SPECTRAL COMPRESSION FOR CUBIC CHANGE DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Changes in third-order interactions can preserve means and covariances. A change spread across many weak cubic coefficients may nevertheless concentrate in a few singular directions. We study when these directions can be learned and used for inference. For bounded marginal densities, unfolding SVD separates a spectral tail from estimation noise weighted by rank. A geometric inequality converts this error into a retained scalar jump, and a common spectral-energy condition supplies both an initial window and an informative direction. Under geometric absolute regularity, we derive the cubic matrix covariance scale and obtain a uniform concentration bound with an improved leading term. A valid dense dependent model exhibits a strict separation between the resulting sufficient conditions and those from residue-class coupling, including coefficient alternatives. For independent observations, validation and anchor conditions yield a calibrated test and a leading inverse-squared-jump localization bound in the small-signal regime. Related guarantees under dependence follow through analytic scanning and a separate joint-law comparison after thinning. Numerical illustrations examine the spectral construction and selected signal.

1 INTRODUCTION

A distribution can change through an interaction that preserves its means and covariances. A density perturbation proportional to $x_1x_2x_3$ under a product-uniform reference measure provides a simple example: every moment of total degree at most two stays fixed. Cubic features expose the change, but their number grows cubically with dimension. A signal spread across many weak coefficients can therefore be difficult to learn one coordinate at a time. Its tensor representation may still concentrate in a few singular directions.

To exploit this structure, a detector must learn a projection with a substantial difference between its pre-change and post-change means. Both the projection and the change location are unknown. A local contrast estimates the direction accurately only when its window contains enough observations from both segments. An initial search must first supply that window. The relevant question is whether one spectral condition can support this search and retain enough signal for the subsequent statistical decision.

We answer this question through an approximation inequality. Unfolding SVD bounds the tensor fitting error by a population spectral tail plus estimation noise weighted by rank. A geometric argument turns the error into a lower bound on the projected jump. Clipped eigenvalue energy then supplies sufficient conditions for the initial window and a useful direction in a finite rank library. The same retained jump controls independent validation and scalar localization. Under the stated density, validation and anchor conditions, the leading localization term is proportional to the inverse squared jump in the small-signal regime. For orthogonally decomposable changes, global least-squares fitting gives a complementary bound in injective tensor noise.

Temporal dependence makes the noise calculation consequential. Coupling separated observations yields a valid concentration bound with a separation factor in its leading term. We instead derive the matrix covariance scale of cubic features from bounded densities and absolute regularity. Under a fixed geometric mixing envelope, the leading local fitting term retains the independent dimension scale, with a separate logarithmic envelope correction. A bounded density construction shows why this matters: its lower-order projections agree, its cubic coefficients are individually weak, and its

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.

unfolding approaches rank one. On a dependent refresh process, the new bound certifies detection and signal retention in a regime where every displayed coefficient and spectral sufficient condition from residue-class coupling fails. This constructive comparison concerns guarantees on the same signal representation.

1.1 Related work

Kernel two-sample testing (Gretton et al., 2012) and kernel segmentation (Arlot et al., 2019; Garreau and Arlot, 2018) detect distributional changes through mean embeddings. Our full Frobenius contrast is a finite Legendre-kernel contrast. Structured approximation determines which directions within that representation can be learned, complementing learned detection kernels (Chang et al., 2019) and sparse projection for high-dimensional mean changes (Wang and Samworth, 2018). Orthogonal decompositions (Anandkumar et al., 2014), HOSVD (De Lathauwer et al., 2000), and tensor SVD (Zhang and Xia, 2018) supply approximation tools. The features here are correlated polynomial functions of vector observations, whereas tensor factor change detection studies native tensor sequences (Zhang et al., 2026).

The dependence analysis uses matrix concentration for absolutely regular sequences (Banna et al., 2016) and polynomial moment control (Montanaro, 2012). We derive their cubic covariance interface and its consequence for a dense change. Online Poisson-process detection (Luo et al., 2026) studies spectral approximation in an intensity model; robust autocovariance estimation (Xu et al., 2026) controls temporal estimation errors and develops inference for covariance entries. Here the statistical object is a cubic density projection, and the central quantity passed from approximation to inference is its retained scalar jump.

2 Cubic representation and three data roles

Let $X_1, \dots, X_n \in [-1,1]^d$ have marginal density f_0 up to sample count η and f_1 thereafter, relative to product-uniform probability measure μ. The signal analysis assumes $f_0, f_1 \le L$. We first define the representation for these marginals, then give independent-sample inference in Section 4 and dependent-data guarantees in Section 5.

Let $\psi_\alpha(x) = \prod_j \phi_{\alpha_j}(x_j)$ use normalized Legendre polynomials, with $|\alpha| \le 3$, $p = d+1$, and $q = \binom{d+3}{3}$. Associate α with a symmetric tensor E_α containing $3 - |\alpha|$ zero indices and α_j copies of j. Normalize its equal permutation entries to unit Frobenius norm. These tensors form an orthonormal basis, and

$$H_3(x) = \sum_{|\alpha| \le 3} \psi_\alpha(x) E_\alpha, \qquad A(f) = \mathbb{E}_f H_3(X), \quad D = A(f_1) - A(f_0), \quad \kappa = \|D\|_{\mathrm{F}}. \tag{1}$$

If P_3 projects onto polynomials of total degree at most three, then

$$\kappa = \|P_3(f_1 - f_0)\|_{L^2(\mu)}. \tag{2}$$

Thus the tensor preserves the density change visible to cubic features. Remove the constant component, $\bar{H}_3 = H_3 - E_0$, and form

$$\widehat{C}^t_{s,e} = \sqrt{\frac{(t-s)(e-t)}{e-s}} \left\{ \frac{\sum_{i=t+1}^{e} \bar{H}_3(X_i)}{e-t} - \frac{\sum_{i=s+1}^{t} \bar{H}_3(X_i)}{t-s} \right\}. \tag{3}$$

Its mean on a single-change interval is $a(t)D$. A unit direction V retains the scalar jump $\langle D, V \rangle$.

The independent-sample procedure assigns three data roles to address the unknown location and direction. A pilot fold places a local window and fits a library of J directions. An independent validation fold selects and orients the direction $\widehat{V}$ with the largest observed jump. Permuting the entire maximum over the fixed library calibrates this selection. The third fold, denoted C, estimates the two means from outer anchors and locates the split using disjoint central observations. Its projected data are $Z_j = \langle \bar{H}_3(X_j^{(C)}), \widehat{V} \rangle$; separating anchors from the searched observations supplies the independence for scalar refinement.

The main spectral candidate truncates the unfolding of the local contrast to rank r, then folds and symmetrizes. Other candidates select large normalized basis coefficients, restrict those coefficients to

marginal cubics, apply the leading left-singular projector in all three modes, or keep the full contrast. Each fitted tensor B_j gives $V_j = B_j/\|B_j\|_{\mathrm{F}}$, with $V_j = 0$ when $B_j = 0$. The procedure cycles the three fold roles, combines their permutation tests by Bonferroni, and takes the median $\widetilde{\eta}$ of the mapped location estimates. Appendix A gives every index, candidate rule, anchor interval, and randomized test formula.

3 Spectral error and retained change signal

A fitted direction is useful when the signal discarded by compression and the noise retained by fitting are both small. Consider a symmetric local contrast $Y = aD + E$, $a > 0$, and let B_r be its symmetrized rank-r unfolding fit. Write $\tau_r(D)^2 = \sum_{j>r} \sigma_j(D_{(1)})^2$ and $\delta = \|E_{(1)}\|_{\mathrm{op}}$.

Theorem 3.1 (Spectral approximation preserves a scalar jump)**.** *For every symmetric D, $1 \le r \le p$, and $a > 0$,*

$$\|B_r - aD\|_{\mathrm{F}} \le a\tau_r(D) + 2\sqrt{2r}\delta =: e_r. \tag{4}$$

If $e_r < a\kappa$, then $B_r \neq 0$ and

$$\langle D, B_r/\|B_r\|_{\mathrm{F}}\rangle \ge \kappa\sqrt{1 - \{e_r/(a\kappa)\}^2}. \tag{5}$$

These bounds include tied singular values.

The proof identifies why the noise cost depends on rank. Put $A = aD_{(1)}$, $Z = E_{(1)}$, and let M_r, A_r be best rank-r approximations of $A + Z, A$. With $u = \|M_r - A\|_{\mathrm{F}}$ and $t = \|A_r - A\|_{\mathrm{F}}$, optimality gives

$$u^2 - t^2 \le 2\langle Z, M_r - A_r\rangle \le 2\sqrt{2r}\delta(u + t). \tag{6}$$

The difference of the fits has rank at most $2r$, so operator/nuclear norm duality gives the second inequality. Dividing by $u + t$ yields the error bound; folding preserves distance and symmetrization contracts it. For the angle, put $K = a\kappa$ and $z = \|B_r\|_{\mathrm{F}}$. Expanding the squared error gives $\langle aD, B_r/z\rangle \ge (K^2 + z^2 - e_r^2)/(2z) \ge \sqrt{K^2 - e_r^2}$. The fitted norm may therefore be random while the retained population jump stays controlled. Appendix C gives the complete proof and the corresponding basis and Tucker bounds.

The relevant spectrum is that of the Gram matrix $K_D = D_{(1)}D_{(1)}^{\top}$. Its eigenvalues $\mu_1 \ge \cdots \ge \mu_p \ge 0$ satisfy $\sum_j \mu_j = \kappa^2$. For a normalized noise bound $\ell \ge \delta/a$, define

$$\mathcal{E}_\ell(D) = \sum_{j=1}^{p} \min\{\mu_j, 8\ell^2\}. \tag{7}$$

This quantity charges $8\ell^2$ for an estimated spectral component and retains smaller eigenvalues as discarded energy. At $r_\ell = \#\{j : \mu_j > 8\ell^2\}$, it equals $\tau_{r_\ell}(D)^2 + 8r_\ell\ell^2$. If $4\mathcal{E}_\ell(D) \le \rho^2\kappa^2$, $0 < \rho < 1$, a dyadic rank grid augmented by p contains a direction with jump at least $\kappa\sqrt{1-\rho^2}$. Squaring the two-term error bound costs a factor two; rounding the threshold rank upward costs at most another factor two. The population spectrum establishes existence, and validation selects from the fitted library.

This criterion also covers approximate low rank. Polynomial decay $\sqrt{\mu_j} \le Aj^{-s}$, $s > 1/2$, gives an error bound of order $A^{1/(2s)}\ell^{1-1/(2s)}$ at an interior rank of order $(A/\ell)^{1/s}$. With $t = 8\ell^2$ and $d_{\mathrm{eff}}(t) = \sum_j \mu_j/(\mu_j + t)$, the inequalities $td_{\mathrm{eff}}(t) \le \mathcal{E}_\ell(D) \le 2td_{\mathrm{eff}}(t)$ identify the effective dimension at the noise scale. Appendix C.4 gives finite-grid endpoints and exponential decay.

Orthogonal tensor structure changes both the tail description and the available noise bound. Define $\|Z\|_{\mathrm{inj}} = \sup_{\|u\|_2=\|v\|_2=\|w\|_2=1} |Z[u,v,w]|$. If $D = \sum_j \lambda_j v_j^{\otimes 3}$ with orthonormal v_j and decreasing $|\lambda_j|$, its component eigenpairs satisfy $D[v_j, v_j, \cdot] = \lambda_j v_j$. In this model $\mu_j = \lambda_j^2$, so the component eigenvalues determine the unfolding tail. A global least-squares fit over tensors with at most r orthogonal components satisfies

$$\|\widehat{T}_r - D\|_{\mathrm{F}} \le \left(\sum_{j>r} \lambda_j^2\right)^{1/2} + 4r\|Z\|_{\mathrm{inj}}, \qquad Y/a = D + Z. \tag{8}$$

The fitted tensor and the true rank-r truncation occupy a common mode subspace of dimension at most $2r$. Restricting Z to this subspace gives Frobenius norm at most $2r\|Z\|_{\mathrm{inj}}$. The optimality factorization in (6) then yields the factor $4r$. Theorem A.3 states this result and its separate supplied-frame version.

For a balanced independent contrast on $2h$ observations, $a \asymp \sqrt{h}$, polynomial moments give the injective bound below, and the unfolding bound permits the displayed choice of ℓ:

$$\ell \lesssim \sqrt{\frac{Lqx}{ph}} + \frac{\sqrt{q}\,x}{h}, \qquad \|Z\|_{\mathrm{inj}} \lesssim \nu_L \left\{ \sqrt{\frac{p+x}{h}} + \frac{(p+x)^{3/2}}{h} \right\}, \quad \nu_L = C\max(1, \sqrt{L}). \tag{9}$$

Here x includes logarithms of the matrix dimensions, the number of contrasts, and inverse failure probability. The leading stochastic upper bounds are $\sqrt{r}\,p/\sqrt{h}$ for unfolding SVD and $r\sqrt{p/h}$ for the orthogonal fit, with the confidence and envelope terms retained above. Stronger tensor structure therefore changes the estimation requirement. This comparison concerns the implemented matrix estimator and a global tensor optimum. Both connect their approximation error to the retained jump through (5).

4 From retained signal to independent-sample inference

Assume now that the observations are independent. A valid local contrast must be obtained before its useful direction can be learned. Uniform noise control handles the selected window, and the same spectral energy can guarantee both stages. The three roles in Section 2 then carry the retained jump into testing and refinement.

The correlated polynomial entries admit an exact matrix variance calculation. With $M(x) = H_3(x)_{(1)}$ and P_2 the projector onto symmetric order-two tensors,

$$\mathbb{E}_\mu M M^\top = (q/p) I_p, \qquad \mathbb{E}_\mu M^\top M = ((p+2)/3) P_2. \tag{10}$$

Partial traces of the symmetric order-three projector give these identities. The density bound makes the centered variance at most Lq/p, while $\sup_x \|H_3(x)\|_{\mathrm{F}} \le b_0\sqrt{q}$, $b_0 = 3\sqrt{3}$, supplies the summand envelope. Matrix Bernstein over all candidate contrasts therefore covers the data-selected window.

Take $n = 3N$, $h \ge 16$ divisible by eight, and require every true segment in every interlaced fold to have length at least $2h$. Put $\beta = \min_b \min(\tau_b, N - \tau_b)/N$, where $\tau_b = \lfloor (\eta + 2 - b)/3 \rfloor$, $b \in \{0, 1, 2\}$. For failure probabilities $\delta_P, \delta_V, \delta_A, \delta_C \in (0, 1)$, define

$$\begin{aligned} &M_0 = 6(N+1), \quad x_P = \log\{2M_0(p+p^2)/\delta_P\}, \\ &\epsilon_P = b_0\sqrt{2\log(4qM_0/\delta_P)}, \quad U_P = \sqrt{2L(q/p)x_P} + \frac{4b_0\sqrt{q}}{3\sqrt{h}} x_P, \\ &a_0 = \frac{15\sqrt{h}}{16\sqrt{2}}, \qquad \ell_P = U_P/a_0. \end{aligned} \tag{11}$$

Use the pilot score $\max\{\max_\alpha |\langle T, E_\alpha\rangle|/\epsilon_P, \|T_{(1)}\|_{\mathrm{op}}/U_P\}$ and fit every dyadic SVD rank, together with p. This is the normalized pilot in (41) when its grids contain one. Its uniform perturbation is at most one.

Theorem 4.1 (A common spectral condition for detection and localization)**.** *Under the independent model and segment conditions above, let $\kappa > 0$, $0 < \rho < 1$, $g = \kappa\sqrt{1-\rho^2}$, and $\nu_L = C\max(1, \sqrt{L})$. Use a finite library of J candidates containing the stated SVD fits and R independent uniform permutations. Set $x_V = \log\{12J(R+1)/\delta_V\}$, $x_A = \log(12/\delta_A)$, and $x_C = \log(6/\delta_C)$. If $3/(R+1) \le \alpha_0$ and*

$$\mathcal{E}_{\ell_P}(D) \le \min\left\{ \frac{\rho^2}{4}, \frac{4\beta^2 h}{225N} \right\} \kappa^2, \tag{12}$$

$$\sqrt{h}\,g \ge 8C\nu_L(\sqrt{x_V} + x_V^{3/2}/\sqrt{h}) + 4\kappa\sqrt{x_V},$$

$$C\nu_L(\sqrt{2x_A/h} + 2x_A^{3/2}/h) \le g/16, \tag{13}$$

then rejection has probability at least $1 - \delta_P - \delta_V$. *Rejection together with*

$$|\widetilde{\eta} - \eta| \le \mathfrak{r}(g) := C\{\nu_L^2 x_C / g^2 + \nu_L x_C^{3/2}/g + 1\} \tag{14}$$

has probability at least $1 - \delta_P - \delta_V - \delta_A - \delta_C$. *Under any iid null, the permutation test has size at most* α_0 *without the signal or density conditions.*

The energy condition closes the link between the unknown window and direction. Put $t = 8\ell_P^2$ and $S = \mathcal{E}_{\ell_P}(D)$. Since $S < \kappa^2$, the largest Gram eigenvalue μ_1 exceeds t, and $S \ge t\kappa^2/\mu_1$. The resulting leading singular value gives enough pilot signal to place the center within $h/16$ of the change. The same condition yields a local candidate with jump at least g. This is Corollary A.6; Appendix D.7 verifies the constants.

Conditional on the pilot, validation compares fixed directions. Uniform directional error b and population amplitude a_B give $\langle D, \widehat{V}\rangle \ge \max_{j \le J} |\langle D, V_j\rangle| - 2b/a_B$. The validation condition retains at least $g/2$ and separates the observed maximum from its permutations. Under the null, exchangeability calibrates the entire maximum, including selection. On the third fold, write $d_\mu = \langle D, \widehat{V}\rangle \ge g/2$. Accurate independent anchors give the scalar loss a drift of order $v d_\mu^2$ at distance v, while centered partial sums contribute at scale $d_\mu \nu_L(\sqrt{vx} + x^{3/2})$. Dividing by d_μ and using its lower bound gives (14). Each ordering's conditional bound is integrated before the three bounds are combined.

At fixed confidence, bounded L, and fixed $\rho < 1$, the leading small-signal refinement term is proportional to κ^{-2}. Dimension and spectrum still determine whether the direction and pilot can be learned. The two-location lower bound in Appendix F.1 has the same jump exponent in its stated subclass and admissible horizon. The present argument uses independent observations; the next section separates dependence-adjusted estimation from calibration based on a comparison of the whole retained data law.

5 Temporal covariance and recoverable cubic signal

The spectral tail measures discarded signal under both independent and dependent observations. Temporal covariance determines the noise against which that tail must be compared. We keep the marginal bound $f_i \le L$ and let $\beta_X(k)$ be the total variation distance between the joint law of $(X_1, \dots, X_t)$ and $(X_{t+k}, \dots, X_n)$ and the product of their marginal laws, maximized over t. Total variation is $\sup_A |P(A) - Q(A)|$. Assume $\beta_X(k) \le Be^{-ck}$, with fixed $B, c > 0$. The condition includes dependence across changes and permits different segment marginals.

5.1 Preserving the unfolding variance scale

Write $M_i = H_3(X_i)_{(1)}$ and $U_i = M_i - \mathbb{E}M_i$. A cubic polynomial satisfies $\|Q\|_{L^4(\mu)} \le K_4 \|Q\|_{L^2(\mu)}$ with $K_4 = 3^{3/2}$. Combining this fourth moment with the exact basis variance controls arbitrary signed subset sums at the same dimension scale.

Lemma 5.1 (Second moments of arbitrary weighted subsets)**.** *Define*

$$V_3^2 = \frac{q}{p}\left[L + 16K_4^2\sqrt{L}\,\frac{\sqrt{B}}{e^{c/2} - 1}\right]. \tag{15}$$

For every $K \subseteq \{1, \dots, n\}$ *and deterministic real weights* w_i, *put* $S_{K,w} = \sum_{i \in K} w_i U_i$. *Then*

$$\max\left\{\|\mathbb{E}S_{K,w}S_{K,w}^\top\|_{\mathrm{op}}, \|\mathbb{E}S_{K,w}^\top S_{K,w}\|_{\mathrm{op}}\right\} \le V_3^2 \sum_{i \in K} w_i^2. \tag{16}$$

The factor $\sqrt{B}/(e^{c/2} - 1)$ *may be replaced by any justified upper bound for* $\sum_{h \ge 1}\sqrt{\beta_X(h)}$.

The factor $q/p \asymp p^2$ survives temporal summation. Pairwise coupling and Hölder's inequality bound each lagged matrix cross moment by $8K_4^2\sqrt{L}(q/p)\sqrt{\beta_X(k)}$. Summing these bounds proves (16). The arbitrary-subset conclusion supplies the variance parameter required by dependent matrix concentration. Appendix G gives the complete moment and covariance arguments.

The contraction argument is essential to this dimension scale. For any unit vector $v \in \mathbb{R}^p$, Minkowski's inequality bounds the vector fourth moment through the scalar coordinates of $U_i^\top v$. The reference second moment then gives

$$\left\| \|U_i^\top v\|_2 \right\|_{L^4} \le 2K_4 L^{1/4} \sqrt{q/p}.$$

Couple X_j to a copy independent of X_i, and use exponents $4, 4, 2$ on the two vector norms and the indicator of a coupling mismatch. This retains the factor q/p and contributes $\sqrt{\beta_X(j-i)}$. The calculation concerns each deterministic covariance expectation. Consequently the sum over lags enters a variance constant, without an additional sample-law comparison probability. This is the mechanism that removes a sample-size-dependent separation factor from the leading concentration term.

The dependent matrix inequality of Banna et al. (2016) then controls a fixed family of M_0 candidate CUSUMs. For $\delta \in (0,1)$, Theorem G.3 gives explicit simultaneous thresholds $U_{\mathrm{BMY}}(w;\delta)$ with failure probability δ. For a contrast of length m, they satisfy

$$U_{\mathrm{BMY}}(w;\delta) \le C\{p\sqrt{x} + p^{3/2}\|w\|_\infty x \log^2(em)\}, \quad x = \log \frac{2a_{\mathrm{mix}} M_0 (p+p^2)}{\delta}, \tag{17}$$

where $a_{\mathrm{mix}} = 1 + \lceil \log \max\{B,1\}/c \rceil$ and C depends only on L, B, c.

The thresholds can be evaluated without an unspecified constant. Put $c_* = a_{\mathrm{mix}} c$, $B_3 = 6\sqrt{3q}$, and

$$\gamma(c_*, m) = \frac{\log m}{\log 2} \max\left\{2, \frac{32 \log m}{c_* \log 2}\right\}, \qquad m \ge 2.$$

The exact group threshold of Theorem G.3 is bounded above by

$$\overline{U}(w;\delta) = 30\sqrt{2a_{\mathrm{mix}}}\, V_3 \sqrt{x} + 2a_{\mathrm{mix}} B_3 \|w\|_\infty \left\{4\sqrt{x/c_*} + \gamma(c_*, m) x\right\}. \tag{18}$$

Using $\overline{U}$ therefore gives a valid, possibly larger, threshold. The two terms distinguish accumulated covariance from the bounded size of one observation. A fixed partition into residue classes absorbs the geometric prefactor B; it retains every observation and depends on the mixing envelope rather than the sample size.

Applying the inequality separately to the two constant-weight sides preserves the normalized CUSUM variance scale, including unbalanced cuts. For a window of length $m \asymp h$, amplitude $a \asymp \sqrt{h}$, and $\|w\|_\infty \lesssim h^{-1/2}$, the rank-$r$ fitting error divided by a is at most $\tau_r(D) + C\sqrt{r}\{p\sqrt{x/h} + p^{3/2} x \log^2(eh)/h\}$. The variance term dominates when $h \gg px\log^4(eh)$. Coupling residue classes instead gives a leading term $p\sqrt{kx}$ and comparison cost $(n-k)\beta_X(k)$ in Theorem G.5. A logarithmic k controls that cost under the geometric envelope.

5.2 A COMMON SPECTRAL CONDITION AFTER DEPENDENT SAMPLING

The same uniform event can place the fitting window and certify its retained signal. This requires a common deterministic noise bound and maximization of the raw operator score.

Corollary 5.2 (Dependent pilot and informative spectral fit)**.** *Suppose one change has $\kappa > 0$ and both segments have length at least $2h$, where $h \ge 16$ is divisible by eight. Include every full-interval cut $c \in [h, n-h]$ and every contrast on $(c-h, c+h]$ at c in the predetermined family. Let $U_* > 0$ bound all their $U_{\mathrm{BMY}}(w;\delta)$, put $\beta = \min(\eta, n-\eta)/n$, and set*

$$a_0 = \frac{15\sqrt{h}}{16\sqrt{2}}, \qquad \ell_* = U_*/a_0, \qquad \xi_* = \min\left\{\frac{\rho^2}{4}, \frac{4\beta^2 h}{225 n}\right\}, \quad 0 < \rho < 1. \tag{19}$$

If $\mathcal{E}_{\ell_*}(D) \le \xi_* \kappa^2$, every maximizer $\widehat{c}$ of $\|\widehat{C}^c_{0,n(1)}\|_{\mathrm{op}}$ over $c \in [h, n-h] \cap \mathbb{Z}$ obeys $|\widehat{c} - \eta| \le h/16$. The library of symmetrized unfolding-SVD fits at all positive dyadic ranks and endpoint p in that local window contains at least one direction with jump at least $\kappa\sqrt{1-\rho^2}$. Both conclusions hold together with probability at least $1-\delta$.

The proof uses $\mathcal{E}_{\ell_*}(D) \ge 8\ell_*^2\kappa^2/\mu_1$ to control the pilot and Corollary A.2 to control fitting. The event already covers the selected local contrast. Thus covariance concentration strengthens the approximation stage through the same spectral quantities as in Section 3.

5.3 A DENSE SIGNAL THAT DISTINGUISHES THE BOUNDS

Many weak cubic coefficients can combine into one informative singular direction. To construct such a change while maintaining positive densities, put $a = d^{-1/2}$ and define

$$g_d(x) = \sin\left(a \sum_j x_j\right) - c_d \sin\left(2a \sum_j x_j\right), \qquad c_d = \frac{r_1(a) r_0(a)^{d-1}}{r_1(2a) r_0(2a)^{d-1}}, \tag{20}$$

where $r_0(z) = \sin(z)/z$ and $r_1(z) = \mathbb{E}[\phi_1(Z)\sin(zZ)]$ for uniform Z on $[-1, 1]$. The coefficient $c_d \to e^{1/2}/2$ keeps g_d uniformly bounded and cancels every linear coefficient exactly. Oddness cancels all even-degree coefficients, so $P_{\le 2} g_d = 0$, where $P_{\le 2}$ denotes polynomial projection through degree two.

Set $T_d = \mathbb{E}_\mu[g_d(X) H_3(X)]$ and $u_d = (0, d^{-1/2}, \ldots, d^{-1/2})$. The coefficients for three distinct coordinates agree with those of $\lambda_d u_d^{\otimes 3}$, where $\lambda_d \to e^{-1/6}/(3\sqrt{2})$. The remaining $O(d^2)$ cubic coefficients each differ by $O(d^{-3/2})$. Hence the squared approximation error is $O(d^{-1})$. Proposition H.1 proves these identities. For any positive sequence $\kappa_d \to 0$, the density jump $D_d = A(f_{+,d}) - A(f_{-,d})$ satisfies

$$f_{\pm,d} = 1 \pm \frac{\kappa_d}{2\|T_d\|_{\mathrm{F}}} g_d \quad \Longrightarrow \quad \begin{cases} 1/2 \le f_{\pm,d} \le 3/2, \\ \|D_d\|_{\mathrm{F}} = \kappa_d, \quad \tau_1(D_d) = O(\kappa_d/\sqrt{d}), \\ m_d := \max_\alpha |\langle D_d, E_\alpha \rangle| = \Theta(\kappa_d d^{-3/2}) \end{cases} \tag{21}$$

for all sufficiently large d. Both densities have the same polynomial projection through degree two. The spectral tail is small even though the signal is spread across the coefficient basis.

Fix $\varrho \in (0, 1)$. Within each segment, retain the preceding vector with probability ϱ and otherwise draw an independent vector from that segment's density. Start each segment with a fresh draw. This process has the stated marginals and $\beta_X(k) \le \varrho^k$. Take $h = 8\lceil d^3/8 \rceil$, $n = 4h$, $\eta = 2h$, and $\kappa_d = d^{-1/2}(\log d)^{3/4}$. At fixed $\delta \in (0, 1/4)$, include all full-interval cuts in $[h, n-h]$ and all their local windows of radius h in a deterministic family. Polynomially many additional predetermined CUSUM contrasts with $\|w\|_2 = 1$ and $\|w\|_\infty \le h^{-1/2}$ preserve the same noise scale.

Let $a_* = \sqrt{h/2}$, $u_{\mathrm{cov}} = U_{\mathrm{cov}}/a_*$ and $u_{\mathrm{coup}} = U_{\mathrm{coup}}/a_*$ denote the explicit uniform thresholds in (115) and (116). Every coupling choice k, ζ with $\zeta + (n-k)\beta_X(k) \le \delta$ obeys

$$\begin{aligned} u_{\mathrm{cov}} &= O\{d^{-1/2}\sqrt{\log d} + d^{-3/2}(\log d)^3\} = o(\kappa_d), \\ u_{\mathrm{coup}} &\gtrsim d^{-1/2} \log d \gg \kappa_d. \end{aligned} \tag{22}$$

The first bound includes the envelope correction. The second holds uniformly over admissible coupling gaps, since the refresh process has $\beta_X(k) = \varrho^k$ for $k < 2h$ and admissibility forces $k \gtrsim \log d$.

Proposition 5.3 (Separation of coupling and covariance sufficient conditions)**.** *For any fixed $\rho_0 \in (0, 1)$, the process and regime above have the following properties for all sufficiently large d.*

1. *The covariance rank-one SVD bound is $o(\kappa_d)$. The raw unfolding-operator pilot places its center within $h/16$ of the change on the covariance noise event, and its local rank-one fit has normalized error $o(\kappa_d)$.*
2. *Every coupling expression in* (117)*, for every $1 \le s \le q$ and $1 \le r \le p$, is at least κ_d, for every admissible (k, ζ). The corresponding coupling Tucker and full-fit bounds also exceed $\rho_0 \kappa_d$. This includes the minimum over these complete candidate libraries.*
3. *The normalized pilot combining all coefficient sizes and unfolding ranks fails its coupling center-accuracy sufficient condition, including its coefficient-only special case.*
4. *Rejecting when the maximum full-interval unfolding-operator CUSUM exceeds U_{cov} has size at most δ under a common marginal null satisfying the density and mixing bounds, and power at least $1 - \delta$ under the constructed alternative.*

The detection statement uses the explicit concentration threshold with the stated density and mixing envelope.

The comparison includes every coefficient support size. Its key step is

$$t_s(D_d) + \frac{2\sqrt{2s}\epsilon_{\mathrm{coup}}}{a_*} \geq \kappa_d + \sqrt{s}\left(\frac{2\sqrt{2}\epsilon_{\mathrm{coup}}}{a_*} - m_d\right) \geq \kappa_d.$$

The first inequality uses $t_s(D_d) \geq \kappa_d - \sqrt{s}\, m_d$. The second follows from $\epsilon_{\mathrm{coup}} \gtrsim \log d$ and the coefficient scale in (21). In contrast, the rank-one spectral tail and covariance noise are both $o(\kappa_d)$. The full proof appears in Appendix H. The explicit constants can outweigh the asymptotic gain at finite sample sizes; that appendix also quantifies their contribution.

5.4 Calibration and temporal resolution

The uniform event also gives analytic detection using all observations: reject when any raw unfolding operator score exceeds its threshold $U_{\mathrm{BMY}}(w;\alpha)$. A constant feature mean gives rejection probability at most α, using justified values of L, B, c. A population contrast aD gives power at least $1-\delta$ when $a\sqrt{\mu_1} > U_{\mathrm{BMY}}(w;\alpha) + U_{\mathrm{BMY}}(w;\delta)$.

Permutation selection and scalar refinement follow on a predetermined grid with spacing k through Theorem G.6. For m retained observations, the joint-law comparison costs $\varepsilon_{\mathrm{mix}} = (m-1)\beta_X(k)$. Under its retained-segment and product-law signal conditions, the adjusted p value is super-uniform and the original-time location radius is $k\mathfrak{r}(g) + k - 1$, with the stated failure probability increased by $\varepsilon_{\mathrm{mix}}$. The adjusted test uses level $\alpha_0 - \varepsilon_{\mathrm{mix}} > 0$. The comparison applies to the complete retained vector and the procedure's independent permutation randomness. Every event defined by fitting, selection, rejection and refinement then changes in probability by at most $\varepsilon_{\mathrm{mix}}$. This preserves the adaptive procedure as a whole. Estimation on all dependent observations uses the covariance event; the retained-grid procedure obtains its inference guarantee from this separate comparison. This route carries the learned-direction inference guarantee to dependent data with explicit sample-size and temporal-resolution costs.

6 Numerical illustrations of the theoretical mechanisms

The dense construction first gives a population comparison: at $d = 4096$, its largest coefficient is 9.346×10^{-6} of the signal norm, while the specified rank-one tensor has relative approximation error 0.009950. Appendix I.1 and Figure 2 give all 11 dimension settings and separate this error from the optimal unfolding tail. These quantities describe the structural advantage used in the dependence theorem. The simulations examine interaction retention, selected jump size, and localization in the coefficient-pilot procedure. The fixed study contains 720 independent-data trials: 480 alternatives and 240 null sequences. Four core alternatives preserve all moments through degree two and change marginal, quadratic–linear, triple, or mixed cubic terms. We use $n = 12000$, $d \in \{8, 16\}$, and 40 repetitions per alternative; Appendix I gives every density, seed, grid, sensitivity setting, and result. Spectral-pilot and temporal-dependence guarantees are developed analytically above.

All cubic families share the fitting window, $R = 199$ permutations, and level $\alpha_0 = .05$. Full-data reference scans use 199 independent known-background null sequences per calibration setting. Table 1 reports detection rates for both calibration settings. Figure 1 measures the selected jump $\langle D, \widehat{V}\rangle/\kappa$ and total localization loss, assigning $n/2$ to a missed detection.

Sparse basis matches the few active coefficients of these models. At $d = 16$, its quadratic–linear and triple detection rates are 100% and 77.5%, compared with 20% and 7.5% for unfolding SVD. Marginal directions have zero population jump on both interactions. Figure 1 relates these structural choices to the selected signal and localization loss.

Broader adaptation changes both selection and calibration: the adaptive library raises mixed-signal detection from 82.5% to 90%, while triple detection changes from 77.5% to 70%. The full-data max basis scan detects all four $d = 16$ signals in all 40 trials, using both the complete sample and known-background reference data. Its median forced-position errors are 7, 40.5, 110.5, and 79. Appendix I reports every null and sensitivity evaluation with uncertainty intervals.

Table 1: Detection rates (%) at nominal level .05, $n = 12000$, $d = 16$, with 40 independent repetitions per column. Every non-dagger row uses the three-fold permutation protocol. Dagger rows use all observations and 199 independent reference sequences from the known null background.

Method	Marginal	Quadratic–linear	Triple	Mixed
Adaptive library	100.0	100.0	70.0	90.0
Sparse basis	100.0	100.0	77.5	82.5
Marginal cubic	100.0	2.5	5.0	85.0
Unfolding SVD	97.5	20.0	7.5	2.5
Symmetric Tucker	100.0	22.5	7.5	7.5
Full direction	50.0	7.5	17.5	5.0
Degree-two direction	5.0	7.5	10.0	2.5
RBF random features	5.0	7.5	12.5	5.0
Max basis scan†	100.0	100.0	100.0	100.0
Full kernel scan†	100.0	62.5	42.5	60.0

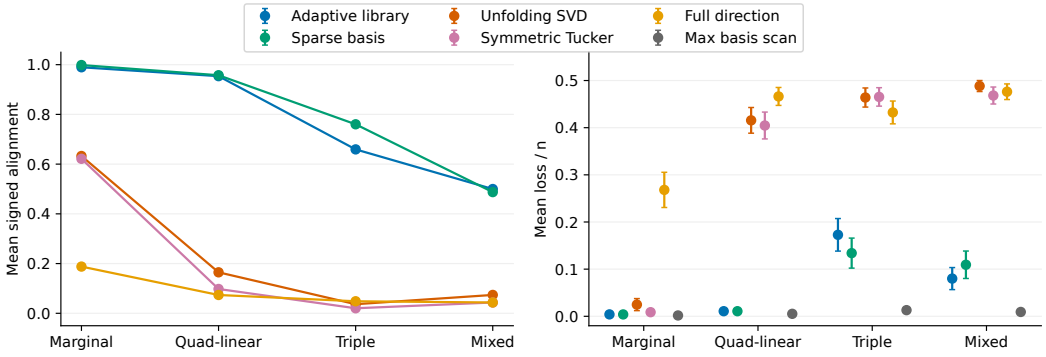

Figure 1: Retained signal and localization at $n = 12000$, $d = 16$. Left: mean normalized selected jump $\langle D, \widehat{V} \rangle / \kappa$, averaged over trials and cyclic orderings. Right: mean total localization loss divided by n, with one standard error over 40 trials. The reference max basis scan uses known-background calibration and has no fitted direction.

7 Discussion

A cubic change can be weak in every coefficient and strong in a singular direction. Spectral tails quantify the signal lost by compression, while covariance bounds quantify the noise incurred by learning that direction. Their joint control supplies an initial window and a retained jump. For independent observations, validation and anchor conditions carry this jump into calibrated testing and scalar localization. Under dependence, covariance control enlarges the proved detection and learning regime in the dense construction; thinning separately supplies a joint-law guarantee for refinement. Calibration and refinement using nearby dependent observations remain a further inferential question. The distinction identifies where additional dependence analysis can improve temporal resolution.

AI USE STATEMENT

AI tools assisted with wording and proofreading of the manuscript.

REFERENCES

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832, 2014. https://www.jmlr.org/papers/v15/anandkumar14b.html.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019. https://jmlr.org/papers/v20/16-155.html.

Heejong Bong and Arun Kumar Kuchibhotla. Tight concentration inequality for sub-Weibull random variables with generalized Bernstein Orlicz norms. *arXiv:2302.03850*, 2023. https://arxiv.org/abs/2302.03850.

Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. Kernel change-point detection with auxiliary deep generative models. *International Conference on Learning Representations*, 2019. https://arxiv.org/abs/1901.06077.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000. https://doi.org/10.1137/S0895479896305696.

Damien Garreau and Sylvain Arlot. Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2):4440–4486, 2018. https://doi.org/10.1214/18-EJS1513.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012. https://www.jmlr.org/papers/v13/gretton12a.html.

Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *TEST*, 27(4):811–825, 2018. https://doi.org/10.1007/s11749-017-0571-1.

Solt Kovács, Peter Bühlmann, Housen Li, and Axel Munk. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023. https://doi.org/10.1093/biomet/asac052.

Arun Kumar Kuchibhotla and Abhishek Chakrabortty. Moving beyond sub-Gaussianity in high-dimensional statistics: applications in covariance estimation and linear regression. *Information and Inference*, 11(4):1389–1456, 2022. https://doi.org/10.1093/imaiai/iaac012.

Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. https://doi.org/10.4007/annals.2010.171.295.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20, 2007. https://papers.nips.cc/paper_files/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html.

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

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

Anru Zhang and Dong Xia. Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, 2018. https://arxiv.org/abs/1703.02724.

Yuqi Zhang, Zetai Cen, and Haeran Cho. Detection and mode-identification of multiple change points in tensor factor models. *arXiv:2604.11300*, 2026. https://arxiv.org/abs/2604.11300.

H. C. P. Berbee. *Random walks with stationary increments and renewal theory*. Mathematical Centre Tracts, 112, 1979. https://ir.cwi.nl/pub/12694.

Xiaokai Luo, Haotian Xu, Carlos Misael Madrid Padilla, and Oscar Hernan Madrid Padilla. Online change point detection for multivariate inhomogeneous Poisson processes time series. *arXiv:2601.20192v2*, 2026. https://arxiv.org/abs/2601.20192v2.

Haotian Xu, Stéphane Guerrier, Runze Li, and Yuan Ke. On tail-robust autocovariance matrix estimation for high-dimensional and potentially nonstationary time series. *arXiv:2609.04418v1*, 2026. https://arxiv.org/abs/2609.04418v1.

Marwa Banna, Florence Merlevède, and Pierre Youssef. Bernstein-type inequality for a class of dependent random matrices. *Random Matrices: Theory and Applications*, 5(2):1650006, 2016. https://arxiv.org/abs/1504.05834.

Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53:122206, 2012. https://arxiv.org/abs/1208.0161.

A ALGORITHM AND COMPLETE INDEPENDENT-SAMPLE RESULTS

The following specification gives the precise three-fold procedure and the complete approximation and inference statements used in Sections 2–4. The power and localization results assume independent observations with bounded marginal densities. Null calibration holds for any common iid law. For dependent data, Section 5 supplies separate estimation bounds and the conditions for a whole-data-law transfer.

A.1 REPRESENTATION AND INDEXED ALGORITHM

Let independent observations $X_1, \dots, X_n \in [-1,1]^d$ have density f_0 up to the sample count η and density f_1 thereafter, relative to product-uniform probability measure μ. The power and localization analysis assumes $f_0, f_1 \le L$. Null calibration requires only a common distribution. Write P_3 for projection onto polynomials of total degree at most three.

A.1.1 A SIGNAL-PRESERVING REPRESENTATION

Use the orthonormal Legendre polynomials

$$\phi_0 = 1, \quad \phi_1(x) = \sqrt{3}x, \quad \phi_2(x) = \tfrac{\sqrt{5}}{2}(3x^2-1), \quad \phi_3(x) = \tfrac{\sqrt{7}}{2}(5x^3-3x).$$

For $|\alpha| \le 3$, let $\psi_\alpha(x) = \prod_j \phi_{\alpha_j}(x_j)$, $p = d+1$, and $q = \binom{d+3}{3}$. Associate α with the symmetric tensor E_α supported on the permutations of three indices containing $3-|\alpha|$ zeros and α_j copies of j. Normalize its nonzero entries by the square root of the number of distinct permutations. Define

$$H_3(x) = \sum_{|\alpha|\le 3} \psi_\alpha(x) E_\alpha, \quad A(f) = \mathbb{E}_f H_3(X), \quad D = A(f_1) - A(f_0), \quad \kappa = \|D\|_{\mathrm{F}}. \tag{23}$$

Orthonormality identifies both the signal energy and its finite kernel:

$$\|D\|_{\mathrm{F}} = \|P_3(f_1 - f_0)\|_{L^2(\mu)}, \qquad K_3(x,y) := \langle H_3(x), H_3(y)\rangle = \sum_{|\alpha|\le 3} \psi_\alpha(x)\psi_\alpha(y). \tag{24}$$

Thus κ measures the density change visible to cubic features. We remove the deterministic constant component, using $\bar{H}_3 = H_3 - E_0$, and form the right-minus-left CUSUM

$$\widehat{C}^t_{s,e} = \sqrt{\frac{(t-s)(e-t)}{e-s}} \left\{ \frac{1}{e-t} \sum_{i=t+1}^{e} \bar{H}_3(X_i) - \frac{1}{t-s} \sum_{i=s+1}^{t} \bar{H}_3(X_i) \right\}. \tag{25}$$

Its expectation on a single-change interval is $a(t)D$, with $a(t) > 0$ given in Appendix D.3. Projection onto a unit Frobenius direction V produces the scalar jump $\langle D, V\rangle$. The purpose of compression is to retain this jump accurately.

A.1.2 LEARNING, SELECTING, AND USING A DIRECTION

For a local pilot contrast Y, fit a finite library B_j, $1 \le j \le J$, and normalize $V_j = B_j/\|B_j\|_{\mathrm{F}}$, setting $V_j = 0$ if $B_j = 0$. Predetermined sparsity and rank grids specify four families. Basis truncation keeps the s largest normalized coefficients, with a marginal variant restricted to E_{3e_i}. Unfolding SVD truncates $Y_{(1)} \in \mathbb{R}^{p\times p^2}$ to rank r, folds, and symmetrizes; r refers to the matrix before symmetrization. Tucker applies the leading left-singular projector P in all three modes. The full candidate keeps Y. These choices retain different signal structures, compared in Section 3.

Take $n = 3N$, folds $X_j^{(b)} = X_{3(j-1)+b+1}$, $b \in \{0,1,2\}$, and $h \ge 16$ divisible by eight. Their true split counts are $\tau_b = \lfloor (\eta + 2 - b)/3 \rfloor$. Each cyclic ordering (A, B, C) assigns three independent data roles:

1. On A, maximize a fixed score of the full-interval CUSUM over $c \in [h, N-h]$ and fit directions from the contrast on $(c-h, c+h]$ at c. The score is either the largest absolute coefficient or the normalized spectral score in (41).

2. On B, compute $T_j = \langle \widehat{C}^c_{c-h,c+h}, V_j \rangle$ and choose $\widehat{j} \in \arg\max_j |T_j|$, with $\widehat{V} = \operatorname{sgn}(T_{\widehat{j}})V_{\widehat{j}}$. Candidates remain fixed. Use deterministic ties and $\operatorname{sgn}(0) = 1$.
3. On C, project onto $\widehat{V}$. Estimate the two means from anchors $(c-h, c-h/2]$ and $(c+h/2, c+h]$, then fit the split on the disjoint central observations.

To calibrate the choice on B, draw R independent uniform permutations of its $2h$ observations and recompute the entire maximum $T^{(r)} = \max_j |T_j^{(r)}|$. Set

$$p_b = \frac{1 + \sum_{r=1}^{R} \mathbf{1}\{T^{(r)} \geq \max_j |T_j|\}}{R+1}, \qquad p_{\text{global}} = \min(1, 3 \min_b p_b). \tag{26}$$

For refinement, set $Z_j = \langle \bar{H}_3(X_j^{(C)}), \widehat{V} \rangle$ and minimize the anchored scalar loss

$$Q(t) = \sum_{j=c-h/2+1}^{t} (Z_j - \widehat{\mu}_L)^2 + \sum_{j=t+1}^{c+h/2} (Z_j - \widehat{\mu}_R)^2, \quad t \in \{c-h/2, \ldots, c+h/2\}. \tag{27}$$

Map the three estimates to $3\widehat{t}_b$ and take their median $\widetilde{\eta}$. Report a change when $p_{\text{global}} \leq \alpha_0$; compute the forced estimate on every sequence for evaluation. Each ordering uses its own pilot and validation folds. Its disjoint anchors supply the independence needed for refinement.

A.2 Complete structural approximation results

We analyze the local contrast $Y = C + E$, where $C = aD$ under one change. The argument separates deterministic approximation from stochastic estimation: the former identifies which structure preserves the signal, and the latter determines the observations needed to learn it. Throughout, $\tau_r(T)^2 = \sum_{j>r} \sigma_j(T_{(1)})^2$ denotes the unfolding spectral tail.

A.2.1 A gap-free approximation bound

Set $\delta = \|E_{(1)}\|_{\text{op}}$, $\epsilon = \max_\alpha |\langle E, E_\alpha \rangle|$, and $\tau_r = \tau_r(C)$. For $1 \leq r \leq p$ and $1 \leq s \leq q$, let $t_s = t_s(C)$ be the best s-coefficient approximation error over the complete normalized basis. The following result compares three ways of compressing the same contrast.

Theorem A.1 (Deterministic approximation bounds)**.** *For basis truncation over the full coefficient set and the two spectral estimators in Section 2,*

$$\|B_{\text{basis}} - C\|_{\text{F}} \leq t_s + 2\sqrt{2s}\epsilon, \tag{28}$$

$$\|B_{\text{svd}} - C\|_{\text{F}} \leq \tau_r + 2\sqrt{2r}\delta, \tag{29}$$

$$\|B_{\text{tuck}} - C\|_{\text{F}}^2 \leq 3(\tau_r + 2\sqrt{2r}\delta)^2 + r\delta^2. \tag{30}$$

The bounds hold for every symmetric C, including tied singular values. If $C = aD$, $a > 0$, and $\|B - C\|_{\text{F}} \leq e < a\kappa$, then

$$\langle D, B/\|B\|_{\text{F}} \rangle \geq \kappa\sqrt{1 - (e/(a\kappa))^2}. \tag{31}$$

The SVD bound follows from an optimality inequality that also explains its dependence on rank. Write $A = C_{(1)}$, $Z = E_{(1)}$, and let M_r and A_r be best rank-r approximations of $A + Z$ and A. With $u = \|M_r - A\|_{\text{F}}$ and $t = \|A_r - A\|_{\text{F}}$, optimality gives

$$u^2 - t^2 \leq 2\langle Z, M_r - A_r \rangle \leq 2\sqrt{2r}\,\delta\,(u + t). \tag{32}$$

The difference of the two fits has rank at most $2r$, which bounds its nuclear norm by $\sqrt{2r}$ times its Frobenius norm. Dividing by $u + t$ yields the error bound. Folding preserves the Frobenius norm, and symmetrization contracts distance to symmetric C. The basis argument uses a support of size at most $2s$ instead of matrix rank; Tucker additionally accounts for signal lost by projection in all three modes. Appendix C gives the complete proofs and the marginal-library variant, whose approximation error includes omitted interactions.

The angle bound is the crucial link to inference. If $K = a\kappa$ and $z = \|B\|_{\mathrm{F}}$, expanding the squared error gives $\langle C, B/z\rangle \geq (K^2 + z^2 - e^2)/(2z) \geq \sqrt{K^2 - e^2}$. Thus a small relative fitting error guarantees a large population jump even though the fitted norm is random. In particular, with $\ell = \delta/a$,

$$\tau_r(D) + 2\sqrt{2r}\ell \leq \rho\kappa \quad \Longrightarrow \quad \langle D, V_r\rangle \geq \kappa\sqrt{1-\rho^2}, \qquad 0 < \rho < 1. \tag{33}$$

This implication gives rank truncation its statistical role: the discarded signal and fitting noise must together leave a useful directional change. The structure of the truncation matters. For $D = E_{e_i+e_j+e_k}$ and a tied leading subspace spanned by $r \in \{1, 2\}$ of these coordinate axes, Tucker gives zero, while symmetrized unfolding gives $(r/3)D$. Proposition C.1 proves this exact interaction example and specifies the subspace choice.

A.2.2 Eigenvalue energy and adaptive rank choice

Let $K_D = D_{(1)}D_{(1)}^{\top}$ have eigenvalues $\mu_1 \geq \cdots \geq \mu_p \geq 0$. Then $\sum_j \mu_j = \kappa^2$ and $\tau_r(D)^2 = \sum_{j>r} \mu_j$. These identities express the error bound entirely through a positive spectrum. If $D = \sum_j \lambda_j v_j^{\otimes 3}$ with orthonormal component directions, its unfolding has singular values $|\lambda_j|$ and $\mu_j = \lambda_j^2$.

Corollary A.2 (Spectral energy controls retained signal)**.** *Suppose $\kappa > 0$, $Y = aD + E$, $a \geq a_0 > 0$, and $\|E_{(1)}\|_{\mathrm{op}} \leq U$. Put $\ell = U/a_0 > 0$ and*

$$\mathcal{E}_\ell(D) = \sum_{j=1}^{p} \min\{\mu_j, 8\ell^2\}, \qquad r_\ell = \#\{j : \mu_j > 8\ell^2\}. \tag{34}$$

If the library contains every rank from 1 to p and $2\mathcal{E}_\ell(D) \leq \rho^2\kappa^2$ for $0 < \rho < 1$, some fitted SVD direction has population jump at least $\kappa\sqrt{1-\rho^2}$. For the dyadic grid $\{1, 2, 4, \ldots\} \cap [1, p]$, augmented by p, the sufficient condition is $4\mathcal{E}_\ell(D) \leq \rho^2\kappa^2$.

The criterion balances signal energy against estimation cost one eigenvalue at a time. At the threshold rank, $\mathcal{E}_\ell(D) = \sum_{j>r_\ell} \mu_j + 8r_\ell\ell^2$. Squaring the SVD error bound costs a factor two; rounding upward to a dyadic rank costs at most another factor two. This proves the corollary and explains why a logarithmic-size grid suffices to contain a useful direction. Validation selects among fitted candidates; the population spectrum establishes their existence.

The same argument gives explicit rates beyond exact low rank. For an interior rank and a polynomial envelope $\sqrt{\mu_j} \leq Aj^{-s}$, $s > 1/2$, the error obeys

$$\frac{\|B_r - aD\|_{\mathrm{F}}}{a} \leq \frac{A}{\sqrt{2s-1}} r^{1/2-s} + \sqrt{8}\ell\sqrt{r}, \qquad r \asymp (A/\ell)^{1/s}. \tag{35}$$

Balancing the terms gives $C_s A^{1/(2s)}\ell^{1-1/(2s)}$. For exponential decay $\sqrt{\mu_j} \leq Ae^{-b(j-1)}$, a rank of order $1 + b^{-1}\log_+(H/\ell)$ gives error at most $5\ell\sqrt{1 + b^{-1}\log_+(H/\ell)}$, where $H = A/\sqrt{1-e^{-2b}}$. Appendix C.4 handles rounding, zero noise, and the strengthened full-rank endpoint $\sqrt{p}\ell$. An informative normalized direction follows whenever the available error bound is below κ.

Clipped energy also identifies the relevant effective dimension. With $t = 8\ell^2$ and $d_{\mathrm{eff}}(t) = \sum_j \mu_j/(\mu_j + t)$,

$$t\, d_{\mathrm{eff}}(t) \leq \mathcal{E}_\ell(D) \leq 2t\, d_{\mathrm{eff}}(t). \tag{36}$$

The spectrum at the noise scale therefore controls the compression requirement. Appendix C contrasts this quantity with stable rank, which can remain small while a substantial fraction of the signal occupies many weak components.

A.2.3 Tensor eigenvalues and injective noise

Orthogonal tensor structure supports a different approximation route. Define $\|Z\|_{\mathrm{inj}} = \sup_{\|u\|_2=\|v\|_2=\|w\|_2=1} |Z[u, v, w]|$. The following bound uses this tensor norm directly and the eigenvalue tail of an orthogonal decomposition.

Theorem A.3 (Tensor eigenvalue tails with injective noise)**.** *Suppose* $D = \sum_{j=1}^{R_0} \lambda_j v_j^{\otimes 3}$ *with orthonormal* v_j *and decreasing* $|\lambda_j|$*. Let* $Y/a = D + Z$ *and* $\omega = \|Z\|_{\mathrm{inj}}$*. For* $1 \le r \le p$*, a global least-squares fit* $\widehat{T}_r$ *over tensors with at most* r *orthogonal components satisfies*

$$\|\widehat{T}_r - D\|_{\mathrm{F}} \le \left(\sum_{j>r} \lambda_j^2\right)^{1/2} + 4r\omega. \tag{37}$$

If the true leading $\min(r, R_0)$ *directions are supplied and the frame is completed to size* r*, projecting onto their cubic tensors instead gives squared error at most* $\sum_{j>r} \lambda_j^2 + r\omega^2$*.*

The proof compares the fitted tensor with the true rank-r truncation. Their component spans lie in a subspace of dimension at most $2r$. Restricting Z to that subspace gives Frobenius norm at most $2r\omega$, and the same optimality factorization as in (32) yields (37). With the true frame supplied, orthogonality instead separates residual signal from projected noise. These are guarantees for different estimation settings, each with its stated optimization premise.

For a balanced independent local contrast with population amplitude $a \asymp \sqrt{h}$, Appendix B.2 gives, up to constants and the logarithmic confidence term x,

$$\ell \lesssim \sqrt{\frac{Lqx}{ph}} + \frac{\sqrt{q}\,x}{h}, \qquad \omega \lesssim \nu_L \left\{ \sqrt{\frac{p+x}{h}} + \frac{(p+x)^{3/2}}{h} \right\}, \quad \nu_L = C\max(1, \sqrt{L}). \tag{38}$$

Thus the leading stochastic upper bounds scale as $\sqrt{r}\,p/\sqrt{h}$ for unfolding SVD and $r\sqrt{p/h}$ for the global orthogonal fit, with confidence, density, and envelope terms retained above. The comparison shows how stronger tensor structure changes the available error bound. The implemented SVD is a matrix approximation; the orthogonal result concerns a global tensor optimization. Each bound connects to the same retained-jump inequality (31).

A.3 Complete independent-sample inference conditions

The spectral analysis supplies a good candidate. The next step is to obtain it from a valid pilot window and preserve its jump through selection. We make the corresponding stochastic scales and inferential conditions explicit.

A.3.1 Uniform noise for the pilot and fitted directions

The entries of $H_3(X)$ share the same observation and are correlated. Orthonormality of the polynomial coefficients nonetheless permits an exact second-moment calculation.

Lemma A.4 (Symmetric-basis matrix variance)**.** *Let* $M(x) = H_3(x)_{(1)}$*,* $q = \binom{p+2}{3}$*, and let* P_2 *be the orthogonal projector onto symmetric order-two tensors in* $\mathbb{R}^{p^2}$*. Under the reference measure,*

$$\mathbb{E}_\mu M M^\top = (q/p) I_p, \qquad \mathbb{E}_\mu M^\top M = ((p+2)/3) P_2. \tag{39}$$

If a marginal density is at most L*, both centered second-moment matrices have operator norm at most* Lq/p*.*

The identity follows by taking partial traces of the orthogonal projector onto symmetric order-three tensors. It provides the variance term for matrix Bernstein, while $\sup_x \|H_3(x)\|_{\mathrm{F}} \le 3\sqrt{3q}$ gives the summand envelope. Applying the resulting bounds simultaneously over all pilot splits and local centers makes them valid after a data-dependent window choice.

Let $\nu_L = C\max(1, \sqrt{L})$, $\delta_P, \delta_V, \delta_A, \delta_C \in (0,1)$, $b_0 = 3\sqrt{3}$, $M = 6(N+1)$, and

$$\epsilon_P = b_0\sqrt{2\log(4qM/\delta_P)}, \quad x_P = \log\{2M(p+p^2)/\delta_P\},$$
$$U_P = \sqrt{2L(q/p)x_P} + \frac{4b_0\sqrt{q}}{3\sqrt{h}}x_P, \quad a_0 = \frac{15\sqrt{h}}{16\sqrt{2}}, \quad \ell_P = U_P/a_0. \tag{40}$$

Lemma A.4 gives the exact basis variance factor $q/p = (p+1)(p+2)/6$. The bounds cover every candidate pilot split and local contrast simultaneously, so they apply to the fitted window selected from the data.

The coefficient pilot has signal $\gamma_\infty = \max_\alpha |\langle D, E_\alpha\rangle|$ and uniform noise at most ϵ_P. A spectral alternative uses predetermined nonempty grids $\mathcal{S}, \mathcal{R}$ and fixed normalizers:

$$\Psi_*(T) = \max\left\{ \max_{s\in\mathcal{S}} \frac{\left(\sum_{j\le s} |T|_{(j)}^2\right)^{1/2}}{\sqrt{s}\epsilon_P}, \quad \max_{r\in\mathcal{R}} \frac{\left(\sum_{j\le r} \sigma_j(T_{(1)})^2\right)^{1/2}}{\sqrt{r}U_P} \right\}. \tag{41}$$

Here $|T|_{(j)}$ orders the absolute normalized coefficients. Triangle inequalities give uniform score noise at most one and population signal $\gamma_* := \Psi_*(D)$. If both grids contain one, the score is the maximum of the normalized coefficient infinity norm and unfolding operator norm. Higher ranks then enrich the fitted library, while the pilot responds to a prominent coefficient or leading singular value.

A.3.2 A COMPLETE FINITE-SAMPLE GUARANTEE

Theorem A.5 (Complete guarantee from structural and pilot conditions)**.** *Let $\varnothing \neq \mathcal{S} \subseteq \{1,\dots,q\}$ and $\varnothing \neq \mathcal{R} \subseteq \{1,\dots,p\}$. Suppose $n = 3N$, $h \ge 16$ is divisible by eight, and each true segment in each fold has length at least $2h$. Set $\beta = \min_b \min(\tau_b, N - \tau_b)/N$. Use either the coefficient pilot with $(\gamma, e) = (\gamma_\infty, \epsilon_P)$ or the pilot in* (41) *with $(\gamma, e) = (\gamma_*, 1)$. Suppose*

$$\begin{aligned} &\beta h\gamma \ge 32\sqrt{N}e, \\ &\min\left\{ \min_{s\in\mathcal{S}} \left[t_s(D) + \frac{2\sqrt{2s}\epsilon_P}{a_0} \right], \min_{r\in\mathcal{R}} \left[\tau_r(D) + 2\sqrt{2r}\ell_P \right] \right\} \le \rho\kappa, \end{aligned} \tag{42}$$

where $0 < \rho < 1$ and the indicated candidates belong to the fitted library. Put $g = \kappa\sqrt{1-\rho^2}$, $x_V = \log\{12J(R+1)/\delta_V\}$, $x_A = \log(12/\delta_A)$, and $x_C = \log(6/\delta_C)$. If $3/(R+1) \le \alpha_0$ and

$$\begin{aligned} \sqrt{h}\, g &\ge 8C\nu_L(\sqrt{x_V} + x_V^{3/2}/\sqrt{h}) + 4\kappa\sqrt{x_V}, \\ C\nu_L(\sqrt{2x_A/h} + 2x_A^{3/2}/h) &\le g/16, \end{aligned} \tag{43}$$

then rejection has probability at least $1 - \delta_P - \delta_V$. Rejection together with

$$|\widetilde{\eta} - \eta| \le \mathfrak{r}(g) := C\{\nu_L^2 x_C/g^2 + \nu_L x_C^{3/2}/g + 1\} \tag{44}$$

has probability at least $1 - \delta_P - \delta_V - \delta_A - \delta_C$. Under any iid null, either pilot gives $\mathbb{P}(p_{\mathrm{global}} \le \alpha_0) \le \alpha_0$ without the signal conditions.

The energy criterion can supply both the pilot and direction conditions, giving a single spectral assumption for these two stages.

Corollary A.6 (A common spectral condition for fitting and inference)**.** *Retain the model, segment-length, validation, anchor, and permutation-resolution assumptions of Theorem A.5. Use Ψ_* with $1 \in \mathcal{S}$ and a rank grid containing every dyadic rank and p, and include their SVD directions. For $\kappa > 0$ and $0 < \rho < 1$, set*

$$\xi = \min\left\{ \frac{\rho^2}{4}, \frac{4\beta^2 h}{225N} \right\}. \tag{45}$$

If $\mathcal{E}_{\ell_P}(D) \le \xi\kappa^2$, the rejection and localization guarantees of Theorem A.5 hold with $g = \kappa\sqrt{1-\rho^2}$.

To see the additional pilot implication, put $t = 8\ell_P^2$. Small clipped energy forces $\mu_1 > t$ and $\mathcal{E}_{\ell_P}(D) \ge t\kappa^2/\mu_1$. Hence the leading singular value is large enough for the spectral pilot, while the same energy condition supplies an informative fitted rank. Appendix D.7 verifies the constants. At fixed $\rho < 1$ and confidence, with bounded L, the resulting leading small-signal localization term is proportional to κ^{-2}.

A.3.3 WHY THE GUARANTEE HOLDS

The proof has three steps, each corresponding to a data role in Section 2. First, the population CUSUM has a deterministic margin. For a score with population signal γ and uniform error e, $|c - \tau| \le 2\sqrt{N}e/(\beta\gamma)$. The first condition places the pilot center within $h/16$ of the change. On

the same uniform event, its local amplitude is at least a_0, and the approximation condition yields a candidate with jump at least g.

Second, conditional on that pilot, validation compares fixed directions. If its population contrast is $a_B D$ and every directional error is at most b, selection and orientation give

$$\langle D, \widehat{V} \rangle \geq \max_{j \leq J} |\langle D, V_j \rangle| - 2b/a_B. \tag{46}$$

Polynomial moments bound b by $C\nu_L(\sqrt{x} + x^{3/2}/\sqrt{h})$ with x logarithmic in J and the failure probability. Condition (43) ensures a selected jump of at least $g/2$ and separates the observed maximum from the permuted maxima. Under the null, exchangeability calibrates that entire maximum, including the selection step.

Third, conditional on this ordering's pilot and validation data, its anchors and central observations remain independent. Write $d_\mu = \langle D, \widehat{V} \rangle \geq g/2$. Accurate anchors give the RSS contrast a drift of order $v d_\mu^2$ at distance v from the change. Centered partial sums contribute at scale $d_\mu \nu_L(\sqrt{vx} + x^{3/2})$. Dividing by d_μ and using its lower bound yields the two terms in (44). The proof integrates each ordering's conditional error before combining the three bounds. Appendix D supplies the complete argument, including permutation balance, dyadic distance shells, and index conversion.

Two lower-bound constructions place these rates in context. A pure-cubic coefficient mixture has detection energy barrier $m\theta^2 \lesssim \log \binom{d+2}{3}$ for m affected observations. A two-location subclass gives error at least a constant times κ^{-2} with fixed positive probability when the horizon accommodates that separation. Appendix F.1 states the models and losses precisely. The latter has the same small-signal jump exponent as (44) when g is comparable to κ and the pilot and anchor conditions hold.

B Representation and probability tools

This appendix establishes the signal and noise quantities used by the method. The isometry identifies the cubic density change with D; polynomial moments control projected observations; matrix second moments give the unfolding noise scale. Appendix C uses these quantities for approximation, and Appendix D carries the resulting signal into inference. Unspecified constants are universal and may be enlarged between displays.

B.1 Representation and linear-time identities

Distinct multiindices have disjoint permutation supports. Each tensor E_α has squared Frobenius norm $m_\alpha / m_\alpha = 1$. There are $\binom{d+3}{3} = \dim \mathrm{Sym}^3(\mathbb{R}^{d+1})$ such tensors, so they form an orthonormal basis. The coefficient of E_α in $A(f)$ is $\int \psi_\alpha f \, d\mu$. Parseval's identity on the polynomial subspace proves (24). More generally, for every symmetric V,

$$\int \langle H_3(x), V \rangle^2 \, d\mu(x) = \|V\|_{\mathrm{F}}^2. \tag{47}$$

Consequently, a fixed unit direction satisfies $\mathrm{Var}_f(\langle H_3(X), V \rangle) \leq L$ when $f \leq L$. The following moment argument upgrades this variance control to the tail bounds used for inference.

For an explicit kernel identity, put $a_j = \phi_1(x_j)\phi_1(y_j)$, $b_j = \phi_2(x_j)\phi_2(y_j)$, $c_j = \phi_3(x_j)\phi_3(y_j)$, and $s_\ell = \sum_j a_j^\ell$. Splitting the multiindices by degree and multiplicity gives

$$\begin{aligned} K_3(x, y) = 1 + s_1 + \sum_j b_j + \tfrac{1}{2}(s_1^2 - s_2) + \sum_j c_j \\ + s_1 \sum_j b_j - \sum_j a_j b_j + \tfrac{1}{6}(s_1^3 - 3 s_1 s_2 + 2 s_3). \end{aligned} \tag{48}$$

Here the quadratic and cubic elementary symmetric polynomials collect the terms using distinct coordinates. Formula (48) costs $O(d)$ operations for one pair (x, y). Removing the constant feature replaces K_3 by $K_3 - 1$. A kernel implementation additionally accounts for the observation pairs it processes.

An analogous identity evaluates cubic contractions. For $u = (a, b_1, \dots, b_d)$, set $y_j = b_j\phi_1(x_j)$ and $s_\ell = \sum_j y_j^\ell$. Then

$$H_3(x)[u,u,u] = a^3 + \sqrt{3}a^2 s_1 + \sqrt{3}a\sum_j b_j^2\phi_2(x_j) + \tfrac{\sqrt{6}}{2}a(s_1^2 - s_2) + \sum_j b_j^3\phi_3(x_j) + \sqrt{3}\sum_j b_j^2\phi_2(x_j)(s_1 - y_j) + \tfrac{\sqrt{6}}{6}(s_1^3 - 3s_1s_2 + 2s_3). \quad (49)$$

The factors $\sqrt{3}$ and $\sqrt{6}$ are the square roots of the corresponding permutation counts. Each displayed sum is linear in d. The explicit SVD and Tucker implementations have the storage and decomposition costs given in Appendix I.

B.2 Polynomial moments and three distinct noise norms

For $a > 0$, use $\|W\|_{\psi_a} = \inf\{c > 0 : \mathbb{E}\exp(|W/c|^a) \le 2\}$. The following argument establishes the moment hypothesis used throughout the appendix under independent observations and the bounded-density assumption. Tensorization is applied under the product reference measure μ, after which the density bound transfers the result to f.

On the reference interval,

$$15 - \sum_{j=1}^{3}\phi_j(x)^2 = \tfrac{5}{4}(1-x^2)(35x^4 + 2x^2 + 11) \ge 0.$$

Every centered linear combination G of these three univariate basis functions therefore satisfies $\|G\|_{L^u} \le \sqrt{15}\|G\|_{L^2}$ for $u \ge 2$. Proposition 3.16 of Mossel et al. (2010) provides the common hypercontractive parameter $(2\sqrt{15}\sqrt{u-1})^{-1}$. Their Propositions 3.11 and 3.12 tensorize the independent coordinate ensembles under μ and bound a polynomial using at most three ensembles. Every polynomial Q of total degree at most three thus satisfies

$$\|Q\|_{L^u(\mu)} \le Cu^{3/2}\|Q\|_{L^2(\mu)}, \qquad u \ge 2.$$

For $Q = \langle H_3, V\rangle$, equation (47) and $f \le L$ give $\|Q\|_{L^u(f\,d\mu)} \le L^{1/u}Cu^{3/2}\|V\|_{\mathrm{F}}$. Subtracting the mean costs at most a factor two. The moment characterization of the Orlicz norm consequently yields

$$\|\langle H_3(X) - A(f), V\rangle\|_{\psi_{2/3}} \le \nu_L\|V\|_{\mathrm{F}}, \qquad \nu_L = C\max(1, \sqrt{L}). \quad (50)$$

This also holds conditionally for a direction independent of the observation.

We use the following consequence of the independent weighted sub-Weibull inequality (Kuchibhotla and Chakrabortty, 2022, Theorem 3.1); the same weight dependence is displayed by Bong and Kuchibhotla (2023). For independent centered W_i with $\|W_i\|_{\psi_{2/3}} \le \nu$, deterministic real weights w_i, and $x \ge 1$,

$$\mathbb{P}\left\{\left|\sum_i w_iW_i\right| > C\nu\big(\sqrt{x}\|w\|_2 + x^{3/2}\|w\|_\infty\big)\right\} \le 2e^{-x}. \quad (51)$$

The variables may have different distributions. For balanced validation weights, $\|w\|_2 = 1$ and $\|w\|_\infty = (2h)^{-1/2}$, giving (63). Conditional on one pilot, apply this inequality to its J fixed directions. Taking $x = \log(6J/\delta_V)$ and integrating before a union over three orderings proves the validation bound in Appendix D.

For clarity, let $E_w = \sum_i w_i\{H_3(X_i) - \mathbb{E}H_3(X_i)\}$. The following bounds control different quantities.

Lemma B.1 (Coordinate, unfolding, and injective errors)**.** *Let $\zeta \in (0,1)$, $b_0 = 3\sqrt{3}$, and $q = \binom{d+3}{3}$. For deterministic weights, each of the following statements holds with probability at least $1 - \zeta$:*

$$\max_\alpha |\langle E_w, E_\alpha\rangle| \le b_0\|w\|_2\sqrt{2\log(2q/\zeta)}, \quad (52)$$

$$\|(E_w)_{(1)}\|_{\mathrm{op}} \le \sqrt{2Lp^2\|w\|_2^2 x} + \frac{4b_0\sqrt{q}}{3}\|w\|_\infty x, \quad x = \log\frac{p+p^2}{\zeta}, \quad (53)$$

$$\|E_w\|_{\mathrm{inj}} \le C\nu_L\{\sqrt{\rho}\|w\|_2 + \rho^{3/2}\|w\|_\infty\}, \quad \rho = 3p\log 17 + \log(2/\zeta). \quad (54)$$

Here $\|T\|_{\mathrm{inj}} = \sup_{\|u\|=\|v\|=\|z\|=1}|T[u,v,z]|$. The coordinate bound needs only independence and bounded support. The other two bounds use $f_i \le L$.

Proof. The largest sup norm of a Legendre product of total degree at most three is at most $\sqrt{3^3} = b_0$. Weighted Hoeffding and a union bound over q coefficients give (52).

For the matrix bound put $M_i = (H_3(X_i))_{(1)}$. Orthogonality gives $\sup_x \|H_3(x)\|_{\mathrm{F}} \le b_0\sqrt{q}$. Thus the centered weighted matrices have operator norms at most $2b_0\sqrt{q}\|w\|_\infty$. For unit vectors $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^{p^2}$, the scalar $a^\top M_i b$ is the contraction against the symmetrization of a tensor of Frobenius norm one. Its second moment is at most L. Summing over a standard basis of $\mathbb{R}^{p^2}$ yields $\|\mathbb{E}M_i M_i^\top\|_{\mathrm{op}} \le Lp^2$; summing over a standard basis of $\mathbb{R}^p$ yields $\|\mathbb{E}M_i^\top M_i\|_{\mathrm{op}} \le Lp$. Centering can only decrease either second-moment matrix in positive-semidefinite order. The rectangular matrix Bernstein variance parameter is therefore at most $Lp^2\|w\|_2^2$. Theorem 1.6 of Tropp (2012), with tail parameter x, gives (53).

For the injective bound choose a $1/8$-net $\mathcal{N}$ of the unit sphere with cardinality at most 17^p. For each triple in $\mathcal{N}^3$, its contraction corresponds to $\mathrm{Sym}(u \otimes v \otimes z)$, of Frobenius norm at most one. Equations (50) and (51) and a union bound give the desired bound at every net triple. Approximating each of three unit vectors by its net point and telescoping gives an error at most $3\|E_w\|_{\mathrm{inj}}/8$. Hence the full injective norm is at most $8/5$ times the largest net contraction, absorbed in C. □

For M predetermined contrasts, replace ζ by ζ/M in a union bound. This is one way to justify an adaptively selected pilot contrast from that finite set. A mean has $\|w\|_2$ of order $N^{-1/2}$, whereas a normalized CUSUM has $\|w\|_2 = 1$. Lemma A.4 sharpens the unfolding variance factor to Lq/p. The unfolding and injective bounds enter the respective estimators in Theorems A.1 and A.3.

B.3 An exact variance calculation for the cubic tensor

The following proof establishes Lemma A.4.

Proof of Lemma A.4. Coefficient orthonormality gives $\mathbb{E}_\mu MM^\top = \sum_\alpha E_{\alpha,(1)}E_{\alpha,(1)}^\top$, and similarly for the other product. The orthogonal projector onto $\mathrm{Sym}^3(\mathbb{R}^p)$ is $P_3^{\mathrm{sym}} = \frac{1}{6}\sum_{\pi\in S_3} U_\pi$. Taking its partial trace over modes two and three gives the first sum. The identity permutation contributes $p^2 I$, the three transpositions together contribute $3pI$, and the two three-cycles contribute $2I$. Their sum divided by six is qI/p.

For the partial trace over mode one, the identity and the swap of modes two and three contribute $p(I + S)$, where S swaps those modes. The other four permutations contribute $2(I + S)$. Thus the second sum is $(p + 2)(I + S)/6 = ((p + 2)/3)P_2$. Integration against a density bounded by L preserves the corresponding positive-semidefinite upper bounds. Centering subtracts a positive-semidefinite matrix. Finally $(p + 2)/3 \le q/p$ for $p \ge 1$. This proves the common variance bound. It improves the constant in the p^2 scale, without changing that dimension order or the summand envelope. □

C Structural approximation and spectral energy

The proofs here support the compression criterion in Section 3. We first prove the approximation-to-jump bound and specify its coefficient classes, then exhibit the difference between a common Tucker subspace and symmetrized unfolding. Spectral energy and decay translate this bound into rank choices. The final argument uses orthogonal tensor eigenvalues and injective noise to supply a second error budget.

C.1 Approximation error and direction geometry

Proof of Theorem A.1. We first prove a matrix inequality. Write $A = C_{(1)}$, $Z = E_{(1)}$, and let A_r and M_r be best rank-at-most-r Frobenius approximations to A and $A + Z$, respectively. Put $u = \|M_r - A\|_{\mathrm{F}}$ and $t = \|A - A_r\|_{\mathrm{F}} = \tau_r$. Optimality of M_r yields

$$\|M_r - A - Z\|_{\mathrm{F}}^2 \le \|A_r - A - Z\|_{\mathrm{F}}^2, \qquad u^2 \le t^2 + 2\langle Z, M_r - A_r\rangle.$$

The difference $M_r - A_r$ has rank at most $2r$. Operator/nuclear norm duality and the triangle inequality therefore give

$$\langle Z, M_r - A_r \rangle \le \delta \|M_r - A_r\|_* \le \sqrt{2r}\delta \|M_r - A_r\|_{\mathrm{F}} \le \sqrt{2r}\delta(u+t).$$

Thus $(u-t)(u+t) \le 2\sqrt{2r}\delta(u+t)$. If $u+t=0$ the claim is immediate; otherwise division proves $u \le t + 2\sqrt{2r}\delta$. Folding is a Frobenius isometry, and symmetrization is the Frobenius orthogonal projection onto symmetric tensors. It fixes C, hence proves (29). No rank claim about the symmetrized output is needed.

For basis truncation, let θ and y be the coefficient vectors of C and Y. Their Euclidean distances equal the corresponding Frobenius distances. Repeat the preceding best-approximation argument with the best s-term vectors θ_s and y_s. Their difference has at most $2s$ nonzero coordinates, so

$$|\langle y - \theta, y_s - \theta_s \rangle| \le \epsilon \|y_s - \theta_s\|_1 \le \sqrt{2s}\epsilon \|y_s - \theta_s\|_2.$$

The same factorization proves (28). This argument concerns the complete normalized coefficient basis. For a restricted marginal library, the omitted non-marginal coefficients contribute an additional approximation error and cannot be discarded from t_s.

For Tucker truncation let P be the specified rank-r orthogonal projector and write $Q = P \otimes P \otimes P$. The matrix $P(A+Z)$ is a best rank-r approximation. Orthogonality of its row-space residual implies

$$a := \|(I-P)A\|_{\mathrm{F}} \le \|P(A+Z) - A\|_{\mathrm{F}} \le \tau_r + 2\sqrt{2r}\delta.$$

The three mode projections commute. Decomposing a tensor into the eight orthogonal blocks obtained from P and $I-P$ on its modes gives

$$\|(I-Q)C\|_{\mathrm{F}}^2 \le \sum_{k=1}^{3} \|C \times_k (I-P)\|_{\mathrm{F}}^2 = 3a^2.$$

The equality uses symmetry of C and the same projector on all three modes. Also

$$\|QE\|_{\mathrm{F}} = \|PZ(P \otimes P)\|_{\mathrm{F}} \le \sqrt{r}\delta.$$

The tensors $(I-Q)C$ and QE are orthogonal, so $\|QY - C\|_{\mathrm{F}}^2 = \|(I-Q)C\|_{\mathrm{F}}^2 + \|QE\|_{\mathrm{F}}^2$. This proves (30), including tied singular values for any chosen leading projector.

Finally set $K = a\kappa$ and $z = \|B\|_{\mathrm{F}}$. The assumption $e < K$ implies $z > 0$. Expanding $\|B - C\|_{\mathrm{F}}^2$ gives

$$\langle C, B/z \rangle \ge \frac{K^2 + z^2 - e^2}{2z} \ge \sqrt{K^2 - e^2},$$

where the last step follows from $z + (K^2 - e^2)/z \ge 2\sqrt{K^2 - e^2}$. Division by a proves (31). □

If $D = \sum_j \lambda_j v_j^{\otimes 3}$ with orthonormal v_j, its unfolding is $\sum_j \lambda_j v_j (v_j \otimes v_j)^\top$. Both sets of vectors are orthonormal, which proves the singular-value and tail identities used in the eigenvalue-energy condition (34). These identities use orthogonality of the component directions. For a deterministic bound U on unfolding noise, Weyl's inequality gives, on the event $\delta \le U$,

$$\tau_r(C) \le \left\{ \sum_{j>r} \big(\sigma_j(Y_{(1)}) + U\big)^2 \right\}^{1/2}.$$

For the marginal estimator, let $\mathcal{M} = \{3e_i : 1 \le i \le d\}$, $\theta_\alpha = \langle C, E_\alpha \rangle$, and let $|\theta^{\mathcal{M}}|_{(j)}$ order the absolute coefficients in $\mathcal{M}$. Its appropriate approximation error is

$$t_s^{\mathcal{M}}(C)^2 = \sum_{\alpha \notin \mathcal{M}} \theta_\alpha^2 + \sum_{j>s} |\theta^{\mathcal{M}}|_{(j)}^2, \qquad 1 \le s \le d. \tag{55}$$

The same optimality argument over this restricted coefficient set gives $\|B_{\mathrm{marginal}} - C\|_{\mathrm{F}} \le t_s^{\mathcal{M}}(C) + 2\sqrt{2s}\epsilon$. Indeed, the empirical and population restricted fits differ on at most $2s$ coordinates, and all coefficients outside $\mathcal{M}$ remain in the residual. For a purely non-marginal signal, this residual already equals the full signal norm.

C.2 A STRUCTURAL COUNTEREXAMPLE

The next example isolates which part of a three-coordinate interaction survives each projection.

Proposition C.1 (An interaction lost by a valid Tucker truncation)**.** *Let $D = E_{e_i+e_j+e_k}$ for distinct spatial coordinates i, j, k. Its three nonzero unfolding singular values are $1/\sqrt{3}$. Let P select any $r \in \{1, 2\}$ of those coordinate axes, which is a legitimate leading singular subspace under the tie. Then*

$$D \times_1 P \times_2 P \times_3 P = 0, \qquad \mathrm{Sym}(D \times_1 P) = \frac{r}{3} D. \tag{56}$$

Moreover $\max_{\|u\|_2=1} |D[u, u, u]| = \sqrt{2}/3 > 0$.

The specified coordinate subspace is a valid choice under the singular-value tie. The proof also exhibits a different leading subspace that retains signal, making the role of the subspace choice explicit.

Proof of Proposition C.1. Set $(i_1, i_2, i_3) = (i, j, k)$ and write

$$D = \frac{1}{\sqrt{6}} \sum_{\pi \in S_3} e_{i_{\pi(1)}} \otimes e_{i_{\pi(2)}} \otimes e_{i_{\pi(3)}}.$$

Its unfolding has exactly three nonzero rows. Each contains two entries $1/\sqrt{6}$, and their supports are disjoint. Hence $D_{(1)} D_{(1)}^\top = (e_i e_i^\top + e_j e_j^\top + e_k e_k^\top)/3$, which proves the singular-value claim. A projector onto any one or two of these coordinate axes is consequently a legitimate leading subspace. Applying it on all modes annihilates every summand, because each summand contains all three coordinates. Applying it only on the first mode preserves $2r$ of the six summands. Symmetrization redistributes each preserved elementary tensor equally among all six permutations, giving $(r/3)D$.

For a unit vector u, $D[u, u, u] = \sqrt{6} u_i u_j u_k$. The arithmetic–geometric mean inequality gives $|u_i u_j u_k| \le \{(u_i^2 + u_j^2 + u_k^2)/3\}^{3/2} \le 1/(3\sqrt{3})$. Equal absolute coordinates attain this bound, proving the stated maximum. The tie qualification is essential: if $u = (e_i + e_j + e_k)/\sqrt{3}$ and $P = uu^\top$, then P is also a valid leading rank-one projector and $D \times_1 P \times_2 P \times_3 P = (\sqrt{2}/3) u^{\otimes 3} \neq 0$. □

C.3 SPECTRAL ENERGY AND ALIGNMENT

Proof of Corollary A.2. Put $t = 8\ell^2$. The normalized error of rank r is at most

$$e_r = \left(\sum_{j > r} \mu_j \right)^{1/2} + \sqrt{rt}.$$

If $r_\ell = 0$, then $\mathcal{E}_\ell(D) = \kappa^2$, which contradicts either sufficient condition when $0 < \rho < 1$ and $\kappa > 0$. Hence $1 \le r_\ell \le p$. The strict threshold convention includes all ties equal to t in the tail, giving the exact identity

$$\mathcal{E}_\ell(D) = r_\ell t + \sum_{j > r_\ell} \mu_j.$$

The inequality $(u + v)^2 \le 2(u^2 + v^2)$ gives $e_{r_\ell}^2 \le 2\mathcal{E}_\ell(D)$. For the dyadic grid augmented by p, choose its smallest member $\bar{r} \ge r_\ell$; then $\bar{r} \le 2r_\ell$, including the final non-dyadic endpoint. Consequently

$$e_{\bar{r}}^2 \le 2 \sum_{j > r_\ell} \mu_j + 2\bar{r}t \le 4\mathcal{E}_\ell(D).$$

The corresponding fitted tensor is nonzero by the strict relative error bound. Equation (31) gives the claimed population jump. The proof selects a population oracle only to establish existence in the library; the actual validation rule observes no population eigenvalue. □

C.4 Rank choices under spectral decay

The following statements concern normalized noise $\ell \geq \|E_{(1)}\|_{\mathrm{op}}/a$ in $Y = aD + E$. Let $\mathcal{G}_p$ contain 0, every positive power of two at most p, and p. Write $\widehat{D}_r = B_r/a$ for positive ranks and $\widehat{D}_0 = 0$ only for approximation comparisons. The zero candidate is never normalized for inference. At rank p, $B_p = Y$ and the strengthened error bound is $\|\widehat{D}_p - D\|_{\mathrm{F}} \leq \sqrt{p}\ell$.

Proposition C.2 (Polynomial and exponential spectral envelopes)**.** *Let $\sigma_j = \sqrt{\mu_j}$ and $\ell > 0$. If $\sigma_j \leq Aj^{-s}$ for $A > 0$, $s > 1/2$, then*

$$\tau_r(D)^2 \leq \frac{A^2}{2s-1}\{r^{1-2s} - p^{1-2s}\}, \quad 1 \leq r \leq p, \tag{57}$$

$$\min_{r \in \mathcal{G}_p} \|\widehat{D}_r - D\|_{\mathrm{F}} \leq \min\{\kappa, \sqrt{p}\ell, C_s A^{1/(2s)} \ell^{1-1/(2s)}\}. \tag{58}$$

If instead $\sigma_j \leq Ae^{-b(j-1)}$ for $b > 0$, put $H = A/\sqrt{1 - e^{-2b}}$ and $\log_+(u) = \max(0, \log u)$. Then

$$\tau_r(D)^2 \leq H^2 e^{-2br}\{1 - e^{-2b(p-r)}\}, \quad 0 \leq r \leq p, \tag{59}$$

$$\min_{r \in \mathcal{G}_p} \|\widehat{D}_r - D\|_{\mathrm{F}} \leq \min\{\kappa, \sqrt{p}\ell, 5\ell\sqrt{1 + b^{-1}\log_+(H/\ell)}\}. \tag{60}$$

An informative direction requires the applicable error bound to be strictly below κ. If $\ell = 0$, full rank gives exact recovery.

Proof. The decreasing integral bound $\sum_{j=r+1}^{p} j^{-2s} \leq \int_r^p x^{-2s}\,dx$ gives (57), including the zero full-rank endpoint. For $0 < \ell < A$, set $u = (A/\ell)^{1/s} > 1$. If $u \leq p$, round upward to $g \in \mathcal{G}_p$ with $u \leq g \leq 2u$. Tail monotonicity and the SVD bound give

$$\|\widehat{D}_g - D\|_{\mathrm{F}} \leq Au^{1/2-s}/\sqrt{2s-1} + 4\ell\sqrt{u} = \{(2s-1)^{-1/2} + 4\}A^{1/(2s)}\ell^{1-1/(2s)}.$$

If $u > p$, full rank gives $\sqrt{p}\ell \leq \sqrt{u}\ell$. If $\ell \geq A$, the zero estimator has error $\kappa \leq A\sqrt{1 + (2s-1)^{-1}}$, which is bounded by the last expression after enlarging C_s. The zero and full-rank endpoints separately give the first two bounds in (58). The minimum is over available candidates; it is not asserted that one fixed rank attains each individual bound simultaneously.

A finite geometric sum proves (59). If $0 < \ell < H$, set $u = \max\{1, b^{-1}\log(H/\ell)\}$. When $u \leq p$, round upward to $g \leq 2u$. The tail is at most $He^{-bu} \leq \ell$, and the stochastic term is at most $4\ell\sqrt{u}$, so their sum is at most $5\ell\sqrt{u}$. If $u > p$, use full rank. If $\ell \geq H$, use the zero estimator and $\kappa \leq H$. Comparing with the zero and full-rank candidates proves (60). □

The positive grid has at most $\lfloor \log_2 p \rfloor + 2$ members. Consequently validation pays for a logarithmic number of rank candidates through J; it does not observe the population rank. For a polynomial envelope the continuous upper bound

$$A(2s-1)^{-1/2}r^{1/2-s} + \sqrt{8}\ell\sqrt{r}$$

has its stationary minimum at $r = \{A\sqrt{2s-1}/(\sqrt{8}\ell)\}^{1/s}$. Clip to $[1, p]$, round upward, and compare the strengthened endpoint bounds. This optimizes an upper envelope, not the unknown actual risk. When $s = 1/2$, the finite tail bound is $A^2\log(p/r)$; when $s < 1/2$ it is $A^2(p^{1-2s} - r^{1-2s})/(1-2s)$. The dimension-free polynomial rate above therefore requires $s > 1/2$.

C.5 Eigenvalue energy, effective dimension, and an ODECO specialization

Let $t = 8\ell^2 > 0$ and define $d_{\mathrm{eff}}(t) = \sum_j \mu_j/(\mu_j + t)$. For every $\mu \geq 0$,

$$\frac{t\mu}{\mu + t} \leq \min\{\mu, t\} \leq \frac{2t\mu}{\mu + t}, \qquad td_{\mathrm{eff}}(t) \leq \mathcal{E}_\ell(D) \leq 2td_{\mathrm{eff}}(t).$$

The scalar inequalities follow by separating $\mu \leq t$ and $\mu \geq t$. In particular, $td_{\mathrm{eff}}(t) \leq \rho^2\kappa^2/8$ implies the dyadic alignment condition. This effective dimension depends on the noise threshold and the entire spectrum.

Stable rank alone cannot replace a relative spectral-tail condition. Consider $\mu_1 = 1$ and $\mu_2 = \cdots = \mu_p = c/(p-1)$ with $0 < c \le p-1$. The stable rank of the unfolding is $1 + c$, but

$$\tau_r(D)^2/\kappa^2 = \frac{c(p-r)}{(p-1)(1+c)}.$$

A tail fraction below $c/(1+c)$ requires rank proportional to p. Diagonal orthogonal tensors realize this example. For a density perturbation in $d = p-1$ spatial dimensions, use this construction on d spatial directions and append one zero eigenvalue for the constant coordinate. Scaling the resulting finite polynomial sufficiently preserves positivity and boundedness without changing the relative spectrum on its spatial support. For a flat spectrum $\mu_j = \kappa^2/p$, the clipped energy is $\min(\kappa^2, pt)$, so a small relative error still requires t of order κ^2/p or smaller.

For an orthogonal decomposition, unfolding gives $D_{(1)} = \sum_j \lambda_j v_j (v_j \otimes v_j)^\top$ with orthonormal left and right factors. Thus $\mu_j = \lambda_j^2$ and

$$\sum_j \min\{\lambda_j^2, 8\ell_P^2\} \le \rho^2\kappa^2/4 \tag{61}$$

is sufficient for some dyadic SVD direction to retain $\kappa\sqrt{1-\rho^2}$. The component vectors satisfy $D[v_j, v_j, \cdot] = \lambda_j v_j$; these are the orthogonal decomposition eigenpairs, not a claim to enumerate all possible tensor eigenpairs. A fixed available rank also suffices when

$$\sum_{j>r} \lambda_j^2 \le \beta_0^2\kappa^2, \quad \ell_P \le \frac{\beta_1\kappa}{\sqrt{8r}}, \quad \beta_0, \beta_1 \ge 0, \quad \beta_0 + \beta_1 < 1.$$

It gives retained signal at least $\kappa\sqrt{1-(\beta_0+\beta_1)^2}$. Under dependent concentration replace ℓ_P by the corresponding bound $U_k(w)/a$. The bounded-envelope contribution in (95) remains part of this condition.

C.6 Tensor noise bounds for orthogonal approximation

Proof of Theorem A.3. A global minimizer exists. For each orthonormal frame $U = (u_1, \ldots, u_r)$, the tensors $u_j^{\otimes 3}$ are Frobenius orthonormal, so the optimal coefficients are $\langle Y/a, u_j^{\otimes 3}\rangle$. Their squared sum is continuous on the compact Stiefel manifold. A frame maximizing this sum attains a least-squares minimum, with zero coefficients allowed.

Let D_r contain the first $\min(r, R_0)$ components, completed by zero coefficients if necessary, and put $z = \|\widehat{T}_r - D\|_{\mathrm{F}}$, $\tau = \|D - D_r\|_{\mathrm{F}}$. Optimality gives

$$z^2 \le \tau^2 + 2\langle Z, \widehat{T}_r - D_r\rangle.$$

The difference is supported on a common mode subspace of dimension $m \le \min(p, 2r)$. Let P project onto it. In an orthonormal basis of that subspace, each of the m matrix slices of $Z \times_1 P \times_2 P \times_3 P$ has operator norm at most ω and Frobenius norm at most $\sqrt{m}\omega$. Hence

$$\|Z \times_1 P \times_2 P \times_3 P\|_{\mathrm{F}} \le m\omega, \quad |\langle Z, \widehat{T}_r - D_r\rangle| \le m\omega(z+\tau).$$

Factor $z^2 - \tau^2$ and divide by $z + \tau$ if nonzero. This gives $z \le \tau + 2m\omega \le \tau + 4r\omega$. The zero case is immediate.

For a specified true leading frame, let Π project onto the span of its cubic tensors. Its signal residual and projected noise are orthogonal, while each noise coefficient has magnitude at most ω. Therefore

$$\|\Pi(D+Z) - D\|_{\mathrm{F}}^2 = \|(I-\Pi)D\|_{\mathrm{F}}^2 + \|\Pi Z\|_{\mathrm{F}}^2 \le \sum_{j>r} \lambda_j^2 + r\omega^2.$$

This proves the known-frame statement. A data-chosen frame still satisfies the same projection identity, but its residual cannot be replaced by the true eigenvalue tail without an additional argument. □

Under the independent bounded-density model, the injective bound in Lemma B.1, with $a \asymp \sqrt{h}$, gives

$$\omega \le C\nu_L\{\sqrt{(p+x)/h} + (p+x)^{3/2}/h\}$$

on a simultaneous event whose x includes the logarithm of the candidate contrasts and failure probability. It can be used in (37), but not in place of unfolding operator noise in (29). At leading order the resulting stochastic upper bounds scale as $r\sqrt{p/h}$ for the global orthogonal fit and $\sqrt{r}\,p/\sqrt{h}$ for unfolding SVD, with their other terms retained. These scales compare the computable unfolding estimator with a theoretical global orthogonal fit. They quantify how additional tensor structure can change the noise dependence under the stated optimization premise. The empirical study evaluates the unfolding estimator and the other candidates in Section 2.

D From a retained jump to calibrated inference

Section 3 bounds the quality of a candidate direction. This appendix explains how the procedure uses it: selection preserves a controlled part of the jump, the pilot places the fitting window, and independent anchors permit scalar refinement. We state the intermediate inference guarantees before proving the complete primitive-parameter result. For each cyclic ordering, conditional arguments use only its own pilot, validation observations, and anchors as specified; integrating these bounds precedes the union over orderings.

D.1 Independent selection and exact null calibration

Theorem D.1 (Selection and null control)**.** *Condition on a pilot fold, including its window and J directions of Frobenius norm at most one. Suppose the validation population contrast is $a_B D$, $a_B > 0$, and $\max_j |\langle \widehat{C}_B - a_B D, V_j\rangle| \le b$. Then the selected and oriented direction satisfies*

$$\langle D, \widehat{V}\rangle \ge \Gamma_A - 2b/a_B, \qquad \Gamma_A = \max_{j\le J} |\langle D, V_j\rangle|. \tag{62}$$

Under any iid null distribution, each p_b in (26) *is super-uniform and $\mathbb{P}(p_{\mathrm{global}} \le \alpha_0) \le \alpha_0$.*

The null statement tests equality of distributions with an unknown common background and also holds without a density bound. Random-permutation validity follows from the group randomization argument (Hemerik and Goeman, 2018). Repeating maximization in each permutation incorporates direction selection into the calibrated statistic.

The selection cost depends on J, conditional on the pilot, rather than directly on the ambient tensor dimension. Directional polynomial moments give

$$b(x) = C\nu_L\{\sqrt{x} + x^{3/2}/\sqrt{h}\}, \qquad \nu_L = C\max(1, \sqrt{L}). \tag{63}$$

For validation alone take $x = \log(6J/\delta_V)$ and union over the three cyclic orderings. This conditional statement leaves the difficulty of learning useful candidates in Γ_A; Theorem A.1 quantifies one route to controlling it.

D.2 Proof of the selection and calibration guarantee

Proof of Theorem D.1. Write $T_j = a_B\langle D, V_j\rangle + e_j$, where $|e_j| \le b$. Let j_* maximize $|\langle D, V_j\rangle|$. By the maximizing and orientation rules,

$$a_B\langle D, \widehat{V}\rangle \ge |T_{\widehat{j}}| - b \ge |T_{j_*}| - b \ge a_B\Gamma_A - 2b.$$

This proves (62), including zero observed values with the stated sign convention.

Under the iid null, condition on the pilot. Its window, finite library, and any algorithmic choices made there are fixed, and its validation observations remain iid. Conditional further on the unordered validation sample, the observed arrangement is uniform on its permutation orbit. Independent uniform group permutations produce arrangements with the same joint exchangeability, including the original arrangement as one of the $R+1$ entries. The upper rank with ties counted by $\ge$ therefore produces a super-uniform p value. This is the random-permutation group argument of Hemerik and

Goeman (2018); repeated values can only make the rank test conservative. The statistic is the entire maximum over the library, recomputed in every permutation. Integrating the conditional validity gives $\mathbb{P}(p_b \le u) \le u$ for every $u \in [0,1]$. Finally,

$$\mathbb{P}(p_{\text{global}} \le \alpha_0) \le \sum_{b=0}^{2} \mathbb{P}(p_b \le \alpha_0/3) \le \alpha_0.$$

No independence between the three p values is used, and no bounded-density assumption is needed for this null statement. □

D.3 Population geometry and pilot sufficient conditions

On a single-change interval $(s, e]$, put $N = e - s$ and let τ be the true split. Summing the piecewise constant means against the right-minus-left weights gives $C^t_{s,e} = a(t)D$, where

$$a(t) = \begin{cases} \dfrac{e-\tau}{\sqrt{N}}\sqrt{\dfrac{t-s}{e-t}}, & s < t \le \tau, \\ \dfrac{\tau-s}{\sqrt{N}}\sqrt{\dfrac{e-t}{t-s}}, & \tau \le t < e. \end{cases} \tag{64}$$

Let Ψ be an absolutely homogeneous score, $\gamma = \Psi(D) > 0$, and $g(t) = a(t)\gamma$. Direct subtraction gives

$$g(\tau)^2 - g(t)^2 = \begin{cases} \gamma^2(\tau - t)(e - \tau)/(e - t), & t \le \tau, \\ \gamma^2(t - \tau)(\tau - s)/(t - s), & t \ge \tau. \end{cases} \tag{65}$$

If $\beta = \min(\tau - s, e - \tau)/N$, this is at least $\beta\gamma^2|t - \tau|$. Suppose a finite candidate set contains τ, a computed score has uniform error at most ϵ from g, and c maximizes that score. Then $g(\tau) - g(c) \le 2\epsilon$ and $g(\tau) + g(c) \le 2g(\tau) \le \gamma\sqrt{N}$. Hence

$$|c - \tau| \le \frac{2\sqrt{N}\epsilon}{\beta\gamma}. \tag{66}$$

For the coefficient pilot, $\Psi(T) = \max_\alpha |\langle T, E_\alpha\rangle|$ and γ is the largest absolute jump coefficient. Lemma B.1, unioned over the predetermined candidate splits, supplies ϵ. The condition $2\sqrt{N}\epsilon/(\beta\gamma) \le h/16$ ensures the required center accuracy. The segment-length condition in Theorem D.2 puts the true split inside $[h, N - h]$.

Candidate quality is a separate condition. On the accurate-center event the local pilot population contrast is $a_A D$ with $a_A = (h - |c - \tau_A|)/\sqrt{2h} \ge 15\sqrt{h}/(16\sqrt{2})$. If at least one candidate obeys a bound $e_j < a_A\kappa$ from Theorem A.1, then

$$\Gamma_A \ge \kappa\sqrt{1 - \{e_j/(a_A\kappa)\}^2}.$$

An explicit sufficient condition for $\Gamma_A \ge g$ is that this right side be at least g. Because the local center is selected from pilot observations, a probabilistic use of the approximation bound requires a simultaneous error event for all local centers that the pilot can select, or another valid selection argument. Theorem A.5 supplies the simultaneous error event and verifies these sufficient conditions from the primitive parameters.

D.4 Power and localization with explicit conditions

Let G_b be measurable on pilot fold b. On G_b, require $|c_b - \tau_b| \le h/16$ and $\Gamma_b \ge g > 0$. Suppose $\mathbb{P}(\cap_b G_b) \ge 1 - \delta_P$, and every true segment in each fold has length at least $2h$.

Theorem D.2 (Detection and local refinement)**.** *Put* $x_V = \log\{12J(R+1)/\delta_V\}$, $b_V = b(x_V)$, $x_A = \log(12/\delta_A)$, *and* $x_C = \log(6/\delta_C)$. *Assume*

$$\sqrt{h}\, g \ge 8b_V + 4\kappa\sqrt{x_V}, \qquad C\nu_L\{\sqrt{2x_A/h} + 2x_A^{3/2}/h\} \le g/16. \tag{67}$$

If $3/(R+1) \le \alpha_0$, *the procedure rejects with probability at least* $1 - \delta_P - \delta_V$. *With probability at least* $1 - \delta_P - \delta_V - \delta_A - \delta_C$, *it also satisfies*

$$|\widetilde{\eta} - \eta| \le C\{\nu_L^2 x_C/g^2 + \nu_L x_C^{3/2}/g + 1\}. \tag{68}$$

These conditions specify when a useful pilot and retained candidate jump are sufficient for the inference procedure. For fixed confidence and bounded L, the leading small-signal localization scale is g^{-2}. The proof controls each fixed cyclic ordering and then takes a union bound. Each conditional bound is integrated before combining the three orderings.

The pilot geometry above supplies the center condition needed by this result. Appendix F.1 situates its localization scale through a two-location lower bound and gives a separate detection lower bound over pure cubic coefficients.

D.5 Random-permutation power and scalar refinement

Lemma D.3 (Balanced random weights)**.** *Let $I \subset \{1, \dots, 2h\}$ be fixed and let (s_i) be uniform among sign vectors with h positive and h negative entries. Then, for $x \geq 0$,*

$$\mathbb{P}\left(\left|\sum_{i\in I} \frac{s_i}{\sqrt{2h}}\right| > \sqrt{x}\right) \leq 2e^{-x}.$$

Proof. Generate a uniform random perfect matching of the positions, and orient each pair independently with one positive and one negative sign. The resulting law is invariant under position permutations and is supported on all balanced sign vectors, so it is uniform. Conditional on the matching, only pairs with exactly one endpoint in I contribute. If there are $k \leq h$ such pairs, the sum is a sum of k independent Rademacher variables divided by $\sqrt{2h}$. Its moment generating function is at most $\exp(kt^2/(4h)) \leq \exp(t^2/4)$. Applying Chernoff to each sign proves the claim. □

Lemma D.4 (Scalar localization with independent anchors)**.** *Consider independent scalar observations on a fixed interval with a single mean change at τ, of magnitude $d_\mu > 0$, and centered $\psi_{2/3}$ norms at most ν. Fix two mean estimates, independent of these observations, whose errors are at most $d_\mu/8$. A least-squares minimizer using these fixed means over a search set containing τ satisfies, for $x \geq 1$,*

$$\mathbb{P}\left\{|\widehat{\tau} - \tau| > C\left(\frac{\nu^2 x}{d_\mu^2} + \frac{\nu x^{3/2}}{d_\mu} + 1\right)\right\} \leq 2e^{-x}. \tag{69}$$

For random independent anchors this holds conditionally on each realization satisfying the stated accuracy.

Proof. Changing the signs of observations and estimated means preserves squared residuals, so take $\mu_R - \mu_L = d_\mu > 0$. For $t > \tau$, write $v = t - \tau$, $Z_j = \mu_R + \xi_j$, and $\widehat{d} = \widehat{\mu}_R - \widehat{\mu}_L$. The exact RSS increment is

$$Q(t) - Q(\tau) = v\widehat{d}(2\mu_R - \widehat{\mu}_L - \widehat{\mu}_R) + 2\widehat{d}\sum_{j=\tau+1}^{t} \xi_j.$$

Both deterministic factors lie in $[3d_\mu/4, 5d_\mu/4]$. The increment is therefore at least $cvd_\mu^2 - Cd_\mu|\sum_{j=\tau+1}^{t} \xi_j|$. For $t < \tau$ reverse the pre-change errors; the deterministic factor becomes $\widehat{\mu}_L + \widehat{\mu}_R - 2\mu_L$ and obeys the same bounds.

Integrating (51) with unit weights gives $\|\sum_{j=1}^{m} \xi_j\|_{L^u} \leq C\nu(\sqrt{mu} + u^{3/2})$ for $u \geq 2$. Partial sums are a martingale. Doob's inequality, with multiplier at most two for $u \geq 2$, followed by Markov's inequality, gives

$$\mathbb{P}\left\{\max_{1\leq t\leq m}\left|\sum_{j=1}^{t} \xi_j\right| > C\nu(\sqrt{mu} + u^{3/2})\right\} \leq e^{-u}.$$

Let M be the ceiling of $C_0(\nu^2 x/d_\mu^2 + \nu x^{3/2}/d_\mu + 1)$ and split distances into shells $2^j M \leq v < 2^{j+1}M$. For each side use the last inequality with $u = x + j + 2$ and with m the smaller of the available side length and $\lceil 2^{j+1}M \rceil$. For $x \geq 1$, $(x + j + 2)/2^j \leq Cx$ and $(x + j + 2)^{3/2}/2^j \leq Cx^{3/2}$. Choosing C_0 large enough makes the stochastic term smaller than the drift in every shell. No minimizer then has distance at least M. The two sides' exceptional probabilities sum to at most $2\sum_{j\geq 0} e^{-(x+j+2)} < 2e^{-x}$. Empty shells contribute nothing. Absorbing the ceiling proves (69). □

Proof of Theorem D.2. Fix a cyclic ordering with pilot fold b, and condition on that pilot. On G_b its center and library are fixed. True split counts in any two interlaced folds differ by at most one. Since $h \geq 16$,

$$|c_b - \tau_B| \leq h/16 + 1 \leq h/8, \qquad |c_b - \tau_C| \leq h/8.$$

The validation window is within the fold by construction. Its population coefficient satisfies

$$a_B = \frac{h - |c_b - \tau_B|}{\sqrt{2h}} \geq \frac{7}{8\sqrt{2}}\sqrt{h} \geq \sqrt{h}/2.$$

The same geometry puts the localization split between the two anchors and inside the central search interval. Each anchor has $h/2$ observations and is pure. The segment-length assumption also ensures that these windows isolate the corresponding change.

For the power argument let I be the post-change positions in the validation window. Write $H_3(X_i) = A(f_0) + \mathbf{1}\{i \in I\}D + Z_i$. For any random permutation and any candidate direction the validation statistic is

$$\langle D, V_j\rangle \sum_{i\in I} w_i^\pi + \sum_i w_i^\pi \langle Z_i, V_j\rangle.$$

Conditional on the pilot and the permutation, the second term is an independent centered sum with deterministic weights. Equation (51) bounds its absolute value by b_V with failure probability at most $2e^{-x_V}$. This noise bound also holds for the unpermuted statistic. Lemma D.3 bounds the random-permutation mean term by $\kappa\sqrt{x_V}$ with the same failure probability, simultaneously over directions because $|\langle D, V_j\rangle| \leq \kappa$. Across three orderings there are at most $3J(R+1)$ noise events and $3R$ random-balance events. Their integrated failure probability is at most

$$\{6J(R+1) + 6R\}e^{-x_V} \leq \delta_V.$$

Only the noise bound includes the identity permutation; the random-balance assertion applies to the R random draws. On the complementary event, in each ordering,

$$T_{\text{obs}} \geq a_B\Gamma_b - b_V \geq \tfrac{1}{2}\sqrt{h}\, g - b_V, \qquad \max_{1\leq r\leq R} T^{(r)} \leq b_V + \kappa\sqrt{x_V}.$$

The first condition in (67) makes the first quantity strictly larger than the second. All three p values equal $1/(R+1)$, so the global test rejects under the stated resolution condition. Adding δ_P proves the detection guarantee.

The same validation noise event and Theorem D.1 imply

$$d_{\mu,b} := \langle D, \widehat{V}_b\rangle \geq g - 4b_V/\sqrt{h} \geq g/2.$$

Now condition on this ordering's pilot and validation folds. Its direction and windows are fixed, while the localization observations remain independent. Applying (51) to an anchor mean of $h/2$ observations bounds its error by

$$C\nu_L\{\sqrt{2x_A/h} + 2x_A^{3/2}/h\} \leq g/16 \leq d_{\mu,b}/8$$

with failure probability at most $2e^{-x_A}$. Integrate separately for each ordering. There are six anchors in total, so their failure probability on the preceding success events is at most $12e^{-x_A} = \delta_A$.

For this same fixed ordering, condition further on its two anchors. The disjoint central observations retain their independent law and the moment bound (50). Lemma D.4 bounds its fold-count localization error by

$$C\{\nu_L^2 x_C/g^2 + \nu_L x_C^{3/2}/g + 1\}$$

with conditional failure probability at most $2e^{-x_C}$. Integrating and summing the three fixed-ordering failure probabilities gives $6e^{-x_C} = \delta_C$. This argument never conditions on the other orderings' complete fitted procedures.

Finally $\tau_b = \lfloor(\eta + 2 - b)/3\rfloor$ implies $|3\tau_b - \eta| \leq 2$. Mapping each successful fold estimate to three times its count multiplies its radius by three and adds at most two. On the simultaneous event all three mapped estimates lie within the resulting radius of η, and so does their median. Absorbing these constants proves (68). The event also contains rejection; no conditioning on rejection or on a data-selected best cyclic ordering is needed. □

D.6 Explicit noise levels and the complete guarantee

Equation (40) states the explicit quantities used here. We prove that they simultaneously cover all pilot choices and then verify each stage of the inference guarantee.

Proof of Theorem A.5. For each of the three pilot folds include every full-interval contrast with split in $[h, N-h]$ and every local contrast with center in that set. Their number is at most $M = 6(N+1)$. Every weight vector has Euclidean norm one and maximum absolute coordinate at most $h^{-1/2}$. The coordinate inequality in Lemma B.1, with failure probability $\delta_P/(2M)$ for each contrast, gives the common bound ϵ_P. Lemma A.4 improves the variance proxy in its matrix inequality to Lq/p. Applying that matrix inequality with the same failure allocation gives U_P. A union bound gives a single event of probability at least $1-\delta_P$ on which both bounds hold at every candidate contrast. In particular, selecting a local window using the pilot does not change this event.

For the coefficient pilot the score is a norm with perturbation at most ϵ_P. For the structured pilot, write

$$\Psi_s^{\mathrm{bas}}(T) = \max_{|I|\le s} \|(\langle T, E_\alpha\rangle)_{\alpha\in I}\|_2, \quad \Psi_r^{\mathrm{svd}}(T) = \max_{U^\top U = I_r} \|U^\top T_{(1)}\|_{\mathrm{F}}.$$

Both are norms: each is a supremum of seminorms, is positively homogeneous, and vanishes only at zero. The second variational identity follows from the eigenvalue variational principle for $T_{(1)}T_{(1)}^\top$. Their triangle inequalities give

$$|\Psi_s^{\mathrm{bas}}(T+E) - \Psi_s^{\mathrm{bas}}(T)| \le \sqrt{s}\max_\alpha |\langle E, E_\alpha\rangle|, \quad |\Psi_r^{\mathrm{svd}}(T+E) - \Psi_r^{\mathrm{svd}}(T)| \le \sqrt{r}\|E_{(1)}\|_{\mathrm{op}}.$$

Taking the maximum after dividing by the stated deterministic bounds gives uniform error at most one for Ψ_*. The segment assumption places the true split in the candidate set. The population margin (66) and the first condition in (42) therefore imply $|c_b - \tau_b| \le h/16$ for each pilot.

On this event the local amplitude satisfies $a_b \ge a_0$. Since $t_s(a_bD) = a_b t_s(D)$ and $\tau_r(a_bD) = a_b\tau_r(D)$, the approximation theorem and the second condition in (42) imply, for at least one fitted candidate,

$$\|B - a_bD\|_{\mathrm{F}}/(a_b\kappa) \le \rho.$$

Equation (31) yields $\Gamma_b \ge g = \kappa\sqrt{1-\rho^2}$. Define G_b by the two resulting pilot conclusions; it is measurable on pilot fold b. We have proved $\mathbb{P}(\cap_b G_b) \ge 1-\delta_P$, rather than assuming it. The conditions of Theorem D.2 now hold, proving both power and localization. Its proof conditions on each cyclic ordering separately, integrates, and then takes a union bound, so overlap between orderings creates no extra independence assumption.

Under the iid null either pilot depends only on its own fold. Conditional on that fold, the validation observations remain exchangeable. The proof of Theorem D.1 therefore applies to either deterministic scoring rule, independently of its tuning or power conditions. □

D.7 A common eigenvalue-energy condition for the complete procedure

Proof of Corollary A.6. All quantities use the definitions in (40). In particular $U_P > 0$, $a_0 = 15\sqrt{h}/(16\sqrt{2}) > 0$, and $\ell_P = U_P/a_0 > 0$. The segment-length conditions give $\beta > 0$, so $\xi > 0$; also $\xi \le \rho^2/4 < 1$. Set $t = 8\ell_P^2$ and $S = \sum_j \min\{\mu_j, t\}$. If $\mu_1 \le t$, then $S = \sum_j \mu_j = \kappa^2$, contradicting $S \le \xi\kappa^2$. Thus $\mu_1 > t$.

For every j, $\min\{\mu_j, t\} \ge t\mu_j/\mu_1$: when $\mu_j \le t$ this uses $t/\mu_1 < 1$, and when $\mu_j > t$ it uses $\mu_j \le \mu_1$. Summing gives

$$S \ge \frac{t\kappa^2}{\mu_1}, \qquad \mu_1 \ge \frac{t}{\xi}, \qquad \gamma_* \ge \frac{\sqrt{\mu_1}}{U_P} \ge \frac{\sqrt{8}}{a_0\sqrt{\xi}} = \frac{64}{15\sqrt{h\xi}}.$$

The pilot includes rank one, so the bound on γ_* applies to the score actually used. Since $\xi \le 4\beta^2 h/(225N)$,

$$\beta h\gamma_* \ge \frac{64\beta\sqrt{h}}{15\sqrt{\xi}} \ge 32\sqrt{N}.$$

This is the first condition in (42) for normalized pilot noise $e = 1$.

For the fitting condition, $4S \le \rho^2\kappa^2$. Let $r_\ell = \#\{j : \mu_j > t\}$, which is positive, and choose the smallest available dyadic rank or endpoint $\bar{r} \ge r_\ell$. As in Corollary A.2, $\bar{r} \le 2r_\ell$ and

$$\{\tau_{\bar{r}}(D) + 2\sqrt{2\bar{r}}\ell_P\}^2 \le 4S \le \rho^2\kappa^2.$$

The corresponding SVD candidate belongs to the fitted library, proving the second condition in (42). The remaining assumptions were retained explicitly. Theorem A.5 now gives the rejection and joint localization probabilities claimed. Threshold ties remain in the spectral tail, and the endpoint $\bar{r} = p$ is included. □

E Dependence bounds and calibrated thinning

The deterministic spectral argument requires a suitable contrast-noise bound; calibration additionally requires a comparison of the data law. A single product-law lemma supplies both probability arguments, with separate costs for all-observation concentration and retained-grid inference.

Lemma E.1 (Product-law comparison on a separated grid)**.** *For deterministic indices $s_1 < \cdots < s_m$ separated by at least k, the joint law P of the retained observations and the product Q of their marginal laws satisfy*

$$d_{\mathrm{TV}}(P, Q) \le (m-1)\beta_X(k).$$

The same bound holds after applying any measurable randomized algorithm with auxiliary randomness independent of the data.

Proof. For the first j retained variables, data processing bounds the distance between their joint law and the product of the first $j-1$ variables' joint law with the last marginal by $\beta_X(k)$. The triangle inequality and $d_{\mathrm{TV}}(P \otimes R, Q \otimes R) = d_{\mathrm{TV}}(P, Q)$ give the claim by induction on j. Appending independent algorithmic randomness and applying a measurable map cannot increase total variation. Equivalently one can couple the two complete retained vectors with failure probability at most this bound and use common auxiliary randomness, as in absolute-regularity coupling (Berbee, 1979). No stationarity is needed. □

Proof of Theorem G.5. Partition $[n]$ into the k residue classes $I_\ell = \{\ell, \ell+k, \ldots\} \cap [n]$. Compare each entire class once to its independent product law using Lemma E.1. The sum of all comparison costs is $\sum_\ell (|I_\ell| - 1)\beta_X(k) = (n-k)\beta_X(k)$. For each weight vector set $a_\ell = (\sum_{i \in I_\ell} w_i^2)^{1/2}$ and $b_\ell = \max_{i \in I_\ell} |w_i|$. Under the product law of a class, Lemma A.4 and the unchanged envelope $2b_0\sqrt{q}$ give

$$\left\| \sum_{i \in I_\ell} w_i \{M_i - \mathbb{E}M_i\} \right\|_{\mathrm{op}} \le \sqrt{2L(q/p)x}\, a_\ell + \tfrac{4b_0\sqrt{q}}{3} x b_\ell$$

except on an event of probability $(p + p^2)e^{-x}$. Coordinate Hoeffding gives the bound $b_0\sqrt{2x}a_\ell$ except on an event of probability $2qe^{-x}$. A union bound over all M_0 weights on this single class precedes the product-law comparison. Thus the comparison cost is incurred once for the entire class, not once per contrast. Summing over classes gives total failure at most

$$M_0 k(p + p^2 + 2q)e^{-x} + (n-k)\beta_X(k) \le \zeta + (n-k)\beta_X(k).$$

On the common event, the triangle inequality, $\sum_\ell a_\ell \le \sqrt{k}\|w\|_2$, and $\sum_\ell b_\ell \le k\|w\|_\infty$ prove the noise bounds. Insert the matrix bound into (29), noting $\tau_r(aD) = a\tau_r(D)$, to obtain the final claim. Dependence between different residue classes is immaterial to this union-bound argument. □

Proof of Theorem G.6. Apply Lemma E.1 to the entire retained vector. Under its product law, the null is iid, so Theorem D.1 gives $\mathbb{P}_Q(p_{\mathrm{global}} \le u) \le u$. Transferring this event gives the null inequality. For $u < 1$, the event $\min(1, p_{\mathrm{global}} + \varepsilon_{\mathrm{mix}}) \le u$ is empty if $u < \varepsilon_{\mathrm{mix}}$, and otherwise is contained in $\{p_{\mathrm{global}} \le u - \varepsilon_{\mathrm{mix}}\}$. Its probability is at most u. At $u = 1$ the assertion is immediate.

Under the alternative, let $\eta' = \#\{j \le m : 1 + (j-1)k \le \eta\}$. The retained segment conditions imply $1 \le \eta' < m$, hence $\eta' = \lceil \eta/k \rceil$ and $|k\eta' - \eta| \le k - 1$. Apply Theorem A.5 under the product

law with sample size m, location η', and the stated retained-sample parameters. Transfer its joint event of rejection and localization once. The triangle inequality gives

$$|k\widetilde{\eta}_{\text{retained}} - \eta| \le k|\widetilde{\eta}_{\text{retained}} - \eta'| + |k\eta' - \eta| \le k\mathfrak{r}(g) + k - 1.$$

Clipping the mapped estimate to $[1, n]$ cannot increase its error. For the adjusted test, rejection at level α_0 is equivalent to the unadjusted rejection at $\alpha_0 - \varepsilon_{\text{mix}}$ when this is positive; in particular the permutation resolution condition becomes $3/(R+1) \le \alpha_0 - \varepsilon_{\text{mix}}$. The theorem excludes changes lying after the last retained point through its segment-length assumption. □

For explicit choices of spacing, suppose first that $\beta_X(k) \le Be^{-ck}$. The deterministic choice

$$k = \max\{1, \lceil c^{-1} \log(Bn/\varepsilon) \rceil\}$$

makes both comparison costs at most ε, provided enough samples remain. For $\beta_X(k) \le Bk^{-a}$ with $a > 0$, the full-data bound can use $k \ge (Bn/\varepsilon)^{1/a}$. Thinning has the smaller cost $(m-1)\beta_X(k) \le Bn/k^{a+1}$ and can instead use $k \ge (Bn/\varepsilon)^{1/(a+1)}$. Integer ceilings and $k \ge 1$ are understood. If a chosen gap leaves too few retained observations, the inference theorem is inapplicable. The full-data concentration and thinning procedures therefore have different costs under polynomial mixing.

E.1 A SEPARATE ROUTE THROUGH FUNCTIONAL DEPENDENCE

The previous results assume absolute regularity. A second description is useful when the observed process is a causal function of iid innovations. Let $X_i = G_i(\varepsilon_i, \varepsilon_{i-1}, \ldots) \in [-1, 1]^d$, where G_i may vary with time. Let $X_i^{(i-j)}$ replace the innovation at time $i - j$ by an independent copy, and define

$$\delta_X(j, u) = \sup_i \max_{a \le d} \|X_{i,a} - X_{i,a}^{(i-j)}\|_{L^u}.$$

Proposition E.2 (Cubic feature dependence and a uniform second-moment bound)**.** *For every coefficient $|\alpha| \le 3$ and $u \ge 1$,*

$$\delta_{\psi_\alpha}(j, u) \le 6\sqrt{7}\, \delta_X(j, u). \tag{70}$$

If $\Delta_2 = \sum_{j \ge 0} \delta_X(j, 2) < \infty$, then for deterministic weights

$$\left\| \sum_i w_i \{\psi_\alpha(X_i) - \mathbb{E}\psi_\alpha(X_i)\} \right\|_2 \le 6\sqrt{7}\, \Delta_2 \|w\|_2. \tag{71}$$

For any M_0 deterministic contrasts, all coefficients simultaneously satisfy $\max_\alpha |\langle E_w, E_\alpha \rangle| \le 6\sqrt{7}\Delta_2 \|w\|_2 \sqrt{qM_0/\zeta}$ with probability at least $1 - \zeta$.

Proof. For the nonzero degree patterns $(1), (2), (1,1), (3), (2,1), (1,1,1)$, the sums of uniform absolute derivative bounds are respectively $\sqrt{3}, 3\sqrt{5}, 6, 6\sqrt{7}, 4\sqrt{15}, 9\sqrt{3}$. Their maximum is $6\sqrt{7}$. Telescope one changed coordinate at a time and use the L^u triangle inequality with the individual derivative bounds. This proves (70) without interchanging a maximum over coordinates with an expectation.

Write $Z_i = \psi_\alpha(X_i) - \mathbb{E}\psi_\alpha(X_i)$ and $\mathcal{F}_t = \sigma(\varepsilon_s : s \le t)$. The projections $\mathcal{P}_t Z = \mathbb{E}(Z \mid \mathcal{F}_t) - \mathbb{E}(Z \mid \mathcal{F}_{t-1})$ are orthogonal across t in L^2. The causal expansion is $Z_i = \sum_{j \ge 0} \mathcal{P}_{i-j} Z_i$; the remote-past conditional expectation equals the mean by the iid tail zero-one law and reverse martingale convergence. Replacing ε_{i-j} gives $\|\mathcal{P}_{i-j} Z_i\|_2 \le \delta_{\psi_\alpha}(j, 2)$, by conditional Jensen. Hence

$$\left\| \sum_i w_i Z_i \right\|_2^2 \le \sum_t \left(\sum_{j \ge 0} |w_{t+j}| \delta_{\psi_\alpha}(j, 2) \right)^2 \le \|w\|_2^2 \left(\sum_{j \ge 0} \delta_{\psi_\alpha}(j, 2) \right)^2.$$

The last step is the ℓ^1–ℓ^2 convolution inequality, which also justifies convergence of the projection sums. Substitute (70). Chebyshev and a union bound over the qM_0 normalized contrasts prove the final statement; zero weight vectors satisfy it deterministically. □

This coefficient result requires no density upper bound. For a fixed tensor direction $V = \sum_\alpha v_\alpha E_\alpha$, the same argument gives a dependence factor $6\sqrt{7}\|v\|_1$, bounded by $6\sqrt{7s}\|V\|_{\mathrm{F}}$ for s nonzero coefficients. A dimension-free coefficient bound therefore does not give the same dimension-free bound for every dense unit direction. The resulting uniform bound has a polynomial confidence dependence and applies to this causal-innovation assumption class.

Functional dependence alone does not imply the absolute regularity used above. Indeed, let iid Bernoulli bits define $U_t = \sum_{j\ge 0} 2^{-j-1}\varepsilon_{t-j}$ and $X_t = 2U_t - 1$. Each X_t has the uniform reference density, and replacing one innovation changes it by at most 2^{-j}. However the binary expansion of U_k determines U_0 after deleting its first k digits, almost surely. The pair (U_0, U_k) lies on the graph of this recovery map. The graph has joint probability one and product probability zero because U_0 is non-atomic. Thus $\beta_X(k) = 1$ for every finite k. This example separates the assumptions needed for coefficient estimation from the absolute-regularity condition used for calibration.

F Further consequences for detection and localization

The preceding inference result concerns one change. This appendix situates its signal and localization scales through explicit lower-bound subclasses, then gives a deterministic construction for several changes from a uniform score bound. Each result specifies the observations, loss, and signal assumptions it uses.

F.1 Detection and localization lower bounds

The first construction studies detection over an orthogonal family of cubic coefficients. The second studies location error for two possible change positions. Both are explicit subclasses of the bounded-density model.

Proposition F.1 (A pure-cubic mixture detection bound)**.** *Let $q_3 = \binom{d+2}{3}$ and enumerate the total-degree-exactly-three Legendre basis functions as $\psi_1, \dots, \psi_{q_3}$. Let P_0 give n independent observations from μ. Under P_j, a known set of m observations has density $f_j = 1+\theta\psi_j$ and all other observations have density one, where $0 < |\theta| \le 1/(6\sqrt{3})$. For $\overline{P} = q_3^{-1}\sum_j P_j$,*

$$\chi^2(\overline{P}, P_0) = \frac{(1+\theta^2)^m - 1}{q_3}. \tag{72}$$

If $q_3 \ge 2$ and $m\theta^2 \le \frac{1}{2}\log q_3$, any test of size at most α has average power at most $\alpha + \frac{1}{2}q_3^{-1/4}$ over these alternatives.

Proof. Each ψ_j integrates to zero, has unit $L^2(\mu)$ norm, and is bounded by $3\sqrt{3}$. Thus $1/2 \le f_j \le 3/2$ and the densities are valid. If I is the changed set, the likelihood ratio is $L_j = \prod_{i\in I}(1+\theta\psi_j(X_i))$. Orthogonality gives

$$\mathbb{E}_0 L_j L_k = (1 + \theta^2 \mathbf{1}\{j = k\})^m.$$

Expanding the square of the averaged likelihood ratio proves (72). Since $(1+\theta^2)^m \le e^{m\theta^2}$, the stated energy condition implies $\chi^2 \le q_3^{-1/2}$. Cauchy–Schwarz gives $\mathrm{TV}(\overline{P}, P_0) \le \frac{1}{2}\sqrt{\chi^2} \le \frac{1}{2}q_3^{-1/4}$. The expectation of any test taking values in $[0,1]$ differs by at most total variation under these two laws, proving the power bound. At least one alternative has power no larger than that average. □

For a known split at $n/2$, set $m = n/2$. The maximum of the q_3 coordinate CUSUMs has a Hoeffding threshold of order $\sqrt{\log(q_3/\delta)}$, while its changed coordinate has population magnitude $\sqrt{n}|\theta|/2$. Hence $n\theta^2 \ge C\log(q_3/\delta)$ suffices for controlled error probabilities. For an unknown location η, the CUSUM at the true split has magnitude $|\theta|\sqrt{\eta(n-\eta)/n}$. A union bound over at most $n-1$ positions therefore gives the sufficient condition

$$\frac{\eta(n-\eta)}{n}\theta^2 \ge C\log(q_3 n/\delta). \tag{73}$$

For $\eta/n \in [\beta_0, 1-\beta_0]$ with fixed $\beta_0 \in (0, 1/2]$, this reduces to $n\theta^2 \ge C_{\beta_0}\log(q_3 n/\delta)$. To make the error probabilities explicit, put $u = b_0\sqrt{2\log\{2q_3(n-1)/\delta\}}$, where $b_0 = 3\sqrt{3}$. Each normalized

CUSUM has squared weight norm one. Weighted Hoeffding and a union bound give simultaneous centered error at most u with probability at least $1-\delta$, under both the null and these alternatives. Rejecting when the maximum exceeds u has size at most δ. If the true-split magnitude is at least $3u$, its observed magnitude is at least $2u$, giving power at least $1-\delta$. Thus $18b_0^2\log\{2q_3(n-1)/\delta\}$ is an explicit sufficient right-hand side in (73). For the known midpoint, the same argument omits the factor $n-1$ from the threshold.

The known-position lower and upper bounds match the logarithmic dependence at fixed error probabilities for this one-coordinate alternative family. For unknown positions the displayed scan bound provides a sufficient condition, and the learned-direction guarantee is given separately by Theorem A.5. Every member of this family has the same degree-at-most-two projection as the null.

Proposition F.2 (A two-location lower bound)**.** *Let $f_0 = 1$ and $f_\kappa = 1+\kappa\phi_3(x_1)$ with $0 < \kappa \le 1/(2\sqrt{7})$. Suppose two admissible change locations differ by $a = \lfloor c_0/\kappa^2 \rfloor$, where $c_0 > 0$ is a sufficiently small universal constant, and both locations satisfy the desired segment-length restrictions. For sufficiently small κ, every location estimator satisfies*

$$\max_{\ell\in\{0,1\}} \mathbb{P}_\ell\left\{|\widehat{\eta}-\eta_\ell| \ge \frac{c_0}{4\kappa^2}\right\} \ge \frac{1}{4}.$$

Consequently its maximum expected absolute error on this two-point family is at least $c_0/(16\kappa^2)$.

Proof. Density validity follows from $|\kappa\phi_3| \le 1/2$. The jump is $D = \kappa E_{3e_1}$ and has Frobenius norm κ, while its degree-at-most-two projection is zero. For $|z| \le 1/2$, $(1+z)\log(1+z) \le z + 2z^2$, so

$$\mathrm{KL}(f_\kappa \| f_0) \le \kappa \int \phi_3 \, d\mu + 2\kappa^2 \int \phi_3^2 \, d\mu = 2\kappa^2.$$

The two sequence laws differ on only a observations. In the orientation from the earlier change to the later change, their KL divergence is at most $2a\kappa^2$. Pinsker's inequality bounds their total variation by $\sqrt{a\kappa^2} \le \sqrt{c_0} \le 1/2$. An estimator induces a test by choosing the nearer candidate location, with deterministic ties. A testing error entails location error at least $a/2$. The sum of the test's two error probabilities is at least one minus total variation, so one model has location-error probability at least $1/4$ at threshold $a/2$. If $\kappa^2 \le c_0/2$, then $a/2 \ge c_0/(4\kappa^2)$. Multiplying the threshold by the positive probability gives the expected-loss statement. □

The admissibility requirement is essential: the observation horizon and allowed segment lengths must accommodate two locations this far apart. This is a probability and expected-loss lower bound on a pure cubic subclass. Comparing it with (68) requires g comparable to κ, successful pilot and anchor conditions, bounded L, and the same parameter regime. Under these conditions, the comparison identifies the same inverse-squared-jump dependence in the upper and lower localization bounds.

F.2 A DETERMINISTIC INTERFACE FOR MULTIPLE CHANGES

A uniform score bound also supports a deterministic extension to several changes. The construction uses seeded segmentation (Kovács et al., 2023) with explicit boundary padding. Its guarantee concerns segmentation on that uniform event; recursive testing additionally requires a suitable calibration rule.

Let Ψ be a fixed seminorm on the mean feature tensor. For piecewise constant means, write D_k for the jumps and $\gamma_k = \Psi(D_k) > 0$. Let Δ be the minimum segment length, including the first and last segments. Choose an integer $h_0 \ge 8$ divisible by eight. For lengths $\ell = \min(2^j h_0, n)$, start seeds on the grid of step $\lfloor \ell/2 \rfloor$ and also at $n-\ell$. Remove duplicates. Scan each seed's central half. Within a recursive segment retain contained seeds, choose a shortest seed whose computed score strictly exceeds a threshold λ, pad its endpoints by $h_0/8$ subject to the segment boundaries, and rescan every integer split at distance at least $h_0/8$ from the padded endpoints. Report a maximizer and recurse.

Proposition F.3 (Deterministic multiple-change interface)**.** *Suppose the computed scores differ from their population seminorm scores by at most ϵ on every seeded triple and every integer triple with length in $[h_0, 5h_0/4]$ and endpoint distances at least $h_0/8$. Let $\gamma_* = \min_k \gamma_k$. If*

$$h_0 \le 3\Delta/4, \qquad \gamma_*\sqrt{h_0} \ge 1600\epsilon, \qquad 100\epsilon \le \lambda \le \gamma_*\sqrt{h_0}/8,$$

the procedure returns exactly K changes and its ordered estimates satisfy

$$|\widehat{\eta}_k - \eta_k| \le 24\sqrt{h_0}\epsilon/\gamma_k =: R_k < h_0/8.$$

For $K = 0$, the uniform event prevents detection whenever $\lambda \ge \epsilon$.

Proof. If a remaining change η has distance at least $3h_0/4$ to both recursive boundaries, the grid start $s = (h_0/2)\lfloor(\eta - h_0/4)/(h_0/2)\rfloor$ gives a contained seed $(s, s + h_0]$ with $h_0/4 \le \eta - s < 3h_0/4$. It isolates the change because $h_0 < \Delta$.

A seed with no remaining change can still contain observations misclassified at an old recursive boundary. Let their counts be R_L, R_R and adjacent jumps be D_L, D_R. Each central CUSUM weight has absolute value at most $2/\sqrt{h_0}$. Subadditivity gives population score at most

$$2\{R_L\Psi(D_L) + R_R\Psi(D_R)\}/\sqrt{h_0}.$$

Assume inductively that every reported change has error at most R_k and leaves remaining changes at distance at least $3h_0/4$ from recursive boundaries. The preceding score is then at most 96ϵ. Its empirical value is at most $97\epsilon \le \lambda$, so it is inactive. Conversely, the isolating seed of each remaining change has population score at least $\gamma_k\sqrt{h_0}/4$. Its empirical score exceeds λ. Hence a shortest active seed has length h_0 and contains a remaining change.

The padded interval has length at most $5h_0/4 < \Delta$ and contains only that true change. Padding places it at least $h_0/8$ from each endpoint; a clipped boundary has the stronger inductive margin. Its balance is at least $1/10$. Applying the single-change score inequality in Appendix D.3 gives error at most $20\sqrt{5h_0/4}\epsilon/\gamma_k \le R_k$. Moreover $R_k \le 24h_0/1600 < h_0/8$. Every remaining change is at distance at least $\Delta - R_k \ge 3h_0/4$ from the new boundary, and any new misclassification is confined to the adjacent segment. This preserves the induction. After K splits all remaining seeds are inactive. The initial margins hold because $\Delta \ge 4h_0/3$. □

A finite family of scoring seminorms Ψ_r with deterministic uniform error bounds $e_r > 0$ can be combined as $\max_r \Psi_r/e_r$. Its uniform error is at most one. The proposition then applies to its retained jump. Establishing such simultaneous bounds for adaptively selected recursive intervals, and allocating false-alarm probability across multiple reported changes, are additional requirements for an inferential extension of the present algorithm.

G Covariance concentration and dependent inference

The unfolding identities in Lemma A.4 determine the marginal variance. We first control temporal cross moments at the same dimension scale, then apply dependent matrix concentration to every candidate CUSUM. The resulting event supplies the pilot and approximation premises of Corollary 5.2. The final subsection states the coupling and thinning results proved in Appendix E.

G.1 A fourth moment for cubic polynomials

Lemma G.1 (Cubic fourth moment)**.** *Every polynomial Q of total degree at most three satisfies*

$$\|Q\|_{L^4(\mu)} \le K_4\|Q\|_{L^2(\mu)}, \qquad K_4 = 3^{3/2}. \tag{74}$$

Proof. Realize each uniform coordinate on $[-1, 1]$ as the third coordinate of an independent uniform point on S^2. The Legendre polynomial ϕ_j becomes a zonal spherical harmonic of degree j. The spherical Poisson operator multiplies each degree-j harmonic by r^j. Beckner's hypercontractive inequality gives its $L^2 \to L^4$ contraction for $r = 1/\sqrt{3}$; see Montanaro (2012, Theorem 7) for this statement and Montanaro (2012, Lemma 10) for tensorization. Apply the product operator to $\widetilde{Q} = \sum_{|\alpha|\le 3} r^{-|\alpha|}Q_\alpha\psi_\alpha$. It returns Q, while orthogonality gives $\|\widetilde{Q}\|_2^2 = \sum_\alpha r^{-2|\alpha|}|Q_\alpha|^2 \le r^{-6}\|Q\|_2^2$. The lifted and interval norms agree. Therefore $\|Q\|_4 \le r^{-3}\|Q\|_2 = 3^{3/2}\|Q\|_2$. □

The reference geometry controls temporal covariance without treating the entries of an unfolding as independent. We transfer the polynomial fourth-moment bound to vector contractions, then

bound temporal cross moments by coupling. The resulting variance parameter retains its quadratic dependence on the feature dimension.

Write $D_{\mathrm{mat}} = p + p^2$, $M_i = H_3(X_i)_{(1)}$, $U_i = M_i - \mathbb{E}M_i$, and $B_3 = 6\sqrt{3q}$. The feature map is fixed and applied separately to each observation. Throughout this section, $f_i \le L$ with respect to μ, where $L \ge 1$. Let K_4 denote the constant in Lemma G.1, so that

$$\|Q\|_{L^4(\mu)} \le K_4 \|Q\|_{L^2(\mu)} \tag{75}$$

for every polynomial of total degree at most three. Lemma A.4 and the bounded feature envelope give

$$\mathbb{E}_\mu M M^\top = (q/p) I_p, \quad \mathbb{E}_\mu M^\top M = ((p+2)/3) P_{\mathrm{sym},2}, \quad \|U_i\|_{\mathrm{op}} \le B_3, \tag{76}$$

where $P_{\mathrm{sym},2}$ projects onto symmetric order-two tensors.

Normalize the absolute-regularity coefficient $\beta(\mathcal{A}, \mathcal{B})$ as the total variation distance between the joint law and the product of its marginals. For the finite sequence, let

$$\beta_X(h) = \sup_{1 \le j \le n-h} \beta\big(\sigma(X_1, \dots, X_j), \sigma(X_{j+h}, \dots, X_n)\big), \quad 1 \le h < n, \tag{77}$$

and set $\beta_X(h) = 0$ for $h \ge n$. Assume that, for given $B, c > 0$,

$$\beta_X(h) \le B e^{-ch}, \qquad h \ge 1. \tag{78}$$

This global condition includes dependence across changes in the marginal distribution. No stationarity is required. Dimension and sample-size comparisons below keep L, B, c fixed.

G.2 A VARIANCE BOUND FOR ARBITRARY WEIGHTED SUBSETS

For $a \in \mathbb{R}^p$, apply Minkowski's inequality to the squared vector norm. Each coordinate of $M_i^\top a$ is a polynomial of total degree at most three. Transferring its fourth moment to μ and then using (76) gives

$$\begin{aligned} \big\| \|M_i^\top a\|_2 \big\|_{L^4}^2 &= \left\| \sum_{j=1}^{p^2} (a^\top M_i e_j)^2 \right\|_{L^2} \le \sum_{j=1}^{p^2} \|a^\top M_i e_j\|_{L^4}^2 \\ &\le K_4^2 \sqrt{L} \sum_{j=1}^{p^2} \|a^\top M e_j\|_{L^2(\mu)}^2 = K_4^2 \sqrt{L}\,(q/p)\|a\|_2^2. \end{aligned} \tag{79}$$

The same calculation for $M_i b$, with $b \in \mathbb{R}^{p^2}$, uses the second reference moment. Subtracting the mean costs at most a factor two by Jensen's inequality. Thus

$$\begin{aligned} \big\| \|U_i^\top a\|_2 \big\|_{L^4} &\le 2K_4 L^{1/4} \sqrt{q/p}\, \|a\|_2, \\ \big\| \|U_i b\|_2 \big\|_{L^4} &\le 2K_4 L^{1/4} \sqrt{(p+2)/3}\, \|b\|_2. \end{aligned} \tag{80}$$

Norms of random variables without a measure subscript use the actual marginal law of X_i.

Proof of Lemma 5.1. For $i < j$, Berbee's coupling provides X_j^* with the same marginal law as X_j, independent of X_i, and

$$\mathbb{P}(X_j^* \ne X_j) = \beta(\sigma(X_i), \sigma(X_j)) \le \beta_X(j - i).$$

This coupling is used separately for each pair, on an enlarged probability space if necessary. Set $U_j^* = H_3(X_j^*)_{(1)} - \mathbb{E}M_j$. For unit $a, b \in \mathbb{R}^p$, centering and independence imply

$$a^\top \mathbb{E}(U_i U_j^\top) b = \mathbb{E}\langle U_i^\top a, (U_j - U_j^*)^\top b \rangle.$$

The integrand vanishes when $X_j = X_j^*$. Hölder's inequality with exponents $4, 4, 2$ and (80) therefore yield

$$\begin{aligned} \|\mathbb{E} U_i U_j^\top\|_{\mathrm{op}} &\le 8K_4^2 \sqrt{L}\,(q/p)\sqrt{\beta_X(j-i)}, \\ \|\mathbb{E} U_i^\top U_j\|_{\mathrm{op}} &\le 8K_4^2 \sqrt{L}\,((p+2)/3)\sqrt{\beta_X(j-i)}. \end{aligned} \tag{81}$$

For the first bound the two L^4 factors are at most $2K_4L^{1/4}\sqrt{q/p}$ and $4K_4L^{1/4}\sqrt{q/p}$. The second bound follows by the same argument with vectors in $\mathbb{R}^{p^2}$. The coupling statement also appears as Lemma 6 of Banna et al. (2016).

Integrating (76) against $f_i \le L$ preserves its positive-semidefinite upper bounds. Centering subtracts a positive-semidefinite matrix. Both diagonal second moments are therefore bounded by Lq/p, since $(p+2)/3 \le q/p$.

Extend the weights by zero outside K and outside $\{1,\dots,n\}$. For each $h \ge 1$, Cauchy–Schwarz gives $\sum_i |w_i w_{i+h}| \le \sum_i w_i^2$. Expand $\mathbb{E}S_{K,w}S_{K,w}^\top$ and pair the two temporal orders of every cross term. Equation (81) bounds its operator norm by

$$\frac{q}{p}\left[L + 16K_4^2\sqrt{L}\sum_{h\ge1}\sqrt{\beta_X(h)}\right]\sum_i w_i^2.$$

The other second moment has the same bound. Finally, $\sum_{h\ge1}\sqrt{\beta_X(h)} \le \sqrt{B}\sum_{h\ge1}e^{-ch/2} = \sqrt{B}/(e^{c/2}-1)$. The argument permits arbitrary signs and arbitrary subsets. □

G.3 An explicit matrix concentration threshold

Theorem 1 of Banna et al. (2016) states a matrix tail inequality with an unspecified universal constant. Their equation (60) supplies an explicit log-Laplace bound. We use that bound to obtain numerical threshold coefficients. Set

$$\gamma(c_0,\ell) = \frac{\log\ell}{\log 2}\max\left\{2, \frac{32\log\ell}{c_0\log 2}\right\}, \qquad c_0>0, \quad \ell\ge 2. \tag{82}$$

Lemma G.2 (Explicit BMY threshold)**.** *Let $A_1,\dots,A_\ell$ be centered self-adjoint matrices of dimension D, with $\ell \ge 2$, $M, v \ge 0$, and $\lambda_{\max}(A_i) \le M$ almost surely. Suppose their absolute-regularity coefficients, with the future starting at $j+h$, are at most $e^{-c_0(h-1)}$ for $h \ge 1$, and*

$$\sup_{\varnothing\ne K\subseteq\{1,\dots,\ell\}} |K|^{-1}\lambda_{\max}\mathbb{E}\left(\sum_{i\in K}A_i\right)^2 \le v^2.$$

For every $x > 0$,

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_{i=1}^{\ell}A_i\right) > 30v\sqrt{\ell x} + 4M\sqrt{x/c_0} + M\gamma(c_0,\ell)x\right\} \le De^{-x}. \tag{83}$$

Proof. The coefficients in Banna et al. (2016) use future indices $i \ge j+k$. With $h=k$, their convention and required bound $e^{-c_0(k-1)}$ agree with the assumption here. For a finite sequence, append deterministic zero matrices. The sigma-fields beyond its endpoint are trivial, so the extension preserves the mixing bound. A subset containing appended zeros has the same sum as its nonzero part and a larger denominator in the variance proxy. Centering and eigenvalue boundedness are preserved as well.

Equation (60) of that paper gives

$$\log\mathbb{E}\,\mathrm{tr}\exp\left(t\sum_{i=1}^{\ell}A_i\right) \le \log D + \frac{t^2a^2}{1-tb}, \quad a = 15v\sqrt{\ell} + 2M/\sqrt{c_0}, \quad b = M\gamma(c_0,\ell),$$

for $0 < t < 1/b$. If $a = 0$, the variance bound for singleton subsets makes every matrix zero almost surely. Otherwise take $t = \sqrt{x}/(a+b\sqrt{x})$ in the matrix Chernoff inequality. This choice lies in the allowed domain and satisfies

$$-t(2a\sqrt{x}+bx) + \frac{t^2a^2}{1-tb} = -x.$$

Expanding the threshold proves the claim. If $b = 0$, the same argument uses the unrestricted positive domain for t. □

For a rectangular U, its self-adjoint dilation is

$$\mathcal{D}(U) = \begin{pmatrix} 0 & U \\ U^\top & 0 \end{pmatrix}, \qquad \lambda_{\max}(\mathcal{D}(U)) = \|U\|_{\mathrm{op}}, \qquad \mathcal{D}(U)^2 = \mathrm{diag}(UU^\top, U^\top U). \tag{84}$$

Lemma 5.1, with weights equal to one on a subset, bounds the BMY variance proxy of a dilated subsequence by V_3^2. A single application of Lemma G.2 thus controls its rectangular operator norm with dimension factor D_{mat}.

G.4 A SIMULTANEOUS CUSUM BOUND

A fixed partition into residue classes puts the general geometric envelope into the form required by BMY and retains every observation. Define

$$a_{\mathrm{mix}} = 1 + \left\lceil \frac{\log \max\{B, 1\}}{c} \right\rceil, \qquad c_* = a_{\mathrm{mix}} c. \tag{85}$$

A gap of h indices within a residue class corresponds to a time gap $a_{\mathrm{mix}} h$. Sigma-field inclusion gives

$$\beta_{\mathrm{residue}}(h) \le B e^{-c a_{\mathrm{mix}} h} \le e^{-c_*(h-1)}. \tag{86}$$

The two constants in (85) depend only on B, c.

Let $\mathcal{T}$ be a deterministic set of $M_0 \ge 1$ triples (s, t, e) with $0 \le s < t < e \le n$. For each triple, write

$$n_L = t - s, \quad n_R = e - t, \quad m = n_L + n_R, \qquad w_L = \sqrt{\frac{n_R}{m n_L}}, \quad w_R = \sqrt{\frac{n_L}{m n_R}}.$$

Define the empirical and population contrasts by

$$\widehat{C}^t_{s,e} = -w_L \sum_{i=s+1}^{t} H_3(X_i) + w_R \sum_{i=t+1}^{e} H_3(X_i), \qquad C^t_{s,e} = \mathbb{E}\widehat{C}^t_{s,e}.$$

The weights sum to zero. Its centered CUSUM unfolding is

$$\begin{aligned} E^{s,e}_{\mathrm{CU}}(t) &= (\widehat{C}^t_{s,e} - C^t_{s,e})_{(1)} = -w_L \sum_{i=s+1}^{t} U_i + w_R \sum_{i=t+1}^{e} U_i, \\ w_{\max} &= \max\{w_L, w_R\}. \end{aligned} \tag{87}$$

These weights satisfy $-n_L w_L + n_R w_R = 0$ and $n_L w_L^2 + n_R w_R^2 = 1$, and $w_L + w_R \le 2 w_{\max}$.

Theorem G.3 (Uniform concentration for dependent CUSUMs)**.** *Under* (75)–(78)*, take* $\delta \in (0, 1)$ *and*

$$x = \log \frac{2 a_{\mathrm{mix}} M_0 D_{\mathrm{mat}}}{\delta} \ge 1. \tag{88}$$

Set $F_0(x) = 0$*,* $F_1(x) = B_3$*, and, for* $\ell \ge 2$*,*

$$F_\ell(x) = \min\left\{ \ell B_3, 30 V_3 \sqrt{\ell x} + 4 B_3 \sqrt{x/c_*} + B_3 \gamma(c_*, \ell) x \right\}. \tag{89}$$

Let $\ell_{L,r}$ *and* $\ell_{R,r}$ *count indices in* $(s, t]$ *and* $(t, e]$ *belonging to residue class* r *modulo* a_{mix}*. With probability at least* $1 - \delta$*, every triple in* $\mathcal{T}$ *satisfies*

$$\|E^{s,e}_{\mathrm{CU}}(t)\|_{\mathrm{op}} \le U^{s,e}_{\mathrm{BMY}}(t; \delta) := w_L \sum_{r=1}^{a_{\mathrm{mix}}} F_{\ell_{L,r}}(x) + w_R \sum_{r=1}^{a_{\mathrm{mix}}} F_{\ell_{R,r}}(x), \tag{90}$$

$$U^{s,e}_{\mathrm{BMY}}(t; \delta) \le 30 \sqrt{2 a_{\mathrm{mix}}}\, V_3 \sqrt{x} + 2 a_{\mathrm{mix}} B_3 w_{\max} \left[4 \sqrt{x/c_*} + \gamma(c_*, \max\{2, m\}) x \right]. \tag{91}$$

Proof. Partition both sides of each triple into residue classes. For each group of length at least two, apply Lemma G.2 to its dilated centered unfoldings. Equation (86) verifies mixing. Lemma 5.1 verifies the variance bound for every subset of that group, even though its original indices are separated. Its summand eigenvalue bound is B_3. The group norm is bounded by the second entry

of (89), except on an event of probability $D_{\mathrm{mat}}e^{-x}$. The deterministic bound ℓB_3 also holds, so their minimum is valid.

An empty group contributes zero. A singleton has norm at most B_3 without a probability argument. There are at most $2a_{\mathrm{mix}}M_0$ groups requiring a tail bound, counting possible repetitions across triples. A union bound gives one event of probability at least $1-\delta$. Triangle inequalities within each side and between the sides prove (90). No independence between different groups or sides is used.

For every nonempty group, its threshold is at most

$$30V_3\sqrt{\ell x} + 4B_3\sqrt{x/c_*} + B_3\gamma(c_*, \max\{2, m\})x.$$

For a singleton this follows from $x \geq 1$ and $\gamma(c_*, 2) \geq 2$. For larger groups it follows from monotonicity of γ. Each side has at most a_{mix} nonempty groups, and

$$\sum_r \sqrt{\ell_{L,r}} \leq \sqrt{a_{\mathrm{mix}} n_L}, \qquad \sum_r \sqrt{\ell_{R,r}} \leq \sqrt{a_{\mathrm{mix}} n_R}.$$

The combined leading term is therefore at most

$$30V_3\sqrt{a_{\mathrm{mix}}x}\left(\sqrt{n_R/m} + \sqrt{n_L/m}\right) \leq 30\sqrt{2a_{\mathrm{mix}}}\, V_3\sqrt{x}.$$

The remaining terms are multiplied by at most $a_{\mathrm{mix}}(w_L + w_R) \leq 2a_{\mathrm{mix}}w_{\max}$. This proves (91), including $m = 2$. □

The leading variance term has no dependence on the balance of the two sides. Their balance remains visible through $w_{\max}$ in the envelope term. The separate left and right sums use the variance proxy for unweighted groups; this keeps the leading term proportional to the normalized CUSUM weight norm.

Corollary G.4 (Dimension and window scales)**.** *For fixed L, B, c, K_4, the event in Theorem G.3 implies*

$$\|E^{s,e}_{\mathrm{CU}}(t)\|_{\mathrm{op}} \leq C\left[p\sqrt{x} + p^{3/2}w_{\max}x\log^2(em)\right] \tag{92}$$

for all candidates, with C depending only on those constants. In a local window with $m \asymp h$, $w_{\max} \lesssim h^{-1/2}$, and population contrast coefficient $a_{\mathrm{CU}} \gtrsim \sqrt{h}$, the stochastic part of the normalized rank-r unfolding approximation is at most

$$C\sqrt{r}\left[p\sqrt{x/h} + p^{3/2}x\log^2(eh)/h\right]. \tag{93}$$

The constants in the window comparisons are included in C.

Proof. Since $q/p = (p+1)(p+2)/6$, equations (15) and (76) give $V_3 \leq Cp$ and $B_3 \leq Cp^{3/2}$. For fixed c_*, $\gamma(c_*, m) \leq C\log^2(em)$ for $m \geq 2$. Since $x \geq 1$, the term $\sqrt{x/c_*}$ is absorbed into $Cx\log^2(em)$. This proves (92). The deterministic rank-truncation inequality multiplies operator noise by a constant times $\sqrt{r}$; normalization divides it by a_{CU}. The stated window conditions yield (93). □

The condition on a_{CU} concerns the true change within the window. Balanced candidate side lengths alone do not ensure it. The ratio of the envelope term to the variance term in (93) is bounded by a constant times $\sqrt{px/h}\log^2(eh)$. Thus $h \gg px\log^4(eh)$ is a sufficient asymptotic regime for the variance term to dominate. The full bound retains both terms at every window length.

The thresholds in (89) are explicit once justified values of L, B, c, K_4 are supplied. The event controls estimation and data-dependent selection within the fixed family $\mathcal{T}$. Using it in a test also requires the assumptions and calibration argument for that test.

G.5 One event for the spectral pilot and fitted direction

Proof of Corollary 5.2. Theorem G.3 supplies an event of probability at least $1-\delta$ on which every declared unfolding error is at most U_*. The full-interval raw operator score therefore has uniform error at most U_* and population value $a(t)\sqrt{\mu_1}$. The candidate set contains η because both segments have length at least $2h$.

Put $t_* = 8\ell_*^2$ and $S = \sum_j \min\{\mu_j, t_*\}$. Since $S \le \xi_*\kappa^2$ with $\xi_* < 1$, we have $\mu_1 > t_*$. For every j, $\min\{\mu_j, t_*\} \ge t_*\mu_j/\mu_1$. Summing and using $a_0 = 15\sqrt{h}/(16\sqrt{2})$ gives

$$\frac{\sqrt{\mu_1}}{U_*} \ge \frac{\sqrt{8}}{a_0\sqrt{\xi_*}} = \frac{64}{15\sqrt{h\xi_*}}, \qquad \beta h\sqrt{\mu_1} \ge 32\sqrt{n}\,U_*.$$

The last step uses $\xi_* \le 4\beta^2 h/(225n)$. The population margin in (66), with score $\Psi(T) = \|T_{(1)}\|_{\mathrm{op}}$ and interval length n, now yields $|\widehat{c}-\eta| \le 2\sqrt{n}\,U_*/(\beta\sqrt{\mu_1}) \le h/16$.

The selected local contrast has amplitude $a(\widehat{c}) = (h - |\widehat{c}-\eta|)/\sqrt{2h} \ge a_0$. It belongs to the same predetermined family and hence obeys the same noise bound. The condition $4S \le \rho^2\kappa^2$ is precisely the dyadic condition of Corollary A.2, with $U = U_*$ and $\ell = \ell_*$. That result supplies one fitted direction with jump at least $\kappa\sqrt{1-\rho^2}$ on the same event. This proof uses neither conditional independence after the pilot nor the three-fold validation argument. □

G.6 Analytic detection on the concentration event

Fix $1 \le r \le p$ and write $S_r(T) = \{\sum_{j\le r}\sigma_j(T)^2\}^{1/2}$. For any simultaneous bound $U_\delta(w)$ from Theorem G.3, the inequality

$$|S_r(T+E) - S_r(T)| \le \sqrt{r}\|E\|_{\mathrm{op}}$$

follows from singular-value perturbation and the Euclidean triangle inequality. Under a constant feature mean, all population CUSUMs vanish. Rejecting if any $S_r(\widehat{C}(w)_{(1)}) > \sqrt{r}\,U_\alpha(w)$ therefore has probability at most α. If a declared contrast has mean aD, $a > 0$, its population score is $a\kappa_r$, where $\kappa_r^2 = \sum_{j\le r}\mu_j$. The same event at failure level δ gives power at least $1-\delta$ whenever

$$a\kappa_r > \sqrt{r}\{U_\alpha(w) + U_\delta(w)\}.$$

The preliminary-location bound (66) applies to the raw score with noise $\sqrt{r}\max_w U_\delta(w)$ and jump κ_r on a single-change interval. Thresholding each cut and maximizing the raw score serve different purposes; the location argument uses the raw population CUSUM geometry.

G.7 Coupling concentration and calibrated thinning

The following results use the global absolute-regularity coefficient defined in Section 5. Their statements allow general decay of that coefficient. Their proofs are given in Appendix E. Set $b_0 = 3\sqrt{3}$. For a deterministic contrast w, write $E_w = \sum_i w_i\{H_3(X_i) - \mathbb{E}H_3(X_i)\}$ and take $\zeta \in (0,1)$. The next bound supplies the noise term for spectral approximation using all observations.

Theorem G.5 (Dependent unfolding noise and spectral approximation)**.** *Let $\mathcal{W}$ contain M_0 deterministic weight vectors in $\mathbb{R}^n$, let $1 \le k \le n$, and put $x = \log\{2M_0k(p+p^2+2q)/\zeta\}$. With probability at least $1-\zeta-(n-k)\beta_X(k)$, simultaneously for $w \in \mathcal{W}$,*

$$\|(E_w)_{(1)}\|_{\mathrm{op}} \le U_k(w) := \sqrt{2L(q/p)kx}\|w\|_2 + \frac{4b_0\sqrt{q}}{3}kx\|w\|_\infty, \tag{94}$$

$$\max_\alpha |\langle E_w, E_\alpha\rangle| \le b_0\sqrt{2kx}\|w\|_2.$$

Consequently, for any contrast with mean aD, $a > 0$, its symmetrized rank-r SVD estimate satisfies $\|B_r - aD\|_{\mathrm{F}} \le a(\sum_{j>r}\mu_j)^{1/2} + 2\sqrt{2r}U_k(w)$.

To prove the bound, partition time into k residue classes. Each class is compared once with its independent product law, after controlling every contrast on that class. The comparison costs sum to $(n-k)\beta_X(k)$, while the number of contrasts enters only through x. Combining the classwise matrix

bounds produces the factors $\sqrt{k}$ and k in the variance and envelope terms. This argument retains all observations.

For a balanced local CUSUM, $a \asymp \sqrt{h}$, its normalized error has order

$$\left(\sum_{j>r} \mu_j\right)^{1/2} + C\sqrt{\frac{rLqkx}{ph}} + C\frac{\sqrt{rq}\,kx}{h}. \tag{95}$$

The spectral tail is unchanged; dependence enters the variance and bounded-envelope terms. For $\beta_X(k) \le Be^{-ck}$, a gap of order $\log(n/\zeta)$ controls the coupling loss. The resulting uniform event also supplies a noise bound for the deterministic multiple-change construction in Appendix F.2.

Calibration requires a comparison of the inference procedure's joint data law. A deterministic retained grid provides that comparison with an explicit total-variation cost. It yields an adjusted p value and a localization radius measured in the original time scale.

Theorem G.6 (Calibrated inference after temporal thinning)**.** *Fix an integer $1 \le k \le n$ before observing data. Retain $X_{1+(j-1)k}$, $1 \le j \le m$, where $m = 3\lfloor(\lfloor(n-1)/k\rfloor + 1)/3\rfloor$, and assume $m = 3N$ with $N \ge 2h$, where $h \ge 16$ is divisible by eight. Apply the three-fold procedure to these observations. Let $\varepsilon_{\mathrm{mix}} = (m-1)\bar{\beta}_X(k)$. Under a common marginal null, $\mathbb{P}(p_{\mathrm{global}} \le u) \le u + \varepsilon_{\mathrm{mix}}$. If the primitive conditions of Theorem A.5 hold under the product of the retained marginal laws, rejection and*

$$|k\widetilde{\eta}_{\mathrm{retained}} - \eta| \le k\mathfrak{r}(g) + k - 1 \tag{96}$$

hold with probability at least $1 - \delta_P - \delta_V - \delta_A - \delta_C - \varepsilon_{\mathrm{mix}}$. The adjusted value $\min(1, p_{\mathrm{global}} + \varepsilon_{\mathrm{mix}})$ is super-uniform; power for that adjusted test uses level $\alpha_0 - \varepsilon_{\mathrm{mix}} > 0$ in the retained-sample theorem.

The proof transfers the whole randomized procedure under one total-variation comparison. Under the product law, Theorem A.5 applies; returning to the dependent law loses at most $\varepsilon_{\mathrm{mix}}$. The retained change count is $\lceil \eta/k \rceil$, which explains the additional $k-1$ in (96). Thus spacing improves the approximation to independence while reducing sample size and temporal resolution.

H A DENSE CUBIC MODEL SEPARATING THE SUFFICIENT CONDITIONS

The variance improvement has a consequence beyond its numerical constant. We construct bounded densities whose change is invisible to linear and quadratic features, spread over cubic coefficients, and close to an unfolding of rank one. A dependent process with these marginals then separates the sufficient conditions supplied by the two concentration arguments. The comparison concerns the displayed sufficient conditions for fitting and detection.

H.1 BOUNDED DENSITIES WITH A DENSE CUBIC PROJECTION

For $Z \sim \mathrm{Unif}[-1,1]$, define the real analytic functions

$$r_j(z) = \mathrm{i}^{-j}\mathbb{E}[\phi_j(Z)e^{\mathrm{i}zZ}], \qquad 0 \le j \le 3. \tag{97}$$

Parity makes these functions real on the real line. In particular $r_0(z) = \sin(z)/z$, with its continuous value at zero. Put $a = d^{-1/2}$ and

$$c_d = \frac{r_1(a)r_0(a)^{d-1}}{r_1(2a)r_0(2a)^{d-1}}, \qquad g_d(x) = \sin\left(a\sum_{j=1}^{d} x_j\right) - c_d \sin\left(2a\sum_{j=1}^{d} x_j\right). \tag{98}$$

The denominator is positive for all sufficiently large d. Write

$$T_d = \mathbb{E}_\mu[g_d(X)H_3(X)], \qquad u_d = (0, d^{-1/2}, \dots, d^{-1/2}) \in \mathbb{R}^{d+1}. \tag{99}$$

The first coordinate of u_d corresponds to the constant tensor coordinate.

Proposition H.1 (A bounded perturbation with a dense cubic projection)**.** *For a sufficiently large fixed d_0, the functions above satisfy*

$$c_d \longrightarrow \tfrac{1}{2}e^{1/2}, \qquad \sup_{d \ge d_0} \|g_d\|_\infty < \infty, \qquad P_{\le 2} g_d = 0. \tag{100}$$

There are positive λ_d such that

$$\lambda_d \longrightarrow \frac{e^{-1/6}}{3\sqrt{2}}, \qquad \|T_d - \lambda_d u_d^{\otimes 3}\|_F = O(d^{-1/2}), \qquad \|T_d\|_F \longrightarrow \frac{e^{-1/6}}{3\sqrt{2}}. \tag{101}$$

Every normalized coefficient of T_d is $O(d^{-3/2})$. For any positive sequence $\kappa_d \to 0$, the densities

$$f_{\pm,d} = 1 \pm \theta_d g_d, \qquad \theta_d = \frac{\kappa_d}{2\|T_d\|_F} \tag{102}$$

are between $1/2$ and $3/2$ for all sufficiently large d. Their jump $D_d = A(f_{+,d}) - A(f_{-,d})$ satisfies

$$\|D_d\|_F = \kappa_d, \qquad \tau_1(D_d) = O(\kappa_d d^{-1/2}), \qquad m_d := \max_\alpha |\langle D_d, E_\alpha\rangle| = \Theta(\kappa_d d^{-3/2}), \tag{103}$$

where $\tau_r(D)^2 = \sum_{j>r} \sigma_j(D_{(1)})^2$. Here $P_{\le 2}$ is the orthogonal projection onto polynomials of total degree at most two. The two densities have identical projections onto this space.

Proof. Taylor expansion and Legendre orthogonality give

$$\begin{aligned} r_0(z) &= 1 - z^2/6 + O(z^4), & r_1(z) &= z/\sqrt{3} + O(z^3), \\ r_2(z) &= z^2/(3\sqrt{5}) + O(z^4), & r_3(z) &= z^3/(15\sqrt{7}) + O(z^5). \end{aligned} \tag{104}$$

The required first nonzero moments are

$$\mathbb{E}[\phi_1(Z)Z] = \frac{1}{\sqrt{3}}, \qquad \mathbb{E}[\phi_2(Z)Z^2] = \frac{2}{3\sqrt{5}}, \qquad \mathbb{E}[\phi_3(Z)Z^3] = \frac{2}{5\sqrt{7}}.$$

Multiplying by the coefficients of the exponential series proves the stated expansions. Also $\log r_0(z) = -z^2/6 + O(z^4)$. Substitution of $a = d^{-1/2}$ proves $c_d \to e^{1/2}/2$, and $1 + c_d$ uniformly bounds $\|g_d\|_\infty$ for large d.

The function g_d changes sign under $x \mapsto -x$. Every product Legendre polynomial of even total degree is invariant under this transformation, so its coefficient in g_d vanishes. All degree-one coefficients are equal by coordinate permutation symmetry. Their common value is $r_1(a)r_0(a)^{d-1} - c_d r_1(2a)r_0(2a)^{d-1} = 0$. This proves $P_{\le 2} g_d = 0$, including the constant coefficient.

Denote the common degree-three coefficients by A_d for three distinct indices, B_d for multiplicities $(2,1)$, and C_d for multiplicity (3). Set $F_d = r_1(a)r_0(a)^{d-1}$. Independence under μ gives

$$\mathbb{E}_\mu[\psi_\alpha(X) e^{\mathrm{i}z\sum_j X_j}] = \mathrm{i}^{|\alpha|} \prod_{j=1}^{d} r_{\alpha_j}(z). \tag{105}$$

Taking imaginary parts, using $\mathrm{i}^3 = -\mathrm{i}$, and substituting c_d yields

$$\begin{aligned} A_d &= F_d \left\{ \left(\frac{r_1(2a)}{r_0(2a)}\right)^2 - \left(\frac{r_1(a)}{r_0(a)}\right)^2 \right\}, \\ B_d &= F_d \left\{ \frac{r_2(2a)}{r_0(2a)} - \frac{r_2(a)}{r_0(a)} \right\}, \\ C_d &= F_d \left\{ \frac{r_3(2a)}{r_1(2a)} - \frac{r_3(a)}{r_1(a)} \right\}. \end{aligned} \tag{106}$$

Since $F_d = e^{-1/6} d^{-1/2}/\sqrt{3}\{1 + O(d^{-1})\}$, these identities imply

$$\begin{aligned} A_d &= \frac{e^{-1/6}}{\sqrt{3}} d^{-3/2}\{1 + O(d^{-1})\}, \\ B_d &= \frac{e^{-1/6}}{\sqrt{15}} d^{-3/2}\{1 + O(d^{-1})\}, \\ C_d &= \frac{e^{-1/6}}{5\sqrt{7}} d^{-3/2}\{1 + O(d^{-1})\}. \end{aligned} \tag{107}$$

The constants do not depend on the coordinate indices.

Choose $\lambda_d = A_d d^{3/2}/\sqrt{6}$. The normalized coefficients of $u_d^{\otimes 3}$ for these three patterns are respectively $\sqrt{6}d^{-3/2}$, $\sqrt{3}d^{-3/2}$, and $d^{-3/2}$. Thus all $\binom{d}{3}$ distinct-index coefficients agree exactly between T_d and $\lambda_d u_d^{\otimes 3}$. Both tensors have zero lower-degree coefficients. Orthonormality gives

$$\|T_d - \lambda_d u_d^{\otimes 3}\|_F^2 = d(d-1)(B_d - A_d/\sqrt{2})^2 + d(C_d - A_d/\sqrt{6})^2 = O(d^{-1}). \tag{108}$$

This proves (101), since $\|u_d^{\otimes 3}\|_F = 1$. The unfolding of $\lambda_d u_d^{\otimes 3}$ has rank one, so the same approximation error bounds $\tau_1(T_d)$.

Finally $\int g_d \, d\mu = 0$, $\theta_d \to 0$, and g_d is uniformly bounded. Hence (102) defines the stated probability densities. Linearity gives $D_d = 2\theta_d T_d$, and (103) follows from the coefficient expansions and the positive limiting norm. Equality of the lower-degree projections follows from $P_{\le 2} g_d = 0$. □

H.2 A DEPENDENT SEQUENCE AND THE COMPLETE NOISE SCALES

Fix $\varrho \in (0,1)$ and put $c = -\log \varrho$. In the first segment, start with density $f_{-,d}$. At each subsequent time, retain the entire preceding vector with probability ϱ, and otherwise draw an independent vector with density $f_{-,d}$. At the change, draw an independent vector with density $f_{+,d}$ and continue with the same refresh rule using that density. All refresh indicators, fresh draws, and the change reset are independent.

Every observation has the specified marginal. Adjacent observations within a segment coincide with probability ϱ, so the process is not independent. A refresh between times t and $t+k$ separates the past up to t from the future from $t+k$ onward, and a change reset does the same. The coupling upper bound is therefore $\beta_X(k) \le \varrho^k$, including across the change. When both segments have length $2h$, the exact global coefficient is

$$\beta_X(k) = \begin{cases} \varrho^k, & 1 \le k < 2h, \\ 0, & 2h \le k < n. \end{cases} \tag{109}$$

Indeed, for $k < 2h$, choose two observations at distance k in one segment. Their joint law gives their equality probability ϱ^k, whereas the product of their non-atomic marginals gives it probability zero. This matches the coupling upper bound. For $k \ge 2h$, every nonempty admissible past and future are separated by the independent reset. Thus the geometric mixing envelope is independent of d.

Set

$$h = 8\lceil d^3/8 \rceil, \qquad n = 4h, \qquad \eta = 2h, \qquad \kappa_d = d^{-1/2}(\log d)^{3/4}. \tag{110}$$

Include the full-interval CUSUM at every integer $t \in [h, n-h]$ and the balanced local contrast on $(t-h, t+h]$ at each such t. This is a predetermined family of $M_0 = O(n)$ contrasts containing the true split. Its weights have $\|w\|_2 = 1$ and $\|w\|_\infty \le h^{-1/2}$. Additional predetermined CUSUM contrasts satisfying these weight bounds may multiply the family size by $M_{\text{extra}} \le n^C$, for a fixed exponent C. At fixed $\delta \in (0, 1/4)$, let

$$x_d = \log\{C_0 M_0 M_{\text{extra}}(p + p^2)/\delta\} \asymp \log d. \tag{111}$$

Here $C_0 = 2$ suffices for the finite union bound in Theorem G.3, since $B = 1$ and $a_{\text{mix}} = 1$.

Use $L = 3/2$, $b_0 = 3\sqrt{3}$, $B_3 = 6\sqrt{3q}$, and

$$V_3^2 = \frac{q}{p}\left\{ L + 16K_4^2\sqrt{L}\,\frac{\sqrt{\varrho}}{1-\sqrt{\varrho}} \right\} \asymp d^2. \tag{112}$$

The constant K_4 is the dimension-independent cubic reference moment constant.

For this particular refresh process, one can use the smaller variance bound

$$V_{\text{refresh}}^2 = \frac{q}{p} L \frac{1+\varrho}{1-\varrho}. \tag{113}$$

To verify it, let $U_i = M_i - \mathbb{E}M_i$. Within one segment, $\mathbb{E}[U_j \mid X_i] = \varrho^{j-i} U_i$, so both lagged matrix cross moments equal ϱ^{j-i} times the corresponding diagonal second moment. Across the reset they are zero. Expanding an arbitrary signed weighted subset sum and using $\sum_i |w_i w_{i+k}| \le \sum_i w_i^2$ gives

the multiplier $1 + 2\sum_{k\geq 1} \varrho^k = (1+\varrho)/(1-\varrho)$. The proof of Theorem G.3 therefore also permits replacing V_3 by V_{refresh}. We use the general bound (112) below to establish the separation under the common model assumptions.

With

$$\gamma(c,n) = \frac{\log n}{\log 2} \max\left\{2, \frac{32\log n}{c\log 2}\right\}, \tag{114}$$

the two-piece matrix concentration bound gives simultaneous operator noise at most

$$U_{\mathrm{cov}} = C_1\left[V_3\sqrt{x_d} + \frac{B_3}{\sqrt{h}}\left\{\sqrt{x_d/c} + \gamma(c,n)x_d\right\}\right] \tag{115}$$

with probability at least $1-\delta$. We may take $C_0 = 2$ and $C_1 = 30\sqrt{2}$ by (91). These constants do not vary with d. Both the extra bounded-summand square term and the logarithmic envelope term remain present. The event covers the pilot and its selected local contrast.

For comparison, the residue-class argument with gap k and allocation ζ gives

$$\begin{aligned} x_{\mathrm{coup}} &= \log\{2M_0M_{\mathrm{extra}}k(p+p^2+2q)/\zeta\},\\ U_{\mathrm{coup}}(k,\zeta) &= \sqrt{2L(q/p)kx_{\mathrm{coup}}} + \frac{4b_0\sqrt{q}}{3\sqrt{h}}kx_{\mathrm{coup}},\\ \epsilon_{\mathrm{coup}}(k,\zeta) &= b_0\sqrt{2kx_{\mathrm{coup}}}. \end{aligned} \tag{116}$$

These formulas follow by comparing each of the k residue subsequences with its independent marginal product law. Couple each entire subsequence once. Summing its internal gaps across the nonempty classes gives total comparison cost at most $(n-k)\beta_X(k)$. The same coupled subsequences serve all predetermined contrasts, so this cost is not multiplied by their number. Within a class, rectangular matrix Bernstein (Tropp, 2012, Theorem 1.6) uses variance at most $L(q/p)\|w_{\mathrm{class}}\|_2^2$ and centered envelope $2b_0\sqrt{q}\|w_{\mathrm{class}}\|_\infty$. Scalar Hoeffding uses the coefficient range of length $2b_0$. Union bounds over the classes, contrasts and coefficients, followed by

$$\sum_{r=1}^{k}\|w_r\|_2 \leq \sqrt{k}, \qquad \sum_{r=1}^{k}\|w_r\|_\infty \leq k/\sqrt{h},$$

produce (116). No independence between the residue classes is required.

Its failure bound is $\zeta + (n-k)\beta_X(k)$. We allow every deterministic $1 \leq k \leq n$ and $\zeta > 0$ making that bound at most δ, with zero comparison cost at $k = n$.

H.3 Separation of approximation, pilot, and detection conditions

At the true local center, the population contrast is a_*D_d, where $a_* = \sqrt{h/2}$. Put $u_{\mathrm{cov}} = U_{\mathrm{cov}}/a_*$ and $u_{\mathrm{coup}} = U_{\mathrm{coup}}/a_*$. Let $t_s(D)$ denote the best error using at most s normalized coefficients from the complete basis. Theorem A.1 gives the normalized sufficient approximation bounds for basis truncation and symmetrized unfolding SVD

$$t_s(D_d) + 2\sqrt{2s}\epsilon/a_*, \qquad \tau_r(D_d) + 2\sqrt{2r}U/a_*. \tag{117}$$

An error bound at most $\rho_0\kappa_d$, for fixed $0 < \rho_0 < 1$, certifies retained jump at least $\kappa_d\sqrt{1-\rho_0^2}$.

Proof of Proposition 5.3. Since $q \asymp d^3$, $q/p \asymp d^2$, $h \asymp d^3$, and $x_d \asymp \log d$, the full covariance bound gives

$$u_{\mathrm{cov}} = O\{d^{-1/2}(\log d)^{1/2} + d^{-3/2}(\log d)^3\}, \qquad u_{\mathrm{cov}}/\kappa_d \longrightarrow 0. \tag{118}$$

The $\sqrt{x_d/c}$ contribution is included in the displayed remainder. Also $\tau_1(D_d)/\kappa_d = O(d^{-1/2})$, so the covariance rank-one expression is $o(\kappa_d)$.

If $k < 2h$, coupling admissibility implies $2h\varrho^k \leq \delta$, hence $k \geq c^{-1}\log(2h/\delta)$. If $k \geq 2h$, this order lower bound is immediate. Thus every admissible choice has $k \gtrsim \log d$ and, since $\zeta \leq \delta$, also $x_{\mathrm{coup}} \gtrsim \log d$. Uniformly over these choices,

$$u_{\mathrm{coup}} \gtrsim d^{-1/2}\log d, \qquad u_{\mathrm{coup}}/\kappa_d \longrightarrow \infty, \qquad \epsilon_{\mathrm{coup}} \gtrsim \log d. \tag{119}$$

For instance $\zeta = \delta/2$ and $k = \lceil c^{-1} \log(8h/\delta) \rceil$ are admissible for large d and give the complete scale

$$u_{\mathrm{coup}} = \Theta\{d^{-1/2} \log d + d^{-3/2} (\log d)^2\}. \tag{120}$$

The leading bounds therefore differ by order $\sqrt{\log d}$.

Every SVD expression with $r \geq 1$ is at least $2\sqrt{2}u_{\mathrm{coup}} \gg \kappa_d$. For basis truncation the triangle inequality gives $t_s(D_d) \geq \kappa_d - \sqrt{s}\, m_d$, so

$$t_s(D_d) + 2\sqrt{2s}\epsilon_{\mathrm{coup}}/a_* \geq \kappa_d + \sqrt{s}\{2\sqrt{2}\epsilon_{\mathrm{coup}}/a_* - m_d\} \geq \kappa_d \tag{121}$$

for all s simultaneously. Indeed $\epsilon_{\mathrm{coup}}/a_* \gtrsim d^{-3/2} \log d$, whereas $m_d = O(\kappa_d d^{-3/2})$ and $\kappa_d \to 0$. A restricted marginal basis has at least the complete-basis approximation error. The Tucker bound $[3\{\tau_r(D_d) + 2\sqrt{2r}u_{\mathrm{coup}}\}^2 + r u_{\mathrm{coup}}^2]^{1/2}$ and the full-fit bound $\sqrt{p}\, u_{\mathrm{coup}}$ also exceed κ_d. The alternative full-fit bound $\sqrt{q}\, \epsilon_{\mathrm{coup}}/a_* \gtrsim \log d$ exceeds κ_d as well, so taking the minimum of these two full-fit bounds does not change the conclusion. The comparison is between the explicit approximation bounds.

We verify pilot accuracy directly. Let $\sigma_d = \|D_{d,(1)}\|_{\mathrm{op}}$. The spectral-tail bound and Parseval give $\sigma_d/\kappa_d \to 1$. The full-interval population contrast is $a(t)D_d$, with

$$a(t) = \begin{cases} (n-\eta)n^{-1/2}\sqrt{t/(n-t)}, & t \leq \eta, \\ \eta n^{-1/2}\sqrt{(n-t)/t}, & t \geq \eta. \end{cases} \tag{122}$$

Its maximum is $a(\eta) = \sqrt{h}$. For any absolutely homogeneous score with jump score $\gamma_0 > 0$, put $G(t) = a(t)\gamma_0$. Direct subtraction gives

$$G(\eta)^2 - G(t)^2 \geq \tfrac{1}{2}\gamma_0^2 |t - \eta|. \tag{123}$$

If the computed score has uniform error e and $\widehat{c}$ maximizes it, then $G(\eta) - G(\widehat{c}) \leq 2e$ and $G(\eta) + G(\widehat{c}) \leq 2\sqrt{h}\gamma_0$. Therefore

$$|\widehat{c} - \eta| \leq 8\sqrt{h}\, e/\gamma_0. \tag{124}$$

For the raw operator pilot, $e = U_{\mathrm{cov}}$ and $\gamma_0 = \sigma_d$. Equation (118) makes this radius at most $h/16$ eventually. The selected local interval lies in $[0, n]$, and summing its two means gives

$$a(\widehat{c}) = \frac{h - |\widehat{c} - \eta|}{\sqrt{2h}} \geq \frac{15}{16} a_*. \tag{125}$$

The simultaneous event covers this contrast after selection. Its normalized rank-one error is at most $\tau_1(D_d) + (32\sqrt{2}/15)u_{\mathrm{cov}} = o(\kappa_d)$.

The coupling normalized pilot with every support size and rank available is

$$\Psi_{\mathrm{coup}}(T) = \max\left\{ \max_{1 \leq s \leq q} \frac{(\sum_{j \leq s} |T|_{(j)}^2)^{1/2}}{\sqrt{s}\, \epsilon_{\mathrm{coup}}}, \max_{1 \leq r \leq p} \frac{(\sum_{j \leq r} \sigma_j(T_{(1)})^2)^{1/2}}{\sqrt{r}\, U_{\mathrm{coup}}} \right\}, \tag{126}$$

where $|T|_{(j)}$ orders absolute normalized coefficients. Each normalized average is at most its first term, attained at size or rank one. Hence

$$\Psi_{\mathrm{coup}}(D_d) = \max\{m_d/\epsilon_{\mathrm{coup}}, \sigma_d/U_{\mathrm{coup}}\}, \qquad \sqrt{h}\, \Psi_{\mathrm{coup}}(D_d) \longrightarrow 0. \tag{127}$$

After multiplication by $\sqrt{h}$, the two terms are $O(\kappa_d/\log d)$ and $O(\kappa_d/u_{\mathrm{coup}})$. The coupling event has uniform normalized score error one. Its radius-$h/16$ sufficient condition from (124) is $\sqrt{h}\, \Psi_{\mathrm{coup}}(D_d) \geq 128$, which is violated eventually. Smaller grids cannot increase its population score.

Finally let $T_{\max} = \max_{t \in [h, n-h] \cap \mathbb{Z}} \|\widehat{C}^t_{0,n(1)}\|_{\mathrm{op}}$. Under a common marginal null, all population contrasts are zero, so the covariance event implies $T_{\max} \leq U_{\mathrm{cov}}$. Strict exceedance therefore has size at most δ. Under the alternative, the same event gives

$$T_{\max} \geq \sqrt{h}\, \sigma_d - U_{\mathrm{cov}} > U_{\mathrm{cov}} \tag{128}$$

for sufficiently large d, proving power at least $1 - \delta$. □

The dense cubic signal makes every displayed coupling coefficient bound at least the signal norm. Its small unfolding tail, combined with covariance concentration, instead certifies an informative fitted direction and an analytic detector using all observations. Theorem G.6 supplies calibrated selection and scalar refinement on a predetermined retained grid under its product-law conditions.

H.4 FINITE-SAMPLE THRESHOLD CONSTANTS

The asymptotic separation holds with fixed constants, so their size determines when the bounds become informative. For $L = 3/2$, $\beta_X(k) \le 2^{-k}$, and $K_4^2 = 27$, the covariance multiplier is

$$G = L + 16K_4^2\sqrt{L}\sum_{k\ge 1} 2^{-k/2} \approx 1278.84.$$

At the same tail parameter x, the leading covariance and coupling thresholds are $30\sqrt{2(q/p)Gx}$ and $\sqrt{2L(q/p)kx}$. The first is smaller when $k > 900G/L \approx 7.67 \times 10^5$. Substituting the process-specific variance (113) changes this leading-term comparison to $k > 2700$. The full thresholds also contain their respective envelope corrections. These constants quantify the potential conservatism of direct analytic calibration, even though the covariance condition is asymptotically informative in Proposition 5.3.

I EXPERIMENTAL AND IMPLEMENTATION DETAILS

The main experiments relate structural choices to direction quality and inference. This appendix provides the data-generating models, computation, diagnostics, and every result needed to assess that comparison.

I.1 POPULATION GEOMETRY OF THE DENSE CONSTRUCTION

We evaluate the two-frequency construction in Proposition H.1 at every $d \in \{4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096\}$. These are deterministic population calculations. They complement the proof by showing the size of the individual coefficients and the approximation error across a fixed dimension grid. Figure 2 distinguishes the error of the explicit tensor $\lambda_d u_d^{\otimes 3}$ from the best rank-one error of the unfolding.

The three coefficient patterns in (106) have multiplicities $\binom{d}{3}$, $d(d-1)$ and d. Summing every pattern gives the exact squared projection norm. We evaluate the Fourier coefficients at 100 and 160 decimal digits using their convergent hypergeometric formulas. Independently integrating all four Legendre polynomials at both frequencies gives 88 integral checks across the grid. For every grid dimension at most 32, we also form the complete tensor and compare all entries, norms and the full Gram spectrum with the pattern formulas. The supplementary script records every value and tolerance.

For the Gram calculation, write $a = A_d/\sqrt{6}$, $b = B_d/\sqrt{3}$ and $c = C_d$. Permutation symmetry makes all spatial diagonal entries equal and all spatial off-diagonal entries equal. Their difference is

$$\zeta_\perp = (c-b)^2 + 2(d-2)(b-a)^2, \qquad \zeta_\parallel = \|T_d\|_{\mathrm{F}}^2 - (d-1)\zeta_\perp.$$

The first eigenvalue has multiplicity $d-1$, the second has multiplicity one, and the constant coordinate contributes a zero eigenvalue. To obtain the difference, compare two unfolding rows: the all-equal entries contribute $(c-b)^2$, and the $2(d-2)$ remaining distinct row-pair positions contribute $(b-a)^2$ each. This also gives a direct check of the reported optimal matrix tail without materializing a cubic array at large d.

At $d = 4096$, the projection norm is 0.199509, the largest coefficient divided by that norm is 9.346×10^{-6}, and the explicit rank-one tensor has relative error 0.009950. The best rank-one unfolding tail is 0.008121 of the norm. The density scale $\theta \le \{2(1 + |c_d|)\}^{-1}$ ensures $1/2 \le 1 \pm \theta g_d \le 3/2$ at every evaluated dimension. These calculations concern the signal geometry; the dependence comparison and detection guarantee follow from Proposition 5.3.

I.2 COMPLETE EXPERIMENTAL PROTOCOL

Every alternative changes at $n/2$ from the uniform density to $1 + g$. The four core signals are

$$g_{\mathrm{marginal}} = .30\phi_3(x_1), \qquad g_{\mathrm{quadratic}} = .22\phi_2(x_1)\phi_1(x_2),$$
$$g_{\mathrm{triple}} = .17\phi_1(x_1)\phi_1(x_2)\phi_1(x_3), \qquad g_{\mathrm{mixed}} = \tfrac{.19}{\sqrt{2}}\{\phi_3(x_1) + \phi_2(x_2)\phi_1(x_3)\}.$$

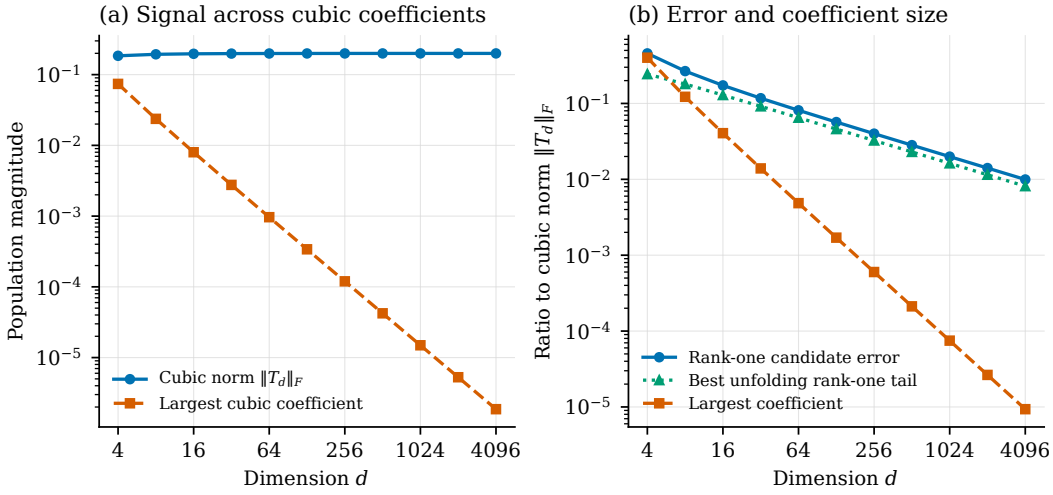

Figure 2: Population geometry of the two-frequency cubic construction on the complete 11-dimension grid. Left: the cubic projection norm $\|T_d\|_{\mathrm{F}}$ and its largest normalized-basis coefficient. Right: the largest coefficient and the two approximation errors, each divided by $\|T_d\|_{\mathrm{F}}$. The approximation curves distinguish the specified rank-one tensor error from the best rank-one unfolding tail. Lines connect deterministic evaluations, with no sampling error bars.

All four signals preserve moments of degree at most two. Their specified amplitudes set different signal strengths, so method comparisons are made within each model. We use $n = 12000$, $d \in \{8, 16\}$, and 40 independent trials per setting. Sensitivity experiments use the triple signal with $n \in \{6000, 24000\}$ at $d = 8$ and with $d = 24$ at $n = 12000$. A weaker signal sets $D = .08u^{\otimes 3}$, where u is uniform on the first three spatial coordinates. All densities are nonnegative by explicit sup-norm bounds. Null evaluations comprise 80 independent sequences for each of uniform $d = 8$, uniform $d = 16$, and $f = 1 + .20\phi_2(x_1)$ at $d = 16$.

The six cubic family comparisons share the pilot, $h = 8\lfloor N/32 \rfloor$, permutation count, and localization rule. The adaptive library has $J = 12$ candidates: basis sparsities $\{1, 4, 16\}$, marginal sparsities $\{1, 4\}$, ranks $\{1, 2, 4\}$ for each spectral family, and the full direction. Family comparisons use the corresponding subsets. We also run a degree-two full direction and 256 Gaussian random Fourier features (Rahimi and Recht, 2007). The latter uses kernel bandwidth squared $d/3$, fixed from the uniform reference scale. Each has its own full-norm pilot and follows the same three-fold protocol. These comparators examine feature choices within the same inference protocol, using explicitly fixed polynomial and random-feature constructions.

Two reference baselines scan every integer split in the central 80% of the full sequence. Max basis scan uses the largest absolute coefficient CUSUM; Full kernel scan uses its Euclidean norm. Their rank p values use 199 independent null sequences for each $(n, d, \text{background})$. These scans combine full-data use with known-background calibration, whereas the three-fold test uses its validation sample under an unknown common null.

We report detection rate, forced-position error even when a test does not reject, and total localization loss, which assigns $n/2$ to a missed detection. Rate intervals are 95% Wilson intervals; loss error bars are one standard error across trials. The records also report localization error conditional on rejection.

I.3 Density generation and randomization

The density is evaluated relative to product-uniform probability measure. If $g = \sum_\alpha \theta_\alpha \psi_\alpha$, define

$$B_g = \sum_\alpha |\theta_\alpha| \prod_j \sqrt{2\alpha_j + 1}.$$

Every simulated model has $B_g < 1$. We generate uniform proposals and accept with probability $(1+g(x))/(1+B_g)$. This gives exact rejection sampling from the specified density. The null quadratic

background uses the same construction. The weaker rank-one model uses $u_1 = u_2 = u_3 = 1/\sqrt{3}$, all other coordinates including index zero equal to zero, and coefficients $\theta_\alpha = .08\langle u^{\otimes 3}, E_\alpha\rangle$. Its code identifier is `dense_rank1`; only three spatial coordinates are active. The active support is therefore fixed as ambient dimension varies.

Settings are indexed from zero in the supplied configuration. Trial r of setting i has seed $2026092500+1000i+r$. The generator uses the seed for data and a separate deterministic offset for analysis. Random Fourier features use a further fixed offset, draw $\omega \sim N(0, 3I_d/d)$ and independent uniform phases, and evaluate $\sqrt{2/256}\cos(\omega^T x + b)$. Features are shared across folds of a trial. The supplementary source records the exact offsets. The reference scan calibration seed is $3016092500 + 1000i + r$, where i indexes the six sorted calibration settings. There are $6 \times 199 = 1194$ independent calibration sequences and 720 paired evaluation sequences.

The permutation procedure samples balanced sign vectors, each with h positive and h negative entries of magnitude $(2h)^{-1/2}$. For a balanced fixed-window CUSUM this has exactly the same distribution as a uniform observation permutation. All candidates in a cyclic ordering use the same permutation weights. The plus-one numerator includes the observed statistic, and ties count against rejection. With 199 permutations and the three-test Bonferroni rule, the smallest global p value is .015 and the largest attainable value below .05 is .045.

I.4 Computational scope and implementation correspondence

The code materializes $q = \binom{d+3}{3}$ normalized coefficients per observation. Feature storage costs $O(nq)$; a full scan using coefficient prefix sums costs $O(nq)$. Dense SVD of a $p \times p^2$ unfolding costs $O(p^4)$, and Tucker contractions are performed sequentially without building a Kronecker projector. Validation first projects $2h$ features onto J fixed directions, then computes the R balanced contrasts, costing $O(hqJ + RhJ)$. Central scalar least squares uses cumulative loss increments and costs $O(h)$. The $O(d)$ identity in Appendix B evaluates one kernel pair; the experiments use the explicit-feature implementation with the costs above.

Numerical normalization returns zero for coefficient norm at most 10^{-14}. No oracle direction replaces a zero estimate. Signal tensors and background means enter diagnostics only. Symmetrization can introduce a constant coefficient into a spectral candidate; all statistical observations have the constant feature removed, so CUSUMs are unchanged and anchor differences retain their meaning. Candidate support ties use stable coefficient ordering. Singular-subspace ties follow the numerical linear algebra routine; Proposition C.1 concerns its specified exact subspaces, rather than a guarantee about that routine's tie convention.

The stored environment is Python 3.12.3 and NumPy 2.4.3 on x86-64 Linux under WSL2. Four worker processes each use one BLAS thread. The 720 main trials took approximately 103 seconds; the 1194 calibration plus 720 reference evaluation jobs took approximately 164 seconds on the execution host. These batch times describe the execution host. Family timing fields include shared library construction, so the records report them as shared-procedure timings. Figure generation uses Matplotlib 3.10.8.

Every reported trial uses significance level .05 and the coefficient-pilot configuration specified in Appendix I.2. The implementation also exposes optional rank grids, the spectral pilot, and thinning; the reported power tables concern the fixed coefficient-pilot configuration. Manifests and supplied source snapshots identify the generating programs. Resume checks enforce agreement of configuration and source versions, and the reproduction guide uses fresh output directories. Complete reruns reproduce every non-timing trial field. All 720 full-kernel scores and estimates computed by the two runners agree exactly.

I.5 Diagnostics and uncertainty

At $d = 16$, the adaptive method selects the one-coefficient basis direction in 108 of 120 marginal orderings, 110 of 120 quadratic–linear orderings, and 72 of 120 triple orderings. Selection frequencies describe overlapping cyclic fits, not 120 independent experiments. The three central localization searches contain the true split in all marginal and quadratic–linear orderings, 90% of triple orderings,

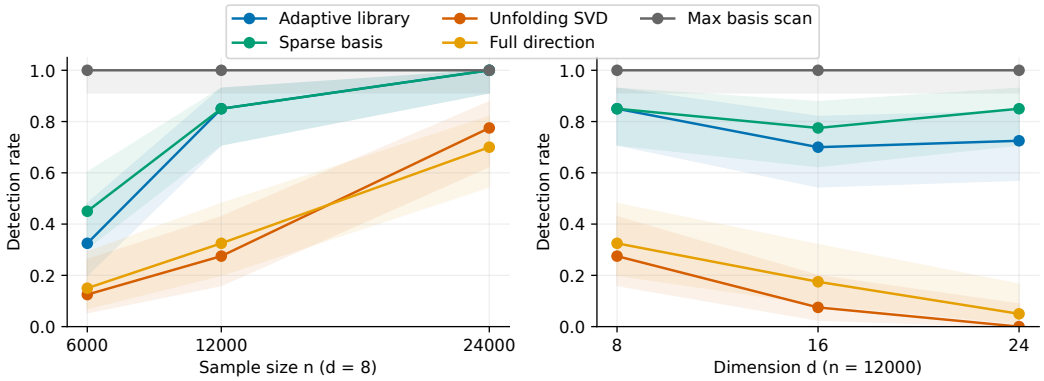

Figure 3: Triple-interaction detection rates with 95% Wilson intervals, 40 trials per point. The left panel varies n at $d = 8$; the right varies d at $n = 12000$. Reference scan rows use known-background calibration. Lines connect discrete experimental settings.

and 90.8% of mixed orderings. These diagnostics connect the empirical pilot to the containment event used by the refinement theorem.

All 720 trials contribute to the summaries. For a detection count k in m independent trials, we use the usual Wilson interval with $z = 1.959963984540054$. Total-loss standard errors use the sample standard deviation divided by $\sqrt{m}$. Under the reference calibration design, test sequences are independent conditional on the common calibration bank; the corresponding rate interval describes that conditional rejection probability. A fresh bank can change the rate. The comparisons report descriptive rates, errors, and their stated uncertainty intervals.

I.6 Additional direction and localization comparisons

Table 1 shows that the sparse basis family retains non-diagonal signals well. At $d = 16$, its quadratic–linear detection rate is 100%, compared with 20% for unfolding SVD and 22.5% for Tucker. On the triple interaction the rates are 77.5%, 7.5%, and 7.5%, respectively. The marginal cubic family has zero population signal on both interactions and correspondingly low rejection rates. All three structural families detect the marginal signal reliably.

Figure 1 links this comparison to the quantity controlled by the theory: the selected direction's population jump. Basis truncation matches the few active coefficients of these models and retains their interaction signal. Its median forced errors at $d = 16$ are 66 for quadratic–linear and 240 for triple changes, compared with 786 and 718.5 for unfolding SVD. Proposition C.1 explains a specific additional mechanism by which a common Tucker subspace can remove an interaction. Together, the comparisons motivate choosing structure by the signal it preserves.

Library size introduces a further tradeoff. The adaptive library raises mixed-signal detection from the sparse family's 82.5% to 90% at $d = 16$, while triple detection changes from 77.5% to 70%. Expanding the library changes both selection and maximum-statistic calibration. Equation (46) identifies the corresponding theoretical requirement: a richer library helps when its useful candidate signal survives the selection error.

I.7 Sample size, signal strength, and null calibration

The full-data max basis scan detects all four $d = 16$ core signals in all 40 trials. Its median errors are 7, 40.5, 110.5, and 79 in table order. This simple scan is particularly effective with the supplied background information and full sample. Its advantage over the full kernel scan on the three non-marginal settings also supports concentrating on the active coefficients. The comparison jointly changes background knowledge and sample use, so it quantifies their combined practical advantage.

Figure 3 shows the sample-size dependence. At $d = 8$, adaptive power rises from 32.5% at $n = 6000$ to 85% at 12000 and 100% at 24000. Sparse basis reaches 45%, 85%, and 100%. The weaker

Table 2: Independent null rejection rates (%), with 95% Wilson intervals in brackets. Each setting has 80 sequences. Calibration reference samples are excluded from these counts.

Method	Uniform, $d = 8$	Uniform, $d = 16$	Quadratic, $d = 16$
Adaptive library	3.8 [1.3, 10.5]	6.2 [2.7, 13.8]	0.0 [0.0, 4.6]
Sparse basis	1.2 [0.2, 6.7]	2.5 [0.7, 8.7]	1.2 [0.2, 6.7]
Marginal cubic	1.2 [0.2, 6.7]	6.2 [2.7, 13.8]	0.0 [0.0, 4.6]
Unfolding SVD	2.5 [0.7, 8.7]	3.8 [1.3, 10.5]	1.2 [0.2, 6.7]
Symmetric Tucker	6.2 [2.7, 13.8]	3.8 [1.3, 10.5]	5.0 [2.0, 12.2]
Full direction	3.8 [1.3, 10.5]	5.0 [2.0, 12.2]	1.2 [0.2, 6.7]
Degree-two direction	6.2 [2.7, 13.8]	5.0 [2.0, 12.2]	3.8 [1.3, 10.5]
RBF random features	2.5 [0.7, 8.7]	2.5 [0.7, 8.7]	5.0 [2.0, 12.2]
Max basis scan†	5.0 [2.0, 12.2]	1.2 [0.2, 6.7]	8.8 [4.3, 17.0]
Full kernel scan†	7.5 [3.5, 15.4]	6.2 [2.7, 13.8]	6.2 [2.7, 13.8]

three-coordinate rank-one setting gives detection rates of 7.5% for adaptive and sparse basis, 10% for Tucker, and 22.5% and 32.5% for the two full-data scans. This model uses amplitude .08, so its comparison with stronger sparse alternatives also changes signal strength.

Adaptive null rejection rates are 3.75%, 6.25%, and 0% across the three backgrounds. The corresponding sparse-basis rates are 1.25%, 2.5%, and 1.25%. Table 2 reports every method with uncertainty intervals. The fixed-library evaluation complements the finite-sample null guarantee proved from exchangeability. Each reference setting shares one calibration bank, so its reported interval describes evaluation uncertainty conditional on that bank.

The following tables include every alternative setting and method. Dagger rows use full-data scans with known-background calibration. The raw summaries additionally include Wilson intervals and localization errors conditional on rejection. Units of position error and loss are original sample indices.

Table 3: Coordinate, $n = 12000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	100.0	51.0	66.6	9.4
Sparse basis	100.0	46.5	64.5	9.5
Marginal cubic	100.0	43.5	60.8	9.4
Unfolding SVD	100.0	64.5	79.4	10.5
Symmetric Tucker	100.0	60.0	71.0	10.5
Full direction	92.5	141.0	639.1	246.4
Degree-two direction	7.5	1104.0	5691.7	184.1
RBF random features	0.0	895.5	6000.0	0.0
Max basis scan†	100.0	13.0	35.0	7.5
Full kernel scan†	100.0	27.0	53.8	12.5

Table 4: Coordinate, $n = 12000$, $d = 16$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	100.0	33.0	48.7	9.5
Sparse basis	100.0	28.5	47.8	9.6
Marginal cubic	100.0	33.0	48.7	9.5
Unfolding SVD	97.5	51.0	297.5	153.4
Symmetric Tucker	100.0	37.5	104.8	21.2
Full direction	50.0	397.5	3216.8	448.1
Degree-two direction	5.0	1330.5	5760.3	167.3
RBF random features	5.0	1273.5	5823.1	127.5
Max basis scan†	100.0	7.0	21.4	5.1
Full kernel scan†	100.0	76.0	155.2	28.8

Table 5: Quadratic linear, $n = 12000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	100.0	60.0	106.7	24.4
Sparse basis	100.0	60.0	106.7	24.4
Marginal cubic	2.5	672.0	5879.1	120.9
Unfolding SVD	85.0	159.0	1088.2	332.4
Symmetric Tucker	75.0	238.5	1790.5	393.1
Full direction	70.0	351.0	2102.6	412.3
Degree-two direction	7.5	1239.0	5714.4	165.8
RBF random features	5.0	1149.0	5808.4	136.1
Max basis scan†	100.0	22.5	70.7	19.0
Full kernel scan†	100.0	82.0	211.9	61.0

Table 6: Quadratic linear, $n = 12000$, $d = 16$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	100.0	66.0	129.5	30.1
Sparse basis	100.0	66.0	129.5	30.1
Marginal cubic	2.5	936.0	5858.0	142.0
Unfolding SVD	20.0	786.0	4986.8	326.8
Symmetric Tucker	22.5	753.0	4855.4	341.7
Full direction	7.5	796.5	5595.1	228.3
Degree-two direction	7.5	1276.5	5635.3	205.2
RBF random features	7.5	1578.0	5688.2	178.2
Max basis scan†	100.0	40.5	63.0	10.7
Full kernel scan†	62.5	410.0	2573.2	430.0

Table 7: Triple, $n = 12000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	85.0	172.5	1115.3	332.2
Sparse basis	85.0	165.0	1093.4	332.8
Marginal cubic	2.5	904.5	5876.7	123.3
Unfolding SVD	27.5	817.5	4501.4	390.6
Symmetric Tucker	25.0	631.5	4647.2	377.4
Full direction	32.5	631.5	4179.2	421.8
Degree-two direction	12.5	1234.5	5536.4	200.9
RBF random features	5.0	1380.0	5762.1	166.1
Max basis scan†	100.0	27.5	93.8	28.8
Full kernel scan†	90.0	128.5	895.2	281.8

Table 8: Triple, $n = 12000$, $d = 16$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	70.0	415.5	2073.4	414.4
Sparse basis	77.5	240.0	1607.9	382.2
Marginal cubic	5.0	930.0	5755.7	170.6
Unfolding SVD	7.5	718.5	5567.6	243.2
Symmetric Tucker	7.5	861.0	5583.1	234.9
Full direction	17.5	1057.5	5187.9	290.3
Degree-two direction	10.0	1479.0	5559.2	217.6
RBF random features	12.5	1437.0	5600.6	178.7
Max basis scan†	100.0	110.5	156.6	28.2
Full kernel scan†	42.5	1109.5	3910.8	403.9

Table 9: Mixed, $n = 12000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	90.0	177.0	877.4	278.3
Sparse basis	87.5	172.5	1020.5	306.8
Marginal cubic	77.5	184.5	1554.6	385.6
Unfolding SVD	40.0	310.5	3709.2	450.4
Symmetric Tucker	30.0	570.0	4374.0	399.4
Full direction	30.0	651.0	4359.6	402.7
Degree-two direction	7.5	948.0	5707.6	176.9
RBF random features	2.5	1426.5	5862.6	137.4
Max basis scan†	100.0	119.5	182.1	32.3
Full kernel scan†	100.0	95.5	237.2	52.6

Table 10: Mixed, $n = 12000$, $d = 16$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	90.0	253.5	959.7	280.7
Sparse basis	82.5	235.5	1311.6	349.3
Marginal cubic	85.0	235.5	1223.8	333.8
Unfolding SVD	2.5	720.0	5859.8	140.2
Symmetric Tucker	7.5	886.5	5618.6	216.3
Full direction	5.0	673.5	5714.9	199.0
Degree-two direction	2.5	1281.0	5870.5	129.5
RBF random features	5.0	1420.5	5776.5	156.7
Max basis scan†	100.0	79.0	110.7	16.8
Full kernel scan†	60.0	444.5	3083.6	415.2

Table 11: Triple, $n = 6000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	32.5	451.5	2143.1	199.1
Sparse basis	45.0	316.5	1730.0	225.3
Marginal cubic	2.5	532.5	2938.2	61.8
Unfolding SVD	12.5	466.5	2674.7	138.0
Symmetric Tucker	10.0	480.0	2747.9	122.2
Full direction	15.0	442.5	2626.7	144.5
Degree-two direction	5.0	625.5	2868.8	91.6
RBF random features	10.0	652.5	2798.6	98.0
Max basis scan†	100.0	20.5	112.2	49.9
Full kernel scan†	45.0	637.5	1992.8	194.4

Table 12: Triple, $n = 24000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	100.0	109.5	188.6	29.4
Sparse basis	100.0	96.0	167.8	26.6
Marginal cubic	5.0	1812.0	11515.8	338.3
Unfolding SVD	77.5	334.5	3042.0	776.0
Symmetric Tucker	72.5	264.0	3664.3	826.9
Full direction	70.0	760.5	4214.9	822.7
Degree-two direction	7.5	2548.5	11183.9	459.9
RBF random features	0.0	1738.5	12000.0	0.0
Max basis scan†	100.0	50.0	97.0	17.4
Full kernel scan†	100.0	140.5	293.1	74.2

Table 13: Triple, $n = 12000$, $d = 24$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	72.5	232.5	1838.2	412.1
Sparse basis	85.0	144.0	1106.8	331.7
Marginal cubic	5.0	957.0	5747.5	176.8
Unfolding SVD	0.0	645.0	6000.0	0.0
Symmetric Tucker	5.0	1033.5	5745.6	177.7
Full direction	5.0	712.5	5734.3	185.7
Degree-two direction	2.5	1401.0	5922.2	77.8
RBF random features	2.5	1512.0	5869.7	130.3
Max basis scan†	100.0	45.0	112.7	24.8
Full kernel scan†	12.5	2191.0	5501.2	225.9

Table 14: Three-coordinate rank-one interaction, $n = 24000$, $d = 8$, 40 repetitions. Error is the median forced-position error, regardless of rejection. Loss assigns $n/2$ to a missed detection.

Method	Power (%)	Error	Mean loss	Loss SE
Adaptive library	7.5	2341.5	11344.7	372.2
Sparse basis	7.5	2094.0	11338.9	376.4
Marginal cubic	0.0	2505.0	12000.0	0.0
Unfolding SVD	5.0	2556.0	11649.2	254.8
Symmetric Tucker	10.0	2080.5	11044.6	463.3
Full direction	5.0	1663.5	11551.6	319.1
Degree-two direction	2.5	2560.5	11784.2	215.8
RBF random features	5.0	2482.5	11524.8	343.1
Max basis scan†	22.5	3249.0	9807.0	669.3
Full kernel scan†	32.5	2645.0	8968.6	744.3