LEARNING SCALAR QUADRATIC CONTRASTS FOR HIGH-DIMENSIONAL CHANGE POINT LOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A high-dimensional distribution can change through many weak interactions while every coordinate mean and variance remains fixed. We study how to aggregate this signal into one quadratic contrast for change point localization. A normalized matrix representation identifies the Frobenius jump with the L^2 change in the degree-two density projection. For bounded densities with low-rank jumps in their degree-two coefficient matrices, a sufficiently strong pilot signal allows the matrix scan to learn a quadratic contrast retaining at least half of the Frobenius jump. With enough data to estimate its two means, independent evaluation observations then yield a small-jump localization error proportional to κ^{-2}, up to confidence factors, for a fixed density bound. The pilot pays for learning the aggregate contrast, while its retained signal governs scalar precision. Explicit matrix bounds extend detection and isolation to three temporal dependence models. Under geometric absolute regularity, deterministic spacing also yields a complete refinement guarantee in original time units. Experiments connect retained signal to localization accuracy, explain the effect of working rank, and quantify the sampling and resolution costs of spacing.

1 INTRODUCTION

When the relationships among many variables change, the transition can be difficult to see in any one coordinate. A sensor array may keep the same marginal means and variances while its measurements become differently associated. Evidence for the transition is then spread across many interactions. To locate the change accurately, an estimator must collect this evidence and separate it from fluctuations in the remaining directions.

A dense interaction change can have a simple spectral structure. Consider two groups of coordinates whose cross-group interactions change together. Every entry of the interaction block changes, but the jump matrix has rank two. A quadratic contrast can combine these weak entries into one scalar mean change. The statistical task is to learn a contrast that retains enough of their aggregate signal for precise localization.

We develop this reduction for changes in degree-two density coefficients. A normalized symmetric feature matrix represents coordinate means, marginal quadratic terms, and pairwise interactions. The Frobenius norm of its mean jump equals the L^2 norm of the corresponding density projection. A normalized quadratic contrast weights the changing spectral components by their signed eigenvalues to collect this full jump into a scalar mean difference. For a fixed density upper bound, the scalar variance is bounded independently of dimension. Low rank limits the number of components to learn while allowing dense changes across the original coordinates.

The main question is how much of this signal survives direction estimation. Under a rank-dependent pilot signal condition, our learned contrast retains at least half of the Frobenius jump. With sufficient samples to estimate the two scalar means, independent evaluation observations support a localization error proportional to $\nu^2/\kappa^2 + B/\kappa$, up to a confidence factor. For a fixed density bound, ν^2 and B are independent of dimension. Learning the aggregate contrast therefore converts a structured multivariate change into a scalar problem whose precision is governed by the retained signal.

Temporal dependence affects both stages. Repeated information increases the matrix noise faced by the pilot, and nearby evaluation observations can remain dependent on the learned direction. Three

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.

concentration bounds quantify the first effect. Under geometric absolute regularity, deterministic spacing and coupling also justify separate scalar evaluation. The final error records both the reduced learning sample and the resolution of retained observations in original time.

Our contributions establish three links from the distributional change to its location:

- A normalized representation identifies the degree-two distributional signal and shows how a quadratic contrast aggregates it with controlled scalar noise.
- A rank-dependent signal condition guarantees isolation and retention of at least half the jump. Explicit matrix bounds quantify learning costs under independence and three temporal dependence assumptions.
- Windows chosen from the pilot supply pure blocks for estimating scalar means. Separate evaluation then gives complete localization guarantees under independence and geometric absolute regularity.

The experiments follow these links. At $d = 128$, rank-two direction learning gives median localization error 9.5, compared with 140.5 when every empirical direction is retained, using the same scalar objective and data allocation. Increasing the total sample size improves signal retention and localization at larger dimension. Dependent sequences then separate the costs of learning from fewer observations and reporting locations on a coarser time grid.

1.1 Relation to projection and multiscale estimation

Projection methods provide the closest context. Sparse projection targets high-dimensional mean changes (Wang & Samworth, 2018). For covariance changes, Wang et al. (2021) learn projection directions from independent data and localize using squared projections. Their signal is measured by the operator norm of the covariance jump. Here, the normalized coefficient matrix joins first moments and quadratic interactions, and the target is the Frobenius signal of its jump. A quadratic contrast combines multiple changing spectral components with their signs intact. The density bound controls the noise of this aggregate, giving a localization guarantee tied to the full degree-two projection signal.

The isolation step draws on seeded segmentation (Kovács et al., 2023) and narrowest-over-threshold estimation (Baranowski et al., 2019). Our discrete deletion rule and pilot window expansion supply pure scalar anchors. Nonparametric methods under mixing estimate changes in smooth multivariate densities (Madrid Padilla et al., 2023). Targeting the degree-two density projection lets us aggregate structured changes without estimating the full density, while tracking the dependence costs of learning and evaluation.

2 Learning a scalar quadratic contrast

The method represents the distributional change as a matrix mean jump, estimates a useful direction, and localizes its scalar projection. The allocation in Figure 1 is part of the estimator: the pilot chooses the windows and contrast, while separate evaluation observations determine the final location.

2.1 Represent the change and identify the ideal contrast

Let $X_1, \ldots, X_n \in [-1, 1]^d$, and let μ be product uniform probability measure on this cube. There are K change points $0 = \eta_0 < \eta_1 < \cdots < \eta_K < \eta_{K+1} = n$. For $\eta_k < i \leq \eta_{k+1}$, the marginal density of X_i relative to μ is $f_k \leq L$. Write $p = d + 1$ and $p_2(x) = (3x^2 - 1)/2$. Define the symmetric feature matrix

$$\begin{aligned} H_{00}(x) &= 1, \qquad & H_{0j}(x) &= H_{j0}(x) = \sqrt{3/2}\, x_j, \\ H_{jj}(x) &= \sqrt{5}\, p_2(x_j), \quad & H_{ij}(x) &= H_{ji}(x) = 3x_i x_j/\sqrt{2} \quad (1 \leq i < j \leq d). \end{aligned} \tag{1}$$

For $M_k = \mathbb{E}_{f_k} H(X)$ and $D_k = M_k - M_{k-1}$, assume

$$\operatorname{rank}(D_k) \leq r, \qquad \kappa_k = \|D_k\|_{\mathrm{F}} > 0, \qquad \kappa = \min_{1 \leq k \leq K} \kappa_k, \qquad \Delta = \min_{0 \leq k \leq K} (\eta_{k+1} - \eta_k). \tag{2}$$

The null case $K = 0$ requires neither a jump size nor a direction. Temporal assumptions are stated separately in Section 3.

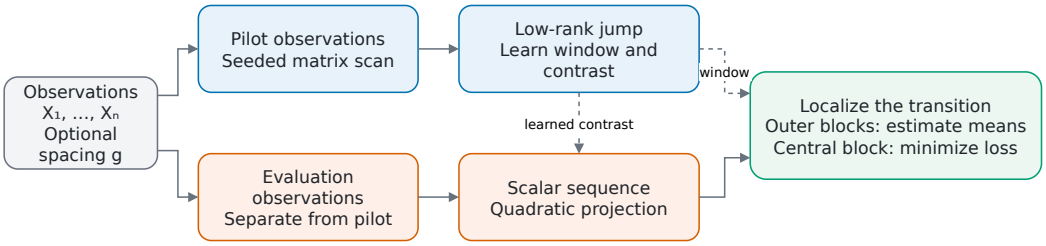

Figure 1: Learning and evaluation have distinct statistical roles. The pilot selects isolating windows and a low-rank quadratic contrast. Separate evaluation observations form a scalar sequence; outer blocks estimate its two means, and the central block determines the change location. Under geometric absolute regularity, deterministic spacing precedes the split.

The normalization gives a precise meaning to the signal. If P_2 denotes orthogonal projection onto polynomials of total degree at most two, then

$$\|M(f) - M(g)\|_{\mathrm{F}} = \|P_2(f-g)\|_{L^2(\mu)}\,. \tag{3}$$

Thus κ_k measures exactly the density change visible in the representation. For a symmetric Frobenius-unit matrix V, let $g_V(x) = \langle H(x), V\rangle_{\mathrm{F}}$. Its mean jump is $\langle D_k, V\rangle_{\mathrm{F}}$, so the ideal direction satisfies

$$V_k^* = \frac{D_k}{\kappa_k}, \qquad \langle D_k, V_k^*\rangle_{\mathrm{F}} = \kappa_k, \qquad \mathrm{Var}_{f_j}(g_V(X)) \le L. \tag{4}$$

The last bound holds for every segment and every Frobenius-unit V, because $\mathbb{E}_\mu g_V^2 = 1$. To see how the signal is collected, write $D_k = \sum_j \lambda_j u_j u_j^\top$. The contrast weights $u_j^\top H(x) u_j$ by λ_j/κ_k, so the component mean jumps contribute λ_j^2/κ_k and sum to κ_k. With these weights, components of either sign contribute positively to the scalar jump. The scalar noise bound remains independent of d for fixed L. Appendix A proves the representation identities. The pilot estimates V_k^*.

The representation contains order d^2 coefficients in a matrix of order $d+1$. Low rank acts on the change among those coefficients. To see the distinction, partition the coordinates into groups A, B of sizes a, b and consider

$$f_\theta(x) = 1 + \theta\ \mathrm{sign}\left(\sum_{i\in A} x_i\right)\mathrm{sign}\left(\sum_{j\in B} x_j\right), \qquad |\theta| < 1. \tag{5}$$

Both group marginals remain product uniform. Changing θ alters every cross-group interaction while preserving all coordinate means and variances. The jump has rank two, and its normalized contrast is

$$g_*(x) = \frac{3}{\sqrt{ab}}\left(\sum_{i\in A} x_i\right)\left(\sum_{j\in B} x_j\right).$$

This contrast aggregates a dense collection of weak interactions. Proposition G.1 gives the exact signal size and the sampling construction used in the experiments.

2.2 Estimate the direction on an isolating window

For $I = (s, e]$ of length $N = e - s$, form the matrix CUSUM

$$\widehat{A}_I(t) = \sqrt{\frac{(t-s)(e-t)}{N}}\left(\frac{\sum_{i=t+1}^{e} H(X_i)}{e-t} - \frac{\sum_{i=s+1}^{t} H(X_i)}{t-s}\right). \tag{6}$$

If I contains one change, its population CUSUM is a positive scalar multiple of that jump. We therefore search for a matrix direction that makes the observed contrast large. For a symmetric A, let $A_{(r)}$ retain the r eigenvalues of largest absolute value. The score has the variational interpretation

$$F_r(A) = \left\|A_{(r)}\right\|_{\mathrm{F}} = \max_{\substack{V = V^\top,\ \mathrm{rank}(V)\le r \\ \|V\|_{\mathrm{F}} \le 1}} \langle A, V\rangle_{\mathrm{F}}, \qquad \widehat{V}_I(t) = \frac{\widehat{A}_I(t)_{(r)}}{\left\|\widehat{A}_I(t)_{(r)}\right\|_{\mathrm{F}}}. \tag{7}$$

Rank truncation therefore learns the aggregate contrast directly. Its signed weights collect changes of either sign, while the working rank limits the number of estimated components. Lemma D.2 bounds the resulting signal loss in terms of matrix noise.

To isolate multiple changes, split independent observations into pilot $U_j = X_{2j-1}$ and evaluation $V_j = X_{2j}$, with $T = \lfloor n/2 \rfloor$. Every window choice uses the pilot. For $m \in 4\mathbb{N}$, take contained intervals of lengths $h = m2^\ell$ starting at multiples of $h/2$. Evaluate every integer split in each interval's central half. Among intervals whose maximum score exceeds τ, select the shortest, delete its interior from further selection, and repeat on the remaining components.

Short intervals give each learned direction a local target: the selected interval is proved to contain one change. If that interval is $I = (s, s+h]$, expand it to $J = (s-h, s+2h]$ and maximize the pilot score again over the central half of J. Normalize the corresponding rank-r matrix to obtain $\widehat{V}$. This expansion both centers the true change for direction learning and supplies outer blocks for scalar mean estimation.

2.3 Localize with separate scalar observations

On the evaluation array, project the observations in J to $Y_j = g_{\widehat{V}}(V_j)$. The outer blocks $(s-h, s]$ and $(s+h, s+2h]$ estimate the means $\widehat{\mu}_L, \widehat{\mu}_R$. The central block supplies the local objective

$$\widehat{b} = \operatorname*{arg\,min}_{t \in \{s, \ldots, s+h\}} \left\{ \sum_{j=s+1}^{t} (Y_j - \widehat{\mu}_L)^2 + \sum_{j=t+1}^{s+h} (Y_j - \widehat{\mu}_R)^2 \right\}. \tag{8}$$

Use the smallest minimizer and report $\widetilde{\eta} = \mathrm{clip}_{[0,n]}(2\widehat{b} + 1)$. The isolation event places one change inside J and makes both anchors pure. Conditioning on the pilot fixes J and $\widehat{V}$, leaving a scalar problem on independent evaluation observations. The anchors are also disjoint from the central block. These separations let the proof control learning, mean estimation, and location error in sequence.

The inputs are the working rank r, minimum scale m, and threshold τ. The theory gives sufficient choices in Section 3; the unknown-count experiment uses a separately calibrated null threshold. Geometrically absolutely regular data first undergo deterministic spacing, then use the same allocation and estimator on retained observations. Appendix G specifies the exact implementation and its computational cost.

3 From learning conditions to localization error

We first connect signal aggregation to localization under independence, then quantify the changes required by temporal dependence. Matrix noise controls isolation and direction learning; scalar noise controls the error after learning. All scan bounds use the same event. For a deterministic collection $\mathcal{G}$ of interval-split triples, let $E_I(t) = \widehat{A}_I(t) - \mathbb{E}\widehat{A}_I(t)$ and

$$\mathcal{E}_{\mathcal{R}}(\delta) = \left\{ \sup_{(s,t,e) \in \mathcal{G}} \|E_I(t)\|_{\mathrm{op}} \le \lambda_{\mathcal{R}}(\delta) \right\}. \tag{9}$$

The collection includes the seeded intervals and all admissible expanded windows. Random selection from this collection is covered by one event. Each candidate is central, each interval has length at least m, and its CUSUM weights satisfy $\sum_i w_i^2 = 1$ and $\max_i |w_i| \le w_{\mathcal{G}} \le \sqrt{3/m}$.

3.1 Noise bounds for learning and evaluation

Set $b_d = 6d$, $v_d = L(1 + d/2)$ and $\ell = \log(2p|\mathcal{G}|/\delta)$. The feature normalization gives $\|H(X_i) - \mathbb{E}H(X_i)\|_{\mathrm{op}} \le b_d$ and $\mathbb{E}(H(X_i) - \mathbb{E}H(X_i))^2 \preceq v_d I_p$. Matrix Bernstein (Tropp, 2012), applied to each weighted sum and both matrix tails, gives the following result.

Theorem 3.1 (Independent scan radius)**.** *For independent observations with the marginal model in* (2)*,* $\mathbb{P}(\mathcal{E}_{\mathrm{iid}}(\delta)) \ge 1 - \delta$ *for*

$$\lambda_{\mathrm{iid}}(\delta) = \sqrt{2 v_d \ell} + \frac{2}{3} b_d w_{\mathcal{G}} \ell. \tag{10}$$

Theorem B.1 gives the proof. The variance term grows as Ld, while the increment term decreases as the minimum window grows.

A second consequence of the representation governs the evaluation stage. For every fixed symmetric V with $\|V\|_{\mathrm{F}} = 1$, the centered scalar $\xi = g_V(X) - \mathbb{E}g_V(X)$ satisfies

$$\mathbb{E}e^{u\xi} \le \exp\left(\frac{\nu^2 u^2}{2(1 - B|u|)}\right), \quad |u| < 1/B, \qquad \nu^2 = CL, \quad B = C', \tag{11}$$

for universal constants C, C'. Proposition B.2 proves this statement by transferring quadratic moment bounds under product measure to $f \le L$, using the Hanson–Wright inequality (Rudelson & Vershynin, 2013). For fixed L, these scalar constants are independent of d. The matrix and scalar bounds govern different stages of the same estimator.

3.2 Detection and direction learning

Define

$$\mathrm{SNR}_{\mathrm{det},\mathcal{R}} = \frac{\Delta\kappa^2}{r\lambda_{\mathcal{R}}(\delta)^2}. \tag{12}$$

For symmetric matrices A, E, the rank score obeys $|F_r(A + E) - F_r(A)| \le \sqrt{r}\|E\|_{\mathrm{op}}$. This deterministic inequality is the interface between concentration and estimation.

Theorem 3.2 (Detection and initial localization)**.** *On $\mathcal{E}_{\mathcal{R}}(\delta)$, write $\varepsilon_r = \sqrt{r}\,\lambda_{\mathcal{R}}(\delta)$. If $K = 0$, a threshold $\tau > \varepsilon_r$ returns no changes. If $K \ge 1$, $m \le \Delta/8$, and*

$$\varepsilon_r < \tau < \frac{\sqrt{\Delta}\,\kappa}{8\sqrt{2}} - \varepsilon_r,$$

the seeded procedure returns exactly K intervals, each shorter than $\Delta/4$ and containing exactly one true change. A sufficient condition for this threshold interval to be nonempty is $\mathrm{SNR}_{\mathrm{det},\mathcal{R}} > 512$. On a one-change window of length N whose true split is an admissible central candidate, its score maximizer satisfies

$$|\widehat{b} - b| \le \frac{2\sqrt{Nr}\,\lambda_{\mathcal{R}}}{\alpha\kappa_k}, \qquad \alpha = \min\{(b - s)/N, (e - b)/N\}.$$

Theorem D.5 and Corollary D.6 prove the result using the population CUSUM margin and a discrete interval induction. Rank enters through $\sqrt{r}\lambda_{\mathcal{R}}$ rather than a Frobenius bound on the entire noise matrix. Under independence, (10) implies

$$r\lambda_{\mathrm{iid}}^2 \lesssim r\left(Ld\ell + \frac{d^2\ell^2}{m}\right). \tag{13}$$

The detection condition compares $\Delta\kappa^2$ with this learning cost. Section 3.4 supplies the corresponding radii under temporal dependence.

3.3 Independent scalar localization

For refinement, write $t_x = \nu^2/x^2 + B/x$. Let Δ_U be the minimum pilot segment length and set $\lambda = \lambda_{\mathrm{iid}}(\delta/2)$, computed for the pilot length T on its augmented grid. The next theorem combines direction learning, pure anchors, and scalar localization.

Theorem 3.3 (Independent scalar refinement)**.** *For independent observations with $K \ge 1$, use $\tau = 6\sqrt{r}\,\lambda$ on the pilot and suppose*

$$m \in 4\mathbb{N}, \quad m \le \Delta_U/8, \quad \frac{\Delta_U\kappa^2}{r\lambda^2} > 6272, \quad m \ge 512 t_\kappa \log(16K/\delta). \tag{14}$$

With probability at least $1 - \delta$, the procedure returns exactly K changes, each learned direction retains a projected jump at least $\kappa_k/2$, and the reported locations satisfy

$$|\widetilde{\eta}_k - \eta_k| \le 2\left\lceil 64 t_{\kappa_k} \log(8K/\delta)\right\rceil + 1. \tag{15}$$

Lemma E.1 supplies the pure anchors and retained signal. Lemma E.2 controls anchor estimation and all excursions of the local objective. Theorem E.3 combines these events, including pilot failure. When $\kappa_k \le \nu^2/B$, the two scalar terms are comparable to ν^2/κ_k^2. For fixed L, the bound therefore has a dimension-free κ_k^{-2} form once (14) holds. A rank-two bounded-density subclass gives an independent expected-error lower bound of order κ^{-2} for small jumps (Theorem E.6); this comparison concerns the jump-strength dependence.

3.4 Temporal dependence and time resolution

Temporal dependence first changes the matrix radius used in Theorem 3.2. Each route below gives $\mathbb{P}(\mathcal{E}_{\mathcal{R}}(\delta)) \ge 1-\delta$ under assumptions uniform across the full sequence, including changes. Geometric absolute regularity also supports the separate coupling step used for scalar refinement.

Under geometric absolute regularity, suppose the centered matrix array is β-mixing with $\beta(h) \le e^{-c_\beta(h-1)}$. The dependent matrix Bernstein inequality of Banna et al. (2016) uses a variance proxy over arbitrary subsets. We bound that proxy explicitly by

$$V_\beta = v_d + 2\sum_{h\ge 1} \min\{v_d,\, 2b_d^2 e^{-c_\beta(h-1)}\} \le v_d\left(1 + 2J_\beta + \frac{2}{1-e^{-c_\beta}}\right), \tag{16}$$

where $J_\beta = \lceil c_\beta^{-1}\log_+(2b_d^2/v_d)\rceil$. Let $\Gamma(c,n) = \frac{\log n}{\log 2}\max\{2, \frac{32\log n}{c\log 2}\}$, $A_\beta = 3V_\beta + 3b_d^2/(c_\beta m)$, and $B_\beta = \sqrt{3}\, b_d\Gamma(c_\beta, n)/\sqrt{m}$. For the universal constant $c_{\mathrm{BMY}} > 0$ in that inequality,

$$\lambda_\beta = \sqrt{A_\beta \ell / c_{\mathrm{BMY}}} + B_\beta \ell / c_{\mathrm{BMY}}. \tag{17}$$

Lemma C.1 and Theorem C.2 verify the weighted subset proxy and both tails. The covariance sum in (16) records how persistence inflates matrix noise.

For the Markov route, let S_i be a stationary latent Markov chain with $\chi = \|P - \Pi\|_{L^2\to L^2} < 1$, and let $X_i = F_k(S_i)$ on segment k. The observation maps can change while the chain remains stationary. Put $\alpha_\chi = (1+\chi)/(1-\chi)$, $b_\chi = 8/[\pi(1-\chi)]$, $c_{\mathrm{M}} = 32/\pi^2$, and $\ell_{\mathrm{M}} = \log(2p^{2-\pi/4}|\mathcal{G}|/\delta)$. Then

$$\lambda_{\mathrm{M}} = \sqrt{c_{\mathrm{M}}\alpha_\chi v_d \ell_{\mathrm{M}}} + c_{\mathrm{M}} b_\chi b_d w_{\mathcal{G}} \ell_{\mathrm{M}}. \tag{18}$$

Theorem C.3 maps the time-varying weighted observations to the matrix inequality of Neeman et al. (2024). A smaller absolute gap increases both variance and increment terms, making the required learning sample explicit.

For physical dependence, suppose $Z_i = H(X_i) - \mathbb{E}H(X_i) = G_i(\ldots, \varepsilon_{i-1}, \varepsilon_i)$ for independent innovations. Replacing only ε_{i-j} by an independent copy gives $\|Z_i - Z_i^{(i-j)}\|_{L^\infty(\mathrm{op})} \le A_d\rho^j$, with $\rho < 1$, uniformly in i. This uses the causal perturbation framework of Wu (2005) with an explicit essential-supremum condition. Conditioning on the last $q+1$ innovations gives a finite-memory approximation with remainder $R_q = A_d\rho^{q+1}/(1-\rho)$. Let $a = q+1$ and $\ell_q = \log(2pa|\mathcal{G}|/\delta)$. Then

$$\lambda_{\mathrm{F}} = \sqrt{2av_d\ell_q} + \frac{2}{3}ab_d w_{\mathcal{G}}\ell_q + \sqrt{n}R_q. \tag{19}$$

The first two terms follow from independent residue classes; the last bounds truncation deterministically. Choosing $q = \lceil \log_+(A_d\sqrt{n}/((1-\rho)\sqrt{v_d}))/|\log\rho|\rceil$ makes $\sqrt{n}R_q \le \sqrt{v_d}$; take $q = 0$ if $A_d = 0$. Lemma C.4 and Theorem C.5 supply the construction and proof.

For scalar refinement under geometric absolute regularity, retain observations at indices $1, 1+g, 1+2g, \ldots$, with $g = 1 + \lceil c_\beta^{-1}\log(4n/\delta)\rceil$. Split this retained sequence into alternating pilot and evaluation arrays. Coupling this entire retained sequence to independent copies has disagreement probability at most $\delta/4$. The retained pilot uses the independent radius $\lambda_{\mathrm{iid}}(\delta/4)$; the scalar stage has budget $\delta/2$. Under the corresponding pilot and anchor conditions, Theorem E.5 gives

$$|\widetilde{\eta}_k - \eta_k| \le 2g\,\lceil 64 t_{\kappa_k}\log(8K/\delta)\rceil + g. \tag{20}$$

The pilot contains $\lfloor N_g/2\rfloor$ observations and has segment spacing at least $\lfloor \Delta/(2g)\rfloor - 1$. Dependence therefore affects both learning feasibility and time resolution. The Markov and physical-dependence results provide scan and isolation guarantees under their own assumptions; the coupled refinement theorem additionally applies when geometric absolute regularity holds.

4 EXPERIMENTS

The experiments test the proposed reduction through three questions: when does a low-rank contrast collect dense interaction changes, how does its retained signal affect localization, and how does temporal spacing change the error? We then evaluate unknown-count detection and a controlled transition between activity-derived distributions.

Synthetic observations follow (5), with equal coordinate groups and opposite segment parameters chosen to fix κ. All densities satisfy $f_\theta \leq 1.8$. Single-change locations are uniform between $0.4n$ and $0.6n$. Methods share the pilot observations; scalar estimators also use separate evaluation observations. Single-change experiments use the full pilot window and outer-quarter anchors, as justified by Corollary G.2. The multiple-change study uses the complete seeded procedure. Independent single-change conditions have 40 trials, and dependence conditions have 32.

4.1 COLLECTING DENSE CHANGES THROUGH A LEARNED DIRECTION

We compare the rank-two scan and scalar estimator with full Frobenius, largest-entry, mean, and Gaussian-kernel feature scans. The kernel comparator uses 512 random Fourier features (Rahimi & Recht, 2007) and a pilot median-distance bandwidth. The population contrast fixes the true direction. Rank-one and full-direction scalar estimators share the data allocation and scalar objective; each working rank selects its own pilot score maximum.

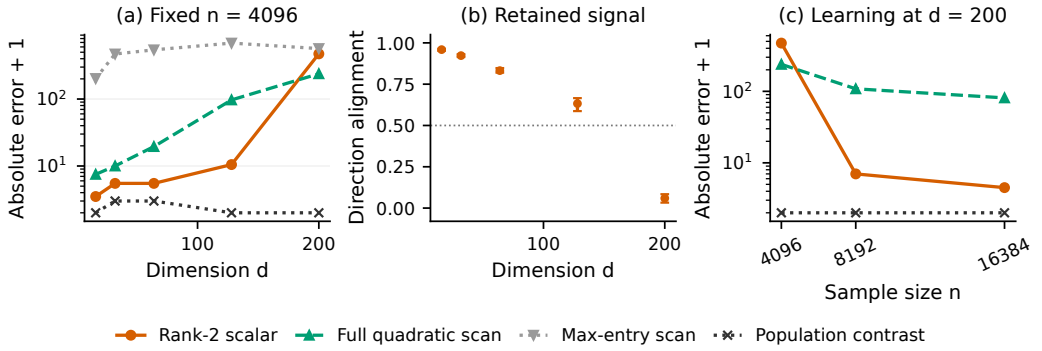

Figure 2: Learning a dense interaction change with $\kappa = 0.9$. (a) Localization at $n = 4096$. (b) Retained signal, measured by $\langle \widehat{V}, V^* \rangle_{\mathrm{F}} = \langle D, \widehat{V} \rangle_{\mathrm{F}} / \kappa$, with medians and interquartile ranges. The 0.5 line marks retention of half the jump. (c) Increasing the sample size restores learning at $d = 200$. Error panels show medians plus one on a logarithmic scale. Each condition has 40 trials.

At $d = 128$ and $n = 4096$, the rank-two scalar estimator has median absolute error 9.5, compared with 96 for the full quadratic scan and 683.5 for the largest-entry scan. Figure 2(b) measures the signal delivered to scalar localization by the learned direction. At $d = 200$, increasing n from 4096 to 8192 and 16384 raises mean signal retention from 0.06 to 0.73 and 0.87. The corresponding scalar median errors fall from 472 to 6 and 3.5.

Table 1 compares working ranks within the same scalar procedure. Rank two gives the smallest median among learned contrasts at $d \leq 128$ and in both larger-sample settings. Rank one captures part of the rank-two jump; retaining every empirical component includes more estimation noise. At $d = 200$, $n = 4096$, the population contrast still has median error 1, locating the gap in direction learning. The recovery with more observations connects the gain from aggregation to the sample requirement for learning it.

Table 1: Direction rank and localization. Each entry is median absolute error over 40 trials. The three learned estimators share the pilot, anchors, and scalar objective. The population contrast uses the true jump direction. Bold marks the smallest median among learned contrasts. Figure 5 shows dispersion in the dimension study.

d	n	Rank one	Rank two	Full direction	Population
16	4096	7	**2.5**	3	1
32	4096	9.5	**4.5**	8.5	2
64	4096	24.5	**4.5**	25.5	2
128	4096	41	**9.5**	140.5	1
200	4096	618.5	472	**221**	1
200	8192	12	**6**	103	1
200	16384	9.5	**3.5**	56	1

4.2 From retained signal to scalar localization

At $d = 32$ and $n = 8192$, increasing κ from 0.35 to 0.9 reduces the rank-two scalar estimator's median error from 62.5 to 1.5 (Figure 3). The population contrast decreases from 21 to 1. The learned and population contrasts are furthest apart at the weaker signals, where pilot error is more consequential. At $\kappa = 0.9$, scalar refinement also improves the rank scan's median error from 4 to 1.5.

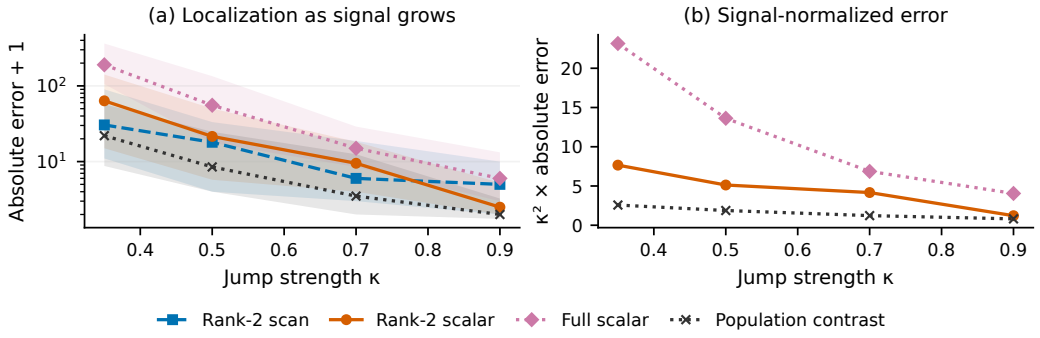

Figure 3: Signal strength and scalar error at $d = 32$, $n = 8192$. (a) Median error with interquartile bands over 40 trials. (b) Multiplication by κ^2 shows the signal scale alongside the population-contrast reference. The latter isolates scalar localization from the cost of learning a direction.

The rescaling in Figure 3(b) relates the experiment to the small-jump term ν^2/κ^2. The learned contrast carries an additional dependence on pilot accuracy, whereas the population contrast holds the direction fixed. This distinction explains why the theorem requires a learning condition before its scalar rate applies. The rank scan remains competitive across the signal settings; scalar evaluation is most accurate here when the pilot has already identified a useful direction.

4.3 Temporal dependence and the resolution of retained data

A stationary latent refresh chain retains its state with probability $1 - a$ and redraws it with probability a. Its absolute spectral gap is a, and $\beta(h) \le (1 - a)^h$. Segment-specific observation maps produce the same marginal density family. We use $n = 131072$, $d = 16$, and $\kappa = 0.9$, with deterministic spacing prescribed by the coupling theorem.

For refresh probabilities $a = 0.8, 0.5, 0.25$, the spacings are $12, 25, 58$, and the retained pilot sizes are $5461, 2621, 1130$. The scalar median errors are $27, 29, 69.5$ in original time units. Dividing by spacing gives $2.25, 1.16, 1.20$. The two panels distinguish precision on retained observations from resolution on the original timeline, as in (20). The smaller pilot samples also expose the second cost of spacing: fewer observations remain for learning the direction.

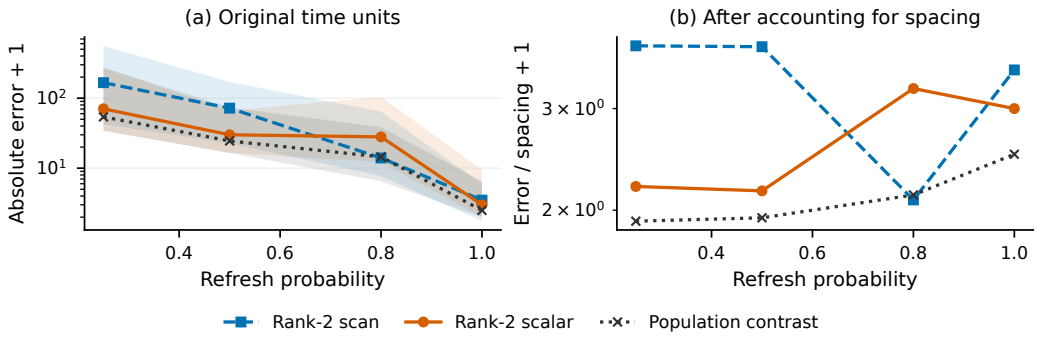

Figure 4: The cost of temporal spacing. (a) Median errors in original time units, with interquartile bands over 32 trials. (b) The same medians divided by spacing, then shifted by one for the logarithmic scale. The independent condition uses spacing one.

4.4 Unknown change counts and activity-derived transitions

For unknown-count estimation, we evaluate every seeded candidate at $n = 16384$, $d = 16$, and $m = 128$. An independent set of 80 null sequences calibrates the threshold. Held-out null trials return no change in $57/60$ cases. With three dense changes, the method returns the correct count in $57/60$ cases, and scalar refinement reduces median Hausdorff error from 12 to 8.5 across all 60 trials. This experiment checks the complete chain from interval selection to localization with an empirically calibrated threshold; the theoretical radius supplies a sufficient high-probability threshold.

The final study uses public UCI human activity recognition features (Reyes-Ortiz et al., 2013). Training data select the coordinates. We resample walking and walking-downstairs observations from the test pools, then apply independent global sign flips. The flips remove mean signals while preserving quadratic features, creating a controlled distribution transition at $d = 128$ and $n = 1600$. Across 40 trials, rank-ten scalar refinement has median error 1, compared with 291.5 for the mean scan and 2 for the kernel-feature scan. The full quadratic scan also has median error 1. These activity-derived distributions provide a second setting in which the quadratic representation collects signal after coordinate means have been removed. Appendix G gives all trial summaries, preprocessing, calibration, and computational details.

5 Discussion and conclusion

The quadratic contrast aggregates changes across spectral components into one scalar mean difference. The normalization identifies the Frobenius jump with the L^2 change in the degree-two density projection, and a density bound controls the noise of every normalized quadratic contrast. Under the pilot and anchor conditions, the learned contrast retains at least half this jump and supports precise localization using separate evaluation observations.

The working rank controls how much signal and estimation noise enter that contrast. For a population CUSUM A and perturbation E, Proposition D.4 gives

$$\left\|(A+E)_{(r)} - A\right\|_{\mathrm{F}} \le \left\|A - A_{(r)}\right\|_{\mathrm{F}} + 2\sqrt{2r}\,\|E\|_{\mathrm{op}}\,.$$

The first term measures discarded signal, while the second measures the cost of estimating more directions. This decomposition extends the representation analysis to approximately low-rank changes and explains the rank comparison in Table 1. The complete localization theorems use exact jump rank at most r.

Temporal dependence changes the cost of learning and the time resolution of the final estimate. Three explicit matrix radii support detection and isolation. Under geometric absolute regularity, spacing and coupling also give a complete scalar refinement guarantee. The resulting analysis specifies when changes spread across many interactions can be aggregated reliably and localized with precision governed by their retained signal.

STATEMENT OF AI USE

AI tools assisted with wording and proofreading of the manuscript.

REFERENCES

Marwa Banna, Florence Merlevède, and Pierre Youssef. Bernstein-type inequality for a class of dependent random matrices. *Random Matrices: Theory and Applications*, 05(02):1650006, 2016. doi: 10.1142/s2010326316500064.

Rafal Baranowski, Yining Chen, and Piotr Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):649–672, 2019. doi: 10.1111/rssb.12322.

Henri Berbee. *Random walks with stationary increments and renewal theory*. MC Tracts. Centrum Voor Wiskunde en Informatica, 1979. URL `https://ir.cwi.nl/pub/12694`.

S Kovács, P Bühlmann, H Li, and A Munk. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023. doi: 10.1093/biomet/asac052.

Carlos Misael Madrid Padilla, Haotian Xu, Daren Wang, Oscar Hernan Madrid Padilla, and Yi Yu. Change point detection and inference in multivariate non-parametric models under mixing conditions. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 21081–21134. Curran Associates, Inc., 2023. doi: 10.52202/075280-0922.

Joe Neeman, Bobby Shi, and Rachel Ward. Concentration inequalities for sums of Markov-dependent random matrices. *Information and Inference: A Journal of the IMA*, 13(4), 2024. doi: 10.1093/imaiai/iaae032.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis (eds.), *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

Jorge Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto, and Xavier Parra. Human activity recognition using smartphones. UCI Machine Learning Repository, 2013. doi: 10.24432/C54S4K.

Mark Rudelson and Roman Vershynin. Hanson–Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18, 2013. doi: 10.1214/ecp.v18-2865.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Optimal covariance change point localization in high dimensions. *Bernoulli*, 27(1):554–575, 2021. doi: 10.3150/20-bej1249.

Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1):57–83, 2018. doi: 10.1111/rssb.12243.

Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005. doi: 10.1073/pnas.0506715102.

A REPRESENTATION AND THE DETECTION PROBLEM

The reduction has three stages. First, a fixed degree-two feature map replaces density estimation by matrix mean estimation. Second, rank truncation retains a small number of changing spectral directions. Third, a learned matrix direction maps each observation to a scalar for local estimation. The intermediate matrix has order $d+1$, but it contains order d^2 coefficients. The statistical compression comes from the retained rank and the final scalar projection.

Let μ be the product uniform probability measure on $[-1,1]^d$, where $d \geq 1$. All densities below are relative to μ. Put

$$p_2(x) = \frac{3x^2-1}{2}, \qquad p = d+1.$$

For $x \in [-1,1]^d$, define $H(x) \in \mathbb{S}^p$ by

$$\begin{aligned} &H_{00}(x) = 1, \qquad H_{0j}(x) = H_{j0}(x) = \sqrt{\frac{3}{2}}x_j \quad (1 \leq j \leq d), \\ &H_{jj}(x) = \sqrt{5}\,p_2(x_j), \quad H_{ij}(x) = H_{ji}(x) = \frac{3}{\sqrt{2}}x_i x_j \quad (1 \leq i < j \leq d). \end{aligned} \tag{21}$$

For a density f, write $M(f) = \mathbb{E}_f H(X)$.

Proposition A.1 (Representation and moment bounds)**.** *Let $f, g \in L^2(\mu)$ be densities and let P_2 denote orthogonal projection onto polynomials of total degree at most two. Then*

$$P_2 f(x) = 1 + \sum_j a_j x_j + \sum_j b_j p_2(x_j) + \sum_{i<j} \gamma_{ij} x_i x_j, \tag{22}$$

$$a_j = 3\mathbb{E}_f X_j, \qquad b_j = 5\mathbb{E}_f p_2(X_j), \qquad \gamma_{ij} = 9\mathbb{E}_f(X_i X_j),$$

$$\|M(f) - M(g)\|_{\mathrm{F}} = \|P_2(f-g)\|_{L^2(\mu)}. \tag{23}$$

Moreover,

$$\|H(x)\|_{\mathrm{op}} \leq 3d, \qquad \mathbb{E}_\mu H(X)^2 = (1+d/2)I_p, \qquad \mathbb{E}_\mu \langle H(X), V\rangle_{\mathrm{F}}^2 = \|V\|_{\mathrm{F}}^2 \quad (V \in \mathbb{S}^p). \tag{24}$$

If $f \leq L$, $Z = H(X) - M(f)$ satisfies

$$\|Z\|_{\mathrm{op}} \leq b_d := 6d, \qquad \mathbb{E}_f Z^2 \preceq v_d I_p, \qquad v_d := L(1+d/2). \tag{25}$$

Proof. Under the univariate uniform probability measure, $1, x, p_2(x)$ are orthogonal and have squared norms $1, 1/3, 1/5$. The product structure gives the basis in (22); the squared norm of $x_i x_j$ is $1/9$. Dividing each inner product with f by the corresponding squared norm gives its coefficient. Consequently,

$$\|P_2(f-g)\|_{L^2}^2 = \sum_j \frac{(a_j(f)-a_j(g))^2}{3} + \sum_j \frac{(b_j(f)-b_j(g))^2}{5} + \sum_{i<j} \frac{(\gamma_{ij}(f)-\gamma_{ij}(g))^2}{9}.$$

The corresponding matrix entries are $a_j/\sqrt{6}$, $b_j/\sqrt{5}$, and $\gamma_{ij}/(3\sqrt{2})$. Counting each off-diagonal entry twice gives (23). The same calculation for

$$\langle H(x), V\rangle_{\mathrm{F}} = V_{00} + \sqrt{6}\sum_j V_{0j}x_j + \sqrt{5}\sum_j V_{jj}p_2(x_j) + 3\sqrt{2}\sum_{i<j} V_{ij}x_i x_j$$

proves the last identity in (24).

Since $|x_j|, |p_2(x_j)| \leq 1$,

$$\|H(x)\|_{\mathrm{F}}^2 \leq 1 + 8d + \frac{9}{2}d(d-1) \leq 9d^2.$$

For the second inequality, subtract the left side from $9d^2$; the difference is $(d-1)(9d/2+1) \geq 0$. The $(0,0)$ entry of $\mathbb{E}_\mu H^2$ is $1+d/2$. Each remaining diagonal entry is $1/2+1+(d-1)/2 = 1+d/2$. Every off-diagonal entry integrates to zero because each of its summands is odd in at least one coordinate. This proves the first two identities in (24). Jensen's inequality gives $\|M(f)\|_{\mathrm{op}} \leq 3d$. Finally, $H(x)^2$ is positive semidefinite, so

$$\mathbb{E}_f Z^2 = \mathbb{E}_f H^2 - M(f)^2 \preceq \mathbb{E}_f H^2 \preceq L\mathbb{E}_\mu H^2 = v_d I_p.$$

The triangle inequality gives the centered envelope. □

Assumption A.2 (Marginal model)**.** There are integers $0 = \eta_0 < \eta_1 < \cdots < \eta_K < \eta_{K+1} = n$. For $\eta_k < i \le \eta_{k+1}$, X_i has density f_k with $0 \le f_k \le L$, where $L \ge 1$. Set

$$M_k = M(f_k), \quad D_k = M_k - M_{k-1}, \quad \kappa_k = \|D_k\|_{\mathrm{F}}, \quad \Delta = \min_{0 \le k \le K} (\eta_{k+1} - \eta_k).$$

For $K \ge 1$, assume $\kappa := \min_{1 \le k \le K} \kappa_k > 0$ and $\operatorname{rank}(D_k) \le r$, where $1 \le r \le p$ is fixed.

The case $K = 0$ is included in null guarantees and does not use κ. Independence across time is imposed only in Section B; each dependence route in Section C replaces it. Within a segment the marginals are identical, while the full sequence has different marginals when $K > 0$. The bounds retain their dependence on L explicitly.

Only changes visible to P_2 are targets. For example, let $p_3(x) = (5x^3 - 3x)/2$ and $f_\theta(x) = 1 + \theta p_3(x_1)$, with $0 < |\theta| \le 1/2$. This is a density bounded between $1/2$ and $3/2$, but $M(f_\theta) = M(1)$ by orthogonality. The proposed matrix-mean scan has zero population signal for this change. This example specifies the target of the matrix-mean statistic. The first row of $H(X)$ also preserves X, so the feature map itself retains the full observation.

A.1 Matrix CUSUM and deterministic scan sets

For integers $0 \le s < t < e \le n$, define

$$\widehat{A}_I(t) = \sqrt{\frac{(t-s)(e-t)}{e-s}} \left\{ \frac{\sum_{i=t+1}^{e} H(X_i)}{e-t} - \frac{\sum_{i=s+1}^{t} H(X_i)}{t-s} \right\}, \qquad I = (s, e], \tag{26}$$

$$A_I(t) = \mathbb{E}\widehat{A}_I(t), \qquad E_I(t) = \widehat{A}_I(t) - A_I(t). \tag{27}$$

Write $Z_i = H(X_i) - \mathbb{E}H(X_i)$ and $N = e - s$. Then

$$E_I(t) = \sum_{i=s+1}^{e} w_i Z_i, \quad w_i = \begin{cases} -\sqrt{(e-t)/(N(t-s))}, & i \le t, \\ \sqrt{(t-s)/(N(e-t))}, & i > t, \end{cases} \tag{28}$$

and direct summation gives

$$\sum_i w_i^2 = 1, \quad w_* := \max_i |w_i| = \sqrt{\frac{\max(t-s, e-t)}{N \min(t-s, e-t)}}. \tag{29}$$

For a candidate in the central half of an interval, $N w_*^2 \le 3$. On any deterministic finite collection $\mathcal{G}$ of such triples with $N \ge m \ge 4$, define

$$w_{\mathcal{G}} = \max_{\mathcal{G}} w_* \le \sqrt{3/m}, \qquad \ell(\epsilon) = \log \frac{2p|\mathcal{G}|}{\epsilon}, \quad 0 < \epsilon < 1.$$

Here and below the collection is nonempty. If no admissible interval exists, the algorithm returns the empty set and no nontrivial recovery claim is made. All candidates are integer-valued. A data-dependent selection from $\mathcal{G}$ is covered by its uniform event.

For $A \in \mathbb{S}^p$, let $A_{(r)}$ be a best rank-r approximation retaining the r eigenvalues of largest absolute value. Define

$$F_r(A) = \left\| A_{(r)} \right\|_{\mathrm{F}} = \left(\sum_{j=1}^{r} \sigma_j(A)^2 \right)^{1/2}, \qquad \widehat{W}_I(t) = F_r(\widehat{A}_I(t)).$$

Tie rules are fixed and measurable throughout. Ties do not affect the score or the bounds below.

B Independent observations

Theorem B.1 (Independent uniform matrix bound)**.** *Under Assumption A.2, suppose $X_1, \ldots, X_n$ are mutually independent. For any deterministic scan set $\mathcal{G}$ described above, put*

$$\lambda_{\mathrm{iid}}(\epsilon) = \sqrt{2 v_d \ell(\epsilon)} + \frac{2}{3} b_d w_{\mathcal{G}} \ell(\epsilon). \tag{30}$$

Then

$$\mathbb{P}\Big\{\max_{(s,t,e)\in\mathcal{G}} \|E_I(t)\|_{\mathrm{op}} \le \lambda_{\mathrm{iid}}(\epsilon)\Big\} \ge 1-\epsilon. \tag{31}$$

In particular, replacing $w_{\mathcal{G}}$ by $2/\sqrt{m}$ gives

$$\lambda_{\mathrm{iid}}(\epsilon) \le \sqrt{2L(1+d/2)\ell(\epsilon)} + \frac{8d\,\ell(\epsilon)}{\sqrt{m}}.$$

Proof. For a fixed triple, the real symmetric matrices $w_i Z_i$ are independent, centered, and bounded in operator norm by $R = b_d w_*$. Moreover,

$$\Big\|\sum_i \mathbb{E}(w_i Z_i)^2\Big\|_{\mathrm{op}} \le v_d \sum_i w_i^2 = v_d.$$

The bounded matrix Bernstein inequality (Tropp, 2012, Theorem 6.1(ii)), applied to both signs, gives

$$\mathbb{P}\{\|E_I(t)\|_{\mathrm{op}} > x\} \le 2p\exp\left(-\frac{x^2}{2(v_d + Rx/3)}\right).$$

For $x = \sqrt{2v_d u} + 2Ru/3$, one has $x^2 \ge 2u(v_d + Rx/3)$ by expansion. Set $u = \ell(\epsilon)$, replace w_* by $w_{\mathcal{G}}$, and sum the tail probabilities over $\mathcal{G}$. No identical-distribution assumption across different segments is used. □

Proposition B.2 (A scalar condition supplied by bounded density)**.** *There is a numerical constant $C_0 \ge 1$, independent of d, L, f, V, such that for $f \le L$, symmetric $\|V\|_{\mathrm{F}} = 1$, and $\xi = \langle H(X) - M(f), V\rangle_{\mathrm{F}}$,*

$$\log \mathbb{E}_f e^{u\xi} \le \frac{u^2\nu^2}{2(1-B|u|)}, \qquad |u| < 1/B, \quad B = 2C_0, \quad \nu^2 = 2LB^2. \tag{32}$$

Also $\mathrm{Var}_f(\langle H(X), V\rangle_{\mathrm{F}}) \le L$. One may instead use any smaller uniform pair (ν^2, B) for which (32) *has been established.*

Proof. The variance bound follows from $\mathbb{E}_f\langle H(X), V\rangle_{\mathrm{F}}^2 \le L\mathbb{E}_\mu\langle H(X), V\rangle_{\mathrm{F}}^2 = L$. To obtain a moment bound, write

$$q(x) = \langle H(x), V\rangle_{\mathrm{F}} - \mathbb{E}_\mu\langle H(X), V\rangle_{\mathrm{F}} = b^\top x + x^\top Q x - \mathbb{E}_\mu(X^\top Q X),$$

where

$$b_j = \sqrt{6}V_{0j}, \quad Q_{jj} = \frac{3\sqrt{5}}{2}V_{jj}, \quad Q_{ij} = \frac{3}{\sqrt{2}}V_{ij} \quad (i \ne j).$$

The Frobenius normalization gives $\|b\|_2 \le \sqrt{3}$ and $\|Q\|_{\mathrm{op}} \le \|Q\|_{\mathrm{F}} \le 3\sqrt{5}/2$. Under μ the coordinates are independent, centered, and bounded. The scalar bounded-sum inequality for $b^\top X$, together with the Hanson–Wright inequality (Rudelson & Vershynin, 2013, Theorem 1.1) for Q, therefore gives

$$\mathbb{P}_\mu\{|q(X)| > t\} \le C\exp[-c\min(t^2, t)] \quad (t \ge 0)$$

for numerical $c, C > 0$. Both matrix norms and all sub-Gaussian coordinate norms used here have numerical bounds. For an integer $a \ge 2$, integrate $a\int_0^\infty t^{a-1}\mathbb{P}(|q| > t)\,dt$. The integral over $(0,1)$ is bounded by a constant, and the integral over $(1,\infty)$ is bounded by a constant times $a!c^{-a}$. Increasing one numerical C_0 gives $\mathbb{E}_\mu|q|^a \le a!C_0^a$.

Now $\xi = q - \mathbb{E}_f q$. Convexity and Jensen's inequality give

$$\mathbb{E}_f|\xi|^a \le 2^a\mathbb{E}_f|q|^a \le La!(2C_0)^a = \frac{a!}{2}\nu^2 B^{a-2}.$$

The absolute moment series converges for $|u| < B^{-1}$. Expanding the exponential, using $\mathbb{E}_f\xi = 0$, and bounding each higher moment in absolute value yields $\mathbb{E}_f e^{u\xi} \le 1 + u^2\nu^2/[2(1-B|u|)]$. Use $\log(1+x) \le x$ to conclude. Independence of coordinates was used only under the reference measure μ, not under the actual density f. □

The proposition supplies a marginal scalar bound. Section E uses independent evaluation observations, or a coupling to such observations, to apply this bound after conditioning on the learned direction.

C TEMPORAL DEPENDENCE

The population CUSUM uses only marginal means. Temporal dependence enters through the centered weighted sum (28). The following three routes provide alternative assumptions for controlling this weighted sum.

C.1 GEOMETRIC ABSOLUTE REGULARITY AND THE BMY VARIANCE PROXY

For sigma fields $\mathcal{A}, \mathcal{B}$, let $\beta(\mathcal{A}, \mathcal{B})$ be one half the supremum of $\sum_{a,b} |\mathbb{P}(A_a \cap B_b) - \mathbb{P}(A_a)\mathbb{P}(B_b)|$ over finite measurable partitions from the two sigma fields. For the centered feature sequence, assume

$$\beta_Z(h) := \sup_n \sup_{1 \le j \le n-h} \beta\big(\sigma(Z_i : 1 \le i \le j), \sigma(Z_i : j+h \le i \le n)\big) \le e^{-c_\beta(h-1)}, \qquad h \ge 1, \tag{33}$$

with $c_\beta > 0$. This is a uniform condition across change boundaries. The BMY envelope in (33) applies to the centered array even when its marginal laws vary across segments.

Lemma C.1 (Explicit dependent variance bound)**.** *Set*

$$c_h = \min\{v_d,\, 2b_d^2 e^{-c_\beta(h-1)}\}, \tag{34}$$

$$V_\beta = v_d + 2\sum_{h=1}^{\infty} c_h, \tag{35}$$

$$J_\beta = \left\lceil \frac{\log_+(2b_d^2/v_d)}{c_\beta} \right\rceil, \qquad \log_+(x) = \max\{\log x, 0\}. \tag{36}$$

Then the convergent series satisfies

$$V_\beta \le v_d \left(1 + 2J_\beta + \frac{2}{1 - e^{-c_\beta}}\right). \tag{37}$$

For a fixed triple, put $A_j = w_{s+j} Z_{s+j}$, $1 \le j \le N$. The variance quantity required by BMY obeys

$$v_A^2 := \sup_{\varnothing \ne J \subseteq \{1,\dots,N\}} \frac{1}{|J|} \lambda_{\max} \mathbb{E}\left(\sum_{j \in J} A_j\right)^2 \le w_*^2 V_\beta. \tag{38}$$

Proof. For unit vectors u, v, matrix Cauchy–Schwarz in its scalar form gives

$$|u^\top \mathbb{E}(Z_i Z_j) v| = |\mathbb{E}\langle Z_i u, Z_j v\rangle| \le (\mathbb{E}\,\|Z_i u\|_2^2\, \mathbb{E}\,\|Z_j v\|_2^2)^{1/2} \le v_d.$$

Berbee's coupling, in the form recorded in Banna et al. (2016, Lemma 6), supplies Z_j^* with the same marginal law as Z_j, independent of Z_i, and $\mathbb{P}(Z_j^* \ne Z_j) \le \beta_Z(j-i)$. Since both matrices are centered,

$$\|\mathbb{E} Z_i Z_j\|_{\mathrm{op}} = \big\|\mathbb{E} Z_i (Z_j - Z_j^*)\big\|_{\mathrm{op}} \le 2b_d^2 \beta_Z(j-i).$$

Thus $\|\mathbb{E} Z_i Z_j\|_{\mathrm{op}} \le c_{j-i}$; the transposed product has the same bound. Expanding the square in (38), bounding $|w_i w_j| \le w_*^2$, and observing that at each positive lag there are at most $|J|$ pairs gives

$$\left\| \mathbb{E}\Big(\sum_{j \in J} A_j\Big)^2 \right\|_{\mathrm{op}} \le |J| w_*^2 \left(v_d + 2\sum_{h \ge 1} c_h\right).$$

This handles arbitrary subsets J, as required by the original theorem, not only contiguous intervals. For (37), the first J_β terms of the series are at most v_d each. The remaining geometric tail is at most $2b_d^2 e^{-c_\beta J_\beta}/(1 - e^{-c_\beta}) \le v_d/(1 - e^{-c_\beta})$. □

Theorem C.2 (BMY uniform CUSUM bound)**.** *Under Assumption A.2 and* (33)*, let $c_{\mathrm{BMY}} > 0$ be the universal exponent constant in Banna et al. (2016, Theorem 1). Define*

$$\Gamma(c, n) = \frac{\log n}{\log 2} \max\left\{2, \frac{32 \log n}{c \log 2}\right\}, \qquad A_\beta = 3V_\beta + \frac{3b_d^2}{c_\beta m}, \qquad B_\beta = \frac{\sqrt{3} b_d}{\sqrt{m}} \Gamma(c_\beta, n).$$

Then (31) *holds with*

$$\lambda_\beta(\epsilon) = \sqrt{\frac{A_\beta \ell(\epsilon)}{c_{\mathrm{BMY}}}} + \frac{B_\beta \ell(\epsilon)}{c_{\mathrm{BMY}}}. \tag{39}$$

Proof. For each fixed triple the matrices $A_j = w_{s+j} Z_{s+j}$ are centered and self-adjoint, and $\lambda_{\max}(\pm A_j) \le b_d w_*$. Deterministic multiplication cannot increase absolute regularity coefficients, so they satisfy the uniform bound required by BMY. Its two-sided consequence is

$$\mathbb{P}\{\|E_I(t)\|_{\mathrm{op}} > x\} \le 2p \exp\left\{ -\frac{c_{\mathrm{BMY}} x^2}{N v_A^2 + c_\beta^{-1} b_d^2 w_*^2 + x b_d w_* \Gamma(c_\beta, N)} \right\}.$$

The subset variance in this display is precisely (38). Use $N w_*^2 \le 3$, $w_*^2 \le 3/m$, and monotonicity of $\Gamma(c, N)$ for $N \ge 4$. The denominator is at most $A_\beta + B_\beta x$. For positive A, B, c, u, $x = \sqrt{Au/c} + Bu/c$ satisfies $cx^2 \ge u(A + Bx)$ by expanding the square. Set $u = \ell(\epsilon)$ and take the union bound. □

The leading variance term is expressed in d, L, c_β through (37). The linear summand correction retains the BMY logarithmic penalty. Neither term follows by simply inserting an effective sample size into the independent theorem.

C.2 A latent Markov chain with an absolute spectral gap

Assume a stationary, time-homogeneous Markov chain S_i on a standard Borel space has invariant law π. Write $\Pi u = (\int u \, d\pi)1$ and

$$\chi = \|P - \Pi\|_{L^2(\pi) \to L^2(\pi)} < 1, \qquad \gamma = 1 - \chi > 0. \tag{40}$$

Here χ is the contraction coefficient and γ is the absolute spectral gap. Let $X_i = T_i(S_i)$ for deterministic measurable maps T_i, constant within each distributional segment, whose induced densities obey Assumption A.2. Equivalently, the result below only needs $Z_i = F_i(S_i)$, with $\mathbb{E}_\pi F_i = 0$, $\|F_i\|_{\mathrm{op}} \le b_d$, and $\left\|\mathbb{E}_\pi F_i^2\right\|_{\mathrm{op}} \le v_d$. The marginal observation law can change even though the latent state law remains stationary.

Theorem C.3 (Markov uniform CUSUM bound)**.** *Under this model, put*

$$c_M = \frac{32}{\pi^2}, \qquad \alpha_\chi = \frac{1+\chi}{1-\chi}, \qquad b_\chi = \frac{8}{\pi(1-\chi)}, \qquad \ell_M(\epsilon) = \log \frac{2p^{\,2-\pi/4}|\mathcal{G}|}{\epsilon}.$$

Then (31) *holds with*

$$\lambda_M(\epsilon) = \sqrt{c_M \alpha_\chi v_d \ell_M(\epsilon)} + c_M b_\chi b_d w_{\mathcal{G}} \ell_M(\epsilon). \tag{41}$$

Proof. For a fixed interval, stationarity makes its first latent state have law π. Apply Neeman et al. (2024, Theorem 2.6) to the functions $s' \mapsto w_i F_i(s')$. They are measurable real symmetric matrices, centered under π, uniformly bounded by $M = b_d w_*$, and their scalar variance bounds sum to $\sigma^2 = \sum_i w_i^2 v_d = v_d$. These are the source theorem's Assumptions 2.1, 2.2 and 2.4. Time dependence of the functions, including a marginal change within the interval, is permitted by that theorem.

The source tail inequality, applied also to the negative matrices, yields

$$\mathbb{P}\{\|E_I(t)\|_{\mathrm{op}} > x\} \le 2p^{\,2-\pi/4} \exp\left\{ -\frac{x^2}{c_M(\alpha_\chi v_d + b_\chi b_d w_* x)} \right\}.$$

For $\chi = 0$ the source linear coefficient is $4/(3\pi)$; replacing it by $b_0 = 8/\pi$ only increases the bound. For $\chi > 0$ it is b_χ. The elementary inversion used in Theorem C.2, followed by a union bound over $\mathcal{G}$, proves the assertion. □

This route separates a variance inflation of order γ^{-1} from a linear correction of that order. The fixed latent transition kernel and changing observation maps allow the spectral-gap inequality to apply directly across segment boundaries.

C.3 PHYSICAL DEPENDENCE AND FINITE MEMORY

Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be independent identically distributed innovations in a standard Borel space, and let

$$Z_i = G_i(\varepsilon_i, \varepsilon_{i-1}, \ldots)$$

be a measurable centered causal representation. The maps may change with i. Let $Z_i^{[j]}$ replace only ε_{i-j} by an independent copy. We impose the following strong, explicit version of physical dependence, using the coupling framework of Wu (2005):

$$\sup_i \operatorname{ess\,sup} \left\| Z_i - Z_i^{[j]} \right\|_{\mathrm{op}} \le A_d \rho^j, \qquad j \ge 0, \quad A_d \ge 0, \quad 0 < \rho < 1. \tag{42}$$

The essential-supremum condition supplies the deterministic remainder used in the exponential bound. Continue to assume the envelope and second moments in (25).

Lemma C.4 (Finite memory approximation)**.** *For an integer $q \ge 0$, define*

$$Z_i^{(q)} = \mathbb{E}[Z_i \mid \varepsilon_i, \ldots, \varepsilon_{i-q}], \qquad R_q = \frac{A_d \rho^{q+1}}{1-\rho}.$$

Then

$$\left\| Z_i - Z_i^{(q)} \right\|_{\mathrm{op}} \le R_q \text{ almost surely}, \quad \mathbb{E} Z_i^{(q)} = 0, \quad \left\| Z_i^{(q)} \right\|_{\mathrm{op}} \le b_d, \quad \mathbb{E}(Z_i^{(q)})^2 \preceq v_d I_p.$$

Within each residue class modulo $q+1$, the matrices $Z_i^{(q)}$ are mutually independent.

Proof. Put $\mathcal{F}_{i,j} = \sigma(\varepsilon_i, \ldots, \varepsilon_{i-j})$. For $j \ge 1$, replacing ε_{i-j} and integrating out the older innovations gives

$$\mathbb{E}[Z_i^{[j]} \mid \mathcal{F}_{i,j}] = \mathbb{E}[Z_i \mid \mathcal{F}_{i,j-1}].$$

The difference between successive conditional expectations therefore has operator norm at most $A_d\rho^j$. The bounded martingale $\mathbb{E}[Z_i \mid \mathcal{F}_{i,j}]$ converges entrywise almost surely to Z_i as $j \to \infty$, because the increasing sigma fields generate its causal inputs. For finite matrix dimension this also implies norm convergence. Telescope from q to a finite Q, bound the finite sum of norms, and then let $Q \to \infty$. The geometric series gives R_q.

Centering and the envelope follow from conditional expectation. For $\mathcal{F} = \mathcal{F}_{i,q}$,

$$\mathbb{E}[Z_i^2 \mid \mathcal{F}] - (\mathbb{E}[Z_i \mid \mathcal{F}])^2 = \mathbb{E}[(Z_i - \mathbb{E}[Z_i \mid \mathcal{F}])^2 \mid \mathcal{F}] \succeq 0.$$

Taking expectations proves the second moment bound. Innovation windows of length $q+1$ are disjoint within each residue class; independence of the innovations proves its final assertion. □

Theorem C.5 (Physical dependence uniform CUSUM bound)**.** *Let $a = q+1$ and $\ell_q(\epsilon) = \log(2pa|\mathcal{G}|/\epsilon)$. Under* (42)*,* (31) *holds with*

$$\lambda_F(q,\epsilon) = \sqrt{2av_d\ell_q(\epsilon)} + \frac{2}{3}ab_d w_{\mathcal{G}} \ell_q(\epsilon) + \sqrt{n}\, R_q. \tag{43}$$

An explicit deterministic choice giving $\sqrt{n}\, R_q \le \sqrt{v_d}$ is

$$q_* = \max\left\{0, \left\lceil \frac{\log_+\!\big(A_d\sqrt{n}/((1-\rho)\sqrt{v_d})\big)}{|\log\rho|} \right\rceil\right\}, \tag{44}$$

where $q_* = 0$ if $A_d = 0$.

Proof. For a fixed triple, split $\sum_i w_i Z_i^{(q)}$ into its a residue classes. In class j, let $U_j = \sum_{i \text{ in class } j} w_i^2$. Independent matrix Bernstein gives a deviation at most $\sqrt{2v_dU_ju} + 2b_dw_*u/3$ except on an event of probability $2pe^{-u}$. Empty classes contribute zero and need no concentration. The triangle inequality and $\sum_j \sqrt{U_j} \le \sqrt{a}$ give the first two terms of (43). Independence between different classes is not asserted or used. The remaining sum is bounded deterministically by

$$\left\| \sum_i w_i (Z_i - Z_i^{(q)}) \right\|_{\mathrm{op}} \le R_q \sum_i |w_i| \le \sqrt{N} R_q \le \sqrt{n} R_q.$$

Use $u = \ell_q(\epsilon)$ and union bound over the classes and all triples. Finally, (44) makes $\rho^{q_*} A_d\sqrt{n}/((1-\rho)\sqrt{v_d}) \le 1$ whenever the ratio exceeds one; otherwise $q_* = 0$ already suffices. □

The parameter q is chosen before viewing the observations. Minimizing over a fixed finite list of values requires another union bound. If q_* is large relative to the segment length, the bound remains valid but can make detection impossible to certify. This is a sample-size limitation retained in the noise term.

C.4 Nonempty models and the difference between the routes

All three routes admit bounded marginal densities and rank-two changes. For $d \geq 2$, consider

$$f_\theta(x) = 1 + 3\theta x_1 x_2, \qquad |\theta| \leq 1/6.$$

Its marginals are uniform, its density lies in $[1/2, 3/2]$, and $M(f_\theta) - M(f_0)$ has only two nonzero entries, both $\theta/\sqrt{2}$, hence rank two and Frobenius norm $|\theta|$. Let T_θ keep coordinates $1, 3, \ldots, d$ unchanged and transform the second uniform coordinate to the conditional density $y \mapsto (1+3\theta x_1 y)/2$ on $[-1, 1]$. This density is at least $1/4$, so its conditional distribution function has a unique inverse.

For the Markov route, take a latent state with law μ and transition $P = (1-a)I + a\Pi$, $0 < a \leq 1$, and use $X_i = T_{\theta_i}(S_i)$ with piecewise constant θ_i. The chain has $\chi = 1 - a$, and refreshing at least once separates the past from the future, giving $\beta_Z(h) \leq (1-a)^h$. The model supplies both a spectral-gap example and an absolute-regularity example, including across a change of θ_i.

For a distinct physical-dependence example, let

$$U_i = \sum_{j=0}^{\infty} 2^{-j-1} B_{i-j}, \qquad B_i \sim \text{Bernoulli}(1/2)$$

with independent bits, and take $S_{i1} = 2U_i - 1$. Use fresh independent uniform inputs for the other coordinates at each time, and set $X_i = T_{\theta_i}(S_i)$. Every S_i has law μ, so the same bounded densities hold. Replacing the bit at lag j changes S_{i1} by at most 2^{-j}. Writing the conditional distribution function as

$$F_{\theta,x}(y) = \frac{y+1}{2} + \frac{3\theta x}{4}(y^2 - 1)$$

gives $|\partial_x F| \leq 1/8$ and $\partial_y F \geq 1/4$. Its inverse is Lipschitz in x with constant at most $1/2$. The entries of H are uniformly Lipschitz on the cube, with maximum entrywise constant at most $3\sqrt{5}$ for the maximum norm on x. The row-sum bound for a symmetric matrix gives $\|H(x) - H(y)\|_{\text{op}} \leq 3\sqrt{5}(d+1)\,\|x - y\|_\infty$. Replacing a whole current innovation has a bound of this same order, so (42) holds with $\rho = 1/2$ and $A_d = 6\sqrt{5}(d+1)$.

Nevertheless, U_i is almost surely the fractional part of $2^h U_{i+h}$, so it is measurable from a future observation at any lag. Partition $[0, 1]$ into M equal intervals. The same partition event can be read from both the past and that future observation, which gives absolute regularity at least $1 - 1/M$. Letting $M \to \infty$ yields $\beta_Z(h) = 1$: the centered first row of H still contains U_i. This process separates the two dependence routes: its finite-memory approximation is geometrically accurate even though its absolute regularity coefficient stays equal to one.

D Estimators and deterministic error bounds

This section is conditional on a single event,

$$\mathcal{E}(\lambda) = \left\{ \max_{(s,t,e)\in\mathcal{G}} \|E_I(t)\|_{\text{op}} \leq \lambda \right\}, \qquad \varepsilon_r = \sqrt{r}\,\lambda.$$

Any of the four concentration theorems supplies its probability. Thus the statistical assumptions and the deterministic estimation argument remain separate.

D.1 Population geometry and low-rank scores

Lemma D.1 (Population shape and exact margin). *If $(s, e]$ has exactly one interior change at b, from M_0 to M_1, and $D = M_1 - M_0$, then $A_I(t) = a_b(t)D$, where*

$$a_b(t) = \begin{cases} \dfrac{e-b}{\sqrt{e-s}}\sqrt{\dfrac{t-s}{e-t}}, & t < b, \\ \sqrt{\dfrac{(b-s)(e-b)}{e-s}}, & t = b, \\ \dfrac{b-s}{\sqrt{e-s}}\sqrt{\dfrac{e-t}{t-s}}, & t > b. \end{cases} \tag{45}$$

Writing $\kappa_I = \|D\|_{\mathrm{F}}$, one has

$$\|A_I(b)\|_{\mathrm{F}}^2 - \|A_I(t)\|_{\mathrm{F}}^2 = \begin{cases} \dfrac{(e-b)(b-t)}{e-t}\kappa_I^2, & t < b, \\ \dfrac{(b-s)(t-b)}{t-s}\kappa_I^2, & t > b. \end{cases} \tag{46}$$

For $\kappa_I > 0$ the population score has a unique maximizer at b.

Proof. When $t < b$, the left mean is M_0 and the right mean is $[(b-t)M_0 + (e-b)M_1]/(e-t)$. Their difference is $(e-b)D/(e-t)$. When $t > b$, the right mean is M_1 and the left mean is $[(b-s)M_0 + (t-b)M_1]/(t-s)$, whose difference from M_1 is $(b-s)D/(t-s)$. Multiplying by the CUSUM normalizer gives (45), including its value at b. For $t < b$, subtracting the squared coefficients gives

$$\frac{(b-s)(e-b)}{e-s} - \frac{(e-b)^2(t-s)}{(e-s)(e-t)} = \frac{(e-b)(b-t)}{e-t}.$$

For $t > b$, the corresponding subtraction gives $(b-s)(t-b)/(t-s)$. Both expressions are strictly positive away from b. No temporal independence is used. □

Lemma D.2 (Score and direction perturbations). *For symmetric A, E,*

$$|F_r(A+E) - F_r(A)| \le \sqrt{r}\,\|E\|_{\mathrm{op}}. \tag{47}$$

If $F_r(A+E) > 0$ and $\widehat{V} = (A+E)_{(r)}/F_r(A+E)$, then

$$\langle A, \widehat{V}\rangle_{\mathrm{F}} \ge F_r(A) - 2\sqrt{r}\,\|E\|_{\mathrm{op}}. \tag{48}$$

Proof. Let $\mathcal{V}_r = \{V = V^\top : \mathrm{rank}(V) \le r, \|V\|_{\mathrm{F}} = 1\}$. If P projects onto the image of V, then $\langle A, V\rangle_{\mathrm{F}} = \langle PA, V\rangle_{\mathrm{F}} \le \|PA\|_{\mathrm{F}}$. In an eigenbasis of A^2, $\|PA\|_{\mathrm{F}}^2 = \mathrm{tr}(PA^2)$ is at most the sum of the largest r eigenvalues of A^2, since the diagonal entries of P lie in $[0,1]$ and sum to at most r. Equality is attained by $A_{(r)}/F_r(A)$ when $F_r(A) > 0$. If $A = 0$, the supremum is zero. Hence $F_r(A) = \sup_{V\in\mathcal{V}_r}\langle A, V\rangle_{\mathrm{F}}$. For each feasible V, spectral decomposition gives $|\langle E, V\rangle_{\mathrm{F}}| \le \|E\|_{\mathrm{op}}\|V\|_* \le \sqrt{r}\,\|E\|_{\mathrm{op}}$. Comparing the two suprema proves (47). If V_* maximizes the population functional, optimality of $\widehat{V}$ gives

$$\langle A, \widehat{V}\rangle_{\mathrm{F}} \ge \langle A, V_*\rangle_{\mathrm{F}} + \langle E, V_* - \widehat{V}\rangle_{\mathrm{F}} \ge F_r(A) - 2\sqrt{r}\,\|E\|_{\mathrm{op}}.$$

The assertion is also valid when $A = 0$. □

Corollary D.3 (Uniform score control and learned signal). *On $\mathcal{E}(\lambda)$, a null interval has $\widehat{W}_I(t) \le \varepsilon_r$ for every candidate. For a one-change interval with $\mathrm{rank}(D) \le r$,*

$$|\widehat{W}_I(t) - a_b(t)\kappa_I| \le \varepsilon_r.$$

Suppose b belongs to its candidate set, and put $S = a_b(b)\kappa_I$. If $S \ge 6\varepsilon_r$ and $\widehat{t}$ maximizes the empirical score, the normalized truncated CUSUM at $\widehat{t}$ is defined and satisfies

$$\langle D, \widehat{V}\rangle_{\mathrm{F}} \ge \kappa_I/2. \tag{49}$$

Proof. The null and one-change statements follow from Lemma D.2. Empirical optimality implies $a_b(\widehat{t})\kappa_I \geq S - 2\varepsilon_r \geq 4\varepsilon_r$, and the empirical maximum is at least $S - \varepsilon_r > 0$. Apply (48) to $A = a_b(\widehat{t})D$ and divide by $a_b(\widehat{t}) > 0$. □

Proposition D.4 (Approximation tail)**.** *Let* $\rho_r(A) = \|A - A_{(r)}\|_{\mathrm{F}}$. *Then*

$$\|(A+E)_{(r)} - A\|_{\mathrm{F}} \leq \rho_r(A) + 2\sqrt{2r}\,\|E\|_{\mathrm{op}}. \tag{50}$$

If $A = aD$, $a \geq 0$, *put* $\kappa_r = F_r(D) = (\|D\|_{\mathrm{F}}^2 - \rho_r(D)^2)^{1/2}$. *The score is centered at* $a\kappa_r$, *so*

$$|F_r(A+E) - a\,\|D\|_{\mathrm{F}}\,| \leq \sqrt{r}\,\|E\|_{\mathrm{op}} + a(\|D\|_{\mathrm{F}} - \kappa_r) \leq \sqrt{r}\,\|E\|_{\mathrm{op}} + a\rho_r(D). \tag{51}$$

Proof. Write $B = (A+E)_{(r)}$ and $A_r = A_{(r)}$. Its least-squares optimality implies

$$\|B - A\|_{\mathrm{F}}^2 \leq \|A - A_r\|_{\mathrm{F}}^2 + 2\langle E, B - A_r\rangle_{\mathrm{F}}.$$

Since $\mathrm{rank}(B - A_r) \leq 2r$, the last inner product is at most $\sqrt{2r}\,\|E\|_{\mathrm{op}}\,(\|B - A\|_{\mathrm{F}} + \|A - A_r\|_{\mathrm{F}})$ in absolute value. For $x = \|B - A\|_{\mathrm{F}}$, $a_0 = \rho_r(A)$, $u = \sqrt{2r}\,\|E\|_{\mathrm{op}}$, this gives $(x-u)^2 \leq (a_0+u)^2$ and hence $x \leq a_0 + 2u$. For the score, apply Lemma D.2 and use $\|D\|_{\mathrm{F}} - \kappa_r = \rho_r(D)^2/(\|D\|_{\mathrm{F}} + \kappa_r) \leq \rho_r(D)$; the zero matrix case follows directly. □

D.2 A DISCRETE SEEDED DETECTION PROCEDURE

Take $m \in 4\mathbb{N}$. For each $h = 2^j m \leq n$, include intervals

$$I_{h,q} = (qh/2, qh/2 + h], \qquad q \in \mathbb{Z}_{\geq 0}, \quad qh/2 + h \leq n.$$

Their candidates are all integers in $[s + h/4, e - h/4]$. Write $\mathcal{I}_m$ for this deterministic collection and $\mathcal{G}_m$ for its triples. An interval has an interior change only when $s < \eta_k < e$; a change exactly at an endpoint gives no within-interval mean jump.

For each interval, compute its largest score $\widehat{S}_I$ and a maximizing split $\widehat{b}_I$. Start with the component $(0, n]$. While an interval contained in a current component has score greater than a fixed threshold τ, select one of shortest length, record $(I, \widehat{b}_I)$, and replace its containing component $(a, c]$ by the nonempty parts $(a, s]$ and $(e, c]$. Ties are resolved by increasing left endpoint and then increasing split. Terminate when no such interval remains.

Theorem D.5 (Detection and isolation)**.** *On* $\mathcal{E}(\lambda)$ *for* $\mathcal{G}_m$, *the procedure returns no changes when* $K = 0$ *and* $\tau > \varepsilon_r$. *For* $K \geq 1$, *suppose*

$$m \leq \Delta/8, \qquad \varepsilon_r < \tau < \frac{\sqrt{\Delta}}{8\sqrt{2}}\kappa - \varepsilon_r. \tag{52}$$

It then returns exactly K *intervals, each of length less than* $\Delta/4$, *with exactly one interior true change. Their recorded splits satisfy* $|\widehat{b}_k - \eta_k| < \Delta/4$ *after matching in order.*

Proof. The central halves of the intervals of length h tile the real line: $[qh/2 + h/4, qh/2 + 3h/4]$ meets the next central half at one endpoint. If a change b is at distance at least h from the endpoints of a component, one such interval is contained in the component and satisfies $b - s, e - b \geq h/4$.

Choose the first dyadic $h \geq \Delta/8$; since $m \leq \Delta/8$, one has $\Delta/8 \leq h < \Delta/4$. If b is at least $3\Delta/4$ from both component boundaries, the preceding construction gives an interval with only this change and

$$a_b(b)\kappa_k \geq \frac{\sqrt{h}}{4}\kappa \geq \frac{\sqrt{\Delta}}{8\sqrt{2}}\kappa.$$

The score at this split exceeds τ by (52).

Inductively maintain that the components contain all and only the undetected changes, each at distance at least $3\Delta/4$ from its component boundaries. Initially the distance is at least Δ. A component interval without an interior change has score at most ε_r, so cannot be selected. Every remaining change has the active isolating interval just constructed. Thus the shortest active interval has length

below $\Delta/4$ and exactly one interior change, say η_j. If η_{j-1} remains on its left, its distance to the new right boundary s is

$$s - \eta_{j-1} = (\eta_j - \eta_{j-1}) - (\eta_j - s) > \Delta - \Delta/4 = 3\Delta/4.$$

The right neighbor satisfies the corresponding inequality at e. Old component boundaries are unchanged, so the invariant persists. The procedure removes one change per iteration and cannot stop with any remaining change. After K iterations every remaining interval is null. Both the reported split and its true change lie in the selected interval, which proves the stated error bound. □

There are at most $2n/h$ intervals at each scale, and at most h candidates per interval. Consequently $|\mathcal{G}_m| \le 2n(1 + \lfloor \log_2(n/m) \rfloor)$. For later direction learning, augment the collection with central-half scans on $(s-h, e+h]$ for every $I = (s, e] \in \mathcal{I}_m$, $h = e - s$, whose expanded endpoints lie in $[0, n]$. Call this deterministic set $\mathcal{G}_m^+$. There are at most $2h$ candidates per expanded interval, and hence

$$|\mathcal{G}_m^+| \le 6n(1 + \lfloor \log_2(n/m) \rfloor). \tag{53}$$

Every triple remains central and has length at least m, so all four concentration theorems apply to this augmentation.

Corollary D.6 (Preliminary location and detection SNR)**.** *For a one-change window of length N, suppose the true split belongs to its candidate set and is at least αN from both endpoints, where $0 < \alpha \le 1/2$. On its uniform score event,*

$$|\widehat{b} - b| \le \frac{2\sqrt{Nr}\lambda}{\alpha\kappa_I}, \qquad \frac{|\widehat{b} - b|}{N} \le \frac{2}{\alpha\sqrt{N\kappa_I^2/(r\lambda^2)}}. \tag{54}$$

For seeded multiple-change detection define

$$\mathrm{SNR}_{\mathrm{det}} = \frac{\Delta\kappa^2}{r\lambda^2}.$$

The threshold interval (52) *is nonempty when* $\mathrm{SNR}_{\mathrm{det}} > 512$*. The specific choice $\tau = 6\sqrt{r}\lambda$, used for the refinement construction below, satisfies* (52) *when* $\mathrm{SNR}_{\mathrm{det}} > 6272$*.*

Proof. Let $W(t) = a_b(t)\kappa_I$. Maximization gives $W(b) - W(\widehat{b}) \le 2\varepsilon_r$. The exact margin is at least $\alpha|\widehat{b} - b|\kappa_I^2$. Since $W(\widehat{b}) \le W(b) \le \sqrt{N}\kappa_I/2$,

$$\alpha|\widehat{b} - b|\kappa_I^2 \le [W(b) - W(\widehat{b})][W(b) + W(\widehat{b})] \le 2\varepsilon_r\sqrt{N}\kappa_I.$$

The SNR claims follow respectively from $\sqrt{\Delta}\kappa/(8\sqrt{2}) > 2\varepsilon_r$ and $> 7\varepsilon_r$. □

The constants provide sufficient thresholds. The selected seed interval supplies isolation through Theorem D.5. The expanded window in the next section places the true split inside its central candidate set, which is the condition required for (54) and direction learning.

E Scalar refinement and its probability guarantee

The previous section gives detection and an isolating interval. A local rate requires an additional construction: learn a direction without using the localization observations, provide pure mean-estimation blocks, and compare the local objective at every candidate with its value at the true split. This section supplies all three steps.

E.1 A training-only window construction

Consider two arrays of length T, a pilot array $U_1, \ldots, U_T$ and a validation array $V_1, \ldots, V_T$. At this point all observations in the two arrays are mutually independent. Their segment distributions are the same $f_0, \ldots, f_K$; their change indices obey

$$b_k^V \in \{b_k^U - 1, b_k^U\}. \tag{55}$$

This is exactly the relation obtained by taking the odd and even observations of one independent sequence. It also holds for a deterministically spaced sequence before such a split. Write Δ_U for the minimum spacing including the endpoints $0, T$ in the pilot array.

Run seeded detection on the pilot array only. Use its augmented deterministic collection $\mathcal{G}_m^+$ for the concentration event and write $\lambda = \lambda_{\mathrm{iid}}(\epsilon_P)$, with $n = T$ and $\mathcal{G} = \mathcal{G}_m^+$ in (30). Use $\tau = 6\sqrt{r}\lambda$. For a selected interval $I_k = (s_k, e_k]$ of length h_k, construct

$$J_k = (s_k - h_k, e_k + h_k].$$

Scan the central half of J_k on the pilot array, and normalize the rank-r matrix at its maximum to obtain $\widehat{V}_k$. The symbol $\widehat{V}_k$ denotes a matrix direction, whereas V_j denotes a validation observation.

Project the validation observations in J_k: $Y_j = \langle H(V_j), \widehat{V}_k \rangle_{\mathrm{F}}$. Estimate its two means from the h_k observations with indices $(s_k - h_k, s_k]$ and $(e_k, e_k + h_k]$, respectively. These are the left and right anchors. Using only the central indices $(s_k, e_k]$, minimize

$$Q_k(t) = \sum_{j=s_k+1}^{t} (Y_j - \widehat{\mu}_L)^2 + \sum_{j=t+1}^{e_k} (Y_j - \widehat{\mu}_R)^2, \qquad t \in \{s_k, s_k + 1, \ldots, e_k\}. \tag{56}$$

An empty sum is zero. Choose the smallest minimizer. This rule, including its endpoints, is part of the estimator.

Lemma E.1 (Isolation supplies valid windows and directions)**.** *Suppose the pilot uniform event holds, $m \le \Delta_U/8$, and the detection conditions* (52) *hold with Δ_U and $\tau = 6\sqrt{r}\lambda$. Then every J_k is within $[0, T]$, has only one pilot and one validation change, and places the pilot change in its central half. Both validation anchor blocks are pure and its true split b_k^V belongs to $\{s_k, \ldots, e_k\}$. In addition, $\langle D_k, \widehat{V}_k \rangle_{\mathrm{F}} \ge \kappa_k/2$.*

Proof. By Theorem D.5, $h_k < \Delta_U/4$ and $s_k < b_k^U < e_k$. Since a true change is at least Δ_U from either global endpoint, expansion by h_k remains in $[0, T]$. Any neighboring pilot change is at distance at least $\Delta_U - 2h_k > \Delta_U/2$ from the expanded boundary on its side. The validation index shift has magnitude at most one. Since $m \ge 4$ and $\Delta_U \ge 8m \ge 32$, this positive margin also excludes a neighboring validation change from J_k.

The central half of J_k, whose length is $3h_k$, is $[s_k - h_k/4, e_k + h_k/4]$. It contains b_k^U. The integer relation (55) and $s_k < b_k^U < e_k$ give $s_k \le b_k^V \le e_k - 1$. Therefore all left anchors have index at most b_k^V, all right anchors have index greater than b_k^V, and the central candidate set contains the true validation split.

Write $\varepsilon_r = \sqrt{r}\lambda$. Selection gives $\widehat{S}_{I_k} > 6\varepsilon_r$, so its population score at the selected maximizer exceeds $5\varepsilon_r$. The peak on an interval of length h_k is at most $\sqrt{h_k}\kappa_k/2$, hence $\sqrt{h_k}\kappa_k > 10\varepsilon_r$. On J_k, the distances from b_k^U to its endpoints lie between h_k and $2h_k$, and sum to $3h_k$. Its population peak is at least $\sqrt{2h_k/3}\kappa_k > 6\varepsilon_r$. The expanded scan lies in $\mathcal{G}_m^+$, so Corollary D.3 gives the direction assertion and a positive normalization denominator. □

E.2 Local loss increments and a finite-horizon crossing bound

Lemma E.2 (Scalar localization with estimated anchors)**.** *Fix a window with one scalar mean change of size $\delta_V = \mu_1 - \mu_0 > 0$ at b. Its candidate set contains b. Suppose independent centered errors within each segment satisfy* (32) *with uniform (ν^2, B). Take pure disjoint outer anchor blocks of size q. For any integer $H \ge 1$, the estimator* (56) *satisfies*

$$\mathbb{P}\{|\widehat{b} - b| \ge H\} \le 4\exp\left[-\frac{q}{128 t_V}\right] + 2\exp\left[-\frac{H}{16 t_V}\right], \qquad t_V = \frac{\nu^2}{\delta_V^2} + \frac{B}{\delta_V}. \tag{57}$$

Proof. For an anchor average error u, the scalar moment bound and a Chernoff argument with parameter $a/(\nu^2 + Ba)$ give

$$\mathbb{P}\{|u| > a\} \le 2\exp\left[-\frac{qa^2}{2(\nu^2 + Ba)}\right].$$

Taking $a = \delta_V/8$, the probability that either anchor error u_L, u_R has magnitude greater than $\delta_V/8$ is at most $4e^{-q/(128t_V)}$. On the complementary event $\mathcal{A}$,

$$\widehat{\delta} = \widehat{\mu}_R - \widehat{\mu}_L = \delta_V + u_R - u_L \in [3\delta_V/4, 5\delta_V/4].$$

For $m > 0$ with $b + m$ in the candidate set, expanding the two squares gives

$$Q(b+m) - Q(b) = m\widehat{\delta}(\delta_V - u_L - u_R) + 2\widehat{\delta} \sum_{j=b+1}^{b+m} \xi_j. \tag{58}$$

For $b - m$ in the candidate set, the corresponding identity is

$$Q(b-m) - Q(b) = m\widehat{\delta}(\delta_V + u_L + u_R) - 2\widehat{\delta} \sum_{j=b-m+1}^{b} \xi_j. \tag{59}$$

Thus, on $\mathcal{A}$, a candidate at distance m can have loss no larger than the true split only if the appropriate one-sided centered sum exceeds $3m\delta_V/8$ in magnitude.

For completeness, let independent centered ζ_j obey the uniform moment bound, and let $a > 0$. With $\theta = a/(\nu^2 + Ba)$ and $\phi(\theta) = \theta^2\nu^2/[2(1 - B\theta)]$, the process

$$L_m = \exp\left(\theta \sum_{j=1}^{m} \zeta_j - m\phi(\theta)\right)$$

is a nonnegative supermartingale with $L_0 = 1$. On $\sum_{j=1}^{m} \zeta_j \geq ma$, it is at least $\exp\{ma^2/[2(\nu^2 + Ba)]\}$. Stopping at the first crossing up to the finite available horizon and using $\mathbb{E}L_{\tau\wedge M} \leq 1$ therefore bounds the probability of a crossing at any $m \geq H$ by

$$\exp[-Ha^2/(2(\nu^2 + Ba))].$$

For the right side use $\zeta_j = -\xi_{b+j}$; for the left side reverse the finite pre-change sequence and use $\zeta_j = \xi_{b-j+1}$. Independence and the same moment bounds hold in each ordering. With $a = 3\delta_V/8$,

$$\frac{2(\nu^2 + Ba)}{a^2} = \frac{128\nu^2}{9\delta_V^2} + \frac{16B}{3\delta_V} \leq 16t_V.$$

Adding the two crossing probabilities and the anchor failure probability proves (57). This is an event-inclusion argument. It does not condition the centered noise on the anchor success event. □

Theorem E.3 (Independent paired-array refinement)**.** *In the paired-array construction, suppose $K \geq 1$, $m \in 4\mathbb{N}$,*

$$m \leq \Delta_U/8, \qquad \frac{\Delta_U\kappa^2}{r\lambda^2} > 6272, \qquad m \geq 512t_\kappa \log\frac{8K}{\epsilon_F}, \quad t_x = \frac{\nu^2}{x^2} + \frac{B}{x}, \tag{60}$$

where $\lambda = \lambda_{\mathrm{iid}}(\epsilon_P)$ is computed on the pilot augmented grid, and $\epsilon_P, \epsilon_F \in (0,1)$. Then with probability at least $1 - \epsilon_P - \epsilon_F$ the procedure returns exactly K locations and, after matching,

$$|\widehat{b}_k^V - b_k^V| \leq H_k := \left\lceil 64t_{\kappa_k} \log\frac{4K}{\epsilon_F} \right\rceil. \tag{61}$$

Proof. The pilot uniform event has probability at least $1 - \epsilon_P$. On it, Theorem D.5 and Lemma E.1 give exactly K valid windows and directions with $\delta_{V,k} \geq \kappa_k/2$. Condition on a pilot realization in this event. The windows, directions, and anchor sets are then fixed, and all validation observations retain their original independent laws. The uniform marginal result Proposition B.2 applies to each fixed direction. Moreover, $t_{V,k} \leq 4t_{\kappa_k} \leq 4t_\kappa$ and $q_k = h_k \geq m$.

The total anchor failure bound over all K windows is at most $4K\exp[-m/(512t_\kappa)] \leq \epsilon_F/2$. For the choice in (61), the total crossing failure bound is at most $\sum_k 2\exp[-H_k/(64t_{\kappa_k})] \leq \epsilon_F/2$. A union bound does not require different windows to be independent. Integrate the conditional probability over the pilot event, and add its failure probability. □

For the original independent sequence, set $U_j = X_{2j-1}$, $V_j = X_{2j}$, $T = \lfloor n/2 \rfloor$, dropping a final unpaired observation. Then $b_k^U = \lceil \eta_k/2 \rceil$, $b_k^V = \lfloor \eta_k/2 \rfloor$, and $\Delta_U \geq \lfloor \Delta/2 \rfloor - 1$ whenever all changes remain interior. Under (60), report

$$\widetilde{\eta}_k = \mathrm{clip}_{[0,n]}(2\widehat{b}_k^V + 1).$$

With $\epsilon_P = \epsilon_F = \delta/2$, the simultaneous error bound is

$$|\widetilde{\eta}_k - \eta_k| \leq 2H_k + 1 \quad \text{with probability at least } 1 - \delta. \tag{62}$$

The probability guarantee includes pilot detection failure as well as scalar estimation error.

E.3 Geometric absolute regularity by spaced observations

Lemma E.4 (Coupling the entire retained sequence)**.** *For deterministic times $t_j = 1 + (j-1)g$, $j = 1, \ldots, M$, there exist mutually independent feature matrices Y_j^* with the same individual laws as $Y_j = H(X_{t_j})$ such that*

$$\mathbb{P}\{\textit{some } Y_j^* \neq Y_j\} \leq (M-1)e^{-c_\beta(g-1)} \tag{63}$$

under (33)*.*

Proof. Adding a deterministic mean preserves the sigma field generated by each feature matrix, so the Y_j satisfy the corresponding absolute regularity bounds. Start with $Y_1^* = Y_1$. At step j, apply Berbee's coupling (Berbee, 1979) to Y_j and the original retained past, augmented by all independent randomization variables used at earlier steps. The added variables are independent of the entire original sequence, so they do not increase its absolute regularity coefficient with Y_j. Construct Y_j^* independent of this augmented past, with its own marginal law and disagreement probability at most $e^{-c_\beta(g-1)}$. Earlier coupled variables are measurable functions of that past and its randomization. Induction makes all Y_j^* mutually independent. Sum the $M-1$ disagreement probabilities. This construction and its cost hold for nonidentical marginals. □

For $0 < \delta < 1$, choose the deterministic integer

$$g = 1 + \left\lceil c_\beta^{-1} \log \frac{4n}{\delta} \right\rceil. \tag{64}$$

There are $N_g = \lfloor (n-1)/g \rfloor + 1$ available retained observations. Set $T = \lfloor N_g/2 \rfloor$ and use only the first $2T$ of them:

$$U_j = X_{1+2(j-1)g}, \qquad V_j = X_{1+(2j-1)g}, \qquad 1 \leq j \leq T.$$

The retained sample must satisfy $T \geq m$ and the segment-spacing conditions stated below.

Theorem E.5 (End-to-end refinement under geometric absolute regularity)**.** *Suppose Assumption A.2,* (33) *and* (32) *hold with $K \geq 1$. Use the preceding deterministic spacing and pilot-only algorithm. Let Δ_U be the actual minimum pilot spacing. Assume* (60)*, now using $\epsilon_P = \delta/4$, $\epsilon_F = \delta/2$, and the independent bound $\lambda_{\mathrm{iid}}(\delta/4)$ on the pilot augmented grid of size T. Then with probability at least $1 - \delta$, the procedure returns exactly K estimates and*

$$\widetilde{\eta}_k = \mathrm{clip}_{[0,n]}(1 + 2g\widehat{b}_k^V), \tag{65}$$

$$|\widetilde{\eta}_k - \eta_k| \leq 2g \left\lceil 64 \left(\frac{\nu^2}{\kappa_k^2} + \frac{B}{\kappa_k} \right) \log \frac{8K}{\delta} \right\rceil + g. \tag{66}$$

The pilot spacing satisfies

$$\Delta_U \geq \left\lfloor \frac{\Delta}{2g} \right\rfloor - 1 \tag{67}$$

when each original segment is represented. In particular, a positive right side at least $8m$ supplies the spacing part of (60)*.*

Proof. Couple all $2T$ used feature matrices at once by Lemma E.4. Since $2T \le n$, (64) bounds disagreement by $\delta/4$. On the product law of the coupled sequence, all observations are independent, have the prescribed marginals, and have the same fixed change indices. The paired-array theorem therefore has failure probability at most $\delta/4 + \delta/2$. The complete estimator, including detection, window selection, anchors, directions and local minimization, is a measurable function of these retained matrices. On the agreement event its output equals the independent-law output. Hence its failure probability under the original dependent law is at most δ. We do not condition on the coupling success event and then claim conditional independence.

The pilot and validation lattices interlace. Counting their observations at or before η_k gives $b_k^V \in \{b_k^U - 1, b_k^U\}$. An interval of Δ consecutive integer times contains at least $\lfloor \Delta/(2g) \rfloor$ pilot lattice points before the possible last unpaired point is dropped. At most one pilot point is dropped; only the final segment can lose it. This proves (67), and a positive lower bound ensures that changes are represented and remain distinct.

Finally, b_k^V is the number of validation observations at or before η_k, so

$$1 + (2b_k^V - 1)g \le \eta_k < 1 + (2b_k^V + 1)g.$$

The midpoint $1 + 2gb_k^V$ is at distance at most g from η_k. A change of H_k validation indices moves this midpoint by $2gH_k$. Clipping cannot increase the distance to a true point in $[0, n]$, which proves (66). □

The BMY theorem supports a preliminary scan on all observations. The refinement theorem uses the retained pilot observations for every window and direction choice. This allocation permits independent coupling of the complete estimation procedure: the matrix bound controls direction learning, and the scalar bound controls location after learning.

The Markov and physical-dependence theorems establish detection, isolation, and preliminary localization. Their subclasses satisfying (33) also inherit the spaced refinement guarantee, since that additional condition supplies the required total-variation coupling.

E.4 An independent localization lower bound

Theorem E.6 (Lower bound on the bounded-density rank-two subclass)**.** *Let $d \ge 2$, $0 < \kappa \le 1/8$, and $h = \lfloor 1/(16\kappa^2) \rfloor$. If $n \ge 4h$, there is a class of independent one-change models with densities bounded by $3/2$, jump rank two and jump norm κ, and with the change at least h from the endpoints, such that*

$$\inf_{\widehat{\eta}} \sup_{P} \mathbb{E}_P |\widehat{\eta} - \eta(P)| \ge \frac{3}{256\kappa^2}.$$

The infimum is over all estimators using the original observations.

Proof. Use $f_0 = 1$ and $f_\kappa = 1 + 3\kappa x_1 x_2$. The normalization and jump calculation were given in Section C.4. Since $\log u \le u - 1$,

$$\mathrm{KL}(f_\kappa, f_0) \le \int f_\kappa (f_\kappa - 1)\, d\mu = \int (f_\kappa - 1)^2\, d\mu = \kappa^2.$$

Let P_0, P_1 change from f_0 to f_κ at h and $2h$. Only h coordinates of their product distributions differ, so $\mathrm{KL}(P_0, P_1) \le h\kappa^2 \le 1/16$. Pinsker's bound gives $\mathrm{TV}(P_0, P_1) \le 1/\sqrt{32} < 1/4$. One can verify the used form directly: Jensen's inequality on an event and its complement reduces relative entropy to Bernoulli relative entropy; the second derivative in its first argument is $1/[u(1-u)] \ge 4$. Therefore it is at least twice the squared difference of event probabilities. Take the event attaining total variation.

For any estimate a, $|a - h| + |a - 2h| \ge h$. Integrating this inequality against the minimum of the two densities gives

$$\mathbb{E}_{P_0} |\widehat{\eta} - h| + \mathbb{E}_{P_1} |\widehat{\eta} - 2h| \ge h(1 - \mathrm{TV}(P_0, P_1)) \ge 3h/4.$$

Their larger risk is at least $3h/8$. Since $\kappa \le 1/8$ makes $1/(16\kappa^2) \ge 4$, $h \ge 1/(32\kappa^2)$, proving the stated constant. Both locations obey the endpoint condition when $n \ge 4h$. □

For fixed L and fixed confidence, the independent upper bound has κ^{-2} dependence in the small-jump regime, consistent with this lower bound whenever the separate pilot sample conditions hold. The comparison matches the independent small-jump dependence on κ. Pilot sample requirements and the dependence penalty g remain explicit in the corresponding upper bounds.

F Signal strength, dependence costs, and scope

The statistical reduction separates two sample requirements. The first pays for detecting a matrix mean change and learning its direction. The second pays for locating a scalar mean change after a direction is fixed. This separation explains how ambient dimension can enter the pilot condition while being absent from the subsequent scalar rate for a fixed density bound.

F.1 Detection and preliminary localization

For route $\mathcal{R}$ let $\lambda_{\mathcal{R}}$ be its explicit uniform bound and define

$$\mathrm{SNR}_{\mathrm{det},\mathcal{R}} = \frac{\Delta\kappa^2}{r\lambda_{\mathcal{R}}^2}.$$

A sufficient threshold-existence condition is $\mathrm{SNR}_{\mathrm{det},\mathcal{R}} > 512$, together with $m \le \Delta/8$. The isolation error is then below $\Delta/4$. On any one-change window with N comparable to Δ and the true split in its central candidate set, Corollary D.6 gives an error divided by N of order $(N\kappa_k^2/(r\lambda_{\mathcal{R}}^2))^{-1/2}$. In particular, this ratio tends to zero if that window's SNR diverges.

The following upper bounds expose every dependence parameter. The logarithms are the ones defined in the respective theorems, including the scan cardinality. Put $a = q + 1$. Squaring each explicit radius and using $(x+y)^2 \le 2(x^2+y^2)$ or $(x+y+z)^2 \le 3(x^2+y^2+z^2)$ gives

$$\lambda_{\mathrm{iid}}^2 \le 4v_d\ell + \frac{8b_d^2\ell^2}{3m}, \tag{68}$$

$$\lambda_\beta^2 \le \frac{6}{c_{\mathrm{BMY}}}\left(V_\beta + \frac{b_d^2}{c_\beta m}\right)\ell + \frac{6b_d^2\Gamma(c_\beta,n)^2\ell^2}{c_{\mathrm{BMY}}^2 m}, \tag{69}$$

$$\lambda_{\mathrm{M}}^2 \le 2c_{\mathrm{M}}\alpha_\chi v_d\ell_{\mathrm{M}} + \frac{6c_{\mathrm{M}}^2 b_\chi^2 b_d^2\ell_{\mathrm{M}}^2}{m}, \tag{70}$$

$$\lambda_{\mathrm{F}}^2 \le 6av_d\ell_q + \frac{4a^2b_d^2\ell_q^2}{m} + 3nR_q^2. \tag{71}$$

These inequalities express the squared radii in terms of the model parameters. Here $v_d = L(1+d/2)$ and $b_d = 6d$. For BMY, V_β includes its displayed covariance sum and c_{BMY} is the universal constant in the cited theorem. For Markov dependence, $\alpha_\chi = (1+\chi)/(1-\chi)$ and $b_\chi = 8/[\pi(1-\chi)]$. For physical dependence, $R_q = A_d\rho^{q+1}/(1-\rho)$. Thus smaller c_β, smaller absolute spectral gap $1-\chi$, or larger memory parameter ρ can increase the sufficient signal requirement. Increasing m reduces the bounded-increment terms, but the isolation proof requires $m \le \Delta/8$.

The choice (44) makes $nR_q^2 \le v_d$. It balances a deterministic truncation term against stochastic fluctuation; the leading fluctuation grows with $q+1$. At $\chi = 0$, the Markov radius recovers independent scaling with larger constants. Independent observations are also zero-memory processes, for which $q = 0$ and $R_0 = 0$ give the independent bound. The BMY route is valid for independent data but need not recover the sharper independent radius, because its general-purpose blocking terms remain present.

F.2 Scalar SNR and the localization benefit

Define the scalar noise-to-signal scale and an anchor SNR by

$$t_{\kappa_k} = \frac{\nu^2}{\kappa_k^2} + \frac{B}{\kappa_k}, \qquad \mathrm{SNR}_{\mathrm{loc},k}(q) = \frac{q}{t_{\kappa_k}}.$$

The anchor condition in (60) requires the minimum anchor length divided by t_κ to dominate $\log(K/\delta)$, with the explicit constant given there. This condition controls the signs and magnitudes of the

estimated mean difference. Once it holds, the crossing bound gives $t_{\kappa_k} \log(K/\delta)$ error on the validation index scale.

Under a fixed upper density bound, Proposition B.2 permits $\nu^2 = CL$ and $B = C'$ with universal constants. If $\kappa_k \le \nu^2/B$, then

$$\frac{\nu^2}{\kappa_k^2} \le t_{\kappa_k} \le \frac{2\nu^2}{\kappa_k^2}.$$

This is the specified regime for a κ_k^{-2} localization rate. Outside it the B/κ_k term remains in the bound. The full procedure retains the dependence on L and the pilot feasibility conditions.

For geometric absolute regularity the final bound in (66) has three explicit costs: the pilot sample is reduced to $T = \lfloor N_g/2 \rfloor$, its segment spacing is at least $\lfloor \Delta/(2g) \rfloor - 1$, and each validation-index error costs $2g$ original time units. The final additional g accounts for time resolution between retained observations. The theorem specifies $g = 1 + \lceil c_\beta^{-1} \log(4n/\delta) \rceil$. Consequently its small-jump error is of order $gL\kappa_k^{-2} \log(K/\delta) + g$, when all pilot and anchor conditions hold. The logarithmic spacing penalty is a proved sufficient price for this coupling construction; an optimality statement for that penalty would require a separate lower bound.

F.3 Coverage of the guarantees

Model	Detection and isolation	Fine localization
Independent	Explicit matrix radius and SNR	Complete paired-array construction
Geometric absolute regularity	BMY radius, covariance proxy and SNR	Complete spaced-data construction
Stationary latent Markov chain	Spectral-gap radius and SNR	Spaced refinement with additional absolute regularity
Geometric physical dependence in L^∞	Memory approximation, remainder and SNR	Spaced refinement with additional absolute regularity

The lower bound in Theorem E.6 supports the independent small-jump dependence on κ on a bounded-density, rank-two subclass. Approximate low rank has the explicit tail control in Proposition D.4; the end-to-end theorems here assume exact jump rank at most r.

The experimental implementation evaluates all specified integer candidates. Section 4 examines the reduction using controlled dense interaction changes and activity-derived distributions.

G Experimental models, protocols, and complete results

G.1 An exact dense rank-two family

Let S_a denote a sum of a independent uniform variables on $[-1, 1]$, and let $m_a = \mathbb{E}|S_a|$. For the partition in (5), define

$$c_{a,b} = \frac{3m_a m_b}{\sqrt{ab}}, \qquad V^*_{ij} = V^*_{ji} = \frac{1}{\sqrt{2ab}} \quad (i \in A, j \in B),$$

with all other entries of V^* zero, including its zeroth row and column.

Proposition G.1 (Dense family and signal normalization)**.** *The function f_θ in* (5) *is a density between $1 - |\theta|$ and $1 + |\theta|$. The marginal distribution on either group is product uniform. For two parameters $\theta_0 \ne \theta_1$,*

$$D = M(f_{\theta_1}) - M(f_{\theta_0}) = (\theta_1 - \theta_0) c_{a,b} V^*, \qquad \operatorname{rank}(D) = 2, \qquad \|D\|_{\mathrm{F}} = |\theta_1 - \theta_0| c_{a,b}.$$

Moreover $\|V^*\|_{\mathrm{F}} = 1$, $g_{V^*}(x) = 3S_a(x)S_b(x)/\sqrt{ab}$, $\mathbb{E}_\theta g_{V^*} = \theta c_{a,b}$, and $\mathbb{E}_\theta g^2_{V^*} = 1$.

Proof. Each group sum has a continuous symmetric distribution, so its sign has mean zero. Independence of the groups under μ gives $\int f_\theta \, d\mu = 1$, and the pointwise bounds give positivity. Integrating

over either group removes the perturbation, proving its other group's uniform marginal law. In particular, the constant, linear, and within-group quadratic coefficients have zero jump.

Exchangeability within the first group gives

$$a\,\mathbb{E}_\mu[X_i \operatorname{sign}(S_a)] = \mathbb{E}_\mu[S_a \operatorname{sign}(S_a)] = m_a.$$

The analogous identity holds for the other group. Therefore, for $i \in A, j \in B$, $\mathbb{E}_\theta X_i X_j = \theta m_a m_b/(ab)$. Multiplication by $3/\sqrt{2}$ gives every off-diagonal block entry of D and hence its stated factorization. The matrix has two nonzero eigenvalues of opposite sign, so its rank is two; its $2ab$ nonzero entries give Frobenius norm one after normalization. Expanding the Frobenius product gives the scalar formula and its mean. Under μ, independence and $\mathbb{E}S_a^2 = a/3$ give $\mathbb{E}_\mu g_{V^*}^2 = 1$. The perturbation in its second moment integrates to zero: reversing every coordinate in the first group preserves $g_{V^*}^2$ and changes the sign of the perturbation. □

We compute the signal factor exactly before simulating. The uniform-sum density, obtained by inclusion and exclusion on the cube, is

$$p_a(s) = \frac{1}{2^a(a-1)!}\sum_{j=0}^{a}(-1)^j\binom{a}{j}(s+a-2j)_+^{a-1}.$$

Symmetry gives $m_a = -2\int_{-a}^{0} s\,p_a(s)\,ds$. For each nonzero term, set $A = a-2j > 0$ and $u = s+A$. Then $\int_0^A (u-A)u^{a-1}\,du = -A^{a+1}/[a(a+1)]$. Consequently

$$m_a = \frac{2}{2^a(a+1)!}\sum_{j=0}^{\lfloor a/2\rfloor}(-1)^j\binom{a}{j}(a-2j)^{a+1}. \tag{72}$$

The code evaluates the integer numerator and denominator before converting to floating point. For equal groups, the central limit theorem and uniform integrability give $c_{a,b} \to 2/\pi$. The experiments set $\theta_0 = -\kappa/(2c_{a,b})$ and $\theta_1 = -\theta_0$, thus comparing the same matrix jump as dimension varies.

Exact sampling. Draw $W \sim \mu$ and an additional independent uniform variable U. Let s_A and s_B be the signs of the two group sums. Choose a target sign $T = s_A$ with probability $(1+\theta)/2$, and $T = -s_A$ otherwise. Replace the entire second block by Ts_BW_B. Conditional on its target sign, this reflected block is uniform on that half of the cube. Given the first block, the two half-cube probabilities are $(1+\theta s_A T)/2$, exactly those under f_θ. This proves that the implemented transport samples the desired density without rejection.

G.2 A SINGLE-CHANGE PROTOCOL IMPLIED BY THE SCALAR PROOF

The single-change experiments use a known-one-change model and a predeclared central search region. This separates the evaluation of representation and localization from unknown-count selection. For paired arrays of length T, take $q = \lfloor T/4\rfloor$. Scan the pilot over its full interval, using all integer candidates in its central half. Learn the normalized rank-r matrix at the score maximum. On evaluation observations, use indices $1,\dots,q$ and $T-q+1,\dots,T$ as anchors, and minimize the scalar loss over splits $q,\dots,T-q$.

Corollary G.2 (Single-window refinement)**.** *Suppose the arrays are mutually independent, each has one change, $b^V \in \{b^U-1, b^U\}$, the pilot split is an admissible central candidate, and $q \le b^V \le T-q$. Let λ control the full pilot scan with failure probability ϵ_P. If its population peak obeys $a_{b^U}(b^U)\kappa \ge 6\sqrt{r}\lambda$ and $q \ge 512t_\kappa\log(8/\epsilon_F)$, then with probability at least $1-\epsilon_P-\epsilon_F$,*

$$\langle D, \widehat{V}\rangle_{\mathrm{F}} \ge \kappa/2, \qquad |\widehat{b}^V - b^V| \le \lceil 64t_\kappa\log(4/\epsilon_F)\rceil\,.$$

Proof. On the pilot event, Corollary D.3 supplies the direction and its signal. Condition on that pilot realization. The two anchor blocks are pure by the assumed position of b^V, and the central evaluation observations are disjoint from both. Proposition B.2 and Lemma E.2 apply with scalar jump at least $\kappa/2$, so $t_V \le 4t_\kappa$. The anchor failure probability is bounded by $4e^{-q/(512t_\kappa)} \le \epsilon_F/2$. At the displayed error radius, the crossing failure probability is at most $\epsilon_F/2$. Integrating the conditional bound and adding the pilot failure gives the claim. □

The corresponding original-time estimate is $2\widehat{b}^V + 1$. Deterministic spacing replaces the conversion by $1 + 2g\widehat{b}^V$, with resolution error at most g, as in Theorem E.5. The experiment's random true locations stay in $[0.4n, 0.6n]$, so the outer-quarter anchors are pure.

G.3 Dependence generator and retained sample sizes

Let S_1 be uniform on $[0,1]^{d+1}$. At each next time, independently replace the state by a fresh uniform draw with probability a, otherwise keep the state. Its transition operator is $P = (1-a)I + a\Pi$. Hence $\|P - \Pi\| = 1 - a$. Conditional dependence across a gap of h disappears if at least one refresh occurs, giving $\beta(h) \le (1-a)^h$. Apply the preceding sampling transport with the segment's value of θ to S_i. The resulting marginals are exactly f_θ, and measurable observation maps preserve the mixing bound even across a change.

For $n = 131072$ and $\delta = 0.05$, the retained spacings for refresh probabilities $1, 0.8, 0.5, 0.25$ are $1, 12, 25, 58$. The corresponding pilot sample sizes are $65536, 5461, 2621, 1130$. All pilot scans, directions, anchors, and scalar fits use only retained observations. The study uses 32 independent repetitions for each refresh probability. At $a = 1$, observations are independent and spacing one is used directly.

G.4 Comparators and fixed numerical choices

The rank score sums the squares of the largest absolute eigenvalues of each symmetric CUSUM, then takes the square root. Rank-one and full-direction refinements use the same scalar objective as rank-two refinement. The full quadratic scan uses the Frobenius norm; the largest-entry scan uses the maximum absolute matrix entry; the mean scan uses the Euclidean norm of the nonconstant zeroth-row entries. Multiplicative feature normalizations do not affect their argmax. The Gaussian-kernel comparator uses 512 random Fourier features with independently seeded frequencies and offsets. Its frequency covariance is $\sigma^{-2}I$, where σ is the median pairwise Euclidean distance among the first 256 pilot observations. The feature CUSUM uses the same central candidates as the matrix scans.

All candidate maximizations and scalar minimizations evaluate every specified integer split. The implementation chooses the smallest maximizer or minimizer when scores tie. Matrix eigenvalues are computed by a symmetric eigensolver. For a zero matrix, the implementation uses the zero direction; the proved signal conditions exclude that case. Prefix sums provide exact interval statistics up to floating-point arithmetic. The implementation stores $O(np^2)$ prefix entries and uses a full $O(p^3)$ eigensolver per candidate. The seeded family has $O(n \log(n/m))$ candidate splits. Faster partial eigensolvers are compatible with the statistic, but the reported experiments use the full solver.

Synthetic dimension conditions use $n = 4096$, $\kappa = 0.9$, and $d = 16, 32, 64, 128, 200$. Signal conditions use $n = 8192$, $d = 32$, and $\kappa = 0.35, 0.5, 0.7, 0.9$. Each has 40 trials. After the fixed-sample experiment identified low direction alignment at $d = 200$, a recorded diagnostic added $n = 8192, 16384$ with 40 fresh seeds each. Estimator parameters were held fixed, and the original condition remains in every relevant summary.

G.5 Unknown-count calibration and evaluation

The seeded experiment uses $n = 16384$, $d = 16$, rank two, $m = 128$, and $\kappa = 0.9$. The three change locations are the quarter, half, and three-quarter indices, each independently perturbed by an integer uniform on $[-150, 150]$. Segment parameters alternate signs. Eighty independent null sequences with constant parameter $\theta = \kappa/(2c_{8,8})$ calibrate the threshold: take order statistic $\lceil (80+1)(1-0.05) \rceil = 77$ of the maximum seeded score. The calibration seeds are separate from 60 null and 60 change trials. This is a practical empirical calibration, while the theorem uses its stated analytic radius.

The seeded implementation applies the proved deletion and expansion rules. If a selected interval's expansion extends outside the global data range, it returns the preliminary estimate for that interval and records the invalid expansion. This totalizes the procedure outside the theorem's successful isolation event. All expansions in the 60 change trials were within range. Three null trials returned a false positive, and three change trials returned four locations; all trials enter the reported summaries. For

nonempty true and estimated sets, Hausdorff distance uses the larger of the two directed nearest-point distances. A one-sided empty set receives error n, while two empty sets receive zero.

G.6 Activity-derived distribution changes

The UCI human activity recognition dataset (Reyes-Ortiz et al., 2013) supplies 561 bounded features and separate training and test splits. We select the 128 coordinates with largest variance among training observations labeled walking (1) or walking downstairs (3). The test pools contain 496 and 420 observations, respectively. A synthetic sequence of length 1600 samples with replacement from the walking pool before its change and from the downstairs pool afterwards. Each observation is multiplied by an independent Rademacher sign.

For a global sign R, $\mathbb{E}(RX) = 0$ and $(RX)(RX)^\top = XX^\top$. The transformation thus removes first-moment information while preserving every quadratic term. We use rank ten, fixed before the trials, and the same separated single-change protocol. This empirical, discrete resampling experiment measures localization in an activity-derived distribution pair. The bounded-density guarantees are established for the synthetic continuous family; dependence guarantees are examined with the stationary latent refresh model. Feature selection never uses the test pools.

G.7 Complete numerical results

Table 2 reports every condition, including interquartile ranges. The accompanying raw records contain every estimate, random seed, runtime, learned alignment, and retained sample count. Figure 5 supplies the rank and comparator ablations.

Table 2: All single-change trials. Entries are median absolute errors in original time units, with interquartile ranges.

Study	Setting	Rank scalar	Full scan	Mean scan
dimension	16	2.5 [1, 5]	6.5 [1, 14]	742 [343, 994.5]
dimension	32	4.5 [1.75, 10.25]	9 [3, 20.25]	578.5 [344.25, 839.5]
dimension	64	4.5 [1, 15.5]	18.5 [8.25, 47.75]	600.5 [291.25, 1002]
dimension	128	9.5 [4, 30.25]	96 [19.75, 241]	546 [198.5, 920.25]
dimension	200	472 [217.5, 843.5]	237.5 [51, 664.5]	690.5 [388.75, 871.5]
sample size	8192	6 [1.75, 20.25]	107.5 [27.5, 205.25]	1019 [798.25, 1696.5]
sample size	16384	3.5 [1, 8.25]	80 [33.75, 145.75]	2768 [1448.75, 4097.75]
signal	0.35	62.5 [14, 140.25]	206 [44.75, 495]	1368 [495.75, 1822.5]
signal	0.5	20.5 [4.75, 49.25]	48.5 [13.75, 101.25]	1146.5 [317.75, 1577.75]
signal	0.7	8.5 [3, 17.25]	12.5 [5.5, 35.25]	1166 [563.75, 1539.75]
signal	0.9	1.5 [1, 4.25]	8 [3, 12.5]	1306.5 [704.75, 1552.5]
dependence	0.25	69.5 [33.25, 271.5]	292 [133, 670]	18688 [14497.2, 26881.8]
dependence	0.5	29 [15.5, 65.25]	114 [35.25, 306.75]	25663 [6093.75, 31562.5]
dependence	0.8	27 [11.25, 103]	27 [6.5, 67.25]	24544.5 [15762.8, 30507.5]
dependence	1	2 [1, 8.5]	2 [0.75, 4.25]	18945.5 [11895, 32328.8]
har	128	1 [1, 3]	1 [0, 2]	291.5 [161, 368.5]

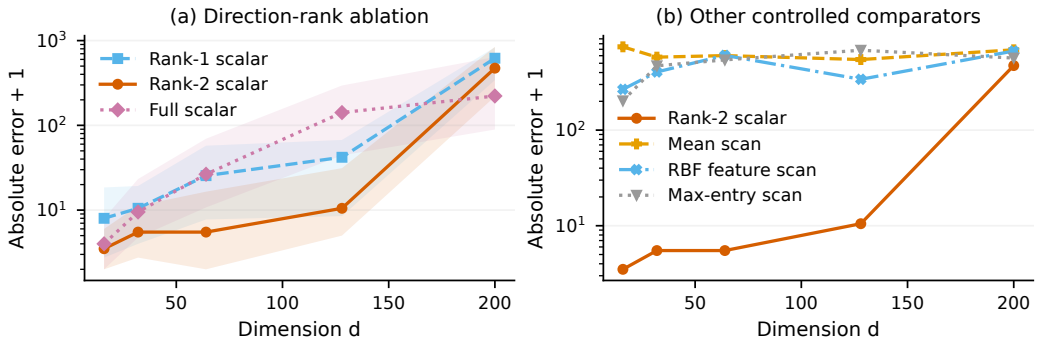

Figure 5: Additional controlled comparisons at $n = 4096$, $\kappa = 0.9$. Rank-one, rank-two, and full-direction scalar estimators use identical evaluation procedures. Bands are interquartile ranges; all medians use 40 trials.