspace movement, not a measure of drift itself. We also give a boundary-driven least-favorable family showing that this r/W^2 scale is sharp for the corresponding rolling drift target.

1 Introduction

1.1 The adjacent subspace increment

Let $\{r_s\}$ be a sequence of mean-zero random vectors in $\mathbb{R}^n$, let $I_t = \{t-W+1, \dots, t\}$ denote the rolling window of length W ending at time t, and define the rolling sample covariance

$$\widehat{\Sigma}_t = \frac{1}{W}\sum_{s \in I_t} r_s r_s^\top.$$

For a symmetric matrix A with a positive top-r eigengap, let $P(A)$ denote its top-r spectral projector, and write $\widehat{P}_t = P(\widehat{\Sigma}_t)$. The object of this paper is the *adjacent subspace increment*

$$\widehat{P}_t - \widehat{P}_{t-1}.$$

Our focus is neither the covariance estimate itself nor single-window principal component analysis, but the increment of the estimated principal subspace between two consecutive windows. The windows I_{t-1} and I_t overlap in $W-1$ observations, so that

$$\widehat{\Sigma}_t - \widehat{\Sigma}_{t-1} = \frac{1}{W}\left(r_t r_t^\top - r_{t-W} r_{t-W}^\top\right). \tag{1}$$

Although each estimate $\widehat{\Sigma}_t$ is built from W observations, their difference has rank at most two and depends only on the entering observation r_t and the exiting observation r_{t-W}. The rank-two boundary identity (1) is the source of the faster decay rate that we establish for the adjacent increment.

1.2 Related work

Principal subspaces and spectral projectors of covariance matrices are central objects in multivariate statistics. Their behavior in high dimensions has been studied extensively under the spiked covariance model of Johnstone [11]; see Paul [17] and Benaych-Georges and Nadakuditi [3] for the eigenstructure of low-rank perturbations of large random matrices. Finite-sample and minimax theory for principal subspace estimation includes Vu and Lei [21], Cai, Ma and Wu [5], and Cai and Zhang [6]. The underlying perturbation theory for spectral projectors goes back to Davis and Kahan [7], Wedin [23], and Kato [12]; see also Stewart and Sun [18] and Bhatia [4]. Sharp operator-norm concentration for sample covariance operators in terms of effective rank is due to Koltchinskii and Lounici [13].

Rolling-window covariance and subspace estimates are widely used to track time-varying second-order structure and to localize changes in covariance [2, 22]. In such analyses the temporal overlap between consecutive windows is typically not exploited. The present paper isolates the effect of this overlap on the projector increment and shows that, under a static null, it produces a drift scale smaller by one factor of W than the single-window estimation error.

The scope is intentionally narrow. The exact exchangeability and one-pair sign-calibration statements use an i.i.d. boundary pair independent of the shared core; serial dependence or local nonstationarity can break the swap argument unless the data are first blocked, resampled, or modeled so that an exchangeable boundary experiment remains valid. Likewise, the constant-level expansion is proved in a strongly spiked regime with fixed r spikes of order n, an $O(1)$ bulk, and a gap of order n, so the effective rank is $O(1)$ and the rates are dimension-free in this class. These assumptions are natural for the clean null calculation, but they are substantive modeling restrictions for rolling-window applications.

1.3 Contributions

We work under a static Gaussian bulk-plus-spikes null, made precise in Section 2. Under this null the population subspace does not move, so the increment $\widehat{P}_t - \widehat{P}_{t-1}$ is pure estimation noise, and its size is the natural benchmark against which true subspace drift must be detected. Our contributions are the following.

1. *Conditional exchangeability* (Proposition 5.1). Conditional on the shared core of $W-1$ common observations, the two adjacent projectors are independent and identically distributed; in particular the increment is conditionally mean-zero.
2. *Exact one-pair randomization calibration* (Theorem 5.2 and Corollary 5.3). Any antisymmetric functional of a fixed adjacent projector pair — in particular any linear contrast of $\widehat{P}_t - \widehat{P}_{t-1}$ whose contrast is measurable with respect to the shared core — has a sign-symmetric conditional law. This gives an exact one-pair randomization calibration for oriented contrasts. It is intentionally a coarse single-pair calibration; using many overlapping rolling times additionally requires blocking, a dependence argument, or resampling calibration.
3. *Adjacent-window null-increment benchmark* (Theorem 5.11). The expected squared Frobenius increment is of order r/W^2, in contrast to the single-window projector error of order r/W.
4. *Constant-level expansion* (Theorem 5.12). In the fixed-rank strongly spiked Gaussian regime, the operative finite-W leading approximation is

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{2}{\alpha_W^2}\frac{\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{W^2} + o(W^{-2}) = \frac{2\,\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{(W-1)^2} + o(W^{-2}).$$

   The limiting shorthand $2\mathfrak{d}_{\mathrm{sub}}/W^2$ is asymptotically equivalent. This is a regime-specific Gaussian constant, not a universal high-dimensional PCA constant.

5. *Boundary-driven least-favorable sharpness* (Proposition 7.1 and Remark 7.2). A boundary-driven hypercube gives a minimax lower bound of order r/W^2 for the rolling drift target $D_t^{\mathrm{roll}} = P(\bar{\Sigma}_t) - P(\bar{\Sigma}_{t-1})$. Over that same local family the drift itself is at the null scale, so the zero estimator matches the lower-bound rate. This section should therefore be read as a sharpness statement at the null floor, not as a claim of nontrivial recovery below that floor.

Standard perturbation and concentration results used as inputs are collected in Section 3.

1.4 Notation

For a symmetric matrix A, $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ are its ordered eigenvalues, $\|A\|_{\mathrm{op}}$ its operator norm, $\|A\|_F$ its Frobenius norm, and $\langle A, B\rangle_F = \mathrm{tr}(A^\top B)$ the Frobenius inner product. We write $a \lesssim b$ if $a \leq Cb$ for a constant C depending only on the fixed spectral constants and the rank r, $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$, and $a = \Theta(b)$ analogously. For a Hilbert-space-valued random variable X and a sub-σ-algebra $\mathcal{G}$,

$$V_F(X \mid \mathcal{G}) := \mathbb{E}[\|X - \mathbb{E}[X \mid \mathcal{G}]\|_F^2 \mid \mathcal{G}]$$

is the conditional Frobenius variance.

2 Setup

Let $P(A)$ denote the top-r spectral projector of a symmetric matrix A whenever it is uniquely defined. On null sets with eigenvalue ties, fix any measurable tie-breaking convention. The static-null results below use the following i.i.d. Gaussian model.

Assumption 2.1 (Static Gaussian null)**.** On $I_t \cup I_{t-1}$, the vectors r_s are independent and identically distributed as $N(0, \Sigma)$.

Assumption 2.2 (Bulk-plus-spikes spectrum)**.** The rank parameter r is fixed. Eigenvalues are ordered as

$$\lambda_1(\Sigma) \geq \lambda_2(\Sigma) \geq \cdots \geq \lambda_n(\Sigma).$$

There exist constants $0 < m < M < \infty$ and $0 < c_\lambda \leq C_\lambda < \infty$, independent of n and W, such that

$$m \leq \lambda_n(\Sigma) \leq \cdots \leq \lambda_{r+1}(\Sigma) \leq M, \qquad c_\lambda n \leq \lambda_r(\Sigma) \leq \cdots \leq \lambda_1(\Sigma) \leq C_\lambda n.$$

Assumption 2.3 (Population eigengap)**.** Let $\delta = \lambda_r(\Sigma) - \lambda_{r+1}(\Sigma)$. Assume

$$\delta \geq c_\delta n.$$

Assumption 2.4 (Asymptotic convention)**.** The spike rank r is fixed, n may grow, and $W = W_n \to \infty$. Whenever exponential sample-covariance concentration is invoked in a theorem statement, we assume $W \geq C_{\log} \log n$ and $W \geq W_0$, with constants $C_{\log}$ and W_0 depending only on the fixed spectral constants and on r. This global convention is stronger than what the effective-rank concentration input alone requires; it is retained to make all later high-probability and union-bound statements uniform.

The bulk-plus-spikes scaling of Assumption 2.2 places the model in the spiked regime studied by Johnstone [11] and Paul [17]: r spikes of order n separated from an $O(1)$ bulk by a gap of order n. This is the benign, well-separated regime; the consequences of relaxing it are discussed in Remark 8.1.

3 Auxiliary standard inputs

We collect the standard perturbation and concentration results used in the sequel. They are stated in the form convenient for our purposes and are included to keep the paper self-contained.

Lemma 3.1 (First-order projector expansion). *Let $A = A^\top$ have eigen-decomposition*

$$A = \sum_{k=1}^{n} \mu_k v_k v_k^\top, \qquad \mu_1 \geq \cdots \geq \mu_n,$$

and suppose the top-r eigengap $\gamma = \mu_r - \mu_{r+1}$ is positive. Let $P(A)$ be the top-r spectral projector of A. Define the reduced-resolvent derivative

$$\mathcal{L}_A(E) = \sum_{i \leq r < j} \frac{v_j v_j^\top E\, v_i v_i^\top + v_i v_i^\top E\, v_j v_j^\top}{\mu_i - \mu_j}.$$

There exist constants $c_0 \in (0, 1/4]$ and $C_r < \infty$, depending at most on the fixed rank r, such that if $\|E\|_{\mathrm{op}} \leq c_0 \gamma$, then the top-$r$ projector $P(A+E)$ is uniquely defined and

$$\|P(A+E) - P(A) - \mathcal{L}_A(E)\|_F \leq C_r \frac{\|E\|_{\mathrm{op}}^2}{\gamma^2}.$$

Consequently,

$$\|P(A+E) - P(A)\|_F \leq C_r \frac{\|E\|_{\mathrm{op}}}{\gamma}.$$

Proof. Let $P = P(A)$ and write $\widetilde{P} = P(A+E)$. By Weyl's inequality [4, Ch. III], if $\|E\|_{\mathrm{op}} \leq \gamma/4$ then the top-r spectral cluster of $A+E$ remains separated from the lower cluster by at least $\gamma/2$, so $\widetilde{P}$ is uniquely defined. We use the standard local perturbation expansion for spectral projectors across an isolated gap [12, 7]: the map $H \mapsto P(A+H)$ is twice Fréchet differentiable on the operator-norm ball $\|H\|_{\mathrm{op}} \leq \gamma/4$, its derivative at zero is the reduced-resolvent operator $\mathcal{L}_A$, and

$$\left\| D^2 P(A+H)[B_1, B_2] \right\|_{\mathrm{op}} \leq C \frac{\|B_1\|_{\mathrm{op}} \|B_2\|_{\mathrm{op}}}{\gamma^2}$$

for all $\|H\|_{\mathrm{op}} \leq \gamma/4$. This is the standard second-order bound for spectral projectors across an isolated spectral gap; equivalently, it follows from the Riesz-projector (contour-integral) representation [12, Ch. II] using any contour that strictly separates the top-r cluster from its complement.

Taylor's formula therefore gives

$$\|P(A+E) - P(A) - \mathcal{L}_A(E)\|_{\mathrm{op}} \leq C \frac{\|E\|_{\mathrm{op}}^2}{\gamma^2}.$$

The linear term has the displayed coordinate form by differentiating the spectral projector in the eigenbasis of A. Finally, $P(A+E) - P(A)$ is the difference of two rank-r projectors and has rank at most $2r$, while $\mathcal{L}_A(E)$ has only the two off-diagonal blocks between P and $P^\perp$ and rank at most $2r$. Hence the remainder has rank at most $4r$, so

$$\|P(A+E) - P(A) - \mathcal{L}_A(E)\|_F \leq 2\sqrt{r}\ \|P(A+E) - P(A) - \mathcal{L}_A(E)\|_{\mathrm{op}} \leq C_r \frac{\|E\|_{\mathrm{op}}^2}{\gamma^2}.$$

The final Lipschitz bound follows from $\|\mathcal{L}_A(E)\|_F \leq C_r \|E\|_{\mathrm{op}} / \gamma$ and the remainder bound, after decreasing c_0 if necessary. □

Lemma 3.2 (Gaussian operator-moment tools)**.** *Let* $S_N = N^{-1}\sum_{k=1}^N x_k x_k^\top$ *with* $x_k \overset{\text{iid}}{\sim} N(0,\Sigma)$*, and suppose* Σ *satisfies Assumption 2.2 with fixed* r*. Then, for all* $N \geq N_0$*,*

$$\mathbb{E}\,\|S_N - \Sigma\|_{\text{op}}^4 \leq C\frac{n^4}{N^2}.$$

In particular $\mathbb{E}\,\|S_N - \Sigma\|_{\text{op}}^2 \leq Cn^2/N$ *by Jensen's inequality. Moreover, for every sufficiently small fixed* $\eta > 0$ *there are constants* $c, C > 0$ *such that*

$$\mathbb{P}\{\|S_N - \Sigma\|_{\text{op}} > \eta n\} \leq Ce^{-cN}$$

whenever $N \geq N_0$*.*

Proof. The effective rank satisfies

$$r_{\text{eff}}(\Sigma) = \frac{\operatorname{tr}\Sigma}{\|\Sigma\|_{\text{op}}} = O(1)$$

under Assumption 2.2, since $\operatorname{tr}\Sigma \asymp n$ and $\|\Sigma\|_{\text{op}} \asymp n$. The Gaussian effective-rank covariance concentration inequality [13, Thm. 9] gives, with probability at least $1 - e^{-u}$,

$$\|S_N - \Sigma\|_{\text{op}} \leq C\,\|\Sigma\|_{\text{op}}\left(\sqrt{\frac{r_{\text{eff}}(\Sigma) + u}{N}} + \frac{r_{\text{eff}}(\Sigma) + u}{N}\right).$$

Since $\|\Sigma\|_{\text{op}} \asymp n$ and $r_{\text{eff}}(\Sigma) = O(1)$, the preceding tail bound can be integrated explicitly as follows. After enlarging constants, it implies the stochastic domination

$$\|S_N - \Sigma\|_{\text{op}} \leq Cn\left(\sqrt{\frac{1+U}{N}} + \frac{1+U}{N}\right)$$

for a nonnegative random variable U with exponential tail $\mathbb{P}\{U > u\} \leq e^{-u}$. Hence, for $N \geq N_0$,

$$\mathbb{E}\,\|S_N - \Sigma\|_{\text{op}}^4 \leq Cn^4\left(\frac{\mathbb{E}(1+U)^2}{N^2} + \frac{\mathbb{E}(1+U)^4}{N^4}\right) \leq C\frac{n^4}{N^2}.$$

The sub-exponential part of the concentration inequality contributes only the N^{-4} term and is dominated for $N \geq N_0$. The second-moment bound follows from $\mathbb{E}\,\|S_N - \Sigma\|_{\text{op}}^2 \leq (\mathbb{E}\,\|S_N - \Sigma\|_{\text{op}}^4)^{1/2}$. Taking $u = c_0 N$ with $c_0 > 0$ sufficiently small gives the exponential deviation bound once $N \geq N_0$. We emphasize that this concentration inequality is *dimension-free*: because $r_{\text{eff}}(\Sigma) = O(1)$, no condition of the form $N \geq C_{\log}\log n$ is required here, and the displayed exponential bound holds for all $N \geq N_0$ uniformly in n. The ambient condition $N \geq C_{\log}\log n$ in Assumption 2.4 is retained only as a convenient sufficient condition for those downstream union bounds that are stated in dimension-dependent form. □

4 Linearized spectral difficulty

Definition 4.1 (Spectral difficulty)**.** Let $\Sigma = \sum_{k=1}^n \lambda_k u_k u_k^\top$ with $\lambda_1 \geq \cdots \geq \lambda_n$ and $\lambda_r > \lambda_{r+1}$. Define

$$\mathfrak{d}_{\text{sub}}(\Sigma, r) = 2\sum_{i \leq r < j}\frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2}.$$

Remark 4.2 (Factor-of-two convention)**.** The leading factor 2 is kept *inside* the definition so that the single-window linearized projector risk is exactly $\mathfrak{d}_{\text{sub}}(\Sigma, r)/W$ (Proposition 4.3). Write

$$\mathsf{s}(\Sigma, r) := \sum_{i \le r < j} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} = \tfrac{1}{2}\, \mathfrak{d}_{\text{sub}}(\Sigma, r)$$

for the un-doubled cross-spectral sum. The adjacent drift constant of Theorem 5.12 is then $2\mathfrak{d}_{\text{sub}}/W^2 = 4\mathsf{s}/W^2$ for the full squared increment, and $\mathfrak{d}_{\text{sub}}/W^2 = 2\mathsf{s}/W^2$ for the half-norm increment $\frac{1}{2}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2$. Applications that adopt the un-doubled sum s as their working spectral-difficulty functional must carry the compensating factor 2 in these benchmarks; mismatching the convention rescales every null benchmark by a factor of two.

Proposition 4.3 (Exact linearized projector risk)**.** *Let P be the top-r projector of a fixed covariance Σ and let*

$$\mathcal{L}_\Sigma(E) = \sum_{i \le r < j} \frac{u_j u_j^\top E\, u_i u_i^\top + u_i u_i^\top E\, u_j u_j^\top}{\lambda_i - \lambda_j}.$$

Then

$$\|\mathcal{L}_\Sigma(E)\|_F^2 = 2 \sum_{i \le r < j} \frac{\left|u_j^\top E u_i\right|^2}{(\lambda_i - \lambda_j)^2}.$$

If $\widehat{\Sigma} = W^{-1} \sum_{s=1}^W r_s r_s^\top$ with i.i.d. $r_s \sim N(0, \Sigma)$, then

$$\mathbb{E}\left\|\mathcal{L}_\Sigma(\widehat{\Sigma} - \Sigma)\right\|_F^2 = \frac{\mathfrak{d}_{\text{sub}}(\Sigma, r)}{W}.$$

If

$$H = \frac{1}{W}(r_+ r_+^\top - r_- r_-^\top), \qquad r_+, r_- \overset{\text{iid}}{\sim} N(0, \Sigma),$$

then

$$\mathbb{E}\,\|\mathcal{L}_\Sigma(H)\|_F^2 = \frac{2\,\mathfrak{d}_{\text{sub}}(\Sigma, r)}{W^2}.$$

Proof. The Frobenius identity follows from the orthonormality of the matrix units $\{u_j u_i^\top,\, u_i u_j^\top : i \le r < j\}$ in the reduced-resolvent expression: each pair (i, j) contributes the coefficient $(u_j^\top E u_i)/(\lambda_i - \lambda_j)$ on both $u_j u_i^\top$ and $u_i u_j^\top$ (equal because E is symmetric), giving the displayed sum. For Gaussian data and $i \ne j$, Isserlis' formula [9, 10] together with the orthogonality $u_i^\top \Sigma u_j = \lambda_i \delta_{ij}$ gives

$$\mathbb{E}\left|u_j^\top(\widehat{\Sigma} - \Sigma)u_i\right|^2 = \frac{\lambda_i \lambda_j}{W}, \qquad \mathbb{E}\left|u_j^\top H u_i\right|^2 = \frac{2\lambda_i \lambda_j}{W^2}.$$

Summing over $i \le r < j$ and using the definition of $\mathfrak{d}_{\text{sub}}$ proves both claims. □

Corollary 4.4 (Bulk-plus-spikes first-order scale)**.** *Under Assumptions 2.2–2.3,*

$$\mathfrak{d}_{\text{sub}}(\Sigma, r) = \Theta(r).$$

Hence the first-order single-window linearized projector scale is r/W, while the first-order adjacent-window linearized scale is r/W^2.

Proof. For $i \le r < j$,

$$\lambda_i \asymp n, \qquad \lambda_j \asymp 1, \qquad \lambda_i - \lambda_j \asymp n.$$

Thus each summand is $\asymp 1/n$. There are $r(n - r)$ summands, giving $\Theta(r)$ for fixed r. □

5 Adjacent-window overlap theory

5.1 Rank-two identity and shared core

For adjacent windows, identity (1) reads

$$\widehat{\Sigma}_t - \widehat{\Sigma}_{t-1} = \frac{1}{W}(r_t r_t^\top - r_{t-W} r_{t-W}^\top),$$

so the covariance difference has rank at most two.

Under the static null $\Sigma_s \equiv \Sigma$ on $I_t \cup I_{t-1}$, define the shared core and its normalization constant by

$$C_t := \frac{1}{W} \sum_{s=t-W+1}^{t-1} r_s r_s^\top, \qquad \alpha_W := \frac{W-1}{W}.$$

Then $\mathbb{E} C_t = \alpha_W \Sigma$, not Σ. This centering is important because

$$\widehat{\Sigma}_t = C_t + \frac{1}{W} r_t r_t^\top, \qquad \widehat{\Sigma}_{t-1} = C_t + \frac{1}{W} r_{t-W} r_{t-W}^\top.$$

Let $\phi_{C_t}(v)$ denote the top-r spectral projector of $C_t + W^{-1} v v^\top$, with the same measurable tie-breaking convention if needed.

Proposition 5.1 (Conditional exchangeability)**.** *Under the static null, conditional on*

$$\mathcal{G}_t = \sigma(r_{t-W+1}, \ldots, r_{t-1}),$$

the projectors

$$\widehat{P}_t = \phi_{C_t}(r_t), \qquad \widehat{P}_{t-1} = \phi_{C_t}(r_{t-W})$$

are independent and identically distributed. Consequently,

$$\mathbb{E}[\widehat{P}_t - \widehat{P}_{t-1} \mid \mathcal{G}_t] = 0, \qquad \mathbb{E} \left\| \widehat{P}_t - \widehat{P}_{t-1} \right\|_F^2 = 2\, \mathbb{E}[V_F(\widehat{P}_t \mid \mathcal{G}_t)].$$

Proof. Conditional on $\mathcal{G}_t$, the core C_t is deterministic and r_t, r_{t-W} are i.i.d. $N(0, \Sigma)$, independent of the core; hence $\widehat{P}_t$ and $\widehat{P}_{t-1}$ are i.i.d. given $\mathcal{G}_t$. For two conditionally i.i.d. square-integrable Hilbert-space-valued variables X_1, X_2 one has $\mathbb{E}[\|X_1 - X_2\|_F^2 \mid \mathcal{G}_t] = 2\, V_F(X_1 \mid \mathcal{G}_t)$, because $\mathbb{E}[\langle X_1, X_2 \rangle_F \mid \mathcal{G}_t] = \|\mathbb{E}[X_1 \mid \mathcal{G}_t]\|_F^2$; taking the unconditional expectation gives the stated identity. □

Theorem 5.2 (Exact one-step conditional sign randomization)**.** *Under the static null, let $\mathcal{A}$ be any measurable Hilbert-space-valued functional of two rank-r projectors satisfying*

$$\mathcal{A}(Q_1, Q_2) = -\mathcal{A}(Q_2, Q_1).$$

Then, conditional on $\mathcal{G}_t$,

$$\mathcal{A}(\widehat{P}_t, \widehat{P}_{t-1}) \overset{d}{=} -\mathcal{A}(\widehat{P}_t, \widehat{P}_{t-1}).$$

This applies to oriented statistics such as $\widehat{P}_t - \widehat{P}_{t-1}$ and to linear contrasts of this difference. It does not by itself calibrate the nonnegative magnitude $\left\| \widehat{P}_t - \widehat{P}_{t-1} \right\|_F$, nor does it give the full finite-sample null law of such norms.

Proof. Swapping the two conditionally i.i.d. boundary vectors r_t and r_{t-W} preserves the conditional law and reverses the order of the two projectors. Antisymmetry of $\mathcal{A}$ negates the statistic. □

Corollary 5.3 (Exact one-pair sign calibration for oriented contrasts)**.** *Under the static null, fix any deterministic symmetric matrix A or any $\mathcal{G}_t$-measurable symmetric matrix A_t, and define*

$$T_t = \langle A_t, \widehat{P}_t - \widehat{P}_{t-1} \rangle_F.$$

Then, conditional on $\mathcal{G}_t$,

$$T_t \overset{d}{=} -T_t.$$

Consequently, if $\mathbb{P}(T_t = 0 \mid \mathcal{G}_t) = 0$, the conditional signs of T_t are exactly balanced under the static null, which yields an exact one-pair randomization calibration for this fixed adjacent pair in the sense of randomization tests [15, Ch. 15]. If there is an atom at zero, exact calibration requires an explicit zero convention, such as randomizing the sign when $T_t = 0$; treating zeros as non-rejections gives a conservative test. The measurability requirement on A_t is essential: a contrast that depends on the boundary swap (for instance, the leading eigenvector of $\widehat{P}_t - \widehat{P}_{t-1}$) co-rotates with the difference and the symmetry fails. The statement concerns oriented contrasts and does not give the full null law of the nonnegative norm $\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F$.

Proof. Apply Theorem 5.2 to the antisymmetric functional $\mathcal{A}(Q_1, Q_2) = \langle A_t, Q_1 - Q_2 \rangle_F$, conditioning on $\mathcal{G}_t$ when A_t is random; the swap leaves A_t fixed precisely because A_t is $\mathcal{G}_t$-measurable. □

Remark 5.4 (Sequential use of the sign symmetry)**.** The sign symmetry in Corollary 5.3 is a one-step conditional randomization statement. It says that, for a fixed adjacent pair (I_{t-1}, I_t) and a contrast A_t measurable with respect to the shared core, the conditional law of

$$T_t = \langle A_t, \widehat{P}_t - \widehat{P}_{t-1} \rangle_F$$

is symmetric about zero.

If the statistic is evaluated over many consecutive times, the resulting signs need not be independent. Adjacent rolling increments share observations, and their shared cores also overlap. Therefore the corollary by itself gives exact calibration for a single oriented contrast, or for collections of contrasts constructed from conditionally independent or otherwise properly blocked adjacent pairs. A simultaneous or time-series sign test over many overlapping times requires an additional dependence argument, blocking scheme, or resampling calibration.

5.2 Shared-core separation

Let

$$C_t = \sum_{k=1}^{n} \widetilde{\lambda}_k \widetilde{u}_k \widetilde{u}_k^\top, \qquad \widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_n,$$

and define the good event

$$\mathcal{E}_t^{\mathrm{core}} = \left\{ \|C_t - \alpha_W \Sigma\|_{\mathrm{op}} \leq \eta\, \alpha_W n \right\}.$$

Under the static null, $\delta_t := \lambda_r(\Sigma) - \lambda_{r+1}(\Sigma) = \delta \geq c_\delta n$. Choose $\eta > 0$ sufficiently small relative to c_λ and c_δ; for example, it is enough to take

$$\eta \leq \frac{c_\delta}{4}, \qquad \eta \leq \frac{c_\lambda}{4}.$$

On $\mathcal{E}_t^{\mathrm{core}}$, Weyl's inequality [4, Ch. III] gives

$$\widetilde{\lambda}_r - \widetilde{\lambda}_{r+1} \geq \alpha_W \delta_t - 2\eta\, \alpha_W n \geq \frac{\alpha_W \delta_t}{2},$$

where the last step uses $\eta \le c_\delta/4 \le \delta_t/(4n)$. Since $W \to \infty$, the factor $\alpha_W = (W-1)/W$ is bounded away from zero and does not affect the order of the bounds below. Define the top-r shared-core projector

$$\widetilde{P}_t^C = \sum_{i=1}^{r} \widetilde{u}_i \widetilde{u}_i^\top$$

whenever the shared-core eigengap is positive. On the complement of this event, all shared-core derivative quantities used below may be assigned arbitrary measurable values, and we take them to be zero. All identities involving $\mathcal{L}_{C_t}$, $Q(C_t, \Sigma)$, and $L_t = \mathcal{L}_{C_t}(H_t)$ are asserted only on the positive-gap event, or after multiplication by an indicator that restricts to that event. This convention is purely notational and avoids undefined expressions on bad events.

Lemma 5.5 (Shared-core separation probability)**.** *Under the static null and Assumptions 2.2–2.3, there exist constants $c, C > 0$ such that*

$$\mathbb{P}\big((\mathcal{E}_t^{\mathrm{core}})^c\big) \le Ce^{-cW}$$

whenever $W \ge W_0$ for a sufficiently large constant W_0.

Proof. Write

$$C_t = \alpha_W S_t^{\mathrm{core}}, \qquad S_t^{\mathrm{core}} = \frac{1}{W-1} \sum_{s=t-W+1}^{t-1} r_s r_s^\top,$$

so that $\|C_t - \alpha_W \Sigma\|_{\mathrm{op}} = \alpha_W \, \|S_t^{\mathrm{core}} - \Sigma\|_{\mathrm{op}}$. The core consists of $W-1$ i.i.d. $N(0,\Sigma)$ vectors, and Lemma 3.2 (with $N = W-1$) gives $\mathbb{P}\{\|S_t^{\mathrm{core}} - \Sigma\|_{\mathrm{op}} > \eta n\} \le Ce^{-cW}$ once $W \ge W_0$. Therefore $\|C_t - \alpha_W \Sigma\|_{\mathrm{op}} \le \eta\, \alpha_W n$ with probability at least $1 - Ce^{-cW}$. ☐

5.3 Conditional linearized risk

On the event $\widetilde{\lambda}_r > \widetilde{\lambda}_{r+1}$, let $\mathcal{L}_{C_t}$ be the Fréchet derivative of the top-r projector map at C_t, expressed in the *shared-core* eigenbasis. Off this event we use the zero extension fixed above:

$$\mathcal{L}_{C_t}(E) = \sum_{i \le r < j} \frac{\widetilde{u}_j \widetilde{u}_j^\top E\, \widetilde{u}_i \widetilde{u}_i^\top + \widetilde{u}_i \widetilde{u}_i^\top E\, \widetilde{u}_j \widetilde{u}_j^\top}{\widetilde{\lambda}_i - \widetilde{\lambda}_j}.$$

Define, also in the shared-core eigenbasis,

$$a_i = \widetilde{u}_i^\top \Sigma \widetilde{u}_i, \qquad b_j = \widetilde{u}_j^\top \Sigma \widetilde{u}_j, \qquad c_{ij} = \widetilde{u}_i^\top \Sigma \widetilde{u}_j.$$

Note that c_{ij} need not vanish, since the shared-core eigenvectors do not in general diagonalize Σ.

Proposition 5.6 (Exact conditional linearized risk)**.** *Under the static null, on the event $\widetilde{\lambda}_r > \widetilde{\lambda}_{r+1}$, with*

$$H_t = \frac{1}{W}\big(r_t r_t^\top - r_{t-W} r_{t-W}^\top\big),$$

one has

$$\mathbb{E}[\|\mathcal{L}_{C_t}(H_t)\|_F^2 \mid \mathcal{G}_t] = \frac{4}{W^2} \sum_{i \le r < j} \frac{a_i b_j + c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2}.$$

Proof. For one boundary term $\Delta = xx^\top/W$ with $x \sim N(0, \Sigma)$, conditional on $\mathcal{G}_t$,

$$\|\mathcal{L}_{C_t}(\Delta)\|_F^2 = \frac{2}{W^2} \sum_{i \le r < j} \frac{\left|\widetilde{u}_j^\top x\right|^2 \left|\widetilde{u}_i^\top x\right|^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2}.$$

By Isserlis' formula [9], with $g_i = \widetilde{u}_i^\top x$ and $g_j = \widetilde{u}_j^\top x$ jointly Gaussian,

$$\mathbb{E}[g_i^2 g_j^2 \mid \mathcal{G}_t] = \mathbb{E}[g_i^2]\,\mathbb{E}[g_j^2] + 2\big(\mathbb{E}[g_i g_j]\big)^2 = a_i b_j + 2c_{ij}^2.$$

Also, since $\mathbb{E}[\mathcal{L}_{C_t}(\Delta) \mid \mathcal{G}_t] = W^{-1}\mathcal{L}_{C_t}(\Sigma)$,

$$\|\mathbb{E}[\mathcal{L}_{C_t}(\Delta) \mid \mathcal{G}_t]\|_F^2 = \frac{2}{W^2} \sum_{i \le r < j} \frac{c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2}.$$

Since the two boundary terms are conditionally i.i.d., the second moment of their difference equals twice the conditional variance of one term, i.e.

$$\mathbb{E}[\|\mathcal{L}_{C_t}(H_t)\|_F^2 \mid \mathcal{G}_t] = 2\Big(\mathbb{E}[\|\mathcal{L}_{C_t}(\Delta)\|_F^2 \mid \mathcal{G}_t] - \|\mathbb{E}[\mathcal{L}_{C_t}(\Delta) \mid \mathcal{G}_t]\|_F^2\Big).$$

Substituting the two displays gives

$$\frac{2 \cdot 2}{W^2} \sum_{i \le r < j} \frac{(a_i b_j + 2c_{ij}^2) - c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2} = \frac{4}{W^2} \sum_{i \le r < j} \frac{a_i b_j + c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2}.$$

□

For later use, on the positive-gap event define the shared-core variance functional

$$Q(C_t, \Sigma) := 4 \sum_{i \le r < j} \frac{a_i b_j + c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2},$$

and set $Q(C_t, \Sigma) = 0$ off that event, with the measurable eigenbasis convention fixed above. Thus, on the positive-gap event,

$$\mathbb{E}[\|\mathcal{L}_{C_t}(H_t)\|_F^2 \mid \mathcal{G}_t] = \frac{Q(C_t, \Sigma)}{W^2}.$$

More generally, if a deterministic symmetric matrix A has an isolated top-r cluster, define

$$Q(A, \Sigma) := \mathbb{E}\left\|\mathcal{L}_A(xx^\top - yy^\top)\right\|_F^2, \qquad x, y \overset{\text{iid}}{\sim} N(0, \Sigma),$$

where $\mathcal{L}_A$ is the Fréchet derivative across that isolated gap. When $A = C_t$ on the shared-core gap event, this basis-free definition agrees with the coordinate formula above.

Lemma 5.7 (Shared-core linearized statistic is order r)**.** *Assume the static null and work on $\mathcal{E}_t^{\text{core}}$. If $W \ge W_0$, so that α_W is bounded away from zero, then*

$$cr \le 4 \sum_{i \le r < j} \frac{a_i b_j + c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2} \le Cr.$$

Consequently, for W large enough,

$$\mathbb{E}[\|\mathcal{L}_{C_t}(H_t)\|_F^2\, \mathbf{1}_{\mathcal{E}_t^{\text{core}}}] \asymp \frac{r}{W^2}.$$

Proof. On $\mathcal{E}_t^{\mathrm{core}}$,

$$\widetilde{\lambda}_r - \widetilde{\lambda}_{r+1} \geq \alpha_W \delta_t - 2\eta\, \alpha_W n \geq \frac{\alpha_W \delta_t}{2} \asymp n.$$

Hence for $i \leq r < j$, $\left|\widetilde{\lambda}_i - \widetilde{\lambda}_j\right| \gtrsim n$. Also

$$\widetilde{\lambda}_i \leq \alpha_W \lambda_i(\Sigma) + \|C_t - \alpha_W \Sigma\|_{\mathrm{op}} \lesssim n \qquad (i \leq r),$$

and $C_t \succeq 0$, so $0 \leq \widetilde{\lambda}_j \leq \widetilde{\lambda}_i \lesssim n$ for $i \leq r < j$. Thus $(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2 \asymp n^2$ for $i \leq r < j$. By Cauchy–Schwarz, $c_{ij}^2 \leq a_i b_j$, so

$$\sum_{i \leq r < j} (a_i b_j + c_{ij}^2) \leq 2 \sum_{i \leq r < j} a_i b_j = 2\Big(\sum_{i \leq r} a_i\Big)\Big(\sum_{j > r} b_j\Big).$$

Since Σ has r spikes of size $O(n)$ and bulk trace $O(n)$,

$$\sum_{i \leq r} a_i \leq r\, \|\Sigma\|_{\mathrm{op}} \lesssim rn, \qquad \sum_{j > r} b_j \leq \operatorname{tr} \Sigma \lesssim n.$$

Dividing by denominators of order n^2 gives the upper bound Cr.

For the lower bound, on $\mathcal{E}_t^{\mathrm{core}}$,

$$\widetilde{\lambda}_i \geq \alpha_W \lambda_i(\Sigma) - \|C_t - \alpha_W \Sigma\|_{\mathrm{op}} \gtrsim n \qquad (i \leq r).$$

Moreover, since $\widetilde{u}_i$ is a unit eigenvector of C_t with eigenvalue $\widetilde{\lambda}_i$,

$$\alpha_W a_i = \widetilde{u}_i^\top (\alpha_W \Sigma) \widetilde{u}_i \geq \widetilde{u}_i^\top C_t \widetilde{u}_i - \|C_t - \alpha_W \Sigma\|_{\mathrm{op}} = \widetilde{\lambda}_i - \|C_t - \alpha_W \Sigma\|_{\mathrm{op}}.$$

Choosing η sufficiently small gives $a_i \gtrsim n$ for $i \leq r$. Also, by the Ky Fan variational principle [4, Ch. III],

$$\sum_{j > r} b_j = \operatorname{tr}\big((I - \widetilde{P}_t^C)\Sigma\big) \geq \sum_{k=r+1}^{n} \lambda_k(\Sigma) \gtrsim n.$$

Therefore, dropping $c_{ij}^2 \geq 0$ and bounding the denominators above by Cn^2,

$$4 \sum_{i \leq r < j} \frac{a_i b_j + c_{ij}^2}{(\widetilde{\lambda}_i - \widetilde{\lambda}_j)^2} \geq \frac{4}{Cn^2} \Big(\sum_{i \leq r} a_i\Big)\Big(\sum_{j > r} b_j\Big) \gtrsim \frac{(rn)(n)}{n^2} \asymp r.$$

This is the lower bound cr. Proposition 5.6 and Lemma 5.5 then give the expectation statement. □

Lemma 5.8 (Boundary tail at an arbitrary fixed threshold)**.** *Under the static null and Assumption 2.2, if $r_\pm \sim N(0, \Sigma)$ independently, then for every fixed $\kappa > 0$ there exist constants $c_\kappa, C_\kappa, W_\kappa > 0$, depending only on κ and on the spectral constants, such that*

$$\mathbb{P}\{\|r_\pm\|_2^2 > \kappa n W\} \leq C_\kappa e^{-c_\kappa W}$$

for all $W \geq W_\kappa$.

Proof. Write $\|r_\pm\|_2^2 = \sum_{k=1}^n \lambda_k Z_k^2$ with independent standard normal Z_k. Assumption 2.2 gives

$$\operatorname{tr} \Sigma \leq C_0 n, \qquad \operatorname{tr}(\Sigma^2) \leq C_1 n^2, \qquad \|\Sigma\|_{\mathrm{op}} \leq C_2 n.$$

For the fixed threshold $\kappa > 0$, choose $W_\kappa \geq 2C_0/\kappa$, so that for $W \geq W_\kappa$,

$$\kappa n W - \operatorname{tr}\Sigma \geq \frac{\kappa}{2} n W.$$

The Bernstein-type bound for weighted chi-square variables [14, 8, 20] gives, for all $u > 0$,

$$\mathbb{P}\left\{\|r_\pm\|_2^2 - \operatorname{tr}\Sigma > u\right\} \leq \exp\left[-c\min\left(\frac{u^2}{\operatorname{tr}(\Sigma^2)}, \frac{u}{\|\Sigma\|_{\mathrm{op}}}\right)\right].$$

Substituting $u = (\kappa/2)nW$ yields, since $u^2/\operatorname{tr}(\Sigma^2) \gtrsim W^2$ and $u/\|\Sigma\|_{\mathrm{op}} \gtrsim W$,

$$\mathbb{P}\{\|r_\pm\|_2^2 > \kappa n W\} \leq \exp\left[-c\min\left\{c'_\kappa W^2,\, c'_\kappa W\right\}\right] \leq C_\kappa e^{-c_\kappa W},$$

after adjusting constants. □

Lemma 5.9 (Fourth moment of the shared-core linearization)**.** *Assume the static null and work on $\mathcal{E}_t^{\mathrm{core}}$. Let*

$$L_t = \mathcal{L}_{C_t}\left(\frac{1}{W}(r_t r_t^\top - r_{t-W} r_{t-W}^\top)\right).$$

Then

$$\mathbb{E}[\|L_t\|_F^4 \mid \mathcal{G}_t] \leq C\frac{r^2}{W^4}.$$

Proof. Conditional on $\mathcal{G}_t$, the core C_t and hence the linear map $\mathcal{L}_{C_t}$ are fixed, and r_t, r_{t-W} are i.i.d. $N(0,\Sigma)$. Writing

$$r_t r_t^\top - r_{t-W} r_{t-W}^\top = (r_t r_t^\top - \Sigma) - (r_{t-W} r_{t-W}^\top - \Sigma),$$

each centered term lies in the second homogeneous Gaussian (Wiener) chaos of the corresponding boundary vector, with no zeroth- or first-order component; applying the fixed linear map $\mathcal{L}_{C_t}$ and subtracting two independent copies keeps L_t in the second chaos. Hypercontractivity for Gaussian chaoses [16, 10] therefore gives

$$\mathbb{E}[\|L_t\|_F^4 \mid \mathcal{G}_t] \leq C\left(\mathbb{E}[\|L_t\|_F^2 \mid \mathcal{G}_t]\right)^2.$$

By Proposition 5.6 and the upper bound in Lemma 5.7, the conditional second moment is at most Cr/W^2 on $\mathcal{E}_t^{\mathrm{core}}$. Squaring gives the claim. □

Lemma 5.10 (Shared-core derivative replacement on the good event)**.** *Assume the static null, Assumptions 2.2–2.3, fixed r, $W \to \infty$, and $W \geq C_{\log}\log n$. Then*

$$\mathbb{E}\left[\left|Q(C_t,\Sigma) - \frac{2}{\alpha_W^2}\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)\right|\mathbf{1}_{\mathcal{E}_t^{\mathrm{core}}}\right] \leq C\frac{\mathbb{E}\,\|C_t - \alpha_W\Sigma\|_{\mathrm{op}}}{n} = O(W^{-1/2}) = o(1).$$

Proof. Write

$$A_0 = \alpha_W\Sigma, \qquad A = C_t, \qquad E_A = A - A_0.$$

By Lemma 3.2, applied to

$$S_t^{\mathrm{core}} := \frac{1}{W-1}\sum_{s=t-W+1}^{t-1} r_s r_s^\top, \qquad C_t = \alpha_W S_t^{\mathrm{core}},$$

we have

$$\mathbb{E}\left\|E_A\right\|_{\mathrm{op}} = \alpha_W \mathbb{E}\left\|S_t^{\mathrm{core}} - \Sigma\right\|_{\mathrm{op}} \le C\frac{n}{\sqrt{W}}.$$

It remains to justify the Lipschitz dependence of the variance functional $Q(A,\Sigma)$ on the base matrix A. We do this in a basis-free way, avoiding any dependence on the possibly unstable eigenvectors inside the lower spectral cluster.

Let Γ be a positively oriented contour enclosing the top-r spectral cluster of $A_0 = \alpha_W\Sigma$ and no lower eigenvalues, chosen so that

$$\mathrm{length}(\Gamma) \le Cn, \qquad \mathrm{dist}(\Gamma, \mathrm{spec}(A_0)) \ge cn.$$

For η sufficiently small, the same contour also separates the top-r cluster of every A satisfying

$$\left\|A - A_0\right\|_{\mathrm{op}} \le \eta\alpha_W n,$$

and for such A,

$$\sup_{z\in\Gamma}\left\|(zI - A)^{-1}\right\|_{\mathrm{op}} \le \frac{C}{n}.$$

The Fréchet derivative of the spectral projector admits the Riesz formula

$$\mathcal{L}_A(B) = \frac{1}{2\pi i}\int_\Gamma (zI - A)^{-1}B(zI - A)^{-1}\,dz.$$

Hence, for any symmetric B,

$$\left\|\mathcal{L}_A(B)\right\|_F \le C_r\frac{\left\|B\right\|_{\mathrm{op}}}{n},$$

because the integral has operator norm at most $C\left\|B\right\|_{\mathrm{op}}/n$ and its image lies in the off-diagonal blocks between a rank-r subspace and its orthogonal complement, hence has rank at most $2r$.

Similarly, by the resolvent identity

$$(zI - A)^{-1} - (zI - A_0)^{-1} = (zI - A)^{-1}(A - A_0)(zI - A_0)^{-1},$$

we obtain

$$\left\|(\mathcal{L}_A - \mathcal{L}_{A_0})(B)\right\|_F \le C_r\frac{\left\|A - A_0\right\|_{\mathrm{op}}\left\|B\right\|_{\mathrm{op}}}{n^2}.$$

Indeed, the integrand difference contains one resolvent difference and two resolvent factors in total, giving the bound $C\left\|A - A_0\right\|_{\mathrm{op}}\left\|B\right\|_{\mathrm{op}}/n^3$ pointwise on Γ, and the contour has length $O(n)$.

Now let

$$Z = xx^\top - yy^\top, \qquad x, y \overset{\mathrm{iid}}{\sim} N(0,\Sigma).$$

By the basis-free definition of $Q(A,\Sigma)$, the identity $Q(A,\Sigma) = \mathbb{E}\left\|\mathcal{L}_A(Z)\right\|_F^2$ is used only for base matrices whose top-r cluster is isolated by the contour Γ. Therefore, on the above neighborhood of A_0,

$$\begin{aligned}\left|Q(A,\Sigma) - Q(A_0,\Sigma)\right| &\le \mathbb{E}\left[\left(\left\|\mathcal{L}_A(Z)\right\|_F + \left\|\mathcal{L}_{A_0}(Z)\right\|_F\right)\left\|(\mathcal{L}_A - \mathcal{L}_{A_0})(Z)\right\|_F\right] \\ &\le C_r\frac{\left\|A - A_0\right\|_{\mathrm{op}}}{n^3}\mathbb{E}\left\|Z\right\|_{\mathrm{op}}^2.\end{aligned}$$

Since

$$\left\|Z\right\|_{\mathrm{op}} \le \left\|x\right\|_2^2 + \left\|y\right\|_2^2$$

and, under Assumption 2.2,

$$\mathbb{E}\left\|x\right\|_2^4 = (\mathrm{tr}\,\Sigma)^2 + 2\,\mathrm{tr}(\Sigma^2) \le Cn^2,$$

we have

$$\mathbb{E}\,\|Z\|_{\mathrm{op}}^2 \le Cn^2.$$

Consequently,

$$|Q(A,\Sigma) - Q(A_0,\Sigma)| \le C\frac{\|A - A_0\|_{\mathrm{op}}}{n}.$$

It remains to evaluate $Q(A_0,\Sigma)$. Since $A_0 = \alpha_W \Sigma$ has the same eigenvectors as Σ and eigenvalues $\alpha_W \lambda_k$, the coordinate formula of Proposition 5.6 gives

$$Q(A_0,\Sigma) = 4\sum_{i\le r<j} \frac{\lambda_i \lambda_j}{\alpha_W^2(\lambda_i - \lambda_j)^2} = \frac{2}{\alpha_W^2}\mathfrak{d}_{\mathrm{sub}}(\Sigma, r).$$

Combining this identity with the preceding Lipschitz bound and integrating over $\mathcal{E}_t^{\mathrm{core}}$ yields

$$\mathbb{E}\left[\left|Q(C_t,\Sigma) - \frac{2}{\alpha_W^2}\mathfrak{d}_{\mathrm{sub}}(\Sigma,r)\right| \mathbf{1}_{\mathcal{E}_t^{\mathrm{core}}}\right] \le C\frac{\mathbb{E}\,\|C_t - \alpha_W\Sigma\|_{\mathrm{op}}}{n} \le CW^{-1/2}.$$

This proves the claim. □

Theorem 5.11 (Adjacent-window null-increment benchmark under the static null)**.** *Assume the static null on $I_t \cup I_{t-1}$ and Assumptions 2.2–2.3, with r fixed. There exist constants $c, C, C_{\log} > 0$ and $W_0 < \infty$, with $C_{\log}$ and W_0 depending on the fixed rank r and the spectral constants, such that if*

$$W \ge C_{\log}\log n \qquad \text{and} \qquad W \ge W_0,$$

then

$$c\,\frac{r}{W^2} \le \mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 \le C\,\frac{r}{W^2} + Cr\,e^{-cW}.$$

Proof. Let $\gamma_t^C = \widetilde{\lambda}_r - \widetilde{\lambda}_{r+1}$ be the top-r shared-core eigengap when positive, and set

$$\Delta_t^+ = \frac{1}{W} r_t r_t^\top, \qquad \Delta_t^- = \frac{1}{W} r_{t-W} r_{t-W}^\top, \qquad H_t = \Delta_t^+ - \Delta_t^-.$$

Choose a small constant $\rho \in (0, c_0]$, with c_0 as in Lemma 3.1, and define

$$\mathcal{H}_t = \mathcal{E}_t^{\mathrm{core}} \cap \left\{ \left\|\Delta_t^+\right\|_{\mathrm{op}} \vee \left\|\Delta_t^-\right\|_{\mathrm{op}} \le \rho\gamma_t^C \right\}.$$

On $\mathcal{E}_t^{\mathrm{core}}$ there is a deterministic constant $c_\gamma > 0$ with $\gamma_t^C \ge c_\gamma n$. Choose the fixed threshold $\kappa = \rho c_\gamma/2$. Then

$$\frac{\|r_t\|_2^2}{W} \vee \frac{\|r_{t-W}\|_2^2}{W} \le \kappa n$$

implies $\left\|\Delta_t^+\right\|_{\mathrm{op}} \vee \left\|\Delta_t^-\right\|_{\mathrm{op}} \le \rho\gamma_t^C$, hence implies $\mathcal{H}_t$ on $\mathcal{E}_t^{\mathrm{core}}$.

The complement of $\mathcal{H}_t$ is contained in the union of $(\mathcal{E}_t^{\mathrm{core}})^c$ and the event that at least one boundary vector exceeds the fixed threshold $\kappa n W$. Lemma 5.5 controls the first event and Lemma 5.8, applied with this κ, controls the second. Therefore $\mathbb{P}(\mathcal{H}_t^c) \le Ce^{-cW}$.

On $\mathcal{H}_t$, Lemma 3.1 applies at C_t to both Δ_t^+ and Δ_t^-, so

$$\widehat{P}_t = P(C_t) + \mathcal{L}_{C_t}(\Delta_t^+) + R_t^+, \qquad \widehat{P}_{t-1} = P(C_t) + \mathcal{L}_{C_t}(\Delta_t^-) + R_t^-,$$

with $\left\|R_t^{\pm}\right\|_F \le C_r \left\|\Delta_t^{\pm}\right\|_{\mathrm{op}}^2/(\gamma_t^C)^2$. Subtracting,

$$\widehat{P}_t - \widehat{P}_{t-1} = \mathcal{L}_{C_t}(H_t) + R_t^{\Delta}, \qquad R_t^{\Delta} = R_t^+ - R_t^-.$$

On $\mathcal{H}_t$, since $\gamma_t^C \asymp n$,

$$\mathbb{E}[\left\|R_t^{\Delta}\right\|_F^2 \mathbf{1}_{\mathcal{H}_t}] \le \frac{C_r}{W^4}.$$

Here we used $\mathbb{E}\|r_{\pm}\|_2^8 \le Cn^4$, which follows directly from the bulk-plus-spikes scaling: if $r_{\pm} = \Sigma^{1/2}g$ with $g \sim N(0, I_n)$, then $\|r_{\pm}\|_2^2 = \sum_k \lambda_k g_k^2$, and the fourth moment of this quadratic form is a universal polynomial in $\operatorname{tr}\Sigma, \operatorname{tr}(\Sigma^2), \operatorname{tr}(\Sigma^3), \operatorname{tr}(\Sigma^4)$. Under Assumption 2.2 and fixed spike rank, $\operatorname{tr}(\Sigma^{\ell}) \le C_{\ell} n^{\ell}$ for $1 \le \ell \le 4$, hence the stated eighth-moment bound.

Let $L_t = \mathcal{L}_{C_t}(H_t)$. By Lemma 5.7, $\mathbb{E}[\|L_t\|_F^2 \mathbf{1}_{\mathcal{E}_t^{\mathrm{core}}}] \asymp r/W^2$, and Lemma 5.9 gives the matching fourth-moment bound. Therefore, by Cauchy–Schwarz and $\mathbb{P}(\mathcal{H}_t^c) \le Ce^{-cW}$,

$$\mathbb{E}[\|L_t\|_F^2 \mathbf{1}_{\mathcal{E}_t^{\mathrm{core}} \cap \mathcal{H}_t^c}] \le \sqrt{\mathbb{E}[\|L_t\|_F^4 \mathbf{1}_{\mathcal{E}_t^{\mathrm{core}}}]}\sqrt{\mathbb{P}(\mathcal{H}_t^c)} \le C\frac{r}{W^2}e^{-cW/2},$$

hence $\mathbb{E}[\|L_t\|_F^2 \mathbf{1}_{\mathcal{H}_t}] \asymp r/W^2$.

For the upper bound, use $\left\|L_t + R_t^{\Delta}\right\|_F^2 \le 2\|L_t\|_F^2 + 2\left\|R_t^{\Delta}\right\|_F^2$ on $\mathcal{H}_t$ and $\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 \le 2r$ on $\mathcal{H}_t^c$, giving

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 \le C\frac{r}{W^2} + Cr\,e^{-cW}.$$

For the lower bound, $\left\|L_t + R_t^{\Delta}\right\|_F^2 \ge \|L_t\|_F^2 - 2\left|\langle L_t, R_t^{\Delta}\rangle\right|$, and by Cauchy–Schwarz

$$2\,\mathbb{E}[\left|\langle L_t, R_t^{\Delta}\rangle\right| \mathbf{1}_{\mathcal{H}_t}] \le 2\sqrt{\mathbb{E}[\|L_t\|_F^2 \mathbf{1}_{\mathcal{H}_t}]}\sqrt{\mathbb{E}[\left\|R_t^{\Delta}\right\|_F^2 \mathbf{1}_{\mathcal{H}_t}]} \le C\frac{\sqrt{r}}{W^3}.$$

Since r is fixed, this is of lower order than r/W^2 for large W. Dropping the nonnegative contribution off $\mathcal{H}_t$ and enlarging W_0 if necessary gives $\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 \ge c\,r/W^2$. □

Theorem 5.12 (Constant-level adjacent-window drift expansion in the strongly spiked Gaussian regime)**.** *Assume the static Gaussian null on $I_t \cup I_{t-1}$, Assumptions 2.2–2.3, fixed r, $W \to \infty$, and $W \ge C_{\log} \log n$. Then*

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{2}{\alpha_W^2}\frac{\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{W^2} + o(W^{-2}) = \frac{2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{(W-1)^2} + o(W^{-2}).$$

Since $\alpha_W = (W-1)/W$ and $\mathfrak{d}_{\mathrm{sub}}(\Sigma, r) = O(1)$ for fixed r, this also implies

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{W^2} + o(W^{-2}).$$

The first display is the finite-W leading approximation; replacing α_W^{-2} by 1 changes the scaled benchmark by a relative $O(W^{-1})$ amount. The $o(W^{-2})$ term is meant uniformly over covariance sequences satisfying Assumptions 2.2–2.3 with the same fixed spectral constants; the leading constant itself is sequence-dependent through $\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)$, and the uniformity is inherited from the shared-core replacement estimate in Lemma 5.10.

Proof. Use the event $\mathcal{H}_t$ and decomposition from the proof of Theorem 5.11. On $\mathcal{H}_t$,

$$\widehat{P}_t - \widehat{P}_{t-1} = L_t + R_t^{\Delta}, \qquad L_t = \mathcal{L}_{C_t}(H_t),$$

with $\mathbb{E}[\left\|R_t^{\Delta}\right\|_F^2 \mathbf{1}_{\mathcal{H}_t}] \le CW^{-4}$ and $\mathbb{P}(\mathcal{H}_t^c) \le Ce^{-cW}$; on the complement $\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 \le 2r$.

By Proposition 5.6, $\mathbb{E}[\|L_t\|_F^2 \mid \mathcal{G}_t] = Q(C_t, \Sigma)/W^2$ on the shared-core gap event. Lemma 5.10, together with Lemma 5.5, gives

$$\mathbb{E}[\|L_t\|_F^2 \, \mathbf{1}_{\mathcal{E}_t^{\mathrm{core}}}] = \frac{1}{W^2}\left(\frac{2}{\alpha_W^2}\mathfrak{d}_{\mathrm{sub}}(\Sigma, r) + o(1)\right).$$

The $o(1)$ term in this display is uniform over the fixed spectral class because Lemma 5.10 gives the replacement error $O(W^{-1/2})$ with constants depending only on r and the spectral constants. The loss from further restricting to $\mathcal{H}_t$ is exponentially small: by Lemma 5.9 and Cauchy–Schwarz,

$$\mathbb{E}[\|L_t\|_F^2 \, \mathbf{1}_{\mathcal{H}_t^c \cap \mathcal{E}_t^{\mathrm{core}}}] \le C\frac{r}{W^2}e^{-cW/2}.$$

Finally,

$$\left\|L_t + R_t^{\Delta}\right\|_F^2 = \|L_t\|_F^2 + 2\langle L_t, R_t^{\Delta}\rangle_F + \left\|R_t^{\Delta}\right\|_F^2,$$

where the remainder square is $O(W^{-4})$ and, by Cauchy–Schwarz,

$$\left|\mathbb{E}[\langle L_t, R_t^{\Delta}\rangle_F \mathbf{1}_{\mathcal{H}_t}]\right| \le \sqrt{\mathbb{E}[\|L_t\|_F^2 \, \mathbf{1}_{\mathcal{H}_t}]}\sqrt{\mathbb{E}[\left\|R_t^{\Delta}\right\|_F^2 \mathbf{1}_{\mathcal{H}_t}]} = O(W^{-3}),$$

since r is fixed. Both terms, and the exponentially small off-event contribution, are $o(W^{-2})$. This proves the stated expansion. □

Remark 5.13 (Explicit leading constant and convention check)**.** In terms of the un-doubled cross-spectral sum $\mathsf{s}(\Sigma, r) = \frac{1}{2}\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)$ of Remark 4.2, Theorem 5.12 reads

$$\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{4}{\alpha_W^2 W^2}\sum_{i \le r < j}\frac{\lambda_i\lambda_j}{(\lambda_i - \lambda_j)^2} + o(W^{-2}) = \frac{4\mathsf{s}(\Sigma, r)}{(W-1)^2} + o(W^{-2}),$$

and

$$\mathbb{E}\,\tfrac{1}{2}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2 = \frac{\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)}{(W-1)^2} + o(W^{-2}) = \frac{2\mathsf{s}(\Sigma, r)}{(W-1)^2} + o(W^{-2}).$$

The limiting constant coincides with the population boundary risk $\mathbb{E}\|\mathcal{L}_\Sigma(H)\|_F^2 = 2\mathfrak{d}_{\mathrm{sub}}/W^2$ of Proposition 4.3, as it must: the random-core gap correction $\alpha_W^{-2} \to 1$ recovers the population linearization. At finite W, however, the more operative leading benchmark is $2\mathfrak{d}_{\mathrm{sub}}/(W-1)^2$, not $2\mathfrak{d}_{\mathrm{sub}}/W^2$. Consistency with the conditional functional of Proposition 5.6 also holds by the law of total expectation, since at the population spectrum $Q(\alpha_W\Sigma, \Sigma) = 2\mathfrak{d}_{\mathrm{sub}}/\alpha_W^2$, which divided by W^2 gives the displayed finite-W leading term.

6 Numerical check of the leading constant

The leading constant in Theorem 5.12 is easy to check by simulation because the adjacent windows differ only through two boundary observations after conditioning on the shared core. Figure 1 uses

$$\Sigma = \mathrm{diag}(n+1, \ldots, n+1, 1, \ldots, 1), \qquad n = 40, \quad r = 3,$$

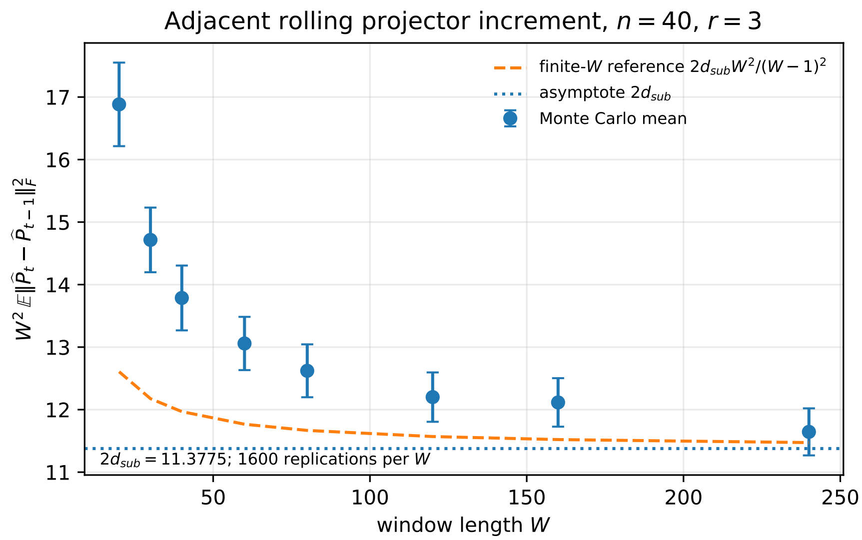

Figure 1: Monte Carlo check of the leading constant in Theorem 5.12. Error bars are approximate 95% Monte Carlo intervals for the scaled mean; the finite-W reference includes the α_W^{-2} correction, while the horizontal asymptote is $2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)$.

with r spikes equal to $n+1$ and an $O(1)$ bulk. For each value of W, the experiment draws a shared core of $W-1$ observations and two independent boundary observations, forms the two adjacent sample covariances, computes their top-r projectors, and estimates $W^2\mathbb{E}\left\|\widehat{P}_t - \widehat{P}_{t-1}\right\|_F^2$ by Monte Carlo. The limiting horizontal reference is $2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)$, equal to 11.3775 for this spectrum. The more accurate finite-W first-order reference for the scaled mean is

$$\frac{2}{\alpha_W^2}\mathfrak{d}_{\mathrm{sub}}(\Sigma, r) = 2\mathfrak{d}_{\mathrm{sub}}(\Sigma, r)\frac{W^2}{(W-1)^2}.$$

The theorem does not predict the sign of the remaining finite-W bias. In moderate-window simulations, the Monte Carlo mean may sit above even this α_W^{-2} reference. That behavior should be attributed to unresolved higher-order perturbation and random-core replacement terms, not primarily to the deterministic core-normalization factor.

7 Boundary-driven sharpness

The adjacent-window lower bound requires a precise target. It is not a lower bound for arbitrary instantaneous drift, and it is not a claim of nontrivial recovery below the null fluctuation floor. Under a known static null the population instantaneous drift $P_t - P_{t-1}$ is zero and the estimator $\widehat{D} \equiv 0$ has zero risk. Under a fully nonstatic class, instantaneous drift may be harder than r/W^2 because only one new observation is available at each endpoint. We therefore state the sharp r/W^2 lower bound for the overlap-driven rolling target

$$D_t^{\mathrm{roll}} := P(\bar{\Sigma}_t) - P(\bar{\Sigma}_{t-1}),$$

where $\bar{\Sigma}_t$ denotes the population window average. Equivalently, the same construction applies to local path models in which adjacent rolling drift is generated by the boundary innovation; it should not be read as a statement about an unrestricted target $P_t - P_{t-1}$.

Proposition 7.1 (Boundary-driven rolling-drift lower bound)**.** *Fix r and suppose $n \geq 2r+1$. There is a finite family of product Gaussian adjacent-window experiments $\{\mathbb{P}_\Theta : \Theta \in \mathcal{V}\}$ with the following explicit form. In experiment $\mathbb{P}_\Theta$, the $W-1$ shared-core observations are independent $N(0, \Sigma_0)$ vectors, the entering boundary observation is $N(0, \Sigma_\Theta)$, and the exiting boundary observation is $N(0, \Sigma_{-\Theta})$, all mutually independent. The covariances are*

$$\Sigma_0 = I_n + nP_0, \qquad \Sigma_{\pm\Theta} = I_n + nP_{\pm\Theta},$$

where P_0 is the coordinate rank-r projector and $P_{\pm\Theta}$ are the graph projectors constructed in the proof. The boundary covariances satisfy the bulk-plus-spikes and eigengap conditions uniformly. The corresponding population window averages

$$\bar{\Sigma}_t(\Theta) = \Sigma_0 + \frac{1}{W}(\Sigma_\Theta - \Sigma_0), \qquad \bar{\Sigma}_{t-1}(\Theta) = \Sigma_0 + \frac{1}{W}(\Sigma_{-\Theta} - \Sigma_0)$$

have a uniformly separated top-r spectral cluster; their lower cluster is bounded by $C(1 + n/W)$. In particular, if one additionally imposes $W \gtrsim n$, then the population window averages also satisfy the full bulk-plus-spikes condition uniformly. There is a threshold $W_0 = W_0(r, \text{spectral constants})$, not depending on n, such that for all $W \geq W_0$, uniformly over $n \geq 2r+1$,

$$\inf_{\widehat{D}} \sup_{\Theta \in \mathcal{V}} \mathbb{E}_\Theta \left\| \widehat{D} - D_t^{\mathrm{roll}}(\Theta) \right\|_F^2 \geq c \frac{r}{W^2}.$$

Proof. Let P_0 be the coordinate rank-r projector,

$$P_0 = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Sigma_0 = I_n + nP_0.$$

For a matrix $\Theta \in \mathbb{R}^{(n-r)\times r}$ with sufficiently small operator norm, define

$$U_\Theta = \begin{pmatrix} I_r \\ \Theta \end{pmatrix} (I_r + \Theta^\top \Theta)^{-1/2}, \qquad P_\Theta = U_\Theta U_\Theta^\top, \qquad \Sigma_\Theta = I_n + nP_\Theta.$$

Every Σ_Θ in this local family has r eigenvalues equal to $n+1$ and $n-r$ eigenvalues equal to 1, so the boundary covariances have uniform spike size, bounded bulk, and eigengap.

We consider the adjacent-window experiment in which the $W-1$ shared-core observations have covariance Σ_0, the entering boundary observation has covariance Σ_Θ, and the exiting boundary observation has covariance $\Sigma_{-\Theta}$. Thus

$$\bar{\Sigma}_t(\Theta) = \Sigma_0 + \frac{1}{W}(\Sigma_\Theta - \Sigma_0), \qquad \bar{\Sigma}_{t-1}(\Theta) = \Sigma_0 + \frac{1}{W}(\Sigma_{-\Theta} - \Sigma_0),$$

and the target is

$$D_t^{\mathrm{roll}}(\Theta) = P(\bar{\Sigma}_t(\Theta)) - P(\bar{\Sigma}_{t-1}(\Theta)).$$

The window averages retain a uniform top-r eigengap but need not retain an $O(1)$ lower cluster unless W is at least of order n. Indeed,

$$\bar{\Sigma}_t(\Theta) = I_n + n\left(\left(1 - \frac{1}{W}\right)P_0 + \frac{1}{W}P_\Theta\right),$$

and Weyl's inequality gives, uniformly over the local chart,

$$\lambda_r(\bar{\Sigma}_t(\Theta)) \geq n + 1 - \frac{2n}{W}, \qquad \lambda_{r+1}(\bar{\Sigma}_t(\Theta)) \leq 1 + \frac{2n}{W}.$$

The same bounds hold for $\bar{\Sigma}_{t-1}(\Theta)$. Hence, for all sufficiently large W, both averages have a top-r eigengap at least a fixed multiple of n, while their lower cluster is bounded by $C(1+n/W)$; in the subregime $W \gtrsim n$ this is the original $O(1)$ bulk condition.

We first record the local geometry of this target. Let

$$K(\Theta) = \begin{pmatrix} 0 & \Theta^\top \\ \Theta & 0 \end{pmatrix}.$$

The graph-projector formula gives, uniformly for $\|\Theta\|_{\mathrm{op}}, \|\Theta'\|_{\mathrm{op}} \le \varepsilon$,

$$P_\Theta - P_0 = K(\Theta) + E(\Theta), \qquad P_{-\Theta} - P_0 = -K(\Theta) + E(-\Theta),$$

with

$$\|E(\Theta) - E(\Theta')\|_F \le C\varepsilon \|\Theta - \Theta'\|_F .$$

Indeed, this follows by direct block computation from $U_\Theta = (I_r, \Theta)^\top (I_r + \Theta^\top\Theta)^{-1/2}$. Writing $M = (I_r + \Theta^\top\Theta)^{-1/2}$, one has

$$P_\Theta = U_\Theta U_\Theta^\top = \begin{pmatrix} M^2 & M^2\Theta^\top \\ \Theta M^2 & \Theta M^2 \Theta^\top \end{pmatrix} = \begin{pmatrix} (I_r + \Theta^\top\Theta)^{-1} & (I_r + \Theta^\top\Theta)^{-1}\Theta^\top \\ \Theta(I_r + \Theta^\top\Theta)^{-1} & \Theta(I_r + \Theta^\top\Theta)^{-1}\Theta^\top \end{pmatrix},$$

so the blocks of $E(\Theta) = P_\Theta - P_0 - K(\Theta)$ are

$$\begin{aligned} E(\Theta)_{11} &= (I_r + \Theta^\top\Theta)^{-1} - I_r = -\Theta^\top\Theta + O(\|\Theta\|^4), \\ E(\Theta)_{22} &= \Theta(I_r + \Theta^\top\Theta)^{-1}\Theta^\top = \Theta\Theta^\top + O(\|\Theta\|^4), \\ E(\Theta)_{12} &= [(I_r + \Theta^\top\Theta)^{-1} - I_r]\Theta^\top = -\Theta^\top\Theta\Theta^\top + O(\|\Theta\|^5), \end{aligned}$$

and $E(\Theta)_{21} = E(\Theta)_{12}^\top$. Thus the diagonal blocks are quadratic in Θ and the off-diagonal blocks are cubic; in either case $E(\Theta) = O(\|\Theta\|^2)$ and the map $\Theta \mapsto E(\Theta)$ has $C\varepsilon$-Lipschitz constant on $\|\Theta\|_{\mathrm{op}} \le \varepsilon$, which is all that is used below. (The off-diagonal block is governed by $(I_r + \Theta^\top\Theta)^{-1}$, not by $(I_r + \Theta^\top\Theta)^{-1/2}$, and its leading correction to $\Theta^\top$ is cubic rather than quadratic.) Since r is fixed, the constants in this Lipschitz bound are uniform in n on the local chart.

We now make the projector Taylor expansion at Σ_0 quantitative. The top-r eigengap of $\Sigma_0 = I_n + nP_0$ equals n, so Lemma 3.1 gives, for symmetric E, E_1, E_2,

$$\|\mathcal{L}_{\Sigma_0}(E)\|_F \le C_r \frac{\|E\|_{\mathrm{op}}}{n}, \qquad \left\|D^2 P(\Sigma_0 + H)[E_1, E_2]\right\|_F \le C_r \frac{\|E_1\|_{\mathrm{op}} \|E_2\|_{\mathrm{op}}}{n^2},$$

the second bound holding uniformly for $\|H\|_{\mathrm{op}} \le n/4$. Since $\Sigma_\Theta - \Sigma_{\Theta'} = n(P_\Theta - P_{\Theta'}) = nK(\Theta - \Theta') + n(E(\Theta) - E(\Theta'))$, the block expansion of E above yields

$$\|\Sigma_\Theta - \Sigma_{\Theta'} - nK(\Theta - \Theta')\|_{\mathrm{op}} \le Cn\varepsilon \|\Theta - \Theta'\|_F, \qquad \|\Sigma_\Theta - \Sigma_0\|_{\mathrm{op}} \le Cn\varepsilon.$$

Because $K(\Theta)$ is supported on the off-diagonal blocks between P_0 and $P_0^\perp$, the reduced-resolvent derivative across the gap n acts by $\mathcal{L}_{\Sigma_0}(nK(\Theta)/W) = K(\Theta)/W$. Applying the first-order expansion to $\bar{\Sigma}_t(\Theta) = \Sigma_0 + W^{-1}(\Sigma_\Theta - \Sigma_0)$, whose perturbation has operator norm $O(n\varepsilon/W) \le n/4$ for ε small and $W \ge W_0$, and controlling the second-order remainder by the displayed D^2P bound together with the $C\varepsilon$-Lipschitz property of E, shows that the Taylor remainder for this projector is Lipschitz of size $C\varepsilon/W$ in the chart. Therefore

$$P(\bar{\Sigma}_t(\Theta)) = P_0 + \frac{1}{W}K(\Theta) + R_+(\Theta), \qquad P(\bar{\Sigma}_{t-1}(\Theta)) = P_0 - \frac{1}{W}K(\Theta) + R_-(\Theta),$$

where, uniformly for $\|\Theta\|_{\mathrm{op}}, \|\Theta'\|_{\mathrm{op}} \le \varepsilon$,

$$\|R_\pm(\Theta) - R_\pm(\Theta')\|_F \le \frac{C\varepsilon}{W} \|\Theta - \Theta'\|_F .$$

Consequently

$$D_t^{\mathrm{roll}}(\Theta) - D_t^{\mathrm{roll}}(\Theta') = \frac{2}{W} K(\Theta - \Theta') + (R_+(\Theta) - R_+(\Theta')) - (R_-(\Theta) - R_-(\Theta')).$$

Because $\|K(A)\|_F = \sqrt{2}\,\|A\|_F$, choosing $\varepsilon > 0$ sufficiently small gives the bi-Lipschitz lower bound

$$\left\|D_t^{\mathrm{roll}}(\Theta) - D_t^{\mathrm{roll}}(\Theta')\right\|_F^2 \ge \frac{c}{W^2} \|\Theta - \Theta'\|_F^2 . \tag{2}$$

Now construct the finite subclass promised in the statement. Let $d = r(n-r)$ and identify $v \in \{-1,1\}^d$ with a matrix $\Theta_v \in \mathbb{R}^{(n-r)\times r}$ whose entries are $\pm h$. Set $\mathcal{V} = \{\Theta_v : v \in \{-1,1\}^d\}$. Choose

$$h^2 = \frac{\kappa}{n},$$

where $\kappa > 0$ is a sufficiently small constant depending only on r and the fixed spectral constants. Since r is fixed,

$$\|\Theta_v\|_{\mathrm{op}} \le \|\Theta_v\|_F = h\sqrt{r(n-r)} \le C_r\sqrt{\kappa},$$

so the whole hypercube lies in the local neighborhood above when κ is small enough. By (2), for any two vertices,

$$\left\|D_t^{\mathrm{roll}}(\Theta_v) - D_t^{\mathrm{roll}}(\Theta_{v'})\right\|_F^2 \ge \frac{c}{W^2} \|\Theta_v - \Theta_{v'}\|_F^2 = \frac{4ch^2}{W^2} d_H(v, v'), \tag{3}$$

where d_H denotes Hamming distance.

It remains to bound the testing affinity between neighboring vertices. The $W-1$ shared-core observations have the same law for every v and therefore do not contribute to the Kullback–Leibler divergence. Only the two boundary observations depend on v.

For two rank-r projectors P, Q, set

$$\Sigma_P = I_n + nP, \qquad \Sigma_Q = I_n + nQ.$$

Since

$$\Sigma_Q^{-1} = I_n - \frac{n}{n+1} Q$$

and $\det(\Sigma_P) = \det(\Sigma_Q) = (n+1)^r$, the Gaussian covariance KL divergence satisfies

$$\begin{aligned} \mathrm{KL}(N(0,\Sigma_P), N(0,\Sigma_Q)) &= \frac{1}{2}\left[\mathrm{tr}(\Sigma_Q^{-1}\Sigma_P) - n\right] \\ &= \frac{n^2}{4(n+1)} \|P - Q\|_F^2 \le Cn \|P - Q\|_F^2 . \end{aligned}$$

In the local chart,

$$\|P_\Theta - P_{\Theta'}\|_F \le C \|\Theta - \Theta'\|_F .$$

Thus, the entering boundary contribution satisfies, if v and v' differ in one coordinate,

$$\mathrm{KL}_{\mathrm{enter}}(\mathbb{P}_v, \mathbb{P}_{v'}) \le Cn \|\Theta_v - \Theta_{v'}\|_F^2 = Cn(4h^2) = C\kappa.$$

The same bound applies to the exiting boundary covariance $\Sigma_{-\Theta_v}$ versus $\Sigma_{-\Theta_{v'}}$, and the shared-core part contributes zero. Hence the full adjacent-window experiment satisfies

$$\mathrm{KL}(\mathbb{P}_v, \mathbb{P}_{v'}) \le 2C\kappa \le C'\kappa.$$

Choosing κ small enough, Pinsker's inequality [19, Ch. 2] gives a uniform edge affinity bounded away from zero.

Assouad's lemma [1, 24] applied to the hypercube $\{\Theta_v : v \in \{-1,1\}^d\}$ and the separation (3) yields

$$\inf_{\widehat{D}} \sup_{v\in\{-1,1\}^d} \mathbb{E}_v \left\| \widehat{D} - D_t^{\mathrm{roll}}(\Theta_v) \right\|_F^2 \ge c\, d\, \frac{h^2}{W^2}.$$

Since $d = r(n-r)$ and $h^2 = \kappa/n$,

$$dh^2 = r(n-r)\frac{\kappa}{n} \asymp r$$

for $n \ge 2r+1$. Therefore

$$\inf_{\widehat{D}} \sup_{v\in\{-1,1\}^d} \mathbb{E}_v \left\| \widehat{D} - D_t^{\mathrm{roll}}(\Theta_v) \right\|_F^2 \ge c\frac{r}{W^2}.$$

This proves the lower bound, with $\mathcal{V} = \{\Theta_v : v \in \{-1,1\}^d\}$. □

Remark 7.2 (Matching upper bound and the limited sharpness claim)**.** The lower bound of Proposition 7.1 is rate-sharp over the same local boundary family, but the matching upper bound is intentionally trivial. The estimator $\widehat{D} \equiv 0$ attains worst-case risk of order r/W^2 because the constructed rolling drift is itself only at the null scale. Indeed, the first-order projector expansion at Σ_0 used in the proof gives, uniformly over the local neighborhood,

$$\left\| D_t^{\mathrm{roll}}(\Theta) \right\|_F = \frac{1}{W} \left\| 2K(\Theta) + W\left(R_+(\Theta) - R_-(\Theta)\right) \right\|_F \le \frac{C}{W} \left\| \Theta \right\|_F,$$

so that on the hypercube

$$\sup_{v\in\{-1,1\}^d} \left\| D_t^{\mathrm{roll}}(\Theta_v) \right\|_F^2 \le \frac{C}{W^2} \left\| \Theta_v \right\|_F^2 = \frac{C}{W^2}\, dh^2 \asymp \frac{r}{W^2}.$$

Consequently

$$\inf_{\widehat{D}} \sup_{v\in\{-1,1\}^d} \mathbb{E}_v \left\| \widehat{D} - D_t^{\mathrm{roll}}(\Theta_v) \right\|_F^2 \le \sup_{v\in\{-1,1\}^d} \left\| D_t^{\mathrm{roll}}(\Theta_v) \right\|_F^2 \le C\frac{r}{W^2},$$

and Proposition 7.1 gives the reverse inequality up to constants. Thus the minimax risk over this least-favorable family is exactly of order r/W^2. This should not be read as a nontrivial estimation theorem below the benchmark of Theorem 5.11; it says that, when the rolling signal is manufactured at the same scale as the static-null fluctuation floor, no estimator can improve the rate, and outputting zero is already rate-optimal on this least-favorable family.

Remark 7.3 (Instantaneous versus rolling drift)**.** Proposition 7.1 concerns D_t^{roll}, not the arbitrary instantaneous target $P_t - P_{t-1}$. A lower bound for instantaneous drift requires an additional local path model linking $P_t - P_{t-1}$ to the boundary difference seen by the two adjacent rolling windows. Without such a link the problem can be either easier, as under a known static null, or harder, as in unrestricted nonstationary models. Thus the sharpness claim in Proposition 7.1 is intentionally a rolling-target claim.

8 Discussion

The central mechanism in this analysis is the shared-core decomposition of adjacent rolling windows. It yields a null increment scale that is smaller by one factor of W than the single-window projector error: the single-window perturbation uses W noisy observations, whereas the adjacent difference uses only the entering and exiting observations after the common core cancels. Under the bulk-plus-spikes scaling this produces the two scales

$$\text{single-window projector scale } r/W, \qquad \text{adjacent overlap drift scale } r/W^2.$$

We stress that, under the static null, the constant $2\mathfrak{d}_{\mathrm{sub}}/W^2$ is a *null-fluctuation floor* rather than a measure of any real movement of the population subspace, which is fixed: it quantifies the noise that genuine drift must exceed to be detectable.

The lower-bound construction in Proposition 7.1 is separate from this null calculation: it concerns the rolling target D_t^{roll} in a boundary-varying family, and its matching upper bound is the zero estimator because the constructed signal is already at the null scale. The construction keeps the boundary covariances in the stated bulk-plus-spikes class. The population window averages retain a uniformly separated top-r cluster, which is what the rolling target needs, but their lower cluster is $O(1+n/W)$; hence the averages themselves satisfy the full $O(1)$-bulk version of Assumption 2.2 only in the additional regime $W \gtrsim n$.

The leading constant is not obtained by applying Davis–Kahan [7] to two independent windows. It requires conditioning on the shared core, expanding each endpoint around that same random core, and then replacing the random-core variance functional by its population limit. Two distinct calibrations follow from the same conditioning: the magnitude benchmark $2\mathfrak{d}_{\mathrm{sub}}/(W-1)^2$ of Theorem 5.12, with limiting shorthand $2\mathfrak{d}_{\mathrm{sub}}/W^2$, which is asymptotic and carries the factor-of-two convention of Remark 4.2, and the exact finite-sample sign symmetry of Theorem 5.2, which needs no constant at all and so remains calibrated for a fixed oriented contrast even when the magnitude scale is in doubt. Sequential use over many overlapping times, however, still requires the additional dependence handling described in Remark 5.4.

Remark 8.1 (Scope: spectral regime and Gaussianity)**.** Three boundaries of the present analysis are worth stating explicitly.

First, Assumption 2.2 fixes the benign, strongly spiked regime: r spikes of order n separated from an $O(1)$ bulk by a gap of order n. It is precisely this $\Theta(n)$ separation that makes the rates dimension-free, the linearization remainder negligible, and the spectral-difficulty functional $\mathfrak{d}_{\mathrm{sub}} = \Theta(r)$. The harder high-dimensional regimes — bounded spikes, a shrinking eigengap, or proportional growth with n/W of constant order, where Baik–Ben Arous–Péché phase-transition phenomena and nontrivial eigenvector inconsistency arise — are deliberately excluded, and the constant $2\mathfrak{d}_{\mathrm{sub}}/W^2$ should not be expected to persist in that setting.

Second, the conditioning argument requires an exchangeable boundary pair independent of the shared core. The i.i.d. Gaussian null supplies this directly; serial dependence, volatility clustering, or deterministic local trends can make r_t and r_{t-W} nonexchangeable after conditioning on the core, in which case the exact sign-randomization statement does not apply without an additional blocking or model-based resampling step.

Third, the *exact constant* is Gaussian. The exact sign-randomization statement uses only conditional exchangeability of the boundary pair. The magnitude rate r/W^2 is expected to extend beyond Gaussian innovations under additional assumptions ensuring shared-core eigengap concentration and suitable quadratic-chaos moment or tail bounds, but that extension is not proved here. At the purely linearized second-moment level, non-Gaussian innovations would introduce fourth-cumulant terms; finite fourth moments alone do not give the full high-dimensional theorem proved here. Thus the constant $2\mathfrak{v}_{\mathrm{sub}}$ is specific to the Gaussian Isserlis calculation, while non-Gaussian boundary innovations generally lead to model-dependent leading constants.

References

[1] P. Assouad. Deux remarques sur l'estimation. *C. R. Acad. Sci. Paris Sér. I Math.*, 296(23):1021–1024, 1983.

[2] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 2009.

[3] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011.

[4] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997.

[5] T. T. Cai, Z. Ma, and Y. Wu. Sparse PCA: optimal rates and adaptive estimation. *Ann. Statist.*, 41(6):3074–3110, 2013.

[6] T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.*, 46(1):60–89, 2018.

[7] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7(1):1–46, 1970.

[8] D. Hsu, S. M. Kakade, and T. Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.*, 17(52):1–6, 2012.

[9] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

[10] S. Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[11] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[12] T. Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, Berlin, reprint of the 1980 edition, 1995.

[13] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017.

[14] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.

[15] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.

[16] E. Nelson. The free Markoff field. *J. Funct. Anal.*, 12(2):211–227, 1973.

[17] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642, 2007.

[18] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

[19] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.

[20] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.

[21] V. Q. Vu and J. Lei. Minimax sparse principal subspace estimation in high dimensions. *Ann. Statist.*, 41(6):2905–2947, 2013.

[22] D. Wang, Y. Yu, and A. Rinaldo. Optimal covariance change point localization in high dimensions. *Bernoulli*, 27(1):554–575, 2021.

[23] P.-Å. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT*, 12(1):99–111, 1972.

[24] B. Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, New York, 1997.