GEOMETRY-AWARE DISTRIBUTIONAL CHANGE-POINT INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

A distributional change in independent manifold-valued observations can preserve location summaries and remain hard to detect even after a representation makes it visible. We study offline testing and localization through the contrast retained by geometric features and the variation assigned to that contrast. Our scan combines multiscale diffusion features with geodesic distance critics and controls finite-sample type-I error by jointly permuting all channels under the iid null. For fixed channels with stable standardizers, we derive local feature-drift limits and consistent localization whenever a retained contrast is nonzero. The leading trace-standardized signal is the fraction of pooled feature variance explained by the change. This characterization motivates projecting distance features off a finite diffusion span: when the diffusion contrast vanishes, projection preserves the distance contrast and yields an explicit variance-reduction factor. Circle, sphere and torus experiments demonstrate sensitivity to matched-moment changes and distinguish detection from localization. A frequency-band comparison recovers the main residual gain at lower cost, showing that variance allocation, as well as signal retention, determines the value of a geometric representation.

1 INTRODUCTION

In wind-farm forecasting, a model fitted to the current wind regime becomes biased when the regime changes; Ezzat et al. (2019) document frequent speed and direction changes and calibrate forecasts by the current regime and the time since its last change. Covariance-matrix sequences likewise motivate change-point inference beyond Euclidean observations (Jiang et al., 2024). Angles wrap around, so their changes require geometry-aware comparisons, and even a geometry-respecting average can miss distributional change: two equally weighted masses at opposite directions have zero ambient mean before and after a rotation, so concentration, multimodality and angular dependence can change without a mean contrast. The offline task for independent observations is to test homogeneity and locate distributional change; rank and graph scans provide nonparametric starting points (Pettitt, 1979; Chen and Zhang, 2015), and the question is which geometric representations retain and amplify the changing features.

Kernel change-point methods address distributional changes by comparing the kernel feature means of candidate segments. Kernel Fisher scans and kernel segmentation already provide tests and location estimates for structured observations (Harchaoui et al., 2008; Garreau and Arlot, 2018; Arlot et al., 2019). A characteristic kernel distinguishes every pair of laws through its mean embedding (Sriperumbudur et al., 2010; Gretton et al., 2012). This population property leaves a statistical choice unresolved: the kernel can assign most feature variation to directions unrelated to the change, and a finite approximation can omit the changing direction altogether, so retaining a nonzero contrast and weighting it effectively are separate requirements.

Geometry offers several ways to construct such contrasts. Energy-distance segmentation uses pairwise distances in an embedding (Matteson and James, 2014); Fréchet scans target geometric location and variation (Dubey and Müller, 2020); and distance-profile scans provide distributional inference for general metric data, including fixed and local alternatives and location rates (Dubey and Zheng, 2026). Wasserstein scans compare empirical laws through optimal transport (Cheng et al., 2020). Diffusion features use additional Riemannian structure, diffusion time weighting spatial frequencies as in diffusion geometry (Coifman and Lafon, 2006). Existing manifold two-sample theory explains

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.

how bandwidth and intrinsic dimension affect kernel tests (Cheng and Xie, 2024), while aggregated maximum mean discrepancy (MMD) tests address uncertainty about the kernel (Schrab et al., 2023). Classification-based scans learn distributional contrasts (Kanrar et al., 2025), while online kernel cumulative-sum (CUSUM) methods address sequential detection (Wei and Xie, 2026). Here representations are compared after a common offline search over splits and joint calibration of their dependent statistics.

We connect these choices through a multichannel CUSUM scan: diffusion channels resolve distributional differences at several diffusion times, distance-cone critics compare mean geodesic distances to landmarks, a pooled variance standardizes each channel, and a common permutation calibrates the maximum over channels and candidate splits. The construction makes the geometric choices explicit and permits finite-feature computation by cumulative sums. The scan detects changes through retained feature contrasts: exact diffusion features identify every change on a closed connected manifold, whereas a computed finite representation detects the contrasts that survive its truncation.

Our theory characterizes signal retention and variance allocation: every feature bridge has the same triangular mean drift, and its leading trace-standardized signal is the share of pooled variance explained by the change. Regressing distance coordinates on the retained diffusion span preserves a distance contrast absent from that span while removing shared variation. We derive the variance gain, fixed-feature null law and consistent localization. Keeping the original diffusion channels also retains contrasts carried by that span.

Experiments on the circle, sphere and torus test matched-moment changes and compare detection with localization: residualization recovers a third-harmonic change missed by raw diffusion–distance aggregation, fixed frequency bands achieve the same gain more cheaply, and transport and energy comparisons show how the representation and joint search affect both tasks.

Section 2 defines the scan, Section 3 explains calibration and signal strength, and Section 4 compares detection and localization. The appendices give implementation details, supporting results and proofs.

2 Geometric features and the joint scan

2.1 Model and feature contrasts

Let $(\mathcal{X}, d)$ be a Polish metric space and let $Z_1, \dots, Z_n$ be independent observations in $\mathcal{X}$. The no-change null is that the observations are independent and identically distributed (iid),

$$H_0 : \quad Z_1, \dots, Z_n \text{ are iid with law } P \tag{1}$$

against a single distributional change,

$$H_1 : \quad Z_i \sim P_n \ (i \le \tau_n), \quad Z_i \sim Q_n \ (i > \tau_n), \quad \theta_n := \tau_n/n \longrightarrow \theta \in (0,1), \tag{2}$$

where $P_n \neq Q_n$; fixed changes have $P_n = P$ and $Q_n = Q$. A trim $\eta \in (0, 1/2)$ restricts the search to $\mathcal{K}_n(\eta) = \{k : \lceil n\eta \rceil \le k \le \lfloor n(1-\eta) \rfloor\}$, with $r_k = k/n$, so each candidate segment has at least $\lceil n\eta \rceil$ observations, and changes to be localized satisfy $\theta \in (\eta, 1-\eta)$. For a feature map $\psi : \mathcal{X} \to \mathcal{H}$ into a separable Hilbert space, write $P\psi = \int \psi \, \mathrm{d}P$ and define its centered cumulative-sum bridge

$$\mathbb{B}_{n,\psi}(r) = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{\lfloor nr \rfloor} \psi(Z_i) - \frac{\lfloor nr \rfloor}{n} \sum_{i=1}^{n} \psi(Z_i) \right\}. \tag{3}$$

At a candidate split this is $\sqrt{n}\, r_k(1-r_k)$ times the difference of the segment feature means, so the change is visible in this representation when $P_n\psi - Q_n\psi \neq 0$.

For the intrinsic diffusion construction, $\mathcal{X} = \mathcal{M}$ is a closed, connected Riemannian manifold with geodesic distance and volume measure μ. If p_t is the heat kernel at time $t > 0$, set $\Phi_t(x) = p_t(x, \cdot) \in L^2(\mathcal{M}, \mu)$ and $k_t(x,y) = \langle \Phi_t(x), \Phi_t(y) \rangle = p_{2t}(x,y)$, so that $\mathrm{MMD}_t(P,Q) = \|P\Phi_t - Q\Phi_t\|$ and, for the eigenpairs (λ_j, φ_j) of $-\Delta$,

$$\mathrm{MMD}_t^2(P,Q) = \sum_{j=0}^{\infty} e^{-2t\lambda_j} \left\{ \int \varphi_j \ \mathrm{d}(P-Q) \right\}^2.$$

Every spectral weight is positive, so the exact embedding separates all laws, $\mathrm{MMD}_t(P, Q) = 0$ if and only if $P = Q$ (Proposition C.1), and increasing t suppresses high-frequency differences more strongly; a finite spectral or Nyström representation retains only its computed span, which exact separation alone does not certify.

Distance contrasts require only a compact metric space. Fix an anchor $x_\star$ and landmarks $a_1, \ldots, a_m$ and define $F_m = (f_1, \ldots, f_m)^\top$ by $f_j(x) = d(x, a_j) - d(x_\star, a_j)$, so that $Pf_j - Qf_j = \mathbb{E}_P d(Z, a_j) - \mathbb{E}_Q d(Z, a_j)$; each critic is anchored and one-Lipschitz. If $\mathcal{F}_{\mathrm{Lip}}$ is the full class of functions with $f(x_\star) = 0$ and $|f(x) - f(y)| \le d(x, y)$, Kantorovich–Rubinstein duality gives

$$\sup_{f \in \mathcal{F}_{\mathrm{Lip}}} |\mathbb{B}_{n,f}(r_k)| = \sqrt{n}\, r_k(1 - r_k) W_1(\widehat{P}_{1:k}, \widehat{P}_{k+1:n}),$$

where the hats denote empirical segment laws (Villani, 2009). The diffusion bridge has the analogous norm identity with MMD_t (Proposition C.2). A dictionary approximating every member of $\mathcal{F}_{\mathrm{Lip}}$ uniformly within ε approximates its population discrepancy within 2ε (Proposition C.3); landmark distance cones measure selected mean-distance contrasts and are not assumed to satisfy this uniform-net condition.

2.2 Standardization and computation

For a finite grid $\mathcal{T} = \{t_1, \ldots, t_L\}$, write $\overline{\Phi}_{n,t} = n^{-1} \sum_i \Phi_t(Z_i)$ and let $\widehat{v}_{n,t}$ be the pooled variance $n^{-1} \sum_i \|\Phi_t(Z_i) - \overline{\Phi}_{n,t}\|^2$. Let $\widehat{\Sigma}_F$ be the pooled covariance of $F_m(Z_i)$ and write $\mathbb{G}_{n,F} = \mathbb{B}_{n,F_m}$. For a nonnegative-definite matrix $A_{n,F}$, the diffusion and critic scans are

$$S^H_{n,t} = \max_{k \in \mathcal{K}_n(\eta)} \frac{\|\mathbb{B}_{n,\Phi_t}(r_k)\|^2}{\widehat{v}_{n,t} r_k(1 - r_k)},$$

$$S^F_n(A_{n,F}) = \max_{k \in \mathcal{K}_n(\eta)} \frac{\mathbb{G}_{n,F}(r_k)^\top A_{n,F} \mathbb{G}_{n,F}(r_k)}{r_k(1 - r_k)}.$$

The default is trace standardization, $A_{n,F} = I_m / \operatorname{tr}(\widehat{\Sigma}_F)$, a zero-variance channel contributes zero, and for positive prespecified weights

$$T_n = \max \left\{ \max_{1 \le \ell \le L} c_\ell S^H_{n,t_\ell},\ c_F S^F_n(A_{n,F}) \right\}. \tag{4}$$

The location estimate $\widehat{\tau}_n$ is the smallest split attaining this joint maximum ((18) gives the splitwise criterion). Other quadratic metrics, including covariance whitening, which requires stable covariance rank for its asymptotic limit, are described in Section A.2.

For finite feature rows $X_i \in \mathbb{R}^q$ with pooled mean $\overline{X} = n^{-1} \sum_i X_i$, compute $u_k = \sum_{i \le k} (X_i - \overline{X})$ and $\widehat{v} = n^{-1} \sum_i \|X_i - \overline{X}\|^2$. The trace-standardized criterion is $\|u_k\|^2 / [n \widehat{v} r_k(1 - r_k)]$. After feature construction, cumulative updates take $O((B + 1)n(Lq + m))$ scan arithmetic for B permutations, L diffusion channels of at most q coordinates and m critics, against $O((B + 1)Ln^2)$ for exact diffusion Gram scans (Section A.4 gives their block-sum formula). The finite Nyström implementation (Williams and Seeger, 2001) and complete test are specified in Section B.1 and Algorithm 1.

Scales, landmarks and covariance estimates may be fitted once to the pooled sample when their construction is permutation invariant; canonical coordinate ordering before landmark selection resolves ties independently of temporal order, after which feature rows return to temporal order, whereas an order-dependent tuning rule must be refitted for every permutation. Repeating the maximum over channels and splits in each permutation calibrates the whole search, so an added channel can raise both the observed statistic and the null critical value.

3 Theoretical results

Calibration controls false alarms from searching over splits and representations, feature drift determines which changes exceed this noise, and pooled variance explains how a representation strengthens or dilutes that drift; these three steps connect the randomization guarantee to power and location accuracy.

3.1 JOINT CALIBRATION AND FEATURE DRIFT

The computed statistic inherits finite-sample validity from the standard randomization principle (Phipson and Smyth, 2010; Hemerik and Goeman, 2018).

Theorem 3.1. *Consider the null case, in which H_0 in* (1) *holds. Let T_n be any measurable statistic obtained by applying the same declared algorithm to an ordered sample, where any data-dependent diffusion grid, critic dictionary, regularization parameter or channel weight is either invariant under permutations of the pooled sample or recomputed by the same rule for every permuted sample. Let $\pi_1, \dots, \pi_B$ be independent uniform random permutations, independent of the data, and let*

$$\widehat{p}_B = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{T_n(Z_{\pi_b(1)}, \dots, Z_{\pi_b(n)}) \geq T_n(Z_{1:n})\}}{B+1}. \tag{5}$$

Then, for every $\alpha \in [0,1]$,

$$\mathbb{P}_{H_0}(\widehat{p}_B \leq \alpha) \leq \alpha. \tag{6}$$

With randomized tie breaking, equality holds at attainable levels under the conditional randomization distribution.

The minimum attainable p-value is $(B+1)^{-1}$, and validity applies directly to the computed features and pooled fitting. For fixed features with finite second moments and stable standardizers, the joint bridge has a Gaussian limit with covariance $(r \wedge s - rs)\Gamma$, Γ the joint feature covariance; common multipliers reproduce the dependence between channels, and finite $2+\delta$ moments with critical-quantile continuity justify their asymptotic calibration. Theorems C.7 and C.9 give these conditions precisely, and Section E treats full critic classes and predictable scores.

Let $\Psi = (\Phi_{t_1}, \dots, \Phi_{t_L}, F_m)$ collect the fixed features in $\mathcal{V} = \bigoplus_{\ell=1}^{L} L^2(\mathcal{M}, \mu) \oplus \mathbb{R}^m$, and write $\delta_n = P_n\Psi - Q_n\Psi$. Under (2), the expected bridge is $\sqrt{n}\, b_{\theta_n}(r_k)\delta_n$ with $b_\theta(r) = r(1-\theta)\mathbf{1}\{r \leq \theta\} + \theta(1-r)\mathbf{1}\{r > \theta\}$. Assumption C.10 collects the conditions: the bridge of $U_{n,i} = \Psi(Z_i) - \mathbb{E}\Psi(Z_i)$ converges to a continuous centered Gaussian process $\mathbb{Z}$ with zero endpoints, the pooled variance limits are positive and finite, and the quadratic standardizer satisfies $A_n \to A$ in probability in operator norm with A bounded, self-adjoint and nonnegative definite. For fixed continuous features on a compact space, Proposition C.11 proves them for trace standardization whenever $P_n \Rightarrow P$, $Q_n \Rightarrow Q$ and the limiting channel variances are positive. This includes fixed P, Q; the law convergence supplies the additional covariance stability for triangular alternatives.

The next result distinguishes a local change, whose $n^{-1/2}$ feature contrast competes with Gaussian noise, from a fixed nonzero contrast, whose drift eventually dominates: part (i) gives the shifted limit governing local power, and parts (ii)–(iii) give the growing signal and consistent location under fixed changes, extending the feature-mean CUSUM argument (Aue et al., 2009).

Theorem 3.2. *Suppose Assumption C.10 holds and $\theta_n \to \theta \in (\eta, 1-\eta)$.*

(i) If $\sqrt{n}\,\delta_n \to h$ in $\mathcal{V}$, then

$$\mathbb{B}_{n,\Psi}(\cdot) \rightsquigarrow \mathbb{Z}(\cdot) + b_\theta(\cdot)h. \tag{7}$$

Thus every continuous scan functional has a Gaussian-bridge limit shifted by the triangular drift, a Brownian bridge only when the pre- and post-change feature covariances coincide, as when $P = Q$ in Proposition C.11.

(ii) If $\delta_n \to \delta$ and $\langle \delta, A\delta \rangle > 0$, then

$$\frac{1}{n} \max_{k \in \mathcal{K}_n(\eta)} \frac{\langle \mathbb{B}_{n,\Psi}(r_k), A_n \mathbb{B}_{n,\Psi}(r_k) \rangle}{r_k(1-r_k)} \xrightarrow{\mathbb{P}} \langle \delta, A\delta \rangle \, \frac{b_\theta(\theta)^2}{\theta(1-\theta)} = \theta(1-\theta)\, \langle \delta, A\delta \rangle . \tag{8}$$

In particular, the unscaled scan diverges in probability.

(iii) Let $\widehat{\theta}_n$ be the smallest maximizing split in (ii), divided by n. Then $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$, and the same conclusion holds for the combined estimator (18) *whenever at least one retained channel has a nonzero limiting signal in its declared quadratic metric.*

The drift's normalized square $b_\theta(r)^2/[r(1-r)]$ increases to θ and then decreases, as in the standard CUSUM localization argument (Aue et al., 2009; Csörgő and Horváth, 1997); every channel shares

this shape, so their maximum preserves the change location (proof in Section I; Theorem C.14 gives fixed-channel test consistency). Under the circle regularity and cutoff conditions of Theorem C.15, pooled Nyström representations converge uniformly up to isometry, and a nonzero limiting contrast then gives $T_n/n \xrightarrow{\mathbb{P}} c > 0$, consistent localization and permutation power tending to one for B permitting rejection.

3.2 Variance allocation and the residual distance channel

Distance coordinates can reveal a contrast omitted by a finite diffusion span but may also repeat large fluctuations that span already represents; we separate the useful contrast from this shared variation and remove the latter by pooled regression, retaining the original diffusion channels. For a fixed trace-standardized score, let $m_P = P\Psi$, $m_Q = Q\Psi$ and $\delta = m_P - m_Q \neq 0$. The variance under the pooled law $\theta P + (1-\theta)Q$ splits into within- and between-segment terms, so the limit in (8) is

$$\frac{\theta(1-\theta)\,\|\delta\|^2}{\theta\mathbb{E}_P\,\|\Psi(Z) - m_P\|^2 + (1-\theta)\mathbb{E}_Q\,\|\Psi(Z) - m_Q\|^2 + \theta(1-\theta)\,\|\delta\|^2}. \tag{9}$$

This fraction lies in $(0, 1]$ and is the share of pooled feature variation explained by the change; directions contributing variation but little mean contrast dilute the leading signal, and for diffusion features diffusion time changes both numerator and denominator, so characteristicness alone does not choose a sensitive scale.

To remove variation already represented by diffusion, let $H : \mathcal{X} \to \mathbb{R}^q$ be a fixed bounded basis for the retained diffusion-coordinate span and let $F_m : \mathcal{X} \to \mathbb{R}^m$ be fixed and bounded. Fit the pooled least-squares regression

$$F^{\perp}_{n,i} = F_m(Z_i) - \overline{F} - \widehat{B}_H\{H(Z_i) - \overline{H}\}, \qquad \widehat{B}_H = \widehat{\Sigma}_{FH}\widehat{\Sigma}^{-1}_{HH}. \tag{10}$$

with pooled means and pooled covariance blocks; set $\widehat{B}_H = 0$ when $\widehat{\Sigma}_{HH}$ is singular and use zero for a zero-variance channel. The trace-standardized residual channel replaces the raw critic channel in (4), diffusion is retained, and the regression fit is permutation invariant and reused across permutations. With several diffusion times, H spans their combined finite coordinate space; an invertible basis change leaves fitted values unchanged, so repeated spectral modes need appear only once, the projection depends only on this span, each original diffusion scan retains its own weights, and pooled fitting uses the same relation at every split and permutation.

For a law ν, let Σ be the covariance of (H, F_m), $B_H = \Sigma_{FH}\Sigma^{-1}_{HH}$, and for a fixed change $\delta_H = PH - QH$ and $\delta_F = PF_m - QF_m$; the residual contrast and pooled variance are

$$\delta_{\perp} = \delta_F - B_H\delta_H, \qquad v_{\perp} = \operatorname{tr}(\Sigma_{FF} - \Sigma_{FH}\Sigma^{-1}_{HH}\Sigma_{HF}). \tag{11}$$

Proposition 3.3. *Suppose Σ_{HH} is positive definite and $v_{\perp}$ in* (11) *is positive.*

(i) Under the iid null ν, the diffusion–residual scan has finite-sample permutation level at most α for every B permutations, using (10)*, with zero regression coefficients for singular covariance and zero channel scores for zero variance. Its fixed-feature null limit is the jointly calibrated maximum of the diffusion and residual bridges. The Gaussian limits of the H and residual bridges are independent.*

(ii) For a fixed change P to Q at $\theta \in (\eta, 1-\eta)$, take $\nu = \theta P + (1-\theta)Q$. The residual scan divided by n converges in probability to $\theta(1-\theta)\|\delta_F - B_H\delta_H\|^2/v_{\perp}$. If either δ_H or $\delta_F - B_H\delta_H$ is nonzero, the joint location estimate is consistent.

(iii) If $\delta_H = 0$ and $\delta_F \neq 0$, the residual contrast equals δ_F and its signal slope is the raw distance slope multiplied by $\operatorname{tr}(\Sigma_{FF})/v_{\perp} \geq 1$. The inequality is strict when $\Sigma_{FH} \neq 0$.

The proof is in Section I.3. Part (i) gives independent limiting Brownian bridges for the retained diffusion span and residuals, and part (iii) a variance reduction without loss of signal; when $\delta_H \neq 0$, projection can also remove signal through $B_H\delta_H$. The slope gain concerns the leading fixed-change signal, while the finite-sample benefit also depends on joint calibration and estimation of the projection.

The circle, identified with $[-\pi, \pi)$ with geodesic distance, makes the variance mechanism explicit: the centered distance to a landmark a has expansion

$$d(x,a) - \frac{\pi}{2} = -\frac{4}{\pi} \sum_{k \geq 1,\ k \text{ odd}} \frac{\cos\{k(x-a)\}}{k^2}.$$

with most of its variance in the first harmonic. Suppose H spans sine and cosine modes $1, \ldots, J$, use m equally spaced landmarks, and let the segment densities be $(1 \pm \varepsilon \cos(jx))/(2\pi)$ with $0 < \varepsilon < 1$ and $\theta = 1/2$, so the pooled law is uniform. For an odd changing frequency $j > J$ and $m \nmid 2j$, Proposition D.1 gives

$$\frac{\theta(1-\theta)\|\delta_\perp\|^2}{v_\perp} = \frac{\varepsilon^2}{4j^4 \sum_{k>J,\ k \text{ odd}} k^{-4}}.$$

The raw distance slope has the same numerator but sums over all positive odd frequencies in the denominator; for $J = 2$, projection removes the first-harmonic variance, a gain of $(\pi^4/96)/(\pi^4/96 - 1) \approx 69.1$. At $\varepsilon = 0.35$ and $j = 3$ the raw and residual slopes are approximately 0.00037 and 0.0258, and at $j = 5$ the residual slope is only 0.00334: the j^{-4} decay remains after removing the dominant irrelevant variance. For even $j > J$, both diffusion and distance mean contrasts vanish. Figure 1 illustrates the preserved contrast and removed variance; splitting diffusion coordinates into frequency bands also reduces irrelevant variation, and the numerical comparison tests both constructions under the same joint permutation threshold.

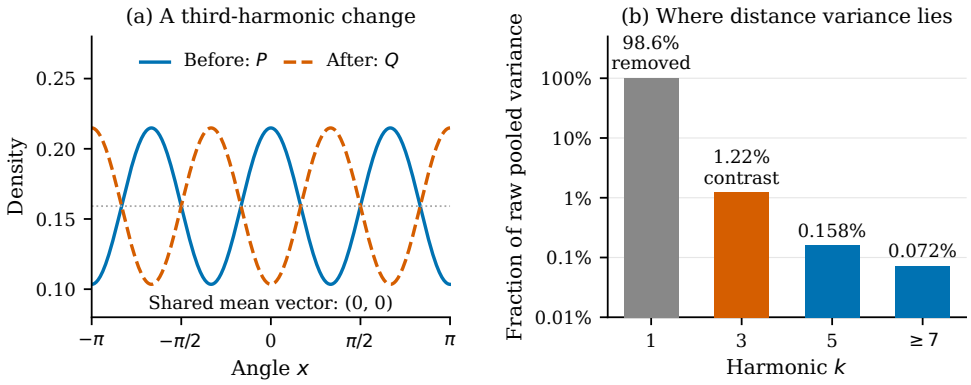


Figure 1: Circle mechanism ($j = 3$, $\varepsilon = 0.35$, $\theta = 1/2$, $J = 2$), using equally spaced landmarks with $m \nmid 6$. Left: densities share their first two Fourier moments. Right: projection removes first-harmonic variance (gray), preserving the changing third harmonic (orange); the final bar sums odd $k \geq 7$. The calculation follows Proposition D.1.

4 Numerical Experiments

All nonnull designs have fixed segment laws: the fixed Fourier experiment tests Theorem 3.2(ii)–(iii) and Proposition 3.3, pooled Nyström experiments assess finite-sample detection and localization, and Corollary C.16 verifies fixed-change consistency for the fitted middle circle design; local alternatives in part (i) have shrinking feature contrast.

4.1 Distributional sensitivity and location accuracy

We use the circle $\mathbb{S}^1$, sphere $\mathbb{S}^2$ and flat torus $\mathbb{T}^2$ to compare representations under the same search. Each primary alternative has $n = 500$, $\theta = 1/2$, trim $\eta = 0.10$, $R = 500$ independent repetitions and $B = 199$ shared permutations at level 0.05. Heat is a multiscale kernel mean-embedding scan maximizing five trace-standardized diffusion channels, Critic trace-standardizes 64 landmark distance cones, and Combined takes their equally weighted joint maximum, testing the added value of distance information. Diffusion features use up to 128 pooled medoid–farthest landmarks and rank at most

128, with times equal to the median squared pairwise distance times $\{2^{-6}, 2^{-4}, 2^{-2}, 1, 4\}$. These pooled choices are permutation invariant.

Extrinsic scans the mean of $(\cos z, \sin z)$ on the circle, $z \in \mathbb{R}^3$ on the sphere, and $(\cos z_1, \sin z_1, \cos z_2, \sin z_2)$ on the torus. Exact empirical circle W_1 and the single-change energy statistic, which uses exponent-one Euclidean distances in these embeddings with the same trim and plus-one calibration (Matteson and James, 2014), provide distributional baselines; full laws, baseline formulas and feature settings are in Sections B.1 and B.4.

The primary grid includes location, concentration and shape changes, with torus dependence replacing concentration. With von Mises density proportional to $\exp\{\kappa \cos(z - \mu)\}$ and sphere von Mises–Fisher density proportional to $\exp(\kappa \mu^\top z)$, for direction μ and concentration κ, circle location shifts have sizes $0.15, 0.30, 0.60$ from concentration 4, sphere shifts use the same angles from concentration 8, and concentration changes start at 8 and end at $4, 2, 1$. For shape changes, each unimodal baseline becomes an equal mixture with directions symmetrically displaced by $\pm a$, where $A(\kappa) \cos a = A(\kappa_0)$ for the mean resultant length $A(\cdot)$ and baseline concentration κ_0, so the population mean vector is exactly unchanged; the circle uses $\kappa_0 = 4$ and $\kappa = 6, 8, 12$, the sphere $\kappa_0 = 8$ and $\kappa = 12, 16, 24$. Torus dependence changes correlations to $0.25, 0.50, 0.75$ with both wrapped-normal marginals fixed; its shape alternatives mix opposite locations with within-component standard deviations $0.20, 0.35, 0.50$ and match the marginal first moments of a wrapped-normal baseline with standard deviation 0.7, and since the displacement decreases as the component variance increases, levels 1–3 approach the baseline rather than increase the change strength.

Rejection frequency estimates power under an alternative and type-I error under a null; localization uses all repetitions, including nonrejections, with mean absolute error (MAE) the average of $|\widehat{\tau} - \tau|/n$, so smaller is better, and the probability of error at most 0.02 measuring point-location accuracy rather than confidence-set coverage. Nine iid-null cells across $n = 200, 500, 1000$ give Combined rejection frequencies from 0.040 to 0.054 (Table 4), with largest null Monte Carlo standard error 0.0101; binomial standard errors quantify individual rejection frequencies, and paired differences use the standard deviation of the replicate-level indicator differences divided by $\sqrt{R}$.

The matched-moment alternatives in Figure 2 test distributional information beyond the mean. A sphere mixture with concentration $\kappa = 24$ preserves the extrinsic mean, yet Combined rejects in 99.0% of repetitions against 4.0% for Extrinsic (paired gain 0.950, Monte Carlo standard error 0.010; Critic 0.052). On the torus, two matched-marginal-moment mixtures are detected in every repetition by diffusion and Combined while the mean scan stays near nominal size, whereas the weaker circle shape designs remain difficult for all methods; these contrasts show which changes the retained features make statistically accessible. Joint calibration also explains the aggregation loss for torus dependence $\rho = 0.75$: Heat has power 0.734 and Combined 0.668 (paired loss 0.066, standard error 0.011) because every Critic rejection in this cell is already a Heat rejection, so the added search raises the threshold without adding detections. Exact circle W_1 is stronger for the concentration change $\kappa = 8 \to 4$, with power 0.838 against 0.616 for Combined (paired difference 0.222, standard error 0.019), while at the strongest circle shape setting the ordering reverses (Combined 0.142, W_1 0.086).

Table 1: Paired comparisons with the energy scan at the middle primary settings ($R = 500$, $n = 500$, $B = 199$); higher power and lower unconditional location MAE are better. Complete comparisons, including nulls, are in Table 7.

		Power		Location MAE	
Geometry	Change	Combined	Energy	Combined	Energy
Sphere	shape	0.226	0.256	0.074	0.121
Torus	dependence	0.088	0.132	0.126	0.150
Torus	shape	1.000	1.000	0.005	0.012

Detection and localization can favor different methods (Table 1): on the middle sphere shape change, Energy rejects more often, but Combined reduces location MAE from 0.121 to 0.074 (paired reduction 0.047, standard error 0.005), so a more concentrated maximizer can coexist with a lower rejection frequency. In the strongest shape designs, Combined localizes within ten indices with probabilities 0.902 on the sphere and 0.998 on the torus, whereas the weakest circle location shift is detected with

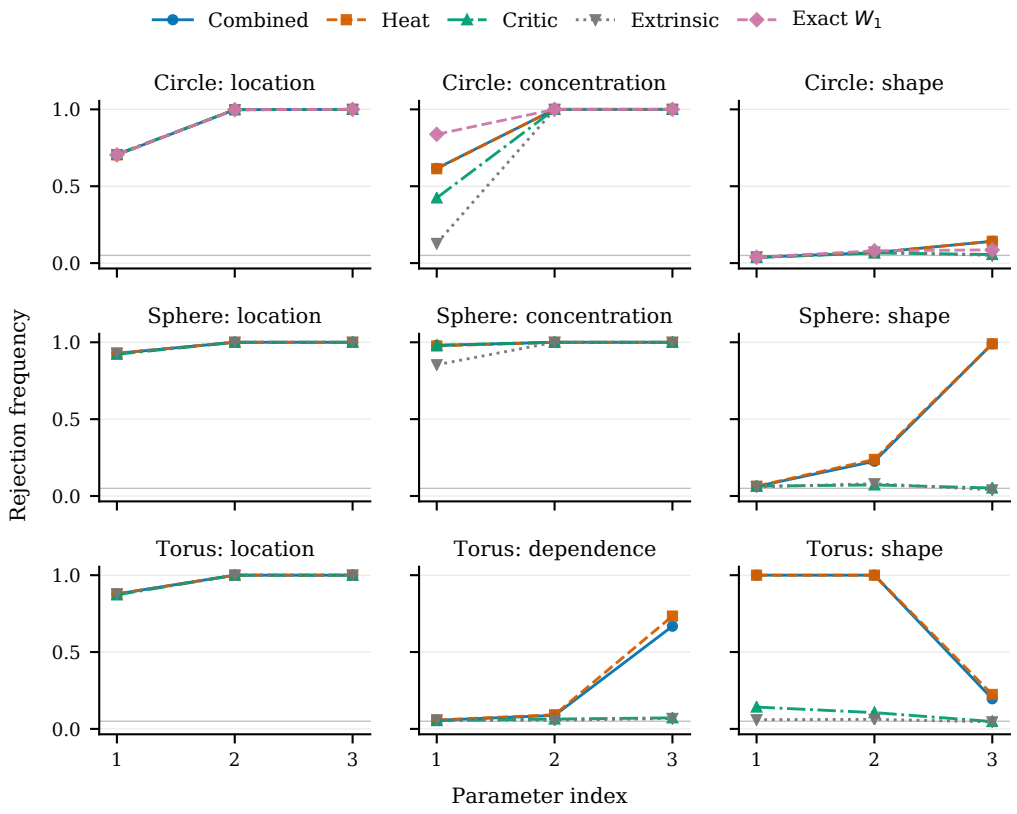


Figure 2: Rejection frequencies in the primary alternatives, with $R = 500$, $n = 500$, $\theta = 1/2$ and $B = 199$. Methods share samples and permutations. The largest binomial Monte Carlo standard error is 0.022. Levels index the parameters defined in Section B.1; torus shape levels move toward the baseline law, explaining their decreasing power.

probability 0.704 but localized within ten indices with probability 0.318. Tables 6 and 9 give the complete primary localization results and the sample-size and unequal-segment comparisons.

4.2 Testing the signal-and-variance mechanism

The spectral experiment uses circle densities $(1 \pm \varepsilon \cos(jx))/(2\pi)$ before and after the change, testing $j = 1, \ldots, 5$ at $\varepsilon = 0.35$, the uniform null $j = 0$ and a weaker third harmonic with $\varepsilon = 0.175$; every cell uses $R = 500$, $n = 500$, $\theta = 1/2$, $\eta = 0.10$ and $B = 199$, the methods share samples and permutations, and their feature rules are fixed across frequencies. H2 retains diffusion modes 1–2 at times $1/16$ and $1/4$; sixteen equally spaced landmarks give the distance vector F; H2+F adds the raw critics, and H2+$F^{\perp}$ replaces them by pooled regression residuals on the four coordinates at the first diffusion time, a basis for the H2 span. H6 retains modes 1–6 at both times, matching the residual combination's 24 scalar coordinates; Band separates modes 1–2 and 3–6 at each time into four trace-standardized channels, also with 24 coordinates; Ridge keeps H2 and adds the twelve H6 coordinates at time $1/16$, transformed by $(\widehat{\Sigma}_X + \lambda_n I)^{-1/2}$, where $\widehat{\Sigma}_X$ is their pooled covariance and $\lambda_n = \mathrm{tr}(\widehat{\Sigma}_X)/(12\sqrt{n})$; it has 20 coordinates and three channels. Section B.5 gives the full fitting rules and timing protocol.

For $j = 3$, residualization raises power from 0.054 to 0.996 (paired gain 0.942, standard error 0.010) and reduces unconditional MAE from 0.232 to 0.029; H6 contains the changing harmonic but has power only 0.170, while Band, with the same coordinates split into frequency bands, reaches power 0.996 and MAE 0.028, isolating the effect of variance allocation, at a median computation time of 5.9 ms against 7.2 ms for residualization. Amplitude and frequency show how the variance gain

Table 2: Spectral rejection frequencies. $j = 0$ is the uniform null; 3_{weak} has half amplitude. The largest binomial Monte Carlo standard error is 0.022.

j	H2	H2+F	H2+$F^{\perp}$	H6	Band	Ridge
0	0.054	0.050	0.044	0.054	0.042	0.054
1	0.998	0.996	0.992	0.998	0.988	0.998
2	0.894	0.808	0.806	0.736	0.768	0.894
3	0.050	0.054	0.996	0.170	0.996	0.226
4	0.060	0.060	0.072	0.064	0.344	0.186
5	0.044	0.048	0.086	0.048	0.044	0.082
3_{weak}	0.056	0.052	0.454	0.056	0.480	0.056

translates into power: at half amplitude, Band has power 0.480 against 0.454 for residualization (paired difference 0.026, standard error 0.011) with MAEs 0.122 and 0.131; for $j = 4$, Band has power 0.344 against 0.072 because distance cones have no even-frequency mean contrast; at $j = 2$, adding the residual channel reduces power relative to H2 from 0.894 to 0.806; and the fifth harmonic remains weak, with residual and Ridge powers 0.086 and 0.082 but MAEs 0.185 and 0.104 (Table 8 reports every location error and total time). Frequency bands are economical when spectral coordinates are available, and pooled projection gives a covariance-based way to remove shared variation from distance coordinates; their relative value depends on the representation and changing direction, not just feature count.

Table 3: Location accuracy and full computation time (median, ms) for the third-harmonic comparison, smaller being better, with the same R, n and B as Table 2.

	$\varepsilon = 0.35$		$\varepsilon = 0.175$	
Method	Location MAE	Time (ms)	Location MAE	Time (ms)
H2	0.237	4.6	0.247	4.6
H2+F	0.232	6.2	0.247	6.2
H2+$F^{\perp}$	0.029	7.2	0.131	7.2
H6	0.092	5.8	0.214	5.8
Band	0.028	5.9	0.122	6.2
Ridge	0.070	6.0	0.212	6.6

Timings in Table 3 use an AMD Ryzen Threadripper 3990X with one thread for NumPy and C++ and cover construction, fitting, permutations, scanning and the decision, excluding data generation and one-time compilation; residualization adds about 1 ms to raw aggregation. For q independent diffusion coordinates and m distance features, canonical sorting costs $O(n(q+m)\log n)$, pooled projection $O(nq^2 + nqm + q^3)$, and the residual scan $O((B+1)nm)$ arithmetic and $O(nm)$ storage, separating representation construction from repeated scanning.

5 Conclusion

Geometric representations determine whether a distributional change is visible and how much variation competes with it. The joint diffusion–distance scan supplies calibrated inference, and its fixed-channel drift and variance formulas explain detection and localization; pooled projection preserves distance contrasts absent from a retained diffusion span and removes shared variance, and the spectral comparisons show that frequency grouping achieves the same principal benefit at lower cost. Features should thus be chosen by retained contrast, variance allocation and joint search cost.

Pooled circle representations also admit fixed-change consistency through convergence of their landmarks and retained spectral spaces; extending this to local alternatives requires rates for the fitted feature bridges. For dependent observations, block resampling is a computational starting point, and calibrating the search across geometric scales would then require control of the joint long-run covariance.

AI USE STATEMENT

Generative AI tools were used to revise wording and improve the clarity and readability of the manuscript; to identify potentially relevant literature, references, and search directions, and to assist with research execution, including drafting, debugging, and refining code. For mathematical proofs, AI was used primarily to review and check the authors' arguments for potential errors, inconsistencies, or gaps. AI-assisted manuscript revisions, literature-search suggestions, feedback on research ideas and proofs, and generated or modified code were reviewed and verified by the authors before incorporation into the final work. The authors independently assessed the accuracy and appropriateness of these outputs and retain full responsibility for the methodology, proofs, implementation, analysis, and conclusions presented in the paper.

REPRODUCIBILITY STATEMENT

The code and an accompanying `README` file are submitted with the manuscript. We provide a dedicated reproduction script for each figure and table, allowing the reported results to be reproduced individually. The `README` provides instructions for running the code.

REFERENCES

A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York, 1980.

S. Arlot, A. Celisse, and Z. Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019.

A. Aue, R. Gabrys, L. Horváth, and P. Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 2009.

B. M. Brown. Martingale central limit theorems. *Annals of Mathematical Statistics*, 42(1):59–66, 1971.

H. Chen and N. Zhang. Graph-based change-point detection. *Annals of Statistics*, 43(1):139–176, 2015.

X. Cheng and Y. Xie. Kernel two-sample tests for manifold data. *Bernoulli*, 30(4):2572–2597, 2024.

K. C. Cheng, S. Aeron, M. C. Hughes, E. Hussey, and E. L. Miller. Optimal transport based change point detection and time series segment clustering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6034–6038, 2020.

A. Chatterjee and A. Ramdas. A martingale kernel two-sample test. In *Proceedings of the 37th International Conference on Algorithmic Learning Theory*, volume 313 of *Proceedings of Machine Learning Research*, pages 1–44, 2026.

R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley, Chichester, 1997.

P. Dubey and H.-G. Müller. Fréchet change point detection. *Annals of Statistics*, 48(6):3312–3335, 2020.

P. Dubey and M. Zheng. Change point inference for non-Euclidean data sequences using distance profiles. arXiv:2311.16025v2, 2026.

A. A. Ezzat, M. Jun, and Y. Ding. Spatio-temporal short-term wind forecast: A calibrated regime-switching method. *The Annals of Applied Statistics*, 13(3):1484–1510, 2019.

D. Garreau and S. Arlot. Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2):4440–4486, 2018.

A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

A. Grigor'yan. *Heat Kernel and Analysis on Manifolds*. AMS/IP Studies in Advanced Mathematics 47, American Mathematical Society, Providence, 2009.

P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980.

Z. Harchaoui, E. Moulines, and F. Bach. Kernel change-point analysis. In *Advances in Neural Information Processing Systems 21*, pages 609–616, 2008.

J. Hemerik and J. Goeman. Exact testing with random permutations. *TEST*, 27:811–825, 2018.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, New York, 2012.

F. Jiang, C. Zhu, and X. Shao. Two-sample and change-point inference for non-Euclidean valued time series. *Electronic Journal of Statistics*, 18(1):848–894, 2024.

R. Kanrar, F. Jiang, and Z. Cai. Model-free change-point detection using AUC of a classifier. *Journal of Machine Learning Research*, 26(190):1–50, 2025.

A. N. Kolmogorov and V. M. Tikhomirov. ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Matematicheskikh Nauk*, 14(2):3–86, 1959.

D. S. Matteson and N. A. James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

A. N. Pettitt. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society, Series C*, 28(2):126–135, 1979.

B. Phipson and G. K. Smyth. Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, 9(1), Article 39, 2010.

J. W. Portegies. Embeddings of Riemannian manifolds with heat kernels and eigenfunctions. *Communications on Pure and Applied Mathematics*, 69(3):478–518, 2016.

S. Rosenberg. *The Laplacian on a Riemannian Manifold*. Cambridge University Press, Cambridge, 1997.

A. Schrab, I. Kim, M. Albert, B. Laurent, B. Guedj, and A. Gretton. MMD aggregated two-sample test. *Journal of Machine Learning Research*, 24(194):1–81, 2023.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, 41(5):2263–2291, 2013.

B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561, 2010.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996.

C. Villani. *Optimal Transport: Old and New*. Springer, Berlin, 2009.

S. Wei and Y. Xie. Online kernel CUSUM for change-point detection. *Journal of the Royal Statistical Society, Series B*, article qkag020, 2026.

C. K. I. Williams and M. Seeger. Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems 13*, pages 682–688, 2001.

APPENDIX

The appendix gives the complete algorithm and computation rules in Section A, followed by experimental designs and supporting results in Section B. The geometric identities, fixed-channel limits, residual mechanism and further process results are in Sections C to E. Proofs follow in Sections F to J.

A ALGORITHM AND GEOMETRIC SCAN DETAILS

Algorithm 1 Geometry-aware multichannel change-point test

Require: Ordered data $Z_{1:n}$; trim η; finite diffusion grid $\mathcal{T}$; critic dictionary F_m; channel weights; number of random permutations B; level α.

1: For every $t_\ell \in \mathcal{T}$, evaluate the diffusion Gram matrix $k_{t_\ell}(Z_i, Z_j) = p_{2t_\ell}(Z_i, Z_j)$ or construct finite diffusion features; compute the corresponding bridge norms by cumulative sums.
2: Evaluate $F_m(Z_i)$, compute $\widehat{\Sigma}_F$, and choose $A_{n,F}$ as in (17) or another declared permutation-invariant matrix.
3: Compute T_n from (4) and $\widehat{\tau}_n$ from (18).
4: **for** $b = 1, \dots, B$ **do**
5: Draw a uniform random permutation π_b of $\{1, \dots, n\}$.
6: Reorder the data as $Z_{\pi_b(1)}, \dots, Z_{\pi_b(n)}$; refit any order-dependent tuning rule and reuse pooled permutation-invariant tuning.
7: Compute $T_n^{(b)}$ using the resulting features and standardizers.
8: **end for**
9: Set $\widehat{p}_B = (1 + \sum_{b=1}^{B} \mathbf{1}\{T_n^{(b)} \geq T_n\})/(B+1)$.
10: Reject H_0 when $\widehat{p}_B \leq \alpha$; report $\widehat{\tau}_n$ only together with the rejection decision and channelwise scan curves.

A.1 DIFFUSION-SEMIGROUP CHANNEL

For the diffusion construction, let $(\mathcal{M}, g)$ be a closed d-dimensional Riemannian manifold without boundary, with Riemannian volume measure μ. Let $-\Delta$ be the nonnegative Laplace–Beltrami operator with eigenpairs

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots, \qquad \{\varphi_j : j \geq 0\} \text{ orthonormal in } L^2(\mathcal{M}, \mu).$$

The heat kernel and diffusion semigroup are

$$p_t(x, y) = \sum_{j=0}^{\infty} e^{-t\lambda_j} \varphi_j(x)\varphi_j(y), \qquad (H_t f)(x) = \int_{\mathcal{M}} p_t(x, y) f(y)\, \mu(\mathrm{d}y). \tag{12}$$

For $t > 0$, define

$$\Phi_t(x) = p_t(x, \cdot) \in L^2(\mathcal{M}, \mu), \qquad k_t(x, y) = p_{2t}(x, y). \tag{13}$$

The semigroup property gives $\langle \Phi_t(x), \Phi_t(y) \rangle_{L^2} = k_t(x, y)$. For a probability measure P, its diffusion mean embedding is

$$m_{P,t} = \int_{\mathcal{M}} \Phi_t(x)\, P(\mathrm{d}x) \in L^2(\mathcal{M}, \mu).$$

The corresponding discrepancy is

$$\mathrm{MMD}_t(P, Q) = \|m_{P,t} - m_{Q,t}\|_{L^2}.$$

The diffusion CUSUM is $\mathbb{B}_{n,t} = \mathbb{B}_{n,\Phi_t}$. At split k, let

$$\widehat{P}_{1:k} = k^{-1} \sum_{i=1}^{k} \delta_{Z_i}, \qquad \widehat{P}_{k+1:n} = (n-k)^{-1} \sum_{i=k+1}^{n} \delta_{Z_i}.$$

Then $\|\mathbb{B}_{n,t}(r_k)\|$ is a deterministic rescaling of $\mathrm{MMD}_t(\widehat{P}_{1:k}, \widehat{P}_{k+1:n})$; the exact formula is stated in Proposition C.2.

Diffusion changes both the mean contrast and the overall size of the feature vectors. To compare scales, normalize by the pooled feature variance. Define

$$\overline{\Phi}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} \Phi_t(Z_i), \qquad \widehat{v}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} \left\|\Phi_t(Z_i) - \overline{\Phi}_{n,t}\right\|_{L^2}^2 .$$

With the convention $0/0 = 0$, the standardized diffusion scan is

$$S_{n,t}^{H} = \max_{k\in\mathcal{K}_n(\eta)} \frac{\|\mathbb{B}_{n,t}(r_k)\|_{L^2}^2}{\widehat{v}_{n,t}\, r_k(1-r_k)}. \tag{14}$$

A finite diffusion grid $\mathcal{T} = \{t_1, \ldots, t_L\}$ is used in the main procedure. A fixed grid is the simplest theoretical choice. A pooled-data grid, such as multiples of the squared median pairwise geodesic distance, is also valid for permutation calibration because it is invariant to temporal reordering.

A.2 Distance critics and the full transport statistic

Assume in this subsection that $(\mathcal{X}, d)$ is compact. Fix an anchor $x_\star \in \mathcal{X}$ and define the anchored unit Lipschitz ball

$$\mathcal{F}_{\mathrm{Lip}} = \{f : \mathcal{X} \to \mathbb{R} : f(x_\star) = 0,\ |f(x) - f(y)| \le d(x, y)\}.$$

Anchoring removes additive constants and implies the uniform envelope $|f(x)| \le \mathrm{diam}(\mathcal{X})$. For $f \in \mathcal{F}_{\mathrm{Lip}}$, define

$$\mathbb{G}_n(r, f) = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{\lfloor nr \rfloor} f(Z_i) - \frac{\lfloor nr \rfloor}{n} \sum_{i=1}^{n} f(Z_i) \right\}. \tag{15}$$

Kantorovich–Rubinstein duality gives an exact empirical W_1 identity at every split (Villani, 2009); see Proposition C.2.

The full critic scan is

$$S_n^{W,\infty} = \max_{k\in\mathcal{K}_n(\eta)} \frac{\sup_{f\in\mathcal{F}_{\mathrm{Lip}}} |\mathbb{G}_n(r_k, f)|}{\sqrt{r_k(1-r_k)}}.$$

Its null weak limit requires an empirical-process condition. The primary computational version instead uses a finite dictionary $F_m = (f_1, \ldots, f_m)^\top$ with $f_j \in \mathcal{F}_{\mathrm{Lip}}$ for each j, and

$$\mathbb{G}_{n,F}(r) = \big(\mathbb{G}_n(r, f_1), \ldots, \mathbb{G}_n(r, f_m)\big)^\top.$$

The coordinatewise dictionary scan

$$S_n^{F,\max} = \max_{k\in\mathcal{K}_n(\eta)} \frac{\max_{1\le j\le m} |\mathbb{G}_n(r_k, f_j)|}{\sqrt{r_k(1-r_k)}} \tag{16}$$

is the direct finite-dictionary approximation to the empirical W_1 scan. For covariance-adaptive aggregation, let

$$\overline{F}_n = n^{-1}\sum_{i=1}^{n} F_m(Z_i), \qquad \widehat{\Sigma}_F = n^{-1}\sum_{i=1}^{n} \{F_m(Z_i) - \overline{F}_n\}\{F_m(Z_i) - \overline{F}_n\}^\top.$$

For a nonnegative-definite, permutation-invariant matrix $A_{n,F}$, define

$$S_n^F(A_{n,F}) = \max_{k\in\mathcal{K}_n(\eta)} \frac{\mathbb{G}_{n,F}(r_k)^\top A_{n,F} \mathbb{G}_{n,F}(r_k)}{r_k(1-r_k)}.$$

Two choices are useful:

$$A_{n,F}^{\mathrm{tr}} = \begin{cases} I_m/\operatorname{tr}(\widehat{\Sigma}_F), & \operatorname{tr}(\widehat{\Sigma}_F) > 0, \\ 0_{m\times m}, & \operatorname{tr}(\widehat{\Sigma}_F) = 0, \end{cases} \quad \text{and} \quad A_{n,F}^{\mathrm{Mah}} = \widehat{\Sigma}_F^{\dagger}, \tag{17}$$

where † denotes the Moore–Penrose inverse. The trace choice is stable and detects any nonzero dictionary mean shift. The Mahalanobis choice is invariant to nonsingular linear reparameterization of the critic vector, but its asymptotic analysis requires convergence of the generalized inverse. Positive definiteness is therefore not imposed globally; rank stability is enough when Mahalanobis whitening is used.

A simple geometry-aware dictionary uses landmarks $a_1, \dots, a_m$:

$$f_j(x) = d(x, a_j) - d(x_\star, a_j), \qquad j = 1, \dots, m.$$

Each function is anchored and one-Lipschitz, and its expectation contrast is the difference in mean distance to the landmark: $Pf_j - Qf_j = \mathbb{E}_P d(Z, a_j) - \mathbb{E}_Q d(Z, a_j)$. The trace-standardized scan aggregates the squared contrasts across landmarks, whereas the coordinatewise scan uses the largest absolute contrast. Thus the same dictionary provides both an aggregate measure of geometric change and a search for an individual separating critic. Landmarks selected by a permutation-invariant rule from the pooled observations preserve exact randomization validity.

These constructions include familiar scans as exact special cases. A single diffusion scale gives a kernel mean CUSUM, the Euclidean heat kernel gives Gaussian radial-basis-function (RBF) MMD, and one scalar critic gives the ordinary scalar CUSUM. Together with the full-class empirical W_1 identity, these reductions connect the geometric features to established discrepancies; Section G.5 records their formulas.

A.3 Location criterion and computational details

Algorithm 1 gives the complete procedure. The formulas below specify its location output and the cumulative-sum implementation of its diffusion and critic scans.

The estimated change location is

$$\widehat{\tau}_n = \min \arg\max_{k \in \mathcal{K}_n(\eta)} \max \left\{ \max_{\ell} c_\ell \frac{\|\mathbb{B}_{n,t_\ell}(r_k)\|^2}{\widehat{v}_{n,t_\ell} r_k(1-r_k)},\ c_F \frac{\mathbb{G}_{n,F}(r_k)^\top A_{n,F} \mathbb{G}_{n,F}(r_k)}{r_k(1-r_k)} \right\}. \tag{18}$$

The smallest maximizer convention makes the estimator total in the presence of ties.

Given the kernel values, the exact diffusion scan costs $O(Ln^2)$ per ordering. With permutation-invariant tuning, Gram matrices are formed once and reordered for each of the B permutations, giving $O((B+1)Ln^2)$ scan arithmetic. Storing every scale simultaneously costs $O(Ln^2)$ memory, whereas processing scales sequentially uses $O(n^2)$ working memory. Cumulative block sums give all split values without recomputing segment MMDs from scratch; Section A.4 gives the identity. A finite dictionary costs $O(mn)$ function evaluations and $O(m^2 n)$ arithmetic for a dense covariance matrix. Trace standardization and each trace-standardized scan require only $O(mn)$ arithmetic. Exact empirical W_1 may replace the dictionary for small samples, but it requires solving a transport problem at every split and permutation.

Finite-feature computation. Finite diffusion features give the same computational form as the critic vector. Once feature vectors $X_i \in \mathbb{R}^q$ have been constructed, replace $\Phi_t(Z_i)$ by X_i in (14). Set $\overline{X} = n^{-1} \sum_i X_i$ and $u_k = \sum_{i=1}^k (X_i - \overline{X})$. The criterion at split k is

$$\frac{\|u_k\|^2}{n \widehat{v}\, r_k(1-r_k)}, \qquad \widehat{v} = \frac{1}{n} \sum_{i=1}^n \left\|X_i - \overline{X}\right\|^2.$$

The update $u_{k+1} = u_k + X_{k+1} - \overline{X}$ evaluates all splits in $O(nq)$ arithmetic. If the feature construction is permutation invariant, both $\overline{X}$ and $\widehat{v}$ are reused, and each permutation only reorders the feature rows. For L diffusion channels of at most q coordinates and m trace-standardized critics, the scan arithmetic is therefore $O((B+1)n(Lq+m))$ after feature construction. This separates the cost of representing geometry from the cost of searching temporal splits.

The implementation in Section B.1 uses Nyström features. At a given diffusion time, form the matrix of kernel values between observations and selected landmarks, and the Gram matrix among those landmarks. Projecting onto retained eigenvectors of the landmark Gram matrix yields a finite feature vector whose inner products approximate the heat kernel. With permutation-invariant feature

construction, all observed and permuted scans use this same representation and Theorem 3.1 calibrates the computed statistic. The approximation determines which feature-mean changes that statistic can resolve. A reproducible way to make pooled landmark calculations invariant is to sort the locations by a fixed coordinate representation before those calculations, then restore temporal row order in the feature matrix; identical locations yield identical features.

A.4 Exact Gram-matrix computation

Let $K^{(t)}$ be the $n \times n$ Gram matrix with $K^{(t)}_{ij} = p_{2t}(Z_i, Z_j)$. For split k, define

$$a_{i,k} = \begin{cases} 1 - r_k, & i \le k, \\ -r_k, & i > k. \end{cases}$$

Then

$$\|\mathbb{B}_{n,t}(r_k)\|^2_{L^2} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,k} a_{j,k} K^{(t)}_{ij}.$$

If

$$A_k = \sum_{i,j \le k} K^{(t)}_{ij}, \qquad B_k = \sum_{i \le k < j} K^{(t)}_{ij}, \qquad C_k = \sum_{i,j > k} K^{(t)}_{ij},$$

then

$$\|\mathbb{B}_{n,t}(r_k)\|^2_{L^2} = \frac{1}{n}\{(1 - r_k)^2 A_k - 2r_k(1 - r_k)B_k + r_k^2 C_k\}. \tag{19}$$

The block sums can be updated from cumulative row and column sums, so all candidate splits at one diffusion scale require $O(n^2)$ arithmetic after the Gram matrix is formed. Equation (19) evaluates the segment discrepancy in (23) using the pooled Gram matrix, so each ordering requires only cumulative block sums.

A.5 Calibration rules

Permutation calibration is the default because it is finite-sample valid under (1), accommodates a maximum over correlated channels, and does not require estimating the eigenvalues of an infinite-dimensional covariance operator. The plus-one Monte Carlo p-value in Algorithm 1 is conservative for every B (Phipson and Smyth, 2010) and becomes exact on its attainable grid after randomized tie breaking; see Theorem 3.1 and Hemerik and Goeman (2018).

For large n, a multiplier bridge avoids repeated reordering. Let $\xi_1, \ldots, \xi_n$ be iid multipliers independent of the data, with mean zero, variance one, and a finite $2 + \delta$ moment. For Hilbert-valued scores X_i with sample mean $\overline{X}_n = n^{-1} \sum_i X_i$, define

$$\mathbb{B}^{\#}_n(r) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{\lfloor nr \rfloor} \xi_i (X_i - \overline{X}_n) - \frac{\lfloor nr \rfloor}{n} \sum_{i=1}^{n} \xi_i (X_i - \overline{X}_n) \right]. \tag{20}$$

Using the same multipliers across all channels retains their joint dependence. Conditional consistency is established in Theorem C.9 for a fixed finite channel collection. Permutation remains preferable when exact finite-sample level is required.

B Experimental designs and supporting comparisons

The primary study compares 27 alternatives on $\mathbb{S}^1$, $\mathbb{S}^2$ and $\mathbb{T}^2$, with nine iid-null settings. Six sample-size and six unequal-segment settings probe how signal strength affects localization. Each design uses $R = 500$ repetitions. The energy and spectral comparisons use paired samples and common permutations to compare both detection and unconditional location error.

B.1 Design and implementation

On $\mathbb{S}^1$, the baseline laws were von Mises distributions. We write $\mathrm{vM}(\mu, \kappa)$ for the circular law with density proportional to $\exp\{\kappa \cos(z - \mu)\}$, where μ is the mean direction and κ is concentration.

Location changes shifted $\mathrm{vM}(0,4)$ by $\delta \in \{0.15, 0.30, 0.60\}$, and concentration changes replaced $\mathrm{vM}(0,8)$ by $\mathrm{vM}(0,\kappa)$ for $\kappa \in \{4,2,1\}$. Let $A_1(\kappa) = I_1(\kappa)/I_0(\kappa)$, where I_j is the modified Bessel function of the first kind of order j. To form matched-moment shape changes, we compared $P = \mathrm{vM}(0,4)$ with

$$Q_\kappa = \tfrac{1}{2}\,\mathrm{vM}(a_\kappa,\kappa) + \tfrac{1}{2}\,\mathrm{vM}(-a_\kappa,\kappa), \qquad a_\kappa = \arccos\{A_1(4)/A_1(\kappa)\},$$

for $\kappa \in \{6,8,12\}$. Hence P and Q_κ have the same first circular moment.

On $\mathbb{S}^2$, the von Mises–Fisher law $\mathrm{vMF}(\mu,\kappa)$ has density proportional to $\exp(\kappa\mu^\top z)$ for unit mean direction μ and concentration κ. Write e_1, e_2 for the first two coordinate unit vectors and $A_3(\kappa) = \coth(\kappa) - \kappa^{-1}$. Location changes rotated the mean direction of $\mathrm{vMF}(e_1,8)$ by $a \in \{0.15, 0.30, 0.60\}$, and concentration changes used $\kappa \in \{4,2,1\}$. The matched-mean shape alternative used the symmetric mixture

$$Q_\kappa = \tfrac{1}{2}\,\mathrm{vMF}(\mu_{\kappa,+},\kappa) + \tfrac{1}{2}\,\mathrm{vMF}(\mu_{\kappa,-},\kappa), \quad \mu_{\kappa,\pm} = \cos(a_\kappa)e_1 \pm \sin(a_\kappa)e_2,$$

where $a_\kappa = \arccos\{A_3(8)/A_3(\kappa)\}$, and $\kappa \in \{12,16,24\}$. This mixture and $\mathrm{vMF}(e_1,8)$ have the same extrinsic mean.

On $\mathbb{T}^2$, we used wrapped bivariate normal laws. The law $\mathrm{WN}_2(\mu,\Sigma)$ reduces each coordinate of a bivariate normal with mean μ and covariance Σ modulo 2π. Location changes shifted the first coordinate of $\mathrm{WN}_2(0, 0.4^2 I_2)$ by $\delta \in \{0.15, 0.30, 0.60\}$. Dependence changes replaced $0.6^2 I_2$ by $0.6^2 \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ for $\rho \in \{0.25, 0.50, 0.75\}$ while preserving the wrapped marginals. For matched marginal first moments, $P = \mathrm{WN}_2(0, 0.7^2 I_2)$ was compared with

$$Q_\sigma = \tfrac{1}{2}\,\mathrm{WN}_2((a_\sigma,a_\sigma)^\top, \sigma^2 I_2) + \tfrac{1}{2}\,\mathrm{WN}_2((-a_\sigma,-a_\sigma)^\top, \sigma^2 I_2),$$
$$a_\sigma = \arccos\left[\exp\{-(0.7^2 - \sigma^2)/2\}\right], \qquad \sigma \in \{0.20, 0.35, 0.50\}.$$

As σ increases toward the baseline value 0.7, a_σ decreases to zero and Q_σ converges to P along this specified mixture path. Levels 1–3 therefore follow this path toward the baseline.

The 27 primary alternative cells used $n = 500$, $\theta = 1/2$, trimming $\eta = 0.10$, and the three parameter levels above. Nine null cells crossed the three manifolds with $n \in \{200, 500, 1000\}$. Six sample-size cells completed the middle matched-shape design at $n = 200$ and 1000, with its $n = 500$ value supplied by the primary grid. Six unequal-segment cells used $\theta \in \{0.25, 0.75\}$.

The exact theoretical diffusion channel in Section A.1 was approximated numerically. For each pooled sample, we selected up to 128 medoid–farthest landmarks and constructed Nyström features of effective rank at most 128 at

$$\widehat{d}^2_{\mathrm{med}}\{2^{-6}, 2^{-4}, 2^{-2}, 1, 2^2\}, \qquad \widehat{d}^2_{\mathrm{med}} = \mathrm{median}_{i<j}\, d(Z_i, Z_j)^2.$$

The pooled landmark rule starts from the medoid minimizing the sum of geodesic distances and then repeatedly takes a point maximizing its distance to the selected set. Coordinates are sorted lexicographically before summing distances and resolving ties; the selected features are then returned to temporal order. Canonical ordering matters even for continuous samples: an even circle sample in an arc shorter than a semicircle has two central observations with equal medoid objectives.

At diffusion time t_ℓ, let K_{nm} contain kernel evaluations p_{2t_ℓ} between all observations and the landmarks, and let K_{mm} be the landmark Gram matrix. If U and Λ contain its retained eigenvectors and eigenvalues, the Nyström feature matrix is $K_{nm} U \Lambda^{-1/2}$. The absolute and relative eigenvalue thresholds retain modes above $\max\{10^{-12}, 10^{-10}\lambda_{\max}(K_{mm})\}$, up to rank 128. The base diffusion-series truncation tolerance was 10^{-10}; each circle factor of the torus product used one quarter of this tolerance.

We used 64 critics $f_j(z) = d(z, a_j) - d(a_1, a_j)$ from the same pooled landmark sequence. The combined statistic is the maximum of the five trace-standardized diffusion scans and the trace-standardized critic scan, with all numerical channel weights fixed at $c_1 = \cdots = c_5 = c_F = 1$. The extrinsic mean scan applies the feature map

$$e(z) = \begin{cases} (\cos z, \sin z), & \mathbb{S}^1, \\ z, & \mathbb{S}^2 \subset \mathbb{R}^3, \\ (\cos z_1, \sin z_1, \cos z_2, \sin z_2), & \mathbb{T}^2, \end{cases}$$

and maximizes $\|\mathbb{B}_{n,e}(r_k)\|^2/[\mathrm{tr}(\widehat{\Sigma}_e)r_k(1-r_k)]$ over the same trimmed splits, using the smallest maximizer. Here $\widehat{\Sigma}_e$ is the pooled covariance of $e(Z_i)$. Every method used the identical temporal permutation stream and plus-one calibration. The null rejection frequencies assess calibration, while the alternative designs probe sensitivity to changes in the retained feature means.

Alternative cells used $B = 199$ random permutations and the plus-one p-value. Null cells used $B = 999$, except the circle cells at $n = 200$ and 500, which used $B = 4999$. Exact empirical W_1 scans were computed only on the circle for $n \leq 500$, using the same samples, trimming interval, geodesic cost, and first 199 permutations.

In the tables, Heat denotes the maximum over the five diffusion scans, Critic denotes the trace-standardized critic scan, and Extrinsic denotes the extrinsic-mean scan.

B.2 Calibration and primary power

Table 4: Rejection frequencies under the iid null at nominal level 0.05. Every cell has $R = 500$ and zero failures. Circle cells at $n = 200, 500$ use $B = 4999$; other cells use $B = 999$. Exact W_1 uses the first 199 permutations.

Geometry	n	Combined	Heat	Critic	Extrinsic	Exact W_1
circle	200	0.040	0.042	0.038	0.040	0.048
circle	500	0.042	0.042	0.032	0.040	0.038
circle	1000	0.046	0.046	0.048	0.046	–
sphere	200	0.040	0.042	0.040	0.042	–
sphere	500	0.050	0.050	0.052	0.050	–
sphere	1000	0.052	0.050	0.048	0.050	–
torus	200	0.044	0.046	0.042	0.044	–
torus	500	0.054	0.052	0.048	0.052	–
torus	1000	0.052	0.054	0.050	0.052	–

Combined rejection frequencies range from 0.040 to 0.054. The largest binomial Monte Carlo standard error is 0.0101 among these null entries and 0.0221 across the primary rejection-frequency entries. The table reports the individual channels as well as their joint maximum.

Location changes are readily detected at the larger shifts, whereas shape and dependence changes separate the representations. Heat supplies most of the power against the stronger sphere and torus shape alternatives. Along the torus mixture path, increasing the parameter level brings the law toward the baseline and reduces power. The full circle transport statistic weights changes differently: for $\kappa = 8 \to 4$, its power exceeds Combined by 0.222, with paired Monte Carlo standard error 0.0186, but Combined is more sensitive at the strongest circle shape setting.

B.3 Localization metrics

The estimand is the change fraction $\theta = \tau/n$, and the estimator is $\widehat{\theta} = \widehat{\tau}_n/n$ from (18). For repetition r, define

$$e_r = |\widehat{\theta}_r - \theta|, \qquad \widehat{C}_{0.02} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{1}\{e_r \leq 0.02\}, \qquad \widehat{\mathrm{MAE}} = \frac{1}{R}\sum_{r=1}^{R} e_r.$$

Thus $\widehat{C}_{0.02}$ estimates the probability of localization within 2% of the sequence length, called tolerance coverage in the tables. At $n = 500$, this means an error of at most 10 observation indices. Both metrics use all $R = 500$ repetitions, including those without rejection, and measure unconditional point-estimation accuracy. The procedure does not construct a confidence interval for θ. Their Monte Carlo standard errors (MCSEs) are $\{\widehat{C}_{0.02}(1-\widehat{C}_{0.02})/R\}^{1/2}$ and $\mathrm{sd}(e_1,\ldots,e_R)/\sqrt{R}$, respectively.

Strong location and concentration changes give small errors, whereas weak shape and dependence changes produce diffuse maximizers. Matched-mean sphere and torus changes can also be located accurately. The middle sphere shape MAE decreases from 0.166 at $n = 200$ to 0.013 at $n = 1000$,

Table 5: Primary rejection frequencies. Every row uses $n = 500$, $\theta = 1/2$, $R = 500$, $B = 199$ and has zero failures. Exact W_1 is computed only on the circle. Parameter levels follow Section B.1, not a common scale of difficulty across geometries.

Geometry	Change	Level	Combined	Heat	Critic	Extrinsic	Exact W_1
circle	location	1	0.704	0.704	0.708	0.706	0.704
circle	location	2	0.998	0.998	0.998	0.998	0.998
circle	location	3	1.000	1.000	1.000	1.000	1.000
circle	concentration	1	0.616	0.614	0.424	0.128	0.838
circle	concentration	2	1.000	1.000	1.000	1.000	1.000
circle	concentration	3	1.000	1.000	1.000	1.000	1.000
circle	shape	1	0.038	0.038	0.036	0.040	0.038
circle	shape	2	0.068	0.068	0.066	0.072	0.080
circle	shape	3	0.142	0.142	0.056	0.050	0.086
sphere	location	1	0.928	0.928	0.922	0.930	–
sphere	location	2	1.000	1.000	1.000	1.000	–
sphere	location	3	1.000	1.000	1.000	1.000	–
sphere	concentration	1	0.980	0.976	0.980	0.854	–
sphere	concentration	2	1.000	1.000	1.000	1.000	–
sphere	concentration	3	1.000	1.000	1.000	1.000	–
sphere	shape	1	0.062	0.064	0.064	0.060	–
sphere	shape	2	0.226	0.238	0.072	0.080	–
sphere	shape	3	0.990	0.990	0.052	0.040	–
torus	location	1	0.878	0.878	0.872	0.878	–
torus	location	2	1.000	1.000	1.000	1.000	–
torus	location	3	1.000	1.000	1.000	1.000	–
torus	dependence	1	0.056	0.058	0.054	0.058	–
torus	dependence	2	0.088	0.092	0.064	0.056	–
torus	dependence	3	0.668	0.734	0.072	0.068	–
torus	shape	1	1.000	1.000	0.142	0.060	–
torus	shape	2	1.000	1.000	0.106	0.062	–
torus	shape	3	0.196	0.224	0.048	0.046	–

Table 6: Unconditional combined-scan localization at $n = 500$ and $\theta = 1/2$, with $R = 500$ and zero failures. Levels follow Section B.1; torus shape moves toward the baseline. The largest Monte Carlo standard errors are 0.0224 for tolerance coverage and 0.0062 for MAE.

Geometry	Change	2% tolerance coverage ↑			MAE ↓		
		Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
circle	location	0.318	0.692	0.972	0.085	0.023	0.005
circle	concentration	0.498	0.896	0.942	0.057	0.009	0.005
circle	shape	0.044	0.084	0.298	0.242	0.215	0.115
sphere	location	0.478	0.864	0.988	0.052	0.010	0.003
sphere	concentration	0.664	0.944	0.982	0.029	0.005	0.003
sphere	shape	0.132	0.508	0.902	0.182	0.074	0.010
torus	location	0.392	0.828	0.996	0.067	0.011	0.003
torus	dependence	0.050	0.272	0.746	0.234	0.126	0.028
torus	shape	0.998	0.960	0.450	0.001	0.005	0.077

while the middle circle shape MAE remains 0.177 at $n = 1000$. A nonzero retained contrast ensures the fixed-change limit, but does not determine its finite-sample magnitude.

B.4 Energy-distance comparison

We compare with the single-change energy scan of Matteson and James (2014), fixing the second segment's endpoint at n and the distance exponent at one. With $d^e_{ij} = \|e(Z_i) - e(Z_j)\|_2$, its two-sample divergence at split k is

$$\widehat{\mathcal{V}}_k = \frac{2\sum_{i\le k<j} d^e_{ij}}{k(n-k)} - \frac{2\sum_{i<j\le k} d^e_{ij}}{k(k-1)} - \frac{2\sum_{k<i<j} d^e_{ij}}{(n-k)(n-k-1)}.$$

The statistic maximizes $k(n-k)\widehat{\mathcal{V}}_k/n$ over $\mathcal{K}_n(0.10)$ and returns the smallest maximizer. Euclidean distances between the standard embedded coordinates use the same observations as the geometric scan. They retain distributional information beyond the embedded mean. For negative-type distances, energy discrepancies correspond to MMD with an induced kernel (Sejdinovic et al., 2013); this comparison therefore probes a different geometric weighting of distributional contrasts. Each permutation recomputes the entire split maximum.

The comparison uses the three $n = 500$ null cells and the middle parameter of each of the nine primary alternative paths. All methods use the same 500 samples per cell and the same first 199 random permutations, with plus-one calibration at level 0.05. These settings cover location, concentration, shape and dependence with a common calibration budget.

Table 7: Energy and combined scans at $n = 500$, with $R = 500$ and $B = 199$. Difference is the energy rejection frequency minus the combined frequency, with its paired Monte Carlo standard error (MCSE). MAE uses all repetitions; C and E denote combined and energy. Alternative rows use the middle parameters in Section B.1.

Geometry	Change	Combined	Energy	Difference	MCSE	MAE (C)	MAE (E)
circle	null	0.046	0.026	-0.020	0.007	–	–
sphere	null	0.060	0.050	-0.010	0.008	–	–
torus	null	0.058	0.058	0.000	0.007	–	–
circle	location	0.998	0.998	0.000	0.000	0.023	0.024
circle	concentration	1.000	1.000	0.000	0.000	0.009	0.008
circle	shape	0.068	0.068	0.000	0.007	0.215	0.220
sphere	location	1.000	1.000	0.000	0.000	0.010	0.010
sphere	concentration	1.000	1.000	0.000	0.000	0.005	0.004
sphere	shape	0.226	0.256	0.030	0.017	0.074	0.121
torus	location	1.000	1.000	0.000	0.000	0.011	0.012
torus	dependence	0.088	0.132	0.044	0.010	0.126	0.150
torus	shape	1.000	1.000	0.000	0.000	0.005	0.012

B.5 Residual distance comparison

The circle spectral experiment follows Proposition D.1. Diffusion features are

$$(e^{-tk^2}\cos(kx), e^{-tk^2}\sin(kx))_{k=1}^{J}, \qquad t \in \{1/16, 1/4\}.$$

The constant mode and common factor $1/\sqrt{\pi}$ cancel from the centered trace-standardized scan. H2 uses $J = 2$ and H6 uses $J = 6$, giving 8 and 24 scalar coordinates across the two scales. The times give different weights within the retained span. Sixteen equally spaced landmarks supply the distance vector; anchor constants disappear upon centering. The residual procedure projects this vector onto the four coordinates at the first diffusion time, which span the same functions as both H2 channels. Ordinary least squares uses the pooled sample, with no change labels or population moments. Covariance calculations use canonical row ordering. A failed Cholesky factorization sets the projection coefficient to zero, and a zero-variance channel contributes zero.

For each $j = 1, \dots, 5$, the segment densities are $(1 \pm \varepsilon\cos(jx))/(2\pi)$ with $\varepsilon = 0.35$; $j = 0$ denotes the common uniform law. Each cell uses 500 independent samples of size $n = 500$, a midpoint

change, trimming 0.10 and 199 shared permutations, with the plus-one test at level 0.05. Rejection sampling draws each segment from its density. Location error uses every repetition, including nonrejections.

Band keeps the two H2 channels and adds modes 3–6 at each of the same two diffusion times. Each block is separately trace-standardized, giving four jointly calibrated channels and 24 scalar coordinates. Ridge keeps H2 and adds the 12 sine/cosine coordinates of H6 at $t = 1/16$, transformed using their pooled covariance $\widehat{\Sigma}_X$. For centered rows $X_i - \overline{X}$, its added channel uses

$$(\widehat{\Sigma}_X + \lambda_n I_{12})^{-1/2}(X_i - \overline{X}), \qquad \lambda_n = \frac{\operatorname{tr}(\widehat{\Sigma}_X)}{12\sqrt{n}}.$$

Thus Ridge has three channels and 20 coordinates. A zero covariance contributes a zero channel; otherwise the positive ridge makes the inverse well defined. Covariance moments use canonical pooled ordering and are fitted once for all permutations. This rule inherits finite-sample permutation validity from Theorem 3.1. Kernel Fisher change-point scans use regularized covariance inverses (Harchaoui et al., 2008); here we use a single pooled covariance and joint permutation calibration to compare representations under a common search. No rule selects the true harmonic.

All six methods use the same samples and permutation indices within each cell. A separate set of 500 third-harmonic samples halves the amplitude to 0.175. Every timing includes feature construction, pooled fitting where used, permutation generation and all scans through the p-value decision. Band's additional work uses the same spectral coordinates as H6 and separates their scan blocks. For 12 spectral coordinates, Ridge fits a 12×12 covariance and Cholesky factor before scanning the transformed rows.

Table 8: Unconditional location MAE and median total wall time in the spectral comparison. Total time includes feature construction, pooled projection where used, generation of all permutations, every scan and the test decision. Both metrics are smaller for better performance. The uniform null has no location target.

j	H2	H2+F	H2+$F^{\perp}$	H6	Band	Ridge
Unconditional location MAE						
1	0.027	0.027	0.027	0.027	0.027	0.027
2	0.042	0.044	0.044	0.049	0.047	0.042
3	0.237	0.232	0.029	0.092	0.028	0.070
4	0.250	0.248	0.254	0.191	0.077	0.082
5	0.250	0.246	0.185	0.239	0.214	0.104
3_{weak}	0.247	0.247	0.131	0.214	0.122	0.212
Median full time (ms)						
0	4.6	6.2	7.3	5.9	5.8	6.0
1	4.6	6.2	7.2	5.8	6.1	6.3
2	4.6	6.2	7.2	5.8	6.0	6.3
3	4.6	6.2	7.2	5.8	5.9	6.0
4	4.6	6.2	7.2	5.8	6.0	6.2
5	4.5	6.1	7.1	5.7	6.3	6.5
3_{weak}	4.6	6.2	7.2	5.8	6.2	6.6

For the third harmonic, the residual procedure's paired power improvement over the raw combination is 0.942 (Monte Carlo standard error 0.010), and its paired MAE improvement is 0.204 (standard error 0.006). Against H6, the corresponding improvements are 0.826 (0.017) and 0.063 (0.005). At the second harmonic, the residual procedure loses 0.088 power relative to H2 (standard error 0.013). These comparisons isolate recovered signal, the larger null search and location accuracy. At the fourth harmonic, the residual contrast is zero; H6 includes that mode but its diffusion weights still give a small signal relative to pooled variation. At the fifth harmonic, the smaller tail contrast yields weak power despite lower mean location error.

Timings in Table 8 use an AMD Ryzen Threadripper 3990X, with one CPU thread for NumPy and the C++ scan. Each method independently constructs its features and scans; the time for generating their

common permutations is charged to every method. Data generation and one-time library compilation are excluded. The full residual procedure adds approximately 1 ms to the raw combination in this setting. For q independent diffusion coordinates and m distance features, canonical sorting costs $O(n(q+m)\log n)$ and pooled projection costs $O(nq^2 + nqm + q^3)$ arithmetic, followed by $O((B+1)nm)$ residual scan arithmetic. Storing the residual rows requires $O(nm)$ memory.

Band and the residual procedure have identical rejection decisions at the original third harmonic. With half amplitude, Band's paired power advantage is 0.026 (Monte Carlo standard error 0.011) and its MAE improvement is 0.0097 (0.0029). For the fourth harmonic, these advantages are 0.272 (0.021) and 0.178 (0.006). Ridge attains power 0.226 at the original third harmonic and 0.056 at half amplitude, reflecting the different weighting of a channel that aggregates all six modes. At the fifth harmonic, residual and Ridge powers are 0.086 and 0.082, with MAE 0.185 and 0.104. These comparisons show how channel weighting and its calibration cost jointly determine sensitivity and localization, even when the retained spectral span is the same.

B.6 Sample size and unequal segment lengths

Table 9: Combined-scan performance for the middle matched-shape alternatives. Rows vary n at $\theta = 1/2$ or vary θ at $n = 500$. Each has $R = 500$, $B = 199$ and zero failures. Coverage is point-location accuracy within 0.02 of the true fraction; it is not confidence-interval coverage.

Geometry	Setting	Power	MCSE	Coverage	MCSE	MAE	MCSE
circle	$n = 200$	0.068	0.011	0.042	0.009	0.236	0.006
circle	$n = 1000$	0.066	0.011	0.146	0.016	0.177	0.006
circle	$\theta = 0.25$	0.036	0.008	0.074	0.012	0.254	0.010
circle	$\theta = 0.75$	0.052	0.010	0.098	0.013	0.269	0.010
sphere	$n = 200$	0.050	0.010	0.198	0.018	0.166	0.006
sphere	$n = 1000$	0.962	0.009	0.784	0.018	0.013	0.001
sphere	$\theta = 0.25$	0.074	0.012	0.430	0.022	0.140	0.009
sphere	$\theta = 0.75$	0.140	0.016	0.430	0.022	0.105	0.008
torus	$n = 200$	0.656	0.021	0.760	0.019	0.028	0.003
torus	$n = 1000$	1.000	0.000	0.996	0.003	0.002	0.000
torus	$\theta = 0.25$	0.996	0.003	0.950	0.010	0.008	0.002
torus	$\theta = 0.75$	1.000	0.000	0.960	0.009	0.005	0.000

Sphere and torus localization improves substantially as n increases. The circle change remains difficult: its power is 0.068 at $n = 200$ and 0.066 at $n = 1000$, despite the reduction in mean location error.

To distinguish weak retained signal from approximation loss in this circle design, we fit the original representation to 12 independent samples at each of $n = 200, 500, 1000$. For each fitted map, we integrate its feature means and pooled variance under the known laws: concentration 4 before the change and the matched-mean concentration-8 mixture afterward. The largest channel signal fraction in (9) lies between 0.00186 and 0.00190 across these fits. It comes from the second diffusion scale; the distance-channel fractions range from 0.00011 to 0.00017. Thus the strongest frozen-map drift contribution grows from about 0.38 at $n = 200$ to 1.9 at $n = 1000$. At the same fitted scales, every diffusion fraction differs from its exact Fourier counterpart by less than 0.5%. This diagnostic identifies weak signal relative to pooled variation, with little loss from the Nyström approximation in these fits.

The exact calculation uses the Fourier moments $I_j(4)/I_0(4)$ before the change and $I_j(8)\cos(ja)/I_0(8)$ afterward, where I_j is the modified Bessel function and a is the mean-matching displacement in Section B.1. Fitted-map moments use equally spaced midpoint quadrature with 2048 points; halving this grid changes every diffusion moment by less than 10^{-12}. These calculations evaluate the population contrast of each frozen fitted map, separately from the rejection frequencies in Table 9.

Moving the change to $\theta = 0.25$ or 0.75 leaves it within the same trimmed search. For fixed contrast and quadratic metric, the leading signal changes with $\theta(1-\theta)$ in (8); trace standardization also

responds to the pooled mixture variance in (9). The torus change retains accurate localization, whereas the weaker circle and sphere changes have larger mean errors than at the midpoint.

C SUPPORTING THEORETICAL RESULTS

The supporting theory starts with conditions for the geometric features and their standardizers. Exact identities relate the scans to distributional discrepancies; joint bridge limits then give calibration and localization. The residual circle mechanism follows in Section D; full critic classes and predictable scores are treated afterward in Section E.

C.1 CONDITIONS AND THEIR ROLES

The assumptions are separated according to the result that uses them.

The Hilbert bridge theorem needs a separable score space and $\mathbb{E}\,\|\psi(Z)\|^2 < \infty$. Bounded kernels and compact state spaces are convenient sufficient conditions, not intrinsic requirements. For a noncompact manifold, the same fixed-t diffusion CUSUM null limit follows whenever

$$\mathbb{E}_P p_{2t}(Z, Z) < \infty,$$

although characteristicness and uniform heat-kernel bounds then require manifold-specific hypotheses. On a closed manifold, the heat kernel is bounded and has a discrete spectral expansion (Rosenberg, 1997; Grigor'yan, 2009; Portegies, 2016). These properties yield the probability-measure separation in Proposition C.1.

The iid null supplies exchangeability for permutation calibration and the covariance of the Brownian-bridge limit. Under weak dependence, calibration must account for long-run covariance, for example through a justified block or dependent multiplier method.

A fixed finite grid permits a direct-sum Hilbert functional central limit theorem (FCLT) and joint multiplier bootstrap. The permutation theorem remains valid for a growing or data-dependent grid if the complete construction is permutation equivariant, but power and bootstrap validity for a growing grid need additional complexity control.

The bracketing entropy condition in Assumption E.1 controls the complexity of the full critic process. It holds for finite dictionaries, regular finite-dimensional parametric critic families, and bounded C^s balls on a compact d-manifold when $s > d/2$. This calculation verifies the full Lipschitz ball in intrinsic dimension one. Finite dictionaries provide a direct route to joint bridge limits in higher dimensions as well.

Positive definiteness of the critic covariance is unnecessary for the permutation test or the trace-standardized statistic. Write $\Sigma_F = \mathrm{Cov}_P\{F_m(Z)\}$. The Mahalanobis limit uses $\widehat{\Sigma}_F^{\dagger} \xrightarrow{\mathbb{P}} \Sigma_F^{\dagger}$. Covariance convergence with eventual rank stability is sufficient: the positive eigenvalues then converge to those of Σ_F, and inversion is continuous on the retained range. Estimated ranks and eigenvalues help diagnose numerical instability, while a ridge inverse offers regularization with a different limiting quadratic metric.

The logarithmic-clock theorem assumes bounded iid centered Hilbert features. Centering features from a bounded kernel, or differencing independent features with the same law, gives bounded mean-zero scores. Boundedness controls the fourth moments used to compare observed and predictable variation and verifies the martingale Lindeberg condition.

C.2 EXACT GEOMETRIC AND RANDOMIZATION PROPERTIES

Proposition C.1 (Diffusion embedding and separation)**.** *Let $(\mathcal{M}, g)$ be closed and let $t > 0$.*

(i) The map Φ_t in (13) *is continuous and bounded as a map from $\mathcal{M}$ into $L^2(\mathcal{M}, \mu)$, and*

$$\langle \Phi_t(x), \Phi_t(y) \rangle_{L^2} = p_{2t}(x, y).$$

Thus $k_t = p_{2t}$ is a bounded continuous positive-definite kernel.

(ii) For finite signed Borel measures ν and ρ,

$$\left\langle \int \Phi_t \, \mathrm{d}\nu, \int \Phi_t \, \mathrm{d}\rho \right\rangle_{L^2} = \iint p_{2t}(x,y) \, \nu(\mathrm{d}x)\rho(\mathrm{d}y).$$

For a signed measure ν, write $H_t\nu = \int \Phi_t \, \mathrm{d}\nu$. In particular,

$$\mathrm{MMD}_t^2(P,Q) = \|H_t(P-Q)\|_{L^2}^2 = \sum_{j=0}^{\infty} e^{-2t\lambda_j} \left\{ \int \varphi_j \, \mathrm{d}(P-Q) \right\}^2. \tag{21}$$

(iii) The kernel p_{2t} is characteristic: $\mathrm{MMD}_t(P,Q) = 0$ if and only if $P = Q$.

The positive-definite and semigroup identities are standard heat-kernel facts (Rosenberg, 1997; Grigor'yan, 2009); the proposition records their statistical implication. Characteristicness follows because the diffusion operator has strictly positive spectral multipliers $e^{-t\lambda_j}$ and the Laplace eigenfunctions determine finite measures on a closed manifold. Hence smoothing does not destroy identifiability at any fixed $t > 0$, although large t can reduce finite-sample power against high-frequency alternatives.

Proposition C.2 (Exact split identities)**.** *For every $k = 1, \ldots, n-1$,*

$$\mathbb{B}_{n,t}(r_k) = \sqrt{n}\, r_k(1-r_k)\{m_{\widehat{P}_{1:k},t} - m_{\widehat{P}_{k+1:n},t}\}, \tag{22}$$

$$\|\mathbb{B}_{n,t}(r_k)\|_{L^2} = \sqrt{n}\, r_k(1-r_k)\, \mathrm{MMD}_t(\widehat{P}_{1:k}, \widehat{P}_{k+1:n}), \tag{23}$$

$$\sup_{f \in \mathcal{F}_{\mathrm{Lip}}} |\mathbb{G}_n(r_k, f)| = \sqrt{n}\, r_k(1-r_k) W_1(\widehat{P}_{1:k}, \widehat{P}_{k+1:n}). \tag{24}$$

At every candidate split, the scan is exactly a weighted two-sample discrepancy. The factor $r(1-r)$ balances segment sizes and produces the triangular alternative drift used below. Proofs of the geometric identities and permutation validity are given in Section G.

Proposition C.3 (Finite critic approximation to W_1)**.** *Let $\{f_1, \ldots, f_m\} \subset \mathcal{F}_{\mathrm{Lip}}$ be an ε_m-net of $\mathcal{F}_{\mathrm{Lip}}$ in the supremum norm. Then, for any probability measures P, Q on $\mathcal{X}$,*

$$0 \le W_1(P,Q) - \max_{1 \le j \le m} |Pf_j - Qf_j| \le 2\varepsilon_m. \tag{25}$$

The same inequality holds simultaneously at every empirical split after replacing P, Q by $\widehat{P}_{1:k}, \widehat{P}_{k+1:n}$.

The bound gives a resolution criterion for the dictionary: if $W_1(P,Q) > 2\varepsilon_m$, at least one retained critic has different expectations under P and Q. This is the separation needed for finite-dictionary consistency in Corollary C.13. The bound approximates W_1 through the coordinatewise supremum in (16); trace and Mahalanobis aggregation use quadratic norms of the same expectation contrasts. The anchored Lipschitz ball is uniformly bounded and equicontinuous, so the Arzelà–Ascoli theorem guarantees a finite uniform net on a compact metric space. Its size can grow rapidly with intrinsic dimension, motivating structured dictionaries such as the distance cones used in the experiments.

Proposition C.4 (Transport domination of diffusion MMD)**.** *Let $(\mathcal{M}, g)$ be closed and connected, equip it with geodesic distance, and fix $t > 0$. Define*

$$L_t = \sup_{x \in \mathcal{M}} \|D\Phi_t(x)\|_{\mathrm{op}}, \tag{26}$$

where $D\Phi_t(x) : T_x\mathcal{M} \to L^2(\mathcal{M}, \mu)$ is the differential of the diffusion feature map. Then $L_t < \infty$ and, for all Borel probability measures P, Q on $\mathcal{M}$,

$$\mathrm{MMD}_t(P,Q) \le L_t \, W_1(P,Q). \tag{27}$$

The inequality relates the channels on a common geometric scale: moving probability mass by a small average geodesic distance also changes the diffusion embedding by a small amount. The spectral expression in (21) explains the dependence on diffusion time. Increasing t suppresses high-frequency differences more strongly, while the Lipschitz constraint defining the transport channel is set directly by geodesic distance. These two weightings motivate their joint use in a multichannel scan.

Proposition C.5 (Special-case reductions)**.** *The following reductions hold.*

(i) On $\mathbb{R}^d$ with its Euclidean metric and Lebesgue measure, the heat kernel channel is a Gaussian-RBF MMD CUSUM up to a deterministic multiplicative constant.

(ii) With one diffusion scale and no critic channel, Algorithm 1 is the ordinary kernel mean-embedding CUSUM for the kernel p_{2t}.

(iii) With one scalar critic, the critic channel is the classical scalar CUSUM applied to $f(Z_i)$.

(iv) With the full anchored Lipschitz class, the unstandardized critic scan is exactly the empirical W_1 scan.

(v) If a predictable-score process has linear variance clock in calendar time, the variance-clock theorem in Theorem E.6 reduces to the standard self-normalized martingale CUSUM bridge.

The finite-sample permutation statement appears as Theorem 3.1 in the main text; its proof is in Section G.

C.3 Joint null limits and multiplier calibration

For a fixed diffusion grid $\mathcal{T} = \{t_1, \dots, t_L\}$ and a fixed dictionary F_m, define the direct-sum Hilbert space

$$\mathcal{V} = \bigoplus_{\ell=1}^{L} L^2(\mathcal{M}, \mu) \oplus \mathbb{R}^m$$

and the score

$$\Psi(z) = \big(\Phi_{t_1}(z), \dots, \Phi_{t_L}(z), F_m(z)\big).$$

Let Γ be the covariance operator of $\Psi(Z)$ under P.

Assumption C.6 (Fixed-channel null conditions)**.** *Under H_0:*

(N1) Ψ is Borel measurable, $\mathcal{V}$ is separable, and $\mathbb{E}_P \|\Psi(Z)\|_{\mathcal{V}}^2 < \infty$.

(N2) For every retained diffusion channel, $v_{t_\ell} = \mathbb{E}\,\|\Phi_{t_\ell}(Z) - m_{P,t_\ell}\|_{L^2}^2 > 0$.

(N3) If the critic scan uses matrices $A_{n,F}$, then $A_{n,F} \xrightarrow{\mathbb{P}} A_F$ in operator norm for a deterministic nonnegative-definite matrix A_F. For $A_{n,F} = \widehat{\Sigma}_F^{\dagger}$, this condition is understood as $\widehat{\Sigma}_F^{\dagger} \xrightarrow{\mathbb{P}} \Sigma_F^{\dagger}$.

On a closed manifold with a finite dictionary, (N1) holds automatically because every channel is bounded. Condition (N2) only removes channels that are constant under P. Condition (N3) is automatic for trace standardization when $\mathrm{tr}(\Sigma_F) > 0$; for a generalized inverse it is a rank-stability condition rather than full positive definiteness.

Theorem C.7 (Joint Hilbert-space CUSUM limit)**.** *Under Assumption C.6, in $D([0,1]; \mathcal{V})$,*

$$\mathbb{B}_{n,\Psi}(\cdot) \rightsquigarrow \mathbb{B}_{\Gamma}(\cdot) := \mathbb{W}_{\Gamma}(\cdot) - \{\cdot\}\mathbb{W}_{\Gamma}(1),$$

where $\mathbb{W}_\Gamma$ is a $\mathcal{V}$-valued Brownian motion with covariance operator Γ. Equivalently, for components a, b and $r, s \in [0,1]$,

$$\mathrm{Cov}\{\mathbb{B}_{\Gamma,a}(r), \mathbb{B}_{\Gamma,b}(s)\} = (r \wedge s - rs)\,\Gamma_{ab}. \tag{28}$$

Consequently,

$$T_n \rightsquigarrow T_\infty, \tag{29}$$

where T_∞ is obtained from (4) *by replacing empirical bridges and standardizers by their Brownian-bridge limits and population values.*

The direct-sum construction retains covariance between channels that share the same observations. Applying the Hilbert-space invariance principle to this joint score (Araujo and Giné, 1980; Horváth and Kokoszka, 2012) therefore gives the null law of their maximum, including correlations between diffusion scales and critics. The unknown operator Γ makes this law nonpivotal. The multiplier method estimates it by applying a common random weight to all features of each observation.

For Mahalanobis aggregation, the critic bridge becomes a standard Brownian bridge on the covariance range. Redundant critics therefore contribute no additional dimensions when that range is estimated consistently. Corollary C.8 records the individual channel limits, proved in Section H.

Corollary C.8 (Single-channel limits)**.** *Under the conditions of Theorem C.7, let $\mathbb{B}_t$ and $\mathbb{B}_F$ denote the diffusion-scale-t and critic components of $\mathbb{B}_\Gamma$. Then*

$$S_{n,t}^H \rightsquigarrow \sup_{r\in[\eta,1-\eta]} \frac{\|\mathbb{B}_t(r)\|_{L^2}^2}{v_t r(1-r)}, \tag{30}$$

$$S_n^F(A_{n,F}) \rightsquigarrow \sup_{r\in[\eta,1-\eta]} \frac{\mathbb{B}_F(r)^\top A_F \mathbb{B}_F(r)}{r(1-r)}, \tag{31}$$

$$S_n^{F,\max} \rightsquigarrow \sup_{r\in[\eta,1-\eta]} \frac{\max_{1\le j\le m}|\mathbb{B}_{F,j}(r)|}{\sqrt{r(1-r)}}. \tag{32}$$

If $A_F = \Sigma_F^\dagger$ and $q = \operatorname{rank}(\Sigma_F)$, the critic limit is the squared norm of a q-dimensional standard Brownian bridge after restriction to the covariance range.

Theorem C.9 (Conditional multiplier consistency)**.** *Suppose Assumption C.6 holds and, for some $\delta > 0$, $\mathbb{E}_P \|\Psi(Z)\|_{\mathcal{V}}^{2+\delta} < \infty$. Let $\xi_1, \dots, \xi_n$ be iid, independent of the data, with $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = 1$, and $\mathbb{E}|\xi_i|^{2+\delta} < \infty$. Construct the joint multiplier bridge by* (20)*, using the same multipliers for every component. Then, conditionally on the observations,*

$$\mathbb{B}_{n,\Psi}^{\#} \rightsquigarrow \mathbb{B}_\Gamma \quad \textit{in probability in } D([0,1];\mathcal{V}).$$

If the distribution function of T_∞ is continuous at its $(1-\alpha)$ quantile, the conditional multiplier critical value gives an asymptotic level-α test.

Common multipliers reproduce the cross-channel covariance in (28); independently resampling each channel would remove it. The theorem justifies simulation of the joint critical value using the original standardizers and fixed features. Bounded diffusion features and critics satisfy the required moment condition. The proof in Section H combines conditional finite-dimensional limits with a uniform bound on the remaining Hilbert coordinates.

C.4 Fixed and local alternatives

The next result is stated for a generic Hilbert score because it applies simultaneously to diffusion channels, finite critic vectors, and their direct sum. Let

$$m_{P_n} = \mathbb{E}_{P_n}\Psi(Z), \qquad m_{Q_n} = \mathbb{E}_{Q_n}\Psi(Z), \qquad \delta_n = m_{P_n} - m_{Q_n}.$$

The magnitude of the expected CUSUM grows as a candidate split approaches the change and declines after it passes the change. Its shape is described by

$$b_\theta(r) = r(1-\theta)\mathbf{1}\{r \le \theta\} + \theta(1-r)\mathbf{1}\{r > \theta\}.$$

Assumption C.10 (Centered triangular-array bridge limit)**.** *Under* (2)*, let $U_{n,i} = \Psi(Z_i) - \mathbb{E}\Psi(Z_i)$. Assume*

$$\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{\lfloor n\cdot\rfloor} U_{n,i} - \frac{\lfloor n\cdot\rfloor}{n}\sum_{i=1}^{n} U_{n,i}\right\} \rightsquigarrow \mathbb{Z} \quad \textit{in } D([0,1];\mathcal{V}) \tag{33}$$

for a centered Gaussian process $\mathbb{Z}$ with continuous paths and $\mathbb{Z}(0) = \mathbb{Z}(1) = 0$ almost surely. For each channel standardizer, assume convergence to a finite positive limit under the pooled mixture $\theta P_n + (1-\theta)Q_n$. For a generic quadratic channel, let $A_n \xrightarrow{\mathbb{P}} A$, where A is bounded, self-adjoint, and nonnegative definite.

The following sufficient condition verifies Assumption C.10 in the compact geometric setting used here. Continuity of the fixed feature map supplies uniform finite-dimensional approximation; weak convergence of the segment laws supplies covariance convergence. Both hold automatically for fixed

laws and the exact diffusion and landmark-distance features. Let Γ_P, Γ_Q be their segment covariance operators and $\mathbb{W}_P, \mathbb{W}_Q$ independent Brownian motions with these covariances. Define

$$\mathbb{W}_{P,Q}(r) = \begin{cases} \mathbb{W}_P(r), & 0 \le r \le \theta, \\ \mathbb{W}_P(\theta) + \mathbb{W}_Q(r-\theta), & \theta < r \le 1, \end{cases}$$

and $\mathbb{Z}(r) = \mathbb{W}_{P,Q}(r) - r\mathbb{W}_{P,Q}(1)$.

Proposition C.11 (Compact-space verification of the alternative limit)**.** *Let $\mathcal{X}$ be compact metric and let $\Psi : \mathcal{X} \to \mathcal{V}$ be fixed and continuous, with $\mathcal{V}$ separable Hilbert. In the independent single-change model, suppose $P_n \Rightarrow P$, $Q_n \Rightarrow Q$ and $\theta_n \to \theta \in (0,1)$. Then* (33) *holds with the process $\mathbb{Z}$ just defined. Each pooled feature variance converges to its variance under $\nu = \theta P + (1-\theta)Q$. Thus Assumption C.10 holds for trace-standardized channels with positive limiting variances. Fixed finite covariance matrices also converge, so covariance whitening is covered when its inverse or generalized inverse is continuous along the estimated matrices. If $P = Q$, $\mathbb{Z}$ is the ordinary bridge $\mathbb{B}_{\Gamma_P}$.*

The proof in Section I.1 uses finite-dimensional invariance principles and a uniform Hilbert-tail bound. It derives the bridge condition rather than imposing a functional limit separately. The high-level assumption remains useful for other scores; convergence of their means alone does not determine covariance or tightness.

For local alternatives with $\sqrt{n}\delta_n \to h$, write $\mathcal{T}_A$ for the limiting quadratic scan functional, or the maximum of the finitely many such functionals with their limiting weights. If a critical value $c_{n,\alpha} \xrightarrow{\mathbb{P}} c_\alpha$ along these alternatives and the shifted scan has no atom at c_α, Theorem 3.2(i) and Slutsky's theorem give

$$\mathbb{P}(T_n > c_{n,\alpha}) \longrightarrow \mathbb{P}\{\mathcal{T}_A(\mathbb{Z} + b_\theta h) > c_\alpha\}.$$

For laws tending to a common P, the noise is the same bridge as under P, and the drift determines this local-power change. This critical-value formulation describes asymptotic calibration; the finite-B permutation guarantee is separately given by Theorem 3.1.

Under these conditions, Theorem 3.2 gives the local drift and fixed-change localization result. The pooled-variance interpretation is (9). The following consequences identify when geometric channels retain a signal.

Corollary C.12 (Fixed diffusion-channel consistency)**.** *Let $(\mathcal{M}, g)$ be closed, fix $t > 0$, and suppose the single-change model has fixed laws $P \neq Q$ and $\theta \in (\eta, 1-\eta)$. Let $\widehat{\theta}_{n,t}$ be the smallest split maximizing the diffusion-channel criterion, divided by n. Then*

$$S^H_{n,t} \to \infty \quad \textit{in probability}, \qquad \widehat{\theta}_{n,t} \xrightarrow{\mathbb{P}} \theta.$$

Under the joint-channel conditions of Theorem 3.2, including the required control of every retained channel and its standardizer, the same conclusion holds for the multichannel estimator whenever this diffusion channel is retained.

Every fixed distributional change on a closed manifold induces a nonzero diffusion-feature mean shift, including changes that preserve selected moments. Diffusion time still controls the magnitude of that shift through the weights in (21). Multiscale aggregation addresses this finite-sample sensitivity to the spectral content of the change. Section I gives the drift argument and its diffusion and critic specializations.

Corollary C.13 (Finite-dictionary consistency)**.** *Under a fixed change from P to Q at fraction $\theta \in (\eta, 1-\eta)$, let $\delta_F = PF_m - QF_m$ and define the pooled covariance*

$$\Sigma_{\theta,F} = \theta \operatorname{Cov}_P\{F_m(Z)\} + (1-\theta)\operatorname{Cov}_Q\{F_m(Z)\} + \theta(1-\theta)\delta_F\delta_F^\top. \tag{34}$$

The trace-standardized critic scan diverges and its location estimator is consistent whenever $\delta_F \neq 0$. The Mahalanobis scan has the same conclusion if $\widehat{\Sigma}_F^\dagger \xrightarrow{\mathbb{P}} \Sigma_{\theta,F}^\dagger$. Under this convergence, $\delta_F \neq 0$ already implies $\delta_F^\top \Sigma_{\theta,F}^\dagger \delta_F > 0$.

A nonzero critic mean contrast supplies a consistent channel under trace standardization. It also lies in the range of the pooled covariance because that covariance includes the between-segment term $\theta(1-\theta)\delta_F\delta_F^\top$. Thus Mahalanobis aggregation retains the signal when the empirical covariance range is consistently estimated. Proposition C.3 gives a sufficient dictionary resolution for detecting a specified transport separation.

Theorem C.14 (Consistency of the permutation-calibrated test)**.** *Suppose a fixed change from P to Q occurs at a fraction $\theta \in (\eta, 1-\eta)$. Let the diffusion grid and critic dictionary be fixed and finite, let every channel construction and standardizer be permutation invariant, and suppose the conditions of Theorem 3.2(ii)–(iii) hold for the joint direct-sum score. Write $\overline{\Psi}_n = n^{-1}\sum_i \Psi(Z_i)$. Assume at least one retained channel has a nonzero limiting signal in its declared quadratic metric and*

$$\frac{1}{n}\sum_{i=1}^{n} \left\|\Psi(Z_i) - \overline{\Psi}_n\right\|_{\mathcal{V}}^2 = O_{\mathbb{P}}(1). \tag{35}$$

For the bounded diffusion and finite-critic channels used by Algorithm 1, (35) *is automatic. For a fixed number B of independent uniform random permutations, let $T_n^{(1)}, \dots, T_n^{(B)}$ and $\widehat{p}_B$ be defined as in Algorithm 1. Then there is a constant $c > 0$ such that*

$$\frac{T_n}{n} \xrightarrow{\mathbb{P}} c, \qquad \max_{1\le b\le B} T_n^{(b)} = O_{\mathbb{P}}(1), \qquad \widehat{p}_B \xrightarrow{\mathbb{P}} \frac{1}{B+1}.$$

Consequently, if $(B+1)^{-1} \le \alpha$, the permutation-calibrated test has power converging to one.

The observed ordering preserves the change and produces a scan of order n. A random permutation disperses observations from the two segments throughout the sequence, leaving a scan of order one under the pooled standardization. This contrast connects the drift result to power of the calibrated test. The number of permutations also sets the resolution of the reported p-value: its smallest value is $(B+1)^{-1}$. Thus exact null validity holds for every B, while consistency at level α requires that this minimum be no larger than α; at level 0.05, at least 19 permutations are needed. The theorem treats fixed B; a growing number requires uniform control of the permutation scans.

C.5 Pooled circle representations under fixed changes

The circle construction in Section B.1 estimates both its scale and its feature space from the observations being scanned. Its fixed-change behavior can be established by convergence of the fitted functions. The geometric step is that farthest-point selection bisects the longest circular gap: after a power-of-two number of selections, its limiting landmark set is equally spaced, even though intermediate selections can tie.

Let $\nu = \theta P + (1-\theta)Q$ be the pooled law on $\mathbb{S}^1$, let $a_\star$ minimize $\int d(x,a)\ \mathrm{d}\nu(x)$, and let d_ν^2 be the median of $d(X,Y)^2$ for independent $X, Y \sim \nu$. Take fixed powers of two $m \le M$ for the critic and diffusion landmark counts, respectively. At scale t, write the evaluated circle kernel as

$$\widetilde{k}_t(x,y) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{h=1}^{J(t)} e^{-2th^2}\cos\{h(x-y)\},$$

where $J(t)$ is the smallest integer $J \ge 0$ satisfying the implemented stopping rule

$$\frac{1}{\pi}\left\{e^{-2t(J+1)^2} + \int_{J+1}^{\infty} e^{-2tx^2}\ \mathrm{d}x\right\} < 10^{-10}. \tag{36}$$

For fixed positive scale multipliers $\gamma_1, \dots, \gamma_L$, put $t_\ell^* = \gamma_\ell d_\nu^2$. Form G_ℓ by evaluating $\widetilde{k}_{t_\ell^*}$ on the M points $a_\star + 2\pi j/M$, $0 \le j < M$. The Nyström cutoff is $c(G) = \max\{\varepsilon_{\mathrm{abs}}, \varepsilon_{\mathrm{rel}}\lambda_{\max}(G)\}$, with the fixed tolerances used in the computation.

Theorem C.15 (Pooled circle representation and localization)**.** *Suppose the independent observations have fixed segment laws P, Q with positive continuous densities on $\mathbb{S}^1$, and $\tau_n/n \to \theta \in (\eta, 1-\eta)$. Suppose the population medoid $a_\star$ is unique. At each t_ℓ^*, assume the implemented Fourier evaluation terminates with a locally constant count $J(t)$ and no eigenvalue of G_ℓ equals $c(G_\ell)$.*

Use the pooled squared-distance median, medoid–farthest landmarks, all Nyström eigenvalues above the cutoff, and the first m distance critics, as in Section B.1. After coordinate permutations and isometric embeddings into fixed Euclidean spaces, the fitted channel maps $\psi_{n,j}$ satisfy

$$\max_j \sup_{x\in\mathbb{S}^1} \|\psi_{n,j}(x) - \psi_j(x)\| \xrightarrow{\mathbb{P}} 0$$

for deterministic continuous limits ψ_j, where j indexes the L diffusion channels and the critic channel. Write $\delta_j = P\psi_j - Q\psi_j$ and $v_j = \int \|\psi_j - \nu\psi_j\|^2 \, \mathrm{d}\nu$. If every $v_j > 0$ and at least one $\delta_j \neq 0$, the trace-standardized joint scan with fixed positive weights c_j obeys

$$\frac{T_n}{n} \xrightarrow{\mathbb{P}} \theta(1-\theta) \max_j \frac{c_j \|\delta_j\|^2}{v_j} > 0, \qquad \frac{\widehat{\tau}_n}{n} \xrightarrow{\mathbb{P}} \theta.$$

For fixed B with $(B+1)^{-1} \leq \alpha$, its permutation rejection probability tends to one.

The medoid condition fixes the limiting grid's rotation. The two spectral conditions keep the evaluated Fourier series and the retained Nyström subspace stable; repeated eigenvalues within that subspace are allowed. Positive densities make the sample dense on the circle and give a unique positive distance median. The proof in Section I.5 controls the complete fitted map on the same sample, so no sample splitting is needed.

Corollary C.16 (The middle circle shape design)**.** *For $P = \mathrm{vM}(0, 4)$, $Q = \frac{1}{2}\{\mathrm{vM}(-a, 8) + \mathrm{vM}(a, 8)\}$ with $\cos a = A(4)/A(8)$, and $\theta = 1/2$, the original construction with $M = 128$, $m = 64$, the five scale multipliers and numerical cutoffs in Section B.1 satisfies Theorem C.15. Thus Combined has consistent location and permutation power tending to one under this fixed alternative.*

Here the limiting grid is centered at zero and the second cosine contrast is between 0.0280 and 0.0281. The proof verifies its retention at all five scales using interval bounds for the population median and spectrum. This asymptotic guarantee complements the small signal fractions and low power at the sample sizes in Section B.6; the theorem supplies no sample-size rate.

D Residual geometry on the circle

Proposition 3.3 states the general pooled-projection result, proved in Section I.3. The following specialization explains its signal gain and its frequency-dependent limitations.

Identify the unit circle with $[-\pi, \pi)$ and use geodesic distance d. For every landmark a, its centered distance has the absolutely convergent Fourier expansion

$$d(x, a) - \frac{\pi}{2} = -\frac{4}{\pi} \sum_{\substack{k \geq 1 \\ k \text{ odd}}} \frac{\cos\{k(x-a)\}}{k^2}. \tag{37}$$

Thus a low-frequency diffusion span can account for most distance variance while omitting a higher-frequency contrast. Let H span the sine and cosine modes $1, \ldots, J$, and use m equally spaced landmarks $a_\ell = 2\pi\ell/m$, $\ell = 0, \ldots, m-1$.

Proposition D.1 (Odd-frequency residual gain)**.** *Let P, Q have circle densities $(1 + \varepsilon\cos(jx))/(2\pi)$ and $(1 - \varepsilon\cos(jx))/(2\pi)$, with $0 < \varepsilon < 1$, and let $\theta = 1/2$. For an odd integer $j > J$ with $m \nmid 2j$,*

$$\frac{\theta(1-\theta)\|\delta_\perp\|^2}{v_\perp} = \frac{\varepsilon^2}{4j^4 \sum_{k>J,\ k \text{ odd}} k^{-4}}. \tag{38}$$

The raw distance slope replaces the tail sum by $\sum_{k \geq 1,\ k \text{ odd}} k^{-4}$. For $J = 2$, the gain factor is $(\pi^4/96)/(\pi^4/96 - 1) \approx 69.1$. For even $j > J$, both diffusion and distance mean contrasts vanish.

The proof is in Section I.3.

For the experiment's $\varepsilon = 0.35$, $J = 2$ and $j = 3$, the raw and residual signal slopes are approximately 0.00037 and 0.0258. Projection preserves the third-harmonic contrast while removing the dominant first-harmonic variance. At $j = 5$ the residual slope is 0.00334, reflecting the j^{-4} decay that remains after projection. These are population slopes; permutation critical values determine finite-sample rejection. Figure 1 shows why the gain is large: the first harmonic accounts for about 98.6% of the pooled distance variance, while the change lies in the third harmonic.

E ADDITIONAL PROCESS RESULTS

E.1 THE FULL LIPSCHITZ PROCESS

For a finite dictionary, joint convergence only has to control finitely many critic coordinates. Passing to the full transport supremum requires uniform control over a function class. Bracketing measures its complexity by the number of upper and lower function pairs needed to enclose every critic to a given $L^2(P)$ accuracy. The condition below makes this uniform stochastic approximation available; the subsequent consistency result uses only the existence of a separating critic.

Let $F(x) = \operatorname{diam}(\mathcal{X})$ be the constant envelope of $\mathcal{F}_{\mathrm{Lip}}$. For a probability law P, write $N_{[]}(\varepsilon, \mathcal{F}, L^2(P))$ for the bracketing number.

Assumption E.1 (Critic entropy)**.** *The class $\mathcal{F}_0 \subseteq \mathcal{F}_{\mathrm{Lip}}$ is pointwise measurable and*

$$J_{[]}(1, \mathcal{F}_0, L^2(P)) := \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon \|F\|_{P,2}, \mathcal{F}_0, L^2(P))}\, \mathrm{d}\varepsilon < \infty. \tag{39}$$

Theorem E.2 (Sequential Lipschitz-critic bridge)**.** *Under H_0 and Assumption E.1,*

$$\mathbb{G}_n \rightsquigarrow \mathbb{G}_P \quad \text{in } \ell^\infty([0,1] \times \mathcal{F}_0), \tag{40}$$

where $\mathbb{G}_P$ is a centered tight Gaussian process, a Brownian bridge in r for each fixed critic, with covariance

$$\operatorname{Cov}\{\mathbb{G}_P(r,f), \mathbb{G}_P(s,g)\} = (r \wedge s - rs) \operatorname{Cov}_P\{f(Z), g(Z)\}. \tag{41}$$

Consequently,

$$\sup_{r\in[\eta,1-\eta]} \sup_{f\in\mathcal{F}_0} \frac{|\mathbb{G}_n(r,f)|}{\sqrt{r(1-r)}} \rightsquigarrow \sup_{r\in[\eta,1-\eta]} \sup_{f\in\mathcal{F}_0} \frac{|\mathbb{G}_P(r,f)|}{\sqrt{r(1-r)}}. \tag{42}$$

The bracketing central limit theorem controls the critic index, while the sequential extension supplies the time index (van der Vaart and Wellner, 1996). Together they yield a null law for optimization over splits and a whole function class. The following examples make the entropy condition explicit for common critic choices.

Proposition E.3 (Critic classes with finite entropy integral)**.** *Each of the following satisfies Assumption E.1.*

(i) Any finite critic dictionary.

(ii) A compact q-dimensional parametric class $\{f_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^q\} \subseteq \mathcal{F}_{\mathrm{Lip}}$ for which $|f_\vartheta(x) - f_{\vartheta'}(x)| \le L(x)\|\vartheta - \vartheta'\|$ and $PL^2 < \infty$.

(iii) An anchored, uniformly bounded C^s or Hölder-s ball contained in $\mathcal{F}_{\mathrm{Lip}}$ on a compact connected d-dimensional smooth Riemannian manifold with its geodesic metric when $s > d/2$.

In particular, the standard Hölder-entropy route verifies the full Lipschitz ball when $d = 1$, but does not by itself verify it when $d \ge 2$.

The smoothness threshold follows from integrating $\varepsilon^{-d/(2s)}$ near zero. At Lipschitz smoothness $s = 1$, this argument covers dimension one; a smoother or finite critic class keeps the same route available in higher dimensions.

Theorem E.4 (Full-critic fixed-change consistency)**.** *Let $(\mathcal{X}, d)$ be compact and suppose a fixed change occurs from P to Q at $\theta \in (\eta, 1-\eta)$. Let $\mathcal{F}_0 \subseteq \mathcal{F}_{\mathrm{Lip}}$ be pointwise measurable, or admit a countable pointwise-dense subclass, and satisfy*

$$\Delta_{\mathcal{F}_0}(P,Q) = \sup_{f\in\mathcal{F}_0} |Pf - Qf| > 0.$$

Then the statistic in (42) *diverges in probability. If $\mathcal{F}_0 = \mathcal{F}_{\mathrm{Lip}}$, then $\Delta_{\mathcal{F}_0}(P,Q) = W_1(P,Q) > 0$ for every $P \ne Q$.*

Consistency only needs one separating bounded critic and does not require the full Donsker condition. The entropy condition is needed for a functional null limit and multiplier calibration over the entire class. Permutation calibration remains finite-sample valid without it, provided the full optimization is performed identically under every permutation.

E.2 PREDICTABLE-SCORE BRIDGES AND THE LOGARITHMIC MARTINGALE-MMD CLOCK

Predictable scores choose the direction of each new observation using past features. Their variance can change with observation index. A bridge must then center by the fraction of accumulated variance. The general result below states this construction; the averaged-score example identifies a logarithmic clock.

For each n, let $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1} \subseteq \cdots \subseteq \mathcal{F}_{n,n}$ be a filtration and let $D_{n,i}$ be $\mathcal{F}_{n,i}$-measurable with $\mathbb{E}(D_{n,i} \mid \mathcal{F}_{n,i-1}) = 0$. Define

$$S_n(k) = \sum_{i=1}^{k} D_{n,i}, \qquad Q_n(k) = \sum_{i=1}^{k} \mathbb{E}(D_{n,i}^2 \mid \mathcal{F}_{n,i-1}), \qquad V_n(k) = \sum_{i=1}^{k} D_{n,i}^2.$$

Let $a_n \to \infty$, let $q : [0,1] \to [0,1]$ be continuous and nondecreasing with $q(0) = 0$, $q(1) = 1$, and let $\kappa_n(s)$ be a deterministic, right-continuous, nondecreasing integer-valued time map with $\kappa_n(0) = 0$ and $\kappa_n(1) = n$.

Assumption E.5 (Variance-clock martingale conditions)**.** *For every $\varepsilon > 0$:*

(M1) $\sup_{s\in[0,1]} |Q_n\{\kappa_n(s)\}/a_n^2 - q(s)| \xrightarrow{\mathbb{P}} 0.$

(M2)

$$a_n^{-2} \sum_{i=1}^{n} \mathbb{E}\{D_{n,i}^2 \mathbf{1}(|D_{n,i}| > \varepsilon a_n) \mid \mathcal{F}_{n,i-1}\} \xrightarrow{\mathbb{P}} 0.$$

(M3) $\sup_{s\in[0,1]} |V_n\{\kappa_n(s)\} - Q_n\{\kappa_n(s)\}|/a_n^2 \xrightarrow{\mathbb{P}} 0.$

Define the observed variance fraction

$$\widehat{q}_n(s) = \frac{V_n\{\kappa_n(s)\}}{V_n(n)}$$

when $V_n(n) > 0$, and set it equal to $q(s)$ otherwise. In the bridge statistic below, set $\mathbb{C}_n \equiv 0$ on the event $V_n(n) = 0$.

Theorem E.6 (Martingale bridge in variance time)**.** *Under Assumption E.5,*

$$\mathbb{C}_n(s) = V_n(n)^{-1/2} \left[S_n\{\kappa_n(s)\} - \widehat{q}_n(s) S_n(n)\right] \rightsquigarrow W\{q(s)\} - q(s)W(1) \tag{43}$$

in $D([0,1])$, where W is standard Brownian motion. If $q(s) = s$, the limit is a standard Brownian bridge.

The martingale functional CLT (Brown, 1971; Hall and Heyde, 1980) yields a bridge after centering by the observed fraction of total variance. When the variance fraction converges to calendar time, $Q_n(\lfloor ns \rfloor)/Q_n(n) \to s$, this is the usual calendar-time bridge. The averaged scores below accumulate variance on a logarithmic scale.

For fixed features, observation fractions index the bridge. Averaging past features to choose each new score changes its variance clock. For iid centered bounded Hilbert elements ζ_i with covariance C, define $D_i = \langle \overline{\zeta}_{i-1}, \zeta_i \rangle$, where $\overline{\zeta}_{i-1} = (i-1)^{-1} \sum_{j<i} \zeta_j$. Independent paired kernel-feature differences under a two-sample equality null give one example (Chatterjee and Ramdas, 2026). Write $S_k = \sum_{i=2}^{k} D_i$ and $V_k = \sum_{i=2}^{k} D_i^2$, both zero at $k = 1$. If $\sigma_D^2 = \mathrm{tr}(C^2) > 0$, each score contributes variance $\sigma_D^2/(i-1)$, and the self-normalized bridge

$$\mathbb{C}_n^{\log}(s) = V_n^{-1/2} \left\{ S_{k_n(s)} - \frac{V_{k_n(s)}}{V_n} S_n \right\}, \qquad k_n(s) = \max\{1, \lfloor n^s \rfloor\}, \tag{44}$$

converges to a standard Brownian bridge, with $\mathbb{C}_n^{\log} \equiv 0$ when $V_n = 0$. The first $\lfloor \sqrt{n} \rfloor$ observations account asymptotically for half the total variance; the first $\lfloor rn \rfloor$ account for almost all of it for every fixed $r > 0$. Centering by the observed variance fraction therefore extends bridge construction to averaged predictable scores. Theorem E.7 states the uniform clock limits, proved in Section J.

For the averaged score above, let $\mathcal{F}_i = \sigma(\zeta_1, \ldots, \zeta_i)$ and $Q_k = \sum_{i=2}^{k} \mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1})$, with $Q_1 = 0$. The following result identifies its logarithmic variance clock.

Theorem E.7 (Logarithmic clock for averaged scores). *Assume $\mathbb{E}\zeta_1 = 0$, $\|\zeta_1\| \le K$ almost surely, and*

$$\sigma_D^2 := \mathrm{tr}(C^2) = \mathbb{E}\,\langle \zeta_1, \zeta_2 \rangle^2 > 0. \tag{45}$$

Then:

(i) $\{D_i, \mathcal{F}_i\}_{i\ge 2}$ is a martingale-difference sequence and

$$\mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1}) = \left\langle C\overline{\zeta}_{i-1}, \overline{\zeta}_{i-1} \right\rangle, \qquad \mathbb{E}D_i^2 = \frac{\sigma_D^2}{i-1}. \tag{46}$$

(ii) Uniformly in $s \in [0,1]$, with $k_n(s) = \max\{1, \lfloor n^s \rfloor\}$,

$$\sup_{s\in[0,1]} \left| \frac{Q_{k_n(s)}}{\sigma_D^2 \log n} - s \right| \xrightarrow{\mathbb{P}} 0, \qquad \sup_{s\in[0,1]} \left| \frac{V_{k_n(s)}}{\sigma_D^2 \log n} - s \right| \xrightarrow{\mathbb{P}} 0. \tag{47}$$

(iii) The self-normalized log-time bridge (44)*, with $\mathbb{C}_n^{\log} \equiv 0$ when $V_n = 0$, converges in $D([0,1])$ to a standard Brownian bridge. Here $D([0,1])$ is the space of right-continuous paths with left limits, equipped with the Skorokhod J_1 topology.*

(iv) For every fixed $r \in (0,1]$,

$$\frac{Q_{\lfloor nr \rfloor}}{Q_n} \xrightarrow{\mathbb{P}} 1. \tag{48}$$

Here the ratio is defined as zero on the event $Q_n = 0$, whose probability tends to zero under (45)*. Thus the calendar-time clock collapses on $(0,1]$ and the standard bridge indexed by $k = \lfloor nr \rfloor$ is not valid for these averaged increments.*

Its proof in Section J applies Theorem E.6 after reindexing the increments.

F Auxiliary Lemmas and Notation

This appendix records the auxiliary facts used in the proofs. The notation follows the main text. All stochastic-process spaces use their Borel sigma-fields; $D([0,1];\mathcal{H})$ carries the Skorokhod J_1 topology, and convergence to a continuous limit is equivalent to uniform convergence after an asymptotically negligible time change.

Lemma F.1 (Bridge transformation). *Let $\mathcal{H}$ be a separable Hilbert space. If $X_n \rightsquigarrow X$ in $D([0,1];\mathcal{H})$ and X has continuous paths almost surely, then*

$$X_n(r) - \frac{\lfloor nr \rfloor}{n} X_n(1) \rightsquigarrow X(r) - rX(1)$$

in $D([0,1];\mathcal{H})$.

Lemma F.2 (Deterministic mean of a one-change bridge). *Under* (2)*, let ψ be Bochner integrable under P_n and Q_n, and put $\delta_n = P_n\psi - Q_n\psi$ and $\theta_n = \tau_n/n$. At each grid point $r_k = k/n$,*

$$\mathbb{E}\mathbb{B}_{n,\psi}(r_k) = \sqrt{n}\, b_{\theta_n}(r_k)\delta_n.$$

Lemma F.3 (Unique maximizer of the normalized drift). *For $\theta \in (0,1)$, define*

$$h_\theta(r) = \frac{b_\theta(r)^2}{r(1-r)}, \qquad r \in (0,1).$$

Then

$$h_\theta(r) = \begin{cases} r(1-\theta)^2/(1-r), & r \le \theta, \\ \theta^2(1-r)/r, & r > \theta, \end{cases}$$

and h_θ has the unique maximizer $r = \theta$, with $h_\theta(\theta) = \theta(1-\theta)$.

Lemma F.4 (Finite-population Hilbert bridge). *Let $y_1, \dots, y_n$ be fixed elements of a Hilbert space with $\sum_{i=1}^n y_i = 0$, let π be a uniform random permutation, and define $S_k^\pi = \sum_{i=1}^k y_{\pi(i)}$. For every integer $K < n$,*

$$\mathbb{E}_\pi \max_{1\le k\le K} \left\| n^{-1/2} S_k^\pi \right\|^2 \le \frac{4K}{n-K} \frac{1}{n-1} \sum_{i=1}^n \|y_i\|^2. \tag{49}$$

Lemma F.5 (Hilbert fourth-moment bound)**.** *Let $X_1, \dots, X_m$ be independent mean-zero random elements in a separable Hilbert space with $\|X_i\| \le K$ almost surely. Then*

$$\mathbb{E}\left\|\sum_{i=1}^{m} X_i\right\|^4 \le 48m^2K^4. \tag{50}$$

Lemma F.6 (Quadratic-form clock fluctuation)**.** *Under the assumptions of Theorem E.7, let $Q_k = \sum_{i=2}^{k} \langle C\overline{\zeta}_{i-1}, \overline{\zeta}_{i-1}\rangle$. There is a finite constant C_0, depending only on the law of ζ_1, such that for every $k \ge 3$,*

$$\mathbb{E}\{Q_k - \sigma_D^2 H_{k-1}\}^2 \le C_0 \log k, \qquad H_m = \sum_{j=1}^{m} j^{-1}. \tag{51}$$

G Proofs of the Exact Results

G.1 Proofs of the auxiliary bridge and drift lemmas

Proof of Lemma F.1. Define $a_n(r) = \lfloor nr \rfloor / n$. Since $\sup_{r\in[0,1]} |a_n(r) - r| \le n^{-1}$, it suffices to study the map

$$x \mapsto \{r \mapsto x(r) - rx(1)\}.$$

The evaluation map $x \mapsto x(1)$ is continuous at every path that is continuous at 1 under the J_1 topology. Multiplication by the continuous scalar function r and subtraction are continuous operations in $D([0,1];\mathcal{H})$ at continuous paths. Because X is continuous almost surely, the continuous mapping theorem gives

$$X_n(\cdot) - \{\cdot\}X_n(1) \rightsquigarrow X(\cdot) - \{\cdot\}X(1).$$

It remains to replace r by $a_n(r)$. The difference has uniform norm

$$\sup_{r\in[0,1]} \|\{a_n(r) - r\}X_n(1)\| \le n^{-1} \|X_n(1)\|.$$

Tightness of $X_n(1)$ follows from $X_n \rightsquigarrow X$, so the right-hand side is $o_{\mathbb{P}}(1)$. Slutsky's theorem in $D([0,1];\mathcal{H})$ proves the claim. □

Proof of Lemma F.2. Write $k = nr_k$ and recall that $\tau_n = n\theta_n$; both are integers. If $k \le \tau_n$, then the first k expectations equal $P_n\psi$, while the full-sample expectation equals $\tau_n P_n\psi + (n - \tau_n)Q_n\psi$. Therefore

$$\begin{aligned}
\mathbb{E}\mathbb{B}_{n,\psi}(r_k) &= \frac{1}{\sqrt{n}}\left[kP_n\psi - \frac{k}{n}\{\tau_n P_n\psi + (n-\tau_n)Q_n\psi\}\right] \\
&= \frac{k(n-\tau_n)}{n\sqrt{n}}(P_n\psi - Q_n\psi) \\
&= \sqrt{n}\, r_k(1-\theta_n)\delta_n = \sqrt{n}\, b_{\theta_n}(r_k)\delta_n.
\end{aligned}$$

If $k > \tau_n$, then

$$\begin{aligned}
\mathbb{E}\mathbb{B}_{n,\psi}(r_k) &= \frac{1}{\sqrt{n}}\left[\tau_n P_n\psi + (k-\tau_n)Q_n\psi - \frac{k}{n}\{\tau_n P_n\psi + (n-\tau_n)Q_n\psi\}\right] \\
&= \frac{\tau_n(n-k)}{n\sqrt{n}}(P_n\psi - Q_n\psi) \\
&= \sqrt{n}\,\theta_n(1-r_k)\delta_n = \sqrt{n}\, b_{\theta_n}(r_k)\delta_n.
\end{aligned}$$

This covers every grid point. □

Proof of Lemma F.3. If $r \le \theta$, then $b_\theta(r) = r(1-\theta)$, so $h_\theta(r) = r(1-\theta)^2/(1-r)$. Its derivative is

$$h'_\theta(r) = \frac{(1-\theta)^2}{(1-r)^2} > 0.$$

If $r > \theta$, then $b_\theta(r) = \theta(1-r)$, so $h_\theta(r) = \theta^2(1-r)/r$ and

$$h'_\theta(r) = -\frac{\theta^2}{r^2} < 0.$$

The two formulas agree at $r = \theta$ and give $h_\theta(\theta) = \theta(1-\theta)$. Hence h_θ increases strictly up to θ and decreases strictly after θ. □

Proof of Lemma F.4. Let $\mathcal{G}_k = \sigma\{\pi(1), \dots, \pi(k)\}$. Conditional on $\mathcal{G}_k$, the next draw is uniform over the $n-k$ remaining elements. Since the total sum is zero,

$$\mathbb{E}_\pi\{y_{\pi(k+1)} \mid \mathcal{G}_k\} = -\frac{S_k^\pi}{n-k}.$$

Consequently,

$$\mathbb{E}_\pi\left\{\frac{S_{k+1}^\pi}{n-k-1}\,\middle|\,\mathcal{G}_k\right\} = \frac{S_k^\pi}{n-k},$$

so $M_k = S_k^\pi/(n-k)$, $0 \le k < n$, is a Hilbert-valued martingale. For $k \le K$, $\|S_k^\pi\| \le n\,\|M_k\|$; therefore Doob's L^2 maximal inequality gives

$$\mathbb{E}_\pi \max_{k\le K}\left\|n^{-1/2}S_k^\pi\right\|^2 \le 4n\mathbb{E}_\pi\,\|M_K\|^2 = \frac{4n}{(n-K)^2}\mathbb{E}_\pi\,\|S_K^\pi\|^2\,. \tag{52}$$

It remains to evaluate the finite-population variance. Expanding the square and using uniform sampling without replacement,

$$\mathbb{E}_\pi\,\|S_K^\pi\|^2 = \frac{K}{n}\sum_{i=1}^n \|y_i\|^2 + \frac{K(K-1)}{n(n-1)}\sum_{i\ne j}\langle y_i, y_j\rangle\,.$$

The centering condition implies $\sum_{i\ne j}\langle y_i, y_j\rangle = -\sum_i \|y_i\|^2$. Hence

$$\mathbb{E}_\pi\,\|S_K^\pi\|^2 = \frac{K(n-K)}{n(n-1)}\sum_{i=1}^n \|y_i\|^2\,. \tag{53}$$

Substitution of (53) into (52) proves (49). □

G.2 Proof of the diffusion embedding result

Proof of Proposition C.1. We prove the three parts in order.

Step 1: the feature map belongs to L^2 and has the stated inner product. For fixed $t > 0$ and $x \in \mathcal{M}$, the spectral expansion in (12) gives

$$p_t(x,\cdot) = \sum_{j=0}^\infty e^{-t\lambda_j}\varphi_j(x)\varphi_j(\cdot) \quad \text{in } L^2(\mathcal{M},\mu).$$

Parseval's identity yields

$$\|p_t(x,\cdot)\|_{L^2}^2 = \sum_{j=0}^\infty e^{-2t\lambda_j}\varphi_j(x)^2 = p_{2t}(x,x),$$

where the last equality is the heat-kernel spectral expansion at time $2t$. The heat kernel is smooth on $(0,\infty)\times\mathcal{M}\times\mathcal{M}$. Compactness of $\mathcal{M}$ therefore implies $\sup_x p_{2t}(x,x) < \infty$. Hence $\Phi_t(x) \in L^2$ and the feature map is uniformly bounded.

For $x, y \in \mathcal{M}$, Parseval again gives

$$\begin{aligned}\langle \Phi_t(x), \Phi_t(y)\rangle_{L^2} &= \sum_{j=0}^\infty e^{-2t\lambda_j}\varphi_j(x)\varphi_j(y)\\ &= p_{2t}(x,y).\end{aligned}$$

This is also the semigroup identity $\int p_t(x,z)p_t(z,y)\,\mu(\mathrm{d}z) = p_{2t}(x,y)$. Continuity of $x \mapsto \Phi_t(x)$ in L^2 follows from

$$\|\Phi_t(x) - \Phi_t(y)\|_{L^2}^2 = p_{2t}(x,x) + p_{2t}(y,y) - 2p_{2t}(x,y)$$

and continuity of p_{2t}. The inner-product representation proves positive definiteness of $k_t = p_{2t}$.

Step 2: signed-measure inner products and the spectral MMD formula. Let ν and ρ be finite signed Borel measures. Boundedness of Φ_t implies Bochner integrability with respect to their total variations. Fubini's theorem for finite signed measures and the bounded integrand gives

$$\begin{aligned}\left\langle \int \Phi_t(x)\,\nu(\mathrm{d}x), \int \Phi_t(y)\,\rho(\mathrm{d}y) \right\rangle_{L^2} &= \iint \langle \Phi_t(x), \Phi_t(y)\rangle_{L^2}\,\nu(\mathrm{d}x)\rho(\mathrm{d}y) \\ &= \iint p_{2t}(x,y)\,\nu(\mathrm{d}x)\rho(\mathrm{d}y).\end{aligned}$$

Take $\nu = \rho = P - Q$. The coefficient of φ_j in $\int \Phi_t\ \mathrm{d}(P-Q)$ is

$$e^{-t\lambda_j} \int \varphi_j(x)\,(P-Q)(\mathrm{d}x).$$

Parseval's identity gives (21).

Step 3: characteristicness. The implication $P = Q \Rightarrow \mathrm{MMD}_t(P,Q) = 0$ is immediate. Conversely, suppose $\mathrm{MMD}_t(P,Q) = 0$ and put $\nu = P - Q$. Every term in the nonnegative series (21) must vanish. Since $e^{-2t\lambda_j} > 0$,

$$\int \varphi_j\ \mathrm{d}\nu = 0 \qquad \text{for every } j \geq 0. \tag{54}$$

Let $f \in C^\infty(\mathcal{M})$. Its Laplace eigenfunction expansion converges to f in every Sobolev norm and therefore uniformly after choosing a Sobolev order larger than $d/2$. If f_N denotes a finite partial sum, (54) gives $\int f_N\ \mathrm{d}\nu = 0$. Uniform convergence and finiteness of $|\nu|$ imply

$$\int f\ \mathrm{d}\nu = \lim_{N\to\infty} \int f_N\ \mathrm{d}\nu = 0.$$

Smooth functions are uniformly dense in $C(\mathcal{M})$. Hence ν annihilates every continuous function. The Riesz representation theorem for finite signed regular Borel measures on the compact space $\mathcal{M}$ yields $\nu = 0$, so $P = Q$. □

G.3 Proofs of the split and critic approximation identities

Proof of Proposition C.2. Let $S_k^t = \sum_{i=1}^k \Phi_t(Z_i)$ and $S_n^t = \sum_{i=1}^n \Phi_t(Z_i)$. At $r_k = k/n$,

$$\begin{aligned}\mathbb{B}_{n,t}(r_k) &= \frac{1}{\sqrt{n}}\left(S_k^t - \frac{k}{n}S_n^t\right) \\ &= \frac{1}{\sqrt{n}}\left[\frac{n-k}{n}S_k^t - \frac{k}{n}(S_n^t - S_k^t)\right] \\ &= \frac{k(n-k)}{n\sqrt{n}}\left\{\frac{S_k^t}{k} - \frac{S_n^t - S_k^t}{n-k}\right\} \\ &= \sqrt{n}\,r_k(1-r_k)\{m_{\widehat{P}_{1:k},t} - m_{\widehat{P}_{k+1:n},t}\}.\end{aligned}$$

This proves (22). Taking L^2 norms and using the definition of MMD_t proves (23).

For any real-valued f, the same algebra gives

$$\mathbb{G}_n(r_k, f) = \sqrt{n}\,r_k(1-r_k)\{\widehat{P}_{1:k}f - \widehat{P}_{k+1:n}f\}. \tag{55}$$

Because both empirical measures have finite first moment on the compact metric space, Kantorovich–Rubinstein duality gives

$$W_1(\widehat{P}_{1:k}, \widehat{P}_{k+1:n}) = \sup_{f\in\mathcal{F}_{\mathrm{Lip}}} |\widehat{P}_{1:k}f - \widehat{P}_{k+1:n}f|.$$

Taking the supremum in (55) proves (24). □

Proof of Proposition C.3. The lower bound follows because every f_j belongs to $\mathcal{F}_{\mathrm{Lip}}$. For the upper bound, fix $f \in \mathcal{F}_{\mathrm{Lip}}$. By the net property, there is an index $j \in \{1, \dots, m\}$ such that $\|f - f_j\|_\infty \le \varepsilon_m$. Then

$$\begin{aligned} |Pf - Qf| &\le |Pf_j - Qf_j| + |P(f - f_j)| + |Q(f - f_j)| \\ &\le \max_{1 \le \ell \le m} |Pf_\ell - Qf_\ell| + 2\varepsilon_m. \end{aligned}$$

Taking the supremum over $f \in \mathcal{F}_{\mathrm{Lip}}$ and applying Kantorovich–Rubinstein duality gives (25). The proof uses only that P and Q are probability measures, so it applies verbatim to every empirical segment pair. The same dictionary works simultaneously for all splits, making the bound simultaneous. □

Proof of Proposition C.4. For fixed $t > 0$, smoothness of the heat kernel implies that $x \mapsto \Phi_t(x) = p_t(x, \cdot)$ is differentiable as an $L^2(\mathcal{M}, \mu)$-valued map, with

$$D\Phi_t(x)v = \langle \nabla_x p_t(x, \cdot), v \rangle_{T_x \mathcal{M}}.$$

The map $(x, v) \mapsto D\Phi_t(x)v$ is continuous from the unit tangent bundle into $L^2(\mathcal{M}, \mu)$. The unit tangent bundle is compact because $\mathcal{M}$ is closed, so the supremum in (26) is finite.

Let $x, y \in \mathcal{M}$. By connectedness and Hopf–Rinow, there is a unit-speed minimizing geodesic $\gamma : [0, d(x, y)] \to \mathcal{M}$ from x to y. The fundamental theorem of calculus for Banach-valued differentiable maps yields

$$\begin{aligned} \|\Phi_t(x) - \Phi_t(y)\|_{L^2} &\le \int_0^{d(x,y)} \|D\Phi_t\{\gamma(u)\}\dot{\gamma}(u)\|_{L^2} \, \mathrm{d}u \\ &\le L_t d(x, y). \end{aligned}$$

Thus the diffusion feature map is L_t-Lipschitz.

Let π be any coupling of P and Q. Bochner integration and the coupling marginals give

$$m_{P,t} - m_{Q,t} = \iint \{\Phi_t(x) - \Phi_t(y)\} \, \pi(\mathrm{d}x, \mathrm{d}y).$$

The triangle inequality for the Bochner integral and the preceding Lipschitz bound imply

$$\mathrm{MMD}_t(P, Q) \le \iint \|\Phi_t(x) - \Phi_t(y)\|_{L^2} \, \pi(\mathrm{d}x, \mathrm{d}y) \le L_t \iint d(x, y) \, \pi(\mathrm{d}x, \mathrm{d}y).$$

Taking the infimum over all couplings proves (27). □

G.4 Proof of finite-sample permutation validity

Proof of Theorem 3.1. Let G be the permutation group on n indices. Under H_0, the ordered sample $Z = Z_{1:n}$ is exchangeable, so

$$Z \overset{d}{=} gZ \qquad \text{for every } g \in G.$$

The tuning requirement makes the complete procedure a single measurable map T_n on ordered samples.

Let $\Pi_0, \Pi_1, \dots, \Pi_B$ be iid uniform on G, independent of Z, and define $X = \Pi_0^{-1} Z$. Exchangeability implies that X is independent of $(\Pi_0, \dots, \Pi_B)$ and has the same distribution as Z. Indeed, for any measurable set A and any $g \in G$,

$$\mathbb{P}\{\Pi_0 = g, \; g^{-1}Z \in A\} = |G|^{-1}\mathbb{P}(g^{-1}Z \in A) = |G|^{-1}\mathbb{P}(Z \in A).$$

The same calculation together with independence of $\Pi_1, \dots, \Pi_B$ gives the asserted joint independence.

Conditional on X, the variables

$$Y_b = T_n(\Pi_b X), \qquad b = 0, \dots, B,$$

are iid and hence exchangeable. Moreover, $Y_0 = T_n(Z)$, while for $b \geq 1$, $Y_b = T_n(\Pi_b \Pi_0^{-1} Z)$. Conditional on Π_0, the permutations $\Pi_b \Pi_0^{-1}$ are iid uniform. Therefore the joint law of $(Y_0, \ldots, Y_B)$ is exactly the joint law of the observed statistic and the B Monte Carlo permutation statistics in (5).

For an exchangeable real vector $(Y_0, \ldots, Y_B)$, define

$$p(Y) = \frac{\#\{b \in \{0, \ldots, B\} : Y_b \geq Y_0\}}{B+1}.$$

This is the plus-one value in (5). Conditional on the unordered multiset of values, index 0 is uniformly distributed among the $B+1$ positions. For any integer j, at most j positions have upper rank at most j. Hence

$$\mathbb{P}\left\{ p(Y) \leq \frac{j}{B+1} \,\middle|\, \{Y_0, \ldots, Y_B\} \right\} \leq \frac{j}{B+1}.$$

Taking $j = \lfloor \alpha(B+1) \rfloor$ proves (6). If ties are broken by an independent uniform ordering within each tied group, the rank of index 0 is exactly uniform on $\{1, \ldots, B+1\}$ conditional on the multiset, which gives equality at attainable levels. □

G.5 Proof of the special-case reductions

Proof of Proposition C.5. For (i), the Euclidean heat kernel is

$$p_t(x, y) = (4\pi t)^{-d/2} \exp\left\{ -\frac{\|x - y\|^2}{4t} \right\}.$$

Hence

$$k_t(x, y) = p_{2t}(x, y) = (8\pi t)^{-d/2} \exp\left\{ -\frac{\|x - y\|^2}{8t} \right\}.$$

Apart from the positive constant $(8\pi t)^{-d/2}$, this is the Gaussian RBF kernel with variance parameter $\sigma^2 = 4t$. A positive rescaling of the kernel rescales the squared MMD and the unstandardized diffusion scan. Permutation calibration and the variance standardization in (14) remove this arbitrary scale.

Statement (ii) follows directly from (4) after retaining one diffusion channel and omitting the critic channel. Statement (iii) follows from (15): for $m = 1$, the vector process is the scalar CUSUM of the transformed observations $f(Z_i)$. Statement (iv) is exactly (24). For (v), choose $\kappa_n(s) = \lfloor ns \rfloor$ and $q(s) = s$ in Theorem E.6; the limit in (43) is then $W(s) - sW(1)$. □

H Proofs of the Null Limits and Multiplier Result

H.1 Proof of the joint Hilbert-space CUSUM limit

Proof of Theorem C.7. Let

$$X_i = \Psi(Z_i) - P\Psi, \qquad i = 1, \ldots, n.$$

By (N1), the X_i are iid mean-zero random elements in the separable Hilbert space $\mathcal{V}$ and $\mathbb{E}\,\|X_i\|_{\mathcal{V}}^2 < \infty$. Their covariance operator is Γ.

Step 1: apply the standard Hilbert-space invariance principle. The iid Hilbert-space functional central limit theorem (see, for example, Araujo and Giné, 1980; Horváth and Kokoszka, 2012) gives

$$\mathbb{W}_n(r) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nr \rfloor} X_i \rightsquigarrow \mathbb{W}_\Gamma(r) \quad \text{in } D([0,1]; \mathcal{V}), \tag{56}$$

where $\mathbb{W}_\Gamma$ is a $\mathcal{V}$-valued Brownian motion satisfying

$$\mathbb{E}\{\mathbb{W}_\Gamma(r)\} = 0, \qquad \mathrm{Cov}\{\mathbb{W}_\Gamma(r), \mathbb{W}_\Gamma(s)\} = (r \wedge s)\Gamma.$$

The covariance operator is trace class because $\mathrm{tr}(\Gamma) = \mathbb{E}\,\|X_1\|^2 < \infty$, which is exactly the condition needed for the Hilbert-valued Gaussian limit.

Step 2: identify the empirical bridge. Centering by $P\Psi$ cancels in (3), so

$$\mathbb{B}_{n,\Psi}(r) = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{\lfloor nr \rfloor} X_i - \frac{\lfloor nr \rfloor}{n} \sum_{i=1}^{n} X_i \right\}$$
$$= \mathbb{W}_n(r) - \frac{\lfloor nr \rfloor}{n} \mathbb{W}_n(1).$$

The Brownian motion in (56) has continuous paths almost surely. Lemma F.1 therefore yields

$$\mathbb{B}_{n,\Psi} \rightsquigarrow \mathbb{W}_\Gamma(r) - r\mathbb{W}_\Gamma(1) = \mathbb{B}_\Gamma(r).$$

For components a, b,

$$\begin{aligned} \mathrm{Cov}\{\mathbb{B}_{\Gamma,a}(r), \mathbb{B}_{\Gamma,b}(s)\} &= \{r \wedge s - rs - rs + rs\}\Gamma_{ab} \\ &= (r \wedge s - rs)\Gamma_{ab}, \end{aligned}$$

which proves (28).

Step 3: consistency of the diffusion standardizers. Fix a diffusion channel t. Put $Y_i = \Phi_t(Z_i)$, $m = m_{P,t}$, and $\overline{Y}_n = n^{-1} \sum_i Y_i$. The identity

$$\widehat{v}_{n,t} = \frac{1}{n} \sum_{i=1}^{n} \|Y_i - m\|^2 - \left\|\overline{Y}_n - m\right\|^2$$

follows by expanding the squared norms and using $n^{-1} \sum (Y_i - m) = \overline{Y}_n - m$. The scalar law of large numbers gives convergence of the first term to $\mathbb{E}\|Y_1 - m\|^2 = v_t$. The Hilbert-space law of large numbers gives $\left\|\overline{Y}_n - m\right\| \xrightarrow{\mathbb{P}} 0$. Therefore $\widehat{v}_{n,t} \xrightarrow{\mathbb{P}} v_t > 0$.

Step 4: continuity of each scan functional. On the trimmed interval $[\eta, 1-\eta]$, the denominator $r(1-r)$ is bounded below by $\eta(1-\eta) > 0$. If $x_n \to x$ uniformly in a Hilbert path and $v_n \to v > 0$, then

$$\sup_{r \in [\eta, 1-\eta]} \left| \frac{\|x_n(r)\|^2}{v_n r(1-r)} - \frac{\|x(r)\|^2}{v r(1-r)} \right| \to 0.$$

Because Skorokhod convergence to a continuous path implies uniform convergence, the diffusion scan functional is almost surely continuous at the Gaussian limit.

For the critic channel, if $x_n \to x$ uniformly in $\mathbb{R}^m$ and $A_{n,F} \to A_F$ in operator norm, then

$$\begin{aligned} &\sup_{r \in [\eta, 1-\eta]} |x_n(r)^\top A_{n,F} x_n(r) - x(r)^\top A_F x(r)| \\ &\quad \le \|A_{n,F}\|_{\mathrm{op}} \sup_r \|x_n(r) - x(r)\| \sup_r \{\|x_n(r)\| + \|x(r)\|\} + \|A_{n,F} - A_F\|_{\mathrm{op}} \sup_r \|x(r)\|^2, \end{aligned}$$

which converges to zero. Thus the critic scan is also continuous at the limit.

Step 5: finite channel maximum. There are finitely many channel functionals. Their joint convergence follows from the direct-sum process convergence and the standardizer convergence. The maximum map on a finite-dimensional vector is continuous. Applying the continuous mapping theorem and Slutsky's theorem proves (29). □

Proof of Corollary C.8. The diffusion and finite-critic limits are the corresponding coordinate projections of Theorem C.7; coordinate projection is continuous in the direct-sum Hilbert space. The continuity of the scan mappings follows from Step 4 of that proof, which gives (30) and (31). The coordinatewise maximum is also a continuous functional on $C([0,1]; \mathbb{R}^m)$ over the trimmed interval, yielding (32).

It remains to identify the Mahalanobis limit. Let $\Sigma_F = U_q \Lambda_q U_q^\top$ be the spectral decomposition on its range, where $U_q \in \mathbb{R}^{m \times q}$ has orthonormal columns and $\Lambda_q = \mathrm{diag}(\lambda_1, \ldots, \lambda_q)$ with all $\lambda_j > 0$. A Brownian bridge with covariance Σ_F can be represented as

$$\mathbb{B}_F(r) = U_q \Lambda_q^{1/2} \mathbb{B}_q^0(r),$$

where $\mathbb{B}_q^0$ has q independent standard Brownian-bridge coordinates. Since $\Sigma_F^\dagger = U_q \Lambda_q^{-1} U_q^\top$,

$$\begin{aligned}\mathbb{B}_F(r)^\top \Sigma_F^\dagger \mathbb{B}_F(r) &= \mathbb{B}_q^0(r)^\top \Lambda_q^{1/2} U_q^\top U_q \Lambda_q^{-1} U_q^\top U_q \Lambda_q^{1/2} \mathbb{B}_q^0(r) \\ &= \left\|\mathbb{B}_q^0(r)\right\|^2.\end{aligned}$$

This proves the final claim. □

H.2 Proof of the multiplier theorem

Proof of Theorem C.9. We verify the conditions of the standard multiplier functional central limit theorem in the direct-sum Hilbert space.

Step 1: centering and the multiplier partial-sum process. Let $X_i = \Psi(Z_i) - P\Psi$ and $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. The score centered by the sample mean is $X_i - \overline{X}_n = \Psi(Z_i) - \overline{\Psi}_n$. Define

$$\mathbb{W}_n^{\#}(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nr \rfloor} \xi_i (X_i - \overline{X}_n).$$

Conditionally on $Z_{1:n}$, the summands are independent, mean zero Hilbert-valued random elements.

Step 2: conditional covariance convergence. For $h, g \in \mathcal{V}$, the conditional covariance at terminal time is

$$\begin{aligned}&\mathbb{E}_\xi\left[\left\langle \mathbb{W}_n^{\#}(1), h\right\rangle \left\langle \mathbb{W}_n^{\#}(1), g\right\rangle \mid Z_{1:n}\right] \\ &\quad = \frac{1}{n}\sum_{i=1}^n \left\langle X_i - \overline{X}_n, h\right\rangle \left\langle X_i - \overline{X}_n, g\right\rangle.\end{aligned}$$

The scalar law of large numbers and $\overline{X}_n \xrightarrow{\mathbb{P}} 0$ imply

$$\frac{1}{n}\sum_{i=1}^n \left\langle X_i - \overline{X}_n, h\right\rangle \left\langle X_i - \overline{X}_n, g\right\rangle \xrightarrow{\mathbb{P}} \mathbb{E}\{\langle X_1, h\rangle \langle X_1, g\rangle\} = \langle \Gamma h, g\rangle.$$

In fact the empirical covariance converges in trace norm:

$$\left\| \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n) \otimes (X_i - \overline{X}_n) - \Gamma \right\|_1 \xrightarrow{\mathbb{P}} 0.$$

Indeed, $\|x \otimes x\|_1 = \|x\|^2$, so the strong law in the separable Banach space of trace-class operators applies to $X_i \otimes X_i$, and the sample-centering correction is $\overline{X}_n \otimes \overline{X}_n$ with trace norm $\left\|\overline{X}_n\right\|^2 = o_{\mathbb{P}}(1)$. This rules out escape of covariance mass into increasing coordinates. The same calculation over the first $\lfloor nr \rfloor$ observations gives covariance $r\Gamma$ on each fixed time grid. On each fixed finite-dimensional projection, the matrix-valued uniform law of large numbers gives this covariance convergence uniformly over $r \in [0, 1]$.

Step 3: conditional Lindeberg condition. Fix $\varepsilon > 0$. By the inequality $\left\|X_i - \overline{X}_n\right\| \le \|X_i\| + \left\|\overline{X}_n\right\|$ and the assumed $2 + \delta$ moments of both X_i and ξ_i, conditional Markov's inequality gives

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \mathbb{E}_\xi\left[\xi_i^2 \left\|X_i - \overline{X}_n\right\|^2 \mathbf{1}\{|\xi_i| \left\|X_i - \overline{X}_n\right\| > \varepsilon\sqrt{n}\} \mid Z_{1:n}\right] \\ &\quad \le \frac{\mathbb{E}|\xi_1|^{2+\delta}}{\varepsilon^\delta n^{\delta/2}} \left\{\frac{1}{n}\sum_{i=1}^n \left\|X_i - \overline{X}_n\right\|^{2+\delta}\right\}.\end{aligned}$$

The sample average in braces is $O_{\mathbb{P}}(1)$ by the law of large numbers and the triangle inequality in $L^{2+\delta}$. Hence the conditional Lindeberg quantity converges to zero in probability.

Step 4: conditional tightness. Let Π_J project onto the first J elements of an orthonormal basis of $\mathcal{V}$. Steps 2–3 give the conditional covariance and Lindeberg conditions for each finite-dimensional

projected process, so the martingale invariance principle applies (Hall and Heyde, 1980, Chapter 4). Conditional Doob's inequality controls the remaining coordinates:

$$\mathbb{E}_\xi \sup_{r\in[0,1]} \left\|(I-\Pi_J)\mathbb{W}_n^{\#}(r)\right\|^2 \le \frac{4}{n}\sum_{i=1}^{n}\left\|(I-\Pi_J)(X_i-\overline{X}_n)\right\|^2.$$

For each J, the right-hand side converges in probability to $4\mathbb{E}\left\|(I-\Pi_J)X_1\right\|^2$, which tends to zero as $J\to\infty$ because $\mathbb{E}\left\|X_1\right\|^2<\infty$. Finite-dimensional convergence and this uniform tail bound yield conditional tightness in the Hilbert space. Therefore, conditionally on the data,

$$\mathbb{W}_n^{\#} \rightsquigarrow \mathbb{W}_\Gamma \quad \text{in probability in } D([0,1];\mathcal{V}). \tag{57}$$

Step 5: bridge transformation. The multiplier bridge in (20) is

$$\mathbb{B}_{n,\Psi}^{\#}(r) = \mathbb{W}_n^{\#}(r) - \frac{\lfloor nr\rfloor}{n}\mathbb{W}_n^{\#}(1).$$

The conditional version of Lemma F.1 applied to (57) gives $\mathbb{B}_{n,\Psi}^{\#} \rightsquigarrow \mathbb{B}_\Gamma$ in conditional probability.

Step 6: statistic and conditional quantiles. The standardizers computed from the original sample converge as in the proof of Theorem C.7; they are held fixed within each multiplier draw. The scan map is continuous at continuous Gaussian paths. Thus the conditional law of the multiplier statistic converges weakly in probability to the law of T_∞. At a continuity point of that distribution, the conditional $(1-\alpha)$ quantile converges to the population quantile. The usual quantile Slutsky argument gives asymptotic rejection probability α under H_0. □

I Proofs under Alternatives and for the Full Critic Process

I.1 Verification from compactness and weak convergence

Proof of Proposition C.11. The image $\Psi(\mathcal{X})$ is compact. For orthogonal projections Π_d onto the first d elements of an orthonormal basis of $\mathcal{V}$, compactness and pointwise convergence imply

$$a_d := \sup_{x\in\mathcal{X}} \|(I-\Pi_d)\Psi(x)\| \longrightarrow 0.$$

Indeed, approximate the compact image by a finite net, choose d to control each net point, and use $\|I-\Pi_d\|\le 1$. Also $\sup_x\|\Psi(x)\|<\infty$. For fixed d, the projected coordinates and their products are bounded continuous functions of x. Weak convergence of P_n, Q_n therefore gives convergence of their means and covariance matrices. Boundedness gives the triangular-array Lindeberg condition. The finite-dimensional invariance principle for independent arrays yields the concatenated Brownian-motion limit $\Pi_d\mathbb{W}_{P,Q}$; equivalently, apply the martingale invariance principle with deterministic convergent cumulative covariance (Hall and Heyde, 1980, Chapter 4).

For the centered independent variables $U_{n,i}=\Psi(Z_i)-\mathbb{E}\Psi(Z_i)$, Doob's inequality gives the uniform tail estimate

$$\mathbb{E}\max_{k\le n}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}(I-\Pi_d)U_{n,i}\right\|^2 \le \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\|(I-\Pi_d)U_{n,i}\|^2 \le 4a_d^2.$$

The Brownian tails satisfy the corresponding bound because both segment covariance tails are at most a_d^2. Finite-dimensional convergence and these uniform tail bounds imply convergence in $D([0,1];\mathcal{V})$ to $\mathbb{W}_{P,Q}$, as in the Hilbert-space approximation argument (Araujo and Giné, 1980). The bridge map $x(r)\mapsto x(r)-rx(1)$ is continuous at continuous paths; replacing r by $\lfloor nr\rfloor/n$ changes the centered process by $o_{\mathbb{P}}(1)$ uniformly. This proves (33).

For any fixed channel ψ of Ψ, independence and boundedness give

$$\mathbb{E}\left\|\frac{1}{n}\sum_i\{\psi(Z_i)-\mathbb{E}\psi(Z_i)\}\right\|^2 = O(n^{-1}).$$

Finite-dimensional approximation as above shows that $P_n\psi \to P\psi$ and $Q_n\psi \to Q\psi$ in norm. Bounded scalar variances and weak convergence similarly give convergence of $n^{-1}\sum_i \|\psi(Z_i)\|^2$ to $\nu\|\psi\|^2$. Subtracting the squared pooled mean proves convergence of the empirical variance to $\nu\|\psi - \nu\psi\|^2$. Applying the same argument entrywise gives convergence of every fixed finite covariance matrix. Taking continuous reciprocals, inverses or rank-stable generalized inverses gives the stated standardizers. Finally, when $P = Q$, the concatenated process has covariance $(r \wedge s)\Gamma_P$, so its bridge is $\mathbb{B}_{\Gamma_P}$. □

I.2 Proof of the triangular-drift theorem

Proof of Theorem 3.2. Let

$$\widetilde{\mathbb{B}}_n(r) = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{\lfloor nr \rfloor} U_{n,i} - \frac{\lfloor nr \rfloor}{n} \sum_{i=1}^{n} U_{n,i} \right\}.$$

By Assumption C.10, $\widetilde{\mathbb{B}}_n \rightsquigarrow \mathbb{Z}$ and therefore

$$\sup_{r\in[0,1]} \left\| \widetilde{\mathbb{B}}_n(r) \right\| = O_{\mathbb{P}}(1). \tag{58}$$

At every grid point, Lemma F.2 gives the exact decomposition

$$\mathbb{B}_{n,\Psi}(r_k) = \widetilde{\mathbb{B}}_n(r_k) + \sqrt{n}\, b_{\theta_n}(r_k)\delta_n. \tag{59}$$

Both empirical processes are extended as step functions between grid points. For $r \in [0,1]$, write $r_n(r) = \lfloor nr \rfloor / n$, so that $\sup_r |r_n(r) - r| \le n^{-1}$ and the deterministic mean term at r is $\sqrt{n}\, b_{\theta_n}\{r_n(r)\}\delta_n$.

Part (i): local alternatives. Assume $\sqrt{n}\delta_n \to h$. Since $\theta_n \to \theta$ and the map $(\vartheta, r) \mapsto b_\vartheta(r)$ is uniformly continuous on $[\eta, 1-\eta] \times [0,1]$,

$$\begin{aligned} &\sup_{r\in[0,1]} \left\| \sqrt{n}\, b_{\theta_n}\{r_n(r)\}\delta_n - b_\theta(r)h \right\| \\ &\quad \le \sup_r |b_{\theta_n}\{r_n(r)\} - b_\theta(r)| \left\| \sqrt{n}\delta_n \right\| + \sup_r |b_\theta(r)| \left\| \sqrt{n}\delta_n - h \right\| \longrightarrow 0. \end{aligned}$$

Combining this deterministic convergence with $\widetilde{\mathbb{B}}_n \rightsquigarrow \mathbb{Z}$ in (59) and applying Slutsky's theorem proves (7).

Part (ii): fixed alternatives. Assume $\delta_n \to \delta$ and $A_n \xrightarrow{\mathbb{P}} A$. The latter convergence and boundedness of A imply $\|A_n\|_{\mathrm{op}} = O_{\mathbb{P}}(1)$. Divide (59) by $\sqrt{n}$:

$$n^{-1/2}\mathbb{B}_{n,\Psi}(r) = b_{\theta_n}\{r_n(r)\}\delta_n + n^{-1/2}\widetilde{\mathbb{B}}_n(r).$$

By (58), the second term is $o_{\mathbb{P}}(1)$ uniformly in r. Uniform continuity of b_θ and convergence of δ_n give

$$\sup_{r\in[\eta,1-\eta]} \left\| n^{-1/2}\mathbb{B}_{n,\Psi}(r) - b_\theta(r)\delta \right\| \xrightarrow{\mathbb{P}} 0.$$

For $x_n(r) = n^{-1/2}\mathbb{B}_{n,\Psi}(r)$ and $x(r) = b_\theta(r)\delta$,

$$\begin{aligned} &\sup_r |\, \langle x_n(r), A_n x_n(r)\rangle - \langle x(r), Ax(r)\rangle \,| \\ &\quad \le \|A_n\|_{\mathrm{op}} \sup_r \|x_n(r) - x(r)\| \sup_r\{\|x_n(r)\| + \|x(r)\|\} + \|A_n - A\|_{\mathrm{op}} \sup_r \|x(r)\|^2 \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Because $r(1-r)$ is bounded away from zero on the trimmed interval,

$$\sup_{r\in[\eta,1-\eta]} \left| \frac{n^{-1}\,\langle \mathbb{B}_{n,\Psi}(r), A_n\mathbb{B}_{n,\Psi}(r)\rangle}{r(1-r)} - \langle \delta, A\delta\rangle\, h_\theta(r) \right| \xrightarrow{\mathbb{P}} 0. \tag{60}$$

The discrete grid becomes dense in $[\eta, 1-\eta]$, and the continuous limit has maximum $\theta(1-\theta)\,\langle \delta, A\delta\rangle$ by Lemma F.3. Taking maxima in (60) proves (8). Positivity of $\langle \delta, A\delta\rangle$ makes the limit positive, so the unscaled statistic grows at rate n.

Part (iii): location consistency. Fix $\varepsilon > 0$. By the unique-maximizer property and compactness, there is $c_\varepsilon > 0$ such that

$$\langle \delta, A\delta \rangle \, h_\theta(\theta) - \sup_{|r-\theta| \geq \varepsilon} \langle \delta, A\delta \rangle \, h_\theta(r) \geq c_\varepsilon.$$

On the event that the uniform difference in (60) is less than $c_\varepsilon/3$ and the grid contains a point r_{k_n} within $1/n$ of θ, the criterion at r_{k_n} exceeds every criterion value with $|r - \theta| \geq \varepsilon$ for all sufficiently large n. The probability of this event tends to one. Hence $\mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon) \to 0$.

For the combined estimator, each signal-bearing channel has a deterministic limit equal to a positive channel constant times the common function $h_\theta(r)$. The maximum over finitely many channels is therefore $Ch_\theta(r)$ with $C > 0$, and it has the same unique maximizer. Channels with zero signal contribute zero at the n scale. Repeating the gap argument proves combined consistency. □

Proof of Corollary C.12. Take $\Psi = \Phi_t$ in Theorem 3.2. Let

$$v_{\theta,t} = \theta \, \mathbb{E}_P \left\| \Phi_t(Z) - m_{P,t} \right\|^2 + (1-\theta) \, \mathbb{E}_Q \left\| \Phi_t(Z) - m_{Q,t} \right\|^2 + \theta(1-\theta) \left\| m_{P,t} - m_{Q,t} \right\|^2.$$

This is the limiting pooled feature variance. Boundedness of Φ_t on a closed manifold gives the centered triangular-array invariance principle for fixed P and Q by applying the Hilbert-space FCLT separately to the pre- and post-change blocks and concatenating the independent limits. The law of large numbers gives $\widehat{v}_{n,t} \xrightarrow{\mathbb{P}} v_{\theta,t}$. Moreover, $v_{\theta,t} > 0$ whenever $P \neq Q$, because the final term is then positive by Proposition C.1. Apply Theorem 3.2 with $A = v_{\theta,t}^{-1} I$.

By Proposition C.1, $P \neq Q$ implies $\delta = m_{P,t} - m_{Q,t} \neq 0$. Therefore $\langle \delta, A\delta \rangle = \|\delta\|^2 / v_{\theta,t} > 0$. Parts (ii)–(iii) of Theorem 3.2 prove divergence and location consistency. □

Proof of Corollary C.13. Take $\Psi = F_m$. For trace standardization, $A = I_m / \operatorname{tr}(\Sigma_{\theta,F})$. If $\delta_F \neq 0$ and the trace is positive, then $\delta_F^\top A \delta_F = \|\delta_F\|^2 / \operatorname{tr}(\Sigma_{\theta,F}) > 0$. If the trace is zero, $F_m(Z)$ is constant under the pooled law and cannot have different segment means, so that case is incompatible with $\delta_F \neq 0$. Theorem 3.2 applies.

For Mahalanobis standardization, assume

$$A_n \xrightarrow{\mathbb{P}} \Sigma_{\theta,F}^\dagger.$$

The covariance in (34) contains $\theta(1-\theta)\delta_F \delta_F^\top$. Every vector in the null space of $\Sigma_{\theta,F}$ is therefore orthogonal to δ_F. Thus δ_F belongs to the covariance range, and, if $\delta_F \neq 0$, then

$$\delta_F^\top \Sigma_{\theta,F}^\dagger \delta_F > 0.$$

The conclusion follows from Theorem 3.2. □

I.3 Proofs for residual distance features

Proof of Proposition 3.3. Pooled means and covariance blocks are invariant to row permutations. Thus the fitted coefficient, rank decision and variance standardizer are shared by all orderings, while residual rows are permuted with their observations. Theorem 3.1 gives finite-sample validity, including the rule $\widehat{B}_H = 0$ when $\widehat{\Sigma}_{HH}$ is singular. This argument also permits pooled permutation-equivariant feature construction.

For the fixed-feature limits, boundedness gives convergence of empirical moments under the null and the fixed two-segment alternative. Positive definiteness implies $\widehat{B}_H \to B_H$ in probability. With $L = (-B_H, I_m)$ and $\widehat{L} = (-\widehat{B}_H, I_m)$, the residual bridge is exactly $\widehat{L}\mathbb{B}_{n,(H,F_m)}$ because centering constants cancel. Its empirical variance converges to

$$\operatorname{tr}(L\Sigma L^\top) = \operatorname{tr}(\Sigma_{FF} - \Sigma_{FH}\Sigma_{HH}^{-1}\Sigma_{HF}) = v_\perp.$$

Under the null, the joint finite-dimensional feature bridge converges to a Brownian bridge. Applying the convergent linear map gives the residual limit. The cross-covariance vanishes since $\Sigma_{HF} - \Sigma_{HH}B_H^\top = 0$, so the limiting diffusion and residual bridges are independent. This independence also covers diffusion channels whose coordinates lie in the span of H.

Under a fixed change, the residual quadratic metric on the joint feature space converges to $L^\top L/v_\perp$. The limiting mean contrast is $L(\delta_H, \delta_F)^\top = \delta_F - B_H\delta_H$. Theorem 3.2 gives the scan slope and localization: each nonzero channel has the same uniquely maximized normalized triangular drift. Finally, the covariance subtraction is positive semidefinite, with positive trace when $\Sigma_{FH} \neq 0$. If $\delta_H = 0$, the numerator is unchanged, proving (iii). □

Proof of Proposition D.1. The pooled law is uniform. Integrating $|x|\cos(kx)$ on $[-\pi, \pi]$ gives coefficient $2\{(-1)^k - 1\}/(\pi k^2)$, proving (37). Orthogonal projection onto H removes exactly the modes $k \leq J$. Under the uniform law, sine and cosine modes are orthogonal and each has variance $1/2$. Consequently,

$$v_\perp = \frac{8m}{\pi^2} \sum_{k>J,\ k \text{ odd}} k^{-4}, \qquad (\delta_\perp)_\ell = -\frac{4\varepsilon}{\pi j^2}\cos(ja_\ell).$$

Since $m \nmid 2j$, the finite geometric sum gives $\sum_\ell \cos^2(ja_\ell) = m/2$. Substitution proves (38). For $J = 2$, only the first odd harmonic is removed and $\sum_{k\geq 1,\ k \text{ odd}} k^{-4} = \pi^4/96$. Even frequencies have zero coefficient in every distance feature; frequencies above J also have zero mean contrast in H. □

I.4 Power of the permutation-calibrated test

Proof of Theorem C.14. The fixed-alternative conclusion of Theorem 3.2 applied to the finite direct-sum score and the finite channel maximum gives

$$T_n/n \xrightarrow{\mathbb{P}} c \tag{61}$$

for some $c > 0$.

It remains to control a random reordering. Conditional on the observed pooled sample, set

$$Y_{n,i} = \Psi(Z_i) - \overline{\Psi}_n, \qquad \overline{\Psi}_n = n^{-1}\sum_{i=1}^n \Psi(Z_i).$$

Then $\sum_i Y_{n,i} = 0$. For a uniform permutation π, the permuted direct-sum bridge at $r_k = k/n$ is exactly

$$\mathbb{B}^\pi_{n,\Psi}(r_k) = n^{-1/2}\sum_{i=1}^k Y_{n,\pi(i)}.$$

Take $K_n = \lfloor n(1-\eta)\rfloor$. By Lemma F.4, conditionally on the data,

$$\mathbb{E}_\pi \sup_{k\in\mathcal{K}_n(\eta)} \left\|\mathbb{B}^\pi_{n,\Psi}(r_k)\right\|^2_{\mathcal{V}} \leq \frac{4K_n}{n-K_n}\frac{1}{n-1}\sum_{i=1}^n \|Y_{n,i}\|^2_{\mathcal{V}}. \tag{62}$$

For fixed diffusion and critic channels on a closed manifold, the score is bounded; more generally, the fixed-alternative moment assumptions and the law of large numbers make the final empirical second moment $O_\mathbb{P}(1)$. Since $K_n/(n-K_n)$ is bounded, conditional Markov's inequality applied to (62) gives

$$\sup_{k\in\mathcal{K}_n(\eta)} \left\|\mathbb{B}^\pi_{n,\Psi}(r_k)\right\|_{\mathcal{V}} = O_\mathbb{P}(1). \tag{63}$$

All channel projections are contractions up to fixed direct-sum constants. The trimming denominator is bounded away from zero, the diffusion variance standardizers converge to positive constants, and the critic matrices are bounded in operator norm in probability. Therefore (63) implies $T^\pi_n = O_\mathbb{P}(1)$. With fixed B, a union bound gives $\max_{b\leq B} T^{(b)}_n = O_\mathbb{P}(1)$. Combining this with (61),

$$\mathbb{P}\left\{\max_{b\leq B} T^{(b)}_n \geq T_n\right\} \longrightarrow 0.$$

On the complementary event, no randomized statistic exceeds the observed one and the plus-one value equals $1/(B+1)$. Thus $\widehat{p}_B \xrightarrow{\mathbb{P}} 1/(B+1)$. If this limit is no larger than α, the rejection probability converges to one. □

I.5 POOLED CIRCLE FEATURES AND FIXED-CHANGE CONSISTENCY

Proof of Theorem C.15. Write $\widehat{\nu}_n = n^{-1}\sum_i \delta_{Z_i}$. Independence and boundedness give $\widehat{\nu}_n \Rightarrow \nu$ in probability. Since distance functions on the circle are uniformly bounded and one-Lipschitz in their center, a finite-net argument gives

$$\sup_a \left| \int d(x,a)\ \mathrm{d}\widehat{\nu}_n(x) - \int d(x,a)\ \mathrm{d}\nu(x) \right| \xrightarrow{\mathbb{P}} 0.$$

The sample covering radius $h_n = \sup_x \min_i d(x, Z_i)$ tends to zero in probability: cover the circle by finitely many short arcs, each having positive probability under both segment laws. The probability that any such arc is empty tends to zero. The empirical medoid is within h_n in objective value of an unrestricted empirical minimizer. Compactness and the unique population minimum therefore imply $a_{n,1} \xrightarrow{\mathbb{P}} a_\star$.

The product empirical measure converges to $\nu \otimes \nu$, and deleting its diagonal changes the squared-distance distribution by $O(n^{-1})$. Positive continuous densities give a continuous, strictly increasing distribution for $d(X,Y)^2$ on $(0,\pi^2)$. Consequently, its sample median converges to $d_\nu^2 > 0$ and every fitted scale converges to t_ℓ^*.

To analyze landmarks, take any subsequence and a further subsequence on which the medoid converges and $h_n \to 0$ almost surely. By compactness, every limit of the first M selected points follows the continuum farthest-point rule: the maximum over the sample differs from the maximum over the whole circle by at most h_n. Distances to a converging finite set converge uniformly, so this remains true at each selection. In the continuum rule, a new point is the midpoint of a longest circular gap. Starting with $a_\star$, all gaps have length $2\pi/2^r$ after 2^r selections, by induction: every longest gap is bisected before any of its shorter children. Thus the unordered first M and first m landmark sets converge to their equally spaced grids, regardless of intermediate ties. Match each finite set to its limiting grid; the largest matched distance tends to zero in probability.

For a diffusion channel, reorder landmarks by this matching and let $G_{n,\ell}$ be their evaluated Gram matrix. Let $g_{n,\ell}(x)$ be the column of kernel values between x and these landmarks. Local constancy of J and convergence of scales and landmarks imply $G_{n,\ell} \xrightarrow{\mathbb{P}} G_\ell$ and uniform convergence of $g_{n,\ell}(x)$ to its deterministic limit $g_\ell(x)$. For a nonnegative-definite matrix G, define

$$R(G) = \sum_{\lambda > c(G)} \lambda^{-1/2} u_\lambda u_\lambda^\top, \qquad \psi_{n,\ell}(x) = R(G_{n,\ell}) g_{n,\ell}(x).$$

The sum uses an orthonormal eigenbasis, and is independent of its choice. Multiplying the usual Nyström coordinate vector by the retained eigenvector matrix gives $\psi_{n,\ell}$; this is an isometric embedding and preserves every scan and pooled variance.

The spectral gap at the cutoff makes R continuous at G_ℓ. Indeed, eigenvalues and the cutoff are continuous, and the limiting spectrum lies in disjoint neighborhoods above and below that cutoff. Choose a continuous scalar function that equals $\lambda^{-1/2}$ on the upper neighborhoods and zero on the lower ones. Its matrix function agrees with R for all nearby matrices; uniform polynomial approximation on a containing compact interval proves continuity. Hence $\psi_{n,\ell}$ converges uniformly to $R(G_\ell)g_\ell$. For the critics, the bound $|d(x,a) - d(x,b)| \le d(a,b)$ gives uniform convergence after matching their landmarks; their common anchor converges to $a_\star$. This proves the asserted simultaneous convergence for all channels without choosing convergent individual eigenvectors.

For each fitted channel the following bound holds pathwise, including when its fit uses every observation:

$$\sup_{0\le r\le 1} \left\| n^{-1/2}\{\mathbb{B}_{n,\psi_{n,j}}(r) - \mathbb{B}_{n,\psi_j}(r)\} \right\| \le 2 \sup_x \|\psi_{n,j}(x) - \psi_j(x)\| \xrightarrow{\mathbb{P}} 0.$$

For the bounded fixed map ψ_j, a maximal second-moment bound for independent centered partial sums and the deterministic mean decomposition give $n^{-1/2}\mathbb{B}_{n,\psi_j}(r) \to b_\theta(r)\delta_j$ uniformly in probability. The same holds for the fitted bridge. Uniform feature convergence and the law of large numbers also give $\widehat{v}_{n,j} \xrightarrow{\mathbb{P}} v_j > 0$. The weighted maximum of the splitwise criteria, divided by n, therefore converges uniformly on the trimmed grid to

$$\frac{b_\theta(r)^2}{r(1-r)} \max_j \frac{c_j \|\delta_j\|^2}{v_j}.$$

Its temporal factor has a unique maximum at θ, which proves the scan limit and location consistency.

Finally, condition on the pooled sample and center each fitted channel's rows. Permutation invariance keeps these rows fixed across reorderings. Apply Lemma F.4 to each channel as in (62). Division by its empirical variance cancels the empirical squared norm on the bound's right side, leaving a constant depending only on the trim and fixed weight (and the bounded factor $n/(n-1)$). A zero-variance channel contributes zero. Thus each permuted maximum is $O_{\mathbb{P}}(1)$, and so is the maximum of fixed B permutations. Since $T_n/n \xrightarrow{\mathbb{P}} c > 0$, with probability tending to one every permuted value is smaller than T_n. On this event $\widehat{p}_B = 1/(B+1) \le \alpha$, proving consistency of the test. □

Proof of Corollary C.16. The pooled density f_ν is positive and even. On $|x| < \pi/2$ it exceeds its antipodal value: the density of P is proportional to $e^{4\cos x}$, and that of Q to $e^{8\cos a\cos x}\cosh(8\sin a\sin x)$, with $\cos a > 0$. For $g(b) = \int d(x,b)\,\mathrm{d}\nu(x)$, differentiation gives $g''(b) = 2\{f_\nu(b) - f_\nu(b-\pi)\}$. Evenness gives $g'(0) = g'(\pi) = 0$. The second derivative is positive on $(0,\pi/2)$ and negative on $(\pi/2,\pi)$, so $g'(b) > 0$ throughout $(0,\pi)$. Thus zero is the unique medoid.

The scale and spectral gaps can be bounded by rapidly convergent series. Let I_h be the modified Bessel function and set

$$p_h = \frac{I_h(4)}{I_0(4)}, \qquad q_h = \frac{I_h(8)}{I_0(8)}\cos(ha), \qquad m_h = \tfrac{1}{2}(p_h + q_h).$$

These are the cosine moments of P, Q, ν; all sine moments vanish. Independence gives the squared-distance distribution

$$H(s) = \frac{\sqrt{s}}{\pi} + \frac{2}{\pi}\sum_{h\ge 1} m_h^2 \frac{\sin(h\sqrt{s})}{h}, \qquad 0 \le s \le \pi^2.$$

Using interval arithmetic with positive Bessel series and explicit remainder bounds yields

$$0.270738 < d_\nu^2 < 0.270739, \qquad 0.0280 < p_2 - q_2 < 0.0281.$$

For completeness, $I_h(\kappa) = \sum_{r\ge 0}(\kappa/2)^{2r+h}/\{r!(r+h)!\}$ has a geometrically bounded positive tail after any term beyond its maximum. The bound $|m_h| \le 4^h e^{16/(h+1)}/h!$ bounds the absolute remainder of H after $h = 40$ by less than 10^{-40}. The enclosed values of H at the displayed endpoints lie strictly below and above $1/2$, respectively. Cosine multiples are evaluated by their polynomial recurrence from $\cos a = p_1 I_0(8)/I_1(8)$.

For the stopping rule (36), put $u = 2t$. Integration by parts bounds the integral with lower endpoint $b > 0$ between $e^{-ub^2}(1 - 1/(2ub^2))/(2ub)$ and $e^{-ub^2}/(2ub)$. Across the entire median interval these bounds give $J = 51, 25, 12, 6, 3$ for the five scales, with strict inequalities at the stopping step and its predecessor. Therefore each stopping count is locally constant at its population scale.

On the uniform 128-point grid, all these counts are below 64. The circulant Gram eigenvalues are $128/(2\pi)$ for the constant mode and $128e^{-2th^2}/(2\pi)$ for each sine and cosine mode $1 \le h \le J$; the remaining eigenvalues are zero. The smallest positive eigenvalue divided by the largest exceeds 2.7×10^{-10} at every scale, above the relative cutoff 10^{-10}, which also dominates the absolute cutoff. Thus every evaluated Fourier mode is retained, the limiting ranks are $103, 51, 25, 13, 7$, and the cutoff gaps are positive. The retained second cosine mode has a nonzero contrast. The diffusion channels and distance critics are nonconstant under the full-support pooled law, so their variances are positive. These observations verify all the theorem's conditions. The supplied script `code/circle_population_bounds.py` reproduces the interval calculations. □

I.6 Proof of the sequential critic bridge theorem

Proof of Theorem E.2. Define the sequential empirical process

$$\alpha_n(r,f) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nr\rfloor}\{f(Z_i) - Pf\}, \qquad (r,f) \in [0,1]\times\mathcal{F}_0.$$

Step 1: ordinary Donsker property. The envelope F is bounded and therefore belongs to $L^2(P)$. The bracketing integral in (39), together with pointwise measurability, is the standard sufficient condition for P-Donsker convergence (van der Vaart and Wellner, 1996, Theorem 2.5.6 and surrounding results). Thus the empirical process indexed by $\mathcal{F}_0$ is asymptotically tight and Gaussian.

Step 2: sequential extension. The sequential empirical-process theorem for a Donsker class gives

$$\alpha_n \rightsquigarrow \mathbb{K}_P \quad \text{in } \ell^\infty([0,1] \times \mathcal{F}_0), \tag{64}$$

where $\mathbb{K}_P$ is the Kiefer–Müller process with covariance

$$\operatorname{Cov}\{\mathbb{K}_P(r,f), \mathbb{K}_P(s,g)\} = (r \wedge s) \operatorname{Cov}_P\{f(Z), g(Z)\}.$$

The same bracketing entropy that controls the class index supplies asymptotic equicontinuity jointly with the time index because the indicator class $\{\mathbf{1}(u \le r) : r \in [0,1]\}$ has finite Vapnik–Chervonenkis dimension.

Step 3: bridge transformation. The critic bridge satisfies

$$\mathbb{G}_n(r,f) = \alpha_n(r,f) - \frac{\lfloor nr \rfloor}{n} \alpha_n(1,f).$$

The limiting Kiefer process admits a version uniformly continuous in the canonical semimetric, and in particular continuous in r. The map $z(r,f) \mapsto z(r,f) - rz(1,f)$ is continuous at such paths in the supremum norm. The discrepancy caused by replacing r with $\lfloor nr \rfloor / n$ is bounded by $n^{-1} \sup_f |\alpha_n(1,f)| = o_{\mathbb{P}}(1)$ because terminal empirical-process tightness implies the supremum is $O_{\mathbb{P}}(1)$. Applying the continuous mapping theorem to (64) gives (40) with

$$\mathbb{G}_P(r,f) = \mathbb{K}_P(r,f) - r\mathbb{K}_P(1,f).$$

A direct covariance calculation yields (41).

Step 4: supremum statistic. The map

$$z \mapsto \sup_{r \in [\eta, 1-\eta]} \sup_{f \in \mathcal{F}_0} \frac{|z(r,f)|}{\sqrt{r(1-r)}}$$

is Lipschitz with respect to the supremum norm, with constant $\{\eta(1-\eta)\}^{-1/2}$. The continuous mapping theorem proves (42). □

Proof of Proposition E.3. For (i), a finite class has finite bracketing number for every positive bracket size; the entropy integral is finite.

For (ii), cover the compact parameter set $\Theta \subset \mathbb{R}^q$ by Euclidean balls of radius ρ. Standard volume comparison gives at most $C\rho^{-q}$ balls. If ϑ lies within ρ of center ϑ_j, then

$$f_{\vartheta_j}(x) - \rho L(x) \le f_\vartheta(x) \le f_{\vartheta_j}(x) + \rho L(x).$$

This is an $L^2(P)$ bracket of size at most $2\rho \|L\|_{P,2}$. Taking ρ proportional to ε gives

$$\log N_{[]}(\varepsilon, \mathcal{F}_0, L^2(P)) \le C_1 + C_2 q \log(1/\varepsilon).$$

The square root of a logarithm is integrable at zero, so (39) is finite.

For (iii), cover the compact manifold by finitely many smooth coordinate charts and use a smooth partition of unity. The classical Kolmogorov–Tikhomirov entropy bound for a bounded Hölder-s ball in d dimensions (Kolmogorov and Tikhomirov, 1959), transferred through the finite atlas, gives

$$\log N(\varepsilon, \mathcal{F}_0, \|\cdot\|_\infty) \le C\varepsilon^{-d/s}$$

for small ε. Supremum-norm covers yield $L^2(P)$ brackets of at most twice the same radius, so the bracketing entropy has the same order. The entropy integral is bounded by a constant multiple of

$$\int_0^1 \varepsilon^{-d/(2s)} \, \mathrm{d}\varepsilon,$$

which is finite exactly when $d/(2s) < 1$, or $s > d/2$. Setting $s = 1$ verifies this sufficient condition for $d = 1$ but not for $d \ge 2$. □

Proof of Theorem E.4. Choose $f_0 \in \mathcal{F}_0$ such that $|Pf_0 - Qf_0| > \Delta_{\mathcal{F}_0}(P,Q)/2$; if the supremum is attained, f_0 may be chosen as a maximizer. Apply Lemma F.2 to the scalar score f_0. With $\theta_n = \lfloor n\theta \rfloor / n$, the exact expectation is

$$\mathbb{E}\mathbb{G}_n(r_k, f_0) = \sqrt{n}\, b_{\theta_n}(r_k)(Pf_0 - Qf_0).$$

Since $\theta_n \to \theta$, the functions b_{θ_n} converge uniformly to b_θ.

Because f_0 is bounded, the centered scalar bridge is $O_{\mathbb{P}}(1)$ uniformly by the ordinary scalar functional central limit theorem. At a grid point r_{k_n} converging to θ,

$$|\mathbb{G}_n(r_{k_n}, f_0)| = \sqrt{n}\,\theta(1-\theta)|Pf_0 - Qf_0| + o_{\mathbb{P}}(\sqrt{n}).$$

The denominator $\sqrt{r_{k_n}(1-r_{k_n})}$ converges to a positive constant. The full supremum statistic is at least this single-critic value and therefore diverges.

If $\mathcal{F}_0 = \mathcal{F}_{\mathrm{Lip}}$, Kantorovich–Rubinstein duality gives $\Delta_{\mathcal{F}_0}(P,Q) = W_1(P,Q)$. On a compact metric space, W_1 is a metric on probability measures, so it is positive whenever $P \neq Q$. □

J Proofs for Predictable Scores and the Logarithmic Clock

J.1 Proofs of the moment and clock lemmas

Proof of Lemma F.5. Let $X_1', \ldots, X_m'$ be an independent copy of $X_1, \ldots, X_m$. Since $\mathbb{E}(X_i' \mid X_1, \ldots, X_m) = 0$, conditional Jensen's inequality gives

$$\begin{aligned}\mathbb{E}\left\|\sum_{i=1}^m X_i\right\|^4 &= \mathbb{E}\left\|\mathbb{E}(\sum_{i=1}^m (X_i - X_i') \mid X_1, \ldots, X_m)\right\|^4 \\ &\leq \mathbb{E}\left\|\sum_{i=1}^m (X_i - X_i')\right\|^4.\end{aligned}$$

Each difference $X_i - X_i'$ is symmetric. If $\epsilon_1, \ldots, \epsilon_m$ are independent Rademacher variables, independent of everything else, then

$$\sum_{i=1}^m (X_i - X_i') \overset{d}{=} \sum_{i=1}^m \epsilon_i (X_i - X_i').$$

For deterministic Hilbert vectors $x_1, \ldots, x_m$, expansion of the fourth power and elimination of Rademacher monomials with an odd exponent give the Hilbert-space Khintchine bound

$$\mathbb{E}_\epsilon \left\|\sum_{i=1}^m \epsilon_i x_i\right\|^4 \leq 3\left(\sum_{i=1}^m \|x_i\|^2\right)^2. \tag{65}$$

Here $\|X_i - X_i'\| \leq 2K$, so conditioning on the samples and applying (65) yields

$$\mathbb{E}_\epsilon \left\|\sum_{i=1}^m \epsilon_i (X_i - X_i')\right\|^4 \leq 3(4mK^2)^2 = 48m^2K^4.$$

Taking expectation proves (50). □

Proof of Lemma F.6. Write $S_m = \sum_{j=1}^m \zeta_j$. From (46), derived below,

$$Q_k = \sum_{m=1}^{k-1} \frac{1}{m^2} \langle CS_m, S_m \rangle. \tag{66}$$

For $1 \leq j \leq k-1$, define

$$w_{j,k} = \sum_{m=j}^{k-1} \frac{1}{m^2}, \quad a_j = \langle C\zeta_j, \zeta_j \rangle, \quad b_{j\ell} = \langle C\zeta_j, \zeta_\ell \rangle \quad (j < \ell).$$

Expanding $\langle CS_m, S_m\rangle$ in (66) and reversing the finite sums gives

$$Q_k = \sum_{j=1}^{k-1} w_{j,k} a_j + 2 \sum_{1\le j<\ell\le k-1} w_{\ell,k} b_{j\ell}.$$

Because $C = \mathbb{E}(\zeta_1 \otimes \zeta_1)$,

$$\mathbb{E}a_j = \mathbb{E}\,\langle C\zeta_j, \zeta_j\rangle = \operatorname{tr}(C^2) = \sigma_D^2, \qquad \mathbb{E}b_{j\ell} = 0.$$

Also,

$$\sum_{j=1}^{k-1} w_{j,k} = \sum_{m=1}^{k-1} \frac{1}{m^2} \sum_{j=1}^{m} 1 = \sum_{m=1}^{k-1} \frac{1}{m} = H_{k-1}.$$

Thus $\mathbb{E}Q_k = \sigma_D^2 H_{k-1}$.

Decompose the centered clock as $Q_k - \mathbb{E}Q_k = A_k + B_k$, where

$$A_k = \sum_{j=1}^{k-1} w_{j,k}(a_j - \sigma_D^2), \qquad B_k = 2\sum_{j<\ell} w_{\ell,k} b_{j\ell}.$$

The a_j are independent. Since $w_{1,k} \le \pi^2/6$ and, for $j \ge 2$,

$$w_{j,k} \le \sum_{m=j}^{\infty} m^{-2} \le \int_{j-1}^{\infty} x^{-2}\,\mathrm{d}x = (j-1)^{-1},$$

we have

$$\mathbb{E}A_k^2 = \operatorname{Var}(a_1) \sum_{j=1}^{k-1} w_{j,k}^2 \le C_A < \infty. \tag{67}$$

The variance is finite because ζ_1 is bounded and C is bounded.

We next examine the off-diagonal terms. Distinct unordered pairs are orthogonal in L^2: if $(j,\ell) \ne (p,q)$ with $j<\ell$ and $p<q$, then

$$\mathbb{E}(b_{j\ell} b_{pq}) = 0.$$

For disjoint pairs this follows by independence and centering. If the pairs share one index, condition on the shared variable and integrate one of the two remaining centered independent variables; the conditional expectation is zero. Moreover,

$$\begin{aligned}\mathbb{E}b_{j\ell}^2 &= \mathbb{E}\left[\mathbb{E}\{\langle C\zeta_j, \zeta_\ell\rangle^2 \mid \zeta_j\}\right] \\ &= \mathbb{E}\,\langle C(C\zeta_j), C\zeta_j\rangle = \mathbb{E}\left\langle \zeta_j, C^3\zeta_j\right\rangle = \operatorname{tr}(C^4) =: \beta < \infty.\end{aligned}$$

Therefore

$$\mathbb{E}B_k^2 = 4\beta \sum_{\ell=2}^{k-1} (\ell-1) w_{\ell,k}^2 \le 4\beta \sum_{\ell=2}^{k-1} \frac{1}{\ell-1} \le C_B \log k. \tag{68}$$

Finally, $\mathbb{E}(A_k B_k) = 0$. Indeed, every product $(a_j - \sigma_D^2) b_{pq}$ has expectation zero: if j is not in the pair, use independence; if $j = p$ or $j = q$, condition on the other member of the pair and use the zero mean of that member. Combining (67) and (68) proves (51). □

J.2 Proof of the general variance-clock bridge

Proof of Theorem E.6. Step 1: martingale functional central limit theorem. Consider the normalized martingale array $a_n^{-1} D_{n,i}$. Assumption (M1) gives convergence of its predictable quadratic-variation process along the time map κ_n to the deterministic continuous clock q. Assumption (M2) is the conditional Lindeberg condition. The martingale functional central limit theorem (Brown, 1971; Hall and Heyde, 1980) therefore yields

$$a_n^{-1} S_n\{\kappa_n(\cdot)\} \rightsquigarrow W\{q(\cdot)\} \quad \text{in } D([0,1]). \tag{69}$$

Because q is continuous, the time-changed Brownian motion has continuous paths.

Step 2: observed total variance. At $s = 1$, (M1) gives $Q_n(n)/a_n^2 \xrightarrow{\mathbb{P}} q(1) = 1$. Assumption (M3) at $s = 1$ gives

$$\frac{V_n(n) - Q_n(n)}{a_n^2} \xrightarrow{\mathbb{P}} 0.$$

Hence

$$V_n(n)/a_n^2 \xrightarrow{\mathbb{P}} 1, \qquad V_n(n)^{1/2}/a_n \xrightarrow{\mathbb{P}} 1. \tag{70}$$

In particular, $\mathbb{P}\{V_n(n) > 0\} \to 1$.

Step 3: observed variance fraction. Uniformly in s,

$$\left| \frac{V_n\{\kappa_n(s)\}}{a_n^2} - q(s) \right| \le \frac{|V_n\{\kappa_n(s)\} - Q_n\{\kappa_n(s)\}|}{a_n^2} + \left| \frac{Q_n\{\kappa_n(s)\}}{a_n^2} - q(s) \right|.$$

The supremum tends to zero in probability by (M1) and (M3). Together with (70), this implies

$$\sup_{s \in [0,1]} |\widehat{q}_n(s) - q(s)| \xrightarrow{\mathbb{P}} 0. \tag{71}$$

Step 4: bridge transformation. By (69), joint convergence holds for the process and its terminal value. The map $x \mapsto x(s) - q(s)x(1)$ is continuous at continuous paths, so

$$a_n^{-1} \left[S_n\{\kappa_n(s)\} - q(s) S_n(n) \right] \rightsquigarrow W\{q(s)\} - q(s) W(1).$$

The difference caused by replacing q with $\widehat{q}_n$ is bounded by

$$\sup_s |\widehat{q}_n(s) - q(s)| \, |S_n(n)|/a_n.$$

The first factor is $o_{\mathbb{P}}(1)$ by (71); the second is $O_{\mathbb{P}}(1)$ by (69) at $s = 1$. Hence the replacement is $o_{\mathbb{P}}(1)$ uniformly. Finally, replace a_n by $V_n(n)^{1/2}$ using (70) and Slutsky's theorem. This proves (43). □

J.3 Proof of the logarithmic-clock theorem

Proof of Theorem E.7. Part (i): martingale difference and conditional variance. The average $\overline{\zeta}_{i-1}$ is $\mathcal{F}_{i-1}$-measurable, while ζ_i is independent of $\mathcal{F}_{i-1}$ and has mean zero. Therefore

$$\mathbb{E}(D_i \mid \mathcal{F}_{i-1}) = \left\langle \overline{\zeta}_{i-1}, \mathbb{E}(\zeta_i \mid \mathcal{F}_{i-1}) \right\rangle = 0.$$

Thus D_i is a martingale difference. For a deterministic $u \in \mathcal{H}$,

$$\mathbb{E} \left\langle u, \zeta_i \right\rangle^2 = \left\langle Cu, u \right\rangle.$$

Conditioning on $\mathcal{F}_{i-1}$ and taking $u = \overline{\zeta}_{i-1}$ gives

$$\mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1}) = \left\langle C\overline{\zeta}_{i-1}, \overline{\zeta}_{i-1} \right\rangle.$$

Let $m = i - 1$ and $S_m = \sum_{j=1}^m \zeta_j$. Independence and centering imply $\mathbb{E}(S_m \otimes S_m) = mC$. Therefore

$$\begin{aligned} \mathbb{E} D_i^2 &= m^{-2} \mathbb{E} \left\langle CS_m, S_m \right\rangle = m^{-2} \operatorname{tr}\{C\mathbb{E}(S_m \otimes S_m)\} \\ &= m^{-1} \operatorname{tr}(C^2) = \frac{\sigma_D^2}{i-1}. \end{aligned}$$

For independent copies ζ_1, ζ_2,

$$\mathbb{E} \left\langle \zeta_1, \zeta_2 \right\rangle^2 = \mathbb{E}\{\langle C\zeta_1, \zeta_1 \rangle\} = \operatorname{tr}(C^2),$$

which verifies the second expression in (45).

Part (ii-a): pointwise predictable clock. By Lemma F.6,

$$\mathbb{E}\{Q_k - \sigma_D^2 H_{k-1}\}^2 \le C_0 \log k.$$

Fix $s \in (0,1]$ and let $k_n(s) = \lfloor n^s \rfloor$ for large n. Chebyshev's inequality gives

$$\frac{Q_{k_n(s)} - \sigma_D^2 H_{k_n(s)-1}}{\sigma_D^2 \log n} \xrightarrow{\mathbb{P}} 0,$$

because the numerator has standard deviation $O(\sqrt{\log n})$ and the denominator is of order $\log n$. The harmonic-number expansion $H_m = \log m + O(1)$ yields

$$\frac{H_{k_n(s)-1}}{\log n} \longrightarrow s.$$

At $s = 0$, $k_n(0) = 1$ and $Q_1 = 0$. Thus

$$Q_{k_n(s)}/(\sigma_D^2 \log n) \xrightarrow{\mathbb{P}} s \qquad \text{for every } s \in [0,1]. \tag{72}$$

Part (ii-b): upgrade to a uniform clock. The process $k \mapsto Q_k$ is nondecreasing because every conditional-variance increment is nonnegative. Fix an integer $J \geq 1$ and grid points $s_j = j/J$. By (72) and a finite union bound,

$$\max_{0 \leq j \leq J} \left| Q_{k_n(s_j)}/(\sigma_D^2 \log n) - s_j \right| \xrightarrow{\mathbb{P}} 0.$$

If $s \in [s_j, s_{j+1}]$, monotonicity gives

$$\frac{Q_{k_n(s_j)}}{\sigma_D^2 \log n} \leq \frac{Q_{k_n(s)}}{\sigma_D^2 \log n} \leq \frac{Q_{k_n(s_{j+1})}}{\sigma_D^2 \log n}.$$

Consequently,

$$\sup_{s \in [0,1]} \left| Q_{k_n(s)}/(\sigma_D^2 \log n) - s \right| \leq \max_{0 \leq j \leq J} \left| Q_{k_n(s_j)}/(\sigma_D^2 \log n) - s_j \right| + J^{-1} + o(1),$$

where the $o(1)$ accounts only for the floor operation. First let $n \to \infty$ for fixed J, and then let $J \to \infty$. This proves the first uniform convergence in (47).

Part (ii-c): observed variation. Let

$$M_k = V_k - Q_k = \sum_{i=2}^{k} \{D_i^2 - \mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1})\}.$$

This is a square-integrable martingale. Doob's L^2 inequality and conditional-variance domination give

$$\begin{aligned} \mathbb{E} \max_{2 \leq k \leq n} |M_k|^2 &\leq 4 \sum_{i=2}^{n} \mathbb{E}\{D_i^2 - \mathbb{E}(D_i^2 \mid \mathcal{F}_{i-1})\}^2 \\ &\leq 4 \sum_{i=2}^{n} \mathbb{E} D_i^4. \end{aligned} \tag{73}$$

Because $\|\zeta_i\| \leq K$, $|D_i| \leq K \left\|\overline{\zeta}_{i-1}\right\|$. With $m = i-1$, Lemma F.5 gives

$$\mathbb{E} D_i^4 \leq K^4 m^{-4} \mathbb{E} \left\|S_m\right\|^4 \leq 48 K^8 m^{-2}.$$

The series $\sum_{m \geq 1} m^{-2}$ converges, so the right-hand side of (73) is bounded uniformly in n. Hence

$$\max_{k \leq n} |V_k - Q_k|/(\log n) \xrightarrow{\mathbb{P}} 0. \tag{74}$$

Combining (74) with the uniform predictable clock proves the second convergence in (47).

Part (iii): functional bridge. The general theorem is indexed from zero, whereas the present scores begin at $i = 2$. Reindex the row by setting, for $1 \leq j \leq n-1$,

$$D^\star_{n-1,j} = D_{j+1}, \qquad \mathcal{F}^\star_{n-1,0} = \mathcal{F}_1, \qquad \mathcal{F}^\star_{n-1,j} = \mathcal{F}_{j+1}, \qquad \kappa^\star_{n-1}(s) = k_n(s) - 1.$$

Then $\kappa^\star_{n-1}(0) = 0$, $\kappa^\star_{n-1}(1) = n - 1$, and the corresponding partial sums and variations satisfy

$$S^\star_{n-1}\{\kappa^\star_{n-1}(s)\} = S_{k_n(s)}, \qquad Q^\star_{n-1}\{\kappa^\star_{n-1}(s)\} = Q_{k_n(s)}, \qquad V^\star_{n-1}\{\kappa^\star_{n-1}(s)\} = V_{k_n(s)}.$$

Apply Theorem E.6 to this reindexed array with $a^2_{n-1} = \sigma^2_D \log n$ and $q(s) = s$. Condition (M1) is the first convergence in (47). For (M2), note that $|D_i| \le K^2$ almost surely because both $\|\overline{\zeta}_{i-1}\|$ and $\|\zeta_i\|$ are at most K. Since $a_{n-1} \to \infty$, for every fixed $\varepsilon > 0$ the Lindeberg indicators $\mathbf{1}(|D_i| > \varepsilon a_{n-1})$ are identically zero for all sufficiently large n. Condition (M3) follows from (74). The general theorem therefore gives

$$V_n^{-1/2}\left\{S_{k_n(s)} - \frac{V_{k_n(s)}}{V_n} S_n\right\} \rightsquigarrow W(s) - sW(1),$$

which is (44).

Part (iv): collapse of calendar time. Fix $r \in (0, 1]$. The pointwise clock argument with $k = \lfloor nr \rfloor$ gives

$$\frac{Q_{\lfloor nr \rfloor}}{\sigma^2_D \log n} = \frac{\sigma^2_D \log(nr) + o_{\mathbb{P}}(\log n)}{\sigma^2_D \log n} \xrightarrow{\mathbb{P}} 1,$$

because $\log(nr)/\log n \to 1$. Part (ii) at $s = 1$ gives $Q_n/(\sigma^2_D \log n) \xrightarrow{\mathbb{P}} 1$. Dividing the two convergences and applying Slutsky's theorem proves (48). □