INTERSECTION-AWARE TANGENT DECOMPOSITION FOR MANIFOLD CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Clustering intersecting manifolds requires distinguishing nearby observations whose generating branches have different tangent directions. We study when local branch representations can make such clustering possible, and present Intersection-aware Tangent-decomposition Clustering (ITC) as a practical construction motivated by that analysis. The abstract theory identifies sufficient conditions for recovery: a local residual margin controls branch assignment, and a graph perturbation bound tolerates arbitrary tangent errors on a vanishing fraction of observations. For balanced orthogonal line segments, we establish a nonempty fixed-radius regime in which imperfect representations and asymptotically optimal two-means rounding yield vanishing misclassification error. This provides an existence result for recoverable intersecting-manifold configurations, together with stability beyond exact tangents. A small-angle oracle example shows why graph recoverability is a separate requirement. The practical method fits affine branches at spatially distributed anchors, assigns one projector per observation, and constructs a radius-supported affinity. Its finite fitting and spectral routines provide a concrete implementation of the design principles, evaluated empirically; the existence guarantee concerns the stated abstract realization. A further conditional result covers the degree-rescaled embedding used by the backend.

1 INTRODUCTION

A neighborhood near a manifold intersection can contain samples from several generating branches. Euclidean proximity alone does not identify those branches, and a single local covariance can combine their tangent directions. This creates a representation problem before the final clustering problem: which local tangent should be attached to each observed point?

Local geometry is already central to multi-manifold clustering. Local-PCA spectral methods can resolve intersections under geometric assumptions (Arias-Castro et al., 2017), and graph-based consistency depends on how within-manifold and cross-manifold connections are constructed (Garcia Trillos et al., 2023). Our focus is a different step within this framework: explicitly fit several affine models in candidate mixed neighborhoods and select a branch for each observation before constructing the graph. This distinguishes local branch estimation from the subsequent question of whether the graph favors the generating partition.

We develop Intersection-aware Tangent-decomposition Clustering (ITC): fit multiple affine models at distributed anchors in candidate mixed neighborhoods, select a branch by each point's residual, and build a radius-supported graph with one tangent projector per observation. The theoretical question is whether there exist intersecting configurations and imperfect representations for which this graph recovers the generating partition. We exhibit such a regime on orthogonal crossing segments, allowing arbitrary errors on a vanishing fraction of samples. The practical algorithm aims to realize these principles through finite fitting and spectral routines; experiments assess its effectiveness.

Our contributions are:

- A local-to-global analysis connecting residual assignment, projector error, and graph recoverability, while allowing arbitrary errors on an unresolved subset under explicit degree conditions.

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.

- An existence and stability result for clustering orthogonal crossing segments at a fixed radius: a nonempty class of imperfect representations admits vanishing misclassification error under abstract spectral rounding. A small-angle oracle obstruction delineates why geometry still matters.
- A fully specified practical ITC construction motivated by these conditions, with distributed anchors, finite affine fitting, pointwise branch selection, and one graph node per observation.

The benchmark datasets, baseline results, and evaluation structure are retained from the previous submission. Updated ITC results are reported on seven synthetic datasets; three-dataset representation ablations assess the effect of branch fitting.

2 Related Work

Geometric multi-manifold clustering. Local-PCA spectral clustering compares estimated principal subspaces in a neighborhood graph and gives guarantees for simplified variants (Arias-Castro et al., 2017). RMMSL combines local tangent estimation with a curved-level similarity (Gong et al., 2012). Sparse Manifold Clustering and Embedding uses local sparse representations (Elhamifar & Vidal, 2011). Simplex Paths instead constructs a locality graph on simplices, with dihedral-angle weights and a largest-angle path distance (Chen et al., 2025). These approaches establish that geometry can distinguish intersecting components; we do not claim that intersections or tangent information have been absent from prior work. ITC investigates explicit local branch fitting followed by pointwise assignment and a single-node position–tangent graph.

Subspaces and tangent estimation. Sparse subspace clustering models unions of linear subspaces through sparse self-representations (Elhamifar & Vidal, 2013). ITC uses affine subspaces locally, without assuming each entire cluster is globally linear. Statistical tangent estimation on a single smooth manifold has its own sampling and regularity conditions (Aamari & Levrard, 2019); those results do not automatically provide a tangent for a mixed neighborhood. We therefore formulate local-model accuracy as a module condition instead of identifying an alternating local fit with its global optimum.

Spectral recovery. Normalized spectral clustering combines a graph embedding with a separate rounding step (Ng et al., 2001; von Luxburg, 2007). For intersecting manifolds, graph geometry itself requires analysis: Garcia Trillos et al. (2023) study sufficient graph conditions and spectral approximation for multi-manifold clustering, including annular proximity graphs with angle constraints. Our analysis concerns the particular radius-supported product affinity used here. It distinguishes an oracle with exact tangents, perturbations of that oracle, and rounding of the resulting embedding. It does not transfer a theorem for a different graph or a different embedding solely by naming the same spectral clustering routine.

3 Problem Setup and Method

We observe $x_i = q_i + \xi_i \in \mathbb{R}^D$, where $q_i \in \mathcal{M}_{y_i}$, $y_i \in \{1, \dots, M\}$, and $\|\xi_i\| \le \nu$. The manifolds have common intrinsic dimension $d < D$; C^2 regularity is used when invoking curvature bounds. The generating label y_i is distinct from set membership: at a shared point it need not be identifiable from location alone. The dimensions d and number of clusters M are inputs. General noisy manifolds motivate the construction; the statistical theorem below is noiseless and restricted to two straight segments.

3.1 Candidate neighborhoods and anchors

For $\mathcal{N}_r(i) = \{j : \|x_i - x_j\| \le r\}$, including i, compute the mean-centered covariance with divisor $|\mathcal{N}_r(i)|$. Require at least $d+2$ samples in every neighborhood. Its descending nonnegative eigenvalues $\widehat{\lambda}_{i,\ell}$ define

$$R_d(i) = \frac{\sum_{\ell > d} \widehat{\lambda}_{i,\ell}}{\sum_{\ell \le d} \widehat{\lambda}_{i,\ell} + \rho_R \sum_{\ell} \widehat{\lambda}_{i,\ell} + t_{\mathrm{tiny}}}. \tag{1}$$

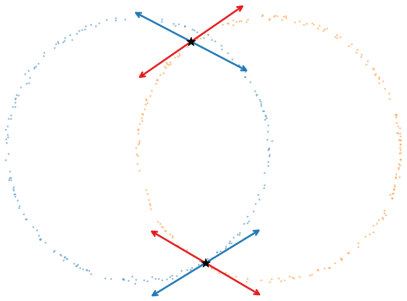


Figure 1: Local branch fitting near an intersection. Arrows show fitted branch directions, and star markers indicate fitting anchors.

Here t_{tiny} is the floating-point positive safeguard. The top-d eigenspace gives $\widehat{P}_i^{\text{loc}}$. Candidates satisfy $R_d(i) \geq \tau_R$. Apply DBSCAN only to these points, with radius r and minimum count $q_{\min}$ including the query itself. Non-candidates and DBSCAN noise retain their local projectors.

Within each retained component, choose its sample medoid, then repeatedly add a point farthest from the current anchors until all component points are within $\gamma r + 10^{-12}$ of a selected anchor. Fit each anchor on its radius-r neighborhood in the full data set. Candidate membership is not a guarantee that this neighborhood contains exactly two branches; the detector is also sensitive to curvature, noise, and mixture balance.

3.2 Finite affine fits and pointwise branch assignment

For an anchor a, fit $K = 1, \ldots, K_{\max}$ affine d-subspaces on $S_a = \{i : \|x_i - a\| \leq r\}$. For fixed $K \geq 2$, each restart initializes all centers at a and obtains each direction from QR decomposition of an independent Gaussian $D \times d$ matrix. The iteration alternates minimum squared-residual assignment with group mean and SVD refitting. A restart is discarded if any group has fewer than $b_{\min}$ points at any size check. The full stopping and final-refit rules appear in Appendix A.

Write $E_K(a)$ for the smallest *achieved* mean residual among the valid restarts. It is not a global optimization oracle. $K = 1$ uses one deterministic mean–SVD fit. Missing fits are skipped. Select

$$\widehat{m}_a = \min \left\{ K : E_K(a) \text{ is finite}, \frac{E_K(a)}{\max\{E_1(a), \epsilon_E\}} \leq \eta_{\text{sel}} \right\}. \tag{2}$$

The anchor is invalid if no K qualifies. For $0 < \eta_{\text{sel}} < 1$, $K = 1$ is selected precisely when its fit exists and $E_1(a) \leq \eta_{\text{sel}} \epsilon_E$. Thus this rule is not a claim of consistent branch counting, nor a general test that one noisy branch is present.

If a component has no valid anchor, the procedure reports failure. Otherwise each retained candidate chooses

$$\begin{aligned} a(i) &\in \operatorname*{arg\,min}_{a \in A_{\ell(i)}^{\text{valid}}} \|x_i - a\|, \\ b(i) &\in \operatorname*{arg\,min}_{1 \leq b \leq \widehat{m}_{a(i)}} \|(I - \widehat{P}_{a(i),b})(x_i - \widehat{\mu}_{a(i),b})\|^2. \end{aligned} \tag{3}$$

Here $\ell(i)$ denotes the retained DBSCAN component of i. There is no assignment-distance cap. Set $\widetilde{P}_i = \widehat{P}_{a(i),b(i)}$ for this path, and $\widetilde{P}_i = \widehat{P}_i^{\text{loc}}$ otherwise. Invalid-anchor removal can enlarge coverage distances. Extra fitted models are competitors in (3); they are not automatically harmless.

Figure 1 illustrates the local branch directions obtained near an intersection.

3.3 One node per sample and a radius-supported graph

For each point with a positive-distance neighbor within r, set $s_i^{\text{A1}} = \text{median}\{\|x_i - x_j\| : 0 < \|x_i - x_j\| \leq r\}$. The A1 spatial scale is $\widehat{\sigma}_s = \text{median}_i\, s_i^{\text{A1}}$ over the defined s_i^{A1}, retaining

Algorithm 1 ITC representation and graph construction

Require: X, d, M, r, τ_R and fitting settings in Appendix A
1: Compute radius-local PCA projectors and scores (1); reject insufficient neighborhoods.
2: Apply DBSCAN to candidates; initialize $\widetilde{P}_i \leftarrow \widehat{P}_i^{\mathrm{loc}}$ for all i.
3: **for** each retained candidate component **do**
4: Select medoid–farthest-point anchors with covering tolerance.
5: Fit affine models and select valid anchors by (2).
6: Reject this run if no valid anchor remains.
7: Assign each component point by (3) and update its single projector $\widetilde{P}_i$.
8: **end for**
9: Estimate A1; reject an undefined scale; form (4).
10: Cluster using the practical backend described below.
11: **return** one label per observation, or an explicit failure

multiplicities. Failure is reported if no finite positive scale exists. Define

$$A_{ij} = \mathbf{1}_{\{\|x_i - x_j\| \le r\}} \exp\left(-\frac{\|x_i - x_j\|^2}{\widehat{\sigma}_s^2} - \frac{\|\widetilde{P}_i - \widetilde{P}_j\|_F^2}{2\sigma_T^2} \right) \quad (i \neq j), \qquad A_{ii} = 1. \tag{4}$$

There is exactly one node per original observation, so no label fusion is required. Radius support permits sparse storage but does not guarantee a sparse graph in every sampling regime.

The tangent discrepancy is basis invariant. If B_i is an orthonormal basis and O_i is orthogonal, then $(B_i O_i)(B_i O_i)^\top = B_i B_i^\top$. Consequently $d_T^2(P_i, P_j) = \frac{1}{2}\|P_i - P_j\|_F^2 = d - \mathrm{tr}(P_i P_j)$ requires no synchronization of basis coordinates. This fact concerns comparison of subspaces in the same ambient space, not transport of vector-valued coordinates between manifolds.

Spectral backend and theoretical realization. Both start from $W = A - \mathrm{diag}(A)$ and $L_{\mathrm{sym}} = I - D_W^{-1/2} W D_W^{-1/2}$, where $D_W = \mathrm{diag}(W\mathbf{1})$ on positive-degree nodes. The implementation provides a default scikit-learn ARPACK path and a dense numerical eigendecomposition option used in the experiments. Both pass $D_W^{-1/2} U$ to finite k-means++/Lloyd runs, without row normalization. The main recovery theorem uses exact bottom-M eigenvectors U, row normalization, and an explicit global objective approximation contract. Theorem D.7 extends the conditional analysis to the degree-rescaled embedding without row normalization. The representation interface and graph construction are shared. The abstract realization establishes a recoverable regime, while the practical backend provides the finite procedure used in the experiments. Appendix A records defaults, failure behavior, and computational costs.

4 Recoverable Regimes and Representation Stability

Let $P_i^* = P_{T_{q_i}\mathcal{M}_{y_i}}$. Our oracle graph uses these projectors but the *same* observed positions, radius support, and empirical A1 scale as (4). It can have cross-manifold edges. We first control the representation, then compare graphs, and finally establish a recoverable oracle regime. The purpose is to identify a nonempty regime in which clustering succeeds, and to characterize the representation errors it tolerates. The practical method is motivated by these sufficient conditions. Complete statements and proofs appear in Appendices B and D.

4.1 Local assignment outside an unresolved region

For two cleanly intersecting branches at p, let $\theta_0 > 0$ be their smallest nonzero principal angle, excluding shared intersection directions. Write $q = p + v + e$, where v lies in the first tangent space orthogonal to the intersection tangent, $t = \|v\| \le t_0$, and $\|e\| \le \kappa t^2/2$. Assume each of the correct and competing fitted planes has offset at most a from p and projector operator-norm error at most ϵ relative to its reference tangent.

Proposition 4.1 (Local residual margin)**.** *For $\|x - q\| \le \nu$, residual assignment prefers the correct model if*

$$t(\sin\theta_0 - 2\epsilon) - \kappa t^2 > 2(a + \nu). \tag{5}$$

If $c_0 = \sin\theta_0 - 2\epsilon - \kappa t_0 > 0$, the condition holds outside $t \le 2(a+\nu)/c_0$.

The correct-model distance is at most $a + \epsilon t + \kappa t^2/2 + \nu$; the competitor's distance is at least $t\sin\theta_0 - a - \epsilon t - \kappa t^2/2 - \nu$. Their comparison proves the claim; see Proposition B.80. With more models the comparison is imposed against every incorrect competitor. The correct model must be available at the anchor actually used. The unresolved region is uncertified, not necessarily misclassified. After correct assignment, tangent transport adds at most $L_T h_i$ to the fitting error (Lemma B.8).

4.2 Arbitrary errors on some samples

Let B contain k observations and assume $\|\widetilde{P}_i - P_i^*\|_F \le \varepsilon$ outside B. On B the projectors may be arbitrary rank-d orthogonal projectors. Let $\overline{W}$ be the hybrid graph that retains their projectors on B and uses exact tangents elsewhere. Suppose the actual, hybrid, and oracle graphs have degree at least $d_{\min} > 0$ and at most $q_{\max}$ supported neighbors per node.

Theorem 4.2 (Representation-to-graph stability)**.** *With $\eta_g = e^{2\sqrt{d}\,\varepsilon/\sigma_T^2} - 1 < 1$,*

$$\|S - S^*\|_{\mathrm{op}} \le \frac{2\eta_g}{1-\eta_g} + \frac{\sqrt{2kq_{\max}}}{d_{\min}}\left(1 + \frac{q_{\max}}{d_{\min}}\right), \tag{6}$$

where $S = D_W^{-1/2} W D_W^{-1/2}$ and S^* is its oracle counterpart.

The first term compares the actual and hybrid graphs multiplicatively. The second uses at most $2kq_{\max}$ changed ordered edge entries and controls the associated degree perturbation; see Theorem B.83. Thus arbitrary errors on B are allowed, but a small fraction k/n alone is insufficient. If $d_{\min} \ge cn$ with fixed $c > 0$, the second term is $O(\sqrt{k/n})$. A complementary aggregate-leakage route, stated in Theorem B.44, compares cross-cluster energy with the within-cluster gap without requiring small maximum row error.

4.3 A concrete partial consistency theorem

For the following theorem only, take $\mathcal{M}_j = [-L, L]e_j \subset \mathbb{R}^2$, $e_1 \perp e_2$, with N independent uniform samples per segment, no noise, and $n = 2N$. Write $x_i = s_i e_{y_i}$, where $s_i \in [-L, L]$ is the signed segment coordinate, distinct from the A1 spacing s_i^{A1}. The two infinite sampling sequences are independent and nested.

For row-normalized spectral rows $Y_i \in \mathbb{R}^2$, define the returned centers-and-labels objective and its global optimum by

$$\begin{aligned} J_N(c, z) &= \frac{1}{n}\sum_{i=1}^{n} \|Y_i - c_{z_i}\|^2, \\ J_N^\star &= \inf_{c_1, c_2 \in \mathbb{R}^2,\ z \in \{1,2\}^n} J_N(c, z). \end{aligned} \tag{7}$$

The rounding contract is $J_N(\widehat{c}, \widehat{y}) \le \alpha_N J_N^\star$ with $\alpha_N \ge 1$. No balance constraint is imposed on the returned clusters.

Theorem 4.3 (Fixed-radius partial recovery)**.** *Fix $\sigma_T = 0.2$ and $5.3 \cdot 10^{-4} \le r/L < 1/4$. Assume representation construction completes. Local-projector-path errors are at most $\varepsilon_{\mathrm{loc},N} \to 0$. At every actually used anchor, all competing lines match a true branch, have origin offset at most $a_N \to 0$, and direction operator-norm error at most $\delta_N \to 0$, with $\delta_N < 1/2$. Each assigned sample's true branch is represented. These bounds are uniform and hold almost surely.*

Use empirical A1, exact bottom-two symmetric-normalized eigenvectors, row normalization, and an α_N-approximate global two-means solution with $\alpha_N \to 1$ almost surely. Then $\mathrm{err}(\widehat{y}_N, y_N) \to 0$ almost surely. Single-model anchors and multiple models of a branch are allowed. No accuracy relative to the generating branch is required of the final projector on

$$B_N = \{i : |s_i| \le w_N\}, \qquad w_N = \frac{2a_N}{1 - 2\delta_N}. \tag{8}$$

The corresponding statement holds in probability under in-probability module and rounding limits.

The local-model assumptions still apply at anchors serving B_N; they need not force the selected model there to represent the generating branch. More generally, Theorem B.84 allows arbitrary rank-one projectors on any set of $o(n)$ observations, provided the final projectors converge uniformly to the true projectors on its complement. This representation-level result does not require anchor or model-count assumptions.

Here err is the misclassification fraction minimized over label permutations. The theorem assumes local-model accuracy rather than proving it for the finite-restart solver. It also does not require all anchors to return two branches or recover a correct branch count. The lower radius endpoint is a conservative sufficient certificate from the population rounding analysis, not a tuning recommendation or a claimed recovery threshold.

Existence and estimator-level consequences. Corollary E.1 shows that the representation conditions describe a nonempty class. Appendix D further connects them to local fitting and to the embedding used by the practical backend. The following consequences use the notation and uniform fitting event defined there.

Corollary 4.4 (A certified local-fitting route)**.** *Consider the balanced noiseless orthogonal-segment model of Theorem 4.3. Suppose that, eventually almost surely, every retained candidate lies in $|s| \le a_{\mathrm{out}} < r$, every retained component has at least one selected anchor, $K_{\max} \ge 2$, and the uniform local-fitting event $\mathcal{G}_N^{\mathrm{fit}}$ in* (116) *occurs. Suppose also that* (121) *and* (122) *hold eventually and that*

$$\frac{m_{\max,N}}{m_{0,N}} = O(1), \qquad \Delta_{\mathrm{opt},N} = o(r^2), \qquad \eta_{c,N} \to 0, \qquad \varepsilon_{\mathrm{loc},N} \to 0 \quad \textit{almost surely}. \tag{9}$$

Then the selection rule chooses the valid $K = 2$ candidate at every selected anchor eventually almost surely. There is a sequence of sets B_N such that

$$\frac{|B_N|}{n} \to 0, \qquad \max_{i \notin B_N} \|\widetilde{P}_i - P_i^*\|_F \to 0 \quad \textit{almost surely}.$$

Consequently, under the spectral and rounding assumptions of Theorem 4.3,

$$\mathrm{err}(\widehat{y}_N, y_N) \to 0 \qquad \textit{almost surely}.$$

The analogous conclusion holds in probability when the stated fitting, local-PCA, selection, and rounding conditions hold in probability.

Corollary 4.5 (Recovery for the backend-aligned embedding)**.** *Suppose the representation conclusion of Corollary 4.4 holds. Let U_N be the exact bottom-two symmetric-normalized Laplacian eigenvectors of the resulting graph. Apply two-means to $D_W^{-1/2}U_N$, without row normalization, with achieved objective ratio $\alpha_N \to 1$ almost surely. Then*

$$\mathrm{err}(\widehat{y}_N, y_N) \to 0 \qquad \textit{almost surely}.$$

The same conclusion holds for orthonormal numerical eigenvectors Q_N if

$$\inf_{O \in O(2)} \|Q_N - U_N O\|_F \to 0 \qquad \textit{almost surely},$$

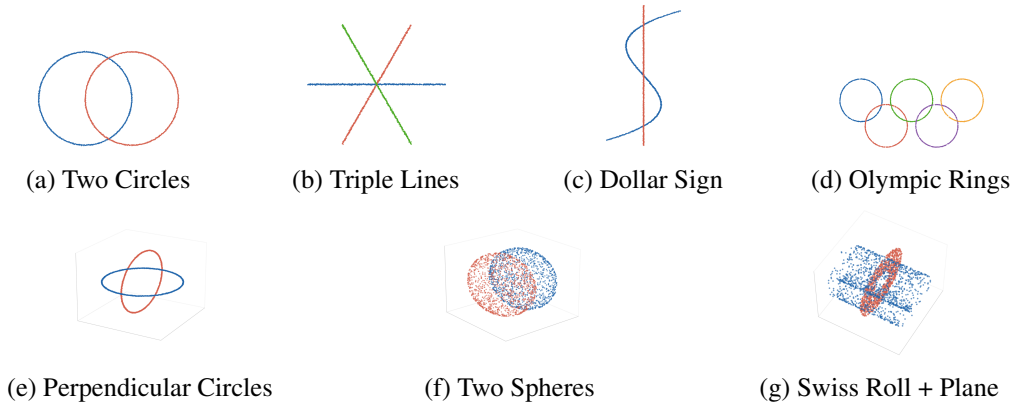
and the objective contract holds on their actual degree-rescaled embedding. The analogous statement holds with all conditions and conclusions in probability.

Corollaries 4.4 and 4.5 narrow two gaps between the abstract recovery theorem and the implementation. They remain conditional on the finite fitter attaining its additive objective contract and on numerical-subspace and rounding accuracy.

4.4 Why exact tangents are not enough

For balanced noiseless segments with acute angle θ and $r(1 + \csc\theta) < L/2$, population A1 is $r/2$. Let $a_\theta = e^{-\sin^2\theta/\sigma_T^2}$. Compare the true branch partition with the partition placing both positive halves together. Both have equal cluster volumes, and

$$\frac{\mathrm{Ncut}(\mathcal{P}_{\mathrm{side}})}{\mathrm{Ncut}(\mathcal{P}_{\mathrm{true}})} = \frac{\sin\theta}{\pi a_\theta} + \frac{\theta}{\pi}. \tag{10}$$



(a) Two Circles (b) Triple Lines (c) Dollar Sign (d) Olympic Rings

(e) Perpendicular Circles (f) Two Spheres (g) Swiss Roll + Plane

Figure 2: Synthetic benchmark geometries. Colors indicate generating manifold labels. The datasets include planar curves, spatial curves, and intersecting surfaces; detailed descriptions appear in Appendix C.1.

Thus the wrong partition has smaller normalized cut when $a_\theta(\pi - \theta) > \sin\theta$ (Proposition B.48). This is a cut-preference statement, not by itself a spectral-rounding counterexample. The distinct-sample construction in Section B.17 separately establishes actual spectral failure for its specified finite-scale configurations. Neither result asserts failure for every small-angle uniform sample. They explain why an additional graph-recoverability condition is needed, even with exact tangents.

5 Experiments

5.1 Synthetic Benchmarks

Datasets We evaluate on seven synthetic datasets: Two Circles, Perpendicular Circles, Two Spheres, Olympic Rings, Dollar Sign, Triple Lines, and Swiss Roll + Plane. They cover planar and spatial curves, surfaces, and multiple crossings. Revised ITC uses $400M$ pooled, randomly subsampled observations and normal-space tubular noise bounded by 0.002; branch counts fluctuate. Figure 2 illustrates the benchmark geometries; Appendix C.1 provides their detailed descriptions, and Appendix C specifies the executed protocol.

Baselines We compare against eight baselines covering standard clustering, subspace methods, and manifold-aware approaches: K-means, Spectral (NJW) (Ng et al., 2001), K-Subspaces (Vidal, 2011), TSC (Arias-Castro et al., 2017), SSC (Elhamifar & Vidal, 2013), SCC, Simplex Paths (Chen et al., 2025), and LocalPCA-Spectral (Arias-Castro et al., 2017).

Metric and repetitions We report clustering accuracy (ACC) under optimal Hungarian label matching. Each configuration is evaluated on ten fresh-data repetitions, with mean and sample standard deviation reported alongside the number of successful runs. Runs without valid clustering outputs are recorded as failures rather than assigned an artificial ACC. Development labels are used for configuration selection; evaluation labels are used only for scoring and post hoc diagnostics.

Quantitative results Table 1 reports the held-out results under two development protocols. On the five curve datasets, the configurations selected by the original search achieve mean ACC between 0.998 and 0.999 on fresh data. For Two Spheres and Swiss Roll + Plane, an expanded development search selects configurations using a stability-oriented criterion. Their held-out ACC is 0.963 ± 0.008 and 0.976 ± 0.008, respectively, with all ten runs completing successfully. The minimum observed ACC is 0.945 for Two Spheres and 0.964 for Swiss Roll + Plane.

These results show high accuracy on the evaluated fresh realizations, but do not establish that instability has been eliminated. For Swiss Roll + Plane, the selected configuration still has four development runs out of twenty below ACC 0.9, including one execution failure. Moreover, the surface datasets receive a broader parameter search than the curve datasets, and no matched baseline

Table 1: Clustering accuracy on seven synthetic datasets. Baseline results are retained from earlier experiments. ITC reports mean and sample standard deviation over ten held-out repetitions; its standard deviations appear on the second line. The two surface datasets use an expanded development search. TC: Two Circles; PC: Perpendicular Circles; OR: Olympic Rings; DS: Dollar Sign; TL: Triple Lines; TS: Two Spheres; SRP: Swiss Roll + Plane.

Method	TC	PC	OR	DS	TL	TS	SRP
K-means	.708	.550	.466	.520	.490	.798	.527
Spectral (NJW)	.705	.557	.724	.520	.507	.800	.537
K-Subspaces	.520	.685	.329	.685	.997	.537	.542
TSC	.507	.980	.356	.818	.602	.537	.625
SSC	.520	.662	.270	.500	.990	.525	.787
SCC	.850	.745	.503	.517	.478	.800	.530
Simplex Paths	.993	.995	.989	.995	.990	.497	.502
LocalPCA-Spec.	.600	.500	.535	.515	.490	.782	.535
ITC (ours)	.998	.999	.998	.998	.998	.963	.976
	$\pm.001$	$\pm.001$	$\pm.001$	$\pm.001$	$\pm.001$	$\pm.008$	$\pm.008$

evaluation is included. Comparisons with earlier ITC results also involve different evaluation seeds, so their differences cannot be attributed solely to the configuration changes.

5.2 Ablation Study: Local Representations

We compare local-PCA projectors, full ITC, and a true-tangent diagnostic on three datasets with three paired fresh-data repetitions. Positions, radius support, A1 scale, tangent bandwidth, and spectral settings are shared. The settings selected for full ITC are held fixed across arms; this isolates representation replacement at those settings rather than comparing independently optimized methods. All 27 runs completed.

Table 2 shows that branch fitting improves both ACC and intersection-region projector error strongly on Two Circles and Triple Lines. In every paired repetition on all three datasets, the mean squared projector error decreases. On Two Spheres, however, ACC does not improve uniformly: one repetition returns ACC 0.5025, below local PCA's 0.85875, despite lower mean projector error.

In that repetition, five candidates share an inaccurate fitted projector and form a group with strong internal but very weak external affinity. The spectral embedding isolates this group, producing a $795 : 5$ partition. Its normalized cut is about 1.31×10^{-4}, versus 0.1253 for the true partition. The same failure persists for ten tested k-means seeds. Thus average representation improvement can coexist with a harmful local graph structure. Appendix C.5 provides the diagnosis. True tangents yield ACC 1 in all nine diagnostic runs, but this is neither a deployable method nor a general performance upper bound. Three repetitions do not establish statistical significance or the frequency of the failure.

6 Discussion and Conclusion

The abstract theory establishes a nonempty recoverable regime for intersecting manifolds and quantifies stability to representation errors. Its statistical guarantee concerns balanced noiseless orthogonal segments at a fixed radius and vanishing misclassification fraction. It does not establish exact label recovery, general noisy-manifold consistency, or shrinking-radius guarantees. The fitting certificate and degree-rescaled extension connect the sufficient conditions to the practical design, while finite fitting restarts and Lloyd optimization remain uncertified.

Empirically, branch fitting reduces mean projector error in every observed paired ablation. However, the Two Spheres failure under the earlier configuration shows that shared local errors can create a weakly attached group that dominates the spectral partition. Thus average representation accuracy alone is insufficient. Main-benchmark ITC configurations are selected using development labels and evaluated on separate fresh realizations. The two surface datasets use an expanded, stability-oriented search; equal tuning budgets across methods are not established. Radius support also does not imply

Table 2: Representation ablation: mean $\pm$ sample SD over three paired repetitions. E_P is mean squared Frobenius projector error. The intersection region is defined from clean geometry with width r. These experiments use the configurations before the later surface retuning.

Representation	ACC	ACC (int.)	E_P (int.)
Two Circles			
Local PCA	0.764 ± 0.093	0.604 ± 0.027	0.201 ± 0.018
Full ITC	0.998 ± 0.001	0.978 ± 0.019	0.037 ± 0.026
True tangent	1.000 ± 0.000	1.000 ± 0.000	0.000 ± 0.000
Triple Lines			
Local PCA	0.660 ± 0.271	0.542 ± 0.188	0.386 ± 0.024
Full ITC	0.999 ± 0.001	0.987 ± 0.012	0.019 ± 0.018
True tangent	1.000 ± 0.000	1.000 ± 0.000	0.000 ± 0.000
Two Spheres			
Local PCA	0.763 ± 0.124	0.646 ± 0.048	0.302 ± 0.022
Full ITC	0.809 ± 0.266	0.760 ± 0.221	0.243 ± 0.025
True tangent	1.000 ± 0.000	1.000 ± 0.000	0.000 ± 0.000

near-linear cost: medoid selection is quadratic within components and fixed-radius graphs can have quadratically many edges.

ITC provides a concrete implementation of the representation-to-graph design. Together, the existence result, controlled ablations, and localized failure analysis identify both its potential and conditions that require further study.

AI USE STATEMENT

Generative AI tools assisted with mathematical formulation and proof drafting, manuscript drafting and revision, LaTeX editing, experimental design, and interpretation of results. Coding agents also assisted with implementing and running experimental and diagnostic scripts, aggregating results, and preparing tables. The authors take responsibility for the final mathematical claims, code, experimental results, and text.

ETHICS STATEMENT

The reported evaluation uses synthetic geometric datasets and involves no human participants or personal data. The study concerns clustering methodology; performance on these benchmarks does not establish suitability for decisions about individuals or groups.

REPRODUCIBILITY STATEMENT

Section 3 and Appendix A specify the algorithm, initialization, default settings, and failure handling. Section 4 states the theoretical assumptions, with proofs in Appendices B and D. Section 5 describes the benchmarks, and Appendix C documents parameter selection, development and evaluation protocols, repetition counts, dispersion definitions, failure accounting, and the representation ablation.

REFERENCES

Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019. doi: 10.1214/18-AOS1685.

Ery Arias-Castro, Gilad Lerman, and Teng Zhang. Spectral clustering based on local PCA. *Journal of Machine Learning Research*, 18(9):1–57, 2017. URL https://jmlr.org/papers/v18/14-318.html.

Haoyu Chen, Anna Little, and Akin Narayan. Robust multi-manifold clustering via simplex paths, 2025. URL https://arxiv.org/abs/2507.10710.

Ehsan Elhamifar and René Vidal. Sparse manifold clustering and embedding. In *Advances in Neural Information Processing Systems*, volume 24, pp. 55–63, 2011.

Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013. doi: 10.1109/TPAMI.2013.57.

Nicolas Garcia Trillos, Pengfei He, and Chenghui Li. Large sample spectral analysis of graph-based multi-manifold clustering. *Journal of Machine Learning Research*, 24(143):1–71, 2023. URL https://www.jmlr.org/papers/v24/21-1254.html.

Dian Gong, Xuemei Zhao, and Gerard Medioni. Robust multiple manifolds structure learning. In *Proceedings of the 29th International Conference on Machine Learning*, pp. 25–32, 2012.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, pp. 849–856, 2001.

Rene Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011. doi: 10.1109/MSP.2010.939739.

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

A Concrete Algorithm and Reproducibility Details

This appendix specifies the supplied execution path `final_mmc.fit_final`, using the primitives in `mmc.py`. It describes the practical algorithm, not a successful realization of the hypotheses in Theorem 4.3. All local PCA, DBSCAN, fitting, A1-neighborhood, and graph-support radii are r.

A.1 Settings and deterministic choices

Setting	Symbol	Code default
Intrinsic dimension; number of clusters	d, M	required
Radius; candidate threshold	r, τ_R	required
DBSCAN minimum count	$q_{\min}$	5
Anchor cover ratio	γ	0.5
Maximum branches	$K_{\max}$	3
Normalized residual cutoff	η_{sel}	0.15
Minimum group size	$b_{\min}$	8
Affine restart count; iteration limit	J, I	$12, 60$
Tangent bandwidth	σ_T	0.2
Score ridge; energy floor	ρ_R, ϵ_E	$10^{-12}, 10^{-12}$
Spectral k-means restarts	J_{km}	10
Random state	seed	0

These are implementation defaults, not experimentally validated optimal settings or sufficient parameter choices for the theorem. Changing the coordinate scale changes the dimensional fitting energies; the fixed energy floor must not be interpreted as a scale-invariant statistical threshold.

All radius comparisons include the boundary. Local PCA uses the neighborhood mean, biased covariance normalization $1/|\mathcal{N}_r(i)|$, and clamps negative numerical eigenvalues to zero for the score. The implementation requires $1 \le d < D$, $2 \le M \le n$, positive radius, cover ratio, and tangent bandwidth, and at least $d+2$ neighbors per sample. DBSCAN is applied to candidates in sample-index order and counts self. Its noise points remain in the data with their original PCA projectors.

For component index set C, the medoid minimizes $\sum_{j \in C} \|x_i - x_j\|$. Each following anchor maximizes its distance to the chosen set, stopping at $\gamma r + 10^{-12}$. Ties use the first index in the relevant ordered array. Components are processed in increasing DBSCAN label order. Anchor seeds use

`SeedSequence([seed, source_index, component])`.

Each anchor has its own generator, shared across K in increasing order. $K = 1$ consumes no random draws. A fixed seed makes the procedure reproducible for the same input order and numerical environment; it does not provide an initialization-success probability.

A.2 Finite affine fitting

For $m_a = |S_a| < K b_{\min}$, the K fit is absent. For $K = 1$, the center is the sample mean and the direction is the top-d right singular subspace of centered data. For $K \ge 2$, independently for each of J restarts:

1. Draw K Gaussian $D \times d$ matrices, take their QR bases B_b, set $P_b = B_b B_b^\top$, and initialize all centers at a. Assign each support point to a minimum-residual model.
2. If any group has fewer than $b_{\min}$ points, discard the restart. Otherwise refit each center as its group mean and each projector by SVD.
3. Reassign all support points by minimum residual and check sizes again. Discard if a group is too small. If assignments are unchanged, return this restart's fit. Otherwise continue, up to I iterations.
4. If the loop reaches its limit without returning or being discarded, refit once to the last assignments, reassign once more, and keep the fit only if the new group sizes are all at least $b_{\min}$.

The squared residual is evaluated as $\max\{\|x-\mu_b\|^2-(x-\mu_b)^\top P_b(x-\mu_b),0\}$; ties choose the first branch. The post-limit return need not be a fixed point: the returned centers fit the pre-reassignment groups, while the recorded labels are the new assignments. The restart energy is the mean minimum residual for the returned models. Among valid restarts the first minimum-energy fit is retained.

The implementation stores absent energies as NaN and scans (2) in increasing K. Failure at one K does not stop larger K from being tested. If E_1 is absent, the support-size test also prevents larger fits, and the anchor is invalid. The selected energy is an achieved objective; neither global optimality nor monotonicity of the retained E_K across K is assumed.

Each component point is assigned to its nearest valid anchor without a radius restriction. Distance ties follow valid-anchor creation order. The selected model supplies only its projector to the graph; the fitted center is used for residual assignment. Non-candidates and DBSCAN noise take the local-projector path. A1 excludes zero distances but retains repeated positive distances and gives equal weight to each eligible center.

A.3 Spectral computation and failure behavior

The default option delegates to scikit-learn `SpectralClustering` with precomputed affinity, k-means labels, `n_init=10`, and ARPACK. The experiments instead select `eigen_solver="dense"`, an explicit option in the supplied executed code. It forms the loop-free normalized Laplacian using SciPy and computes its bottom M eigenvectors, including the trivial eigenvector, with `scipy.linalg.eigh`. This is numerical eigendecomposition, not the theorem's exact-arithmetic operation.

Both paths use $D_W^{-1/2}U$, deterministic column signs, and no row normalization. The dense path calls scikit-learn k-means with k-means++, Lloyd iterations, `n_init=10`, `max_iter=300`, and `tol=1e-4`. Its seeded random state is not consumed by an iterative eigensolver first. In the default ARPACK path, eigensolver random draws can affect the subsequent k-means initialization, and internal fallbacks are possible. Finite Lloyd runs do not certify the global approximation factor assumed by the abstract theorem. Versions and the executed option are recorded in Appendix C.

The implementation explicitly reports failure for insufficient local PCA neighborhoods, a retained component without any valid anchor, an undefined/nonpositive A1 scale, and spectral exceptions of types `ValueError`, `LinAlgError`, or `ArithmeticError`. Other exceptions can propagate. Disconnected-graph or eigensolver warnings need not stop the run. Evaluation must record failed runs and warnings rather than only successful clusterings.

A.4 Computational cost and storage

Let $m_i=|\mathcal{N}_r(i)|$, c_ℓ be candidate-component sizes, q_ℓ their numbers of anchors, $m_a=|S_a|$, and E_r the number of undirected supported graph pairs. Radius-query cost depends on dimension and the data structure; it is not assumed to be logarithmic in general. Dense covariance formation and eigendecomposition cost $O(\sum_i m_iD^2+nD^3)$. The supplied medoid–FPS routine explicitly forms each component's pairwise distance matrix, costing $O(\sum_\ell c_\ell^2 D)$ time and up to $O(\max_\ell c_\ell^2)$ additional memory, followed by $O(\sum_\ell q_\ell c_\ell)$ cover updates.

For a given K at anchor a, a conservative bound using full projector residuals and dense SVDs is

$$O\big(J(I+1)[Km_aD^2+m_aD\min\{m_a,D\}]\big),$$

apart from initialization and low-order operations. Sum over attempted K and anchors. This is a cost bound, not a convergence guarantee. Constructing projector-based edge weights costs $O(E_rD^2)$ with the current representation. The sparse graph itself stores $O(n+E_r)$ entries, but stored projectors require $O(nD^2)$ memory and the current vectorized edge differences can require $O(E_rD^2)$ temporary memory. Iterative spectral cost depends on convergence and cannot be reduced to a fixed near-linear bound here. In the theorem's fixed-radius regime, E_r can be of order n^2. Thus radius support removes mandatory all-pairs affinity evaluation only when the observed support graph is sparse.

A.5 What is and is not established

The residual rule does not need exact branch-count recovery for the theory, but it does need a correct model and control of every incorrect competitor. Basis invariance is algebraic and is proved in the main text. Fitting convergence, practical spectral accuracy, parameter tuning, real-data utility, and improvements over cluster-first tangent estimation are not consequences of that invariance or of the conditional recovery theorem. They remain separate estimator or empirical questions.

B Technical Proofs and Auxiliary Results

This appendix supplies the definitions, intermediate results, and proofs supporting Section 4. Proposition 4.1 is Proposition B.80; Theorem 4.2 is Theorem B.83; and Theorem 4.3 is the local-model specialization Theorem B.1. The small-angle comparison is Proposition B.48. These results establish the recoverable regime and stability properties of the abstract representation and spectral construction. Corollary E.1 gives an explicit admissible witness; the practical implementation is motivated by these conditions.

B.1 Scope, notation, and the abstract algorithm

Let $\mathcal{M}_1, \dots, \mathcal{M}_M \subset \mathbb{R}^D$ be smooth manifolds of common dimension d. We observe

$$x_i = q_i + \xi_i, \qquad q_i \in \mathcal{M}_{y_i}, \qquad y_i \in \{1, \dots, M\}, \qquad \|\xi_i\| \le \nu, \quad i = 1, \dots, n.$$

The latent generating label y_i is fixed, including when a point belongs to an intersection. Such a label need not be identifiable from location alone. Define

$$P_i^* = P_{T_{q_i}\mathcal{M}_{y_i}}.$$

All tangent projectors in this document are rank-d orthogonal projectors. Vector norms are Euclidean. Matrix norms carry a subscript F or op. The local estimation and finite-graph recovery results are deterministic conditional statements about a realized sample and realized estimator outputs. Later sections separately state population integral-operator results and fixed-radius almost-sure sampling and oracle clustering results for specified crossing-segment models. These do not supply joint shrinking-radius sampling guarantees or probability bounds for the concrete estimators.

B.1.1 Pipeline

The abstract algorithm takes the observed sample, d, M, and a common radius $r > 0$ as inputs, together with module-specific settings.

1. A local geometry estimator returns projectors $\widehat{P}_i^{\mathrm{loc}}$ and a candidate set.
2. A component routine groups candidates. Candidates not retained in a component follow the local-projector path.
3. An anchor routine places anchors in each component. At each anchor a, a branch estimator returns a finite collection $\{(\widehat{\mu}_{a,k}, \widehat{P}_{a,k})\}_{k=1}^{\widehat{m}_a}$, or declares the anchor invalid.
4. Write A_ℓ^{valid} for the valid anchors of component ℓ. If a retained component has no valid anchor, the algorithm declares decomposition failure and produces no clustering.
5. For each candidate assigned to component $\ell(i)$, choose

$$a(i) \in \operatorname*{arg\,min}_{a \in A_{\ell(i)}^{\mathrm{valid}}} \|x_i - a\|, \tag{11}$$

$$b(i) \in \operatorname*{arg\,min}_{1 \le k \le \widehat{m}_{a(i)}} \|(I - \widehat{P}_{a(i),k})(x_i - \widehat{\mu}_{a(i),k})\|^2. \tag{12}$$

Ties are resolved by a fixed rule. Set

$$\widetilde{P}_i = \begin{cases} \widehat{P}_{a(i),b(i)}, & i \text{ follows the branch-assignment path,} \\ \widehat{P}_i^{\mathrm{loc}}, & \text{otherwise.} \end{cases}$$

6. Form exactly one node $(x_i, \widetilde{P}_i)$ for each observation. No anchor nodes or sample copies are added.
7. A spatial scale estimator returns $\widehat{\sigma}_s > 0$. Its concrete realization is A1, defined below.
8. With $\sigma_T > 0$, form the symmetric affinity

$$A_{ij} = \mathbf{1}_{\{\|x_i - x_j\| \le r\}} \exp\left(-\frac{\|x_i - x_j\|^2}{\widehat{\sigma}_s^2} - \frac{d_T^2(\widetilde{P}_i, \widetilde{P}_j)}{\sigma_T^2} \right) \quad (i \neq j), \qquad A_{ii} = 1, \tag{13}$$

where

$$d_T^2(P, Q) = d - \operatorname{tr}(PQ) = \tfrac{1}{2}\|P - Q\|_F^2.$$

9. A clustering routine returns $\widehat{y} = \mathcal{C}(A, M)$.

The common concrete choices are radius-local PCA, R_d detection, DBSCAN, medoid/farthest-point anchor covering, affine K-subspaces, normalized-E_1 branch selection, and spectral clustering. Only the assignment, single-node representation, and affinity rules are fixed in the abstract analysis. The tangent bandwidth is a symbolic positive constant here; the concrete choice is 0.2.

B.1.2 Concrete A1 scale and status of its analysis

Let

$$\mathcal{D}_i = \{\|x_i - x_j\| : 0 < \|x_i - x_j\| \le r\}, \qquad I_+ = \{i : \mathcal{D}_i \neq \varnothing\}.$$

Here $\mathcal{D}_i$ is an indexed multiset: each eligible neighbor index j contributes one entry, and equal positive distances retain their multiplicities. The outer median likewise counts one entry per eligible center index. No distance deduplication is performed. A1 uses

$$s_i^{\mathrm{A1}} = \operatorname{median} \mathcal{D}_i \quad (i \in I_+), \qquad \widehat{\sigma}_s = \operatorname{median}_{i \in I_+} s_i^{\mathrm{A1}}. \tag{14}$$

If $I_+ = \varnothing$, scale estimation fails. Otherwise $0 < \widehat{\sigma}_s \le r$ follows directly from the definition. A lower bound of the form $c_s r \le \widehat{\sigma}_s$ does not follow without a condition on the empirical local distance distributions. The perturbation results below compare graphs using the same observed positions, radius support, and estimated scale, so they do not require such a lower bound. Section B.10 provides an empirical sufficient condition for the lower bound used in geometric analysis.

B.2 Main partial recovery guarantee

The central claim concerns the abstract representation and row-normalized spectral realization. Local-model accuracy is an input condition; the conclusion tolerates unresolved samples near the intersection. In particular, no fixed branch count at every anchor is assumed.

Theorem B.1 (Local models imply partial recovery at a fixed radius)**.** *Let $\mathcal{M}_j = [-L, L]e_j \subset \mathbb{R}^2$ for $j = 1, 2$, where $e_1 \perp e_2$. Use N independent uniform samples per branch, independently across branches, on nested sample sequences, with $n = 2N$ and no noise. Write $x_i = s_i e_{y_i}$ for the signed segment coordinates. Fix $\sigma_T = 0.2$ and $5.3 \cdot 10^{-4} \le r/L < 1/4$. Assume the following three conditions.*

1. ***A completed representation.*** *Every observation receives a rank-one orthogonal projector through the abstract pipeline. Every component actually used has a valid anchor; assignment uses its selected same-component nearest valid anchor. No condition on unused anchors or their model counts is imposed.*

2. ***Accuracy of the models actually used.*** *Local-projector-path outputs have Frobenius error at most $\varepsilon_{\mathrm{loc},N}$. Every model competing at an actually used anchor is associated with one of the two true lines, has distance at most a_N from the origin, and has direction operator-norm error at most $\delta_N < 1/2$ relative to that line. At least one model represents each assigned sample's generating branch. The uniform bounds satisfy*

$$\varepsilon_{\mathrm{loc},N} \to 0, \qquad a_N \to 0, \qquad \delta_N \to 0 \quad \textit{almost surely}.$$

Single-model anchors and multiple models of a branch are allowed.

3. ***An abstract spectral realization.*** *Construct the radius-supported position–tangent affinity using empirical A1. Remove self-loops for degrees, compute exact bottom-two eigenvectors of its symmetric normalized Laplacian, and normalize rows. Returned two-means centers and labels satisfy a global objective approximation contract $J_N(\widehat{c},\widehat{y}) \le \alpha_N J_N^\star$ as in* (7)*, with $\alpha_N \ge 1$ and $\alpha_N \to 1$ almost surely.*

Define

$$w_N = \frac{2a_N}{1-2\delta_N}, \qquad B_N = \{i : |s_i| \le w_N\}, \qquad \varepsilon_N = \max\{\varepsilon_{\mathrm{loc},N}, \sqrt{2}\delta_N\}.$$

Outside B_N, every branch-path assignment selects its generating branch and every final projector has Frobenius error at most ε_N. No accuracy relative to the generating branch is required of the final projectors inside B_N; the assumed accuracy of the competing models still applies at anchors serving these samples. The final labels satisfy

$$\mathrm{err}(\widehat{y}_N, y_N) \longrightarrow 0 \quad \textit{almost surely}.$$

The analogous conclusion holds in probability when the module and rounding limits are only in probability.

More quantitatively, let z_N be the maximum two-branch empirical CDF error, and let $\mathcal{B}_N, h_N, \omega_\Phi$ denote the oracle sampling operator bound, transport displacement, and population row modulus from Section B.20. For all sufficiently large N on the almost-sure event above,

$$\mathrm{err}(\widehat{y}_N, y_N) \le C\left[(\alpha_N - 1) + \varepsilon_N + \sqrt{w_N/L + 2z_N} + \mathcal{B}_N + \omega_\Phi(h_N)\right], \tag{15}$$

where $C < \infty$ depends on the fixed geometry, bandwidth, population spectral gap, and rounding separation, but not on N. This is an error budget; no estimator-specific rate for its terms is asserted.

Proof. Corollary B.85 proves the assignment and projector conclusions and $|B_N|/n \le w_N/L + 2z_N$ eventually. Theorem B.84 proves consistency. For the displayed budget, its proof establishes a degree lower bound cn in the actual, hybrid, and oracle graphs with fixed $c > 0$. Equations (89) and (96) then give

$$H_N \le C_1\{\varepsilon_N + \sqrt{w_N/L + 2z_N}\}$$

eventually, since $\exp(2\varepsilon_N/\sigma_T^2) - 1 = O(\varepsilon_N)$ in the rank-one case. Equation (93) bounds $\sqrt{E_N}$ by a constant times $\mathcal{B}_N + H_N + \omega_\Phi(h_N)$ once their sum is sufficiently small. In (79), use $J_{0,n} \le 1$, $\alpha_N \le 2$ eventually, and $E_N \le \sqrt{E_N}$ once $E_N \le 1$. This proves (15). The same deterministic bounds yield the in-probability statement on events of probability tending to one. □

Logical scope. The theorem is conditional on local-model accuracy and the abstract rounding contract. It is neither a guarantee for a named fitting solver nor eventual exact recovery of all labels. It allows inaccurate final projectors in the unresolved region. The stronger representation-level Theorem B.84 also allows an arbitrary vanishing fraction of bad vertices, even when they are not in a tube. The curved local margin in Proposition B.80 is useful separately; it does not turn this straight-line sampling theorem into a general curved-manifold theorem.

B.3 Affine model perturbation and hard assignment

Lemma B.2 (Stability of distance to an affine plane)**.** *Let*

$$L^* = \mu^* + \mathrm{range}(P^*), \qquad \widehat{L} = \widehat{\mu} + \mathrm{range}(\widehat{P}).$$

Suppose

$$\|\widehat{P} - P^*\|_{\mathrm{op}} \le \varepsilon_P, \qquad \|\widehat{\mu} - \mu^*\| \le \varepsilon_\mu, \qquad \|x - \mu^*\| \le R.$$

Then

$$|\operatorname{dist}(x, \widehat{L}) - \operatorname{dist}(x, L^*)| \le R\varepsilon_P + \varepsilon_\mu. \tag{16}$$

Proof. Subtract the residual vectors:

$$\begin{aligned}(I-\widehat{P})(x-\widehat{\mu})-(I-P^*)(x-\mu^*)\\ =(P^*-\widehat{P})(x-\mu^*)+(I-\widehat{P})(\mu^*-\widehat{\mu}).\end{aligned}$$

Its norm is at most $R\varepsilon_P+\varepsilon_\mu$ since $\|I-\widehat{P}\|_{\mathrm{op}}\le 1$. The reverse triangle inequality gives (16). □

Remark B.3 (Centers and equivalent affine representations). An affine plane has many possible centers. The lemma uses fixed, corresponding center representations. A center-error hypothesis must therefore be justified for that choice; a displacement along the same plane need not represent a geometric estimation error. Frobenius control of a projector also provides the operator-norm control used here.

Lemma B.4 (Assignment under a residual margin)**.** *Fix a candidate i and its actually selected anchor $a(i)$. Associate to every estimated model at that anchor a reference affine model $L^*_{a(i),k}$. Define*

$$d^*_{i,k}=\mathrm{dist}(x_i,L^*_{a(i),k}),\qquad \widehat{d}_{i,k}=\mathrm{dist}(x_i,\widehat{L}_{a(i),k}).$$

Suppose the correct branch k_i^* is represented among these models and, for every represented branch k, including k_i^*,

$$|\widehat{d}_{i,k}-d^*_{i,k}|\le\delta_i.$$

If

$$\min_{k\neq k_i^*}(d^*_{i,k}-d^*_{i,k_i^*})>2\delta_i, \tag{17}$$

then the assignment rule (12) *selects k_i^* uniquely. For a single available model, the comparison condition is vacuous.*

Proof. For any competitor k,

$$\widehat{d}_{i,k}-\widehat{d}_{i,k_i^*}\ge d^*_{i,k}-d^*_{i,k_i^*}-2\delta_i>0.$$

Distances and squared distances have the same minimizers because the squaring map is strictly increasing on the nonnegative real numbers. □

Remark B.5 (Missing and extra branches). Exact branch-count recovery is not an explicit requirement of this lemma. However, the correct branch must be present, and every additional model must satisfy the competitor condition. Extra branches are not automatically harmless. Likewise, no residual rule can select a missing branch. All conditions concern the nearest valid anchor actually selected by the algorithm, not some better anchor that happens to exist elsewhere.

Proposition B.6 (A geometric sufficient condition for the margin)**.** *For the correct local branch choose $q_a\in\mathcal{M}_{y_i}$ and define*

$$L^*_{a,k_i^*}=q_a+T_{q_a}\mathcal{M}_{y_i}.$$

Assume $\|q_i-q_a\|\le h_i$ and a local quadratic approximation bound

$$\mathrm{dist}(q_i,L^*_{a,k_i^*})\le C_\kappa h_i^2.$$

Write

$$s_i^{\mathrm{sep}}=\min_{k\neq k_i^*}\mathrm{dist}(q_i,L^*_{a,k}).$$

Under the distance-estimation bound in Lemma B.4, the condition

$$s_i^{\mathrm{sep}}>C_\kappa h_i^2+2\nu+2\delta_i \tag{18}$$

guarantees correct branch assignment.

Proof. Distance to a fixed nonempty closed set is 1-Lipschitz. Thus

$$d^*_{i,k_i^*}\le C_\kappa h_i^2+\nu,\qquad \min_{k\neq k_i^*}d^*_{i,k}\ge s_i^{\mathrm{sep}}-\nu.$$

Subtracting establishes (17), and Lemma B.4 applies. □

Remark B.7 (Shared directions and ambiguous points). Condition (18) does not require a strictly positive smallest principal angle between full tangent spaces. It can accommodate shared tangent directions at positive-dimensional intersections. It is a pointwise margin condition, which may fail at or very near the intersection. This proposition does not establish a bound on the number of such ambiguous samples.

B.4 Propagation of tangent error from anchors

Lemma B.8 (Anchor-to-sample tangent error)**.** *Suppose the tangent projector field on the relevant manifold patch obeys*

$$\|P_{T_q\mathcal{M}_k} - P_{T_{q'}\mathcal{M}_k}\|_F \le L_T\|q - q'\|.$$

For candidate i, suppose the correct branch is selected and

$$\|\widehat{P}_{a(i),k_i^*} - P_{T_{q_a}\mathcal{M}_{y_i}}\|_F \le \varepsilon_{\mathrm{fit}}, \qquad \|q_i - q_a\| \le h_i.$$

Then

$$\|\widetilde{P}_i - P_i^*\|_F \le \varepsilon_{\mathrm{fit}} + L_T h_i. \tag{19}$$

Proof. Insert $P_{T_{q_a}\mathcal{M}_{y_i}}$ between the two projectors and apply the triangle inequality followed by the two hypotheses. □

Remark B.9 (Coverage after invalid-anchor removal). An initial cover of radius γr need not remain a cover of the same radius after invalid anchors are discarded. Existence of one valid anchor per component is a termination condition, not a locality guarantee. If the actually selected valid anchor satisfies

$$\|x_i - a(i)\| \le h_{\mathrm{cov}}, \qquad \|a(i) - q_a\| \le h_{\mathrm{anchor}},$$

then

$$h_i \le \nu + h_{\mathrm{cov}} + h_{\mathrm{anchor}}.$$

The second condition is branch-specific: an observed anchor need not lie on every manifold represented by its local model.

B.5 From tangent error to affinity error

Definition B.10 (Reference affinity)**.** Use exactly the same observed positions, radius support, spatial scale, and tangent bandwidth as the algorithm, but replace estimated tangent projectors by P_i^*:

$$A_{ij}^* = \mathbf{1}_{\{\|x_i - x_j\| \le r\}} \exp\left(-\frac{\|x_i - x_j\|^2}{\widehat{\sigma}_s^2} - \frac{d - \mathrm{tr}(P_i^* P_j^*)}{\sigma_T^2}\right) \quad (i \ne j), \qquad A_{ii}^* = 1. \tag{20}$$

This reference isolates tangent-estimation error. It is not assumed to be block diagonal, nor is its clustering correctness automatic. Both its locations and its spatial scale remain data-dependent.

Lemma B.11 (Chordal-distance and affinity perturbation)**.** *Let $\varepsilon_i = \|\widetilde{P}_i - P_i^*\|_F$. Then*

$$|d_T^2(\widetilde{P}_i, \widetilde{P}_j) - d_T^2(P_i^*, P_j^*)| \le \sqrt{d}\,(\varepsilon_i + \varepsilon_j). \tag{21}$$

On every supported off-diagonal pair,

$$e^{-\zeta_{ij}} A_{ij}^* \le A_{ij} \le e^{\zeta_{ij}} A_{ij}^*, \qquad \zeta_{ij} = \frac{\sqrt{d}\,(\varepsilon_i + \varepsilon_j)}{\sigma_T^2}. \tag{22}$$

In particular, if $\varepsilon_i \le \varepsilon$ for every i, then

$$|A_{ij} - A_{ij}^*| \le \eta A_{ij}^*, \qquad \eta = \exp\left(\frac{2\sqrt{d}\,\varepsilon}{\sigma_T^2}\right) - 1. \tag{23}$$

Proof. Expand

$$\begin{aligned}\mathrm{tr}(\widetilde{P}_i\widetilde{P}_j) - \mathrm{tr}(P_i^* P_j^*) = \mathrm{tr}((\widetilde{P}_i - P_i^*)\widetilde{P}_j) \\ + \mathrm{tr}(P_i^*(\widetilde{P}_j - P_j^*)).\end{aligned}$$

Apply the Frobenius Cauchy–Schwarz inequality and $\|P_i^*\|_F = \|\widetilde{P}_j\|_F = \sqrt{d}$. The spatial factors cancel in the ratio A_{ij}/A_{ij}^*, yielding (22). Finally, $1 - e^{-t} \le e^t - 1$ for $t \ge 0$ gives (23). Unsupported pairs and diagonals have zero perturbation. □

B.6 NORMALIZED GRAPH PERTURBATION

For this analysis remove self-loops explicitly:

$$W = A - \mathrm{diag}(A), \qquad W^* = A^* - \mathrm{diag}(A^*).$$

Set $d_i = \sum_j W_{ij}$, $d_i^* = \sum_j W_{ij}^*$, and $D = \mathrm{diag}(d_i)$, $D^* = \mathrm{diag}(d_i^*)$. When degrees are positive, define

$$S = D^{-1/2} W D^{-1/2}, \qquad S^* = (D^*)^{-1/2} W^* (D^*)^{-1/2}.$$

The corresponding symmetric normalized Laplacians are $I - S$ and $I - S^*$, with the same perturbation norm. An implementation using a different treatment of diagonal entries must use matching degree and normalization definitions when invoking the results.

Proposition B.12 (Weighted row error controls normalized affinity)**.** *Let W, W^* be symmetric nonnegative adjacency matrices and suppose $d_i^* > 0$ for all i. If*

$$\max_i \frac{\sum_j |W_{ij} - W_{ij}^*|}{d_i^*} \le \eta < 1, \tag{24}$$

then all $d_i > 0$ and

$$\|S - S^*\|_{\mathrm{op}} \le \frac{2\eta}{1-\eta}. \tag{25}$$

Proof. Put $E = W - W^*$. First,

$$|d_i - d_i^*| \le \eta d_i^*, \qquad (1-\eta) d_i^* \le d_i \le (1+\eta) d_i^*.$$

For any vector v, symmetry and $2|ab| \le a^2 + b^2$ give

$$\begin{aligned} |v^\top (D^*)^{-1/2} E (D^*)^{-1/2} v| &\le \frac{1}{2} \sum_{ij} |E_{ij}| \left(\frac{v_i^2}{d_i^*} + \frac{v_j^2}{d_j^*} \right) \\ &\le \eta \|v\|^2. \end{aligned}$$

Thus $F = (D^*)^{-1/2} E (D^*)^{-1/2}$ satisfies $\|F\|_{\mathrm{op}} \le \eta$. Let $H = D^{-1/2} (D^*)^{1/2}$ and $a = (1-\eta)^{-1/2}$. Then $S = H(S^* + F)H$, $\|H\|_{\mathrm{op}} \le a$, and $\|H - I\|_{\mathrm{op}} \le a - 1$. The latter follows from the diagonal entry interval $[(1+\eta)^{-1/2}, (1-\eta)^{-1/2}]$ and convexity of $t \mapsto t^{-1/2}$. Since W^* is nonnegative and symmetric, the same quadratic-form inequality with row sums d_i^* gives $\|S^*\|_{\mathrm{op}} \le 1$. Consequently,

$$\begin{aligned} \|S - S^*\|_{\mathrm{op}} &\le \|HFH\|_{\mathrm{op}} + \|HS^*H - S^*\|_{\mathrm{op}} \\ &\le a^2 \eta + (a-1)(a+1) \\ &= \frac{2\eta}{1-\eta}. \end{aligned}$$

□

B.7 A COMBINED DETERMINISTIC THEOREM

Theorem B.13 (Local models imply normalized graph stability)**.** *Suppose the abstract algorithm completes its representation and scale construction successfully. Assume:*

1. *Every sample following the local-projector path satisfies $\|\widehat{P}_i^{\mathrm{loc}} - P_i^*\|_F \le \varepsilon_{\mathrm{loc}}$.*
2. *At each candidate's actually selected anchor, the correct branch is present and the model errors and margins satisfy Lemma B.4.*
3. *For each such candidate, the selected correct branch satisfies the hypotheses of Lemma B.8, with uniform constants $\varepsilon_{\mathrm{fit}}$, L_T, and $h_i \le h$.*
4. *All reference off-diagonal degrees d_i^* are positive.*

Define

$$\varepsilon = \max\{\varepsilon_{\mathrm{loc}}, \varepsilon_{\mathrm{fit}} + L_T h\}, \qquad \eta = \exp\left(\frac{2\sqrt{d}\,\varepsilon}{\sigma_T^2}\right) - 1.$$

Then $\|\widetilde{P}_i - P_i^*\|_F \le \varepsilon$ *for all* i*. If* $\eta < 1$*, then*

$$\|S - S^*\|_{\mathrm{op}} \le \frac{2\eta}{1-\eta}.$$

Proof. Lemma B.4 establishes correct candidate assignments. Lemma B.8 bounds candidate tangent errors; the first hypothesis bounds all remaining tangent errors. Lemma B.11 implies $|W_{ij} - W_{ij}^*| \le \eta W_{ij}^*$. Summing over j proves (24). Proposition B.12 completes the proof. □

Remark B.14 (Strength and interpretation). This is a sufficient condition, not a sharp parameter prescription. It does not assert that local PCA, DBSCAN, or K-subspaces always satisfy the hypotheses. In particular, samples sent to the local-projector path by DBSCAN noise handling are not automatically geometrically regular. The hypothesis must include their actual tangent errors.

B.8 Allowance for ambiguous samples

Let G be a set on which tangent errors are bounded by ε, and let $B = \{1, \dots, n\} \setminus G$. Define

$$\eta_G = \exp\left(\frac{2\sqrt{d}\,\varepsilon}{\sigma_T^2}\right) - 1, \qquad \beta = \max_i \frac{\sum_{j:\, i\in B \text{ or } j\in B} |W_{ij} - W_{ij}^*|}{d_i^*}.$$

Corollary B.15 (Good-node error and bad-incident-edge error)**.** *If all* $d_i^* > 0$ *and* $\eta_G + \beta < 1$*, then*

$$\|S - S^*\|_{\mathrm{op}} \le \frac{2(\eta_G + \beta)}{1 - \eta_G - \beta}.$$

Proof. On $G \times G$, Lemma B.11 bounds the absolute edge error by $\eta_G W_{ij}^*$. On the remaining pairs use the definition of β. The total row error divided by d_i^* is at most $\eta_G + \beta$, and Proposition B.12 applies. □

Remark B.16 (Small cardinality alone is insufficient). The condition $|B|/n \ll 1$ does not by itself imply that β is small. A few bad nodes can introduce strong bridges or suffer large relative degree perturbations. In particular, the maximum defining β includes rows indexed by bad nodes themselves. This extension is a weighted-edge perturbation result, not yet an error bound depending only on the fraction of ambiguous samples.

B.9 Conditional connection to clustering

For hard labels define permutation-invariant error by

$$\mathrm{err}(\widehat{y}, y) = \min_{\pi \in \mathfrak{S}_M} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{\pi(\widehat{y}_i) \ne y_i\}}.$$

Assumption B.17 (An explicit, separate clustering stability contract)**.** For the reference instance (W^*, y), let Ψ be nondecreasing on $[0, t_0]$. The chosen clustering routine, acting on the graph and using the stated normalization convention, has the property that every admissible W with positive degrees satisfies

$$\|S(W) - S(W^*)\|_{\mathrm{op}} \le t \le t_0 \quad \Longrightarrow \quad \mathrm{err}(\mathcal{C}(W + I, M), y) \le \Psi(t).$$

Any solver or rounding approximation requirements are part of this contract. For a randomized routine, a probabilistic version must explicitly state its event and probability.

Corollary B.18 (Clustering implication, conditional on the contract)**.** *Under Theorem B.13 and Assumption B.17, if* $2\eta/(1-\eta) \le t_0$*, then*

$$\mathrm{err}(\widehat{y}, y) \le \Psi\left(\frac{2\eta}{1-\eta}\right).$$

The analogous statement holds with $\eta_G + \beta$ *in place of* η *under Corollary B.15.*

Proof. Substitute the established normalized graph bound into the stability contract and use monotonicity of Ψ. □

What is and is not established. The lemmas and graph perturbation theorem above are proved here. The clustering contract is not. In particular, it includes reference graph recoverability even at $t = 0$, and cannot be inferred merely from using true tangent projectors. A concrete spectral-clustering guarantee must establish within-manifold connectivity, controlled cross-manifold weights, a suitable spectral separation, and a justified rounding rule. Arbitrary k-means initialization is not automatically covered. Section B.11 supplies a proved sufficient realization with an explicit embedding and approximate-rounding requirement; it does not establish this contract for an arbitrary clustering routine.

B.10 A1 SCALE AND GEOMETRIC REFERENCE-GRAPH SEPARATION

Lemma B.19 (A deterministic A1 scale interval)**.** *Suppose $I_+ \neq \varnothing$. For some $c_s \in (0, 1)$, assume that for strictly more than half the indices $i \in I_+$, strictly more than half the distances in $\mathcal{D}_i$ are at least $c_s r$. Then*

$$c_s r \leq \widehat{\sigma}_s \leq r.$$

This holds for the conventional median averaging the two central order statistics when the number of entries is even.

Proof. Each qualifying inner median is at least $c_s r$. A strict majority of the outer entries therefore have this property, so their median also does. Every entry is positive and at most r. □

Remark B.20 (Interpretation). This empirical sufficient condition prevents most local distance lists from concentrating almost entirely near zero. A sampling theorem for it is not established here. A1 summarizes distances throughout a radius-r neighborhood; it is not a nearest-neighbor spacing estimator and need not vanish with sample size when r is fixed.

Lemma B.21 (Within-manifold edge lower bound)**.** *Assume the scale interval of Lemma B.19. For relevant same-manifold pairs suppose*

$$\|P_i^* - P_j^*\|_F \leq L_T \|q_i - q_j\|.$$

Fix $\vartheta \in (0, 1]$. If $y_i = y_j$, $i \neq j$, and $\|x_i - x_j\| \leq \vartheta r$, then

$$W_{ij}^* \geq w_{\text{in}} := \exp\left(-\frac{\vartheta^2}{c_s^2} - \frac{L_T^2(\vartheta r + 2\nu)^2}{2\sigma_T^2}\right). \tag{26}$$

Proof. The noise bound gives $\|q_i - q_j\| \leq \vartheta r + 2\nu$, so $d_T^2(P_i^*, P_j^*) \leq L_T^2(\vartheta r + 2\nu)^2/2$. Also $\|x_i - x_j\|^2/\widehat{\sigma}_s^2 \leq \vartheta^2/c_s^2$. Substitute these into the reference affinity. □

Lemma B.22 (Cross-manifold edge upper bound)**.** *Suppose every supported cross-manifold pair satisfies*

$$y_i \neq y_j, \quad \|x_i - x_j\| \leq r \quad \Longrightarrow \quad d_T(P_i^*, P_j^*) \geq \Delta_T > 0.$$

Then every cross-manifold pair obeys

$$W_{ij}^* \leq w_{\text{out}} := \exp(-\Delta_T^2/\sigma_T^2).$$

Proof. The spatial factor is at most one and the tangent factor is bounded by the displayed quantity. Unsupported pairs have zero weight. □

Remark B.23 (Uniform separation is additional). A tangent mismatch at the intersection itself does not automatically imply this uniform bound for all supported observed cross-manifold pairs. Local geometry and neighborhood size must justify that extension. Positive chordal separation allows shared tangent directions.

Proposition B.24 (Cross-degree relative to within-degree)**.** *Define*

$$m_i = \#\{j \neq i : y_j = y_i,\ \|x_i - x_j\| \leq \vartheta r\},$$
$$b_i = \#\{j : y_j \neq y_i,\ \|x_i - x_j\| \leq r\},$$
$$d_i^{\text{in}} = \sum_{j:y_j=y_i} W_{ij}^*, \qquad d_i^{\text{out}} = \sum_{j:y_j\neq y_i} W_{ij}^*.$$

Under Lemmas B.21 and B.22, if $m_i > 0$, then

$$\frac{d_i^{\text{out}}}{d_i^{\text{in}}} \leq \frac{b_i}{m_i} \exp\left(\frac{\vartheta^2}{c_s^2} + \frac{L_T^2(\vartheta r + 2\nu)^2/2 - \Delta_T^2}{\sigma_T^2}\right). \tag{27}$$

Proof. Summing the edge bounds gives $d_i^{\text{in}} \geq m_i w_{\text{in}}$ and $d_i^{\text{out}} \leq b_i w_{\text{out}}$. Divide them. □

B.10.1 The block reference graph

Define an analysis-only graph

$$W_{ij}^0 = W_{ij}^* \mathbf{1}_{\{y_i=y_j\}}, \qquad d_i^0 = d_i^{\text{in}}, \qquad D^0 = \text{diag}(d_i^0), \qquad S^0 = (D^0)^{-1/2} W^0 (D^0)^{-1/2}.$$

The algorithm does not use true labels to construct this graph. If the observed radius-ϑr graph is connected inside each true cluster, the positive lower bound in Lemma B.21 ensures connected blocks of W^0. This alone does not imply a uniform spectral gap.

Proposition B.25 (Distance to the block reference graph)**.** *Assume $d_i^{\text{in}} > 0$ for every i, and let*

$$\rho = \max_i \frac{d_i^{\text{out}}}{d_i^{\text{in}}} < 1.$$

Then

$$\|S^* - S^0\|_{\text{op}} \leq \frac{2\rho}{1-\rho}.$$

Under Theorem B.13, with $\eta < 1$, consequently

$$\|S - S^0\|_{\text{op}} \leq \delta, \qquad \delta := \frac{2\eta}{1-\eta} + \frac{2\rho}{1-\rho}. \tag{28}$$

Alternatively replace η by $\eta_G + \beta < 1$ under Corollary B.15.

Proof. The absolute row error between W^* and W^0 equals d_i^{out}. Apply Proposition B.12 using W^0 as the reference graph. Then apply the triangle inequality and the previously proved estimation bound. □

B.11 A spectral realization of the abstract clustering stage

The abstract module remains a general clustering routine. This section provides one explicit sufficient realization: bottom-M eigenvectors of the symmetric normalized Laplacian, row normalization, and approximate M-means. A concrete implementation must be checked for identical embedding and normalization conventions before invoking these results.

B.11.1 Quantitative within-cluster connectivity

Write $I_k = \{i : y_i = k\}$ and $n_k = |I_k| \geq 2$. Assume every block of W^0 is connected. Define

$$L^0 = I - S^0, \qquad L = I - S, \qquad g = \min_{1\leq k\leq M} \lambda_2(L_k^0) > 0.$$

Here g denotes the spectral gap, to avoid confusion with the anchor cover ratio γ. Then L^0 has exactly M zero eigenvalues and $\lambda_{M+1}(L^0) = g$. Suppose $\|L - L^0\|_{\text{op}} \leq \delta$.

Proposition B.26 (Stability of the bottom eigenspace)**.** *Let $U^0 \in \mathbb{R}^{n\times M}$ be an orthonormal basis of $\ker L^0$ and $U \in \mathbb{R}^{n\times M}$ contain orthonormal bottom-M eigenvectors of L. If $\delta < g/2$, some orthogonal matrix $O \in \mathbb{R}^{M\times M}$ satisfies*

$$\|U - U^0 O\|_F^2 \leq \frac{2M\delta^2}{(g-\delta)^2}. \tag{29}$$

Proof. On the nullspace of L^0, every unit vector has Rayleigh quotient for L at most δ. The variational eigenvalue characterization gives $\lambda_M(L) \le \delta$. Applying the same characterization to an operator-norm perturbation gives $\lambda_{M+1}(L) \ge g - \delta > \delta$.

Let $Q = I - U^0(U^0)^\top$. For a bottom eigenvector u_j with eigenvalue $0 \le \lambda_j \le \delta$,

$$(L^0 - \lambda_j I)Qu_j = Q(L^0 - L)u_j.$$

On the range of Q, the left operator has eigenvalues at least $g - \delta$, so $\|Qu_j\| \le \delta/(g-\delta)$. Thus $\|QU\|_F^2 \le M\delta^2/(g-\delta)^2$. If θ_j are the principal angles of the column spaces, the singular value decomposition of $(U^0)^\top U$ gives

$$\min_{O^\top O = I} \|U - U^0 O\|_F^2 = 2\sum_j (1 - \cos\theta_j) \le 2\sum_j \sin^2\theta_j = 2\|QU\|_F^2.$$

□

B.11.2 Reference row geometry

Let $\mathrm{vol}_k = \sum_{i \in I_k} d_i^0$. Choose

$$U_{ik}^0 = \sqrt{d_i^0 / \mathrm{vol}_k}\, \mathbf{1}_{\{i \in I_k\}}.$$

The columns are orthonormal and span the nullspace. For $i \in I_k$, set $a_i = \sqrt{d_i^0/\mathrm{vol}_k} > 0$. Then $U_i^0 = a_i e_k^\top$, and its row-normalized version is $V_i^0 = e_k^\top$. Distinct reference centers are separated by $\sqrt{2}$. Let $a_{\min} = \min_i a_i$.

Lemma B.27 (Error after row normalization)**.** *Normalize nonzero rows of U to form V, assigning a fixed unit vector to any zero row. For the alignment above,*

$$\|V_i - V_i^0 O\| \le \frac{2\|U_i - U_i^0 O\|}{a_i}.$$

Consequently

$$E := \|V - V^0 O\|_F^2 \le 4\sum_i \frac{\|U_i - U_i^0 O\|^2}{a_i^2} \le \frac{8M\delta^2}{a_{\min}^2 (g-\delta)^2}. \tag{30}$$

Proof. For $u \ne 0$ and $\|v\| = a_i$,

$$\left\| \frac{u}{\|u\|} - \frac{v}{a_i} \right\| \le \frac{|a_i - \|u\||}{a_i} + \frac{\|u - v\|}{a_i} \le \frac{2\|u - v\|}{a_i}.$$

If $u = 0$, the left side is at most two, equal to the right-side bound. Square, sum, and use Proposition B.26. □

Proposition B.28 (Rounding by approximate M-means)**.** *Let reference centers satisfy $\|c_k^* - c_\ell^*\| = \sqrt{2}$ for $k \ne \ell$, and set $E = \sum_i \|V_i - c_{y_i}^*\|^2$. Suppose returned centers and labels satisfy*

$$\sum_i \|V_i - \widehat{c}_{\widehat{y}_i}\|^2 \le \alpha \min_{c_1,\dots,c_M,\ell_1,\dots,\ell_n} \sum_i \|V_i - c_{\ell_i}\|^2, \qquad \alpha \ge 1.$$

If $2(1+\sqrt{\alpha})^2 E < n_{\min} := \min_k n_k$, then

$$\mathrm{err}(\widehat{y}, y) \le \frac{2(1+\sqrt{\alpha})^2 E}{n}. \tag{31}$$

Proof. The reference centers and labels are feasible, so the returned objective is at most αE. The triangle inequality for stacked row vectors gives

$$\left(\sum_i \|\widehat{c}_{\widehat{y}_i} - c_{y_i}^*\|^2 \right)^{1/2} \le (1 + \sqrt{\alpha})\sqrt{E}.$$

Let B_{round} be the rows with $\|\widehat{c}_{\widehat{y}_i} - c^*_{y_i}\| \geq 1/\sqrt{2}$. Then $|B_{\text{round}}| \leq 2(1+\sqrt{\alpha})^2 E < n_{\min}$, so every true cluster contains a good row. One estimated center cannot serve good rows from two distinct true clusters: the triangle inequality would make the reference centers closer than $\sqrt{2}$. Each of the M true clusters therefore requires a distinct estimated center. Since there are only M estimated centers, this establishes a bijection. After label permutation, all good rows are correct; only bad rows can be errors. □

Theorem B.29 (Explicit conditional spectral recovery)**.** *Suppose W^0 has M connected positive-degree blocks with spectral gap g and $\|S - S^0\|_{\text{op}} \leq \delta < g/2$. Use bottom-$M$ eigenvectors of L, row normalization, and the α-approximate M-means procedure in Proposition B.28. If*

$$\frac{16(1+\sqrt{\alpha})^2 M\delta^2}{a_{\min}^2 (g-\delta)^2} < n_{\min}, \tag{32}$$

then

$$\text{err}(\widehat{y}, y) \leq \frac{16(1+\sqrt{\alpha})^2 M}{n a_{\min}^2} \frac{\delta^2}{(g-\delta)^2}. \tag{33}$$

Equation (28) *supplies one sufficient value of δ when its hypotheses hold.*

Proof. Choose the reference cluster centers as the rows of O, which have pairwise distance $\sqrt{2}$. Apply Lemma B.27 and Proposition B.28. Condition (32) implies the rounding small-error condition. □

Remark B.30 (Solver obligations). This theorem specifies one sufficient realization, not an automatic guarantee for a named library implementation. Its exact embedding and normalization conventions must match. A finite number of Lloyd k-means restarts is not proved here to achieve any stated α. The theorem also uses exact eigenvectors, without a numerical eigensolver-error term.

B.12 Degree balance and conditional consistency

Lemma B.31 (Degree balance and reference row norms)**.** *If for every cluster*

$$\frac{\max_{i \in I_k} d_i^0}{\min_{i \in I_k} d_i^0} \leq \chi, \qquad \chi \geq 1,$$

then for $i \in I_k$,

$$a_i^2 \geq \frac{1}{\chi n_k}, \qquad a_{\min}^2 \geq \frac{1}{\chi n_{\max}}.$$

Proof. $\text{vol}_k \leq n_k \max_{j \in I_k} d_j^0 \leq \chi n_k d_i^0$. Divide d_i^0 by this upper bound. □

Corollary B.32 (Balanced-degree recovery bound)**.** *Under the graph and solver hypotheses of Theorem B.29 and the degree-ratio condition, sufficient small-error conditions are*

$$\delta < g/2, \qquad 16(1+\sqrt{\alpha})^2 M\chi \frac{n_{\max}}{n_{\min}} \frac{\delta^2}{(g-\delta)^2} < 1.$$

Then

$$\text{err}(\widehat{y}, y) \leq 16(1+\sqrt{\alpha})^2 M\chi \frac{n_{\max}}{n} \frac{\delta^2}{(g-\delta)^2}.$$

Proof. Substitute $a_{\min}^{-2} \leq \chi n_{\max}$ into (32) and (33). □

Corollary B.33 (Conditional misclassification consistency)**.** *Consider a sequence of instances satisfying the preceding representation, reference-graph, and solver hypotheses. Suppose M_n, α_n, χ_n, and $n_{\max,n}/n_{\min,n}$ are uniformly bounded, and*

$$\delta_n / g_n \longrightarrow 0.$$

Then the small-error conditions eventually hold and $\text{err}(\widehat{y}_n, y_n) \to 0$.

Proof. $\delta_n/(g_n - \delta_n) = (\delta_n/g_n)/(1 - \delta_n/g_n) \to 0$. Apply Corollary B.32 and $n_{\max,n}/n \leq 1$. □

Remark B.34 (Scope of consistency). This is a deterministic conditional sequence result. If its hypotheses hold on events of probability tending to one and the bounds vanish on those events, it yields convergence in probability of the error fraction. It is not yet a sampling consistency theorem for the concrete algorithm. Vanishing error fraction is weaker than exact recovery of every label. For example, a strict fraction bound below $1/n$ would force zero errors by integrality; convergence to zero alone does not.

B.13 Shrinking-radius regimes and limitations of the bound

Proposition B.35 (A spectral-gap upper bound for a local graph)**.** *Consider one connected positive-degree block of W^0, with edges only between observed points at distance at most r. For a 1-Lipschitz function f of the observed coordinates, let $f_i = f(x_i)$ and*

$$\bar{f} = \frac{\sum_i d_i^0 f_i}{\mathrm{vol}}, \qquad v_f = \frac{\sum_i d_i^0 (f_i - \bar{f})^2}{\mathrm{vol}} > 0.$$

Then the block's symmetric normalized Laplacian satisfies

$$\lambda_2 \le \frac{r^2}{2v_f}.$$

Proof. The Rayleigh test vector $z_i = \sqrt{d_i^0}(f_i - \bar{f})$ is orthogonal to the null vector. Its quotient is

$$\frac{\frac{1}{2}\sum_{ij} W_{ij}^0 (f_i - f_j)^2}{\sum_i d_i^0 (f_i - \bar{f})^2}.$$

On every edge $|f_i - f_j| \le r$, and $\sum_{ij} W_{ij}^0 = \mathrm{vol}$. The numerator is at most $r^2\,\mathrm{vol}\,/2$, giving the claim. □

If a suitable v_f stays bounded below as $r \to 0$, the gap cannot be treated as a fixed positive constant. This is an upper bound only: a matching lower bound is not established here. If a subsequent analysis shows $g_n \asymp r_n^2$, the sufficient condition $\delta_n/g_n \to 0$ would require $\delta_n = o(r_n^2)$.

B.13.1 Fixed tangent bandwidth and maximum cross-degree

The present leakage control uses

$$\rho_n = \max_i d_i^{\mathrm{out}}/d_i^{\mathrm{in}}.$$

Near an intersection the within- and cross-manifold neighbor counts can be comparable. At fixed tangent bandwidth, a suppression factor such as $\exp(-\Delta_T^2/0.2^2)$ is a positive constant rather than a vanishing quantity. Increasing sample size and shrinking r alone therefore do not establish $\rho_n \to 0$. In suitable crossing configurations, leakage at the most difficult nodes can remain nonzero even when most samples lie far from the intersection.

This is a limitation of the maximum-row operator-norm argument, not a proof of inconsistency of the concrete algorithm. The finite-sample results remain valid. A regime $\sigma_{T,n} \to 0$ would differ from the fixed-bandwidth implementation and would also require controlling within-manifold affinity loss and increased estimation sensitivity.

B.13.2 Motivation for intersection-localized analysis

Let

$$\mathcal{I} = \bigcup_{k \ne \ell} (\mathcal{M}_k \cap \mathcal{M}_\ell).$$

The next section controls aggregate leakage without requiring a small maximum row error everywhere. Under an explicit localization assumption, it bounds leakage by the weighted mass of a neighborhood of $\mathcal{I}$. Deriving localization, mass rates, and estimator accuracy from sampling and manifold geometry remains a separate task. Small sample mass alone is insufficient: these nodes may create strong cross-cluster bridges or provide essential within-cluster connections. Localization also requires assumptions excluding other near-contacts.

B.14 CLUSTER-LEVEL LEAKAGE ON THE ACTUAL GRAPH

This is an alternative to the operator-norm perturbation route. It acts directly on the actual estimated adjacency W, retains intersection nodes, and permits substantial leakage at individual nodes. It still requires quantitatively connected within-cluster graphs and bounded degree inflation. It does not by itself establish that local estimators produce graphs satisfying these conditions.

B.14.1 ACTUAL WITHIN-CLUSTER DEGREES AND AGGREGATE LEAKAGE

To distinguish these quantities from the reference degrees used earlier, define

$$p_i = \sum_{j:y_j=y_i} W_{ij}, \qquad t_i = \sum_{j:y_j\neq y_i} W_{ij}, \qquad d_i = p_i + t_i,$$

$$v_k = \sum_{i\in I_k} d_i, \qquad q_k = \frac{\sum_{i\in I_k} t_i}{v_k}, \qquad \tau = \sum_{k=1}^{M} q_k.$$

Thus p_i, t_i are actual within- and cross-degrees, whereas $d_i^{\mathrm{in}}, d_i^{\mathrm{out}}$ earlier concerned W^*. Assume $n_k \geq 2$, $p_i > 0$, and connected actual within-cluster graphs. Let L_k^{in} denote their normalized Laplacians, normalized using p_i, and define

$$g_{\mathrm{in}} = \min_k \lambda_2(L_k^{\mathrm{in}}) > 0.$$

Assume bounded, but not necessarily vanishing, degree inflation:

$$t_i \leq R_{\mathrm{out}} p_i \quad \text{for every } i, \qquad 0 \leq R_{\mathrm{out}} < \infty. \tag{34}$$

The constant R_{out} is unrelated to the local affine radius R in Lemma B.2. It need not be less than one.

Define the true-partition subspace using the actual total degrees:

$$F_{ik} = \sqrt{d_i/v_k}\,\mathbf{1}_{\{i\in I_k\}}, \qquad F^\top F = I_M.$$

The partition and F are analysis-only objects. As before, $L = I - D^{-1/2}WD^{-1/2}$, and U contains orthonormal bottom-M eigenvectors of L.

Lemma B.36 (Energy of the partition subspace)**.** *The partition subspace satisfies*

$$\mathrm{tr}(F^\top LF) = \tau, \qquad \mathrm{tr}(U^\top LU) = \sum_{j=1}^{M} \lambda_j(L) \leq \tau.$$

Proof. For column f_k of F, the degree-rescaled entries are $d_i^{-1/2}(f_k)_i = \mathbf{1}_{\{i\in I_k\}}/\sqrt{v_k}$. In the quadratic form

$$f_k^\top L f_k = \tfrac{1}{2}\sum_{ij} W_{ij}\left(\frac{(f_k)_i}{\sqrt{d_i}} - \frac{(f_k)_j}{\sqrt{d_j}}\right)^2,$$

within-cluster terms vanish, while crossing terms give exactly $\sum_{i\in I_k} t_i/v_k = q_k$. Summing proves the equality. The variational principle for the sum of the smallest M eigenvalues, applied to the feasible orthonormal matrix F, proves the inequality. □

Lemma B.37 (Within-cluster coercivity)**.** *Under* (34)*, every $z \in \mathbb{R}^n$ satisfies*

$$z^\top Lz \geq \frac{g_{\mathrm{in}}}{1+R_{\mathrm{out}}}\|(I-FF^\top)z\|^2. \tag{35}$$

Proof. Set $h_i = z_i/\sqrt{d_i}$. Discarding the nonnegative cross-edge energy,

$$z^\top Lz \geq \sum_k \tfrac{1}{2}\sum_{i,j\in I_k} W_{ij}(h_i-h_j)^2.$$

For each block the variational characterization of its normalized spectral gap gives

$$\tfrac{1}{2}\sum_{i,j\in I_k} W_{ij}(h_i-h_j)^2 \geq g_{\mathrm{in}} \min_{c\in\mathbb{R}} \sum_{i\in I_k} p_i(h_i-c)^2.$$

Since $d_i \le (1 + R_{\text{out}})p_i$, this is at least

$$\frac{g_{\text{in}}}{1 + R_{\text{out}}} \min_{c \in \mathbb{R}} \sum_{i \in I_k} d_i (h_i - c)^2.$$

Summing the minima gives precisely the squared distance of z to the space of vectors constant on each cluster after degree rescaling. This space is $\text{range}(F)$, proving the claim. □

Theorem B.38 (Eigenspace control by aggregate leakage)**.** *Under the hypotheses of this section,*

$$\|(I - FF^\top)U\|_F^2 \le \frac{(1 + R_{\text{out}})\tau}{g_{\text{in}}}. \tag{36}$$

There is an orthogonal matrix $O \in \mathbb{R}^{M \times M}$ such that

$$\|U - FO\|_F^2 \le \frac{2(1 + R_{\text{out}})\tau}{g_{\text{in}}}. \tag{37}$$

Proof. Apply Lemma B.37 to each column of U and sum. Lemma B.36 bounds the resulting energy by τ. For the second inequality apply the principal-angle alignment inequality proved in Proposition B.26, now to U and F. □

Remark B.39 (No operator-norm closeness is required). This result does not require L to be close in operator norm to a block-diagonal reference Laplacian. Nor does it require a vanishing maximum ratio t_i/p_i. It replaces that requirement by bounded degree inflation, within-cluster coercivity, and small aggregate leakage relative to g_{in}. The bound holds for any orthonormal selection of bottom-M eigenvectors, even without a separately assumed eigengap of L.

B.14.2 Row normalization and label recovery

Let

$$b_i^{\text{row}} = \sqrt{d_i / v_{y_i}}, \qquad b_{\min}^{\text{row}} = \min_i b_i^{\text{row}} > 0.$$

These actual-degree row norms differ from the reference row norms a_i in Section B.11. Normalize rows of U to get V, using the same zero-row convention as before. Normalize rows of F to get $V_i^{\text{part}} = e_{y_i}^\top$. Lemma B.27, applied with F, implies

$$E_{\text{agg}} := \|V - V^{\text{part}} O\|_F^2 \le \frac{8(1 + R_{\text{out}})\tau}{(b_{\min}^{\text{row}})^2 g_{\text{in}}}. \tag{38}$$

Corollary B.40 (Clustering error from aggregate leakage)**.** *Use the α-approximate M-means rounding of Proposition B.28. If*

$$\frac{16(1 + \sqrt{\alpha})^2(1 + R_{\text{out}})\tau}{(b_{\min}^{\text{row}})^2 g_{\text{in}}} < n_{\min}, \tag{39}$$

then

$$\text{err}(\widehat{y}, y) \le \frac{16(1 + \sqrt{\alpha})^2(1 + R_{\text{out}})}{n(b_{\min}^{\text{row}})^2} \frac{\tau}{g_{\text{in}}}. \tag{40}$$

Proof. Apply Proposition B.28 with the orthonormal rows of O as reference centers and the error bound (38). □

If actual total degrees satisfy

$$\max_{i \in I_k} d_i / \min_{i \in I_k} d_i \le \chi \quad \text{for every } k,$$

the proof of Lemma B.31 gives $(b_{\min}^{\text{row}})^2 \ge 1/(\chi n_{\max})$. Consequently a sufficient version of (39) is

$$16(1 + \sqrt{\alpha})^2(1 + R_{\text{out}})\chi \frac{n_{\max}}{n_{\min}} \frac{\tau}{g_{\text{in}}} < 1. \tag{41}$$

Under this condition,

$$\text{err}(\widehat{y}, y) \le 16(1 + \sqrt{\alpha})^2(1 + R_{\text{out}})\chi \frac{n_{\max}}{n} \frac{\tau}{g_{\text{in}}}. \tag{42}$$

Thus, along sequences with uniformly bounded α, R_{out}, χ, and $n_{\max}/n_{\min}$, the condition $\tau_n / g_{\text{in},n} \to 0$ implies vanishing error fraction under the graph and solver hypotheses. This follows directly from (41)–(42).

B.14.3 LEAKAGE LOCALIZED NEAR AN INTERSECTION

Proposition B.41 (Leakage bounded by neighborhood mass)**.** *Let $B \subseteq \{1, \dots, n\}$ be a designated intersection-neighborhood sample set. Suppose every positive-weight cross-manifold edge has both endpoints in B. Then*

$$q_k \le \frac{\sum_{i \in I_k \cap B} d_i}{v_k}. \tag{43}$$

If the actual total-degree ratio within each cluster is at most χ, then

$$q_k \le \chi \frac{|I_k \cap B|}{n_k}, \qquad \tau \le \chi \sum_k \frac{|I_k \cap B|}{n_k}. \tag{44}$$

Proof. Localization gives $t_i = 0$ for $i \notin B$, and always $t_i \le d_i$. Hence $\sum_{i \in I_k} t_i \le \sum_{i \in I_k \cap B} d_i$, proving (43). Bound its numerator by $|I_k \cap B| \max_{i \in I_k} d_i$ and denominator by $n_k \min_{i \in I_k} d_i$ to obtain (44). □

Remark B.42 (What localization still requires). The proposition does not show that radius-supported cross edges actually have this localization. Manifolds can pass close to each other away from their intersections. A geometric separation condition, together with noise control, is needed to identify an appropriate B and prove that all supported cross pairs lie there. The set B is used only in analysis; the clustering algorithm neither knows it nor deletes its nodes.

B.14.4 RATE IMPLICATIONS AND UNRESOLVED LOW-CODIMENSION CASES

For illustration only, suppose a later geometric and sampling analysis establishes, for fixed M and bounded degree ratios,

$$|I_k \cap B|/n_k = O(r^c), \qquad g_{\text{in}} \asymp r^2.$$

The mass bound then implies

$$\tau / g_{\text{in}} = O(r^{c-2}).$$

This sufficient bound vanishes for $c > 2$, but does not imply convergence when $c \le 2$, including codimension-one intersection neighborhoods. These rates are not general geometric or sampling results. Section B.16 below proves their codimension-one population analogue for uniform straight segments. A nonvanishing upper bound alone does not establish algorithmic inconsistency. The crude mass bound also discards any additional suppression of cross weights by the tangent Gaussian. A fixed suppression factor can improve finite-sample constants without supplying an extra vanishing power of r.

The following sections compare crossing-edge energy with internal connections at fixed tangent bandwidth. They identify genuine graph failures at small angles and a certified population regime at a right angle, while keeping the very-small-radius surrogate analysis separate from results about the original graph.

B.15 INTEGRATED CONDITIONAL RECOVERY THROUGH MULTIPLICATIVE COMPARISON

This section combines the established local representation bounds with aggregate leakage. It compares the estimated graph to an oracle graph on the same observed sample, with the same radius support and empirical A1 scale. It does not assume that the oracle graph is block diagonal. All graphs in this section use the loop-free normalization defined above.

Proposition B.43 (Transfer of graph conditions from the oracle)**.** *Suppose, entrywise,*

$$e^{-\xi} W^* \le W \le e^{\xi} W^*, \qquad 0 \le \xi < \infty.$$

Assume every oracle within-cluster graph is connected with positive degrees. Write $p_i^*, t_i^*, d_i^* = p_i^* + t_i^*$ for its within, cross, and total degrees, and define

$$v_k^* = \sum_{i \in I_k} d_i^*, \qquad \tau^* = \sum_k \frac{\sum_{i \in I_k} t_i^*}{v_k^*}, \qquad R^* = \max_i \frac{t_i^*}{p_i^*}, \qquad \chi^* = \max_k \frac{\max_{i \in I_k} d_i^*}{\min_{i \in I_k} d_i^*}.$$

Let $g^* > 0$ be the minimum normalized gap of the oracle within-cluster blocks, normalized by p_i^*. The actual within-cluster graphs are connected with positive degrees. Taking $R_{\mathrm{out}} = \max_i t_i/p_i$ and $\chi = \max_k(\max_{i\in I_k} d_i/\min_{i\in I_k} d_i)$, the corresponding quantities obey

$$g_{\mathrm{in}} \geq e^{-2\xi} g^*, \qquad \tau \leq e^{2\xi}\tau^*, \qquad R_{\mathrm{out}} \leq e^{2\xi} R^*, \qquad \chi \leq e^{2\xi}\chi^*. \tag{45}$$

Proof. Positive edge support is identical, so connectivity is preserved. Summing the entrywise bounds gives the same multiplicative bounds for within degrees, cross degrees, total degrees, and cluster volumes. Taking ratios proves the last three inequalities, with zero cross numerators causing no difficulty. For a nonconstant scalar vector h on a block, write

$$\mathcal{E}_W(h) = \tfrac{1}{2}\sum_{i,j\in I_k} W_{ij}(h_i - h_j)^2, \qquad V_p(h) = \min_c \sum_{i\in I_k} p_i(h_i - c)^2.$$

Then $\mathcal{E}_W(h) \geq e^{-\xi}\mathcal{E}_{W^*}(h)$ and $V_p(h) \leq e^{\xi} V_{p^*}(h)$. The variational gap characterization therefore gives the first inequality. □

Theorem B.44 (Integrated conditional spectral recovery)**.** *Suppose representation and A1 construction complete successfully, and the local-projector, actually selected anchor, branch-margin, and tangent transport hypotheses of Theorem B.13 hold. Set*

$$\varepsilon = \max\{\varepsilon_{\mathrm{loc}}, \varepsilon_{\mathrm{fit}} + L_T h\}, \qquad \xi = \frac{2\sqrt{d}\,\varepsilon}{\sigma_T^2}.$$

Assume the oracle graph satisfies the connected positive-degree block hypotheses of Proposition B.43, with $n_k \geq 2$. Use exact bottom-M eigenvectors of the symmetric normalized Laplacian of W, row normalization, and the α-approximate M-means contract of Proposition B.28. Define

$$C_{\mathrm{rec}} = 16(1 + \sqrt{\alpha})^2 \chi^* e^{6\xi}(1 + e^{2\xi} R^*).$$

If

$$C_{\mathrm{rec}} \frac{n_{\max}}{n_{\min}} \frac{\tau^*}{g^*} < 1, \tag{46}$$

then

$$\mathrm{err}(\widehat{y}, y) \leq C_{\mathrm{rec}} \frac{n_{\max}}{n} \frac{\tau^*}{g^*}. \tag{47}$$

There is no additional requirement that $e^{\xi} - 1 < 1$.

Proof. The local hypotheses imply uniform tangent error at most ε. Lemma B.11 gives the multiplicative comparison with exponent ξ. Proposition B.43 implies

$$(1 + R_{\mathrm{out}})\chi \frac{\tau}{g_{\mathrm{in}}} \leq (1 + e^{2\xi} R^*)\chi^* e^{6\xi} \frac{\tau^*}{g^*}.$$

Substitute into (41) and (42). This uses multiplicative energy comparison, not the small-row-error normalized-operator bound. □

Corollary B.45 (Conditional consistency of the integrated bound)**.** *Along a sequence satisfying the theorem's local, graph, and solver hypotheses, suppose ξ_n, α_n, R_n^*, χ_n^*, and $n_{\max,n}/n_{\min,n}$ are uniformly bounded. If $\tau_n^*/g_n^* \to 0$, then the smallness condition eventually holds and $\mathrm{err}(\widehat{y}_n, y_n) \to 0$.*

Proof. The constants $C_{\mathrm{rec},n}$ are uniformly bounded, so apply (46)–(47). □

Remark B.46 (Interpretation and limits). This route does not require tangent error to be $o(r_n^2)$; bounded ξ_n is sufficient for the stated conditional sequence result. Its exponential constants can nevertheless be very large. It neither proves the local estimation hypotheses nor makes oracle recoverability automatic. In particular it does not resolve the orthogonal crossing's very-small-radius regime, where τ^*/g^* does not vanish. Probability statements require these hypotheses on sampling events with controlled probabilities. The conclusion is vanishing error fraction, not unconditional exact recovery. The theorem also does not establish a finite-restart Lloyd guarantee or account for numerical eigenvector error.

B.16 Oracle crossing segments: population cuts and leakage

The examples in this and the following sections separate graph geometry from estimation error. Every observation is given its exact generating tangent projector; neither local PCA nor a branch-fitting solver is assumed to produce it. The Gaussian convention remains exactly (13), including the absence of a factor two in the exponents. These examples do not modify the algorithm.

B.16.1 Compact domains, sampling, and population A1

Let $e_1, e_2 \in \mathbb{R}^2$ be unit vectors with $e_1^\top e_2 = \cos\theta$, where $0 < \theta \le \pi/2$, and set

$$\mathcal{M}_k = \{se_k : -L \le s \le L\}, \qquad k = 1, 2.$$

These are smooth compact one-dimensional manifolds with boundary. Their endpoints are not identified. The statistical model takes N independent uniform observations on each segment, independently between segments, with no noise. Thus the two generating clusters are balanced. The origin has probability zero. Population integrals below use arclength on each segment; multiplying both measures by the same constant does not change normalized cuts or Rayleigh quotients. Population calculations are not silently identified with simultaneous $N \to \infty$, $r_N \to 0$ limits. Assume

$$r(1 + \csc\theta) < L/2. \tag{48}$$

Proposition B.47 (Population A1 for crossing segments)**.** *Under* (48)*, population A1 equals* $r/2$*. For fixed* L, r, θ *satisfying this condition, empirical A1 under the specified sampling model converges almost surely to* $r/2$ *as* $N \to \infty$*.*

Proof. A point at least r from its segment's endpoints and more than r from the other segment sees a uniform interval of length $2r$. Its conditional neighbor-distance distribution is uniform on $[0, r]$, with median $r/2$. On either segment the exceptional fraction is at most $r/L + r/(L\sin\theta) < 1/2$: the first term accounts for endpoints, and the second for $|s| \le r/\sin\theta$. Hence strictly more than half of the population local medians equal $r/2$, proving the first assertion. For the second assertion, the empirical distribution functions of the two one-dimensional samples converge uniformly. Ball intersections with either segment are intervals, so their empirical probabilities converge uniformly as well. At the regular centers above the limiting conditional distance distribution has positive slope at $r/2$ and a positive fixed neighborhood mass. Its empirical median therefore converges uniformly to $r/2$ on that set. Self-exclusion changes counts by at most one. The fraction of sampled centers in the regular set converges to a number strictly above one half; their converging medians force the outer median to converge to $r/2$. □

Define

$$K_r(u) = e^{-4u^2/r^2}\mathbf{1}_{\{0\le u\le r\}}, \qquad a_\theta = e^{-\sin^2\theta/0.2^2},$$
$$H = \int_0^1 ue^{-4u^2}\,du = \frac{1-e^{-4}}{8}, \qquad m_0 = \int_{-1}^1 e^{-4u^2}\,du,$$
$$J_\theta = \frac{2\pi H}{\sin\theta}.$$

The population within kernel is $K_r(|s-t|)$ and the cross kernel is $a_\theta K_r(\|se_1 - te_2\|)$. Cuts count each undirected edge once; volumes integrate the actual total degree, including cross degrees.

B.16.2 An exact comparison with a balanced competing partition

Let $\mathcal{P}_{\mathrm{true}}$ separate the two entire segments. Let $\mathcal{P}_{\mathrm{side}}$ put the positive halves of both segments in one cluster and the negative halves in the other.

Proposition B.48 (Exact cut comparison)**.** *Under* (48)*,*

$$C_{\mathrm{true}} = a_\theta J_\theta r^2, \tag{49}$$

$$C_{\mathrm{side}} = 2Hr^2 + \frac{\theta}{\pi}a_\theta J_\theta r^2. \tag{50}$$

Both partitions divide the actual total volume V equally. Consequently

$$\frac{\mathrm{Ncut}(\mathcal{P}_{\mathrm{side}})}{\mathrm{Ncut}(\mathcal{P}_{\mathrm{true}})} = \frac{\sin\theta}{\pi a_\theta} + \frac{\theta}{\pi}. \tag{51}$$

The competing partition has smaller normalized cut exactly when

$$a_\theta(\pi - \theta) > \sin\theta. \tag{52}$$

Its population misclassification fraction is $1/2$.

Proof. The linear change of variables $z = se_1 - te_2$ has absolute determinant $\sin\theta$. Every supported cross pair satisfies $|s|, |t| \le r/\sin\theta < L$, so the transformed integration domain is the full radius-r disk. Polar integration gives

$$C_{\mathrm{true}} = \frac{a_\theta}{\sin\theta} \int_{\|z\| \le r} e^{-4\|z\|^2/r^2}\, dz = a_\theta J_\theta r^2.$$

Splitting one segment at zero cuts weight $\int_{s,t\ge 0} K_r(s+t)\, ds\, dt = Hr^2$. The cross pairs with $st < 0$ map to two sectors of angle θ each, and hence contribute the fraction θ/π of the cross cut. Segment exchange and simultaneous sign reversal preserve total degrees; they imply equal volumes for both partitions. Each normalized cut is $4C/V$, proving the ratio and its strict inequality criterion. Each side cluster contains half of each true cluster, giving error $1/2$ under either label permutation. □

Angle	a_θ	Side/true normalized-cut ratio
10°	0.47056	0.17302
15°	0.18737	0.52303
20°	0.05369	2.13866
90°	$1.38879 \cdot 10^{-11}$	$2.29199 \cdot 10^{10}$

Angles in formulas are in radians. The common radial integral cancels from (51); merely changing the spatial Gaussian scale does not reverse this comparison when a common radial kernel is used and the same full-disk geometry applies.

Corollary B.49 (Fixed-radius empirical cut preference)**.** *For fixed L, r, θ satisfying* (48) *and the strict condition* (52)*, the empirical side partition has smaller normalized cut than the true partition eventually almost surely as $N \to \infty$.*

Proof. At the population bandwidth, bounded within-sample and between-sample pair averages converge almost surely to the cut and volume integrals. Proposition B.47 and uniform continuity of the Gaussian in a neighborhood of the positive bandwidth $r/2$ allow replacement by empirical A1. Volumes have positive limits. The strict limiting inequality therefore persists. □

Remark B.50 (A cut counterexample is not a rounding theorem). This proves a genuine preference of the normalized-cut objective for an incorrect partition at small crossing angles. It does not identify a particular spectral embedding's rounded labels. Spectral relaxation and subsequent M-means need separate analysis. No joint shrinking-radius sampling rate is asserted by Corollary B.49.

B.16.3 Leakage, localization, and the population internal gap

For one segment the within-degree volume and total graph volume are

$$B_{\mathrm{in}} = 2Lm_0 r - 2Hr^2, \qquad V = 2B_{\mathrm{in}} + 2a_\theta J_\theta r^2. \tag{53}$$

Thus the population analogue of aggregate leakage is

$$\tau = \frac{4a_\theta J_\theta r^2}{V} = \mathrm{Ncut}(\mathcal{P}_{\mathrm{true}}) \asymp \frac{a_\theta}{\sin\theta}\frac{r}{L} \quad (r/L \to 0,\ \theta \text{ fixed}). \tag{54}$$

Every supported cross pair has both endpoints in $|s| \le r/\sin\theta$. That tube has fraction $r/(L\sin\theta)$ on either segment. These are proved localization and mass statements for this model, not universal manifold assertions.

The population internal gap below is defined variationally by

$$g_{\mathrm{in}} = \inf_{f \text{ nonconstant}} \frac{\frac{1}{2}\int_{-L}^{L}\int_{-L}^{L} K_r(|s-t|)(f(s)-f(t))^2\,ds\,dt}{\min_c \int_{-L}^{L} p_r(s)(f(s)-c)^2\,ds}, \qquad p_r(s) = \int_{-L}^{L} K_r(|s-t|)\,dt.$$

Both segments have the same gap. Set $\varrho = r/L$ to avoid confusing this dimensionless radius with the earlier maximum cross-degree ratio.

Proposition B.51 (Explicit population interval-gap bounds)**.** *For* $0 < \varrho < 1/4$*,*

$$c_g \varrho^2 \le g_{\mathrm{in}} \le 3\varrho^2, \qquad c_g = \frac{8e^{-1}}{m_0(8.25)^3} > 0.0059418. \tag{55}$$

Proof. Partition $[-L, L]$ into $N = \lceil 8L/r \rceil$ equal intervals of length $\ell = 2L/N$. Within the same or adjacent intervals the distance is at most $2\ell \le r/2$, so the kernel is at least e^{-1}. Write m_j for interval means and $V_j = \int_{I_j}(f - m_j)^2$. The energy is at least

$$e^{-1}\left[\ell \sum_j V_j + \ell^2 \sum_{j=1}^{N-1} (m_{j+1} - m_j)^2\right].$$

The path Poincare inequality gives $\sum_j (m_j - \bar{m})^2 \le (N^2/4)\sum_j (m_{j+1} - m_j)^2$; indeed the path gap is $4\sin^2(\pi/(2N)) \ge 4/N^2$. Hence the energy is at least $(4e^{-1}\ell/N^2)\min_c \int (f-c)^2$. Since $p_r \le m_0 r$, division gives $g_{\mathrm{in}} \ge 4e^{-1}\ell/(m_0 r N^2)$. Using $N \le 8.25/\varrho$ yields the lower bound. For the upper bound use $f(s) = s$. Symmetry centers it at zero, and $m_0 r/2 \le p_r(s) \le m_0 r$ implies its degree-weighted variance is at least $L^2/6$. The local-support Rayleigh argument of Proposition B.35, with sums replaced by integrals, yields $g_{\mathrm{in}} \le 3(r/L)^2$. □

Consequently $\tau/g_{\mathrm{in}} \asymp (a_\theta/\sin\theta)L/r$ at fixed angle. This verifies the previously hypothetical codimension-one scaling for these population graphs. Neither this divergence nor a gap estimate for a population integral operator is, by itself, a finite-sample inconsistency result.

B.17 An exact spectral failure and its stability under distinct samples

This section uses the same segments but a different radius and sample design from Section B.16. It is a finite-scale counterexample, not a uniform-sampling shrinking-radius theorem.

Proposition B.52 (Four-point oracle spectral failure)**.** *Take* $\theta = 10°$*, the four points* $x_1 = Le_1, x_2 = -Le_1, x_3 = Le_2, x_4 = -Le_2$*, exact tangent projectors, and* $r = 2L$*. Use empirical A1 and remove self-loops before normalization. Bottom-two eigenvectors, row normalization, and any finite* α*-approximate two-means return the side partition, with error* $1/2$*.*

Proof. Each positive-distance list is $2L\sin(\theta/2), 2L\cos(\theta/2), 2L$, so $\widehat{\sigma}_s = 2L\cos(\theta/2)$. Put

$$w = e^{-\sec^2(\theta/2)}, \qquad b = a_\theta e^{-\tan^2(\theta/2)}, \qquad c = a_\theta e^{-1}.$$

Then

$$W = \begin{pmatrix} 0 & w & b & c \\ w & 0 & c & b \\ b & c & 0 & w \\ c & b & w & 0 \end{pmatrix}, \qquad d = w + b + c.$$

The nonconstant eigenvectors $(1,-1,1,-1)$, $(1,1,-1,-1)$, $(1,-1,-1,1)$ have normalized-Laplacian eigenvalues, respectively, $2(w+c)/d$, $2(b+c)/d$, $2(w+b)/d$. At $10°$,

$$(w, b, c) \approx (0.36507434, 0.46696708, 0.17310754),$$

and these eigenvalues are approximately 1.070850, 1.273592, 1.655558. The unique second eigenvector is the side direction. The normalized embedding has only the two distinct rows $(1,1)/\sqrt{2}$ and $(1,-1)/\sqrt{2}$. Two-means has optimum zero; a finite multiplicative approximation must also have objective zero and separate these two row types. Each contains one point from each true manifold, giving error $1/2$. □

Remark B.53 (Diagonal convention and estimator scope). For these four points the degrees are equal, so including unit self-loops also preserves the eigenvectors and their order. The result does not claim that the concrete branch estimator succeeds on this small sample. Exact tangents are supplied to isolate the graph and clustering stages.

Theorem B.54 (Failure persists for arbitrarily many distinct points)**.** *Keep $\theta = 10^\circ$ and $r = 2L$. In each of the four endpoint intervals*

$$\{se_k : (1-\epsilon)L \le s \le L\}, \qquad \{se_k : -L \le s \le -(1-\epsilon)L\}, \qquad k = 1, 2,$$

place m distinct points, giving $n = 4m$. Give them their exact tangent projectors and compute empirical A1. Use the loop-free spectral realization in Proposition B.52. Define

$$\gamma_0 = \frac{2(b-w)}{1+w+b+c} \approx 0.1016311, \qquad \delta_\epsilon = \frac{32\epsilon}{1-16\epsilon},$$

and

$$B_\epsilon = 16(1+\sqrt{\alpha})^2 \frac{\delta_\epsilon^2}{(\gamma_0 - \delta_\epsilon)^2}.$$

If $0 < \epsilon < 1/16$, $\delta_\epsilon < \gamma_0/2$, and $B_\epsilon < 1/2$, then

$$\mathrm{err}(\widehat{y}, y_{\mathrm{true}}) \ge \frac{1}{2} - B_\epsilon. \tag{56}$$

In particular, for any fixed finite α one can choose $\epsilon > 0$ independently of m so that the error is at least 0.4 for every m. For $\alpha = 1$ and $\epsilon = 10^{-5}$ it is greater than 0.49936.

Proof. As a comparison matrix only, collapse each group to its endpoint, giving weight 1 between different vertices in that group. Its common degree is $d_m = (m-1) + m(w+b+c)$. The three group-constant nonzero eigenvalues are

$$\frac{2m(w+c)}{d_m}, \quad \frac{2m(b+c)}{d_m}, \quad \frac{2m(w+b)}{d_m}.$$

The remaining eigenvalues, when $m > 1$, equal $1 + 1/d_m$. Direct comparison shows the second eigenvector is the side direction and $\lambda_3 - \lambda_2 \ge \gamma_0$ for every m. Its orthonormal bottom-two embedding has rows $n^{-1/2}(1, \pm 1)$.

Every actual pair distance differs from its collapsed value by at most $2\epsilon L$, and all pairs remain supported. The actual lists contain $4m - 1$ positive distances. In their limits there are $m - 1$ zero distances, followed by m distances of each of the three types in the preceding proof; the middle order statistic is still $\sigma_0 = 2L\cos(\theta/2)$. Stability of medians in the maximum norm therefore gives $|\widehat{\sigma}_s - \sigma_0| \le 2\epsilon L$. In particular the bandwidth along interpolation is at least L. For $G(u, \sigma) = e^{-u^2/\sigma^2}$,

$$|\partial_u G| \le \sqrt{2/e}/L, \qquad |\partial_\sigma G| \le 2/(eL).$$

Thus every weight changes by at most 4ϵ, including the change in A1. Since $w + b + c > 1$, the relative row error is at most 16ϵ. Proposition B.12 gives operator error at most δ_ϵ.

The projection-resolvent proof of Proposition B.26 also applies to a bottom-two reference space with nonzero second eigenvalue and gap γ_0: on its orthogonal complement the separation from each perturbed bottom eigenvalue is at least $\gamma_0 - \delta_\epsilon$. Hence the squared aligned embedding error is at most $4\delta_\epsilon^2/(\gamma_0 - \delta_\epsilon)^2$. Reference row norms are $\sqrt{2/n}$, so row normalization gives

$$E \le 8n \frac{\delta_\epsilon^2}{(\gamma_0 - \delta_\epsilon)^2}.$$

Apply Proposition B.28 with the side labels as reference. Its two centers are orthonormal and its cluster sizes are $n/2$. The stated smallness condition gives $\mathrm{err}(\widehat{y}, y_{\mathrm{side}}) \le B_\epsilon$. Permutation-invariant label distance satisfies the triangle inequality, and $\mathrm{err}(y_{\mathrm{side}}, y_{\mathrm{true}}) = 1/2$, proving the claim. At $\epsilon = 10^{-5}, \alpha = 1$, the bound is $B_\epsilon \approx 0.000638713$. □

Remark B.55 (Sampling interpretation of the robust example). The actual points in Theorem B.54 are distinct; collapsed points are only an analysis device. Both true within graphs are connected, sizes are balanced, and the reference eigengap is bounded away from zero uniformly in sample size. The failure is not caused by an eigenspace tie or a poor two-means local optimum. Balanced stratified sampling in the four short intervals satisfies the sample design almost surely. Under uniform sampling on the entire segments, the required placement has positive probability for every fixed finite sample size, but this does not give high-probability asymptotic failure for uniform sampling. The example uses $r = 2L$, not a shrinking local radius.

B.18 Orthogonal crossing: exact population sector analysis

Return to equal-density population segments, now with $\theta = \pi/2$ and $0 < \varrho = r/L < 1/4$. Write

$$a = e^{-25}, \qquad J = 2\pi H = \frac{\pi(1 - e^{-4})}{4}.$$

Population A1 is exactly $r/2$. Define

$$p_r(s) = \int_{-L}^{L} K_r(|s - t|)\, dt, \quad c_r(s) = \int_{-L}^{L} K_r(\sqrt{s^2 + t^2})\, dt, \quad d_r(s) = p_r(s) + ac_r(s).$$

For $|s| > r$, $c_r(s) = 0$. For $|s| \le r$, $p_r(s) = m_0 r$ and $c_r(s) \le m_0 r$. Consequently

$$ac_r(s) \le ap_r(s), \qquad \frac{\max_s d_r(s)}{\min_s d_r(s)} \le 2(1 + a). \tag{57}$$

The exact leakage and a uniform bound are

$$\tau = \frac{aJ\varrho}{m_0 - H\varrho + (aJ/2)\varrho} \le c_\tau a\varrho, \qquad c_\tau = \frac{J}{m_0 - H/4} \approx 0.905579. \tag{58}$$

For a pair of scalar functions $h = (h_1, h_2)$, define

$$\mathcal{E}_r(h) = \frac{1}{2} \sum_{k,\ell=1}^{2} \int_{-L}^{L} \int_{-L}^{L} W_{k\ell}(s, t)(h_k(s) - h_\ell(t))^2\, ds\, dt, \qquad \mathcal{D}_r(h) = \sum_k \int_{-L}^{L} d_r(s) h_k(s)^2\, ds.$$

Then $\mathcal{R}_r(h) = \mathcal{E}_r(h)/\mathcal{D}_r(h)$ is the normalized-Laplacian Rayleigh quotient after degree rescaling; its mean-zero constraint is $\sum_k \int d_r h_k = 0$. This is the continuum counterpart of writing a symmetric-normalized vector as $z_i = \sqrt{d_i} h_i$.

B.18.1 An exact comparison of label-preserving trial functions

Put $\phi(u) = \sin(\pi u/(2L))$ for $0 \le u \le L$ and define

$$h_{\mathrm{man},1}(s) = \phi(|s|), \quad h_{\mathrm{man},2}(s) = -\phi(|s|), \qquad h_{\mathrm{side},1}(s) = h_{\mathrm{side},2}(s) = \sin(\pi s/(2L)).$$

The first function has the correct label signs except at the origin; the second has side-partition signs. Both are degree-centered and have the same pointwise squares and hence the same denominator.

Proposition B.56 (A label-preserving trial function has smaller energy)**.** *Define*

$$A_r = \int_0^L \int_0^L K_r(s + t)\phi(s)\phi(t)\, ds\, dt, \qquad B_r = \int_0^L \int_0^L K_r(\sqrt{s^2 + t^2})\phi(s)\phi(t)\, ds\, dt.$$

Then

$$\mathcal{E}_r(h_{\mathrm{side}}) - \mathcal{E}_r(h_{\mathrm{man}}) = 8(A_r - aB_r), \qquad A_r/B_r \ge \frac{4}{3\pi^2} > a. \tag{59}$$

In particular $\mathcal{R}_r(h_{\mathrm{man}}) < \mathcal{R}_r(h_{\mathrm{side}})$.

Proof. Separating the four sign combinations on each line shows that within energy increases by $8A_r$ for the side function; its cross energy decreases by $8aB_r$. This gives the identity. On $[0, L]$, $u/L \le \phi(u) \le \pi u/(2L)$. If $J_3 = \int_0^1 u^3 e^{-4u^2}\, du$, the corresponding integrals with st instead of $\phi(s)\phi(t)$ are $r^4 J_3/6$ for the triangle $s + t \le r$ and $r^4 J_3/2$ for the quarter disk. Combining these integrals with the sine bounds yields the ratio bound. □

Let $m_2 = \int_{-1}^{1} u^2 e^{-4u^2}\, du$ and $D_{\mathrm{diff}} = m_2/(2m_0) \approx 0.0599045$. Taylor expansion on each half segment, treating the length-$O(r)$ endpoint and origin regions separately, gives for either trial function

$$\mathcal{R}_r(h) = D_{\mathrm{diff}} \frac{\pi^2}{4} \varrho^2 + O(\varrho^3).$$

More precisely, the exact difference above gives

$$\mathcal{R}_r(h_{\mathrm{side}}) - \mathcal{R}_r(h_{\mathrm{man}}) = \frac{\pi^2 J_3}{6m_0}(1-3a)\varrho^3 + o(\varrho^3), \quad J_3 = \frac{1-5e^{-4}}{32}. \tag{60}$$

Indeed $A_r = (\pi/(2L))^2 r^4 J_3/6 + O(r^6/L^4)$, $B_r = (\pi/(2L))^2 r^4 J_3/2 + O(r^6/L^4)$, and their common denominator is $2Lm_0 r(1+O(\varrho))$. The coefficient $\pi^2 J_3/(6m_0)$ is approximately 0.0529392. These are trial-function calculations, not identifications of the actual second eigenfunction.

B.18.2 Exact decomposition into one-dimensional sectors

For $s, t \in [0, L]$, set

$$A(s,t) = K_r(|s-t|)+K_r(s+t), \quad O(s,t) = K_r(|s-t|)-K_r(s+t), \quad B(s,t) = K_r(\sqrt{s^2+t^2}).$$

The degree-normalized random-walk integral operator is self-adjoint in $L^2(d_r(s)\, ds)$ and is unitarily equivalent to the symmetric-normalized operator. Reflections on either segment and segment exchange decompose it into the following invariant sectors:

Sector	Function form	Half-interval integral kernel
Manifold	Even on each segment, opposite between segments	$(A(s,t) - 2aB(s,t))/d_r(s)$
Odd	Odd on one segment; two identical copies	$O(s,t)/d_r(s)$
Common	Even and identical between segments	$(A(s,t) + 2aB(s,t))/d_r(s)$

Cross integrals of odd functions vanish exactly. Let λ_{man} and λ_{odd} be the lowest Laplacian eigenvalues in their respective sectors, and let $\lambda_{\mathrm{common},2}$ be the first nonconstant one in the common sector. Then

$$\lambda_2 = \min\{\lambda_{\mathrm{man}}, \lambda_{\mathrm{odd}}, \lambda_{\mathrm{common},2}\}. \tag{61}$$

The statements concern the low discrete spectrum of $I - T_r$; the integral operator T_r is compact, and in the regime below the relevant eigenvalues are separated from its Laplacian accumulation point 1.

Theorem B.57 (Population Fiedler sign separation at a quantified radius)**.** *If*

$$\tau < \frac{g_{\mathrm{in}}}{1+a}, \tag{62}$$

then the second population eigenvalue is simple and belongs to the manifold sector. After degree rescaling its eigenfunction has the form

$$h_1(s) = f(|s|), \qquad h_2(s) = -f(|s|), \qquad f > 0.$$

Its signs recover the generating labels almost everywhere. Moreover,

$$\lambda_2 \le \tau, \qquad \lambda_3 - \lambda_2 \ge \frac{g_{\mathrm{in}}}{1+a} - \tau > 0. \tag{63}$$

A sufficient explicit radius condition is

$$\frac{c_\tau}{c_g}(1+a)a < \varrho < \frac{1}{4}. \tag{64}$$

Its lower endpoint is approximately $2.11660 \cdot 10^{-9}$.

Proof. The manifold-sector trial function $(1, -1)$ has Rayleigh quotient τ, so $\lambda_{\mathrm{man}} \le \tau$. Every odd-sector function has zero degree-weighted mean separately on each segment. The same holds for common-sector functions orthogonal to the constant function. The integral version of Lemma B.37,

using (57), therefore bounds both competing sector quotients below by $g_{\mathrm{in}}/(1+a)$. Condition (62) selects the manifold sector.

Since $B(s,t) \le K_r(|s-t|)$ and $a < 1/2$,

$$A(s,t) - 2aB(s,t) \ge (1-2a)K_r(|s-t|) \ge 0.$$

This kernel connects the whole half interval through positive local steps. The compact positive integral operator is irreducible; its top eigenfunction is strictly positive and its top eigenvalue is simple. This proves the stated sign structure. Alternatively, positivity follows by replacing a maximizer by its absolute value and propagating positivity through its eigenfunction equation; equality and connectivity give simplicity. The integral coercivity bound on the orthogonal complement of the two-dimensional partition space and the min–max principle yield $\lambda_3 \ge g_{\mathrm{in}}/(1+a)$, giving the asserted gap. Equations (55) and (58) imply the explicit sufficient condition. □

Remark B.58 (What the radius threshold does not mean). This is a conservative sufficient condition in a population oracle model, not a phase transition or a parameter recommendation. Failure of the condition below its lower endpoint does not prove failure of sign recovery. For fixed $a > 0$, a sequence $\varrho \to 0$ eventually leaves this certified regime, so the theorem alone is not unrestricted shrinking-radius consistency. Finite samples also need a sampling and spectral approximation analysis.

B.18.3 Population approximate two-means, separately from signs

The sign rule in Theorem B.57 is not the concrete row-normalized two-means procedure. The latter has the following separate conditional bound. Use the row-normalized bottom-two population eigenfunctions, and assume centers and labels achieve a finite factor α of the optimal two-means objective integrated against the balanced uniform probability measure. Define population error by the probability of a wrong label, minimized over permutations.

Corollary B.59 (Population two-means error bound)**.** *If*

$$32(1+\sqrt{\alpha})^2(1+a)^2\frac{\tau}{g_{\mathrm{in}}} < 1,$$

then

$$\mathrm{err}_{\mathrm{pop}} \le 16(1+\sqrt{\alpha})^2(1+a)^2\frac{\tau}{g_{\mathrm{in}}} \le 16(1+\sqrt{\alpha})^2(1+a)^2\frac{c_\tau}{c_g}\frac{a}{\varrho}. \tag{65}$$

Proof. Replace sums in the aggregate energy, coercivity, alignment, and rounding proofs by integrals over the disjoint parameterized segments. The two cluster masses are $1/2$, and the degree ratio is at most $\chi = 2(1+a)$. For an orthonormal partition basis in L^2 of this probability measure, its squared row norm is at least $2/\chi$. Row normalization and the integral triangle and Markov inequalities give exactly the same rounding argument: if the bad set has mass less than $1/2$, both true clusters contain good points and force a bijection of centers. Substitution in (41) and (42), with size ratios replaced by mass ratios, gives the stated constants. The last bound uses (55) and (58). □

For $\alpha = 1$ and $\varrho = 10^{-3}$ the last error bound is less than $1.36 \cdot 10^{-4}$. This is an analytic population bound, not an observed experimental error. Finite Lloyd restarts are not automatically an α-approximate population or empirical solver.

B.18.4 Direct comparison with the odd sector at every small radius

The radius lower bound in Theorem B.57 is sufficient but is not needed for the following direct sector comparison. The proof uses the original nonlocal kernel throughout.

Proposition B.60 (Uniform ordering of the manifold and odd sectors)**.** *In the orthogonal population model, let*

$$c_0 = e^{-9/4}(1-e^{-2}), \qquad \kappa(a) = \frac{e^{-4}}{\left(1+4(1+a)m_0/c_0\right)^2}.$$

If $a < \kappa(a)$, then for every $0 < r/L < 1/4$,

$$\lambda_{\mathrm{man}} < \lambda_{\mathrm{odd}}. \tag{66}$$

In particular, this condition holds at $a = e^{-25}$, where $\kappa(a) \approx 1.16119 \cdot 10^{-5}$.

Proof. The odd-sector kernel $O(s,t)$ is nonnegative and connects $(0,L]$. Its principal random-walk eigenfunction can therefore be chosen positive almost everywhere, denoted by ϕ, with

$$\mu_{\mathrm{odd}} d_r(s)\phi(s) = \int_0^L O(s,t)\phi(t)\,dt, \qquad 0 < \mu_{\mathrm{odd}} = 1 - \lambda_{\mathrm{odd}} \le 1.$$

The upper bound on μ_{odd} follows because this sector is a restriction of the full self-adjoint Markov operator. Put

$$I_0 = \int_0^{r/2} \phi(s)\,ds, \qquad I_1 = \int_{r/2}^{r} \phi(s)\,ds.$$

For $s \in [r/4, r/2]$ and $t \in [r/2, r]$, we have $|s-t| \le 3r/4$ and $st \ge r^2/8$. If $s+t \le r$, the ratio of the second Gaussian in O to the first is $e^{-16st/r^2} \le e^{-2}$; if $s+t > r$, the second term vanishes. In either case $O(s,t) \ge c_0$. Integrating the eigenfunction equation over $[r/4, r/2]$, and using $d_r \le (1+a)m_0 r$, gives

$$\frac{rc_0}{4} I_1 \le \mu_{\mathrm{odd}} \int_{r/4}^{r/2} d_r\phi \le (1+a)m_0 r I_0.$$

Consequently $I_1 \le 4(1+a)m_0 I_0/c_0$.

Define

$$A_\phi = \int_0^L \int_0^L K_r(s+t)\phi(s)\phi(t)\,ds\,dt, \qquad B_\phi = \int_0^L \int_0^L B(s,t)\phi(s)\phi(t)\,ds\,dt, \qquad D_\phi = \int_0^L d_r(s)\phi(s)^2\,ds.$$

On $[0, r/2]^2$, $K_r(s+t) \ge e^{-4}$, so $A_\phi \ge e^{-4} I_0^2$. Since $0 \le B \le 1$ and its support is contained in $[0,r]^2$, $B_\phi \le (I_0 + I_1)^2$. Therefore $A_\phi \ge \kappa(a) B_\phi$. Use the same ϕ as a manifold-sector trial function. Its quotient and the odd-sector eigenvalue differ exactly by

$$\mathcal{R}_{\mathrm{man}}(\phi) - \lambda_{\mathrm{odd}} = \frac{2(aB_\phi - A_\phi)}{D_\phi}.$$

As $B_\phi > 0$, the variational principle yields

$$\lambda_{\mathrm{odd}} - \lambda_{\mathrm{man}} \ge \frac{2(\kappa(a) - a)B_\phi}{D_\phi} > 0. \tag{67}$$

□

Remark B.61 (Strict ordering versus a quantitative gap rate). Equation (67) is a positive lower bound involving the actual odd eigenfunction. It does not yet identify a power of r/L or prove the interface surrogate's ϱ^3/a gap asymptotic. Its hypothesis is comfortably satisfied by the fixed bandwidth, but the gap itself need not stay uniformly positive as $r \to 0$.

B.18.5 Direct exclusion of the common sector

Let $\mathcal{E}_A(f) = \frac{1}{2}\int_0^L \int_0^L A(s,t)(f(s) - f(t))^2\,ds\,dt$; its degree is $p_r(s)$. Introduce an analysis-only reflected kernel

$$R(s,t) = A(s,t) + K_r(2L - s - t), \qquad s,t \in [0,L].$$

It has constant degree $m_0 r$. Since $2L - s - t \ge |s-t|$, the added term is at most $K_r(|s-t|) \le A(s,t)$. Consequently $\mathcal{E}_A \ge \mathcal{E}_R/2$.

Lemma B.62 (A reflected-kernel gap bound)**.** *For $0 < \varrho < 1/4$, the reflected normalized gap satisfies*

$$g_R \ge c_R \varrho^2, \qquad c_R = D_{\mathrm{diff}}\pi^2(1 - \pi^2/192) \approx 0.56084169.$$

Proof. Even periodic extension on an interval of period $2L$ diagonalizes R in the cosine basis $\cos(k\pi s/L)$. For $k \ge 1$ the eigenvalues of its normalized Laplacian are

$$\ell_k = \frac{1}{m_0}\int_{-1}^{1} e^{-4u^2}(1 - \cos(k\pi\varrho u))\,du.$$

For $k = 1$, use $1 - \cos x \geq x^2/2 - x^4/24$ and $\int u^4 e^{-4u^2}\, du \leq m_2$ to get $\ell_1 \geq D_{\text{diff}}\pi^2\varrho^2(1 - \pi^2\varrho^2/12) \geq c_R\varrho^2$. For $k \geq 2$ with $\omega = k\pi\varrho \leq \pi$, use $1 - \cos x \geq 2x^2/\pi^2$ on $[-\pi, \pi]$, obtaining $\ell_k \geq 4D_{\text{diff}}k^2\varrho^2 \geq 16D_{\text{diff}}\varrho^2 > c_R\varrho^2$. For $\omega > \pi$, integration by parts gives

$$\left| \int_{-1}^{1} e^{-4u^2} \cos(\omega u)\, du \right| \leq 2/\omega,$$

so $\ell_k \geq 1 - 2/(m_0\pi)$. To verify that this is also enough, $m_0 \geq 2/3$ follows by integrating $1 - 4u^2$ on $[-1/2, 1/2]$, and $m_2 = m_0/8 - e^{-4}/4$ implies $D_{\text{diff}} < 1/16$. Thus $1 - 2/(m_0\pi) \geq 1 - 3/\pi > \pi^2/256 > c_R/16 \geq c_R\varrho^2$. The cosine basis is complete, proving the gap bound. □

Proposition B.63 (Uniform exclusion of the common sector)**.** *At $a = e^{-25}$ and every $0 < \varrho < 1/4$,*

$$\lambda_{\text{common},2} \geq \frac{c_R}{2(1+a)}\varrho^2 > 0.280420\varrho^2, \tag{68}$$

$$\lambda_{\text{man}} \leq C_{\text{man}}\varrho^2 < 0.200649\varrho^2, \tag{69}$$

where

$$\eta_* := \frac{1}{3} - \frac{H}{4m_0}, \qquad C_{\text{man}} = \frac{D_{\text{diff}} + a\pi J_3/(4m_0)}{\eta_*}.$$

In particular, $\lambda_{\text{common},2} - \lambda_{\text{man}} > 0.079771\varrho^2$.

Proof. Since $p_r \leq m_0 r$, the variance comparison and $\mathcal{E}_A \geq \mathcal{E}_R/2$ give the even within-sector gap $g_{\text{even,in}} \geq g_R/2$. Common-sector cross energy is nonnegative. Moreover $d_r \leq (1+a)p_r$, so the weighted-variance argument used in Lemma B.37 gives $\lambda_{\text{common},2} \geq g_{\text{even,in}}/(1+a)$. This proves (68).

For the upper bound use the degree-centered manifold trial function $h_1(s) = |s|$, $h_2(t) = -|t|$. The total within energy is at most $2Lm_2r^3$, because $||s| - |t|| \leq |s - t|$. Its cross energy is at most

$$a \int_{s^2+t^2 \leq r^2} e^{-4(s^2+t^2)/r^2} 2(s^2 + t^2)\, ds\, dt = 4\pi a J_3 r^4.$$

On $[0, L]$ the integrated deficit of p_r from $m_0 r$ is Hr^2, all at the outer endpoint. Hence

$$\int_0^L p_r(s)s^2\, ds \geq \frac{m_0 r L^3}{3} - Hr^2L^2 \geq m_0 r L^3 \eta_*.$$

The full denominator is at least $4m_0 r L^3 \eta_*$. Division and $\varrho < 1/4$ prove (69). The displayed numerical inequalities evaluate the explicitly given constants and have been rounded conservatively. □

Theorem B.64 (Population Fiedler signs throughout the small-radius regime)**.** *For the equal-length, equal-density orthogonal crossing segments with exact tangents, population A1, and fixed $\sigma_T = 0.2$, the second population eigenvalue is simple and lies in the manifold sector for* every $0 < r/L < 1/4$. *Its degree-rescaled eigenfunction can be chosen as*

$$h_1(s) = f_r(|s|), \qquad h_2(t) = -f_r(|t|), \qquad f_r > 0 \quad \textit{almost everywhere}.$$

Its signs recover the generating manifold labels almost everywhere.

Proof. Propositions B.60 and B.63 strictly exclude both competing sectors in (61). The positivity and simplicity argument for the manifold kernel in Theorem B.57 then applies without its additional radius lower bound. □

Remark B.65 (Scope of the completed sign analysis). This theorem concerns the original nonlocal Gaussian population graph; no interface approximation is used in its proof. It strengthens the sign conclusion of Theorem B.57, while that earlier theorem retains its useful explicit gap bound in the stated radius regime. A positive gap for each fixed radius is not a uniform gap rate for shrinking radii. The common-sector separation is quantified, but a sharp odd-sector gap and the next manifold-sector eigenvalue still matter for finite-sample eigenspace perturbation.

Here $\tau/g_{\text{in}} \asymp a/\varrho$ diverges as $\varrho \to 0$, yet population Fiedler signs remain correct. Thus the failure of the aggregate sufficient condition is conservative for sign recovery in this specific model. The sign theorem alone does not prove that row-normalized approximate two-means has the same labels. Section B.19 adds a certified global population two-means result in a specified radius range. Section B.20 gives a separate fixed-radius oracle sampling result; concrete estimator guarantees remain unproved. Small-angle counterexamples in Sections B.16 and B.17 remain valid.

B.19 Population two-means: geometry, optimality, and a certified radius range

This section concerns the row-normalized embedding of the original orthogonal population graph. Its results are separate from the interface surrogate below. Population two-means uses the balanced uniform sampling measure, not a degree-weighted measure. Degree weights enter the normalization of eigenfunctions only.

B.19.1 Exact row geometry and the true Lloyd fixed point

Let $Z_d = \int_0^L d_r(s)\,ds$, and let π_d be the probability measure on $[0, L]$ with density d_r/Z_d. Normalize the positive manifold-sector profile from Theorem B.64 so that

$$\mathbb{E}_{\pi_d} f_r^2 = 1.$$

The constant and contrast eigenfunctions in symmetric-normalized coordinates have the common degree factor. It cancels under row normalization, giving exactly

$$V_{\pm}(s) = \frac{(1, \pm f_r(s))}{\sqrt{1 + f_r(s)^2}}, \qquad 0 \le s \le L. \tag{70}$$

Here the sign is the generating manifold label; each sign has mass $1/2$, and conditional on it s is uniform on $[0, L]$. Write $X(s) = (1 + f_r(s)^2)^{-1/2}$, $Y(s) = f_r(s)(1 + f_r(s)^2)^{-1/2}$, and let bars denote uniform averages on this interval.

Proposition B.66 (True partition as a population Lloyd fixed point)**.** *The true centroids are $c_{\pm} = (\bar{X}, \pm\bar{Y})$, with $\bar{Y} > 0$. Every row is strictly nearer to its own true centroid almost everywhere. The true-partition objective, with optimal centroids, is*

$$J_{\text{true}} = 1 - \bar{X}^2 - \bar{Y}^2. \tag{71}$$

Proof. Symmetry gives the centroids. For an upper row the difference between its squared distance to the lower and upper centroids is $4\bar{Y}Y(s) > 0$; the lower case is symmetric. Expanding the within-cluster variance and using $X^2 + Y^2 = 1$ gives the objective formula. □

A Lloyd fixed point need not be globally optimal, and this proposition makes no claim about convergence from arbitrary initial centers.

For any competing measurable cluster B, put $p = \mathbb{P}(B) \in (0, 1)$, $m_B = \mathbb{E}[V\mathbf{1}_B]$, and $m = \mathbb{E}V = (\bar{X}, 0)$. Optimizing its two centroids gives

$$J(B) = 1 - \|m\|^2 - \frac{\|m_B - pm\|^2}{p(1-p)}.$$

Consequently the exact criterion for global optimality of the true partition is

$$\frac{\|m_B - pm\|^2}{p(1-p)} \le \bar{Y}^2 \quad \text{for every such } B. \tag{72}$$

This criterion does not follow from sign separation or symmetry alone.

Remark B.67 (Limitation of an objective-only approximation contract). Suppose the true partition is globally optimal and $J_{\text{true}} > 0$. For any $\alpha > 1$, retain its centroids but assign an upper-manifold set of sufficiently small positive probability $p < 1/2$ to the lower center. The objective increases by at most $4p$, because its pointwise increase is $4\bar{Y}Y \le 4$. Taking $4p \le (\alpha - 1)J_{\text{true}}$ gives an α-approximate solution with positive label error. Thus the objective-only contract cannot guarantee exact recovery at $\alpha > 1$ without additional information. These altered assignments need not be

nearest-center assignments; this is a limitation of the stated contract, not a demonstrated failure of a Lloyd implementation. Nor does it rule out vanishing error along a sequence where the available objective slack vanishes. In the present graph the continuous profile is nonconstant: a constant positive profile would force $\lambda = 0$ away from the intersection and a positive value near it in its eigenfunction equation. Hence its row curve is nonconstant and $J_{\mathrm{true}} > 0$.

B.19.2 A GEOMETRIC SUFFICIENT CONDITION AND EXCESS-OBJECTIVE CONTROL

The following elementary result applies to a nonatomic population or a finite balanced sample. Centroid distances below concern all nonempty subsets of the respective true clusters.

Theorem B.68 (Balanced two-means under separated centroid sets)**.** *Suppose the two true clusters each have probability $1/2$. Their within supports have diameter at most D_{in}, and every pair of centroids from opposite true clusters has distance at least B_{out}. Define $\Delta_{\mathrm{round}} = B_{\mathrm{out}}^2 - D_{\mathrm{in}}^2 > 0$. For any competing partition $\mathcal{P}$, with its own optimal centroids, and permutation-invariant error $e \in [0, 1/2]$,*

$$J(\mathcal{P}) - J_{\mathrm{true}} \geq \Delta_{\mathrm{round}} \frac{e}{1+2e} \geq \frac{\Delta_{\mathrm{round}}}{2} e. \tag{73}$$

Thus the true partition is the unique global optimum, up to label permutation and null sets. For centers and labels satisfying the α-approximate objective contract,

$$e \leq \min\left\{\frac{1}{2}, \frac{2(\alpha-1)J_{\mathrm{true}}}{\Delta_{\mathrm{round}}}\right\}. \tag{74}$$

No balance is imposed on the competing clusters.

Proof. Let u, v be the fractions of the first and second true clusters assigned to competing cluster 1. Splitting each true cluster into its two cells can reduce objective by at most $D_{\mathrm{in}}^2 S$, where

$$S = \tfrac{1}{2}[u(1-u) + v(1-v)].$$

Merging cells from opposite true clusters into the competing clusters increases objective by at least $B_{\mathrm{out}}^2 M$, where

$$M = \tfrac{1}{2}\left[\frac{uv}{u+v} + \frac{(1-u)(1-v)}{2-u-v}\right].$$

Terms with zero cell mass are taken as zero. These formulas are the standard variance decomposition for merging two weighted groups; the within-centroid distance is at most the support diameter. When both competing clusters are nonempty, put $z = (u+v)(2-u-v)$ and $b = |u-v|$. Direct algebra yields

$$M = S/z = \tfrac{1}{4}(1 - b^2/z), \qquad 0 < z \leq 1, \qquad b(2-b) \leq z.$$

Thus $M \geq S$, and with $e = (1-b)/2$ one obtains $M \geq e/(1+2e)$ (the case $b = 0$ gives $M = 1/4$ directly). Therefore $J(\mathcal{P}) - J_{\mathrm{true}} \geq \Delta_{\mathrm{round}} M$, proving the first inequality. For a single nonempty competing cluster, $M = 1/4, S = 0, e = 1/2$, and the same conclusion holds. Strict positivity proves uniqueness. Finally, reoptimizing centroids for returned labels cannot increase objective. Since $J_{\mathrm{opt}} = J_{\mathrm{true}}$, their excess is at most $(\alpha-1)J_{\mathrm{true}}$; apply the second inequality. □

Corollary B.69 (An angular sufficient condition for the actual row curve)**.** *Let $\vartheta_r(s) = \arctan f_r(s)$ have range $[\vartheta_{\min}, \vartheta_{\max}] \subset (0, \pi/2)$. One can use*

$$D_{\mathrm{in}} = 2\sin((\vartheta_{\max} - \vartheta_{\min})/2), \qquad B_{\mathrm{out}} = 2\sin\vartheta_{\min}.$$

If $\vartheta_{\max} < 3\vartheta_{\min}$, the conclusions of Theorem B.68 hold. A simpler sufficient condition is

$$\frac{\max f_r}{\min f_r} < \sqrt{3}. \tag{75}$$

Proof. The within arc's largest chord has the displayed length. Any centroid of an upper subset has vertical coordinate at least $\sin\vartheta_{\min}$, and any lower centroid has coordinate at most its negative, proving the cross-centroid bound. Sine is increasing on the relevant interval, so $B_{\mathrm{out}} > D_{\mathrm{in}}$ is equivalent to $\vartheta_{\max} < 3\vartheta_{\min}$. Under any initial profile scaling, normalization divides it by its π_d root-mean-square, which lies between its minimum and maximum. Condition (75) therefore implies the normalized minimum is greater than $1/\sqrt{3}$. Hence $\vartheta_{\min} > \pi/6$ while $\vartheta_{\max} < \pi/2$, as required. □

B.19.3 POINTWISE CONTROL OF THE ORIGINAL EIGENFUNCTION

The remaining step is to verify a range condition from the actual kernel, not to impose it as an unexplained profile assumption.

Proposition B.70 (From weighted variance to a uniform profile bound)**.** *In the orthogonal model, set* $c_r = \mathbb{E}_{\pi_d} f_r > 0$ *and*

$$\varepsilon_r^2 = \frac{(1+a)\tau}{g_{\mathrm{in}}}.$$

If $\varepsilon_r^2 < 1$ *and* $\tau < 1$*, then*

$$\left\| \frac{f_r}{c_r} - 1 \right\|_\infty \le \zeta_r := \frac{\sqrt{\frac{8(1+a)}{m_0 \varrho}} \frac{\varepsilon_r}{\sqrt{1-\varepsilon_r^2}} + \tau + 2a}{1-\tau}. \tag{76}$$

The norm may equivalently be read as essential supremum; the positive profile has a continuous representative.

Proof. Within coercivity applied to the manifold eigenfunction and $\lambda_{\mathrm{man}} \le \tau$ give $\mathrm{Var}_{\pi_d}(f_r) \le \varepsilon_r^2$. Since $\mathbb{E}_{\pi_d} f_r^2 = 1$, this implies $c_r \ge \sqrt{1-\varepsilon_r^2}$. Let $T = T_{\mathrm{man}}$ have kernel $G(s,t)/d_r(s)$, where $G = A - 2aB \ge 0$. Its row sum is $T1 = 1 - 2t_r^{\mathrm{out}}(s)/d_r(s)$, with $t_r^{\mathrm{out}} = a c_r^{\mathrm{cross}}$ denoting the cross degree; here $c_r^{\mathrm{cross}}(s) = \int_{-L}^{L} K_r(\sqrt{s^2+t^2})\, dt$ is distinct from the scalar mean c_r. Thus $t_r^{\mathrm{out}}/d_r \le a$ and

$$(1-\lambda_{\mathrm{man}})(f_r - c_r) = T(f_r - c_r) + c_r(\lambda_{\mathrm{man}} - 2t_r^{\mathrm{out}}/d_r).$$

To control the first term, its squared row norm as a functional on $L^2(\pi_d)$ is

$$\frac{Z_d}{d_r(s)^2} \int_0^L \frac{G(s,t)^2}{d_r(t)}\, dt.$$

Since $0 \le G \le A \le 2$, $G^2 \le 2A$, and $\int_0^L A(s,t)\, dt = p_r(s) \le d_r(s)$. Also $d_{\min} \ge m_0 r/2$ and $Z_d \le (1+a) m_0 r L$. The squared norm is consequently at most $2Z_d/d_{\min}^2 \le 8(1+a)/(m_0\varrho)$. Cauchy–Schwarz bounds $\|T(f_r - c_r)\|_\infty$ by this square root times ε_r. Divide the eigenfunction identity by c_r, and use $\lambda_{\mathrm{man}} \le \tau$, to obtain the result. Continuity follows from the eigenfunction equation: kernel rows vary continuously in L^2 with s, including across the moving support boundary, and d_r is continuous and strictly positive. □

Theorem B.71 (Certified population two-means recovery for the original graph)**.** *In the equal-length, equal-density orthogonal oracle population model, if the bound in Proposition B.70 satisfies*

$$\zeta_r < 2 - \sqrt{3}, \tag{77}$$

then the true manifold partition is the unique population two-means optimum for the actual row-normalized bottom-two embedding, up to label permutation and null sets. Approximate solutions satisfy (74) *with the angular separation constant from Corollary B.69. For fixed* $\sigma_T = 0.2$*, a sufficient explicit range is*

$$5.3 \cdot 10^{-4} \le r/L < 1/4. \tag{78}$$

Proof. The uniform bound gives $\max f_r / \min f_r \le (1+\zeta_r)/(1-\zeta_r)$. Condition (77) makes this smaller than $\sqrt{3}$, so Corollary B.69 applies. For the explicit range, write

$$\bar{v}(\varrho) = \frac{(1+a)c_\tau a}{c_g \varrho}, \qquad \bar{\tau}(\varrho) = c_\tau a \varrho.$$

Equations (55) and (58) allow $\varepsilon_r^2 \le \bar{v}$ and $\tau \le \bar{\tau}$ in (76). Put $\varrho_0 = 5.3 \cdot 10^{-4}$ and $v_0 = \bar{v}(\varrho_0)$. Over $[\varrho_0, 1/4)$ a uniform upper bound is

$$Z_* = \frac{\sqrt{\frac{8(1+a)}{m_0 \varrho_0}} \sqrt{\frac{v_0}{1-v_0}} + c_\tau a/4 + 2a}{1 - c_\tau a/4}.$$

Using the explicitly defined constants gives $Z_* < 0.26142 < 2 - \sqrt{3}$; also $v_0 < 4 \cdot 10^{-6}$, so all denominators are positive. This verifies the whole stated range, not just a single numerical radius. □

Remark B.72 (Interpretation and remaining rounding questions). The numerical lower endpoint in (78) is a conservative certificate, not a success/failure transition or a recommended algorithm parameter. At $r/L = 10^{-3}$ the displayed bounds give $\zeta_r < 0.13856$ and $\max f_r / \min f_r < 1.322$. Exact global two-means ($\alpha = 1$) has zero population error in the certified range. Fixed $\alpha > 1$ yields a quantitative error bound, not automatic exact recovery or vanishing error along arbitrary sequences. Finite Lloyd restarts are not proved to meet a given α.

Fiedler sign recovery holds throughout $0 < r/L < 1/4$, but two-means global optimality at smaller radii remains unresolved by this range argument. The sine-shaped interface-surrogate profile does not meet its uniform separation condition; that is a limitation of this sufficient condition, not a counterexample to global two-means optimality. These are population statements. Section B.20 subsequently addresses fixed-radius oracle sampling. Numerical eigensolver errors and concrete tangent-estimator accuracy remain separate.

B.20 Rounding stability and fixed-radius oracle sampling consistency

This section connects the orthogonal population analysis to balanced finite samples. Tangents are supplied exactly. The radius is fixed, the spatial bandwidth is empirical A1, and self-loops are removed before normalization. No guarantee for a concrete tangent estimator or a finite-restart Lloyd solver is assumed implicitly.

B.20.1 Finite-sample rounding under embedding error

For rows $X_i \in \mathbb{R}^2$ and a two-cluster label partition P, let

$$J_X(P) = \min_{c_1, c_2} \frac{1}{n} \sum_i \|X_i - c_{P_i}\|^2.$$

Empty clusters are allowed, with their zero contributions omitted.

Proposition B.73 (Rounding stability relative to separated ideal rows)**.** *Let ideal rows Z_i have two equally sized true clusters. Suppose the support-diameter and opposite-subset-centroid conditions of Theorem B.68 hold, with $\Delta_{\mathrm{round}} > 0$. Write P_0 for the true partition and $J_{0,n} = J_Z(P_0)$. For empirical rows $\widehat{V}_i$, put*

$$E_n = \min_{O^\top O = I_2} \frac{1}{n} \sum_i \|\widehat{V}_i O - Z_i\|^2.$$

If returned centers and labels on $\widehat{V}$ satisfy an α-approximate two-means objective contract, then

$$\mathrm{err}(\widehat{y}, y) \le \min\left\{ \frac{1}{2}, \frac{2}{\Delta_{\mathrm{round}}} \big[(\alpha - 1) J_{0,n} + 2\sqrt{\alpha}(1 + \sqrt{\alpha}) \sqrt{J_{0,n} E_n} + (1 + \sqrt{\alpha})^2 E_n \big] \right\}. \tag{79}$$

For $\alpha = 1$, this reduces to

$$\mathrm{err}(\widehat{y}, y) \le \min\left\{ \frac{1}{2}, \frac{8}{\Delta_{\mathrm{round}}} (\sqrt{J_{0,n} E_n} + E_n) \right\}.$$

Proof. Choose an optimal alignment and set $X = \widehat{V} O$. For each fixed partition P, the square root of its optimized objective is the Euclidean distance, divided by $\sqrt{n}$, to the linear space of rows constant on the cells of P. Distances to a fixed set are 1-Lipschitz, so $|\sqrt{J_X(P)} - \sqrt{J_Z(P)}| \le \sqrt{E_n}$. Rotation leaves the approximation contract unchanged. Reoptimizing the returned centers can only decrease objective, giving

$$\sqrt{J_Z(\widehat{P})} \le \sqrt{J_X(\widehat{P})} + \sqrt{E_n} \le \sqrt{\alpha} \sqrt{J_X(P_0)} + \sqrt{E_n} \le \sqrt{\alpha J_{0,n}} + (1 + \sqrt{\alpha}) \sqrt{E_n}.$$

Square and subtract $J_{0,n}$. Theorem B.68 bounds this excess below by $\Delta_{\mathrm{round}} \, \mathrm{err}(\widehat{y}, y)/2$, proving the claim. □

For orthogonal population rows $Z_i = (1, \ell_i f_r(|s_i|))/\sqrt{1 + f_r(|s_i|)^2}$, where $\ell_i \in \{+1, -1\}$ is the generating manifold sign, Corollary B.69 supplies the same positive Δ_{round} for every balanced sample whenever its angular condition holds. Subset centroids are finite averages within the same arcs. Also $0 \le J_{0,n} \le 1$, since the ideal rows have unit norm. This observation does not require the empirical conditional distribution of $|s_i|$ to be exactly uniform.

B.20.2 A COMMON PROBABILITY SPACE FOR OPERATORS

Fix $L > 0$ and $0 < r/L < 1/4$. On each of the two orthogonal segments, take an infinite independent sequence of uniform parameters, and use its first N entries at sample size $n = 2N$. The two sequences are independent. There is no observation noise. Every sample receives its exact generating tangent, so the cross tangent factor remains $a = e^{-25}$. Let

$$\Omega = \{1, 2\} \times [-L, L], \qquad d\mu(k, s) = \frac{ds}{4L}, \qquad x(k, s) = se_k.$$

The two parameter copies of the origin are distinct but have zero mass. Let $\mathcal{K}$ be the population affinity kernel with bandwidth $r/2$:

$$\mathcal{K}(\omega, \omega') = a^{\mathbf{1}_{\{k \neq k'\}}} e^{-4\|x(\omega) - x(\omega')\|^2/r^2} \mathbf{1}_{\{\|x(\omega) - x(\omega')\| \le r\}}.$$

Define the probability-normalized degree $\mathfrak{d}(\omega) = \int \mathcal{K}(\omega, \omega')\, d\mu(\omega')$. Thus $\mathfrak{d}(k, s) = d_r(s)/(4L)$, using the arclength degree from Section B.18. In particular,

$$\mathfrak{d} \ge d_\circ := \frac{m_0 r}{8L} > 0. \tag{80}$$

Partition each parameter interval into N equal-length cells. Send its jth cell to the jth order statistic of that segment's sample, retaining the segment index. This defines a measurable map T_N; endpoint conventions affect only null sets. Each cell has μ-mass $1/n$. The pushforward of μ is the empirical measure with equal point weights. Put

$$h_N = \operatorname*{ess\,sup}_{\omega} \|x(T_N\omega) - x(\omega)\|, \qquad b_N = |\widehat{\sigma}_{s,N} - r/2|.$$

Almost surely, $h_N \to 0$ and $b_N \to 0$. To justify the first claim, the empirical distribution functions converge uniformly: the strong law on a countable collection of finite uniform grids, followed by monotonicity between grid points, proves this directly. Since the uniform population distribution has slope $1/(2L)$, inversion and the cell width bound give $h_N \to 0$. The second claim is Proposition B.47; its radius condition holds because $2r < L/2$. Distinctness and avoidance of the origin hold almost surely simultaneously for all sample sizes.

Let $\mathcal{K}_N$ be the step kernel equal to the actual loop-free empirical weight W_{ij} on cell pair (i, j), using empirical A1. On cell pairs with $i = j$, it is zero. Write $\mathfrak{d}_N = \int \mathcal{K}_N\, d\mu$.

Lemma B.74 (A uniform shell bound for orthogonal segments)**.** *For $0 < \delta < r$,*

$$\sup_{\omega \in \Omega} \mu\{\omega' : |\|x(\omega) - x(\omega')\| - r| \le \delta\} \le \frac{\delta + \sqrt{r\delta}}{L}.$$

Proof. On the center's own segment, the shell is contained in two intervals of total length at most 4δ. On the other segment it satisfies $(r - \delta)^2 \le s^2 + t^2 \le (r + \delta)^2$. On the positive half-line its endpoints $0 \le u \le v$ obey $v^2 - u^2 \le 4r\delta$, including the case where the lower squared endpoint is truncated at zero. Thus $v - u \le 2\sqrt{r\delta}$. Including negative parameters gives length at most $4\sqrt{r\delta}$. Truncation to the compact segments cannot increase these lengths. Multiply their sum by $1/(4L)$. □

Proposition B.75 (Empirical kernel and normalized-operator bounds)**.** *Suppose $h_N < r/2$ and $b_N \le r/4$, and define*

$$\mathcal{A}_N = 8\sqrt{2/e}\frac{h_N}{r} + \frac{8}{e}\frac{b_N}{r} + \frac{2h_N + \sqrt{2rh_N}}{L} + \frac{1}{2N}. \tag{81}$$

Then, with essential supremum for step-kernel rows,

$$\sup_{\omega} \int |\mathcal{K}_N - \mathcal{K}|(\omega, \omega')\, d\mu(\omega') \le \mathcal{A}_N,$$
$$\|\mathcal{K}_N - \mathcal{K}\|^2_{L^2(\mu \times \mu)} \le \mathcal{A}_N, \qquad \|\mathfrak{d}_N - \mathfrak{d}\|_\infty \le \mathcal{A}_N.$$

If $\mathcal{A}_N < d_\circ/2$, both normalized integral operators are well-defined and satisfy

$$\|\mathcal{S}_N - \mathcal{S}\|_{\mathrm{op}} \le \|\mathcal{S}_N - \mathcal{S}\|_{\mathrm{HS}} \le \mathcal{B}_N := \frac{2\sqrt{\mathcal{A}_N}}{d_\circ} + \frac{4\mathcal{A}_N}{d_\circ^2}, \tag{82}$$

where their kernels are $\mathcal{K}_N/\sqrt{\mathfrak{d}_N \mathfrak{d}'_N}$ and $\mathcal{K}/\sqrt{\mathfrak{d}\mathfrak{d}'}$, respectively.

Proof. For $G(u,\sigma) = e^{-u^2/\sigma^2}$ and $\sigma \ge r/4$,

$$|\partial_u G| \le 4\sqrt{2/e}/r, \qquad |\partial_\sigma G| \le 8/(er).$$

Transport changes pair distance by at most $2h_N$, while bandwidth changes by b_N. The exact tangent factor is unchanged and at most one. A support indicator can change only in the shell of width $2h_N$. Lemma B.74 bounds its row mass; the case $h_N = 0$ follows directly or by taking a limit. Suppressing the cell containing the same sample changes a row integral by at most $1/n$. These observations give (81). Since both kernels lie in $[0,1]$, their squared difference is at most their absolute difference. Integration yields the L^2 bound, and integration of the signed difference yields the degree bound.

Now $\mathfrak{d}_N, \mathfrak{d} \ge d_\circ/2$. Split the normalized kernel difference into a numerator perturbation and a denominator perturbation. The former has L^2 norm at most $2\sqrt{\mathcal{A}_N}/d_\circ$. On $[d_\circ/2, \infty)^2$, each partial derivative of $(x,y) \mapsto (xy)^{-1/2}$ has absolute value at most $2/d_\circ^2$. The denominator perturbation is therefore pointwise at most $4\mathcal{A}_N/d_\circ^2$, since $\mathcal{K} \le 1$. The probability measure has total mass one, giving the Hilbert–Schmidt bound. Operator norm is at most Hilbert–Schmidt norm. □

Lemma B.76 (Exact correspondence with the empirical matrix)**.** *On the subspace of cellwise constant functions, $\mathcal{S}_N$ is unitarily equivalent to the actual empirical normalized adjacency $D_N^{-1/2} W_N D_N^{-1/2}$. It is zero on the orthogonal complement. An orthonormal matrix of empirical eigenvectors U_N corresponds to orthonormal functions with cell values $\sqrt{n}(U_N)_i$.*

Proof. Write $d_i^{\mathrm{mat}} = \sum_j W_{ij}$. On cell i, $\mathfrak{d}_N = d_i^{\mathrm{mat}}/n$. The normalized kernel on (i,j) therefore equals $nW_{ij}/\sqrt{d_i^{\mathrm{mat}} d_j^{\mathrm{mat}}}$. Integrating over a cell multiplies this by $1/n$. The basis $\sqrt{n}\mathbf{1}_{\mathrm{cell}\ i}$ is orthonormal in $L^2(\mu)$, which proves the claimed matrix and scaling. Functions in the orthogonal complement have zero average on every cell and are annihilated. □

B.20.3 Eigenspace and row convergence at a fixed radius

Let $\Phi = (\phi_0, \phi_1)$ be the population bottom-two symmetric-normalized eigenfunctions, orthonormal in $L^2(\mu)$. They may be chosen as

$$\phi_0(\omega) = \sqrt{\mathfrak{d}(\omega)/\bar{d}}, \qquad \phi_1(k,s) = \ell_k \sqrt{\mathfrak{d}(k,s)/\bar{d}}\, f_r(|s|), \quad \bar{d} = \int \mathfrak{d}\, d\mu, \quad \ell_1 = 1,\ \ell_2 = -1.$$

The normalization $\mathbb{E}_{\pi_d} f_r^2 = 1$ gives the second column's unit norm, and segment symmetry makes the columns orthogonal. Their row normalization is exactly the population row curve. Both columns are continuous on each closed parameter interval, and

$$c_\Phi := \inf_\omega \|\Phi(\omega)\| \ge \sqrt{d_\circ/\bar{d}} > 0.$$

Let $\omega_\Phi(h)$ be the supremum of $\|\Phi(k,s) - \Phi(k,t)\|$ over the same segment with $|s-t| \le h$. It tends to zero as $h \to 0$.

Let P project onto the span of these two columns. Write $\beta_2 = 1 - \lambda_2 > 0$ for the second population adjacency eigenvalue, and define

$$\gamma_r = \beta_2 - \max\{0, \sup \mathrm{spec}(\mathcal{S}|_{\mathrm{range}(I-P)})\} > 0. \tag{83}$$

Positivity holds by Theorem B.64, compactness of $\mathcal{S}$, and $\lambda_2 < 1$ (for example, (69) gives this last inequality). The population graph is connected, so its constant random-walk mode is simple: zero Dirichlet energy forces constancy along each segment and agreement through the positive cross kernel. Definition (83) includes the adjacency accumulation point zero, equivalently the Laplacian accumulation point one. No uniform lower bound on γ_r as $r \to 0$ is asserted.

Theorem B.77 (Fixed-radius convergence of the empirical row embedding)**.** *In the specified oracle sampling model, compute empirical A1 and the exact bottom-two eigenvectors of the loop-free symmetric normalized Laplacian, and normalize their rows to obtain $\widehat{V}_i$. Set $Z_i = \Phi(\omega_i)/\|\Phi(\omega_i)\|$ at each sample parameter. For all N satisfying the preceding kernel-bound conditions and $\mathcal{B}_N < \gamma_r/2$,*

$$E_N := \min_{O^\top O = I_2} \frac{1}{n} \sum_i \|\widehat{V}_i O - Z_i\|^2 \le \frac{4}{c_\Phi^2} \left[\frac{2\mathcal{B}_N}{\gamma_r - \mathcal{B}_N} + \omega_\Phi(h_N) \right]^2 . \tag{84}$$

These conditions hold eventually almost surely, and $E_N \to 0$ almost surely for every fixed $0 < r/L < 1/4$.

Proof. All operators act on $L^2(\mu)$. The two largest eigenvalues of $\mathcal{S}_N$ differ from their population counterparts by at most $\mathcal{B}_N$, by the variational characterization. They are positive since $\gamma_r \le \beta_2$. The third variational spectral value is at most the complementary population spectral supremum plus $\mathcal{B}_N$, so the leading two-dimensional subspace is separated. By Lemma B.76, it corresponds to exactly the bottom-two empirical matrix eigenvectors.

For each of its unit eigenfunctions u with eigenvalue $\widehat{\beta}$, put $Q = I - P$. The equation

$$(\widehat{\beta} I - \mathcal{S}) Q u = Q(\mathcal{S}_N - \mathcal{S}) u$$

has spectral separation at least $\gamma_r - \mathcal{B}_N$ on $\mathrm{range}(Q)$. Therefore $\|Qu\| \le \mathcal{B}_N/(\gamma_r - \mathcal{B}_N)$. Summing for two columns and using the principal-angle alignment argument of Proposition B.26 gives an orthogonal O with

$$\|\Phi_N O - \Phi\|_{L^2(\mu;\mathbb{R}^2)} \le \frac{2\mathcal{B}_N}{\gamma_r - \mathcal{B}_N}, \qquad \Phi_N|_{\mathrm{cell}\ i} = \sqrt{n}(U_N)_i .$$

Continuity gives $\|\Phi \circ T_N - \Phi\|_{L^2} \le \omega_\Phi(h_N)$. The triangle inequality bounds $\|\Phi_N O - \Phi \circ T_N\|_{L^2}$ by the bracket in (84). Apply the row-normalization inequality with reference norms at least c_Φ, square, and integrate. Each cell has mass $1/n$, and the factor $\sqrt{n}$ cancels under row normalization. This proves the bound; the zero-row convention, if needed, is the one in Lemma B.27. Finally, $h_N, b_N \to 0$ implies $\mathcal{A}_N, \mathcal{B}_N \to 0$. The radius, $d_\circ$, and γ_r are fixed, so all required smallness conditions hold eventually and the right side tends to zero. □

Corollary B.78 (Fixed-radius oracle misclassification consistency)**.** *In the preceding sampling model, fix $5.3 \cdot 10^{-4} \le r/L < 1/4$. Suppose the centers and labels returned on the empirical row embedding satisfy the objective approximation contract with $\alpha_N \to 1$; exact global two-means is included. Then $\mathrm{err}(\widehat{y}_N, y_N) \to 0$ almost surely.*

Proof. Theorem B.71 and its angular proof supply a fixed $\Delta_{\mathrm{round}} > 0$ for every balanced ideal sample. Apply Proposition B.73, using $J_{0,n} \le 1$, $E_N \to 0$, and $\alpha_N \to 1$. □

Remark B.79 (Scope and remaining sampling questions). The approximation contract in the corollary is an explicit solver hypothesis (eventually along the sample sequence); it is not guaranteed by finite Lloyd restarts. A fixed $\alpha > 1$ leaves a nonvanishing slack term in the available bound and does not itself imply consistency. The result concerns vanishing misclassification fraction, not eventual exact recovery of every sample. It supplies no finite-N probability rate, shrinking-radius sampling theorem, or concrete tangent-estimator analysis. At radii outside the certified rounding range, the embedding convergence theorem still holds, but does not establish two-means label consistency. If r varies, the degree floor, spectral gap, profile continuity, and rounding separation must all be controlled together; the displayed fixed-radius bounds cannot be transferred automatically.

B.21 Abstract recovery with an unresolved intersection neighborhood

This section connects a local geometric margin, graph stability allowing arbitrary projectors on an unresolved set, and fixed-radius consistency for imperfect representations of orthogonal crossing segments. Local model accuracy remains an explicit module hypothesis. No medoid rule, restart event, branch-count selector, or numerical solver property is used. The spectral realization uses exact symmetric-normalized eigenvectors, row normalization, and an objective approximation contract.

B.21.1 Local disambiguation outside a certified width

Proposition B.80 (Transverse margin with an unresolved neighborhood)**.** *Let two d-dimensional branches meet cleanly along $\mathcal{I}$ at p, and suppose their smallest nonzero principal angle is at least $\theta_0 > 0$. Write a noiseless point on the first branch as*

$$q = p + v + e, \qquad v \in T_p\mathcal{M}_1 \cap (T_p\mathcal{I})^\perp, \qquad t = \|v\| \le t_0, \qquad \|e\| \le \kappa t^2/2,$$

and let $x = q + \xi$ with $\|\xi\| \le \nu$. This representation and remainder bound are assumed on the patch under consideration. At the anchor actually selected for x, let an available correct model $\widehat{L}_1$ and a competing model $\widehat{L}_2$ satisfy

$$\operatorname{dist}(p, \widehat{L}_j) \le a, \qquad \|\widehat{P}_j - P_{T_p\mathcal{M}_j}\|_{\mathrm{op}} \le \epsilon, \quad j = 1, 2.$$

The correct model has strictly smaller residual whenever

$$t(\sin\theta_0 - 2\epsilon) - \kappa t^2 > 2(a + \nu). \tag{85}$$

In particular, if $c_0 = \sin\theta_0 - 2\epsilon - \kappa t_0 > 0$, all points on this patch with

$$t > w := \frac{2(a+\nu)}{c_0} \tag{86}$$

are correctly assigned against this competitor. If the same bounds hold against every incorrect competing model, the selected model belongs to the correct branch. Multiple accurate models of the correct branch and the absence of an incorrect competitor are both allowed.

Proof. For any affine plane $\widehat{L}$ with direction projector $\widehat{P}$, the difference between $\operatorname{dist}(p + v, \widehat{L})$ and $\|(I - \widehat{P})v\|$ is at most $\operatorname{dist}(p, \widehat{L})$, by writing the affine residual at p and applying the triangle inequality. The principal-angle assumption gives $\|(I - P_{T_p\mathcal{M}_2})v\| \ge t\sin\theta_0$. Projector perturbation and the 1-Lipschitz property of distance imply

$$\operatorname{dist}(x, \widehat{L}_1) \le a + \epsilon t + \kappa t^2/2 + \nu,$$
$$\operatorname{dist}(x, \widehat{L}_2) \ge t\sin\theta_0 - a - \epsilon t - \kappa t^2/2 - \nu.$$

Subtracting proves (85). Since $\kappa t^2 \le \kappa t_0 t$, condition (86) is sufficient. Compare the available correct model against each incorrect competitor. □

Remark B.81 (Meaning of the unresolved set). The region $t \le w$ is uncertified, not necessarily misclassified. The coordinate t is a transverse tangent coordinate; identifying its sublevel set with a geometric tube requires an appropriate patch representation. For instance, if p is a nearest point on $\mathcal{I}$, then $t \le w$ implies $\operatorname{dist}(q, \mathcal{I}) \le w + \kappa w^2/2$ under the stated remainder bound. No tube-mass claim for general manifolds is implicit here. Coverage, presence of a correct model, and control of all incorrect competitors concern the anchor actually used. Once a correct model is selected, Lemma B.8 supplies $\varepsilon_{\mathrm{fit}} + L_T h_i$ as its sample-level tangent error. Fixed-scale curvature bias is not asserted to vanish with sample size.

B.21.2 Graph stability under arbitrary errors at some vertices

Lemma B.82 (Sparse changed rows and weighted degrees)**.** *Let $W, W^\star$ be symmetric, loop-free, supported on the same radius graph, with weights in $[0, 1]$. Suppose their entries agree unless an endpoint belongs to a set B of k vertices. At most $q_{\max}$ neighbors are supported at any vertex, and degrees of both matrices are at least $d_{\min} > 0$. For $S = D^{-1/2}WD^{-1/2}$ and $S^\star = (D^\star)^{-1/2}W^\star(D^\star)^{-1/2}$,*

$$\|S - S^\star\|_{\mathrm{op}} \le \frac{\sqrt{2kq_{\max}}}{d_{\min}}\left(1 + \frac{q_{\max}}{d_{\min}}\right). \tag{87}$$

The degree difference obeys $\|d - d^\star\|_2 \le q_{\max}\sqrt{2k}$.

Proof. At most $2kq_{\max}$ ordered entries differ, so $\|W - W^\star\|_F \le \sqrt{2kq_{\max}}$. If b_i is the number of potentially changed entries in row i, $\sum_i b_i \le 2kq_{\max}$ and $\sum_i b_i^2 \le 2kq_{\max}^2$. The degree bound follows. Add and subtract $D^{-1/2}W^\star D^{-1/2}$ and use the mean-value bound for $(xy)^{-1/2}$ on $[d_{\min}, \infty)^2$; summing the squared degree contributions over at most $q_{\max}$ entries per row costs at most $\sqrt{q_{\max}}\,\|d - d^\star\|_2/d_{\min}^2$. This gives (87). □

Theorem B.83 (Stability of an imperfect abstract representation)**.** *Let W be the loop-free affinity of the abstract representation and W^* its exact-tangent oracle on the same observations, radius support, and positive spatial scale. Every projector has rank d. Suppose a set B of k vertices satisfies*

$$\max_{i \notin B} \|\widetilde{P}_i - P_i^*\|_F \le \varepsilon;$$

projectors on B are unrestricted apart from being rank-d orthogonal projectors. Form $\overline{W}$ by retaining the projectors on B and replacing all other projectors by their oracle values. Assume at most $q_{\max}$ supported neighbors per vertex and degree at least $d_{\min} > 0$ in each of $W, \overline{W}, W^*$. Put

$$\eta = \exp(2\sqrt{d}\,\varepsilon/\sigma_T^2) - 1.$$

If $\eta < 1$, their symmetric normalized adjacencies satisfy

$$\|S - S^*\|_{\mathrm{op}} \le H := \frac{2\eta}{1-\eta} + \frac{\sqrt{2kq_{\max}}}{d_{\min}} \left(1 + \frac{q_{\max}}{d_{\min}}\right). \tag{88}$$

In particular, if $d_{\min} \ge cn$ for a fixed $c > 0$, then

$$\|S - S^*\|_{\mathrm{op}} \le \frac{2\eta}{1-\eta} + \frac{\sqrt{2}}{c}(1 + c^{-1})\sqrt{k/n}. \tag{89}$$

Proof. Write $\overline{P}_i$ for the hybrid projectors. At every vertex, $\|\widetilde{P}_i - \overline{P}_i\|_F \le \varepsilon$, including the vertices in B where the difference is zero. Since $\|P\|_F = \sqrt{d}$ for a rank-d projector,

$$|\operatorname{tr}(\widetilde{P}_i\widetilde{P}_j) - \operatorname{tr}(\overline{P}_i\overline{P}_j)| \le 2\sqrt{d}\,\varepsilon.$$

The position term is unchanged. Hence the affinity ratio on every supported edge lies between $e^{-2\sqrt{d}\varepsilon/\sigma_T^2}$ and $e^{2\sqrt{d}\varepsilon/\sigma_T^2}$, and Proposition B.12 gives $\|S - \overline{S}\|_{\mathrm{op}} \le 2\eta/(1-\eta)$. Only edges incident to B can differ between $\overline{W}$ and W^*. Lemma B.82 and the triangle inequality prove (88). Substitute $q_{\max} \le n$ and $d_{\min} \ge cn$ to obtain the dense-degree bound. □

This theorem permits arbitrary, data-dependent errors on B without independence between B, the sample, and the estimated projectors. It does not infer graph stability from k/n alone: the degree and support assumptions are essential. Operator stability must still be combined with oracle spectral and rounding separation.

B.21.3 Fixed-radius consistency beyond the exact-tangent oracle

Theorem B.84 (Consistency of imperfect representations on orthogonal segments)**.** *Use the balanced, noiseless orthogonal sampling model and common probability space of Section B.20, with fixed $\sigma_T = 0.2$ and*

$$5.3 \cdot 10^{-4} \le r/L < 1/4.$$

At each $n = 2N$, suppose the representation module returns a rank-one orthogonal projector for every observation. For possibly data-dependent sets B_N, assume

$$\max_{i \notin B_N} \|\widetilde{P}_i - P_i^*\|_F \le \varepsilon_N \longrightarrow 0, \qquad |B_N|/n \longrightarrow 0 \quad \textit{almost surely}. \tag{90}$$

Use empirical A1, exact bottom-two eigenvectors of the loop-free symmetric normalized Laplacian, and row normalization. Suppose returned centers and labels satisfy an α_N-approximate two-means objective contract with $\alpha_N \to 1$ almost surely. Then

$$\operatorname{err}(\widehat{y}_N, y_N) \longrightarrow 0 \quad \textit{almost surely}. \tag{91}$$

If the representation and approximation-ratio assumptions instead hold in probability, the conclusion holds in probability. Arbitrary rank-one projectors are allowed on B_N.

Proof. Let z_N be the maximum empirical CDF error on the two segments; almost surely $z_N \to 0$. The intersection of any radius-r ball centered on a segment with that same segment has parameter length at least r. Its empirical count, excluding the center, is therefore at least

$$N\left(\frac{r}{2L} - 2z_N\right) - 1 \ge \frac{Nr}{4L} = \frac{nr}{8L}$$

eventually, uniformly over centers. Proposition B.47 gives $\widehat{\sigma}_{s,N} \geq r/4$ eventually almost surely. For any pair of rank-one projectors, $0 \leq d_T^2 \leq 1$. Consequently every supported edge in the actual, hybrid, and oracle graphs has weight at least $e^{-16-\sigma_T^{-2}}$, and all three degree minima satisfy

$$d_{\min,N} \geq cn, \qquad c = \frac{r}{8L} e^{-16-\sigma_T^{-2}} > 0. \tag{92}$$

Theorem B.83 now yields a bound $H_N \to 0$ on $\|S_N - S_N^*\|_{\mathrm{op}}$.

Let C_i be the equal-mass cells of Section B.20, ordered according to the same sample permutation for both empirical graphs. The map $J_N : \mathbb{R}^n \to L^2(\mu)$ given by $J_N u = \sqrt{n} \sum_i u_i \mathbf{1}_{C_i}$ is an isometry. The lifted actual and oracle normalized operators are respectively $J_N S_N J_N^\dagger$ and $J_N S_N^* J_N^\dagger$, where $J_N^\dagger$ is the adjoint and S_N^* denotes the oracle matrix. Consequently their difference has norm exactly $\|S_N - S_N^*\|_{\mathrm{op}}$, and both operators vanish on the orthogonal complement of cell-constant functions. Denote the lifted actual operator by $\widetilde{\mathcal{S}}_N$. By Proposition B.75,

$$\|\widetilde{\mathcal{S}}_N - \mathcal{S}\|_{\mathrm{op}} \leq \mathcal{T}_N := \mathcal{B}_N + H_N \longrightarrow 0.$$

The scalar $\mathcal{T}_N$ is distinct from the transport map T_N.

Write P for the population leading-two spectral projection and $Q = I - P$. By definition of γ_r in (83), $\sup \operatorname{spec}(\mathcal{S}|_{\operatorname{range} Q}) \leq \beta_2 - \gamma_r$ and $\beta_2 \geq \gamma_r > 0$. The variational principle gives $\widehat{\beta}_2 \geq \beta_2 - \mathcal{T}_N > 0$ and bounds the third variational value by $\beta_2 - \gamma_r + \mathcal{T}_N$. Thus when $\mathcal{T}_N < \gamma_r/2$, the leading-two subspace is separated, lies in the cell-constant subspace, and corresponds to the bottom-two eigenvectors of the actual matrix Laplacian. The zero eigenvalues on the orthogonal complement cannot intervene.

For each normalized leading eigenfunction u with eigenvalue $\widehat{\beta}$, the identity

$$(\widehat{\beta} I - \mathcal{S}) Q u = Q(\widetilde{\mathcal{S}}_N - \mathcal{S}) u$$

and spectral separation give $\|Qu\| \leq \mathcal{T}_N/(\gamma_r - \mathcal{T}_N)$. Summing over two eigenfunctions and using orthogonal Procrustes alignment yields

$$\min_{O^\top O = I_2} \|\widetilde{\Phi}_N O - \Phi\|_{L^2(\mu;\mathbb{R}^2)} \leq \frac{2\mathcal{T}_N}{\gamma_r - \mathcal{T}_N}, \qquad \widetilde{\Phi}_N|_{C_i} = \sqrt{n}(\widetilde{U}_N)_i.$$

Here $\widetilde{U}_N$ contains the actual matrix eigenvectors. Continuity along each parameter branch implies $\|\Phi - \Phi \circ T_N\|_{L^2} \leq \omega_\Phi(h_N)$. The reference row norms of $\Phi \circ T_N$ are at least c_Φ. The triangle inequality and Lemma B.27 now give

$$E_N := \min_{O^\top O = I_2} \frac{1}{n} \sum_i \|\widehat{V}_i O - Z_i\|^2 \leq \frac{4}{c_\Phi^2} \left[\frac{2\mathcal{T}_N}{\gamma_r - \mathcal{T}_N} + \omega_\Phi(h_N) \right]^2 \longrightarrow 0. \tag{93}$$

Here h_N, Φ, c_Φ, ω_Φ, and Z_i retain their sampling-section meanings, and $\widehat{V}_i$ are the estimated-projector graph's normalized rows. In particular h_N here denotes transport displacement, not graph error.

The radius range guarantees a fixed $\Delta_{\mathrm{round}} > 0$ for the ideal rows by Theorem B.71. Their true-partition objective obeys $J_{0,n} \leq 1$. Proposition B.73 then gives (91). For exact global two-means the explicit bound is

$$\operatorname{err}(\widehat{y}_N, y_N) \leq \min \left\{ \frac{1}{2}, \frac{8}{\Delta_{\mathrm{round}}} (\sqrt{E_N} + E_N) \right\}. \tag{94}$$

For $\alpha_N \to 1$ use the full bound (79). If the representation or rounding limits are in probability, the same deterministic inequalities on the sampling events give convergence in probability; independence is not needed. □

Corollary B.85 (Local-model conditions sufficient for the representation contract)**.** *In the preceding orthogonal model, suppose representation construction completes and every sample using the local-projector path has Frobenius error at most $\varepsilon_{\mathrm{loc},N} \to 0$. At the actually selected anchor of each sample using the branch-assignment path, suppose every returned affine line can be associated with one of the two true lines, with origin offset at most a_N and direction operator-norm error at most $\delta_N < 1/2$. At least one model must represent the sample's own branch. Assume $a_N \to 0$ and*

$\delta_N \to 0$, uniformly over these actually used models. Multiple models of a branch and single-model anchors are allowed; every returned model is subject to these bounds. Put

$$w_N = \frac{2a_N}{1-2\delta_N}, \qquad B_N = \{i : |s_i| \le w_N\}, \qquad \varepsilon_N = \max\{\varepsilon_{\mathrm{loc},N}, \sqrt{2}\delta_N\}. \tag{95}$$

Then all assignment-path samples outside B_N select a model of their own branch, and all samples outside B_N have Frobenius projector error at most ε_N. If z_N is the two-branch CDF error, then eventually

$$\frac{|B_N|}{n} \le \frac{w_N}{L} + 2z_N \longrightarrow 0. \tag{96}$$

Consequently Theorem B.84 applies under its rounding contract. Almost-sure module limits give almost-sure recovery; limits in probability give recovery in probability, and the displayed tube-count bound holds on the events $w_N \le L$.

Proof. Apply Proposition B.80 with $p = 0$, $\theta_0 = \pi/2$, $\kappa = \nu = 0$, and $t = |s_i|$. Every incorrect competitor is strictly worse outside B_N. For rank-one projectors, $\|\widehat{P} - P\|_F = \sqrt{2}\|\widehat{P} - P\|_{\mathrm{op}}$. This proves the error bound, including the separately controlled local-projector path. On each segment the uniform probability of $[-w_N, w_N]$ is w_N/L once $w_N \le L$. Uniform CDF control bounds its empirical mass by $w_N/L + 2z_N$. Averaging the two balanced counts proves (96). The CDF event holds for all intervals, so the bound also applies to data-dependent w_N. All required limits follow, and the theorem supplies recovery. □

Remark B.86 (Scope of the partial guarantee). The new conclusion goes beyond exact-tangent oracle consistency by allowing arbitrary projectors on a vanishing fraction of observations. It remains conditional on representation accuracy and the abstract rounding contract; it does not prove those conditions for a chosen estimator. It is a vanishing-error-fraction result, not eventual exact label recovery. The positive constant in (92) is conservative and can be very small. Fixed radius is essential to this deduction: no uniform degree, spectral, or rounding constants are asserted as $r \to 0$. The curved margin proposition is a separate local statement and does not extend this statistical theorem to curved, noisy, or nonorthogonal models.

C Experimental Details

C.1 Benchmark geometries

We evaluate on seven synthetic union-of-manifolds datasets spanning a range of intersection geometries, dimensionalities, and multiplicities (Figure 2):

- **Two Circles** ($\mathbb{R}^2$, K=2, 2 intersection points): two overlapping planar circles.
- **Triple Lines** ($\mathbb{R}^2$, K=3, 1 intersection point): three lines meeting at a single point, testing triple-intersection multiplicity.
- **Dollar Sign** ($\mathbb{R}^2$, K=2, 3 intersection points): a sinusoidal curve intersecting a vertical line, testing varying tangent angles at multiple intersection points along the same pair of manifolds.
- **Olympic Rings** ($\mathbb{R}^2$, K=5, 4 intersection points): five circles arranged as the Olympic rings, testing large manifold count.
- **Perpendicular Circles** ($\mathbb{R}^3$, K=2, 2 intersection points): two circles in orthogonal planes, requiring 3D tangent decomposition.
- **Two Spheres** ($\mathbb{R}^3$, K=2, intersection along a circle): two spheres intersecting along a circle, testing high-dimensional intersection loci.
- **Swiss Roll + Plane** ($\mathbb{R}^3$, K=2, intersection along a curve): a Swiss roll surface pierced by a flat plane, testing curved manifold geometry.

Revised ITC uses $n = 400M$ observations for M generating manifolds, obtained by random subsampling of the pooled generated data; individual branch counts are not fixed. Perturbations lie in the normal space with norm at most 0.002 (tubular noise). Appendix C specifies generation, radius selection, and the numerical backend.

Table 3: Selected ITC parameters and successful benchmark runs out of ten attempts.

Dataset	k_r	τ_R	σ_T	Successes
Two Circles	15	0.05	0.1	10/10
Perpendicular Circles	20	0.1	0.1	10/10
Two Spheres	20	0.1	0.3	10/10
Olympic Rings	15	0.15	0.1	10/10
Dollar Sign	30	0.05	0.2	10/10
Triple Lines	40	0.35	0.1	10/10
Swiss Roll + Plane	60	0.15	0.3	9/10

Development and evaluation protocols For the five curve datasets, configurations are selected by mean ACC on development data seeds 42–51 and evaluated on data seeds 1042–1051.

For Two Spheres and Swiss Roll + Plane, an expanded search additionally varies the minimum branch size and residual-selection cutoff. Configurations are selected using twenty development data seeds, 42–61, by first minimizing the number of runs below ACC 0.9, counting execution failures as unsuccessful runs, and then maximizing the recorded mean-ACC selection score to break ties. The selected configurations are frozen before evaluation on previously unused data seeds 2042–2051. Earlier results on seeds 1042–1051 informed development and are not treated as final evaluation results for these two datasets.

All selection uses development labels. The final evaluation labels are used only for scoring and post hoc diagnostics. The two protocols have different search spaces and development budgets.

Let $\delta_{i,(k)}$ be the kth ordered observed distance, including the query itself at distance zero. For each new data realization, set

$$r = \max\left\{\operatorname{median}_i \delta_{i,(k_r)},\ (1+10^{-9}) \max_i \delta_{i,(d+2)}\right\}.$$

The second term ensures sufficient local-PCA support, with a numerical boundary allowance. This common radius is used for local PCA, DBSCAN, anchor fitting, A1, and graph support. The median term is active in all nine ablation data realizations. All other synthetic-data fitting parameters use the defaults in Appendix A.

C.2 Data, repetitions, and failure accounting

The generator uses tubular perturbations in the normal space, bounded by 0.002, rather than Gaussian noise. It generates a pooled sample and subsamples $400M$ points without replacement, so branch counts fluctuate. The ablation datasets have 800 (Two Circles), 1200 (Triple Lines), and 800 (Two Spheres) observations. For repetition s, the data seed is $42 + s$ and the fitting and spectral seeds are s. The main ITC benchmark uses $s = 0, \ldots, 9$; the ablation uses $s = 0, 1, 2$. Dispersion is sample standard deviation with divisor one less than the repetition count.

All selected main-benchmark configurations complete ten runs except Swiss Roll + Plane, which has one decomposition failure. The benchmark code assigns such failures the single-cluster score $\max_m n_m/n$ and includes it in the mean, SD, and grid-selection objective. Thus its reported ACC is a failure-adjusted score, not a mean over ten successful clusterings. All 27 ablation runs succeed; no fallback score enters the ablation. The raw results supplied for the three ablation datasets reproduce the reported aggregates and the corresponding benchmark seeds.

The preserved baseline results come from the earlier submission. Their implementation descriptions are retained: scikit-learn for K-means and Spectral (NJW), the `ksubspaces` package for K-Subspaces, the public SSC implementation of Elhamifar & Vidal (2013), public TSC/SCC/Simplex Paths implementations, and an author implementation of LocalPCA-Spectral following Arias-Castro et al. (2017). The current results do not establish identical tuning budgets or repetition protocols for those historical baseline entries.

Table 4: Projector errors for the three paired repetitions, reported as mean and sample SD. All arms succeed on all three repetitions.

Representation	E_P (all)	E_P (int.)
Two Circles		
Local PCA	0.0158 ± 0.0023	0.2014 ± 0.0183
Full ITC	0.0032 ± 0.0020	0.0366 ± 0.0261
True tangent	0.0000 ± 0.0000	0.0000 ± 0.0000
Triple Lines		
Local PCA	0.0423 ± 0.0020	0.3856 ± 0.0240
Full ITC	0.0022 ± 0.0019	0.0193 ± 0.0178
True tangent	0.0000 ± 0.0000	0.0000 ± 0.0000
Two Spheres		
Local PCA	0.1087 ± 0.0043	0.3022 ± 0.0220
Full ITC	0.0922 ± 0.0105	0.2426 ± 0.0251
True tangent	0.0000 ± 0.0000	0.0000 ± 0.0000

C.3 Executed numerical path and environment

The revised benchmark and ablation use the explicit dense spectral option in Appendix A, not the default ARPACK option. Dense numerical eigendecomposition is followed by degree rescaling and finite k-means without row normalization. The recorded environment is Python 3.13.5, NumPy 2.1.3, SciPy 1.15.3, and scikit-learn 1.6.1, with one BLAS thread per worker process. No GPU acceleration is used. The earlier submission recorded an AMD Ryzen 9 7900 CPU and 64 GB RAM; that historical hardware description is not an independently verified hardware specification for the new run logs.

C.4 Representation ablation protocol

Local PCA uses the radius-local projectors already computed by full ITC. Full ITC uses the selected affine branch projector on its candidate path. The diagnostic uses the exact tangent at each clean generating point, while retaining noisy observed positions. Support, A1, tangent bandwidth, and the practical spectral settings and seeds are held fixed across arms. The execution report verifies that the adapter's full-ITC labels match the original path on all nine ablation data realizations.

The intersection region is $B_{\mathrm{eval}} = \{i : \mathrm{dist}(q_i, \mathcal{I}) \le r\}$. Here $\mathcal{I}$ consists of $(0, \pm\sqrt{3}/2)$ for Two Circles, the origin for Triple Lines, and the small circle $\{x = 0,\ y^2 + z^2 = 3/4\}$ for Two Spheres. Regional ACC uses the same Hungarian permutation as overall ACC. The average regional sample counts are 58.7, 131.3, and 268.0, respectively. The radii across the three repetitions are $(0.10864, 0.11012, 0.11069)$, $(0.09916, 0.09840, 0.09783)$, and $(0.40744, 0.40133, 0.40037)$ in that order.

The paired full-ITC minus local-PCA ACC differences have mean and range $+0.235$ $[0.175, 0.344]$ on Two Circles, $+0.339$ $[0.025, 0.503]$ on Triple Lines, and $+0.046$ $[-0.356, 0.341]$ on Two Spheres. These values are computed before rounding the individual table entries. Both global and regional mean projector errors decrease in every paired repetition.

C.5 A localized failure on Two Spheres

In repetition zero, full ITC returns a $795 : 5$ partition. The five points all receive the second model of anchor 133, which selects $K = 2$ with fitting group sizes 18 and 8, and $E_2/E_1 \simeq 0.091 < 0.15$.

Table 5: Mean computation times in seconds. PCA and A1 are shared; ITC extra is detection and decomposition. Graph and spectral entries list local PCA / full ITC / true tangent. Cached costs are not standalone end-to-end timings.

Dataset	PCA	A1	ITC extra
Two Circles	0.065	0.020	0.063
Triple Lines	0.149	0.046	0.272
Two Spheres	0.078	0.023	0.834

Dataset	Graph	Spectral
Two Circles	.0025/.0018/.0015	.075/.066/.067
Triple Lines	.0060/.0054/.0047	.177/.176/.174
Two Spheres	.0030/.0024/.0024	.065/.059/.058

Their squared projector errors are approximately 1.94–2.00. For rank-two planes in $\mathbb{R}^3$, the maximum squared Frobenius projector distance is 2: writing $P = I - nn^\top$ gives $\|P - Q\|_F^2 = 2(1-(n^\top m)^2) \leq 2$. Two assigned points lie at $1.04r$ and $1.05r$ from the anchor, outside its fitting support; assignment has no distance cap. This observation alone does not establish that assignment distance caused the failure.

The five points share one projector, so their mutual tangent factors are one. Their internal undirected weight sum is about 3.818, while their cut to the remaining points is about 0.00100. The graph remains connected. With self-loops excluded, their partition has normalized cut 1.30859×10^{-4}, very close to, but not equal to, $\lambda_2 = 1.30833 \times 10^{-4}$. The generating partition has normalized cut 0.125252. The second embedding coordinate concentrates on the five points. The returned-label k-means objective is approximately 3.95×10^{-7}, compared with 0.6521 for the generating labels in the same embedding. All ten tested k-means seeds return ACC 0.5025; initialization sensitivity was not observed in this test.

Global mean projector error decreases from 0.111 to 0.104, yet the number of samples with squared error above one grows from 21 to 40. Cross-branch edge weight also decreases, from 197.44 to 96.86. Thus neither average tangent error nor total cross-branch weight alone explains the observed partition. The evidence supports a localized, weakly attached group of mutually consistent erroneous projectors as the mechanism in this repetition. It does not explain why the fit converged to this plane or how frequently this mechanism occurs. Repetitions one and two also contain large errors but do not exhibit this five-point split. This finite noisy-surface example does not contradict the abstract fixed-radius orthogonal-segment theorem: its module and graph conditions have not been established for this execution.

D Fitting Certificates and Degree-Rescaled Spectral Recovery

This appendix supplements the abstract representation analysis in two directions. Section D.1 studies a noiseless two-branch local model: it records the population mixing or degeneracy of single PCA, establishes population identifiability of a two-line fit, and converts an approximate empirical fit into representation accuracy. The subsequent section analyzes the degree-rescaled embedding used by the practical spectral backend. Neither analysis certifies finite random-restart fitting or finite Lloyd optimization.

D.1 Population demixing and empirical fitting certificates

This subsection isolates a local mechanism that is specific to noiseless crossing line segments. It serves two purposes. First, it shows that one-component PCA at an intersection generally produces a mixed direction, whereas a two-line population objective identifies the two branches. Second, it converts a sufficiently accurate empirical two-line fit into the local representation conditions used by Theorem 4.3. The results do not establish that a finite-restart fitting routine attains the required empirical objective.

Centered crossing distribution. Let $e_1, e_2 \in \mathbb{R}^2$ be unit vectors with

$$e_1^\top e_2 = \cos\theta, \qquad 0 < \theta \le \frac{\pi}{2},$$

and write $P_j = e_j e_j^\top$. Let $Q_\circ$ be the equally weighted mixture of the uniform distributions on $[-r, r]e_1$ and $[-r, r]e_2$. Its mean is zero, and its covariance is

$$\Sigma_\circ = \frac{r^2}{6}(P_1 + P_2). \tag{97}$$

Corollary D.1 (Population mixing of single PCA)**.** *If $0 < \theta < \pi/2$, the leading population PCA direction of $Q_\circ$ is the angular bisector*

$$u_\circ = \frac{e_1 + e_2}{\|e_1 + e_2\|}.$$

Consequently, with $P_\circ = u_\circ u_\circ^\top$,

$$\|P_\circ - P_1\|_F = \|P_\circ - P_2\|_F = \sqrt{2}\sin(\theta/2) > 0. \tag{98}$$

The two nonzero covariance eigenvalues are

$$\lambda_\pm(\Sigma_\circ) = \frac{r^2}{6}(1 \pm \cos\theta). \tag{99}$$

For $\theta = \pi/2$, these eigenvalues coincide, so the leading eigenspace is $\mathrm{span}\{e_1, e_2\}$ and population rank-one PCA has no uniquely identified branch direction.

Proof. On $\mathrm{span}\{e_1, e_2\}$, the eigenvectors of $P_1 + P_2$ are proportional to $e_1 + e_2$ and $e_1 - e_2$, with eigenvalues $1 + \cos\theta$ and $1 - \cos\theta$, respectively. This proves (99). For $0 < \theta < \pi/2$, the leading eigenvector is therefore the bisector. Its angle from either generating line is $\theta/2$, and the Frobenius distance between rank-one projectors whose directions have angle φ is $\sqrt{2}|\sin\varphi|$. At $\theta = \pi/2$, the two eigenvalues are equal. □

Population and empirical fitting criteria. Let $\mathfrak{L}$ denote the collection of affine lines in $\mathbb{R}^2$. For an anchor $a = se_1$ and fitting radius b satisfying $|s| < b$ and $b + |s| < L$, let $Q_{a,b}$ be normalized arclength measure on

$$B(a, b) \cap (\mathcal{M}_1 \cup \mathcal{M}_2), \qquad \mathcal{M}_j = [-L, L]e_j, \qquad e_1 \perp e_2.$$

Thus the two local support intervals are

$$[s - b, s + b]e_1 \quad \text{and} \quad [-\ell_b, \ell_b]e_2, \qquad \ell_b := \sqrt{b^2 - s^2}.$$

For $\boldsymbol{\mathcal{L}} = (\mathcal{L}_1, \dots, \mathcal{L}_K) \in \mathfrak{L}^K$, define

$$\mathcal{E}_K^{\mathrm{pop}}(a, b; \boldsymbol{\mathcal{L}}) := \int \min_{1 \le k \le K} \mathrm{dist}^2(x, \mathcal{L}_k)\, dQ_{a,b}(x) \tag{100}$$

and

$$E_K^{\mathrm{pop}}(a, b) := \inf_{\boldsymbol{\mathcal{L}} \in \mathfrak{L}^K} \mathcal{E}_K^{\mathrm{pop}}(a, b; \boldsymbol{\mathcal{L}}). \tag{101}$$

For a realized fitting neighborhood $S_a = \{i : \|x_i - a\| \le r\}$ with $m_a := |S_a|$, define

$$\widehat{\mathcal{E}}_{K,a}(\boldsymbol{\mathcal{L}}) := \frac{1}{m_a} \sum_{i \in S_a} \min_{1 \le k \le K} \mathrm{dist}^2(x_i, \mathcal{L}_k) \tag{102}$$

and

$$\widehat{E}_{K,a}^\star := \inf_{\boldsymbol{\mathcal{L}} \in \mathfrak{L}^K} \widehat{\mathcal{E}}_{K,a}(\boldsymbol{\mathcal{L}}). \tag{103}$$

If the fitting routine returns $\widehat{\boldsymbol{\mathcal{L}}}_{K,a}$, write

$$E_K(a) := \widehat{\mathcal{E}}_{K,a}\big(\widehat{\boldsymbol{\mathcal{L}}}_{K,a}\big).$$

An additive empirical optimization contract means

$$E_K(a) \le \widehat{E}_{K,a}^\star + \Delta_{\mathrm{opt},N}. \tag{104}$$

Theorem D.2 (Population one-versus-two-line separation)**.** *In the noiseless orthogonal model above,*

$$E_1^{\mathrm{pop}}(a,b) = \frac{\ell_b^3}{3(b+\ell_b)}, \qquad E_2^{\mathrm{pop}}(a,b) = 0. \tag{105}$$

Moreover, the unique global minimizer of the two-line population criterion, up to permutation, is

$$\{\mathcal{L}_1, \mathcal{L}_2\} = \{\mathbb{R}e_1, \mathbb{R}e_2\}.$$

In particular, fix $b = r$ and suppose all retained anchors satisfy $|s| \le a_{\mathrm{out}} < r$. Put

$$\ell_* := \sqrt{r^2 - a_{\mathrm{out}}^2} = r v_*, \qquad v_* := \sqrt{1 - (a_{\mathrm{out}}/r)^2} > 0,$$

and

$$B_* := \frac{v_*^3}{3(1+v_*)}.$$

Then uniformly over these anchors,

$$E_1^{\mathrm{pop}}(a,r) - E_2^{\mathrm{pop}}(a,r) = E_1^{\mathrm{pop}}(a,r) \ge r^2 B_*. \tag{106}$$

Proof. The two local branch masses are proportional to b and ℓ_b. Their conditional means are se_1 and 0, and their directional variances are $b^2/3$ and $\ell_b^2/3$.

Put $v = \ell_b/b$. The covariance eigenvalue in the e_2 direction is

$$\frac{\ell_b}{b+\ell_b}\frac{\ell_b^2}{3} = \frac{\ell_b^3}{3(b+\ell_b)}.$$

The difference between the e_1 and e_2 covariance eigenvalues is

$$\frac{b^2(1-v)(1+4v+v^2)}{3(1+v)} \ge 0.$$

Hence the smaller covariance eigenvalue, which equals the best affine one-line residual, is the first quantity in (105).

The two generating lines attain zero two-line residual. Conversely, any population minimizer with zero residual must cover a positive-length interval of each generating line. An affine line distinct from a given generating line intersects it in at most one point. Hence one fitted line must equal $\mathbb{R}e_1$ and the other must equal $\mathbb{R}e_2$, proving uniqueness up to permutation.

Finally, $\ell_b \ge \ell_* = r v_*$ when $b = r$ and $|s| \le a_{\mathrm{out}}$. Since $v \mapsto v^3/[3(1+v)]$ is increasing on $[0,\infty)$, (106) follows. □

Witness-bin condition. For an anchor $a = se_1$ with $|s| \le a_{\mathrm{out}}$, define

$$I_{1,a} := [s-r, s+r]e_1, \qquad I_{2,a} := [-\ell_*, \ell_*]e_2. \tag{107}$$

Partition each interval into nine equal bins. Let $\mathcal{W}_a(m_0)$ denote the event that each bin numbered $1,3,5,7,9$ on each interval contains at least m_0 observations. For anchors on the second branch, use the definition with the two coordinates interchanged.

For an achieved two-line energy E, define

$$\epsilon_a(E) := \sqrt{\frac{m_a E}{m_0}}, \qquad \kappa_a(E) := \frac{9\epsilon_a(E)}{2\ell_*}, \qquad A_a(E) := \epsilon_a(E) + 2r\kappa_a(E). \tag{108}$$

Theorem D.3 (Deterministic recovery from a low-energy empirical fit)**.** *Suppose $\mathcal{W}_a(m_0)$ holds and two affine lines $\widehat{\mathcal{L}}_{1,a}, \widehat{\mathcal{L}}_{2,a}$ have achieved empirical energy E. If*

$$\kappa_a(E) < \frac{1}{\sqrt{2}}, \tag{109}$$

then there is a permutation π of $\{1,2\}$ such that, for $j = 1, 2$,

$$\left\| P_{\widehat{\mathcal{L}}_{\pi(j),a}} - P_j \right\|_{\mathrm{op}} \le \kappa_a(E), \tag{110}$$

$$\left\|P_{\widehat{\mathcal{L}}_{\pi(j),a}} - P_j\right\|_F \le \sqrt{2}\,\kappa_a(E), \tag{111}$$

and

$$\operatorname{dist}\big(0, \widehat{\mathcal{L}}_{\pi(j),a}\big) \le A_a(E). \tag{112}$$

Here $P_{\mathcal{L}}$ denotes the projector onto the direction space of the affine line $\mathcal{L}$.

For a noiseless observation $x = ue_j$, residual assignment selects the fitted line matched to branch j whenever

$$|u| > \frac{2A_a(E)}{\sqrt{1-\kappa_a(E)^2} - \kappa_a(E)}. \tag{113}$$

Proof. If every observation in one witness bin had distance greater than $\epsilon_a(E)$ from both returned lines, that bin alone would contribute more than

$$m_0\epsilon_a(E)^2 = m_a E$$

to the total squared residual. Hence each witness bin contains a point whose distance from at least one returned line is at most $\epsilon_a(E)$.

Consider the five witnesses on one generating line. At least three are close to the same returned line. Two of these three are separated by at least 2Δ, where $\Delta = 2\ell_*/9$. Their perpendicular displacements from the returned line differ by at most $2\epsilon_a(E)$. The sine of the angle between the returned and generating directions is therefore at most

$$\frac{\epsilon_a(E)}{\Delta} = \kappa_a(E).$$

The same argument applies to the other generating line. A single direction cannot have sine-angle error strictly below $1/\sqrt{2}$ from both orthogonal directions. Thus the two matchings are distinct and give a bijection.

For rank-one projectors, the operator distance equals the sine of the angle and the Frobenius distance is $\sqrt{2}$ times that sine, proving (110)–(111). Every selected witness has norm at most $2r$. Combining its residual error with the directional error gives

$$\operatorname{dist}\big(0, \widehat{\mathcal{L}}_{\pi(j),a}\big) \le \epsilon_a(E) + 2r\kappa_a(E) = A_a(E),$$

which proves (112).

For $x = ue_j$, the residual to the matched line is at most

$$A_a(E) + |u|\kappa_a(E),$$

whereas the residual to the competing line is at least

$$|u|\sqrt{1-\kappa_a(E)^2} - A_a(E).$$

The strict comparison of these two quantities gives (113). □

Exact-N witness-count scale. Take exactly N independent uniform observations on each segment and put $n = 2N$. For N sufficiently large that the following integer is positive, define

$$\mu_{*,N} := \frac{(N-1)\ell_*}{9L}, \qquad m_{0,N} := \left\lfloor \frac{\mu_{*,N}}{2} \right\rfloor. \tag{114}$$

Proposition D.4 (Uniform availability of witnesses)**.** *Take exactly N independent uniform observations on each segment, independently across the two branches, and let $\mu_{*,N}$ and $m_{0,N}$ be as in* (114)*. Then*

$$\Pr\left(\bigcap_{\substack{a \text{ sample-centered} \\ |s(a)|\le a_{\mathrm{out}}}} \mathcal{W}_a(m_{0,N})\right) \ge 1 - 10n\exp(-\mu_{*,N}/8). \tag{115}$$

The conclusion holds simultaneously over all eligible sample-centered anchors and hence also over anchors selected adaptively from the data. For anchors lying on the second branch, $\mathcal{W}_a(m_{0,N})$ is understood with the two branch coordinates interchanged.

Proof. Fix a sample-centered anchor a and condition on its location and generating branch. After omitting the center itself, the remaining observations on each branch are independent uniform draws.

Every tested bin on the truncated branch interval has length $2\ell_*/9$. Every tested bin on the anchor's own branch has length $2r/9 \geq 2\ell_*/9$. Since each complete segment has length $2L$, the expected number of observations in every tested bin is therefore at least

$$\mu_{*,N} = \frac{(N-1)\ell_*}{9L}.$$

For a binomial random variable with mean at least $\mu_{*,N}$, the multiplicative Chernoff bound gives

$$\Pr\left\{\text{bin count} < \frac{\mu_{*,N}}{2}\right\} \leq \exp(-\mu_{*,N}/8).$$

There are ten tested bins at each anchor and at most n possible sample-centered anchors. A union bound therefore yields (115). Independence between overlapping bins or between different anchors is not required.

Because the event holds simultaneously over all eligible sample-centered anchors, it continues to hold for any data-dependent subset of them, including the anchors returned by the medoid–farthest-point selection procedure. □

Uniform local-fitting event. Let $\mathcal{A}_N^{\mathrm{anc}}$ denote the nonempty set of anchors selected for local branch fitting. For each $a \in \mathcal{A}_N^{\mathrm{anc}}$, let m_a be its fitting-neighborhood size, let $\widehat{\Sigma}_a$ be its empirical centered covariance, and let $\Sigma(a)$ be the corresponding population covariance.

Write

$$\widehat{\boldsymbol{\mathcal{L}}}_{2,a} = (\widehat{\mathcal{L}}_{1,a}, \widehat{\mathcal{L}}_{2,a})$$

for the returned two-line candidate. For deterministic tolerances $\eta_{c,N} \geq 0$ and $\Delta_{\mathrm{opt},N} \geq 0$, define $\mathcal{G}_N^{\mathrm{fit}}$ by

$$\begin{aligned}\mathcal{G}_N^{\mathrm{fit}} := \bigcap_{a \in \mathcal{A}_N^{\mathrm{anc}}} \Big\{ & \|\widehat{\Sigma}_a - \Sigma(a)\|_{\mathrm{op}} \leq r^2 \eta_{c,N}, \\ & \mathcal{W}_a(m_{0,N}) \text{ holds}, \\ & \widehat{\boldsymbol{\mathcal{L}}}_{2,a} \text{ is a valid two-line candidate}, \\ & E_2(a) \leq \widehat{E}^{\star}_{2,a} + \Delta_{\mathrm{opt},N} \Big\}.\end{aligned} \tag{116}$$

Here validity refers to the minimum-group-size and other admissibility checks imposed by the fitting routine. In the noiseless model, $\widehat{E}^{\star}_{2,a} = 0$ because the two generating lines give zero empirical residual. If the reference optimum is restricted to valid fits, the same conclusion holds whenever $b_{\min} \leq 5m_{0,N}$.

Set

$$\begin{aligned} m_{\max,N} &:= \max_{a \in \mathcal{A}_N^{\mathrm{anc}}} m_a, \\ \epsilon_{N,\mathrm{fit}} &:= \sqrt{\frac{m_{\max,N}\Delta_{\mathrm{opt},N}}{m_{0,N}}}, \\ \delta_N &:= \frac{9\epsilon_{N,\mathrm{fit}}}{2\ell_*}, \\ a_N &:= \epsilon_{N,\mathrm{fit}} + 2r\delta_N. \end{aligned} \tag{117}$$

Also define the uniform empirical one-line floor

$$G_N := r^2(B_* - \eta_{c,N}) \tag{118}$$

and, whenever $\delta_N < 1/2$, the certified unresolved set

$$w_N := \frac{2a_N}{1-2\delta_N}, \qquad \mathcal{B}_N := \{i : |s_i| \leq w_N\}. \tag{119}$$

Finally, let

$$\widehat{F}_{j,N}(u) := \frac{1}{N} \sum_{i:y_i=j} \mathbf{1}_{\{s_i \le u\}}, \qquad F(u) := \frac{u+L}{2L}, \quad -L \le u \le L,$$

and define the branchwise empirical CDF error

$$z_N := \max_{j=1,2} \sup_{-L \le u \le L} |\widehat{F}_{j,N}(u) - F(u)|. \tag{120}$$

Corollary D.5 (Approximate empirical fitting implies representation accuracy)**.** *Assume the balanced noiseless orthogonal-segment model with fixed $r/L < 1/4$, and suppose every retained candidate satisfies $|s| \le a_{\mathrm{out}} < r$. Assume that every retained component contains at least one selected anchor, that $K_{\max} \ge 2$, and that the uniform local-fitting event $\mathcal{G}_N^{\mathrm{fit}}$ in* (116) *occurs.*

If

$$\delta_N < \frac{1}{2}, \tag{121}$$

and

$$G_N > \eta_{\mathrm{sel}} \epsilon_E, \qquad \Delta_{\mathrm{opt},N} \le \eta_{\mathrm{sel}} \max\{G_N, \epsilon_E\}, \qquad 0 < \eta_{\mathrm{sel}} < 1, \tag{122}$$

then the selection rule (2) *rejects $K = 1$ and selects the valid $K = 2$ candidate at every selected anchor.*

Moreover, after a permutation of the two fitted lines at each anchor,

$$\max_{a \in \mathcal{A}_N^{\mathrm{anc}}} \max_{j=1,2} \big\| P_{\widehat{\mathcal{L}}_{j,a}} - P_j \big\|_F \le \sqrt{2}\, \delta_N, \qquad \max_{a \in \mathcal{A}_N^{\mathrm{anc}}} \max_{j=1,2} \operatorname{dist}\big(0, \widehat{\mathcal{L}}_{j,a}\big) \le a_N. \tag{123}$$

Suppose additionally that every output following the local-PCA path has Frobenius projector error at most $\varepsilon_{\mathrm{loc},N}$. Then the final representations satisfy

$$\max_{i \notin \mathcal{B}_N} \|\widetilde{P}_i - P_i^*\|_F \le \max\big\{\varepsilon_{\mathrm{loc},N}, \sqrt{2}\, \delta_N\big\}. \tag{124}$$

The unresolved set obeys

$$\frac{|\mathcal{B}_N|}{n} \le \frac{w_N}{L} + 2 z_N. \tag{125}$$

Consequently, if

$$\frac{m_{\max,N}}{m_{0,N}} = O(1), \qquad \Delta_{\mathrm{opt},N} = o(r^2), \qquad \eta_{c,N} \to 0, \qquad \varepsilon_{\mathrm{loc},N} \to 0, \qquad z_N \to 0, \tag{126}$$

then

$$\max_{i \notin \mathcal{B}_N} \|\widetilde{P}_i - P_i^*\|_F \to 0, \qquad \frac{|\mathcal{B}_N|}{n} \to 0. \tag{127}$$

Thus, taking $B_N = \mathcal{B}_N$, the representation hypotheses of Theorem B.84 hold. Under the spectral and rounding assumptions of Theorem 4.3, the misclassification fraction satisfies

$$\operatorname{err}(\widehat{y}_N, y_N) \to 0.$$

Finally, let $\mathcal{H}_N^{\mathrm{fit}}$ denote the intersection of the covariance, validity, and additive-optimization clauses in (116)*, excluding only the witness-bin clauses. Then Proposition D.4 and a union bound give*

$$\Pr(\mathcal{G}_N^{\mathrm{fit}}) \ge 1 - \Pr\big((\mathcal{H}_N^{\mathrm{fit}})^c\big) - 10n \exp(-\mu_{*,N}/8). \tag{128}$$

No bound on $\Pr((\mathcal{H}_N^{\mathrm{fit}})^c)$ is asserted for the finite-restart fitting routine. The bound also does not cover retained-component coverage or the local-PCA-path accuracy assumption.

Proof. For $K = 1$, the deterministic mean–SVD fit attains the smallest eigenvalue of $\widehat{\Sigma}_a$. By Theorem D.2 and Weyl's inequality, on $\mathcal{G}_N^{\mathrm{fit}}$,

$$E_1(a) \ge r^2 B_* - r^2 \eta_{c,N} = G_N$$

uniformly over selected anchors. The first inequality in (122) therefore excludes $K = 1$ under (2).

Because $\widehat{E}^{\star}_{2,a} = 0$, the additive fitting contract gives

$$E_2(a) \le \Delta_{\mathrm{opt},N} \le \eta_{\mathrm{sel}} \max\{G_N, \epsilon_E\} \le \eta_{\mathrm{sel}} \max\{E_1(a), \epsilon_E\}.$$

Thus the valid $K = 2$ candidate qualifies. Since $K = 1$ has been rejected, the smallest qualifying value is $K = 2$.

For every selected anchor,

$$\epsilon_a(E_2(a)) \le \epsilon_{N,\mathrm{fit}}, \qquad \kappa_a(E_2(a)) \le \delta_N, \qquad A_a(E_2(a)) \le a_N.$$

Theorem D.3 therefore proves (123). Since $\delta_N < 1/2$,

$$\sqrt{1-\delta_N^2} - \delta_N \ge 1 - 2\delta_N.$$

Hence every branch-assignment-path observation outside $\mathcal{B}_N$ is assigned to its matching fitted line. Combining this with the local-PCA-path assumption proves (124).

The population mass of $[-w_N, w_N]$ on either branch is w_N/L. The definition of z_N therefore gives

$$\frac{|\mathcal{B}_N|}{n} \le \frac{w_N}{L} + 2z_N,$$

also when w_N is data-dependent. Under (126), $\epsilon_{N,\mathrm{fit}} = o(r)$, $\delta_N \to 0$, $a_N = o(r)$, and $w_N/r \to 0$. This proves (127) and permits application of the cited recovery theorems.

Finally, Proposition D.4 and a union bound with $\mathcal{H}_N^{\mathrm{fit}}$ prove (128). □

Scope of the certificate. The population and witness results are specific to straight, noiseless, two-branch configurations. They verify that low empirical two-line energy is geometrically meaningful and supply one sufficient route to the representation assumptions. They do not show that finitely many random initializations attain the additive optimization contract. The general representation theorem does not require every anchor to return two models and permits arbitrary returned projectors on a vanishing fraction of observations.

D.2 Rounding without row normalization

Lemma D.6 (Objective transfer between two embeddings)**.** *Let $Y_i, Z_i \in \mathbb{R}^2$ be two embeddings of a balanced two-class sample. Suppose the ideal rows Z_i satisfy the separated-centroid condition of Theorem B.68 with margin $\Delta > 0$. Write J_0 for their true-partition cost and $E = n^{-1}\sum_i \|Y_i - Z_i\|^2$. If returned centers and labels on Y have cost at most α times its global optimum, $\alpha \ge 1$, then*

$$\mathrm{err} \le \frac{2}{\Delta}\left[(\alpha-1)J_0 + 2\sqrt{\alpha}(\sqrt{\alpha}+1)\sqrt{J_0 E} + (\sqrt{\alpha}+1)^2 E\right].$$

Proof. For a fixed partition, centroid fitting is an orthogonal projection onto vectors constant on each cluster. The triangle inequality for its residual therefore gives $|\sqrt{J_Y(\mathcal{P})} - \sqrt{J_Z(\mathcal{P})}| \le \sqrt{E}$. For the returned partition $\widehat{\mathcal{P}}$,

$$\sqrt{J_Z(\widehat{\mathcal{P}})} \le \sqrt{\alpha}\sqrt{J_Y(\mathcal{P}_0)} + \sqrt{E} \le \sqrt{\alpha}\sqrt{J_0} + (\sqrt{\alpha}+1)\sqrt{E}.$$

Squaring, subtracting J_0, and using the separated-centroid excess-cost bound proves the result. No unit-row assumption is used. □

Theorem D.7 (Conditional recovery with degree-rescaled eigenvectors)**.** *Use the balanced, noiseless orthogonal sampling model, fixed $\sigma_T = 0.2$, and $5.3 \cdot 10^{-4} \le r/L < 1/4$ of Theorem B.84. Assume its representation conditions: arbitrary rank-one projectors on B_N, $|B_N|/n \to 0$, and uniform Frobenius error $\varepsilon_N \to 0$ elsewhere, almost surely. Build the same loop-free graph using empirical A1. Let U_N be its exact bottom-two symmetric-normalized Laplacian eigenvectors. Apply two-means to $D_N^{-1/2}U_N$, without row normalization, with achieved objective ratio $\alpha_N \to 1$ almost surely. Then the misclassification fraction tends to zero almost surely.*

The same conclusion holds for orthonormal numerical columns Q_N if $\inf_{O\in O(2)} \|Q_N - U_N O\|_F \to 0$ almost surely and the objective contract holds on their actual degree-rescaled embedding. The analogous statement holds with all limits in probability.

Proof. Throughout this proof $d_{N,i} = \sum_j W_{ij}$ is the raw graph degree and $\widehat{d}_{N,i} = d_{N,i}/n$ the empirical kernel degree. Stars denote the exact-tangent oracle on the same observations and A1 scale. The fixed-radius degree argument in Theorem B.84 gives, eventually, $d_{N,i}, d^*_{N,i} \geq np_0$ for some fixed $p_0 > 0$, also for the hybrid graph. It applies to arbitrary rank-one projectors because the tangent weight has a positive lower bound and A1 is bounded below.

Retain actual projectors on B_N and use exact ones elsewhere to form the hybrid. Every actual-to-hybrid projector change is at most ε_N, including zero change on B_N. Thus the multiplicative comparison applies to all supported edges, including good–bad edges. With $\eta_N = \exp(2\varepsilon_N/\sigma_T^2) - 1$, graph stability gives $\|S_N - S^*_N\|_{\mathrm{op}} \to 0$. In addition, the multiplicative degree error has ℓ_2 norm at most $\eta_N n^{3/2}$, and the bad-incident-edge term has norm at most $n\sqrt{2|B_N|}$. Consequently

$$\frac{\|d_N - d^*_N\|_2}{n^{3/2}} \leq \eta_N + \sqrt{2|B_N|/n} \longrightarrow 0.$$

The oracle fixed-radius gap and the sampling argument of Section B.20 imply aligned exact eigenspace convergence $\|U_N - U^*_N O_N\|_F \to 0$.

A uniform row bound follows here rather than being assumed. Since weights are at most one and degrees at least np_0, $\|(S^*_N)_{i,:}\|_2 \leq 1/(\sqrt{n}p_0)$. The top two adjacency eigenvalues are bounded below by some fixed $b > 0$ eventually: their population limits are 1 and $1 - \lambda_2 > 0$. From $S^*_N U^*_N = U^*_N \Lambda_N$ one obtains $\sqrt{n}\|(U^*_N)_{i,:}\|_2 \leq 1/(p_0 b)$.

Let $\bar{d} = \int d(\omega)\, d\mu(\omega) > 0$ be the mean population kernel degree. The correctly rescaled empirical rows are

$$Y_{N,i} = n\sqrt{\bar{d}}\, d_{N,i}^{-1/2}(U_N)_i = \sqrt{n\bar{d}}\, \widehat{d}_{N,i}^{-1/2}(U_N)_i.$$

The common scalar leaves labels and objective ratios unchanged. Splitting $Y_N - Y^*_N O_N$ into the eigenspace and degree terms, using the degree lower bound and the Lipschitz estimate for $x^{-1/2}$, gives

$$\frac{1}{n}\|Y_N - Y^*_N O_N\|_F^2 \leq C\left(\|U_N - U^*_N O_N\|_F^2 + \frac{\|d_N - d^*_N\|_2^2}{n^3}\right) \longrightarrow 0.$$

The uniform oracle kernel-degree convergence and common-space eigenfunction convergence in Section B.20 likewise imply that Y^*_N approaches the ideal rows $Z_i = (1, \ell_i f_r(|s_i|))$ in mean square, up to alignment, where $\ell_i = 1$ or -1 is the branch sign. Here division by the kernel degree is continuous above its positive floor; continuity of the population profile transfers cell values to sample values.

Write $m = \min f_r > 0$ and $F = \max f_r < \infty$. The certified population profile bound from the two-means analysis gives $F/m < \sqrt{3} < 3$ in this radius range. Each ideal branch has diameter at most $F - m$, whereas opposite-branch subset centroids differ by at least $2m$ in their second coordinate. Thus $\Delta = 4m^2 - (F - m)^2 > 0$, uniformly in sample size, and $J_0 \leq F^2$. Lemma D.6 proves consistency. For numerical Q_N, its additional rescaled mean-square error is bounded by $(\bar{d}/p_0)\inf_O \|Q_N - U_N O\|_F^2$, so the same proof applies. □

This theorem aligns the abstract embedding with the practical backend, but still assumes numerical subspace accuracy and asymptotically optimal rounding. Finite restarts and a solver termination flag do not establish these contracts. For exact eigenvectors on a connected two-cluster graph, the first degree-rescaled coordinate is constant, so global two-means can be realized by sorting the second coordinate and checking all splits with prefix sums. This provides an abstract optimization realization, not a claim about the implemented finite Lloyd runs.

D.3 Proofs of the main-text corollaries

Proof of Corollary 4.4. On the eventual event $\mathcal{G}^{\mathrm{fit}}_N$, Corollary D.5 shows that (121) and (122) exclude $K = 1$ and make the valid $K = 2$ candidate qualify at every selected anchor.

More precisely, taking

$$B_N = \mathcal{B}_N,$$

the same corollary gives

$$\max_{i\notin B_N} \|\widetilde{P}_i - P_i^*\|_F \le \max\{\varepsilon_{\mathrm{loc},N}, \sqrt{2}\,\delta_N\}, \qquad \frac{|B_N|}{n} \le \frac{w_N}{L} + 2z_N. \tag{129}$$

By (117) and (9),

$$\epsilon_{N,\mathrm{fit}} = o(r), \qquad \delta_N \to 0, \qquad a_N = o(r), \qquad \frac{w_N}{r} \to 0 \quad \text{almost surely.}$$

The two branchwise empirical distributions converge uniformly almost surely under the nested uniform sampling model, so $z_N \to 0$ almost surely. Equation (129) therefore implies

$$\max_{i\notin B_N} \|\widetilde{P}_i - P_i^*\|_F \to 0, \qquad \frac{|B_N|}{n} \to 0 \quad \text{almost surely.}$$

These are the representation conditions required by Theorem 4.3. Applying that theorem under its spectral and rounding assumptions proves

$$\mathrm{err}(\widehat{y}_N, y_N) \to 0 \quad \text{almost surely.}$$

The argument is deterministic conditional on the stated module events. Replacing eventual almost-sure occurrence and convergence by their in-probability counterparts gives the corresponding in-probability conclusion. □

Proof of Corollary 4.5. Corollary 4.4 supplies a sequence B_N satisfying

$$\frac{|B_N|}{n} \to 0, \qquad \max_{i\notin B_N} \|\widetilde{P}_i - P_i^*\|_F \to 0.$$

These are precisely the representation hypotheses of Theorem D.7. Applying that theorem to the exact degree-rescaled embedding $D_W^{-1/2}U_N$ and the objective contract $\alpha_N \to 1$ proves the first conclusion.

For numerical eigenvectors, the additional assumption

$$\inf_{O\in O(2)} \|Q_N - U_N O\|_F \to 0$$

is the numerical-subspace condition in Theorem D.7. Its numerical version therefore gives the same conclusion for the actual degree-rescaled embedding $D_W^{-1/2}Q_N$. The in-probability statement follows from the corresponding part of that theorem. □

E Recovery Proof Roadmap and Explicit Existence Witness

The following expands the existence and transfer argument summarized in Section 4; it retains the same hypotheses.

Proof structure. Proposition 4.1 gives correct branch selection outside B_N, with Frobenius error at most $\varepsilon_N = \max\{\varepsilon_{\mathrm{loc},N}, \sqrt{2}\delta_N\}$. Uniform empirical CDF error z_N gives $|B_N|/n \le w_N/L + 2z_N \to 0$, even for data-dependent w_N. At fixed radius, A1 converges to $r/2$ and same-branch neighborhood counts yield degree at least cn in all three graphs. The tangent Gaussian has a positive lower bound for arbitrary rank-one projectors, so this degree argument also includes the unresolved points. Equation (6) implies operator convergence to the empirical oracle.

Population analysis of the original product kernel establishes a positive bottom-two spectral separation and, in the stated radius range, a positive global two-means rounding separation. A common-space sampling comparison transfers these properties to the imperfect empirical embedding; objective stability then yields label consistency. The full argument is Theorem B.1, with the transfer detailed in Theorem B.84. For fixed geometry its error budget is

$$\mathrm{err}(\widehat{y}_N, y_N) \le C\left[(\alpha_N - 1) + \varepsilon_N + \sqrt{w_N/L + 2z_N} + \mathcal{B}_N + \omega_\Phi(h_N)\right]$$

eventually almost surely. The last two terms are the oracle sampling operator error and population eigenfunction transport error, both tending to zero. Their definitions and explicit finite-sample bounds are given in Appendix B.20. No estimator-specific convergence rate is inferred from this decomposition.

Existence of admissible imperfect representations. The sufficient conditions above describe a nonempty class, rather than only a hypothetical implication. The following explicit witness uses the representation-level theorem and keeps the same empirical graph construction.

Corollary E.1 (A nonempty recoverable class)**.** *In the sampling and radius regime of Theorem 4.3, let $u_N \downarrow 0$ and $\beta_N \to 0$ be deterministic, with $0 < u_N < L$. Set $B_N = \{i : |s_i| \le u_N\}$ and let R_{β_N} be a planar rotation. The rank-one representations*

$$\widetilde{P}_i = \begin{cases} R_{\beta_N} P_i^* R_{\beta_N}^\top, & i \notin B_N, \\ Q_{i,N}, & i \in B_N, \end{cases}$$

where $Q_{i,N}$ are arbitrary rank-one orthogonal projectors, yield $\mathrm{err}(\widehat{y}_N, y_N) \to 0$ almost surely under empirical A1 and the abstract spectral realization with exact global two-means.

Proof. Uniform convergence of the two empirical CDFs gives $|B_N|/n \le u_N/L + 2z_N \to 0$ almost surely. Outside B_N, $\|\widetilde{P}_i - P_i^*\|_F = \sqrt{2}|\sin\beta_N| \to 0$. A global two-means optimum exists for a finite embedding (minimize over its finitely many labelings, using centroids for nonempty clusters), so $\alpha_N = 1$ is admissible. The conclusion follows from Theorem B.84. □

This is an idealized representation witness using the true tangents, not a prescription for estimating them from unlabeled observations. For example, $u_N = L/\sqrt{N+1}$ and $\beta_N = 1/(N+1)$ give a concrete sequence with imperfect representations. The substantive conclusion is that the specified graph and rounding recover the labels despite these errors and the intersection, not merely that true labels can be supplied. At the local-model level, both exact generating lines offered at every used anchor also satisfy the margin assumptions with $a_N = \delta_N = 0$; the general theorem does not require every anchor to have two models. These witnesses establish attainability of the abstract accuracy conditions. The practical construction seeks comparable representations through fitting.

Fitting certificates and the practical embedding. The representation conditions also admit a local energy certificate. For two noiseless orthogonal lines, if separated witness bins on both branches contain enough samples, a sufficiently small achieved two-line residual controls direction and offset errors and hence residual assignment (Theorem D.3). Corollary D.5 is one sufficient route to the representation hypotheses; it does not prove that finite restarts attain that residual or require all anchors in the general theory to return two models.

Theorem D.7 further establishes the same fixed-radius consistency conclusion using $D_W^{-1/2}U$, without row normalization, under the general imperfect-representation conditions and an asymptotically optimal two-means objective contract. It also accommodates numerical eigenvectors whose aligned subspace error vanishes. The proof controls degrees as well as eigenspaces and transfers two-means objectives between embeddings. Thus a recoverable abstract regime exists for the backend's embedding too; its numerical and optimization contracts remain separate from the finite routines used in the experiments.