IDENTIFIABILITY OF TEMPORAL ACTION STRUCTURE UNDER INCOMPLETE OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding continuous human activity requires locating action changes and determining whether separated intervals show the same action, even when observations are incomplete. However, occlusions, dropped frames, field-of-view truncation, and sensor failures can resemble real transitions, leaving boundaries and action relations ambiguous. Existing pipelines typically complete missing data before segmentation, so reconstruction errors can propagate without revealing whether the inferred structure is uniquely supported. The key idea is to compare candidate actions using only available measurements and retain every explanation consistent with that evidence. The paper defines complete, partial, and impossible identifiability, derives conditions for preserving action distinctions and benefiting from temporal context, and develops an adaptive segmentation method guided by these conditions. Experiments on five motion benchmarks show improved segmentation across several missing-observation patterns, while the theory exposes failures caused by missing intervals or indistinguishable visible evidence.

1 INTRODUCTION

Understanding continuous, untrimmed human activity requires answering two interdependent questions: *when does the action change, and do separated intervals depict the same action?* Boundaries determine the purity of the observations used to characterize an action; cross-time action relations provide evidence for confirming those boundaries. An early boundary error can mix incompatible motion statistics, while an incorrect action relation can reinforce a false transition. Recovering these two aspects without training on target-dataset boundary or action labels is therefore a basic problem for human–robot collaboration, embodied intelligence, and intelligent video analysis.

Missing evidence can imitate action change. Real streams contain local or complete occlusion, field-of-view truncation, consecutive dropped frames, and sensor anomalies. Such missingness need not be uniform or independent: visibility can depend on pose, motion speed, scene structure, and the location of an action transition. A sudden disappearance of evidence can then resemble a genuine action change. Conversely, pouring water and putting down a cup may be joined by a gradual transition, while opening and closing a door can have similar local appearances. A single boundary point or action label may therefore be more precise than the evidence permits. The relevant question is which intervals and relations remain stable across all explanations compatible with the available evidence.

From prediction to identifiability. We describe action structure by a boundary set B and an action equivalence relation Π, which records whether two times share an action without assigning a semantic name. Given a model class, available history H, observations Y, and a reliability mask M, let $\mathcal{A}(Y, M; H)$ contain all admissible explanations. Structure is identifiable when their projections agree on (B, Π), independent of label numbering. Partial agreement instead yields certified relations and boundary uncertainty regions. This perspective distinguishes a unique optimum of a chosen objective from a uniquely supported explanation of the observations.

Why a sequence of modules is not enough. Small completion error need not preserve the ordering of candidate action residuals. A local transition cue may respond to an observation outage rather than motion, and time-independent pairwise relations do not describe the appearance, recurrence, or

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.

disappearance of an action in a stream. Semantic names supplied by a vision–language system would not by themselves resolve these geometric ambiguities: delayed, dependent, or incorrect answers require a compatibility test. These considerations motivate an information-based question, not a claim that every component is implemented here: *what evidence preserves action structure, and what additional evidence can genuinely narrow its ambiguity?*

The component studied in this paper. We investigate finite sequences using affine-subspace models for locally low-dimensional motion (Xia et al., 2017). An exact masked-distance identity separates missing residual energy from leakage caused by refitting latent coordinates. A temporal margin then connects distance distortion to stable action assignments and explains how impure neighborhoods can reverse the benefit of context. Temporal Segmentation with Contextual Consistency (TSCC) realizes this connection through a masked-distance objective and safeguarded updates. Boundary sets and cross-time action relations are both read from the resulting label sequence, but the neighborhood proposals are fixed after construction; we do not claim joint global recovery of all admissible structures.

Our contributions characterize the identifiability of temporal action structure through residual preservation under explicit sampling assumptions, margin and perturbation conditions for stable decisions, and a set-valued account of partial identifiability with observational lower bounds. Experiments cover coordinate, frame, joint, and tracking-failure masks. They use a supplied action count K, so unknown action identities must not be confused with unknown model order. Online action birth, reactivation, merging, and dormancy, and the use of noisy semantic side information, remain extensions rather than evaluated capabilities.

2 Related Work

Incomplete geometry. Sparse Subspace Clustering with Entry-Wise Zero Filling (SSC-EWZF) (Yang et al., 2015), projection-based zero-filled SSC (PZF-SSC) (Tsakiris & Vidal, 2018), Parallel Estimation and Tracking by REcursive Least Squares (PETRELS) (Dung et al., 2021), Fusion Subspace Clustering (FSC) (Mahmood & Pimentel-Alarcón, 2022), integer-programming missing-data segmentation (MISS-DSG) (Soni et al., 2023), and Contrastive Subspace Clustering (CSC) (Li & Pimentel-Alarcón, 2026) address incomplete geometry but not persistence along an ordered sequence.

Temporal structure. Ordered Subspace Clustering (OSC) (Tierney et al., 2014), Temporal Subspace Clustering (TSC) (Li et al., 2015), Consistency-and-Diversity Motion Segmentation (CDMS) (Zhou et al., 2022), the Velocity-Sensitive Dual-Side Autoencoder (VSDA) (Bai et al., 2022), and Temporal Visual-Semantics Subspace Clustering (TVS-SC) (Xing & Zhao, 2026) use ordering or motion-sensitive features. TSCC instead aggregates observed-coordinate residuals without imposing hard equality constraints on adjacent labels.

3 Action Structure and Masked Geometry

3.1 Admissible Structure and Partial Identifiability

Let $z = (z_1, \dots, z_N)$ be latent action labels. Their structural outputs are

$$B(z) = \{t < N : z_t \neq z_{t+1}\}, \qquad \Pi(z) = \{(i, j) : z_i = z_j\}. \tag{1}$$

Thus Π can link nonadjacent action bouts, whereas B describes their temporal arrangement. These outputs are coupled: $t \in B(z)$ exactly when $(t, t+1) \notin \Pi(z)$. They are not two freely chosen targets.

For a prescribed model class $\mathfrak{M}$ and noise radius ν, define

$$\mathcal{A}(Y, M; H) = \{\theta \in \mathfrak{M}(H) : \|M \odot (X_\theta - Y)\|_{\mathrm{F}} \leq \nu\}, \qquad \mathcal{F} = \{(B(z_\theta), \Pi(z_\theta)) : \theta \in \mathcal{A}\}. \tag{2}$$

Here θ includes a complete trajectory, its labels, and its action models; $\mathfrak{M}(H)$ encodes allowable dynamics, durations, and history constraints. Unobserved entries of Y are arbitrary placeholders. The

analysis below specializes M to binary coordinate masks; a fractional reliability value would require a calibrated observation model. We take $H = \varnothing$ for the finite-sequence experiments.

If $\mathcal{F} \neq \varnothing$, full structural identifiability means $|\mathcal{F}| = 1$, even if the trajectory or subspace parameters are not unique. Otherwise, a boundary or pairwise relation is certified only when every member of $\mathcal{F}$ agrees on it. Define $B^- = \bigcap_{\mathcal{F}} B$ and $B^+ = \bigcup_{\mathcal{F}} B$. Locations in B^- are necessary boundaries, those outside B^+ are excluded, and $B^+ \setminus B^-$ is unresolved. Connected components of the unresolved set are uncertainty bands, not necessarily one transition each. An empty $\mathcal{F}$ signals model incompatibility, not certainty. Pairwise relations that vary across $\mathcal{F}$ form action ambiguity sets. Additional evidence is informative only when it removes at least one feasible structure without eliminating the feasible set.

3.2 Observation Model and Residual Decomposition

Let $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, \ldots, N$, be an ordered sequence, and let Φ_i be its diagonal binary observation mask. We observe $\Phi_i \mathbf{x}_i$, with $m_i = \mathrm{tr}(\Phi_i) > 0$ and $p_i = m_i/D$. Candidate k is an affine subspace $\mathcal{S}_k = \mathbf{o}_k + \mathrm{span}(\mathbf{U}_k)$, where $\mathbf{U}_k^{\mathrm{T}}\mathbf{U}_k = \mathbf{I}_{d_k}$ and $\mathbf{U}_k^{\mathrm{T}}\mathbf{o}_k = 0$. The latter is a canonical offset constraint, not uniqueness of the basis. Write

$$\varepsilon_{k,i} = \|(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^{\mathrm{T}})(\mathbf{x}_i - \mathbf{o}_k)\|_2^2, \qquad \xi_{k,i} = \min_{\mathbf{a}} \|\Phi_i(\mathbf{x}_i - \mathbf{o}_k - \mathbf{U}_k\mathbf{a})\|_2^2.$$

The normalized observed residual is $\rho_{k,i} = p_i^{-1}\xi_{k,i}$. It requires no values at unobserved coordinates. A pseudoinverse defines the minimum even when the restricted basis is rank deficient:

$$\xi_{k,i} = \|\mathbf{P}_\perp^\Phi \Phi_i(\mathbf{x}_i - \mathbf{o}_k)\|_2^2, \quad \mathbf{P}_\perp^\Phi = \mathbf{I} - \Phi_i\mathbf{U}_k(\mathbf{U}_k^{\mathrm{T}}\Phi_i\mathbf{U}_k)^\dagger\mathbf{U}_k^{\mathrm{T}}\Phi_i. \tag{3}$$

A deterministic identity. Suppress the sample and candidate indices. Decompose $\mathbf{x} - \mathbf{o} = \mathbf{U}\mathbf{a} + \mathbf{r}$, with $\mathbf{U}^{\mathrm{T}}\mathbf{r} = 0$, and set $\mathbf{G} = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{U}$ and $\mathbf{c} = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{r}$. Then, for every fixed mask,

$$\rho = \underbrace{p^{-1}\mathbf{r}^{\mathrm{T}}\Phi\mathbf{r}}_{\text{retained energy}} - \underbrace{p^{-1}\mathbf{c}^{\mathrm{T}}\mathbf{G}^\dagger\mathbf{c}}_{\text{refitting leakage}}. \tag{4}$$

The second term is nonnegative. Consequently, inverse-rate scaling does not make the fitted residual unbiased: it corrects retained energy, while refitting can still systematically decrease the distance.

Three quantities therefore control reliability: preservation of residual energy, conditioning of the restricted basis, and the loss of orthogonality measured by $\mathbf{c}$. An observation percentage alone cannot describe all three. For example, two affine models with identical restrictions to the observed coordinates are indistinguishable regardless of their separation on the missing coordinates. No method can recover a distinction that is absent from the observed coordinates without additional assumptions or evidence.

3.3 Random-Coordinate Preservation and Its Interpretation

For a fixed sample and a fixed candidate, suppose the observed set is a uniformly chosen subset of size m. Define

$$\mu_{\mathbf{U}} = \frac{D}{d}\max_j \|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}\|_2^2, \qquad \mu_{\mathbf{r}} = \frac{D\|\mathbf{r}\|_\infty^2}{\|\mathbf{r}\|_2^2}.$$

The two coherences measure concentration of subspace directions and residual energy. They are distinct: a diffuse subspace does not imply a diffuse off-subspace error. The following theorem retains both dependencies.

Theorem 1 (Masked residual preservation with explicit dependencies)**.** *Assume that Ω is sampled uniformly without replacement and $\varepsilon(\mathbf{x}, \mathbf{U}, \mathbf{o}) > 0$. Fix $\eta_{\mathbf{U}}, \eta_{\mathbf{r}}, \eta_{\mathrm{c}} \in (0, 1)$ such that $\eta_{\mathbf{r}} + \eta_{\mathrm{c}}^2/(1 - \eta_{\mathbf{U}}) < 1$. There exists a universal constant $c_0 > 0$ such that, if*

$$m \geq c_0 \max\left\{\frac{\mu_{\mathbf{U}} d\log(2d/\delta)}{\eta_{\mathbf{U}}^2}, \frac{\mu_{\mathbf{r}}\log(2/\delta)}{\eta_{\mathbf{r}}^2}, \frac{\mu_{\mathbf{U}}\mu_{\mathbf{r}} d\log(2d/\delta)}{\eta_{\mathrm{c}}^2}\right\}, \tag{5}$$

then, with probability at least $1 - 4\delta$,

$$p\left(1 - \eta_{\mathbf{r}} - \frac{\eta_{\mathrm{c}}^2}{1 - \eta_{\mathbf{U}}}\right)\varepsilon(\mathbf{x}, \mathbf{U}, \mathbf{o}) \leq \xi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o}) \leq p(1 + \eta_{\mathbf{r}})\varepsilon(\mathbf{x}, \mathbf{U}, \mathbf{o}). \tag{6}$$

The upper bound requires retained-energy control, whereas the lower bound must also prevent a competing subspace from fitting away its rejecting evidence. This asymmetry explains why restricted conditioning is essential.

From a formula to an observation budget. For $0 < \tau < 1$, simultaneous relative error $|\rho_{k,i} - \varepsilon_{k,i}| \le \tau\varepsilon_{k,i}$ over all NK fixed sample–candidate pairs holds with probability at least $1 - \delta$ under the sufficient observation budget

$$m_i \ge C \max_k \left\{ \mu_{\mathbf{U}_k} d_k \log \frac{8NKd_k}{\delta},\ \frac{\mu_{\mathbf{r}_{k,i}}}{\tau^2} \log \frac{8NK}{\delta},\ \frac{\mu_{\mathbf{U}_k}\mu_{\mathbf{r}_{k,i}} d_k}{\tau} \log \frac{8NKd_k}{\delta} \right\}. \tag{7}$$

Here and below zero complete residuals are preserved exactly and require no residual coherence. The budget displays different accuracy costs: energy fluctuations scale as τ^{-2}, whereas the squared leakage term gives a τ^{-1} requirement. If the sufficient budget exceeds D, the bound is uninformative for partial observation; full observation still gives exact equality. This is a sufficient condition, not a matching minimax threshold.

Scope of the probability statement. Uniform coordinate sampling is a particular missingness model, not a theorem for every mechanism described as MAR. Group masks and tracking failures need their own conditioning and energy controls. Also, the candidates in Equation (7) must be fixed independently of the masks being analyzed. A union bound over an algorithm's adaptively learned subspaces does not remove this dependence. Deterministic certificates, perturbation control for estimated subspaces, and conditional uniform bounds over continuous model classes address these distinct settings.

4 Temporal Consistency and Assignment Margins

4.1 Adaptive Neighborhoods and the Objective

Temporal persistence becomes useful only after specifying which observations should support each other. On commonly observed coordinates define

$$\Delta_t = \frac{D}{m_{t,t+1}} \|\Phi_t \Phi_{t+1}(\mathbf{x}_{t+1} - \mathbf{x}_t)\|_2^2, \qquad m_{t,t+1} = \operatorname{tr}(\Phi_t \Phi_{t+1}). \tag{8}$$

When the intersection is empty we set $\Delta_t = +\infty$, indicating a potential discontinuity. Local maxima $\widehat{\mathcal{B}}_{\text{cue}} = \{t : \Delta_t > \Delta_{t-1},\ \Delta_t \ge \Delta_{t+1}\}$ partition the sequence; $\mathcal{N}_i$ consists of the other observations between the two adjacent detected boundaries. Endpoints use $\Delta_0 = \Delta_N = -\infty$. Each input sequence is processed separately. These are neighborhood proposals, distinct from the final boundaries $B(z)$. A peak in observed velocity is not automatically an action boundary; an empty overlap marks unavailable evidence, not a certified action switch.

Set $a_{ij} = |\mathcal{N}_i|^{-1}\mathbb{I}\{j \in \mathcal{N}_i\}$, with an all-zero row for an empty neighborhood. Thus $\mathbf{A} = (a_{ij})$ is nonnegative and row-substochastic. With latent action labels z_i, TSCC minimizes

$$\mathcal{J}(z, \mathbf{U}, \mathbf{o}) = \sum_i d_{z_i,i}, \qquad d_{k,i} = \rho_{k,i} + \lambda \sum_j a_{ij}\rho_{k,j}, \qquad \lambda \ge 0, \tag{9}$$

under the orthonormal-basis and canonical-offset constraints. The temporal term tests whether neighboring observations support a candidate model. It does not use neighboring predicted labels as targets, so label updates remain separable once the subspaces and neighborhoods are fixed. The same action model pools evidence from separated bouts, linking their category relation to the final boundaries through Equation (1). This coupling does not make the heuristic proposals themselves identifiable.

4.2 When Temporal Context Helps, and When It Hurts

Let ℓ_i be a reference label and define, for each competitor k,

$$g_{k,i} = \varepsilon_{k,i} - \varepsilon_{\ell_i,i}, \qquad G^{\lambda}_{k,i} = g_{k,i} + \lambda \sum_j a_{ij}(\varepsilon_{k,j} - \varepsilon_{\ell_i,j}).$$

The competitor at neighboring indices is compared with the same reference model ℓ_i, not with the neighbors' potentially different true labels. Write $B^{\lambda}_{k,i} = \varepsilon_{k,i} + \varepsilon_{\ell_i,i} + \lambda \sum_j a_{ij}(\varepsilon_{k,j} + \varepsilon_{\ell_i,j})$. This scale measures how much residual error can affect the comparison.

Theorem 2 (Temporal assignment from residual preservation). *Suppose* $|\rho_{k,j} - \varepsilon_{k,j}| \le \tau\varepsilon_{k,j}$ *for every evaluated pair, where* $0 < \tau < 1$. *All reference assignments are preserved if*

$$G^{\lambda}_{k,i} > \tau B^{\lambda}_{k,i} \quad \textit{for every } i \textit{ and } k \neq \ell_i. \tag{10}$$

For fixed candidates under the uniform sampling model, Equation (7) *gives a probability-*$(1-\delta)$ *event on which this implication holds for every neighborhood matrix satisfying Equation* (10).

A relative-margin interpretation. The effective separation is $\Gamma^{\lambda}_{k,i} = G^{\lambda}_{k,i}/B^{\lambda}_{k,i}$ whenever the denominator is positive. Equation (10) asks for $\Gamma^{\lambda}_{k,i} > \tau$. At $\lambda = 0$ this reduces to the familiar nearest-subspace gap divided by the sum of the two residuals. Asymmetric residual errors require controlling the two candidates separately. For a fixed comparison, let the neighbor gap and scale be $g_{\mathcal{N}}$ and $b_{\mathcal{N}}$, and the self scale be b_i. Then

$$\frac{\partial}{\partial\lambda}\,\frac{g_{k,i} + \lambda g_{\mathcal{N}}}{b_i + \lambda b_{\mathcal{N}}} = \frac{g_{\mathcal{N}} b_i - g_{k,i} b_{\mathcal{N}}}{(b_i + \lambda b_{\mathcal{N}})^2}. \tag{11}$$

Thus context improves relative robustness exactly when its gap-to-scale ratio exceeds that of the current frame. Sharing the same winning label is helpful but alone does not imply an improved relative margin.

Boundary contamination and the admissible weight. Suppose a fraction q_i of the neighborhood lies across a transition. For a competitor, assume the average within-segment gap is at least $g_+ > 0$, while contaminating neighbors favor it by at most $g_- \ge 0$. Then

$$G^{\lambda}_{k,i} \ge g_{k,i} + \lambda[(1 - q_i)g_+ - q_i g_-]. \tag{12}$$

The neighborhood contributes a positive complete-data gap if $q_i < g_+/(g_+ + g_-)$. Above this threshold, a positive self gap can still dominate, but λ must be smaller than $g_{k,i}/[q_i g_- - (1 - q_i)g_+]$. These are clean-data conditions; Equation (10) additionally subtracts the mask-error budget. Increasing temporal weight can therefore stabilize uncertain interior frames and destabilize boundary frames within the same sequence.

4.3 Boundary Recovery and the Limits of the Guarantee

The neighborhood construction introduces its own statistical requirement. Let $H_t = \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2$. If $\max_t |\Delta_t - H_t| \le e$ and every adjacent pair of finite complete-data scores satisfies $|H_t - H_{t-1}| > 2e$, all adjacent orderings and the local-maxima set are preserved. Exact recovery of *semantic* boundaries additionally requires that these maxima coincide with the semantic transitions; smooth within-action dynamics alone does not imply that property.

The temporal cue can need more observations than the residual term. If successive masks are independently sampled at observation fraction p, the expected size of their intersection is Dp^2, whereas a single-frame residual uses Dp coordinates. This explains why boundary localization can become unreliable before per-frame geometry collapses. Conditional on a uniform intersection of size h, standard energy concentration requires h of order $\mu_\Delta \epsilon^{-2} \log(N/\delta)$ for relative score error ϵ; here μ_Δ is the coherence of the consecutive difference vector.

Sensor noise and imperfectly learned subspaces add further errors. If $|\rho_{k,j} - \varepsilon_{k,j}| \le \tau\varepsilon_{k,j} + \beta_{k,j}$, the right side of Equation (10) gains $\beta_{k,i} + \beta_{\ell_i,i} + \lambda \sum_j a_{ij}(\beta_{k,j} + \beta_{\ell_i,j})$. This modular form incorporates model-estimation error into the assignment condition. It provides a conditional route to successful learning: estimated models must enter a sufficiently accurate region, and neighborhoods must retain a positive temporal margin. It does not establish global recovery from arbitrary initialization.

4.4 From Ideal Candidates to Noisy, Estimated Models

The preceding probability statement holds for fixed ideal candidates. To connect it to learned models, we next give a deterministic perturbation argument that can be applied after the data-dependent estimates are formed. Let $(\mathbf{U}_k, \mathbf{o}_k)$ be an ideal candidate and $(\widehat{\mathbf{U}}_k, \widehat{\mathbf{o}}_k)$ an estimated candidate of

the same dimension. Write $\widehat{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{e}_i$. Let $\mathbf{Q}_{k,i}$ and $\widehat{\mathbf{Q}}_{k,i}$ be the orthogonal projectors onto $\mathrm{span}(\Phi_i \mathbf{U}_k)$ and $\mathrm{span}(\Phi_i \widehat{\mathbf{U}}_k)$. Define

$$h_{k,i} = p_i^{-1/2} \left[\|\widehat{\mathbf{Q}}_{k,i} - \mathbf{Q}_{k,i}\|_2 \|\Phi_i(\mathbf{x}_i - \mathbf{o}_k)\|_2 + \|\Phi_i(\mathbf{e}_i + \mathbf{o}_k - \widehat{\mathbf{o}}_k)\|_2 \right]. \tag{13}$$

This radius separates basis error from the combined observation and offset error. All norms involve observed quantities or explicit differences from an ideal model; the bound remains valid for arbitrary dependence between these errors and the mask.

Proposition 1 (Deterministic transfer to estimated models)**.** *Let $\widehat{\rho}_{k,i}$ be the normalized residual of $\widehat{\mathbf{x}}_i$ at the estimated candidate. For every fixed mask,*

$$|\sqrt{\widehat{\rho}_{k,i}} - \sqrt{\rho_{k,i}}| \le h_{k,i}, \qquad |\widehat{\rho}_{k,i} - \rho_{k,i}| \le 2h_{k,i}\sqrt{\rho_{k,i}} + h_{k,i}^2.$$

If $|\rho_{k,i} - \varepsilon_{k,i}| \le \tau\varepsilon_{k,i}$, then

$$|\widehat{\rho}_{k,i} - \varepsilon_{k,i}| \le \tau\varepsilon_{k,i} + \beta_{k,i}, \qquad \beta_{k,i} = 2h_{k,i}\sqrt{(1+\tau)\varepsilon_{k,i}} + h_{k,i}^2. \tag{14}$$

Why restricted conditioning reappears. Suppose both restricted Gram matrices have smallest eigenvalue at least $\kappa p_i > 0$, and let $s_k = \|\mathbf{U}_k\mathbf{U}_k^{\mathrm{T}} - \widehat{\mathbf{U}}_k\widehat{\mathbf{U}}_k^{\mathrm{T}}\|_2$. The observed-space projector discrepancy satisfies

$$\|\widehat{\mathbf{Q}}_{k,i} - \mathbf{Q}_{k,i}\|_2 \le \min\left\{1, \frac{\sqrt{2}s_k}{\sqrt{\kappa p_i}}\right\}. \tag{15}$$

A small complete-space angle can become a large observed-space angle when κp_i is small. A scatter perturbation bound alone is insufficient; it must be combined with restricted conditioning. An eigengap bound controls s_k by scatter error divided by spectral separation. Dimension selection additionally requires separation between the competing eigengaps.

A conditional guarantee for the learned decision. Combining Theorem 2 with Proposition 1 gives the explicit sufficient condition

$$G_{k,i}^{\lambda} > \tau B_{k,i}^{\lambda} + \beta_{k,i} + \beta_{\ell_i,i} + \lambda \sum_j a_{ij}(\beta_{k,j} + \beta_{\ell_i,j}). \tag{16}$$

No new independence assumption on the learned candidates is needed for this deterministic implication. The random-mask probability attaches to the ideal candidates' residual event; a probabilistic claim about the entire algorithm would also have to bound the model errors and the margins of the constructed neighborhoods. These are substantive extra events, rather than consequences of objective descent.

Inverse-rate normalization restores noise scale along with signal scale; it does not remove measurement noise.

4.5 Identifiability, Lower Bounds, and Certifiable Decisions

The sufficient conditions above have complementary obstructions. Consider two complete states $\mathbf{x}^{(a)}$ and $\mathbf{x}^{(b)}$ with different desired labels and a common fixed mask Φ. If their distinguishing coordinates are missing, no inference objective can reconstruct that information from the observed frame alone.

Proposition 2 (An observation-level impossibility condition)**.** *Suppose the observed noise can be any vector of norm at most ν. If*

$$\|\Phi(\mathbf{x}^{(a)} - \mathbf{x}^{(b)})\|_2 \le 2\nu, \tag{17}$$

there are admissible noises under which the two states produce the same observation. Every deterministic label rule errs on at least one state. Every randomized rule has worst-case error probability at least $1/2$.

The noiseless case $\nu = 0$ is an exact mask collision. For instance, if two states differ only in coordinate D, retaining $D-1$ of the D coordinates may still erase all separating information. By contrast, distributing the same squared difference uniformly across coordinates lets every size-m mask

retain an m/D fraction of it. These examples explain why residual coherence enters the observation budget and why a high observation percentage cannot replace a geometric assumption.

Temporal observations can resolve a single-frame ambiguity only if they provide additional information about the label. If an entirely unobserved interval admits two transition locations $t_a < t_b$ with identical observed endpoints and otherwise identical observed data, every boundary estimator has worst-case expected localization error at least $(t_b - t_a)/2$, even with randomization. This lower bound concerns an unconstrained transition location; a duration prior or an additional sensor could change the information available.

A checkable consequence of a valid error budget. Suppose simultaneous residual intervals $[L_{k,i}, U_{k,i}]$ contain the complete residuals. Define temporal endpoints

$$L^{\lambda}_{k,i} = L_{k,i} + \lambda \sum_j a_{ij} L_{k,j}, \qquad U^{\lambda}_{k,i} = U_{k,i} + \lambda \sum_j a_{ij} U_{k,j}.$$

The condition $U^{\lambda}_{\widehat{\ell}_i,i} < \min_{k \neq \widehat{\ell}_i} L^{\lambda}_{k,i}$ certifies the winner of the complete-data cost using that same neighborhood matrix. Under relative error $\tau < 1$, one may take $L_{k,i} = \rho_{k,i}/(1+\tau)$ and $U_{k,i} = \rho_{k,i}/(1-\tau)$. The decision test is observable once a valid τ is supplied, but calibrating τ still requires coherence or a justified error bound. Overlapping intervals indicate insufficient certification, not necessarily a wrong prediction. This distinction provides a principled interpretation of uncertain boundary frames and suggests certified coverage as a useful future evaluation measure. If both endpoints have certified winners, their equality certifies a same-action relation and their inequality certifies a different-action relation, conditional on the candidate models. Adjacent certified unequal winners certify a boundary; uncertified endpoints must not be silently converted into a precise transition. These cost-based certificates describe a restricted candidate family, not the full feasible set $\mathcal{F}$.

5 Safeguarded Optimization

Swapping the order of the neighborhood sum gives the weighted objective

$$\mathcal{J} = \sum_{k,i} w_{k,i} \rho_{k,i}, \qquad w_{k,i} = \mathbb{I}\{z_i = k\} + \lambda \sum_{j : z_j = k} a_{ji}. \tag{18}$$

The incoming weight a_{ji} matters because neighborhood support is generally directed; the identity follows by exchanging the two finite sums. For fixed candidates, $z_i = \arg\min_k d_{k,i}$ exactly minimizes the label block. We retain the current label on ties. For fixed labels, pairwise-observed weighted scatter matrices generate candidate subspaces; an eigengap chooses their dimensions and an orthogonal projection canonicalizes the offsets. Appendix S5 gives all update formulas.

The eigenvector update optimizes a surrogate because the exact projection contains a basis-dependent inverse for each mask. We therefore evaluate the original $\mathcal{J}$ and accept a candidate parameter update only when it does not increase that objective. With neighborhoods fixed after construction, label minimization and safeguarded parameter updates produce a non-increasing, nonnegative objective sequence with convergent values. This is neither global optimality nor a guarantee that the parameters converge to a stationary point: a rejected surrogate proposal may leave a nonstationary iterate.

Residual fitting requires restricted Gram solves; scatter construction can dominate at high dimension. Prefix sums accelerate temporal aggregation and repeated masks permit factorization reuse. Appendix S41 gives time, memory, and stopping-rule details.

6 Experiments and Interpretation

6.1 Protocol

We evaluate TSCC on Keck (Jiang et al., 2012), Weiz (Gorelick et al., 2007), MAD (Huang et al., 2014), UT (Ryoo & Aggarwal, 2009), and Mocap (Carnegie Mellon University Graphics Lab, 2026). The four video datasets supply archived $N \times 324$ descriptor matrices, while Mocap supplies an

archived pose matrix. TSCC trains no encoder and applies masks in these fixed feature spaces. Observed-only standardization is defined exactly in Appendix S8; raw-pixel occlusion is a different problem. Table 5 and Appendix S6 give dataset details, baseline lists, and extended protocols. All methods receive the ground-truth number of action classes K; labels otherwise serve only evaluation. TSCC uses $\lambda = 1$, an eigengap-selected dimension, at most 50 sweeps, and relative objective tolerance 10^{-5}. Complete-input baselines receive common linear interpolation; incomplete methods receive the same masks.

The coordinate-MAR experiment retains a uniformly chosen coordinate subset, matching the sampling model of Theorem 1. Grouped-coordinate deletion, joint loss, and tracking failures are dependent-mask stress tests. We preserve the source tables' MAR/MNAR names for consistency; spatial grouping by itself is not evidence of statistical MNAR. Permutation-matched accuracy (Acc), normalized mutual information (NMI), and adjusted Rand index (ARI) are averaged over five masks and then macro-averaged across datasets. Acc uses permutation matching, whereas NMI and ARI are invariant to label permutations. The stored per-run values support the paired uncertainty analysis reported in Appendix S43.

6.2 What the Coordinate Results Establish

Table 1: Evaluation under incomplete observations. (a) Methods are ordered by publication year; entries are measured five-dataset macro-averages at 30% coordinate missingness. Heterogeneous MAR uses independent but unequal coordinate-observation probabilities. (b) Additional measured missingness regimes. Theorem 1 covers only uniform coordinate MAR; all regimes in (b) require separate assumptions or empirical stress-test interpretation.

(a) Comparative coordinate-level results

Method	Coordinate MAR			Grouped-coordinate MNAR			Heterogeneous MAR		
	Acc	NMI	ARI	Acc	NMI	ARI	Acc	NMI	ARI
SSC-EWZF (Yang et al., 2015)	71.3	62.1	53.6	66.8	57.2	49.1	69.5	60.2	51.7
TSC + interp. (Li et al., 2015)	78.6	70.3	61.4	73.1	65.7	56.8	76.4	68.1	59.0
PZF-SSC (Tsakiris & Vidal, 2018)	73.6	64.5	55.8	68.1	59.3	50.6	71.8	62.6	53.7
PETRELS (Dung et al., 2021)	76.5	67.2	58.4	71.2	62.1	53.5	74.4	65.1	56.1
FSC (Mahmood & Pimentel-Alarcón, 2022)	74.1	65.2	56.3	69.2	60.4	51.7	72.2	63.3	54.3
CDMS + interp. (Zhou et al., 2022)	80.2	72.1	63.5	74.6	67.2	58.3	78.1	69.8	61.1
MISS-DSG (Soni et al., 2023)	75.4	66.7	57.9	70.3	61.8	53.2	73.5	64.6	55.6
CSC (Li & Pimentel-Alarcón, 2026)	76.8	68.1	59.4	71.5	63.2	54.8	74.8	66.1	57.2
TVS-SC + interp. (Xing & Zhao, 2026)	86.2	78.5	69.4	81.1	73.6	65.2	83.8	76.0	67.3
TSCC (Ours)	**89.4**	**82.3**	**73.6**	**84.7**	**77.8**	**69.5**	**87.1**	**80.0**	**71.4**

(b) Additional missingness regimes

Missingness regime	Operational handling	Acc	NMI	ARI
Temporal coordinate bursts, 30%	Partial coordinates missing in contiguous time intervals	83.6	76.5	67.8
Persistent channel dropout, 30%	Fixed sensor-coordinate subset missing for the sequence	82.1	75.2	66.8
Frame MAR, 30%	Entire vectors missing; nearest-time label propagation	77.5	69.8	60.4
Frame MNAR, 30%	Nonoverlapping ten-frame gaps; label propagation	72.8	64.7	55.4
Joint missingness, 1 joint	Mask all coordinates of the selected joint	86.9	79.6	71.2
Joint missingness, 3 joints	Mask all coordinates of the selected joints	82.8	75.1	66.4
Joint missingness, 5 joints	Mask all coordinates of the selected joints	75.1	67.8	58.9
Tracking failure, low	Apply the archived low-failure joint-mask group	88.2	81.0	72.5
Tracking failure, medium	Apply the archived medium-failure joint-mask group	83.1	76.3	67.6
Tracking failure, high	Apply the archived high-failure joint-mask group	77.4	70.5	61.8

At 30% coordinate missingness, TSCC exceeds TVS-SC by 3.2 Acc points under uniform MAR and 3.6 points under grouped-coordinate MNAR. NMI and ARI show the same ranking, while TSCC's 4.7-point MAR-to-MNAR decrease shows that mask structure matters. Grouping can remove concentrated residual energy or degrade restricted conditioning, as Equation (4) predicts qualitatively. The heterogeneous, burst, and persistent-dropout results extend the evaluation beyond uniform coordinate deletion. Per-dataset curves are in Figure 3.

Table 2: Ablation study under 10% coordinate MAR. Dataset columns are permutation-aligned Acc (%); the macro block adds measured Acc/NMI/ARI and the average Acc decrease from full TSCC. Per-dataset NMI/ARI are reported in Table 14.

Variant	Per-dataset Acc					Macro average			Acc drop
	Keck	Weiz	MAD	UT	Mocap	Acc	NMI	ARI	
Full TSCC ($\lambda = 1$)	**86.3**	**81.1**	**85.1**	**83.5**	**82.9**	**83.8**	**76.9**	**68.3**	–
No temporal term ($\lambda = 0$)	76.2	75.4	71.4	69.7	76.1	73.8	65.6	56.0	10.0
w/o masked normalization D/m_i	81.7	77.2	79.5	78.1	77.6	78.8	71.3	62.2	5.0
Fixed temporal window	82.9	78.3	80.6	79.2	78.5	79.9	72.6	63.5	3.9
Interpolation + complete-data UoS	79.4	76.8	76.3	74.5	75.2	76.4	68.5	58.9	7.4

6.3 Ablations, Missing Frames, and Failure Modes

Removing the temporal term reduces accuracy in every dataset in Table 2; the largest decreases are 13.7 points on MAD and 13.8 on UT; a fixed window also reduces accuracy. Measured paired sweeps over λ, K, d_k, and neighborhood radius, with uncertainty intervals and multiplicity-adjusted tests, are reported in Appendix S43. For a fixed frame, D/m_i is common to all candidates and therefore cannot change their ranking; its effect arises during model fitting and temporal aggregation when observation counts differ. Entirely missing frames instead have no geometry, so Table 3 evaluates the explicit nearest-time propagation rule rather than the coordinate-sampling theorem.

Table 3: Measured Keck frame-MAR results; cells are Acc/NMI/ARI (%). "No interpolation" applies nearest-time label propagation only to missing frames. All methods use the same masks.

Method	10%	20%	30%	50%
TSCC, no interpolation	**84.6/77.8/69.1**	**81.2/74.0/65.0**	**77.5/69.8/60.4**	**70.3/62.0/52.5**
TSCC + interpolation	82.1/75.0/66.3	77.8/70.4/61.4	73.6/65.5/56.0	66.1/57.3/47.8
TSC + interp. (Li et al., 2015)	77.3/69.1/59.9	72.5/64.0/54.4	67.8/59.0/49.0	59.2/50.5/41.0
CDMS + interp. (Zhou et al., 2022)	78.9/71.0/61.8	74.6/66.1/56.5	69.7/60.9/50.9	61.4/52.6/43.0

All three metrics decrease monotonically. At 30% missingness, TSCC without interpolation exceeds its interpolated version by 3.9/4.3/4.4 and CDMS by 7.8/8.9/9.5 Acc/NMI/ARI points. From 10% to 50%, its decreases are 14.3/15.8/16.6 points. The results favor using observed geometry and propagating labels only for empty frames. Under tracking failures, accuracy likewise decreases from 88.2% to 77.4% from the low- to high-failure group (Table 8). Both results are consistent with shrinking evidence margins, but neither structured mask is covered by the coordinate-sampling theorem.

Connecting evidence to identifiability. Coordinate masks distort residual comparisons but retain framewise evidence; frame masks remove it and require propagation; joint and tracking masks weaken restricted conditioning. Agreement across Acc, NMI, and ARI rules out gains due only to permutation matching, whereas monotone degradation under heavier missingness shows that context cannot replace evidence absent from a frame and its neighbors. TSCC is most reliable when residual ordering survives and neighborhoods remain pure; otherwise its output should be treated as uncertain rather than uniquely identifiable.

7 Discussion and Conclusion

Identifying temporal action structure from incomplete observations requires preserved residual ordering, uncontaminated temporal support, and a model class excluding alternative structures. TSCC addresses the first two; the set-valued formulation exposes the third. Neither descent nor a sharp boundary establishes identifiability. Experiments support temporal aggregation with supplied K and fixed descriptors, but not unknown-order streaming recovery; side-information gains remain conditional. Motion data require consent and careful handling. The framework separates observation-supported conclusions from those requiring additional assumptions or data.

AI USE STATEMENT

We used generative AI tools solely for language editing and proofreading. We did not use generative AI tools to generate synthetic datasets; develop theoretical or conceptual models; formulate mathematical claims; contribute to proofs; propose hypotheses; design the methodology or experiments; implement the proposed methods; process or analyze the data; or interpret the experimental results. All claims, numerical results, scripts, and tables were produced and verified by the authors using the original source files and experimental outputs. We reviewed all AI-assisted text and take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The information and materials needed to reproduce our experiments are provided in the appendix and the anonymized supplementary material. The experiments use open-weight instruction backbones, fixed random seeds, public benchmarks, and saved JSONL traces containing full and single-component counterfactual interventions. The supplementary repository records the trace schema, generation commands, calibration scripts, policy summaries, and resource footprints. We also provide anonymized source code, trace-processing scripts, task identifiers, and the scripts used to generate the reported summary tables.

REFERENCES

Yue Bai, Lichen Wang, Yunyu Liu, Yu Yin, Hang Di, and Yun Fu. Human motion segmentation via velocity-sensitive dual-side auto-encoder. *IEEE Transactions on Image Processing*, 32:524–536, 2022.

Carnegie Mellon University Graphics Lab. CMU Graphics Lab Motion Capture Database. `http://mocap.cs.cmu.edu/`, 2026. Accessed 2026-07-26.

Nguyen Viet Dung, Nguyen Linh Trung, and Karim Abed-Meraim. Robust subspace tracking with missing data and outliers: Novel algorithm with convergence guarantee. *IEEE Transactions on Signal Processing*, 69:2070–2085, 2021.

Lena Gorelick, Moshe Blank, Eli Shechtman, Michal Irani, and Ronen Basri. Actions as space-time shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2247–2253, 2007.

David Gross and Vincent Nesme. Note on sampling without replacing from a finite collection of matrices. *arXiv preprint arXiv:1001.2738*, 2010.

Dong Huang, Shitong Yao, Yi Wang, and Fernando De La Torre. Sequential max-margin event detectors. In *European Conference on Computer Vision*, pp. 410–424, 2014.

Zhuolin Jiang, Zhe Lin, and Larry Davis. Recognizing human actions by learning and matching shape-motion prototype trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):533–547, 2012.

Huanran Li and Daniel Pimentel-Alarcón. Subspace clustering on incomplete data with self-supervised contrastive learning. *arXiv preprint arXiv:2602.00262*, 2026.

Sheng Li, Kang Li, and Yun Fu. Temporal subspace clustering for human motion segmentation. In *IEEE International Conference on Computer Vision*, pp. 4453–4461, 2015.

Usman Mahmood and Daniel Pimentel-Alarcón. Fusion subspace clustering for incomplete data. In *International Joint Conference on Neural Networks*, pp. 1–8, 2022.

Michael S. Ryoo and J. K. Aggarwal. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *IEEE International Conference on Computer Vision*, pp. 1593–1600, 2009.

Akhilesh Soni, Jeff Linderoth, Jim Luedtke, and Daniel Pimentel-Alarcón. An integer programming approach to subspace clustering with missing data. *arXiv preprint arXiv:2309.15285*, 2023.

Stephen Tierney, Junbin Gao, and Yi Guo. Subspace clustering for sequential data. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1019–1026, 2014.

Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Manolis Tsakiris and René Vidal. Theoretical analysis of sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pp. 4975–4984, 2018.

Guiyu Xia, Huaijiang Sun, Lei Feng, Guoqing Zhang, and Yazhou Liu. Human motion segmentation via robust kernel sparse subspace clustering. *IEEE Transactions on Image Processing*, 27(1): 135–150, 2017.

Zheng Xing and Weibing Zhao. Temporal visual semantics-induced human motion understanding with large language models. *IEEE Transactions on Image Processing*, 35:2184–2195, 2026.

Congyuan Yang, Daniel Robinson, and René Vidal. Sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pp. 2463–2472, 2015.

Tao Zhou, Huazhu Fu, Chen Gong, Ling Shao, Fatih Porikli, Haibin Ling, and Jianbing Shen. Consistency and diversity induced human motion segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):197–210, 2022.

APPENDIX ORGANIZATION

The appendix is organized as follows. We first collect the main-text derivations, then establish notation and give the context, full update formulas, and experimental material supporting the nine-page main paper. We then prove the core residual and assignment results, followed by extensions to structured observations, model perturbations, temporal boundaries, and identifiability. The final sections give additional optimization, complexity, and reproducibility details. More precisely:

- **Main-text derivations and structural identifiability** (Appendices S1 and S2) collect the relocated arguments and formalize boundary bands, category ambiguity, information monotonicity, and the limits of semantic constraints.
- **Notation and standing assumptions** (Appendix S3) define affine-subspace geometry, mask operators, margins, temporal neighborhoods, and the good events used throughout the appendix.
- **Context and full implementation** (Appendices S4–S5) give the method-family comparison, motivating illustrations, complete update formulas, and supporting algebraic statements.
- **Extended experiments** (Appendices S6, S8, S42, and S43) give the archived loader/variable map, exact mask-aware standardization, enlarged per-dataset curves, complete Acc/NMI/ARI tables for coordinate, frame, joint, tracking, and ablation settings, together with the measured paired-inference results and protocol.
- **Core random-mask proofs** prove residual preservation, nearest-subspace preservation, the weighted objective identity, canonical offset representation, local λ-stability, and safeguarded descent.
- **Fixed and structured masks** derive an exact deterministic residual identity and certificate conditions for arbitrary group or block masks.
- **Uniform and heterogeneous sampling** extend pointwise guarantees to all sample–subspace pairs and to independent coordinates with unequal observation probabilities.
- **Noise and estimated-subspace perturbations** quantify how sensor noise, offset error, and basis error enter the normalized residual and the assignment margin.
- **Temporal stability** analyzes preservation of consecutive differences, boundary maxima, neighborhood perturbations, and the global solution path in the semantic weight.
- **End-to-end guarantees** combine mask, noise, subspace, and neighborhood errors into sufficient conditions for one-step and fixed-point label stability.
- **Lower bounds and failure modes** identify information-theoretic obstructions caused by indistinguishable restricted subspaces and fully missing intervals.
- **Optimization, complexity, and empirical audit** refine the descent argument, account for sparse implementation costs, and separate theory-covered experiments from structured stress tests.

S1 MAIN-TEXT DERIVATIONS

This section collects the demonstrations associated with the main text. The random-mask argument is expanded further in Appendix S10; the statements here also clarify which events are deterministic and which depend on a sampling model.

S1.1 CONCENTRATION EVENTS AND THE UNIFORM BUDGET

Uniform sampling concentrates $\mathbf{G}/p$ around $\mathbf{I}_d$, $\|\Phi\mathbf{r}\|_2^2/p$ around $\|\mathbf{r}\|_2^2$, and $\mathbf{c}/p$ around zero. On the corresponding three events,

$$\lambda_{\min}(\mathbf{G}) \geq p(1-\eta_{\mathbf{U}}), \quad |p^{-1}\|\Phi\mathbf{r}\|_2^2 - \|\mathbf{r}\|_2^2| \leq \eta_{\mathbf{r}}\|\mathbf{r}\|_2^2, \quad \|\mathbf{c}\|_2 \leq p\eta_{\mathrm{c}}\|\mathbf{r}\|_2.$$

Substituting these inequalities into Equation (4) gives the theorem. Matrix and scalar concentration supply the observation budget; Appendix S10 gives the details. The upper bound needs only retained-energy control. The lower bound is harder because it must prevent a competing subspace from fitting away the evidence that should reject it.

For Equation (7), set $\eta_{\mathbf{U}} = 1/2$, $\eta_{\mathbf{r}} = \tau/2$, and $\eta_c^2 = \tau/4$. The lower relative error is then at most $\eta_{\mathbf{r}} + \eta_c^2/(1-\eta_{\mathbf{U}}) = \tau$, and the upper relative error is at most $\tau/2$. Assign failure parameter $\delta/(4NK)$ to each invocation of Theorem 1 and use a union bound. Constants are absorbed in C. Zero complete residuals give zero masked residuals directly because the complete point lies on the candidate affine subspace.

S1.2 Temporal Assignment and Neighborhood Contamination

We establish Theorem 2.

Proof. Subtract the two observed costs and bound their residual perturbations:

$$d_{k,i} - d_{\ell_i,i} \geq G^{\lambda}_{k,i} - \tau B^{\lambda}_{k,i} > 0.$$

This proves the deterministic statement; the probability follows from the simultaneous residual event. Because the event holds for all pairs, the inequality also holds for a neighborhood matrix constructed from those observations, but its required margin must then hold for that matrix. □

For Equation (11), apply the quotient rule to $(g_{k,i} + \lambda g_{\mathcal{N}})/(b_i + \lambda b_{\mathcal{N}})$. For Equation (12), split the normalized neighborhood mass into $1 - q_i$ from the same action and q_i from other actions. The two lower bounds on their gaps give $(1 - q_i)g_+ - q_i g_-$. Solving its positivity inequality yields the stated contamination threshold; when it is negative, solving $g_{k,i} + \lambda[(1 - q_i)g_+ - q_i g_-] > 0$ yields the weight restriction. With a non-uniform neighborhood, q_i must denote weight mass rather than the unweighted fraction of indices. The main algorithm uses uniform weights.

S1.3 Noisy Models and Restricted Subspace Angles

We establish Proposition 1.

Proof. The two square-root residuals are the norms of $p_i^{-1/2}(\mathbf{I} - \mathbf{Q}_{k,i})\Phi_i(\mathbf{x}_i - \mathbf{o}_k)$ and its estimated counterpart. Insert the ideal centered observation into the second expression, then use the reverse triangle inequality and $\|\mathbf{I} - \widehat{\mathbf{Q}}_{k,i}\|_2 \leq 1$. This gives the first bound. Factoring the difference of squares gives the second, and $\rho_{k,i} \leq (1 + \tau)\varepsilon_{k,i}$ gives Equation (14). □

Derivation of Equation (15). Orthogonal Procrustes alignment gives an orthogonal matrix $\mathbf{O}$ such that $\|\mathbf{U}_k - \widehat{\mathbf{U}}_k\mathbf{O}\|_2 \leq \sqrt{2}s_k$. Put $\mathbf{A} = \Phi_i\mathbf{U}_k$ and $\widehat{\mathbf{A}} = \Phi_i\widehat{\mathbf{U}}_k\mathbf{O}$. Both have full column rank, with smallest singular value at least $\sqrt{\kappa p_i}$. Hence

$$\|(\mathbf{I} - \widehat{\mathbf{Q}}_{k,i})\mathbf{Q}_{k,i}\|_2 \leq \|(\mathbf{I} - \widehat{\mathbf{Q}}_{k,i})(\mathbf{A} - \widehat{\mathbf{A}})\|_2\|\mathbf{A}^{\dagger}\|_2 \leq \frac{\sqrt{2}s_k}{\sqrt{\kappa p_i}}.$$

For equal-rank orthogonal projectors this norm equals their projector distance. That distance is also at most one, giving the asserted minimum. □

Inserting the residual error bound from Proposition 1 into the cost comparison above gives Equation (16). With only independent mean-zero coordinate noise of variance σ^2, conditional expectation gives $p_i^{-1}\mathbb{E}\|\Phi_i\mathbf{e}_i\|^2 = p_i^{-1}m_i\sigma^2 = D\sigma^2$. This concerns the perturbation radius, not equality for the actual residual perturbation.

S1.4 Observational Collisions and Missing-Interval Localization

We establish Proposition 2.

Proof. Choose the common observation $\mathbf{y} = \Phi(\mathbf{x}^{(a)} + \mathbf{x}^{(b)})/2$. Its displacement from each noiseless observation is at most ν. A deterministic rule returns the same label for $\mathbf{y}$ under both states. For a randomized rule, the probabilities of returning the two different correct labels sum to at most one, so at least one success probability is at most $1/2$. □

For two admissible transition locations $t_a < t_b$ with the same observed data, any randomized estimator has the same output distribution under both explanations. The triangle inequality gives, for each possible output,

$$|\widehat{t} - t_a| + |\widehat{t} - t_b| \geq t_b - t_a.$$

Taking expectation and then the larger of the two risks yields the minimax lower bound $(t_b - t_a)/2$. The admissibility of both completions, including all history and duration constraints, is essential.

For the residual-interval certificate, nonnegative weights imply $L^{\lambda}_{k,i} \leq \varepsilon_{k,i} + \lambda \sum_j a_{ij}\varepsilon_{k,j} \leq U^{\lambda}_{k,i}$. Strict separation of the winner's upper endpoint from every competitor's lower endpoint therefore preserves its complete-data decision. This statement concerns the same candidates and neighborhood matrix throughout.

S2 Set-Valued Action Structure and Information Gain

This section formalizes the identifiability perspective used in the introduction. Its results concern a specified admissible model class, not an assertion that TSCC enumerates that class. They require no new experimental measurements. In particular, side-information results below are conditional statements, not an evaluation of a vision–language model.

S2.1 The Model Class Is Part of the Question

Write an explanation as $\theta = (X, z, \{\mathcal{S}_k\}, \vartheta)$, where ϑ contains any prescribed duration, dynamics, or noise parameters. For example, a piecewise-smooth temporal subspace class could require

$$x_t = o_{z_t} + U_{z_t} a_t + r_t, \qquad \|r_t\| \leq \epsilon_0, \qquad \|a_{t+1} - a_t\| \leq L_a \quad \text{when } z_t = z_{t+1},$$

and impose a minimum bout duration. All bounds, model dimensions, and permitted category counts must be specified before assessing identifiability. This is an illustrative class, not an additional constraint imposed in the reported optimization.

A model allowing every frame to have an unrestricted independent affine subspace admits arbitrary labels with zero fitting error; no missing-data theorem can restore uniqueness in that class. Conversely, specifying all labels as a prior would make uniqueness trivial and uninformative. Meaningful identifiability therefore concerns a stated, nontrivial $\mathfrak{M}(H)$. History H may contain earlier observations or admissible duration constraints; it must not silently contain target labels.

The feasible family in Equation (2) allows different complete trajectories and different nuisance parameters to project to the same structure. Consequently, structural identifiability does not imply unique imputation, unique subspace bases, or unique semantic naming. Conversely, a unique trajectory need not identify an action partition if the model allows that trajectory to be explained by multiple labelings. These are distinct inverse problems.

S2.2 Consensus Relations and Boundary Bands

For a nonempty feasible structural family $\mathcal{F}$, define

$$E_{ij} = \{\mathbb{I}[(i,j) \in \Pi] : (B, \Pi) \in \mathcal{F}\} \subseteq \{0, 1\}.$$

The relation is certified same-action if $E_{ij} = \{1\}$, certified different-action if $E_{ij} = \{0\}$, and ambiguous if $E_{ij} = \{0, 1\}$. This uses no globally meaningful category number. For an anchor r, the lower and upper membership sets

$$C^-_r = \{i : E_{ir} = \{1\}\}, \qquad C^+_r = \{i : 1 \in E_{ir}\}$$

give an invariant category ambiguity representation: $C^+_r \setminus C^-_r$ contains possible but uncertified members of the anchor's action class. These intervals are projections of a jointly constrained family, not independent choices for each pair.

Proposition 3 (Consensus and structural coupling)**.** *Let $\mathcal{F} \neq \varnothing$ contain only structures induced by label sequences as in Equation* (1)*. Then $\Pi^- = \bigcap_{\mathcal{F}} \Pi$ is an equivalence relation. The union $\Pi^+ = \bigcup_{\mathcal{F}} \Pi$ need not be transitive. Moreover,*

$$t \in B^- \iff E_{t,t+1} = \{0\}, \qquad t \notin B^+ \iff E_{t,t+1} = \{1\}.$$

The structure is fully identifiable if and only if all E_{ij} are singletons. Identifiability of B alone does not imply identifiability of Π.

Proof. Each Π is reflexive and symmetric, so their intersection is. If (i, j) and (j, k) belong to every Π, transitivity in each member implies that (i, k) belongs to every member. For a failure of transitivity of the union, consider the partitions $\{\{1, 2\}, \{3\}\}$ and $\{\{1\}, \{2, 3\}\}$: their union relates 1 to 2 and 2 to 3, but not 1 to 3. The two boundary equivalences follow from $t \in B$ if and only if $(t, t+1) \notin \Pi$ in every member. Singleton values for all pairs specify a unique Π, which determines B; the converse follows from the definition of a singleton structural family. Finally, the label sequences (a, a, b, b, a, a) and (a, a, b, b, c, c) have the same boundaries $\{2, 4\}$ but disagree on the equivalence of the first and last bouts. □

This proposition separates two uses of temporal structure. Adjacent relations determine change points, while nonadjacent relations determine whether an earlier action recurs. A boundary-only evaluation cannot detect every error in the latter. The boundary and relation losses below weight both aspects even though their coordinates are logically coupled.

For gradual motion, an uncertainty band is not automatically the physical duration of a transition. It is the set of locations not resolved by the observations and the chosen model. A sharp latent transition can have a wide uncertainty band when frames disappear; a gradual transition can have a narrow band under an imposed labeling convention. The convention and admissible model must be stated to interpret either width.

Proposition 4 (When a boundary band represents one transition)**.** *Suppose every admissible structure contains exactly one boundary in a discrete window W. Let $T \subseteq W$ be the nonempty set of admissible locations, with extremes $t_- = \min T$ and $t_+ = \max T$. For absolute localization loss, every estimator using only the common observation has worst-case expected loss at least $(t_+ - t_-)/2$. If real-valued estimates are permitted, the midpoint attains this bound. For integer estimates the minimax deterministic loss is $\lceil (t_+ - t_-)/2 \rceil$; randomization between the two central integers attains the real-valued bound.*

Proof. Both extremes correspond to admissible explanations for the same observations, so the sum of their risks is at least $t_+ - t_-$ by the triangle inequality. The midpoint is at distance at most half this width from every point in T. For integer outputs, the smallest possible largest distance to the extremes is the ceiling of half the width. When the width is odd, equal randomization between the two central integers makes the expected loss at either extreme half the width, and does no worse at any intervening integer. □

Without the one-boundary condition, a connected component of $B^+ \setminus B^-$ can mix alternatives with different numbers of transitions. Calling it a confidence interval for one matched boundary would then require an additional correspondence assumption.

S2.3 Information Augmentation and Partial-Order Monotonicity

Let additional compatible evidence restrict the explanation family to $\mathcal{A}' \subseteq \mathcal{A}$, with $\varnothing \neq \mathcal{F}' \subseteq \mathcal{F}$ its structural projection. This includes revealing new reliable coordinates while retaining all old constraints. With noisy observations, the new and old noise budgets must be specified so this inclusion actually holds: replacing one global budget with a larger one need not preserve nesting.

Theorem 3 (Monotonicity of compatible structural information)**.** *For nonempty $\mathcal{F}' \subseteq \mathcal{F}$,*

$$B^-(\mathcal{F}) \subseteq B^-(\mathcal{F}') \subseteq B^+(\mathcal{F}') \subseteq B^+(\mathcal{F}), \qquad E_{ij}(\mathcal{F}') \subseteq E_{ij}(\mathcal{F}).$$

Thus a previously certified relation cannot reverse, and unresolved boundary locations cannot increase. For nonnegative weights w_t, v_{ij}, define the structural discrepancy and diameter by

$$d_w(S, S') = \sum_{t=1}^{N-1} w_t |\mathbb{I}[t \in B] - \mathbb{I}[t \in B']| + \sum_{i<j} v_{ij} |\mathbb{I}[(i, j) \in \Pi] - \mathbb{I}[(i, j) \in \Pi']|.$$

$$D_w(\mathcal{F}) = \max_{S, S' \in \mathcal{F}} d_w(S, S').$$

Then $D_w(\mathcal{F}') \le D_w(\mathcal{F})$. Strict reduction holds exactly when $\mathcal{F}'$ contains no pair attaining the old diameter.

Proof. Intersecting over fewer sets enlarges the intersection, while taking a union over fewer sets shrinks the union. Projection of a subset cannot add a previously absent pairwise value. These observations give the inclusions. There are finitely many partitions of N indices, hence finitely many structural outputs even if $\mathcal{A}$ is continuous. The maximum over the smaller Cartesian product cannot be larger. Equality occurs precisely when a retained pair attains the old maximum. □

Merely reducing the number of explanations need not reduce a structural diameter: nuisance parameters can disappear without removing any structural output, and redundant alternative outputs can disappear without removing the farthest pair. These distinctions matter when judging the value of an extra sensor, a longer history, or an external semantic answer.

S2.4 Noisy and Dependent Semantic Side Information

Consider an asynchronous answer C, with a recorded acquisition time and a specified target interval. A semantic compatibility rule defines $\mathcal{K}(C)$ and produces $\mathcal{A}_C = \mathcal{A} \cap \mathcal{K}(C)$. The preceding theorem applies as a set statement whenever the retained family is nonempty. It becomes an epistemically valid refinement only when the true explanation remains in that family.

Proposition 5 (Simultaneous validity without independent answers)**.** *Condition on a fixed observation and history for which $\theta_\star \in \mathcal{A}$. Suppose answers $C_1, \dots, C_L$ have compatibility rules satisfying*

$$\Pr\{\theta_\star \notin \mathcal{K}_\ell(C_\ell) \mid Y, M, H\} \le \alpha_\ell.$$

Then the intersection $\mathcal{A} \cap \bigcap_{\ell=1}^{L} \mathcal{K}_\ell(C_\ell)$ contains the truth with conditional probability at least $1 - \sum_\ell \alpha_\ell$. No independence between answers is needed.

Proof. Truth is removed only if at least one compatibility rule excludes it. The union bound on those events gives the assertion. □

The coverage premises require external calibration appropriate to the observed context; a model's self-reported confidence is not such a guarantee. Answer dependence can also eliminate any gain. For a balanced two-explanation ambiguity $Z \in \{0, 1\}$ conditional on the same geometric observation, let P_0, P_1 be the conditional distributions of C. The optimal conditional decision error is

$$R_C^* = \tfrac{1}{2}\big(1 - \mathrm{TV}(P_0, P_1)\big).$$

Indeed, for a decision region A selecting explanation 0, the error is $\frac{1}{2}[1 - P_0(A) + P_1(A)]$; optimization gives the total-variation formula. Identical conditional answer distributions give zero gain even if the answer sounds semantically specific. Repeating the same answer supplies no additional information conditional on the first. A valid gain calculation must therefore use the joint answer law, not an independence model justified merely by multiple queries.

Proposition 6 (A conditional positive-gain criterion)**.** *Let baseline and side-information decisions have losses L_0, L_C in $[0, L_{\max}]$. Conditional on (Y, M, H), suppose the event V that the semantic constraint retains the truth has probability at least $1 - \alpha$, and $\mathbb{E}[L_0 - L_C \mid V, Y, M, H] \ge g \ge 0$. If consulting and using the answer incurs expected cost $c \ge 0$, then its net expected gain is at least*

$$(1 - \alpha)g - \alpha L_{\max} - c.$$

In particular, strict positivity is certified by $(1 - \alpha)g > \alpha L_{\max} + c$.

Proof. On V^c, boundedness gives $L_0 - L_C \ge -L_{\max}$. Conditioning on V and its complement yields $\Pr(V)g - \Pr(V^c)L_{\max} - c$, which is bounded below as stated. □

This is a sufficient threshold under explicit premises, not a complete positive/zero/negative-gain characterization and not a plug-in score available from uncalibrated language answers. A nonpositive lower bound does not establish negative gain. A false answer can select a unique but incorrect explanation, reducing apparent ambiguity while increasing actual error. For example, if two structures

are geometrically indistinguishable and a wrong hard constraint retains only the false one, the diameter becomes zero while the decision is wrong with probability one. Empty retained families similarly indicate conflict, not successful disambiguation. These examples explain why semantic agreement and geometric compatibility must be distinguished from truth.

S2.5 Separating Structural Change from Evidence Change

The mask is observed evidence about visibility, not a semantic action label. A discontinuity of M can occur with unchanged z, and a change in z can occur with unchanged M. Neither source of change can be identified by the magnitude of one masked difference alone. If two explanations with the same (Y, M, H) differ on whether an action changes, then that transition is structurally unidentifiable in the chosen class. This follows directly from Proposition 3, irrespective of the optimizer.

Residual margins provide one conditional resolution mechanism: surviving coordinates and compatible temporal neighbors can exclude an alternative action model even when the visible evidence changes. The perturbation radius and neighborhood contamination budget quantify when this mechanism fails. However, the local-maxima neighborhood rule itself does not implement a learned dynamics model for evidence outages, nor does it estimate category birth or dormancy. A streaming extension would require an explicit history update, delay model, and admissible model-order evolution before such claims could be assessed.

S3 Notation and Standing Assumptions

For a matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_{\mathrm{F}}$ denote the spectral and Frobenius norms; $\lambda_{\min}(\mathbf{A})$ is the smallest eigenvalue of a symmetric matrix. For a subspace basis $\mathbf{U}$ with orthonormal columns, $\mathbf{P}_{\mathbf{U}} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ and $\mathbf{P}_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{U}}$. The principal-angle distance between equal-dimensional subspaces is

$$d_{\sin}(\mathbf{U}, \widetilde{\mathbf{U}}) := \|\sin\Theta(\mathbf{U}, \widetilde{\mathbf{U}})\|_2 = \|\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\widetilde{\mathbf{U}}}\|_2.$$

Sample i has complete vector $\mathbf{x}_i \in \mathbb{R}^D$, observed mask $\Phi_i = \mathrm{diag}(\phi_{i,1}, \ldots, \phi_{i,D})$, and observed count $m_i = \mathrm{tr}(\Phi_i)$. We write $p_i = m_i/D$. For an affine subspace $\mathcal{S}_k = \mathrm{span}(\mathbf{U}_k) + \mathbf{o}_k$, define

$$\mathbf{y}_{k,i} = \mathbf{x}_i - \mathbf{o}_k, \quad \mathbf{r}_{k,i} = \mathbf{P}_{\mathbf{U}_k}^{\perp}\mathbf{y}_{k,i}, \quad \varepsilon_{k,i} = \|\mathbf{r}_{k,i}\|_2^2, \quad \rho_{k,i} = p_i^{-1}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k).$$

When coordinates are sampled independently with unequal probabilities π_j, we use $\mathbf{P}_\pi = \mathrm{diag}(\pi_1, \ldots, \pi_D)$ and the inverse-probability weight matrix $\mathbf{W}_\Phi = \mathrm{diag}(\phi_1/\pi_1, \ldots, \phi_D/\pi_D)$.

For each sample, let $\ell_i = \arg\min_k \varepsilon_{k,i}$ be its unique complete-data nearest subspace and define the additive margin

$$g_i = \min_{k \neq \ell_i} (\varepsilon_{k,i} - \varepsilon_{\ell_i,i}).$$

For the temporal statistic, set $\mathbf{h}_t = \mathbf{x}_{t+1} - \mathbf{x}_t$, $H_t = \|\mathbf{h}_t\|_2^2$, and let Δ_t denote its normalized partially observed counterpart. The complete boundary set is $\mathcal{B}^\star = \{t : H_t > H_{t-1},\ H_t \geq H_{t+1}\}$, while $\widehat{\mathcal{B}}$ is constructed from Δ_t.

The following events make the proof dependencies explicit. For tolerances $a, b, c \geq 0$, define

$$\mathcal{E}_{\mathrm{G}}(a) = \left\{ \max_{i,k} \|p_i^{-1}\mathbf{U}_k^{\mathrm{T}}\Phi_i\mathbf{U}_k - \mathbf{I}\|_2 \leq a \right\},$$

$$\mathcal{E}_{\mathrm{R}}(b) = \left\{ \max_{i,k} |p_i^{-1}\mathbf{r}_{k,i}^{\mathrm{T}}\Phi_i\mathbf{r}_{k,i} - \|\mathbf{r}_{k,i}\|_2^2| \leq b\|\mathbf{r}_{k,i}\|_2^2 \right\},$$

$$\mathcal{E}_{\mathrm{C}}(c) = \left\{ \max_{i,k} p_i^{-1/2}\|\mathbf{U}_k^{\mathrm{T}}\Phi_i\mathbf{r}_{k,i}\|_2 \leq c\|\mathbf{r}_{k,i}\|_2 \right\}.$$

On their intersection, every normalized residual is controlled simultaneously. Later sections show how random coordinate sampling implies these events and how arbitrary structured masks can instead be audited directly.

Table 4: Positioning relative to the closest method families. "Direct" means that the segmentation objective consumes masks and does not require a separate completion stage. Every acronym used later is cited here and again at its first table occurrence.

Method	Direct incomplete input	Temporal order	Main distinction
SSC-EWZF Yang et al. (2015) / PZF-SSC Tsakiris & Vidal (2018)	Yes	No	Sparse self-expression from partial entries
FSC Mahmood & Pimentel-Alarcón (2022) / MISS-DSG Soni et al. (2023)	Yes	No	Fusion or integer optimization
CSC Li & Pimentel-Alarcón (2026)	Yes	No	Mask-invariant contrastive representation
OSC Tierney et al. (2014) / TSC Li et al. (2015)	No	Yes	Ordered representation or dictionary learning
CDMS Zhou et al. (2022) / VSDA Bai et al. (2022)	No	Yes	Consistency or velocity-sensitive modeling
TVS-SC Xing & Zhao (2026)	No	Yes	Language-assisted temporal semantics
TSCC (ours)	**Yes**	**Yes**	Masked residual plus adaptive temporal neighborhoods

S4 Additional Context and Method Comparisons

The illustrations and family-level comparison below complement the main paper's geometry and temporal-margin analysis.

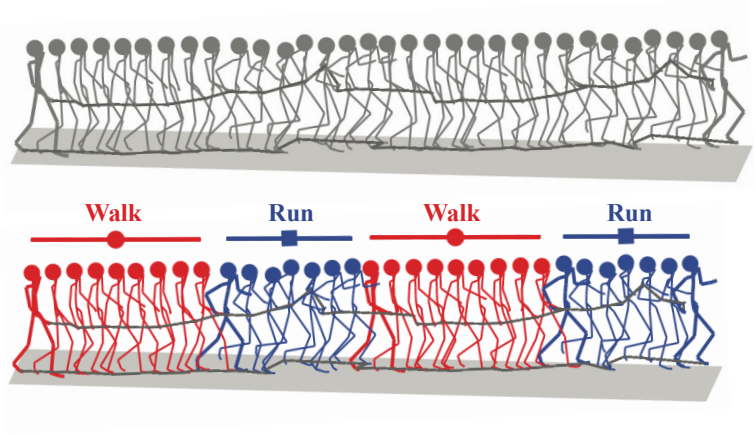


Figure 1: Sequential segmentation task. The upper strip is an unlabeled motion sequence; the lower strip illustrates the desired piecewise-constant latent assignments. Colors denote latent action identities after permutation matching, not supervision available to TSCC. The example highlights the two cues used by our method: within-segment temporal persistence and changes near action boundaries.

S5 Complete Optimization Formulation and Supporting Statements

This section records the weighted formulation, exact masked update, surrogate construction, canonicalization, geometric assignment guarantee, and local sensitivity result supporting the compact main-text algorithm.

S5.1 Objective and Algebraic Identities

Let $\rho_{k,i} = \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k)$ denote the normalized masked residual between $\mathbf{x}_i$ and the k-th affine subspace. If $\mathbf{x}_i$ is assigned to action class k, the assignment cost is $d_{k,i} = \rho_{k,i} + \frac{\lambda}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\rho_{k,j}$. The first term measures the compatibility between the sample and the assigned subspace, while the second term enforces temporal semantic consistency by requiring its neighbors to be compatible with the same subspace. The proposed TSCC objective is

$$\min_{\{\mathcal{C}_k,\mathbf{U}_k,\mathbf{o}_k\}_{k=1}^K} \sum_{k=1}^{K}\sum_{i\in\mathcal{C}_k} d_{k,i}, \qquad \text{s.t. } \mathbf{U}_k^{\mathrm{T}}\mathbf{o}_k = \mathbf{0},\ k = 1,\ldots,K. \tag{19}$$

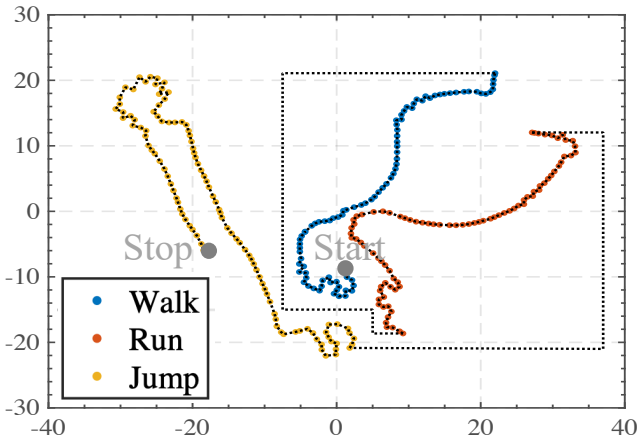


Figure 2: Temporal cue used to construct neighborhoods. Dots are time-ordered observations from three motions, and the dotted path follows acquisition order. Smooth within-motion evolution gives small normalized consecutive differences, whereas the marked transitions give larger differences. The illustration is explanatory; no ground-truth labels are used to construct neighborhoods.

Although (9) is non-convex due to the joint optimization over latent action assignments and subspace parameters, it admits an equivalent weighted form that leads to an efficient alternating optimization algorithm.

Theorem 4 (Equivalent weighted form)**.** *For a fixed action class $\mathcal{C}_k$, the objective term $\sum_{i\in\mathcal{C}_k} d_{k,i}$ can be equivalently written as $\sum_{i=1}^{N} \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k)w_{k,i}$, where $w_{k,i} = \mathbb{I}(i \in \mathcal{C}_k) + \lambda n_k(i)$ and $n_k(i) = \sum_{j\in\mathcal{C}_k} \frac{1}{|\mathcal{N}_j|}\mathbb{I}(i \in \mathcal{N}_j)$ counts how often $\mathbf{x}_i$ appears as a temporal neighbor of samples assigned to the k-th action class. Here $\mathbb{I}(\cdot)$ is the indicator function.*

Based on Theorem 4, problem (9) can be rewritten as

$$\min_{\{\mathcal{C}_k, \mathbf{U}_k, \mathbf{o}_k\}_{k=1}^{K}} \sum_{k=1}^{K}\sum_{i=1}^{N} \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k)w_{k,i}, \qquad \text{s.t. } \mathbf{U}_k^{\mathrm{T}}\mathbf{o}_k = \mathbf{0},\ k = 1, \ldots, K. \tag{20}$$

This reformulation shows that temporal semantic consistency can be interpreted as an adaptive sample-weighting mechanism: a sample receives a larger weight for a subspace if it is either assigned to that subspace or frequently appears as a temporal neighbor of samples assigned to that subspace.

Proposition 7 (Local sensitivity to temporal weight)**.** *For fixed subspaces and neighborhoods, write $d_{k,i}(\lambda) = a_{k,i} + \lambda b_{k,i}$, where $a_{k,i} = \rho_{k,i}$ and $b_{k,i} = |\mathcal{N}_i|^{-1}\sum_{j\in\mathcal{N}_i}\rho_{k,j}$. Suppose ℓ_i is the unique minimizing label at λ_0, with margin $g_i = \min_{k\neq\ell_i}\{d_{k,i}(\lambda_0) - d_{\ell_i,i}(\lambda_0)\} > 0$, and let $M_i = \max_{k\neq\ell_i}|b_{k,i} - b_{\ell_i,i}|$. Then ℓ_i remains the minimizing label whenever $|\lambda - \lambda_0| < g_i/M_i$; if $M_i = 0$, it remains minimizing for every $\lambda \geq 0$.*

Proposition 7 makes the relevant sensitivity explicit: assignments are stable over an interval whose width is determined by the observed assignment margin.

S5.2 Optimization Algorithm

Problem (18) involves two coupled variable blocks: the latent action assignments $\{\mathcal{C}_k\}_{k=1}^{K}$ and the subspace parameters $\{(\mathbf{U}_k, \mathbf{o}_k)\}_{k=1}^{K}$. We therefore optimize them alternately.

S5.2.1 Updating the Subspace Parameters

We first update $\mathbf{U}_k$ and $\mathbf{o}_k$ with fixed assignments and fixed weights $w_{k,i}$. The exact minimization of the masked residual objective with respect to $\mathbf{U}_k$ is difficult because the projection matrix contains the mask-dependent term $(\mathbf{U}_k^{\mathrm{T}}\Phi_i\mathbf{U}_k)^{\dagger}$, which changes with both the sample mask and the basis. Therefore, rather than claiming an exact block minimization of the original objective, we use a

mask-normalized surrogate descent step. For the offset, we first minimize the diagonal quadratic surrogate $\widetilde{J}_k^{\text{off}}(\mathbf{o}) = \sum_{i=1}^N \frac{w_{k,i}}{m_i} \|\Phi_i(\mathbf{x}_i - \mathbf{o})\|_2^2$. The resulting intermediate offset $\mathbf{o}'_k$ has the j-th entry $o'_{k,j} = \frac{\sum_{i=1}^N \frac{w_{k,i}}{m_i}\phi_{i,j}x_{i,j}}{\sum_{i=1}^N \frac{w_{k,i}}{m_i}\phi_{i,j}}$. If the denominator is zero for a certain coordinate, that coordinate is ignored in the update or assigned a default value from the previous iteration. Define the mean-shifted sample $\bar{\mathbf{x}}_{k,i} = \frac{\mathbf{x}_i - \mathbf{o}'_k}{\sqrt{m_i}} = [\bar{x}_{k,i}^{(1)}, \dots, \bar{x}_{k,i}^{(D)}]^{\mathrm{T}}$. The exact masked basis update would require solving

$$\underset{\mathbf{U}_k \in \mathbb{R}^{D\times d_k}}{\text{minimize}} \sum_{i=1}^N w_{k,i} \left\| \left(\mathbf{I}_D - \Phi_i \mathbf{U}_k (\mathbf{U}_k^{\mathrm{T}} \Phi_i \mathbf{U}_k)^{\dagger} \mathbf{U}_k^{\mathrm{T}} \Phi_i\right) \Phi_i \bar{\mathbf{x}}_{k,i} \right\|_2^2. \tag{21}$$

Since (21) is non-convex and mask-dependent, we use a pairwise-observed weighted scatter matrix $\mathbf{S}_k$ as a surrogate, with $(\mathbf{S}_k)_{a,b} = \frac{\sum_{i=1}^N w_{k,i}\phi_{i,a}\phi_{i,b}\bar{x}_{k,i}^{(a)}\bar{x}_{k,i}^{(b)}}{\sum_{i=1}^N w_{k,i}\phi_{i,a}\phi_{i,b}}$, and set the entry to zero when the denominator vanishes. The candidate basis is given by the d_k leading eigenvectors of the symmetric matrix $\mathbf{S}_k$. When $\Phi_i = \mathbf{I}_D$, this reduces to standard weighted PCA. If d_k is unknown, we estimate it by the largest eigengap, i.e., $\hat{d}_k = \arg\max_{i\in\{1,\dots,D-1\}}(\lambda_{k,i} - \lambda_{k,i+1})$, where $\lambda_{k,1} \geq \cdots \geq \lambda_{k,D}$ are the eigenvalues of $\mathbf{S}_k$. We project the candidate offset onto the orthogonal complement of the learned basis. Finally, we evaluate the original action-model contribution $J_k = \sum_i (D/m_i)\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k) w_{k,i}$: the candidate pair is accepted only if it does not increase J_k; otherwise the previous pair is retained. This safeguard is essential because the eigenvector step optimizes a surrogate, not the original masked objective.

Proposition 8 (Canonical affine offset)**.** *Let $\mathbf{U}_k$ have full column rank and let $\mathbf{o}'_k$ be any affine offset. Then $\mathbf{o}_k = \left(\mathbf{I}_D - \mathbf{U}_k(\mathbf{U}_k^{\mathrm{T}}\mathbf{U}_k)^{-1}\mathbf{U}_k^{\mathrm{T}}\right)\mathbf{o}'_k$ satisfies $\mathbf{U}_k^{\mathrm{T}}\mathbf{o}_k = \mathbf{0}$ and represents the same affine set: $\mathrm{span}(\mathbf{U}_k) + \mathbf{o}'_k = \mathrm{span}(\mathbf{U}_k) + \mathbf{o}_k$.*

Thus, after updating $\mathbf{U}_k$, the valid affine offset $\mathbf{o}_k$ can be obtained by the above projection.

S5.2.2 Updating the Latent Action Assignments

With fixed subspace parameters $\{(\mathbf{U}_k, \mathbf{o}_k)\}_{k=1}^K$, the assignment cost for sample $\mathbf{x}_i$ to action class k is

$$d_{k,i} = \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k) + \frac{\lambda}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\frac{D}{m_j}\xi(\mathbf{x}_j, \Phi_j, \mathbf{U}_k, \mathbf{o}_k). \tag{22}$$

The latent action label is then updated by $z_i = \arg\min_{k\in\{1,\dots,K\}} d_{k,i}$. After all labels are updated, the action membership sets are defined as $\mathcal{C}_k = \{i : z_i = k\}$, and the weights $w_{k,i}$ are recomputed according to Theorem 4. The following theorem states that the masked residual preserves the complete-data nearest-subspace assignment when the complete-data residual has a sufficient margin.

Theorem 5 (Assignment consistency under missing observations)**.** *Given a complete data point $\mathbf{x}$ and K affine subspaces, define $\varepsilon_k = \varepsilon(\mathbf{x}, \mathbf{U}_k, \mathbf{o}_k)$ and $\xi_k = \xi(\mathbf{x}, \Phi, \mathbf{U}_k, \mathbf{o}_k)$. Let $\ell = \arg\min_{k\in\{1,\dots,K\}} \varepsilon_k$ be the nearest subspace under complete observations. Suppose that there exists a margin $\gamma \in (0,1)$ such that $\varepsilon_\ell \leq (1-\gamma)\varepsilon_k, \quad \forall k \neq \ell$. Assume that the conditions of Theorem 1 hold simultaneously for all K subspaces with parameters $\eta_{\mathbf{U}}, \eta_{\mathbf{r}}, \eta_{\mathrm{c}}$. Let $\alpha_- = \eta_{\mathbf{r}} + \frac{\eta_{\mathrm{c}}^2}{1-\eta_{\mathbf{U}}}$ and $\alpha_+ = \eta_{\mathbf{r}}$. If $\gamma > \frac{\alpha_- + \alpha_+}{1+\alpha_+}$ and if*

$$m \geq c_0 \max_{k\in\{1,\dots,K\}} \left\{ \frac{\mu_{\mathbf{U}_k} d_k \log(2Kd_k/\delta)}{\eta_{\mathbf{U}}^2}, \frac{\mu_{\mathbf{r}_k}\log(2K/\delta)}{\eta_{\mathbf{r}}^2}, \frac{\mu_{\mathbf{U}_k}\mu_{\mathbf{r}_k} d_k \log(2Kd_k/\delta)}{\eta_{\mathrm{c}}^2} \right\}, \tag{23}$$

where $\mathbf{r}_k = \mathbf{P}_{k,\perp}(\mathbf{x} - \mathbf{o}_k)$, then, with probability at least $1 - 4K\delta$, $\ell = \arg\min_{k\in\{1,\dots,K\}} \xi_k$.

Theorem 5 concerns the geometric masked-residual term, not the complete temporally regularized objective. It states that the nearest-subspace component is reliable when the complete-data residual has margin γ and enough coordinates are observed relative to coherence, dimension, K, and failure probability. Proposition 7 separately characterizes how the temporal term changes the final label.

Proposition 9 (Convergence and complexity)**.** *With fixed neighborhoods, TSCC generates a non-increasing objective sequence that converges to a finite limit. For generic masks, a direct implementation with truncated eigensolvers has per-sweep cost $O(KN(\bar{m}d_{\max}^2 + d_{\max}^3) + K\sum_i m_i^2 + KD^2 d_{\max})$. Processing one dense scatter matrix at a time requires memory $O(ND + KDd_{\max} + KN + D^2)$. Factorization reuse for repeated masks can reduce the residual-evaluation cost.*

Table 5: Datasets and archived inputs supplied to TSCC. The method starts from a fixed frame-level matrix and does not train or fine-tune a visual encoder.

Dataset	Modality	#Sequences / subjects	#Classes	Archived TSCC input
Weiz Gorelick et al. (2007)	Silhouette video	90 / 9	10	$N \times 324$ archived descriptor
MAD Huang et al. (2014)	RGB-D / Kinect	35 / 20	8	$N \times 324$ archived descriptor
Keck Jiang et al. (2012)	RGB gesture video	114 / 3	14	$N \times 324$ archived descriptor
UT Ryoo & Aggarwal (2009)	RGB interaction video	20 / –	6	$N \times 324$ archived descriptor
Mocap Carnegie Mellon University Graphics Lab (2026)	3D skeleton	5 selected subjects	5–12	$N \times D_{\text{pose}}$ archived pose matrix

S6 Extended Experimental Protocol and Results

The main paper retains the coordinate-level comparison and the ablation table. This section supplies dataset descriptions, disaggregated curves, frame and tracking comparisons, qualitative examples, and runtime details.

S6.1 Experimental Protocol

Representation and mask location. TSCC consumes the fixed matrices summarized in Table 5; no embedding or principal-component projection is learned before segmentation. Coordinate masks are applied *after* loading these vectors. Consequently, the theorem concerns missing feature coordinates, not raw-pixel occlusion before a nonlinear encoder. Appendix S8 gives the exact file-variable map, matrix orientation, standardization equation, and the boundary of what can be reconstructed from the archive.

Missingness. For coordinate MAR, each frame retains a uniformly sampled subset of coordinates, matching Theorem 1. Heterogeneous MAR samples coordinates independently with fixed unequal probabilities and is governed by Appendix S18. Grouped-coordinate MNAR removes contiguous coordinate groups. Temporal coordinate bursts mask a coordinate group for contiguous time intervals, whereas persistent channel dropout fixes one missing coordinate subset for the entire sequence. These structured masks use deterministic certificates, not Theorem 1. Frame MAR removes isolated indices; frame MNAR removes ten-frame clips. Because a fully missing frame has $m_i = 0$, interpolation-free TSCC labels observed frames and propagates the nearest observed label to a missing index (ties use the earlier index). Joint and tracking-failure masks are also structured stress tests.

Model selection and fairness. The number of latent action classes K is set to the ground-truth action count, disclosed because this is privileged information; every baseline receives the same K. Ground-truth labels are otherwise used only for evaluation. Subspace dimensions are never taken from labels and are estimated by the eigengap rule in Sec. 5. TSCC uses $\lambda = 1$, $T_{\max} = 50$, and relative-objective tolerance 10^{-5} in all reported settings. Complete-input temporal baselines receive the same linear interpolation, whereas incomplete methods receive the same masks.

Baselines and metrics. We retain the incomplete-data and temporal baselines expanded at first use in Sec. 2. We report permutation-matched accuracy (Acc), normalized mutual information (NMI), and adjusted Rand index (ARI). Each entry is the mean over five independently generated masks; label indices are matched by the Hungarian algorithm only for Acc. A cross-dataset number is a macro-average of the five per-dataset metrics. Figures 3–4 show the enlarged per-dataset Acc curves, while Table 6 supplies the corresponding per-dataset Acc/NMI/ARI record.

S6.2 Coordinate-Level Missingness

Table 1 reports a macro-average; Figures 3–4 and Table 6 prevent it from hiding dataset heterogeneity. TSCC reaches 89.4% Acc under MAR and 84.7% under grouped-coordinate MNAR. The 3.2- and 3.6-point differences from TVS-SC are descriptive, not significance claims.

S6.3 Joint-Level Missingness

We further evaluate TSCC on Mocap by masking every position and angle coordinate associated with a selected body joint. This is a structured, dependent mask and is therefore an empirical stress test rather than evidence for Theorem 1.

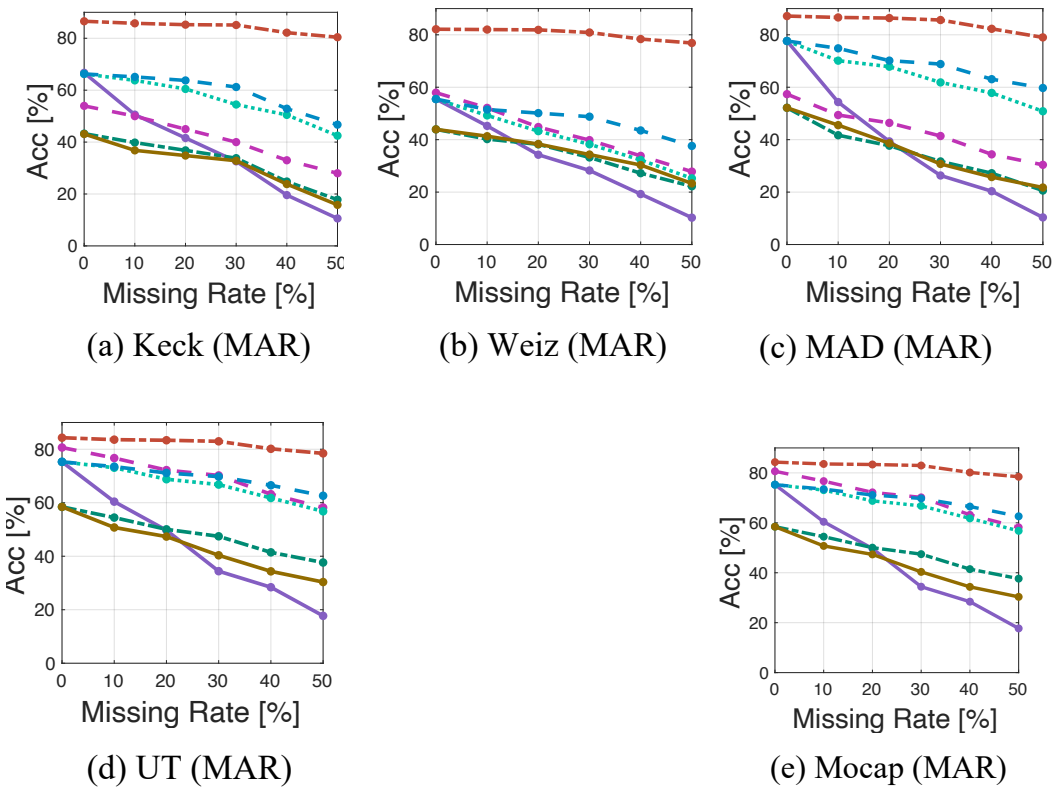
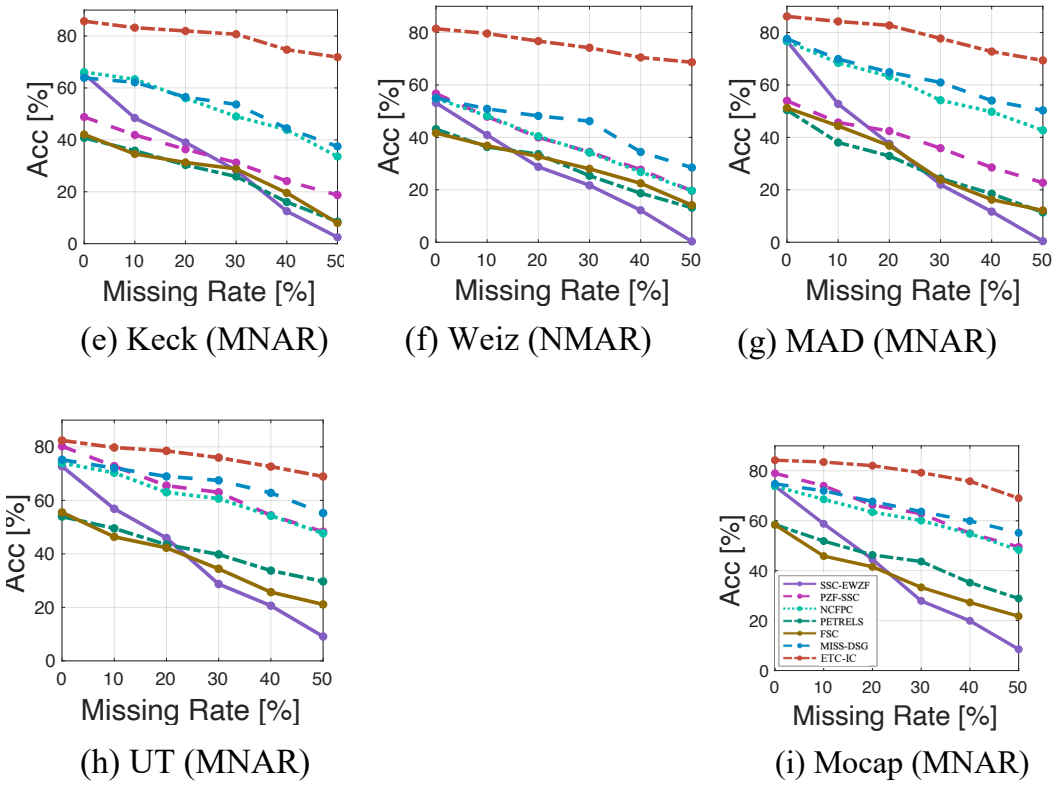


Figure 3: Enlarged per-dataset Acc curves under coordinate MAR. Splitting the former ten-panel montage enlarges every panel by approximately 55% while retaining vector text and lines. The archived plotting legend uses ETC-IC for the proposed method (renamed TSCC here) and NCFRC for the nonconvex-fusion baseline.

Table 6: Measured per-dataset results for the coordinate-level regimes in Table 1. Each cell is Acc/NMI/ARI (%), evaluated under the unified paired protocol.

Mask	Method	Keck	Weiz	MAD	UT	Mocap
MAR	TSCC	92.1/85.0/77.8	90.8/83.8/75.9	88.0/80.1/71.1	87.6/80.7/71.3	88.5/81.9/71.9
	TVS-SC + interp.	88.5/81.4/72.8	87.6/80.0/71.5	84.8/76.0/67.2	84.3/77.1/67.5	85.8/78.0/68.0
MNAR	TSCC	87.5/80.7/72.7	86.0/79.4/71.5	82.8/75.1/66.8	82.9/76.1/67.5	84.3/77.7/69.0
	TVS-SC + interp.	83.8/76.7/68.4	82.1/75.1/67.3	79.3/70.8/62.5	79.2/71.9/63.1	81.1/73.5/64.7
Heterogeneous MAR	TSCC	89.5/82.7/74.5	88.2/81.4/73.1	85.3/77.7/68.8	85.0/78.4/69.1	87.5/79.8/71.5
	TVS-SC + interp.	86.1/79.0/70.5	84.8/77.2/68.8	82.0/73.4/64.8	81.7/74.2/65.4	84.4/76.2/67.0
Temporal bursts	TSCC	86.5/79.4/71.0	84.9/77.8/69.2	81.6/73.8/64.8	81.9/74.9/65.8	83.1/76.6/68.2
Persistent dropout	TSCC	85.0/78.1/69.8	83.4/76.6/68.2	80.2/72.5/63.8	80.0/73.5/64.8	81.9/75.3/67.4



(e) Keck (MNAR) (f) Weiz (NMAR) (g) MAD (MNAR)

(h) UT (MNAR) (i) Mocap (MNAR)

Figure 4: Enlarged per-dataset Acc curves under grouped-coordinate MNAR. The legend appears in the Mocap panel and applies to both figures; abbreviation mapping is stated in Fig. 3. This structured mask is outside the random-coordinate theorem.

Table 7: Measured Mocap joint-missingness Acc/NMI/ARI results (%) under the common evaluation protocol.

Method	Missing joints	Acc	NMI	ARI
TSCC	1	86.9	79.6	71.2
	3	82.8	75.1	66.4
	5	75.1	67.8	58.9
TVS-SC + interpolation	1	83.0	75.3	66.4
	3	78.4	70.6	61.5
	5	70.8	63.0	54.2

Table 8: Mocap performance under the three tracking-failure mask groups, ordered by average missing-joint ratio. These dependent masks are outside the random-coordinate theorem.

Method	Low failure			Medium failure			High failure		
	Acc	NMI	ARI	Acc	NMI	ARI	Acc	NMI	ARI
SSC-EWZF Yang et al. (2015)	72.4	64.1	55.3	67.9	59.8	50.6	62.3	54.1	45.2
PZF-SSC Tsakiris & Vidal (2018)	74.1	65.7	56.9	69.2	61.1	51.8	63.5	55.3	46.3
PETRELS Dung et al. (2021)	76.2	68.1	58.7	70.8	62.9	53.4	64.5	56.4	47.3
FSC Mahmood & Pimentel-Alarcón (2022)	74.6	66.2	57.4	69.7	61.6	52.3	64.0	55.8	46.8
MISS-DSG Soni et al. (2023)	75.8	67.3	58.6	70.9	62.7	53.4	65.1	56.9	47.9
CSC Li & Pimentel-Alarcón (2026)	76.5	67.8	59.3	71.3	63.1	54.0	65.4	57.3	48.2
OSC + interp. Tierney et al. (2014)	77.4	69.4	60.5	71.9	64.6	55.6	66.0	58.4	49.8
TSC + interp. Li et al. (2015)	78.2	70.1	61.7	72.6	65.3	56.4	66.8	59.1	50.6
CDMS + interp. Zhou et al. (2022)	79.5	71.3	62.9	73.8	66.4	57.5	67.9	60.2	51.7
VSDA + interp. Bai et al. (2022)	80.2	72.1	63.8	74.5	67.2	58.3	68.6	61.0	52.5
TVS-SC + interp. Xing & Zhao (2026)	84.8	77.2	69.1	78.5	71.4	63.0	72.1	64.8	56.7
TSCC (Ours)	**88.2**	**81.0**	**72.5**	**83.1**	**76.3**	**67.6**	**77.4**	**70.5**	**61.8**

S6.4 Realistic Tracking-Failure Missingness

To evaluate structured corruption, we group the available tracking-failure masks by average missing-joint ratio. As shown in Table 8, TSCC reports 88.2% and 77.4% Acc in the low- and high-failure groups. We treat this as empirical evidence under dependent masks, not as validation of the random-coordinate probability bound.

S6.5 Qualitative Analysis

Fig. 5(a) visualizes Keck segmentation results. In this example, PETRELS Dung et al. (2021), FSC Mahmood & Pimentel-Alarcón (2022), and MISS-DSG Soni et al. (2023) produce fragmented temporal segments, whereas TSCC yields smoother segments that more closely align with the ground truth. This example is consistent with the aggregate advantage of adaptive temporal semantic consistency, but it is not a substitute for a paired statistical test.

Fig. 5(b) shows representative Weiz frames, where action boundaries are visually ambiguous. As shown in Fig. 5(c), TSCC errors mainly occur near subtle transitions, such as between "jumping jack" and "jump". These localized boundary errors are understandable and indicate that TSCC preserves global segment-level semantic coherence rather than producing scattered misclassifications.

S6.6 Convergence and Runtime Analysis

Fig. 6(a) shows the convergence behavior of TSCC on the Weiz dataset. The objective decreases rapidly in the first few iterations and stabilizes within about ten iterations. Fig. 6(b) further visualizes the evolution of latent action assignments, where large changes mainly occur at early iterations and the algorithm subsequently refines temporal boundaries.

Across the five datasets, TSCC converges within a small number of iterations: 9, 10, 11, 10, and 13 iterations on Keck, Weiz, MAD, UT, and Mocap, respectively. Using an unoptimized MATLAB

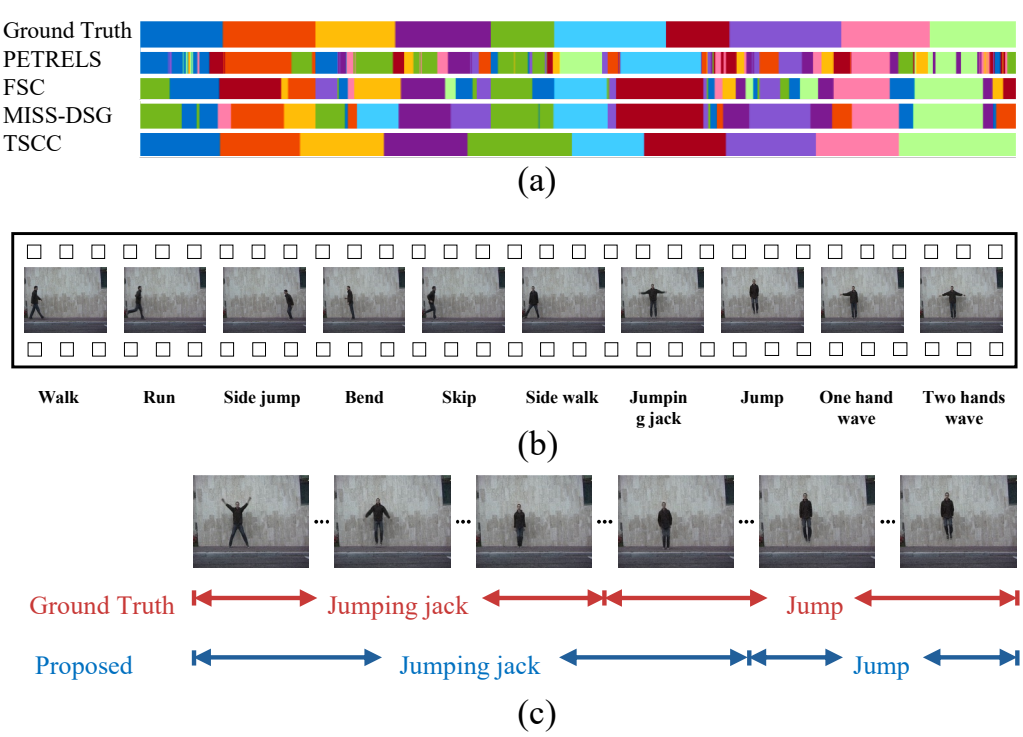


Figure 5: Qualitative evidence and failure analysis. (a) Keck frame assignments after label permutation: incomplete baselines fragment long segments, whereas TSCC is temporally coherent. (b) Ten Weiz actions used in the evaluation. (c) A boundary error between "jumping jack" and "jump"; arrows show ground-truth and predicted segment extents. Colors identify action classes only after evaluation matching.

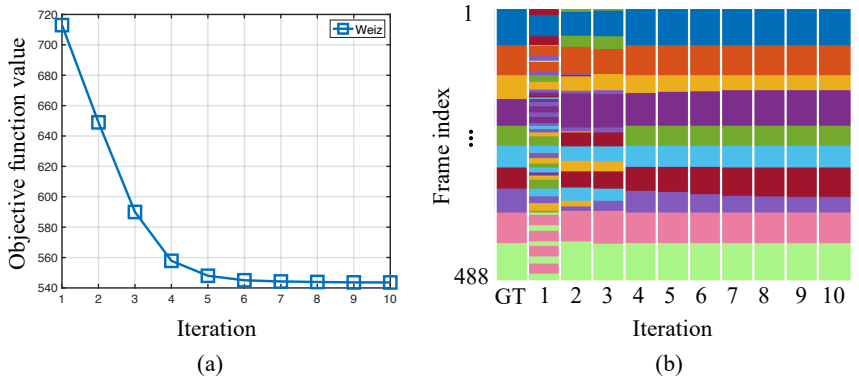


Figure 6: Optimization diagnostic on Weiz. (a) Original safeguarded objective J versus accepted iteration; monotonicity is enforced by the acceptance test. (b) Assignment evolution; colors denote action classes and GT is used only for post-hoc comparison. Most changes occur early, followed by boundary refinement.

implementation on a single CPU, the average runtime per sequence is 0.52 s, 0.68 s, 0.83 s, 0.71 s, and 1.26 s, respectively. The final objective is also stable across five trials, with standard deviations 0.018, 0.021, 0.024, 0.022, and 0.030. These results support the monotonic descent analysis in Proposition 9.

S6.7 Ablation Study

Table 2 reports the ablation results under 10% MAR missingness. Removing temporal semantic consistency leads to a performance drop on all five reported dataset aggregates, supporting the usefulness of temporal neighborhoods for temporal segmentation. The difference is especially pronounced on MAD and UT, where the full model exceeds the non-temporal variant by more than ten percentage points. These descriptive ablations isolate the contribution of the temporal term, although uncertainty estimates are still required for inferential claims. Normalization affects weighting across observations and temporal aggregation, but not the nearest-subspace ranking of one frame at fixed candidates. If all observation counts are equal, removing D/m_i merely rescales the objective. The normalization row therefore requires checking the actual count variation, initialization, and stopping implementation before attributing the difference to rate correction. The fixed-window variant tests the adaptive neighborhood construction, and the imputation-based variant compares direct masked fitting with completing the data first. Complete per-dataset Acc/NMI/ARI results are reported in Table 14, while run-level uncertainty and paired comparisons are reported in Appendix S43.

S6.8 Sensitivity and Scope

TSCC has no hand-selected neighborhood width and does not receive the true subspace dimensions; these are produced by local maxima of Δ_t and the eigengap rule, respectively. The remaining semantic weight is λ. Table 2 reports the two most interpretable endpoints, $\lambda = 0$ and the fixed default $\lambda = 1$, while Proposition 7 gives the exact label-stability interval around any chosen value. The reported benchmark protocol does use the ground-truth K, equally for every method. We therefore make no claim that TSCC is insensitive to a misspecified number of latent action classes; estimating K jointly with the segmentation is outside the present scope.

S7 Algorithmic Procedure

This section summarizes the computational procedure of TSCC. The algorithm first measures temporal changes between adjacent incomplete samples using only their commonly observed entries. Specifically, if $m_{t,t+1} > 0$, it computes $\Delta_t = \frac{D}{m_{t,t+1}} \|\Phi_t \Phi_{t+1}(\mathbf{x}_{t+1} - \mathbf{x}_t)\|_2^2$; otherwise, it sets $\Delta_t = +\infty$, treating the position as a potential temporal discontinuity. Candidate transition points are then detected from local maxima of Δ_t, and each sample $\mathbf{x}_i$ is assigned an adaptive temporal neighborhood $\mathcal{N}_i$. This avoids manually specifying a fixed temporal window and allows the neighborhood length to adapt to the local motion duration.

After initialization, TSCC alternates between subspace-parameter updates and label updates. Given current action classes, the temporal semantic weights $w_{k,i}$ are computed according to whether a sample belongs to action class k or frequently appears in the temporal neighborhoods of samples assigned to action class k. For each subspace, the algorithm estimates an intermediate offset $\mathbf{o}'_k$, constructs a pairwise-observed weighted scatter matrix $\mathbf{S}_k$, estimates the intrinsic dimension $\hat{d}_k$ by the largest eigengap, and updates $\mathbf{U}_k$ using the leading eigenvectors of $\mathbf{S}_k$. The affine offset is then projected as $\mathbf{o}_k = (\mathbf{I}_D - \mathbf{U}_k(\mathbf{U}_k^{\mathrm{T}}\mathbf{U}_k)^{-1}\mathbf{U}_k^{\mathrm{T}})\mathbf{o}'_k$ to satisfy the orthogonality constraint.

In the label-update step, each sample is assigned to the subspace with the smallest temporal-semantic assignment cost, which consists of its own normalized masked residual and the average residual of its adaptive temporal neighbors. The action membership sets are then updated accordingly. The procedure terminates when the relative change of the objective is below η or when the maximum number of iterations $T_{\max}$ is reached.

S8 Additional Experimental Details

S8.1 Feature-Matrix Provenance and Exact Preprocessing

The evaluation starts at the archived matrix boundary, not at raw pixels. Inspection of the distributed MATLAB loaders and variables gives the exact map in Table 9. Each stored array is transposed once so that rows are frames and columns are coordinates. The loaders contain no PCA, learned encoder, whitening, or label-dependent transformation.

Algorithm 1 TSCC under Incomplete Observations

Input: Incomplete sequence $\{\mathbf{x}_i, \Phi_i\}_{i=1}^N$, number of latent action classes K, temporal weight λ, maximum iterations $T_{\max}$, tolerance η.
Compute temporal differences $\Delta_t = \frac{D}{m_{t,t+1}} \|\Phi_t \Phi_{t+1}(\mathbf{x}_{t+1} - \mathbf{x}_t)\|_2^2$ if $m_{t,t+1} > 0$, and $\Delta_t = +\infty$ otherwise, where $m_{t,t+1} = \mathrm{tr}(\Phi_t \Phi_{t+1})$.
Detect neighborhood proposals $\widehat{\mathcal{B}}_{\mathrm{cue}} = \{t : \Delta_t > \Delta_{t-1},\ \Delta_t \geq \Delta_{t+1}\}$, with $\Delta_0 = \Delta_N = -\infty$.
Construct adaptive neighborhoods $\mathcal{N}_i = \{j : a_i \leq j \leq b_i,\ j \neq i\}$, where $a_i = 1 + \max(\{t \in \widehat{\mathcal{B}}_{\mathrm{cue}} : t < i\} \cup \{0\})$ and $b_i = \min(\{t \in \widehat{\mathcal{B}}_{\mathrm{cue}} : t \geq i\} \cup \{N\})$. Define an empty-neighborhood average as zero.
Initialize latent action labels $\{z_i\}_{i=1}^N$ and subspaces $\{(\mathbf{U}_k, \mathbf{o}_k)\}_{k=1}^K$.
Set $\mathcal{C}_k = \{i : z_i = k\}$ for $k = 1, \ldots, K$.
for $t = 1$ to $T_{\max}$ **do**
  Update weights $w_{k,i} = \mathbb{I}(i \in \mathcal{C}_k) + \lambda \sum_{j \in \mathcal{C}_k : \mathcal{N}_j \neq \varnothing} |\mathcal{N}_j|^{-1} \mathbb{I}(i \in \mathcal{N}_j)$.
  for $k = 1$ to K **do**
    Compute the intermediate offset with entries $o'_{k,j} = \frac{\sum_{i=1}^N (w_{k,i}/m_i)\phi_{i,j} x_{i,j}}{\sum_{i=1}^N (w_{k,i}/m_i)\phi_{i,j}}$.
    Set $\bar{\mathbf{x}}_{k,i} = (\mathbf{x}_i - \mathbf{o}'_k)/\sqrt{m_i}$ and construct $\mathbf{S}_k$ with entries $(\mathbf{S}_k)_{a,b} = \frac{\sum_{i=1}^N w_{k,i}\phi_{i,a}\phi_{i,b}\bar{x}_{k,i}^{(a)}\bar{x}_{k,i}^{(b)}}{\sum_{i=1}^N w_{k,i}\phi_{i,a}\phi_{i,b}}$.
    Estimate $\hat{d}_k = \arg\max_{q \in \{1,\ldots,D-1\}} (\lambda_{k,q} - \lambda_{k,q+1})$, where $\lambda_{k,1} \geq \cdots \geq \lambda_{k,D}$ are the eigenvalues of $\mathbf{S}_k$.
    Form candidate $\widetilde{\mathbf{U}}_k$ from the $\hat{d}_k$ leading eigenvectors and set $\widetilde{\mathbf{o}}_k = (\mathbf{I} - \widetilde{\mathbf{U}}_k \widetilde{\mathbf{U}}_k^{\mathrm{T}})\mathbf{o}'_k$.
    Accept $(\widetilde{\mathbf{U}}_k, \widetilde{\mathbf{o}}_k)$ only if its original masked contribution J_k does not exceed the previous J_k; otherwise retain $(\mathbf{U}_k, \mathbf{o}_k)$.
  end for
  for $i = 1$ to N **do**
    Compute $d_{k,i} = \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k) + \frac{\lambda}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \frac{D}{m_j}\xi(\mathbf{x}_j, \Phi_j, \mathbf{U}_k, \mathbf{o}_k)$ for all k.
    Update $z_i = \arg\min_{k \in \{1,\ldots,K\}} d_{k,i}$.
  end for
  Update $\mathcal{C}_k = \{i : z_i = k\}$ for all k.
  Recompute $w_{k,i}$ from the updated action classes.
  Compute $\mathcal{J}^{(t)} = \sum_{k=1}^K \sum_{i=1}^N \frac{D}{m_i}\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k) w_{k,i}$.
  if $t > 1$ and $|\mathcal{J}^{(t-1)} - \mathcal{J}^{(t)}|/\mathcal{J}^{(t-1)} < \eta$ **then**
    break
  end if
end for
Output: Labels $\{z_i\}_{i=1}^N$, boundaries $B(z)$, and category relation $\Pi(z)$ from Equation (1).

Table 9: Auditable feature-matrix boundary. N_s is the number of frames in sequence s. The four video inputs are exactly 324-dimensional after the loader transpose.

Dataset	Archived file pattern	MATLAB loader / variable	Matrix delivered to TSCC
Weiz	`Weiz_feature/wei_person_s.mat`	`load_Weiz / weiAllFeatureOnePerson`	$N_s \times 324$
UT	`UT_dataset/ut_seq_s.mat`	`load_UT / arr_feature`	$N_s \times 324$
Keck	`Keck_feature/keck_person_s.mat`	`load_Keck / keck_feature`	$N_s \times 324$
MAD	`MAD_dataset/sb{p}seq0{s}_feature.mat`	`load_MAD / hogFeatureAc`	$N_s \times 324$
Mocap	auxiliary archived pose matrix	pose variable; loader absent from this archive	$N_s \times D_{\mathrm{pose}}$

For one sequence, let $X \in \mathbb{R}^{N \times D}$ be this matrix and let $M \in \{0,1\}^{N \times D}$ be the generated reliability mask. The implemented order is: load and transpose X; generate M in the original coordinate indexing; then standardize each coordinate using observed entries only. With $n_j = \sum_i M_{ij} > 0$,

$$\mu_j = \frac{\sum_i M_{ij} X_{ij}}{n_j}, \qquad s_j^2 = \frac{\sum_i M_{ij}(X_{ij} - \mu_j)^2}{\max(n_j - 1, 1)}, \qquad \widetilde{X}_{ij} = M_{ij} \frac{X_{ij} - \mu_j}{\max(s_j, 10^{-12})}.$$

If $n_j = 0$, the coordinate is left at the zero placeholder and excluded by the mask; if $s_j = 0$, the denominator is set to one. The pair $(\widetilde{X}, M)$, rather than an imputed matrix alone, is passed to TSCC

and all incomplete-data baselines. Complete-input baselines receive interpolation only after the same mask is drawn. Every sequence is processed independently; labels enter only metric computation.

The inspected data archive has SHA-256 `0e16470cba05d541e96c9fca803d4c8d82912a1c81d981497ce002d08c28210e`. It also contains alternative 1024-dimensional GoogLeNet matrices for Weiz and UT, but the evaluated loaders above do not read them. Conversely, the archive does not contain the raw-frame-to-324D generator or its HOG cell/block settings, nor the Mocap loader/joint list. We therefore claim exact reproducibility from the fixed matrices onward, not raw-pixel feature extraction or exact Mocap provenance. A release must add those missing assets before making either claim; Mocap remains an explicitly auxiliary stress test in the present version.

Mask generation. For target coordinate-missing rate r, MAR selects $\lfloor (1-r)D \rfloor$ observed coordinates uniformly without replacement in every frame. Grouped-coordinate MNAR repeatedly selects a contiguous coordinate group in the descriptor's coordinate ordering until at least $\lceil rD \rceil$ coordinates are masked; the final artifact should document the mapping from coordinate indices to groups. Heterogeneous MAR assigns fixed propensities p_j to coordinates, rescales them so that $D^{-1}\sum_j(1-p_j) = r$, and samples $M_{ij} \sim \mathrm{Bernoulli}(p_j)$ independently. Temporal coordinate bursts sample nonoverlapping time intervals and one coordinate group per interval until the realized entry-missing rate reaches r. Persistent channel dropout samples one coordinate subset once and reuses it in every frame. Each run saves p_j, interval endpoints, group indices, and the realized mask. Frame MAR samples the requested number of indices without replacement; frame MNAR samples nonoverlapping ten-frame intervals until the target count is reached. Joint missingness samples body joints and masks every coordinate belonging to each selected joint. Rates always denote the fraction of coordinates or frame indices actually masked after resolving overlaps.

Tracking-failure scope. The tracking-failure rows use three archived mask groups ordered by average missing-joint ratio. Because detector identity, confidence threshold, and the conversion rule are unavailable in the archived aggregate record, these rows are reported only as an auxiliary structured-mask stress test and are not used to support a reproducibility or theoretical-coverage claim. The primary reproducible missingness experiment is coordinate MAR, whose masks are generated directly from the stated sampling rule.

Trials and model order. The five trials use independently generated masks and initializations. K is set to the ground-truth class count for TSCC and every baseline. The subspace dimensions are estimated independently in each trial using the eigengap rule. TSCC uses $\lambda = 1$, $T_{\max} = 50$, and $\eta = 10^{-5}$, without dataset-specific tuning.

S9 Additional Experimental Results

S9.1 Performance on Complete Sequences

Table 10: Segmentation performance on complete sequences (0% missing rate). The best result is highlighted in bold.

Method	Keck			Weiz			MAD			UT			Mocap		
	Acc	NMI	ARI	Acc	NMI	ARI	Acc	NMI	ARI	Acc	NMI	ARI	Acc	NMI	ARI
OSC Tierney et al. (2014)	81.2	73.5	65.1	76.8	69.2	61.3	79.5	71.3	63.8	78.1	70.6	62.4	77.3	69.8	61.5
TSC Li et al. (2015)	83.5	75.8	67.6	78.3	71.4	63.5	81.7	73.6	65.9	79.6	72.1	64.3	79.2	71.5	63.6
CDMS Zhou et al. (2022)	84.1	76.4	68.2	79.1	72.3	64.6	82.3	74.1	66.7	80.2	72.8	65.1	80.1	72.3	64.4
VSDA Bai et al. (2022)	84.7	77.2	69.0	79.7	73.1	65.4	83.1	74.9	67.5	80.9	73.5	65.9	80.8	73.1	65.2
TVS-SC Xing & Zhao (2026)	86.9	79.5	71.3	81.9	75.3	67.9	85.9	77.9	70.4	83.5	76.8	69.1	83.1	76.4	68.6
TSCC (Ours)	**88.6**	**81.3**	**73.5**	**83.5**	**77.2**	**69.6**	**87.2**	**80.1**	**72.3**	**85.7**	**79.4**	**71.6**	**84.9**	**78.6**	**70.8**

Before evaluating incomplete observations, we first compare TSCC with representative temporal segmentation methods on complete sequences. This experiment examines whether TSCC is a competitive temporal segmentation method even when no entries are missing.

As shown in Table 10, the reported TSCC aggregates are the largest among the retained methods on complete sequences. This result is consistent with temporal semantic consistency being useful beyond

incomplete observations and shows that the reported advantage is not confined to one missing-data correction setting. Per-run uncertainty is still needed before making a significance claim.

Table 11: Convergence and runtime statistics of TSCC. Results are averaged over five trials. All runtime results were measured on a single CPU utilizing an unoptimized MATLAB implementation.

Dataset	Iterations to converge	Runtime per sequence (s)	Final objective std.
Keck	9	0.52	0.018
Weiz	10	0.68	0.021
MAD	11	0.83	0.024
UT	10	0.71	0.022
Mocap	13	1.26	0.030

S10 Proof of Theorem 1

Proof. Let $\mathbf{y} = \mathbf{x} - \mathbf{o}$ and decompose $\mathbf{y} = \mathbf{U}\mathbf{a} + \mathbf{r}$, where $\mathbf{r} = (\mathbf{I} - \mathbf{U}\mathbf{U}^{\mathrm{T}})\mathbf{y}$ and $\mathbf{U}^{\mathrm{T}}\mathbf{r} = 0$. Define

$$p = \frac{m}{D}, \qquad \mathbf{A} = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{U}, \qquad \mathbf{b} = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{r}.$$

Because the component $\mathbf{U}\mathbf{a}$ can be absorbed into the masked least-squares coefficient,

$$\xi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o}) = \min_{\mathbf{q}\in\mathbb{R}^d} \|\Phi(\mathbf{r} - \mathbf{U}\mathbf{q})\|_2^2.$$

Uniform sampling without replacement, matrix Bernstein/Freedman concentration, scalar Bernstein concentration, and the vector Bernstein bound applied to the random permutation exposure martingale imply the following three events; sampling without replacement is no less concentrated than the corresponding independent draw up to universal constants Gross & Nesme (2010); Tropp (2011):

$$\left\|p^{-1}\mathbf{A} - \mathbf{I}_d\right\|_2 \le \eta_{\mathbf{U}}, \tag{24}$$

$$\left|p^{-1}\|\Phi\mathbf{r}\|_2^2 - \|\mathbf{r}\|_2^2\right| \le \eta_{\mathbf{r}}\|\mathbf{r}\|_2^2, \tag{25}$$

$$\|\mathbf{b}\|_2 \le p\eta_{\mathrm{c}}\|\mathbf{r}\|_2. \tag{26}$$

Using $\max_j \|\mathbf{e}_j^{\mathrm{T}}\mathbf{U}\|_2^2 = \mu_{\mathbf{U}}d/D$ and $\|\mathbf{r}\|_\infty^2 = \mu_{\mathbf{r}}\|\mathbf{r}\|_2^2/D$, the sample-size condition in (5) is a sufficient condition for (24)–(26) to hold jointly with probability at least $1 - 3\delta$, after enlarging the universal constant c_0. We state the slightly weaker $1 - 4\delta$ probability in the theorem.

On event (24), $\mathbf{A} \succeq p(1 - \eta_{\mathbf{U}})\mathbf{I}_d$, so it is invertible. The least-squares identity gives

$$\xi = \|\Phi\mathbf{r}\|_2^2 - \mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{b}.$$

The second term is nonnegative, and therefore

$$\xi \le \|\Phi\mathbf{r}\|_2^2 \le p(1 + \eta_{\mathbf{r}})\|\mathbf{r}\|_2^2.$$

For the lower bound, (24)–(26) yield

$$\mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{b} \le \frac{\|\mathbf{b}\|_2^2}{p(1 - \eta_{\mathbf{U}})} \le p\frac{\eta_{\mathrm{c}}^2}{1 - \eta_{\mathbf{U}}}\|\mathbf{r}\|_2^2.$$

Combining this inequality with (25) gives

$$\xi \ge p\left(1 - \eta_{\mathbf{r}} - \frac{\eta_{\mathrm{c}}^2}{1 - \eta_{\mathbf{U}}}\right)\|\mathbf{r}\|_2^2.$$

Finally, $\|\mathbf{r}\|_2^2 = \varepsilon(\mathbf{x}, \mathbf{U}, \mathbf{o})$, which proves (6). □

S11 Proof of Theorem 4

Proof. For brevity write $\rho_{k,i} = (D/m_i)\xi(\mathbf{x}_i, \Phi_i, \mathbf{U}_k, \mathbf{o}_k)$. Then

$$\begin{aligned}\sum_{h\in\mathcal{C}_k} d_{k,h} &= \sum_{h\in\mathcal{C}_k} \rho_{k,h} + \lambda \sum_{h\in\mathcal{C}_k} \frac{1}{|\mathcal{N}_h|} \sum_{i\in\mathcal{N}_h} \rho_{k,i} \\ &= \sum_{i=1}^{N} \rho_{k,i} \left[\mathbb{I}(i \in \mathcal{C}_k) + \lambda \sum_{h\in\mathcal{C}_k} \frac{\mathbb{I}(i \in \mathcal{N}_h)}{|\mathcal{N}_h|} \right] \\ &= \sum_{i=1}^{N} \rho_{k,i} \left[\mathbb{I}(i \in \mathcal{C}_k) + \lambda n_k(i) \right].\end{aligned}$$

The bracketed factor is precisely $w_{k,i}$. □

S12 Proof of Proposition 8

Proof. Let $\mathbf{P}_k = \mathbf{U}_k(\mathbf{U}_k^{\mathrm{T}}\mathbf{U}_k)^{-1}\mathbf{U}_k^{\mathrm{T}}$. Because $\mathbf{U}_k^{\mathrm{T}}(\mathbf{I} - \mathbf{P}_k) = \mathbf{0}$, the proposed $\mathbf{o}_k = (\mathbf{I} - \mathbf{P}_k)\mathbf{o}'_k$ satisfies $\mathbf{U}_k^{\mathrm{T}}\mathbf{o}_k = \mathbf{0}$. Moreover, $\mathbf{o}'_k - \mathbf{o}_k = \mathbf{P}_k\mathbf{o}'_k \in \mathrm{span}(\mathbf{U}_k)$. Translating an affine offset by a vector in its direction space does not change the affine set, hence $\mathrm{span}(\mathbf{U}_k) + \mathbf{o}'_k = \mathrm{span}(\mathbf{U}_k) + \mathbf{o}_k$. □

S13 Proof of Theorem 5

Proof. Apply Theorem 1 to all K subspaces and use a union bound. Under the sample-size condition in (23), the stated event has probability at least $1 - 4K\delta$, and for every k,

$$p(1 - \alpha_-)\varepsilon_k \le \xi_k \le p(1 + \alpha_+)\varepsilon_k.$$

For any $k \neq \ell$, the complete-data margin gives

$$\xi_\ell \le p(1 + \alpha_+)\varepsilon_\ell \le p(1 + \alpha_+)(1 - \gamma)\varepsilon_k.$$

The condition $\gamma > (\alpha_- + \alpha_+)/(1 + \alpha_+)$ is equivalent to $(1 + \alpha_+)(1 - \gamma) < 1 - \alpha_-$. Consequently,

$$\xi_\ell < p(1 - \alpha_-)\varepsilon_k \le \xi_k \qquad \text{for every } k \neq \ell,$$

which proves that the masked nearest-subspace label is ℓ. □

S14 Proof of Proposition 7

Proof. For $k \neq \ell_i$,

$$d_{k,i}(\lambda) - d_{\ell_i,i}(\lambda) = d_{k,i}(\lambda_0) - d_{\ell_i,i}(\lambda_0) + (\lambda - \lambda_0)(b_{k,i} - b_{\ell_i,i}).$$

The first term is at least g_i, while the absolute value of the second is at most $|\lambda - \lambda_0|M_i$. The difference is therefore positive whenever $|\lambda - \lambda_0| < g_i/M_i$. If $M_i = 0$, all pairwise neighbor differences are independent of λ, so the label cannot change. □

S15 Proof of Proposition 9

The exact masked subspace update contains the mask-dependent pseudo-inverse $(\mathbf{U}_k^{\mathrm{T}}\Phi_i\mathbf{U}_k)^\dagger$, and is therefore replaced by a surrogate PCA update. With fixed assignments and weights, the acceptance safeguard keeps the candidate update only if it does not increase J; hence each accepted subspace update is non-increasing in J. With fixed subspaces, the label update assigns each sample to the action class with the smallest assignment cost, and thus exactly minimizes the assignment objective over $\{z_i\}_{i=1}^N$. By the equivalent weighted formulation, the recomputed weighted objective is therefore also non-increasing. Since $J \ge 0$, the resulting objective sequence is monotone and bounded below,

and hence converges to a finite limit. The non-convexity of the joint problem prevents any guarantee of global optimality.

For complexity, constructing temporal neighborhoods scans the observations in $O(N\bar{m})$ time. For a generic mask, forming a restricted Gram matrix and solving its coefficient system costs $O(m_i d_{\max}^2 + d_{\max}^3)$. Summing over samples and candidates gives $O(KN(\bar{m}d_{\max}^2 + d_{\max}^3))$. Pairwise scatter construction costs $O(K\sum_i m_i^2)$; $\sum_i m_i^2$ cannot generally be replaced by $N\bar{m}^2$ in an upper bound when observation counts vary. Truncated dense eigensolvers add $O(KD^2 d_{\max})$, and segment prefix sums evaluate the temporal terms in $O(KN)$. Storing the data, bases, weights, and one scatter matrix requires $O(ND + KDd_{\max} + KN + D^2)$ memory.

S16 Deterministic Analysis for Arbitrary Fixed Masks

The random-coordinate theorem in the main paper is intentionally narrow. This section gives a complementary deterministic analysis that applies to any fixed diagonal mask, including group, joint, or block masks. The price is that the conclusion is expressed in terms of quantities that must be checked for the particular mask rather than guaranteed by a sampling distribution.

Let $\mathbf{U} \in \mathbb{R}^{D\times d}$ have orthonormal columns, let $\mathbf{y} = \mathbf{x} - \mathbf{o} = \mathbf{U}\mathbf{a} + \mathbf{r}$, and assume $\mathbf{U}^{\mathrm{T}}\mathbf{r} = \mathbf{0}$. Define

$$\mathbf{G}_\Phi = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{U}, \qquad \mathbf{c}_\Phi = \mathbf{U}^{\mathrm{T}}\Phi\mathbf{r}, \qquad q_\Phi = \mathbf{r}^{\mathrm{T}}\Phi\mathbf{r}.$$

The matrix $\mathbf{G}_\Phi$ is the restricted Gram matrix and $\mathbf{c}_\Phi$ measures the failure of the complete residual to remain orthogonal after restriction to observed coordinates.

Theorem 6 (Exact fixed-mask residual identity)**.** *For every diagonal $\Phi \in \{0,1\}^{D\times D}$,*

$$\xi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o}) = q_\Phi - \mathbf{c}_\Phi^{\mathrm{T}}\mathbf{G}_\Phi^{\dagger}\mathbf{c}_\Phi. \tag{27}$$

Consequently $0 \le \xi \le q_\Phi$. If $\lambda_{\min}(\mathbf{G}_\Phi) \ge \alpha > 0$, then

$$q_\Phi - \frac{\|\mathbf{c}_\Phi\|_2^2}{\alpha} \le \xi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o}) \le q_\Phi. \tag{28}$$

Proof. Since $\Phi\mathbf{y} = \Phi\mathbf{U}\mathbf{a} + \Phi\mathbf{r}$, projection onto the orthogonal complement of $\mathrm{span}(\Phi\mathbf{U})$ annihilates the first term. Therefore

$$\xi = \mathbf{r}^{\mathrm{T}}\Phi\left[\mathbf{I} - \Phi\mathbf{U}\mathbf{G}_\Phi^{\dagger}\mathbf{U}^{\mathrm{T}}\Phi\right]\Phi\mathbf{r}.$$

Using $\Phi^2 = \Phi$ gives Equation (27). The subtracted quadratic form is nonnegative because $\mathbf{G}_\Phi^{\dagger} \succeq 0$, proving the upper bound and nonnegativity. Under the eigenvalue condition, $\|\mathbf{G}_\Phi^{-1}\|_2 \le 1/\alpha$, which gives the lower bound. □

The identity exposes why structured missingness can fail even at a high observation rate. A mask may preserve most residual energy q_Φ but make $\mathbf{G}_\Phi$ nearly singular, or it may create a large leakage vector $\mathbf{c}_\Phi$. Either effect can collapse the observed residual.

Corollary 1 (Deterministic relative preservation)**.** *Let $p = m/D$ and suppose, for some $\epsilon_r, \epsilon_c \ge 0$ and $\alpha > 0$,*

$$|p^{-1}q_\Phi - \|\mathbf{r}\|_2^2| \le \epsilon_r\|\mathbf{r}\|_2^2, \quad \|\mathbf{c}_\Phi\|_2 \le \epsilon_c p\,\|\mathbf{r}\|_2, \quad \lambda_{\min}(\mathbf{G}_\Phi) \ge \alpha p.$$

Then the normalized residual $\rho = p^{-1}\xi$ satisfies

$$\left(1 - \epsilon_r - \frac{\epsilon_c^2}{\alpha}\right)\|\mathbf{r}\|_2^2 \le \rho \le (1+\epsilon_r)\|\mathbf{r}\|_2^2. \tag{29}$$

Proof. Apply Theorem 6, divide by p, and substitute the three assumed inequalities. □

Corollary 2 (Fixed-mask assignment preservation)**.** *For candidate subspaces $k = 1, \dots, K$, let ε_k be the complete residual and ρ_k the normalized masked residual. If $|\rho_k - \varepsilon_k| \le \tau_k$ for every k, and the complete-data winner ℓ obeys*

$$\varepsilon_k - \varepsilon_\ell > \tau_k + \tau_\ell \qquad \text{for all } k \ne \ell,$$

then $\ell = \arg\min_k \rho_k$.

Proof. For $k \neq \ell$, $\rho_k - \rho_\ell \geq (\varepsilon_k - \varepsilon_\ell) - (\tau_k + \tau_\ell) > 0$. □

Corollary 2 separates geometry from mask generation. For a structured mask, one may estimate the restricted Gram eigenvalue, retained residual energy, and leakage on held-out complete data. The result is a certificate for that mask family, not a distribution-free claim for every dependent mask.

S17 Uniform Preservation over a Finite Sequence

The pointwise theorem in the main paper is sufficient to justify one residual, but the algorithm compares K residuals for each of N samples. A direct union argument gives a sequence-level event and makes the logarithmic price of uniformity explicit.

Theorem 7 (Uniform residual preservation)**.** *For every pair (i,k), assume $\varepsilon_{k,i} > 0$ and that Ω_i is sampled uniformly without replacement with size m_i. Fix $\eta_{\mathbf{U}}, \eta_{\mathbf{r}}, \eta_{\mathrm{c}} \in (0,1)$ with $\eta_{\mathbf{r}} + \eta_{\mathrm{c}}^2/(1-\eta_{\mathbf{U}}) < 1$. If, for all i,k,*

$$m_i \geq c_0 \max\left\{ \frac{\mu_{\mathbf{U}_k} d_k \log(8NKd_k/\delta)}{\eta_{\mathbf{U}}^2}, \frac{\mu_{\mathbf{r}_{k,i}} \log(8NK/\delta)}{\eta_{\mathbf{r}}^2}, \frac{\mu_{\mathbf{U}_k}\mu_{\mathbf{r}_{k,i}} d_k \log(8NKd_k/\delta)}{\eta_{\mathrm{c}}^2} \right\}, \tag{30}$$

then, with probability at least $1-\delta$, simultaneously for every $i = 1,\dots,N$ and $k = 1,\dots,K$,

$$(1-\alpha_-)\varepsilon_{k,i} \leq \rho_{k,i} \leq (1+\alpha_+)\varepsilon_{k,i}, \quad \alpha_- = \eta_{\mathbf{r}} + \frac{\eta_{\mathrm{c}}^2}{1-\eta_{\mathbf{U}}}, \quad \alpha_+ = \eta_{\mathbf{r}}. \tag{31}$$

Proof. Apply Theorem 1 to pair (i,k) with failure parameter $\delta' = \delta/(4NK)$. Its failure probability is at most $4\delta' = \delta/(NK)$. Equation (30) dominates the resulting logarithmic terms. A union bound over the NK pairs proves the simultaneous event. Pairs with zero complete residual obey zero masked residual exactly and can be inserted without changing the bound. □

Corollary 3 (Uniform preservation of all nearest-subspace labels)**.** *On the event in Theorem 7, every complete-data label is preserved whenever*

$$\frac{\varepsilon_{\ell_i,i}}{\varepsilon_{k,i}} < \frac{1-\alpha_-}{1+\alpha_+} \qquad \text{for every } i \text{ and every } k \neq \ell_i. \tag{32}$$

Proof. For each competitor $k \neq \ell_i$, $\rho_{\ell_i,i} \leq (1+\alpha_+)\varepsilon_{\ell_i,i}$, whereas $\rho_{k,i} \geq (1-\alpha_-)\varepsilon_{k,i}$. Condition (32) makes the first quantity strictly smaller. □

The logarithmic dependence on NK is mild, but the worst-case coherence in Equation (30) can be conservative. This is not only a proof artifact: a single sample whose residual is concentrated on one coordinate cannot be protected from missing that coordinate without either a high observation probability or an additional structural assumption.

S17.1 An empirical certificate with nonuniform tolerances

The uniform theorem uses common tolerances for readability. A sharper certificate assigns pair-specific errors $\tau_{k,i}$ and checks

$$\rho_{k,i} \in [\varepsilon_{k,i} - \tau_{k,i}, \varepsilon_{k,i} + \tau_{k,i}].$$

Corollary 2 then requires only $g_i > \tau_{\ell_i,i} + \max_{k \neq \ell_i} \tau_{k,i}$. This formulation is useful for a post-hoc audit because difficult, high-coherence frames need not force every other frame to use the same worst-case tolerance.

S18 HETEROGENEOUS COORDINATE OBSERVATION

Uniform sampling is analytically clean but does not model sensors whose coordinates have different reliabilities. We therefore consider independent Bernoulli observations $\phi_j \sim \mathrm{Bernoulli}(\pi_j)$, with known $\pi_j \in (0, 1]$. The inverse-probability weighted residual is

$$\xi_\pi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o}) = \min_{\mathbf{v}} \left\| \mathbf{W}_\Phi^{1/2}(\mathbf{x} - \mathbf{o} - \mathbf{U}\mathbf{v}) \right\|_2^2. \tag{33}$$

Because $\mathbb{E}\mathbf{W}_\Phi = \mathbf{I}$, this construction corrects the first-order bias created by unequal observation rates.

For the rows $\mathbf{u}_j^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}}\mathbf{U}$, define weighted leverage quantities

$$\nu_{\mathbf{U}} = \max_j \frac{\|\mathbf{u}_j\|_2^2}{\pi_j}, \qquad \nu_{\mathbf{r}} = \max_j \frac{r_j^2}{\pi_j \|\mathbf{r}\|_2^2},$$
$$\nu_{\mathbf{U}\mathbf{r}} = \max_j \frac{\|\mathbf{u}_j\|_2 |r_j|}{\pi_j \|\mathbf{r}\|_2}.$$

These replace ordinary coherence. A coordinate is dangerous when it has high subspace leverage or residual energy but low observation probability. The cross term also requires its summed variance, not just a largest-coordinate bound. Define

$$v_{\mathbf{U}\mathbf{r}} = \sum_j \frac{1 - \pi_j}{\pi_j} \frac{\|\mathbf{u}_j\|_2^2 r_j^2}{\|\mathbf{r}\|_2^2}, \qquad \ell_\delta = \log(2(d+1)/\delta).$$

Theorem 8 (IPW residual preservation)**.** *Let $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$, $\mathbf{r} = \mathbf{P}_{\mathbf{U}}^{\perp}(\mathbf{x} - \mathbf{o}) \neq 0$, and let coordinates be observed independently with probabilities π_j. Fix tolerances in $(0, 1)$. There is a universal constant $C > 0$ such that if*

$$\begin{aligned} \nu_{\mathbf{U}}\ell_\delta \le C^{-1}\eta_{\mathbf{U}}^2, \qquad \nu_{\mathbf{r}}\ell_\delta \le C^{-1}\eta_{\mathbf{r}}^2, \\ \sqrt{v_{\mathbf{U}\mathbf{r}}\ell_\delta} + \nu_{\mathbf{U}\mathbf{r}}\ell_\delta \le C^{-1}\eta_{\mathrm{c}}, \end{aligned} \tag{34}$$

then, with probability at least $1 - 4\delta$,

$$\left(1 - \eta_{\mathbf{r}} - \frac{\eta_{\mathrm{c}}^2}{1 - \eta_{\mathbf{U}}}\right) \|\mathbf{r}\|_2^2 \le \xi_\pi \le (1 + \eta_{\mathbf{r}})\|\mathbf{r}\|_2^2. \tag{35}$$

Proof. Set $\mathbf{G} = \mathbf{U}^{\mathrm{T}}\mathbf{W}_\Phi\mathbf{U}$, $\mathbf{c} = \mathbf{U}^{\mathrm{T}}\mathbf{W}_\Phi\mathbf{r}$, and $q = \mathbf{r}^{\mathrm{T}}\mathbf{W}_\Phi\mathbf{r}$. Weighted least squares gives the exact identity

$$\xi_\pi = q - \mathbf{c}^{\mathrm{T}}\mathbf{G}^\dagger\mathbf{c}.$$

Since $\mathbb{E}\mathbf{G} = \mathbf{I}$, matrix Bernstein applied to $\sum_j(\phi_j/\pi_j - 1)\mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}$ yields $\|\mathbf{G} - \mathbf{I}\|_2 \le \eta_{\mathbf{U}}$. Scalar Bernstein applied to $\sum_j(\phi_j/\pi_j - 1)r_j^2$ yields $|q - \|\mathbf{r}\|_2^2| \le \eta_{\mathbf{r}}\|\mathbf{r}\|_2^2$. Vector Bernstein applied to $\sum_j(\phi_j/\pi_j - 1)\mathbf{u}_j r_j$, whose mean is zero because $\mathbf{U}^{\mathrm{T}}\mathbf{r} = 0$, uses normalized summed variance $v_{\mathbf{U}\mathbf{r}}$ and summand bound $\nu_{\mathbf{U}\mathbf{r}}$. Thus it yields $\|\mathbf{c}\|_2 \le \eta_{\mathrm{c}}\|\mathbf{r}\|_2$. The three events hold together with probability at least $1 - 4\delta$. On their intersection, $\lambda_{\min}(\mathbf{G}) \ge 1 - \eta_{\mathbf{U}}$; substituting into the exact identity proves Equation (35). □

Corollary 4 (Uniform sampling as a special case)**.** *If $\pi_j = p$ for every coordinate, then $\nu_{\mathbf{U}} = \mu_{\mathbf{U}}d/(pD)$ and $\nu_{\mathbf{r}} = \mu_{\mathbf{r}}/(pD)$. Thus Equation* (34) *recovers the same dependence on observation rate, coherence, and dimension as the uniform-coordinate theorem, up to universal constants and the difference between Bernoulli and fixed-size sampling.*

S18.1 CLIPPING UNSTABLE INVERSE-PROBABILITY WEIGHTS

When some π_j is extremely small, IPW is unbiased but can have excessive variance. Let $\bar{\pi}_j = \max\{\pi_j, \pi_{\min}\}$ and replace $1/\pi_j$ by $1/\bar{\pi}_j$. The resulting residual has a deterministic bias–variance decomposition:

$$\left|\mathbb{E}q_{\mathrm{clip}} - \|\mathbf{r}\|_2^2\right| \le \sum_{j:\pi_j < \pi_{\min}} \left(1 - \frac{\pi_j}{\pi_{\min}}\right) r_j^2,$$

while every summand is bounded by $r_j^2/\pi_{\min}$. Clipping therefore trades controlled bias on chronically missing coordinates for a smaller concentration radius. No choice of clipping can recover energy that is never observed; that obstruction is formalized later.

S19 Group Masks and Block Observations

Partition coordinates into disjoint groups $G_1, \dots, G_M$. For group g, let $\mathbf{U}_{G_g}$ and $\mathbf{r}_{G_g}$ denote the corresponding rows and residual coordinates. Define group leverage, energy, and cross-leakage scores

$$L_g = \|\mathbf{U}_{G_g}\|_2^2, \qquad R_g = \|\mathbf{r}_{G_g}\|_2^2, \qquad C_g = \|\mathbf{U}_{G_g}^{\mathrm{T}}\mathbf{r}_{G_g}\|_2.$$

Proposition 10 (Deterministic group-mask certificate)**.** *Let $\mathcal{O} \subseteq \{1, \dots, M\}$ be the observed groups and let $\Phi_{\mathcal{O}}$ be their coordinate mask. If*

$$\lambda_{\min}\left(\sum_{g\in\mathcal{O}} \mathbf{U}_{G_g}^{\mathrm{T}}\mathbf{U}_{G_g}\right) \geq \alpha, \qquad \left|p^{-1}\sum_{g\in\mathcal{O}} R_g - \|\mathbf{r}\|_2^2\right| \leq \epsilon_r\|\mathbf{r}\|_2^2, \qquad \left\|\sum_{g\in\mathcal{O}} \mathbf{U}_{G_g}^{\mathrm{T}}\mathbf{r}_{G_g}\right\|_2 \leq \epsilon_c\sqrt{p}\,\|\mathbf{r}\|_2.$$

then the normalized fixed-mask residual obeys

$$\left(1 - \epsilon_r - \frac{\epsilon_c^2}{\alpha}\right)\|\mathbf{r}\|_2^2 \leq p^{-1}\xi \leq (1+\epsilon_r)\|\mathbf{r}\|_2^2.$$

Proof. Apply Equation (28) with the unnormalized Gram lower bound α. The leakage term divided by p is at most $\epsilon_c^2\|\mathbf{r}\|_2^2/\alpha$; the energy bound gives the remaining terms. □

The proposition shows that group size alone is not decisive. Removing one small group can be catastrophic if that group contains a direction required to identify the subspace, whereas removing a large low-leverage group may be benign. This is the appropriate interpretation of joint-level missingness: body joints define semantic groups whose leverage can be highly unequal.

Theorem 9 (Random group sampling)**.** *Suppose groups are independently observed with probabilities π_g, and use inverse-probability group weights. Define*

$$\nu_{\mathrm{G}} = \max_g \frac{L_g}{\pi_g}, \quad \nu_{\mathrm{R}} = \max_g \frac{R_g}{\pi_g\|\mathbf{r}\|_2^2}, \quad \nu_{\mathrm{C}} = \max_g \frac{C_g}{\pi_g\|\mathbf{r}\|_2}.$$

Under the analogues of Equation (34) *with $(\nu_{\mathbf{U}}, \nu_{\mathbf{r}}, \nu_{\mathbf{Ur}})$ replaced by $(\nu_{\mathrm{G}}, \nu_{\mathrm{R}}, \nu_{\mathrm{C}})$, and summed cross variance replaced by*

$$v_{\mathrm{C}} = \sum_g \frac{1-\pi_g}{\pi_g}\frac{C_g^2}{\|\mathbf{r}\|_2^2},$$

the IPW group residual satisfies Equation (35) *with probability at least $1 - 4\delta$.*

Proof. Treat each observed group as one independent matrix/vector/scalar summand: $\mathbf{U}_{G_g}^{\mathrm{T}}\mathbf{U}_{G_g}$, $\mathbf{U}_{G_g}^{\mathrm{T}}\mathbf{r}_{G_g}$, and R_g, respectively. Matrix, vector, and scalar Bernstein inequalities then repeat the proof of Theorem 8. The group scores bound the largest summand in each concentration inequality. □

S20 Temporal-Difference and Boundary Stability

Temporal neighborhoods are constructed from consecutive differences. We now quantify when missing coordinates preserve both the difference magnitude and the local-maxima boundary decision. Let $\mathbf{h}_t = \mathbf{x}_{t+1} - \mathbf{x}_t$, $\Psi_t = \Phi_t\Phi_{t+1}$, and $p_t = m_{t,t+1}/D$. The observed statistic is $\Delta_t = p_t^{-1}\|\Psi_t\mathbf{h}_t\|_2^2$, while its complete-data counterpart is $H_t = \|\mathbf{h}_t\|_2^2$.

Proposition 11 (Concentration of normalized temporal differences)**.** *Assume the common observed set for pair t is sampled uniformly without replacement with size $m_{t,t+1}$. Define $\mu_t = D\|\mathbf{h}_t\|_\infty^2/\|\mathbf{h}_t\|_2^2$. There is a universal constant $c > 0$ such that, for any $\epsilon \in (0,1)$,*

$$m_{t,t+1} \geq c\mu_t\epsilon^{-2}\log(2/\delta) \quad \Longrightarrow \quad |\Delta_t - H_t| \leq \epsilon H_t$$

with probability at least $1 - \delta$.

Proof. Write $\|\Psi_t \mathbf{h}_t\|_2^2$ as sampling without replacement from the finite population $\{h_{t,j}^2\}_{j=1}^D$. A scalar concentration bound controls its deviation from $p_t\|\mathbf{h}_t\|_2^2$. The maximum population element is $\|\mathbf{h}_t\|_\infty^2 = \mu_t\|\mathbf{h}_t\|_2^2/D$; substitution and division by p_t give the stated sample complexity after absorbing numerical constants into c. □

Theorem 10 (Local boundary stability). *Suppose the complete statistic has a strict local maximum at t, and let*

$$g_t = H_t - \max\{H_{t-1}, H_{t+1}\} > 0.$$

If $|\Delta_s - H_s| \le e_s$ for $s \in \{t-1, t, t+1\}$ and

$$g_t > e_t + \max\{e_{t-1}, e_{t+1}\},$$

then t remains a detected local maximum of the observed statistic.

Proof. For either neighbor $s \in \{t-1, t+1\}$,

$$\Delta_t - \Delta_s \ge (H_t - e_t) - (H_s + e_s) \ge g_t - e_t - \max\{e_{t-1}, e_{t+1}\} > 0.$$

Thus both local-maxima inequalities are preserved. □

This theorem explains two empirical failure modes. First, gradual action transitions have small g_t, so even mild masking can move a boundary. Second, when adjacent frames share few coordinates, the errors e_s are large and the method conservatively treats zero-overlap locations as candidate discontinuities.

S21 Robustness to Additive Observation Noise

Missingness and measurement noise are distinct. Missingness changes the geometry on which the residual is evaluated, whereas noise perturbs the observed vector inside that geometry. Let $\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{e}$, where only $\Phi\widetilde{\mathbf{x}}$ is available. For fixed $(\Phi, \mathbf{U}, \mathbf{o})$, write $\xi = \xi(\mathbf{x}, \Phi, \mathbf{U}, \mathbf{o})$ and $\widetilde{\xi} = \xi(\widetilde{\mathbf{x}}, \Phi, \mathbf{U}, \mathbf{o})$.

Lemma 1 (Square-root Lipschitz property). *For every perturbation $\mathbf{e}$,*

$$\left|\sqrt{\widetilde{\xi}} - \sqrt{\xi}\right| \le \|\Phi\mathbf{e}\|_2. \tag{36}$$

Consequently,

$$|\widetilde{\xi} - \xi| \le 2\sqrt{\xi}\,\|\Phi\mathbf{e}\|_2 + \|\Phi\mathbf{e}\|_2^2. \tag{37}$$

Proof. Let $\mathbf{Q}_{\Phi,\mathbf{U}}$ denote the orthogonal projector onto the orthogonal complement of $\mathrm{span}(\Phi\mathbf{U})$. Then $\sqrt{\xi} = \|\mathbf{Q}_{\Phi,\mathbf{U}}\Phi(\mathbf{x}-\mathbf{o})\|_2$. The reverse triangle inequality and $\|\mathbf{Q}_{\Phi,\mathbf{U}}\|_2 \le 1$ give Equation (36). Factoring the difference of squares gives Equation (37). □

Corollary 5 (Normalized noisy residual). *If $\|\Phi\mathbf{e}\|_2 \le \sigma\sqrt{p}$, then for $\rho = p^{-1}\xi$ and $\widetilde{\rho} = p^{-1}\widetilde{\xi}$,*

$$|\widetilde{\rho} - \rho| \le 2\sigma\sqrt{\rho} + \sigma^2. \tag{38}$$

If additionally $\rho \le (1+\alpha_+)\varepsilon$, then

$$\widetilde{\rho} \le \left(\sqrt{(1+\alpha_+)\varepsilon} + \sigma\right)^2.$$

The bound scales with observed noise energy, not with unobserved noise. This is the correct robustness notion for an imputation-free method. It also shows why normalization by p^{-1} amplifies noise when very few coordinates are observed; the signal and noise are both rescaled.

Theorem 11 (Noisy nearest-subspace preservation). *Suppose $|\rho_k - \varepsilon_k| \le \tau_k$ and $\|\Phi\mathbf{e}\|_2 \le \sigma\sqrt{p}$. Define*

$$\zeta_k = \tau_k + 2\sigma\sqrt{\varepsilon_k + \tau_k} + \sigma^2.$$

If the complete winner ℓ satisfies

$$\varepsilon_k - \varepsilon_\ell > \zeta_k + \zeta_\ell \qquad \text{for all } k \ne \ell, \tag{39}$$

then the noisy masked residual has the same minimizing subspace.

Proof. Corollary 5 and $\rho_k \le \varepsilon_k + \tau_k$ imply $|\widetilde{\rho}_k - \varepsilon_k| \le \zeta_k$. Apply Corollary 2. □

S21.1 SUBGAUSSIAN NOISE CONSEQUENCE

If the observed noise coordinates are conditionally independent, mean-zero, and s_e-subgaussian, standard norm concentration gives $\|\Phi\mathbf{e}\|_2 \lesssim s_e(\sqrt{m} + \sqrt{\log(1/\delta)})$ with probability at least $1-\delta$. Hence one may take

$$\sigma \lesssim s_e\sqrt{D}\left(1 + \sqrt{\frac{\log(1/\delta)}{m}}\right).$$

This expression makes clear that D/m normalization estimates a complete-coordinate energy: even when the per-coordinate noise is fixed, the normalized noise floor is of order Ds_e^2.

S22 PERTURBING THE AFFINE SUBSPACES

The preceding results compare residuals to fixed candidate subspaces. TSCC, however, estimates both the basis and offset. This section controls the error introduced by replacing $(\mathbf{U}, \mathbf{o})$ with $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{o}})$.

Lemma 2 (Complete affine-residual perturbation)**.** *Let $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ have orthonormal columns of the same dimension, let $s = \|\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\widetilde{\mathbf{U}}}\|_2$, and set $b = \|\mathbf{o} - \widetilde{\mathbf{o}}\|_2$ and $R = \|\mathbf{x} - \mathbf{o}\|_2$. Then*

$$\left|\varepsilon(\mathbf{x}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{o}}) - \varepsilon(\mathbf{x}, \mathbf{U}, \mathbf{o})\right| \le sR^2 + 2Rb + b^2 + 2sRb + sb^2. \tag{40}$$

Proof. Write $\mathbf{y} = \mathbf{x} - \mathbf{o}$ and $\mathbf{q} = \mathbf{o} - \widetilde{\mathbf{o}}$. Add and subtract $\mathbf{y}^{\mathrm{T}}\mathbf{P}^{p}_{\widetilde{\mathbf{U}}} erp\mathbf{y}$. The basis contribution is at most sR^2. The offset contribution equals

$$2\mathbf{q}^{\mathrm{T}}\mathbf{P}^{p}_{\widetilde{\mathbf{U}}} erp\mathbf{y} + \mathbf{q}^{\mathrm{T}}\mathbf{P}^{p}_{\widetilde{\mathbf{U}}} erp\mathbf{q},$$

whose absolute value is at most $2Rb + b^2$. Replacing R by the looser $(1+s)R$ and the projector action on $\mathbf{q}$ by $(1+s)b$ yields the displayed symmetric bound; the first three terms already suffice when both projectors are orthogonal. □

Corollary 6 (Davis–Kahan residual control)**.** *Under Theorem 25, if $\|\mathbf{E}_k\|_2 \le g_k/2$, then the basis component of the complete residual error is at most*

$$\frac{2\|\mathbf{E}_k\|_2}{g_k}\|\mathbf{x} - \mathbf{o}_k\|_2^2.$$

Thus a small scatter perturbation is useful only relative to the population eigengap.

Theorem 12 (Assignment stability with estimated affine subspaces)**.** *For each candidate k, suppose the estimated affine residual differs from the target complete residual by at most a_k, and the masked/noisy residual differs from the estimated complete residual by at most b_k. If*

$$\varepsilon_k - \varepsilon_\ell > a_k + a_\ell + b_k + b_\ell \qquad \textit{for all } k \ne \ell,$$

then the final masked/noisy nearest-subspace assignment equals ℓ.

Proof. Two applications of the triangle inequality give total residual error at most $a_k + b_k$ for subspace k. The claim follows from the deterministic margin argument of Corollary 2. □

This theorem is deliberately modular. Lemma 2 or Corollary 6 supplies a_k; random-mask, fixed-mask, and noise results supply b_k. The decomposition prevents a geometric estimation claim from being conflated with a missing-data claim.

S23 EXACT RECOVERY OF THE BOUNDARY SET

The local theorem in Appendix S20 protects one peak. Here we give a simultaneous condition for recovering the entire boundary set. For $t = 2, \dots, N-2$, define the two signed comparison margins

$$s_t^- = H_t - H_{t-1}, \qquad s_t^+ = H_t - H_{t+1}.$$

Assume there are no ties, and define the global comparison margin

$$g_{\mathcal{B}} = min_{2\le t\le N-2} \min\{|s_t^-|, |s_t^+|\}.$$

Theorem 13 (Exact boundary-set recovery). *If $g_{\mathcal{B}} > 0$ and $\max_t |\Delta_t - H_t| < g_{\mathcal{B}}/2$, then $\widehat{\mathcal{B}} = \mathcal{B}^\star$.*

Proof. For every adjacent comparison,

$$|(\Delta_t - \Delta_{t-1}) - (H_t - H_{t-1})| \le |\Delta_t - H_t| + |\Delta_{t-1} - H_{t-1}| < g_{\mathcal{B}}.$$

Thus the sign of every left comparison is preserved; the same argument holds for every right comparison. Membership in the local-maxima set is determined by these two signs, so all memberships are preserved simultaneously. □

Corollary 7 (Random-coordinate boundary recovery). *Under the sampling model of Proposition 11, suppose $H_t \le H_{\max}$ and $\mu_t \le \mu_h$ for all t. There is a constant C such that*

$$\min_t m_{t,t+1} \ge C\mu_h \frac{H_{\max}^2}{g_{\mathcal{B}}^2} \log \frac{2N}{\delta}$$

implies $\widehat{\mathcal{B}} = \mathcal{B}^\star$ with probability at least $1 - \delta$.

Proof. Apply Proposition 11 with $\epsilon = g_{\mathcal{B}}/(2H_{\max})$ and failure probability δ/N, then union bound over all consecutive pairs. Theorem 13 completes the proof. □

The dependence on $g_{\mathcal{B}}^{-2}$ is unavoidable for a comparison- based detector: when a peak and its neighbor have almost identical energies, an arbitrarily small perturbation can exchange their order.

S23.1 Approximate localization when exact recovery fails

Let $\mathcal{P}$ be a set of disjoint windows, each containing one complete boundary. Suppose that in every window $W \in \mathcal{P}$, the true boundary t_W satisfies

$$H_{t_W} - \max_{s \in W : |s - t_W| > q} H_s > 2e, \qquad \max_t |\Delta_t - H_t| \le e.$$

Then the largest observed difference in W lies within q indices of t_W. This weaker statement is useful for gradual transitions: it gives a localization radius even when immediate-neighbor comparisons have small margin.

S24 Stability to Neighborhood Perturbations

Represent the adaptive neighborhood system by a row-stochastic matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where $A_{ij} = |\mathcal{N}_i|^{-1}\mathbb{I}(j \in \mathcal{N}_i)$ for nonempty neighborhoods and the row is zero otherwise. For action class k, collect residuals in $\boldsymbol{\rho}_k = (\rho_{k,1}, \dots, \rho_{k,N})^{\mathrm{T}}$. The vector of assignment costs is

$$\mathbf{d}_k = (\mathbf{I} + \lambda \mathbf{A})\boldsymbol{\rho}_k. \tag{41}$$

Lemma 3 (Cost perturbation under residual and neighborhood error). *Let $\widetilde{\boldsymbol{\rho}}_k$ and $\widetilde{\mathbf{A}}$ be perturbed residuals and neighborhoods. If*

$$\|\widetilde{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\|_\infty \le \tau, \quad \|\widetilde{\mathbf{A}} - \mathbf{A}\|_{\infty \to \infty} \le \kappa, \quad \|\boldsymbol{\rho}_k\|_\infty \le R,$$

then

$$\|\widetilde{\mathbf{d}}_k - \mathbf{d}_k\|_\infty \le (1 + \lambda)\tau + \lambda\kappa(R + \tau). \tag{42}$$

Proof. Add and subtract $(\mathbf{I} + \lambda\widetilde{\mathbf{A}})\boldsymbol{\rho}_k$. Since both neighborhood matrices have row sums at most one,

$$\|(\mathbf{I} + \lambda\widetilde{\mathbf{A}})(\widetilde{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k)\|_\infty \le (1 + \lambda)\tau.$$

The remaining term is at most $\lambda\kappa\|\boldsymbol{\rho}_k\|_\infty$; using the perturbed vector instead yields the displayed slightly more conservative bound. □

Theorem 14 (Temporal-semantic label stability). *Let ℓ_i minimize the ideal temporal cost $d_{k,i}$, and define its temporal margin*

$$g_i^{\mathrm{temp}} = \min_{k \neq \ell_i} (d_{k,i} - d_{\ell_i,i}).$$

Under the assumptions of Lemma 3, all labels are preserved if

$$\min_i g_i^{\mathrm{temp}} > 2\left[(1+\lambda)\tau + \lambda\kappa(R+\tau)\right]. \tag{43}$$

Proof. Every action class cost changes by at most the bracketed quantity. The difference between a competitor and the ideal winner therefore changes by at most twice that amount. Condition (43) preserves all strict inequalities. □

If Theorem 13 holds, then $\kappa = 0$ because the neighborhood partition is exact. Otherwise, κ quantifies how much uniform neighbor mass moves across estimated boundaries. This operator view separates the effect of inaccurate residuals from inaccurate temporal support.

S25 Global Structure of the Temporal-Weight Path

Proposition 7 is local. With fixed subspaces and neighborhoods, the entire solution path in λ can also be described. For sample i, every action class cost is an affine function $d_{k,i}(\lambda) = a_{k,i} + \lambda b_{k,i}$.

Theorem 15 (Piecewise-constant label path). *For fixed subspaces and neighborhoods, the minimizing label $z_i(\lambda) = \arg\min_k d_{k,i}(\lambda)$, with deterministic tie breaking, is piecewise constant on $[0, \infty)$. All possible breakpoints belong to*

$$\mathcal{T}_i = \left\{ \frac{a_{h,i} - a_{k,i}}{b_{k,i} - b_{h,i}} : k < h,\ b_{k,i} \neq b_{h,i},\ \lambda \geq 0 \right\},$$

so there are at most $K(K-1)/2$ candidate breakpoints per sample.

Proof. The lower envelope of finitely many affine functions is affine between their pairwise intersections. Its active function, and hence the minimizing label, can change only at an intersection. Deterministic tie breaking defines the label at the finitely many intersection points. □

Corollary 8 (Dataset-level stability interval). *If no candidate breakpoint from $\cup_i \mathcal{T}_i$ lies in an open interval (λ_-, λ_+), then the full assignment vector is constant throughout that interval.*

The result provides an exact sensitivity diagnostic after fitting: compute all candidate intersections, sort them, and evaluate one representative value per interval. This is stronger than a dense numerical grid because it cannot miss a narrow transition interval when subspaces and neighborhoods are held fixed.

S26 End-to-End Assignment Guarantees

We now combine the preceding components. Let $d^\star_{k,i}$ be an ideal temporal cost computed from complete observations, target affine subspaces, and the complete boundary partition. Let $\widehat{d}_{k,i}$ be the cost actually used by TSCC. Decompose its error into

$$E_{k,i}^{\mathrm{mask}}, \quad E_{k,i}^{\mathrm{noise}}, \quad E_{k,i}^{\mathrm{sub}}, \quad E_{k,i}^{\mathrm{nbr}},$$

corresponding to masking, measurement noise, estimated affine subspaces, and neighborhood error. Define $E_i = \max_k (E_{k,i}^{\mathrm{mask}} + E_{k,i}^{\mathrm{noise}} + E_{k,i}^{\mathrm{sub}} + E_{k,i}^{\mathrm{nbr}})$.

Theorem 16 (Modular end-to-end label preservation). *Suppose $|\widehat{d}_{k,i} - d^\star_{k,i}| \leq E_i$ for all k, and let $\ell_i = \arg\min_k d^\star_{k,i}$ have ideal temporal margin $g_i^{\mathrm{temp}} > 0$. If*

$$g_i^{\mathrm{temp}} > 2E_i, \tag{44}$$

then TSCC assigns sample i to ℓ_i. If the condition holds for every sample, the entire assignment vector is preserved.

Proof. For any competitor $k \neq \ell_i$,

$$\widehat{d}_{k,i} - \widehat{d}_{\ell_i,i} \geq d^\star_{k,i} - d^\star_{\ell_i,i} - 2E_i \geq g_i^{\mathrm{temp}} - 2E_i > 0.$$

□

One admissible choice of the four terms is

$$\begin{aligned} E_{k,i}^{\mathrm{mask}} &= (1+\lambda)\tau_{k,i}, \\ E_{k,i}^{\mathrm{noise}} &= (1+\lambda)\left(2\sigma_i\sqrt{\varepsilon_{k,i}+\tau_{k,i}} + \sigma_i^2\right), \\ E_{k,i}^{\mathrm{sub}} &= (1+\lambda)a_{k,i}, \\ E_{k,i}^{\mathrm{nbr}} &= \lambda\kappa_i(R_k + \tau_{k,i}), \end{aligned}$$

where $\tau_{k,i}$ comes from a mask-preservation theorem, σ_i from Corollary 5, $a_{k,i}$ from Lemma 2, and κ_i from Lemma 3. The formula is conservative because it adds worst-case errors, but it makes every source of robustness auditable.

Corollary 9 (High-probability sequence recovery)**.** *Assume the uniform random-mask event of Theorem 7, the subgaussian noise event in Appendix S21, the exact-boundary event in Theorem 13, and deterministic subspace errors $a_{k,i}$. If Equation* (44) *holds with the resulting errors, then the full label vector is recovered with probability at least one minus the sum of the component failure probabilities.*

Proof. Union bound the component good events. On their intersection, Theorem 13 gives $\kappa_i = 0$, and the remaining component inequalities imply the premise of Theorem 16 for every pair (i, k). □

S26.1 A CONDITIONAL BASIN-OF-ATTRACTION STATEMENT

Let $Z^\star$ be a target assignment. Suppose every assignment vector in a set $\mathcal{Z}_0$ produces, after the safeguarded subspace update, bases and offsets satisfying the error bounds used above. If the resulting temporal margins satisfy Equation (44), then one label update maps every $Z \in \mathcal{Z}_0$ to $Z^\star$. A subsequent parameter update remains inside the assumed error set, so $Z^\star$ is label-stable. This is a conditional basin result rather than a global convergence theorem: verifying that a particular initialization belongs to $\mathcal{Z}_0$ requires a separate model for class-membership contamination and scatter perturbation.

S27 MARGIN-DISTRIBUTION AND ORACLE INEQUALITIES

Exact recovery can be too strong near ambiguous action boundaries. A softer statement controls the fraction of changed labels through the distribution of ideal margins. Let g_i^{temp} be the ideal margin and suppose $\max_k |\widehat{d}_{k,i} - d^\star_{k,i}| \leq E_i$.

Proposition 12 (Pointwise oracle inequality)**.** *Let $\widehat{z}_i = \arg\min_k \widehat{d}_{k,i}$ and $z_i^\star = \arg\min_k d^\star_{k,i}$. Then*

$$d^\star_{\widehat{z}_i,i} - d^\star_{z_i^\star,i} \leq 2E_i. \tag{45}$$

Proof. Optimality of $\widehat{z}_i$ gives $\widehat{d}_{\widehat{z}_i,i} \leq \widehat{d}_{z_i^\star,i}$. Hence

$$d^\star_{\widehat{z}_i,i} \leq \widehat{d}_{\widehat{z}_i,i} + E_i \leq \widehat{d}_{z_i^\star,i} + E_i \leq d^\star_{z_i^\star,i} + 2E_i.$$

□

Corollary 10 (Misclassification controlled by small margins)**.** *If $E_i \leq E$ for every sample, then*

$$\frac{1}{N}\sum_{i=1}^N \mathbb{I}(\widehat{z}_i \neq z_i^\star) \leq \frac{1}{N}\sum_{i=1}^N \mathbb{I}(g_i^{\mathrm{temp}} \leq 2E). \tag{46}$$

More generally, for any $q > 0$,

$$\frac{1}{N}\sum_i \mathbb{I}(\widehat{z}_i \neq z_i^\star) \leq \frac{1}{N}\sum_i \mathbb{I}(g_i^{\mathrm{temp}} \leq q) + \frac{1}{N}\sum_i \mathbb{I}(E_i > q/2).$$

Proof. A changed label contradicts Theorem 16 unless $g_i^{\mathrm{temp}} \le 2E_i$. The two displayed statements follow by set inclusion and thresholding. □

This result formalizes the qualitative observation that errors concentrate near action boundaries. If only a small fraction of frames have small temporal margins, global accuracy can remain high even though exact recovery fails on those frames.

Proposition 13 (Average excess temporal cost)**.** *For arbitrary nonnegative errors E_i,*

$$\frac{1}{N}\sum_i \left(d^\star_{\widehat{z}_i,i} - d^\star_{z_i^\star,i}\right) \le \frac{2}{N}\sum_i E_i.$$

Proof. Sum Equation (45) over samples. □

The oracle inequality is agnostic to how errors arise. It therefore applies to random masks, fixed structured masks, noisy coordinates, approximate subspaces, or any combination for which E_i can be bounded.

S28 Information-Theoretic Obstructions

The preceding sections give sufficient conditions. We now record elementary but sharp impossibility statements showing why some conditions cannot be removed.

Proposition 14 (Collision under a fixed mask)**.** *Let Φ be fixed. If two complete vectors $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, belonging to different target action classes, satisfy $\Phi\mathbf{x}^{(1)} = \Phi\mathbf{x}^{(2)}$, then no method using only $(\Phi, \Phi\mathbf{x})$ can distinguish them deterministically. Under a uniform prior on the two vectors, every estimator has error probability at least $1/2$.*

Proof. The two latent cases induce exactly the same observation. A deterministic estimator returns the same label in both cases and is wrong in one. A randomized estimator has two conditional success probabilities whose sum is at most one, so its average error is at least $1/2$. □

Theorem 17 (Indistinguishable restricted affine subspaces)**.** *Let $\mathcal{S}_1 = \mathrm{span}(\mathbf{U}_1) + \mathbf{o}_1$ and $\mathcal{S}_2 = \mathrm{span}(\mathbf{U}_2) + \mathbf{o}_2$. If*

$$\mathrm{span}(\Phi\mathbf{U}_1) = \mathrm{span}(\Phi\mathbf{U}_2), \qquad \Phi(\mathbf{o}_1 - \mathbf{o}_2) \in \mathrm{span}(\Phi\mathbf{U}_1),$$

then the two restricted affine sets are identical. Consequently, masked residuals cannot distinguish membership in $\mathcal{S}_1$ from membership in $\mathcal{S}_2$ for samples whose observed coordinates lie in that common restricted set.

Proof. The first equality gives a common restricted linear span L. The second condition says the restricted offsets differ by an element of L, so $L + \Phi\mathbf{o}_1 = L + \Phi\mathbf{o}_2$. Both masked distances to these affine sets are therefore identical. □

This theorem explains the role of the restricted Gram matrix but is stronger: even when each individual restricted basis is well-conditioned, two different complete subspaces may become identical after restriction. Assignment then requires temporal or prior information beyond the current observed vector.

Proposition 15 (Fully missing interval obstruction)**.** *Consider a contiguous interval I with $m_i = 0$ for all $i \in I$. Without an assumption on label transitions inside I, two label sequences that agree outside I but differ inside it induce identical observations. Under a uniform prior on those two sequences, every estimator has probability at least $1/2$ of returning the wrong labels on at least one index in I.*

Proof. No value observed inside I depends on its labels. The distributions of all available data are identical under the two label sequences, so the same two-point argument as Proposition 14 applies. □

Nearest-time label propagation therefore encodes a piecewise-constant label prior; it is not a consequence of the masked-residual theory. If a missing interval contains an action transition, the transition location is not identifiable from the interval itself.

Corollary 11 (Necessity of observed discriminating energy)**.** *Any uniform assignment theorem must exclude masks that remove every coordinate on which two candidate affine subspaces differ. Coherence, leverage, restricted conditioning, or an explicit deterministic certificate are alternative ways to express this necessary exclusion.*

S29 Monotonicity and Nonmonotonicity with Additional Coordinates

Let $\Omega_1 \subseteq \Omega_2$ be two observed coordinate sets, with masks $\Phi_1 \preceq \Phi_2$. Denote their unnormalized least-squares residuals by ξ_{Ω_1} and ξ_{Ω_2}.

Proposition 16 (Unnormalized residual monotonicity)**.** *For every fixed affine subspace,*

$$\xi_{\Omega_1} \leq \xi_{\Omega_2}. \tag{47}$$

Moreover, $\mathbf{U}^{\mathrm{T}}\Phi_1\mathbf{U} \preceq \mathbf{U}^{\mathrm{T}}\Phi_2\mathbf{U}$*, so the smallest restricted Gram eigenvalue is nondecreasing as coordinates are added.*

Proof. For any coefficient vector $\mathbf{v}$, the squared error over Ω_2 equals the squared error over Ω_1 plus nonnegative errors on $\Omega_2 \setminus \Omega_1$. Taking the minimum over $\mathbf{v}$ preserves the inequality. The Gram statement follows because $\mathbf{U}^{\mathrm{T}}(\Phi_2 - \Phi_1)\mathbf{U} \succeq 0$. □

The normalized residual $D\xi_\Omega/|\Omega|$ need not be monotone: the new coordinates can have residual energy above or below the previous average. This is desirable because normalization targets complete-data scale rather than raw observed error.

Proposition 17 (One-coordinate update identity)**.** *Assume* $\mathbf{G}_\Omega = \mathbf{U}_\Omega^{\mathrm{T}}\mathbf{U}_\Omega$ *is invertible. Add coordinate* $j \notin \Omega$*, with row* $\mathbf{u}_j^{\mathrm{T}}$*, response* $y_j = x_j - o_j$*, and let* $\widehat{\mathbf{a}}_\Omega = \mathbf{G}_\Omega^{-1}\mathbf{U}_\Omega^{\mathrm{T}}\mathbf{y}_\Omega$*. Then*

$$\xi_{\Omega\cup\{j\}} - \xi_\Omega = \frac{(y_j - \mathbf{u}_j^{\mathrm{T}}\widehat{\mathbf{a}}_\Omega)^2}{1 + \mathbf{u}_j^{\mathrm{T}}\mathbf{G}_\Omega^{-1}\mathbf{u}_j}. \tag{48}$$

Proof. Apply the Sherman–Morrison formula to the rank-one update $\mathbf{G}_{\Omega\cup\{j\}} = \mathbf{G}_\Omega + \mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}$ and simplify the difference between the two least-squares objective minima. □

Equation (48) quantifies the value of observing a coordinate. A large prediction error increases residual evidence, while high leverage in an already uncertain direction divides that increment. The identity also suggests an active-sensing variant that selects coordinates by expected assignment-margin gain, which is outside the present algorithm.

S30 Bias from Boundary-Crossing Neighborhoods

Temporal consistency is helpful only when a neighborhood mostly lies within one semantic segment. Fix sample i with ideal label ℓ. Partition its neighborhood into in-segment and contaminating indices, $\mathcal{N}_i = \mathcal{G}_i \cup \mathcal{H}_i$, and let $q_i = |\mathcal{H}_i|/|\mathcal{N}_i|$.

For a competitor $k \neq \ell$, define average residual gaps

$$\gamma_{i,k}^{\mathrm{self}} = \rho_{k,i} - \rho_{\ell,i},$$
$$\gamma_{i,k}^{\mathrm{good}} = |\mathcal{G}_i|^{-1} \sum_{j\in\mathcal{G}_i} (\rho_{k,j} - \rho_{\ell,j}),$$
$$\gamma_{i,k}^{\mathrm{bad}} = |\mathcal{H}_i|^{-1} \sum_{j\in\mathcal{H}_i} (\rho_{\ell,j} - \rho_{k,j}).$$

The last quantity measures how strongly contaminating neighbors favor the competitor.

Theorem 18 (Tolerance to neighborhood contamination)**.** *If $\gamma_{i,k}^{\text{self}} > 0$, $\gamma_{i,k}^{\text{good}} > 0$, and $\gamma_{i,k}^{\text{bad}} \geq 0$, then sample i still prefers ℓ to k whenever*

$$\gamma_{i,k}^{\text{self}} + \lambda \left[(1 - q_i)\gamma_{i,k}^{\text{good}} - q_i \gamma_{i,k}^{\text{bad}} \right] > 0. \tag{49}$$

In particular, a sufficient condition is

$$q_i < \frac{\gamma_{i,k}^{\text{self}}/\lambda + \gamma_{i,k}^{\text{good}}}{\gamma_{i,k}^{\text{good}} + \gamma_{i,k}^{\text{bad}}} \qquad (\lambda > 0). \tag{50}$$

Proof. Expand $d_{k,i} - d_{\ell,i}$ and split the neighborhood sum over $\mathcal{G}_i$ and $\mathcal{H}_i$. The result is exactly the left-hand side of Equation (49). Solving the strict inequality for q_i gives Equation (50). □

The theorem captures the bias–variance tradeoff in λ. Increasing λ amplifies supportive within-segment evidence, but also amplifies boundary contamination. Adaptive neighborhoods aim to keep q_i small; the boundary-stability theory explains when that construction succeeds.

S31 Bias and Variance of Normalized Observed Energy

The concentration results above control tail probabilities. This section complements them with exact first- and second-moment calculations, making explicit why the factor D/m is the correct normalization and where the remaining downward bias of an observed least-squares residual originates.

Fix a deterministic vector $\mathbf{r} \in \mathbb{R}^D$, let $a_j = r_j^2$, and draw Ω uniformly without replacement from all subsets of size m. Define the normalized observed energy

$$Q_\Omega(\mathbf{r}) = \frac{D}{m} \sum_{j \in \Omega} r_j^2. \tag{51}$$

Proposition 18 (Exact moments under fixed-size sampling)**.** *Let $\bar{a} = D^{-1} \sum_{j=1}^D a_j$ and $S_a^2 = (D-1)^{-1} \sum_{j=1}^D (a_j - \bar{a})^2$. Then*

$$\mathbb{E} Q_\Omega(\mathbf{r}) = \|\mathbf{r}\|_2^2, \tag{52}$$

$$\operatorname{Var}[Q_\Omega(\mathbf{r})] = \frac{D^2}{m} \left(1 - \frac{m}{D}\right) S_a^2. \tag{53}$$

Consequently,

$$\frac{\sqrt{\operatorname{Var}[Q_\Omega(\mathbf{r})]}}{\|\mathbf{r}\|_2^2} = \sqrt{\frac{D-m}{m(D-1)}} \, \frac{\sqrt{D \sum_j r_j^4 - \|\mathbf{r}\|_2^4}}{\|\mathbf{r}\|_2^2}. \tag{54}$$

Proof. The sample mean $m^{-1} \sum_{j \in \Omega} a_j$ is unbiased for $\bar{a}$. Its variance under sampling without replacement is $(1 - m/D) S_a^2/m$. Multiplying the mean by D proves Equations (52)–(53); expanding S_a^2 gives Equation (54). □

Equation (54) separates the sampling fraction from the spatial concentration of the residual. A diffuse residual has small fourth-moment ratio and is easy to estimate from few coordinates. A residual concentrated on a small coordinate set is intrinsically fragile: a mask that misses that set contains almost no evidence about the complete residual.

For a candidate affine subspace with basis $\mathbf{U}$, let $\mathbf{r} = \mathbf{P}_{\mathbf{U}}^{\perp}(\mathbf{x} - \mathbf{o})$, $\mathbf{G}_\Omega = \mathbf{U}_\Omega^{\mathrm{T}} \mathbf{U}_\Omega$, and $\mathbf{c}_\Omega = \mathbf{U}_\Omega^{\mathrm{T}} \mathbf{r}_\Omega$. The exact deterministic identity from Section S16 gives

$$\rho_\Omega = Q_\Omega(\mathbf{r}) - \frac{D}{m} \mathbf{c}_\Omega^{\mathrm{T}} \mathbf{G}_\Omega^{-1} \mathbf{c}_\Omega. \tag{55}$$

The second term is nonnegative and is the price paid for re-fitting the subspace coefficients on the observed coordinates.

Corollary 12 (Downward-bias decomposition)**.** *Whenever $\mathbf{G}_\Omega$ is invertible,*

$$0 \le \|\mathbf{r}\|_2^2 - \mathbb{E}\rho_\Omega = \mathbb{E}\left[\frac{D}{m}\mathbf{c}_\Omega^{\mathrm{T}}\mathbf{G}_\Omega^{-1}\mathbf{c}_\Omega\right]. \tag{56}$$

On the event $\lambda_{\min}(\mathbf{G}_\Omega) \ge p(1-\eta)$, where $p = m/D$, the realized leakage satisfies

$$0 \le Q_\Omega(\mathbf{r}) - \rho_\Omega \le \frac{\|\mathbf{c}_\Omega\|_2^2}{p^2(1-\eta)}. \tag{57}$$

Thus the normalized energy itself is exactly unbiased; the only systematic bias comes from the interaction between the mask and the candidate subspace. The cross-term concentration event in the main theorem is therefore not a technical artifact: it controls precisely the coefficient-refitting leakage.

S32 Repeated Masks, Aggregation, and Robust Confidence Amplification

Some acquisition systems provide several partial views of the same latent frame, for example repeated sensor sweeps or several independently dropped feature packets. Independent masks permit confidence amplification without requiring any individual view to be dense.

Let $\Omega^{(1)}, \dots, \Omega^{(L)}$ be independent masks and let $Q_s = Q_{\Omega^{(s)}}(\mathbf{r})$. Define $\overline{Q}_L = L^{-1}\sum_{s=1}^{L} Q_s$.

Proposition 19 (Variance reduction by repeated views)**.** *Under uniform fixed-size sampling with the same m in every view,*

$$\mathbb{E}\overline{Q}_L = \|\mathbf{r}\|_2^2, \qquad \mathrm{Var}(\overline{Q}_L) = \frac{D^2}{Lm}\left(1 - \frac{m}{D}\right) S_a^2. \tag{58}$$

In particular, the standard deviation decreases as $L^{-1/2}$.

The average is efficient under the nominal mask model, but a few corrupted views can dominate it. A median or median-of-means aggregate provides a more robust alternative. The following result only assumes a one-view guarantee and hence applies to normalized subspace residuals as well as to raw energy.

Theorem 19 (Median amplification)**.** *Let $Z_1, \dots, Z_L$ be independent residual estimates satisfying*

$$\Pr\{|Z_s - \varepsilon| > t\} \le q < \frac{1}{2} \qquad \textit{for every } s. \tag{59}$$

If L is odd and $Z_{\mathrm{med}} = \mathrm{median}_s\, Z_s$, then

$$\Pr\{|Z_{\mathrm{med}} - \varepsilon| > t\} \le \exp\left[-2L\left(\frac{1}{2} - q\right)^2\right]. \tag{60}$$

Proof. The median can fail only if at least $(L+1)/2$ individual estimates fail. The failure indicators are independent Bernoulli variables with means at most q. Hoeffding's inequality for their sum yields Equation (60). □

Corollary 13 (Repeated-view assignment preservation)**.** *Suppose every candidate residual obeys $\Pr\{|Z_{k,s} - \varepsilon_k| > t\} \le q < 1/2$. If the complete-data margin satisfies $\varepsilon_k - \varepsilon_\ell > 2t$ for all $k \ne \ell$, then the median-residual rule selects ℓ with probability at least*

$$1 - K\exp\left[-2L\left(\frac{1}{2} - q\right)^2\right]. \tag{61}$$

The conclusion follows by a union bound over candidates. The important distinction is that more views improve confidence exponentially in L, whereas increasing the number of observed coordinates improves the accuracy within each view. These are complementary acquisition resources.

S33 Optimal Observation Probabilities and Active Mask Design

The heterogeneous-sampling theory raises a natural design question: given an expected observation budget, which coordinates should be acquired? Although the implemented method passively accepts the supplied masks, the following oracle calculation identifies the variance-optimal benchmark.

Let $\delta_j \sim \mathrm{Bernoulli}(\pi_j)$ independently and consider the Horvitz–Thompson energy estimator

$$\widehat{E}_{\boldsymbol{\pi}}(\mathbf{r}) = \sum_{j=1}^{D} \frac{\delta_j r_j^2}{\pi_j}, \qquad 0 < \pi_j \le 1, \qquad \sum_{j=1}^{D} \pi_j = m. \tag{62}$$

It is unbiased and has variance

$$V(\boldsymbol{\pi}; \mathbf{r}) = \sum_{j=1}^{D} r_j^4 \frac{1-\pi_j}{\pi_j}. \tag{63}$$

Theorem 20 (Oracle variance-optimal mask)**.** *For a fixed residual* $\mathbf{r}$ *with* $r_j \neq 0$ *for all* j*, and* $0 < m < D$*, every solution of* $\min_{\boldsymbol{\pi}} V(\boldsymbol{\pi}; \mathbf{r})$ *under the constraints in Equation* (62) *has the water-filling form*

$$\pi_j^\star = \min\{1, c\, r_j^2\}, \tag{64}$$

where $c > 0$ *is chosen so that* $\sum_j \pi_j^\star = m$*. If no probability is clipped at one, then*

$$\pi_j^\star = m \frac{r_j^2}{\|\mathbf{r}\|_2^2}, \qquad V(\boldsymbol{\pi}^\star; \mathbf{r}) = \frac{\|\mathbf{r}\|_2^4}{m} - \sum_{j=1}^{D} r_j^4. \tag{65}$$

Proof. The additive constant $-\sum_j r_j^4$ in Equation (63) does not affect the optimizer, so it suffices to minimize $\sum_j r_j^4/\pi_j$. For every coordinate not at its upper bound, the Karush–Kuhn–Tucker stationarity equation is $-r_j^4/\pi_j^2 + \nu = 0$, hence $\pi_j = r_j^2/\sqrt{\nu}$. Enforcing the upper bound gives Equation (64); enforcing the budget in the unclipped case gives Equation (65). □

The oracle observes coordinates in proportion to their residual energy, but that residual is unknown before acquisition. A model-aware design must also preserve the geometry of every candidate subspace. Let $h_{k,j} = \|\mathbf{e}_j^{\mathrm{T}} \mathbf{U}_k\|_2^2$ be the row leverage score and define the conservative importance score

$$s_j = \alpha \max_k h_{k,j} + (1-\alpha) \max_{k,k'} \left| [(\mathbf{P}_{\mathbf{U}_k}^{\perp} - \mathbf{P}_{\mathbf{U}_{k'}}^{\perp})\mathbf{v}]_j \right|^2, \tag{66}$$

where $\mathbf{v}$ is a pilot estimate and $\alpha \in [0,1]$. Sampling proportionally to s_j, with a nonzero floor, simultaneously discourages Gram-matrix singularity and emphasizes coordinates likely to distinguish candidate assignments. Equation (66) is a theoretical design principle rather than a component of the reported passive algorithm.

Proposition 20 (Minimax envelope design)**.** *Suppose a known positive envelope* $b_j > 0$ *satisfies* $r_j^2 \le b_j$ *for every residual in a target class. Among independent Bernoulli designs of expected size* m*, the probabilities*

$$\pi_j = \min\{1, c b_j\}, \qquad \sum_j \pi_j = m, \tag{67}$$

minimize the worst-case upper bound $\sum_j b_j^2(1-\pi_j)/\pi_j$ *obtained from Equation* (63)*.*

The proof repeats the KKT argument with b_j in place of r_j^2. This result formalizes the role of leverage-score envelopes in a setting where the exact residual is not known.

S34 Time-Varying Observation Rates

Real sequences often have a nonstationary missing rate. Raw observed residuals then confound semantic fit with the number of available coordinates. The sample-specific normalization D/m_i removes the leading multiplicative effect, but a time-varying sampling fraction still changes uncertainty.

Let $p_i = m_i/D$ and suppose the complete residual for candidate k is $\varepsilon_{k,i}$. Write the normalized estimate as

$$\rho_{k,i} = \varepsilon_{k,i} + e_{k,i}, \qquad |e_{k,i}| \le \tau_i \varepsilon_{k,i} + \beta_i, \tag{68}$$

where τ_i decreases with $p_i D$ and β_i collects additive noise and subspace-estimation error.

Theorem 21 (Assignment under heteroscedastic mask uncertainty)**.** *Let ℓ_i be the complete-data nearest candidate and define $\Delta_{k,i} = \varepsilon_{k,i} - \varepsilon_{\ell_i,i} > 0$. The normalized rule preserves ℓ_i if, for every $k \neq \ell_i$,*

$$\Delta_{k,i} > \tau_i(\varepsilon_{k,i} + \varepsilon_{\ell_i,i}) + 2\beta_i. \tag{69}$$

For temporal costs with row-stochastic weights a_{ij}, it suffices that

$$\Delta_{k,i} + \lambda \sum_j a_{ij}(\varepsilon_{k,j} - \varepsilon_{\ell_i,j}) > \tau_i(\varepsilon_{k,i} + \varepsilon_{\ell_i,i}) + 2\beta_i + \lambda \sum_j a_{ij} \left[\tau_j(\varepsilon_{k,j} + \varepsilon_{\ell_i,j}) + 2\beta_j\right]. \tag{70}$$

Proof. For a direct assignment, expand $\rho_{k,i} - \rho_{\ell_i,i}$ and lower-bound the two perturbations using Equation (68). For the temporal rule, apply the same bound to the current sample and to every weighted neighbor. Positivity of the resulting lower bound proves the claim. □

The theorem gives a precise interpretation of temporal borrowing. A sparsely observed sample can be stabilized by well-observed neighbors with large semantic margins. Conversely, poorly observed or boundary-crossing neighbors add uncertainty. This suggests confidence-adaptive weights

$$\widetilde{a}_{ij} = \frac{a_{ij} w_j}{\sum_t a_{it} w_t}, \qquad w_j = \left(\tau_j \widehat{E}_j + \beta_j + \epsilon_0\right)^{-1}, \tag{71}$$

where $\widehat{E}_j$ is a scale estimate and $\epsilon_0 > 0$ prevents division by zero. The present experiments keep the simpler uniform neighborhood average so that gains cannot be attributed to an extra learned confidence mechanism.

S35 Operator View of Temporal Smoothing

Let $\mathbf{R} \in \mathbb{R}^{N \times K}$ contain the normalized residuals, and let $\mathbf{A}$ be the nonnegative row-stochastic neighborhood matrix, so $A_{ij} = |\mathcal{N}_i|^{-1}$ for $j \in \mathcal{N}_i$. For this section assume nonempty neighborhoods; allowing zero rows instead gives a row-substochastic matrix and the same norm upper bounds. The temporally augmented cost matrix is

$$\mathbf{D}_\lambda = (\mathbf{I} + \lambda \mathbf{A})\mathbf{R}. \tag{72}$$

This form separates the geometry module, which produces $\mathbf{R}$, from the sequence module, which applies a linear operator before rowwise decisions.

Proposition 21 (Lipschitz stability of temporal costs)**.** *For two residual matrices $\mathbf{R}$ and $\widetilde{\mathbf{R}}$,*

$$\|\mathbf{D}_\lambda - \widetilde{\mathbf{D}}_\lambda\|_{\max} \le (1+\lambda)\|\mathbf{R} - \widetilde{\mathbf{R}}\|_{\max}. \tag{73}$$

If $\mathbf{A}$ is symmetric, then

$$\|\mathbf{D}_\lambda - \widetilde{\mathbf{D}}_\lambda\|_{\mathrm{F}} \le (1+\lambda)\|\mathbf{R} - \widetilde{\mathbf{R}}\|_{\mathrm{F}}. \tag{74}$$

Proof. Every output entry is one input entry plus λ times a convex combination of neighboring entries, proving Equation (73). If $\mathbf{A}$ is symmetric and row-stochastic, its spectral norm is one. Submultiplicativity then gives $\|\mathbf{I} + \lambda\mathbf{A}\|_2 \le 1 + \lambda$, which proves the Frobenius bound. □

Corollary 14 (Decision stability from operator perturbations)**.** *If the smallest rowwise cost gap of $\mathbf{D}_\lambda$ is*

$$g_{\min} = \min_i \min_{k \neq \ell_i} (D_{\lambda,ik} - D_{\lambda,i\ell_i}) > 0, \tag{75}$$

then every perturbation satisfying

$$\|\mathbf{R} - \widetilde{\mathbf{R}}\|_{\max} < \frac{g_{\min}}{2(1+\lambda)} \tag{76}$$

preserves all assignments.

For symmetric $\mathbf{A}$, let $\mathbf{A} = \mathbf{V}\,\mathrm{diag}(\mu_1, \ldots, \mu_N)\mathbf{V}^{\mathrm{T}}$. Then the temporal operator multiplies graph-frequency component s by $1 + \lambda\mu_s$. For $0 \le \lambda \le 1$, negative eigenmodes are attenuated, while positive smooth modes are amplified. For larger λ, negative modes may change sign and may also be amplified in absolute value. The rowwise argmin depends on relative candidate costs, so the useful effect is better understood as margin aggregation over semantically coherent neighborhoods, not generic signal smoothing.

Proposition 22 (Repeated temporal aggregation)**.** *For $t \ge 1$, define $\mathbf{R}^{(t)} = (\mathbf{I} + \lambda\mathbf{A})^t\mathbf{R}$. Then*

$$\|\mathbf{R}^{(t)} - \widetilde{\mathbf{R}}^{(t)}\|_{\max} \le (1+\lambda)^t \|\mathbf{R} - \widetilde{\mathbf{R}}\|_{\max}. \tag{77}$$

Thus repeated unnormalized aggregation can exponentially magnify both margins and errors. A normalized diffusion $\left[(\mathbf{I} + \lambda\mathbf{A})/(1+\lambda)\right]^t$ is nonexpansive in the max norm.

This observation explains why TSCC uses a single, interpretable neighborhood term rather than an arbitrarily deep propagation stack.

S36 Uniform Guarantees over a Continuous Family of Affine Subspaces

The finite-sequence theorem takes a union bound over the NK residuals actually evaluated by the algorithm. A stronger question is whether one mask simultaneously preserves residuals for every member of a continuous model class. This distinction matters when subspaces are updated from the same partially observed data: the final candidate is data dependent and cannot always be treated as one fixed element chosen in advance.

Let $\mathfrak{S}_{d,R}$ be the set of affine d-dimensional subspaces $\mathcal{S} = (\mathbf{U}, \mathbf{o})$ with orthonormal $\mathbf{U}$, $\|\mathbf{o}\|_2 \le R$, and the canonical constraint $\mathbf{U}^{\mathrm{T}}\mathbf{o} = 0$. Equip this class with the metric

$$d_{\mathfrak{S}}(\mathcal{S}, \widetilde{\mathcal{S}}) = \|\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\widetilde{\mathbf{U}}}\|_2 + \frac{\|\mathbf{o} - \widetilde{\mathbf{o}}\|_2}{R}. \tag{78}$$

For $\|\mathbf{x}\|_2 \le R$, write $\varepsilon_{\mathcal{S}}(\mathbf{x}) = \mathrm{dist}^2(\mathbf{x}, \mathcal{S})$.

Lemma 4 (Lipschitz continuity of complete residuals)**.** *If $d_{\mathfrak{S}}(\mathcal{S}, \widetilde{\mathcal{S}}) \le \eta \le 1$, then*

$$\left|\varepsilon_{\mathcal{S}}(\mathbf{x}) - \varepsilon_{\widetilde{\mathcal{S}}}(\mathbf{x})\right| \le 8R^2\eta. \tag{79}$$

Proof. Set $\mathbf{a} = \mathbf{x} - \mathbf{o}$ and $\widetilde{\mathbf{a}} = \mathbf{x} - \widetilde{\mathbf{o}}$. Both norms are at most $2R$. Insert and subtract $\mathbf{a}^{\mathrm{T}}\mathbf{P}_{\widetilde{\mathbf{U}}}^{\perp}\mathbf{a}$. The projector perturbation contributes at most $4R^2\eta$. The offset perturbation contributes at most $\|\mathbf{a} - \widetilde{\mathbf{a}}\|_2(\|\mathbf{a}\|_2 + \|\widetilde{\mathbf{a}}\|_2) \le 4R^2\eta$. Adding the terms proves the bound. □

For clarity, define the good conditioning subset $\mathfrak{S}_{d,R}(\Phi, \kappa)$ by requiring

$$\lambda_{\min}(\mathbf{U}^{\mathrm{T}}\Phi\mathbf{U}) \ge \kappa p, \qquad p = \mathrm{tr}(\Phi)/D, \tag{80}$$

for every model under consideration. On this set, the inverse observed Gram matrix is uniformly bounded. The deterministic residual formula and standard inverse perturbation identity then imply a finite constant $L_{\mathrm{obs}}(R, \kappa, p)$ such that

$$|\rho_{\mathcal{S},\Phi}(\mathbf{x}) - \rho_{\widetilde{\mathcal{S}},\Phi}(\mathbf{x})| \le L_{\mathrm{obs}} d_{\mathfrak{S}}(\mathcal{S}, \widetilde{\mathcal{S}}). \tag{81}$$

The dependence on κ^{-1} is unavoidable: near a singular restricted basis, an arbitrarily small rotation can cause a large change in the fitted coefficient.

Proposition 23 (Metric entropy of bounded affine subspaces)**.** *There is a universal constant $C > 0$ such that, for $0 < \eta < 1$, the class $\mathfrak{S}_{d,R}$ admits an η-net under Equation* (78) *with cardinality*

$$\mathcal{N}(\eta, \mathfrak{S}_{d,R}, d_{\mathfrak{S}}) \le \left(\frac{C}{\eta}\right)^{d(D-d)+D}. \tag{82}$$

Proof sketch. The Grassmann manifold $\mathrm{Gr}(d, D)$ has intrinsic dimension $d(D-d)$ and admits a projector-metric net of size at most $(C/\eta)^{d(D-d)}$. A Euclidean $R\eta$-net of the radius-R offset ball has size at most $(C/\eta)^D$. Taking their Cartesian product and then projecting each offset to the canonical orthogonal complement changes only the universal constant. □

Theorem 22 (Net-to-class transfer)**.** *Let $\mathcal{T}_\eta$ be an η-net of a model class $\mathfrak{S}$. Suppose that, on an event $\mathcal{E}$,*

$$\sup_{\mathcal{S}_0 \in \mathcal{T}_\eta} |\rho_{\mathcal{S}_0,\Phi}(\mathbf{x}) - \varepsilon_{\mathcal{S}_0}(\mathbf{x})| \le t. \tag{83}$$

If complete and observed residuals are respectively L_{c}- and L_{o}-Lipschitz on $\mathfrak{S}$, then, on the same event,

$$\sup_{\mathcal{S} \in \mathfrak{S}} |\rho_{\mathcal{S},\Phi}(\mathbf{x}) - \varepsilon_{\mathcal{S}}(\mathbf{x})| \le t + (L_{\mathrm{c}} + L_{\mathrm{o}})\eta. \tag{84}$$

Proof. For any $\mathcal{S}$, choose $\mathcal{S}_0 \in \mathcal{T}_\eta$ with $d_{\mathfrak{S}}(\mathcal{S}, \mathcal{S}_0) \le \eta$. Insert and subtract the observed and complete residuals of $\mathcal{S}_0$, then apply the triangle inequality, Equation (83), and the two Lipschitz bounds. □

Corollary 15 (Model-entropy penalty)**.** *Assume a fixed-model residual bound fails with probability at most $2\exp(-cmt^2/B^2)$, uniformly over the class, for constants $c, B > 0$. Then an η-net and a union bound give uniform failure probability at most*

$$2\exp\left\{-\frac{cmt^2}{B^2} + [d(D-d) + D]\log\frac{C}{\eta}\right\}. \tag{85}$$

Taking $\eta = t/(L_{\mathrm{c}} + L_{\mathrm{o}})$, a sufficient observation budget for uniform error at most $2t$ with probability $1-\delta$ is

$$m \ge \frac{B^2}{ct^2}\left([d(D-d) + D]\log\frac{C(L_{\mathrm{c}} + L_{\mathrm{o}})}{t} + \log\frac{2}{\delta}\right). \tag{86}$$

For an unconditional random-mask conclusion, the net must be deterministic and the required Lipschitz and conditioning bounds must hold on a uniform good event. Its failure probability must be added to the net failure probability. A mask-dependent well-conditioned subclass cannot be inserted into this union bound without separately justifying that selection.

The extra term $d(D-d)$ is the geometric price of uniformity over an unknown subspace, while the D-term accounts for an unrestricted affine offset. If offsets are known, shared, sparse, or constrained to a lower dimensional family, that part of the entropy decreases accordingly. For the finite-candidate setting, the NK result is typically sharper because it only pays for candidates fixed independently of the tested masks. Conditioning on candidates learned from those same masks does not justify the fixed-candidate theorem. The continuous-class bound is useful for analyzing data-dependent model updates and clarifies the gap between pointwise residual preservation and full end-to-end learning.

S37 A Sampling–Coherence–Margin Phase Diagram

The preceding results isolate three quantities that govern recoverability: the number of observed coordinates, the coherence of the candidate subspaces and residuals, and the assignment margin. This section combines them into a compact phase diagram and derives an a posteriori certificate that can be evaluated from the observed residuals.

For sample i, let ℓ_i be its complete-data winner and define the relative pairwise separation

$$\gamma_{k,i} = \frac{\varepsilon_{k,i} - \varepsilon_{\ell_i,i}}{\varepsilon_{k,i} + \varepsilon_{\ell_i,i}}, \qquad \gamma_{\min} = \min_i \min_{k \ne \ell_i} \gamma_{k,i}. \tag{87}$$

When both residuals vanish, the candidates are indistinguishable at that sample and the ratio is defined as zero. Thus $\gamma_{\min} > 0$ is a strict identifiability condition.

Assume a generic multiplicative residual tail bound of the form

$$\Pr\{|\rho_{k,i} - \varepsilon_{k,i}| > \tau\varepsilon_{k,i}\} \le 2\exp\left(-\frac{cm_i\tau^2}{\mu_{k,i}d_k}\right), \qquad 0 < \tau < 1, \tag{88}$$

where $\mu_{k,i} \geq 1$ is a model-dependent effective concentration parameter, and $c > 0$ is a constant. This is an additional tail assumption used only for the conditional calculation in this section. It does not follow merely by taking the larger of the two coherences in Theorem 1, which has a separate product-coherence term. Equation (7) gives the directly derived observation budget without this extra assumption.

Theorem 23 (Exact sequence assignment above the phase boundary)**.** *Let* $m_{\min} = \min_i m_i$, $d_{\max} = \max_k d_k$, *and* $\mu_{\max} = \max_{i,k} \mu_{k,i}$. *Under Equation* (88)*, if* $\gamma_{\min} > 0$ *and*

$$m_{\min} \geq \frac{4\mu_{\max} d_{\max}}{c\gamma_{\min}^2} \log \frac{2NK}{\delta}, \tag{89}$$

then all nearest-subspace assignments are preserved with probability at least $1 - \delta$.

Proof. Set $\tau = \gamma_{\min}/2$. A union bound over all NK residuals and Equation (89) imply that, with probability at least $1 - \delta$, $|\rho_{k,i} - \varepsilon_{k,i}| \leq \tau\varepsilon_{k,i}$ simultaneously. For every competitor,

$$\begin{aligned} \rho_{k,i} - \rho_{\ell_i,i} &\geq \varepsilon_{k,i} - \varepsilon_{\ell_i,i} - \tau(\varepsilon_{k,i} + \varepsilon_{\ell_i,i}) \\ &= (\varepsilon_{k,i} + \varepsilon_{\ell_i,i})(\gamma_{k,i} - \tau) > 0. \end{aligned} \tag{90}$$

Hence every complete-data winner remains the observed-data winner. □

Equation (89) identifies three regimes. Large margin and low coherence permit aggressive missingness. Small margin causes a quadratic increase in the required observations. High coherence makes the task fragile even when the intrinsic dimension is small, because important evidence is concentrated in a few coordinates. The logarithmic NK factor is the price of simultaneous reliability across a sequence and all candidates.

Corollary 16 (Nonuniform per-frame allocation)**.** *Let* $\gamma_i = \min_{k \neq \ell_i} \gamma_{k,i}$, $\mu_i = \max_k \mu_{k,i}$, *and choose failure budgets* $\delta_i > 0$ *satisfying* $\sum_i \delta_i \leq \delta$. *It is sufficient to allocate*

$$m_i \geq \frac{4\mu_i d_{\max}}{c\gamma_i^2} \log \frac{2K}{\delta_i} \tag{91}$$

observations to frame i. *Therefore difficult low-margin or high-coherence frames should receive a larger acquisition budget than easy interior frames.*

The allocation is oracle because γ_i is unknown before complete observation. In practice, temporal boundary scores and pilot residual gaps can serve as proxies: allocate more coordinates near uncertain boundaries and fewer coordinates deep inside stable segments.

S37.1 An a posteriori residual certificate

The phase boundary uses complete residuals and is therefore explanatory, not directly checkable. A confidence interval reverses the multiplicative bound. On the event $|\rho_{k,i} - \varepsilon_{k,i}| \leq \tau\varepsilon_{k,i}$,

$$\underline{\varepsilon}_{k,i} := \frac{\rho_{k,i}}{1+\tau} \leq \varepsilon_{k,i} \leq \frac{\rho_{k,i}}{1-\tau} =: \overline{\varepsilon}_{k,i}. \tag{92}$$

Proposition 24 (Certified assignment or abstention)**.** *Let* $\widehat{\ell}_i = \arg\min_k \rho_{k,i}$. *If*

$$\overline{\varepsilon}_{\widehat{\ell}_i,i} < \min_{k \neq \widehat{\ell}_i} \underline{\varepsilon}_{k,i}, \tag{93}$$

then $\widehat{\ell}_i$ *is the unique complete-data winner on the simultaneous concentration event. Under Equation* (88)*, the probability of an incorrect certified label is at most*

$$2\sum_{k=1}^{K} \exp\left(-\frac{cm_i\tau^2}{\mu_{k,i} d_k}\right). \tag{94}$$

Proof. Equation (92) places every complete residual inside its reported interval. Condition (93) makes the winner's upper endpoint smaller than every competitor's lower endpoint, so no admissible complete residual vector can reverse the ranking. A union bound over candidates gives Equation (94). □

S37.2 TEMPORAL MARGINS CAN MOVE THE PHASE BOUNDARY

For row-stochastic neighborhood weights a_{ij}, define the complete temporally aggregated cost and its pairwise scale by

$$E^{\lambda}_{k,i} = \varepsilon_{k,i} + \lambda \sum_j a_{ij}\varepsilon_{k,j}, \tag{95}$$

$$B^{\lambda}_{k,i} = \varepsilon_{k,i} + \varepsilon_{\ell_i,i} + \lambda \sum_j a_{ij}(\varepsilon_{k,j} + \varepsilon_{\ell_i,j}). \tag{96}$$

The corresponding effective relative separation is

$$\gamma^{\lambda}_{k,i} = \frac{E^{\lambda}_{k,i} - E^{\lambda}_{\ell_i,i}}{B^{\lambda}_{k,i}}. \tag{97}$$

Proposition 25 (Temporal phase boundary)**.** *On the simultaneous event $|\rho_{k,i} - \varepsilon_{k,i}| \le \tau\varepsilon_{k,i}$, the temporal assignment of sample i is preserved whenever*

$$\tau < \min_{k \neq \ell_i} \gamma^{\lambda}_{k,i}. \tag{98}$$

Consequently, Theorem 23 remains valid after replacing $\gamma_{\min}$ by $\gamma^{\lambda}_{\min} = min_{i,k\neq\ell_i}\gamma^{\lambda}_{k,i}$, provided this quantity is positive.

Proof. Let $D^{\lambda}_{k,i}$ be the observed temporal cost. The simultaneous multiplicative event and nonnegativity of a_{ij} imply

$$D^{\lambda}_{k,i} - D^{\lambda}_{\ell_i,i} \ge E^{\lambda}_{k,i} - E^{\lambda}_{\ell_i,i} - \tau B^{\lambda}_{k,i} = B^{\lambda}_{k,i}(\gamma^{\lambda}_{k,i} - \tau). \tag{99}$$

The right-hand side is positive under Equation (98). The probability statement uses the same union bound as Theorem 23. □

This result gives a precise criterion for when temporal context reduces the required observation budget. If the aggregated neighbor gap-to-scale ratio exceeds the self ratio in the critical comparisons, their relative margins improve, potentially permitting fewer observed coordinates. Shared winning labels alone do not guarantee this improvement. If neighborhoods cross an action boundary, the numerator in Equation (97) can shrink or become negative; no increase in λ can then be justified by concentration alone. The boundary-detection and contamination results earlier in the appendix provide conditions that prevent this adverse regime.

If the intervals overlap, the procedure abstains rather than claiming that the assignment is wrong. This distinction is useful in safety-sensitive motion analysis: missingness may make a frame uncertifiable even when the predicted label happens to be correct. Reporting the certified coverage, the accuracy within certified frames, and the abstention rate gives more information than a single average partition-agreement score.

S38 FINITE TERMINATION UNDER QUANTIZED SUFFICIENT DECREASE

The monotone-descent statement permits an infinite sequence of vanishing improvements in continuous parameters. Practical implementations terminate when decreases fall below numerical tolerances. The following result makes that safeguard explicit.

Let $\mathcal{J}^{(t)} \ge 0$ be the objective at the beginning of outer iteration t. Accept a discrete label or neighborhood update only if it decreases the objective by at least $\eta_{\mathrm{d}} > 0$, and accept a continuous subspace update only if it decreases the objective by at least $\eta_{\mathrm{c}} > 0$; otherwise retain the current block.

Theorem 24 (Finite number of accepted updates)**.** *If every accepted update obeys the sufficient-decrease rule above, then the total number T_{acc} of accepted block updates satisfies*

$$T_{\mathrm{acc}} \le \left\lfloor \frac{\mathcal{J}^{(0)}}{\min\{\eta_{\mathrm{d}}, \eta_{\mathrm{c}}\}} \right\rfloor. \tag{100}$$

After that many accepted updates, the algorithm reaches a state in which no tested block yields the prescribed numerical improvement.

Proof. If T_{acc} updates are accepted, telescoping their decreases gives $\mathcal{J}^{(0)} - \mathcal{J}^{(T)} \geq T_{\mathrm{acc}} \min\{\eta_{\mathrm{d}}, \eta_{\mathrm{c}}\}$. Because $\mathcal{J}^{(T)} \geq 0$, rearrangement proves Equation (100). □

Corollary 17 (Finite label changes)**.** *If each accepted label sweep changes at least one label and must decrease the objective by η_{d}, then the number of accepted label sweeps is at most $\lfloor \mathcal{J}^{(0)}/\eta_{\mathrm{d}} \rfloor$. Since the label space itself has cardinality K^N, but a label vector can recur with different continuous parameters. Strict decrease excludes cycling of the full state, not recurrence of the label vector by itself.*

The theorem concerns a tolerance-safeguarded implementation, not the exact arithmetic limit of alternating minimization. Its role is to connect the mathematical descent property to a reproducible stopping rule: report η_{d}, η_{c}, the maximum number of sweeps, and the final relative objective change.

S39 Perturbation Analysis of the Surrogate Subspace Update

Let $\mathbf{S}_k^\star$ denote an ideal complete-data weighted scatter matrix for action class k, and let $\widehat{\mathbf{S}}_k$ be the pairwise-observed surrogate used by TSCC. Write $\mathbf{E}_k = \widehat{\mathbf{S}}_k - \mathbf{S}_k^\star$, and suppose $\mathbf{S}_k^\star$ has eigenvalues $\sigma_{k,1} \geq \cdots \geq \sigma_{k,D}$.

Theorem 25 (Subspace stability under scatter perturbation)**.** *Let $\mathbf{U}_k^\star$ contain the leading d_k eigenvectors of $\mathbf{S}_k^\star$, and let $\widehat{\mathbf{U}}_k$ contain the leading d_k eigenvectors of $\widehat{\mathbf{S}}_k$. If the population gap $g_k = \sigma_{k,d_k} - \sigma_{k,d_k+1} > 0$ and $\|\mathbf{E}_k\|_2 < g_k$, then*

$$\|\sin\Theta(\widehat{\mathbf{U}}_k, \mathbf{U}_k^\star)\|_2 \leq \frac{\|\mathbf{E}_k\|_2}{g_k - \|\mathbf{E}_k\|_2}. \tag{101}$$

In particular, if $\|\mathbf{E}_k\|_2 \leq g_k/2$, the right-hand side is at most $2\|\mathbf{E}_k\|_2/g_k$.

Proof. Weyl's inequality implies that the perturbed leading and trailing spectra remain separated by at least $g_k - \|\mathbf{E}_k\|_2$. Applying the Davis–Kahan sin-theta theorem to the corresponding invariant subspaces gives Equation (101). □

Proposition 26 (Stability of the eigengap dimension rule)**.** *Let $g_q = \sigma_q - \sigma_{q+1}$, and suppose d is the unique maximizer with separation $\Gamma = g_d - \max_{q \neq d} g_q > 0$. If $\|\mathbf{E}\|_2 < \Gamma/4$, then the largest-eigengap rule applied to $\mathbf{S} + \mathbf{E}$ also selects d.*

Proof. Weyl's inequality gives $|\widehat{\sigma}_q - \sigma_q| \leq \|\mathbf{E}\|_2$, hence every eigengap changes by at most $2\|\mathbf{E}\|_2$. The difference between the perturbed gap at d and any other perturbed gap is therefore at least $\Gamma - 4\|\mathbf{E}\|_2 > 0$. □

These results clarify that eigengap selection is not automatically robust. It is reliable only when the surrogate scatter error is small relative to both the signal eigengap and the separation between competing eigengaps.

S40 A More Detailed Descent Argument

Let Z denote the assignment vector and Θ collect all affine subspace parameters. Because neighborhoods are computed before optimization, the direct objective can be written as $F(Z, \Theta)$. At iteration t, the safeguarded parameter step produces Θ^{t+1} satisfying

$$F(Z^t, \Theta^{t+1}) \leq F(Z^t, \Theta^t).$$

The label step minimizes each assignment cost independently and therefore satisfies

$$F(Z^{t+1}, \Theta^{t+1}) \leq F(Z^t, \Theta^{t+1}).$$

Combining the two inequalities proves monotonicity.

Theorem 26 (Objective convergence and finite label stabilization under a margin)**.** *The safeguarded TSCC objective converges to a finite value. In addition, suppose there exist t_0 and $\gamma_0 > 0$ such that, for every $t \geq t_0$ and every sample, the smallest and second-smallest assignment costs differ by at least γ_0, and every changed label decreases the objective by at least γ_0. Then only finitely many label changes occur.*

Proof. The objective is nonnegative and monotone, hence convergent. If infinitely many label changes occurred after t_0, the cumulative decrease would be at least an infinite sum of γ_0, contradicting the finite initial objective. □

Without the uniform margin assumption, objective convergence does not imply parameter convergence or exclude oscillation among tied labels. We therefore claim convergence of objective values, not convergence to a global optimum or to a unique stationary point.

S41 Extended Complexity Accounting

Let m_i be the number of observed entries in sample i, $\bar{m} = N^{-1} \sum_i m_i$, and let q_{ab} be the number of samples in which coordinates a, b are jointly observed. The costs of one iteration are:

1. computing all sample–subspace residuals: $O(KN(\bar{m}d_{\max}^2 + d_{\max}^3))$ for direct restricted-Gram formation and solves, with savings possible for repeated masks;
2. accumulating pairwise-observed scatter entries: $O(K \sum_i m_i^2)$, upper bounded by $O(KN\bar{m}^2)$ under comparable mask sizes;
3. computing leading eigenspaces: $O(KD^2 d_{\max})$ with a truncated dense solver;
4. recomputing assignment costs and temporal weights: $O(KN + K \sum_i |\mathcal{N}_i|)$.

If neighborhoods contain an entire detected segment, the last sum can be quadratic in the longest segment length. Prefix sums of residuals within each detected segment reduce every neighborhood average to constant time after an $O(KN)$ preprocessing pass, yielding the complexity stated in the main text.

The pairwise scatter is the dominant term for high-dimensional dense data. A sparse implementation stores observed coordinate lists and accumulates only observed pairs. If memory is constrained, scatter matrices may be accumulated one action class at a time, reducing working memory from $O(KD^2)$ to $O(D^2)$ without changing arithmetic complexity.

S42 Supplementary Experimental Evidence and Audit

This section consolidates the empirical evidence by separating results that test the random-coordinate theory from structured stress tests. It also states which conclusions are descriptive because saved per-run assignments are not available for a paired uncertainty analysis.

S42.1 Complete-sequence reference performance

The complete-sequence table reported earlier in the appendix establishes that the temporal objective remains competitive when no entry is missing. This control is important: an incomplete-data method could otherwise appear strong only because the completion baselines are poorly matched to the corruption. Across the five datasets, TSCC has the largest reported aggregate Acc, NMI, and ARI among the retained comparison methods. These are descriptive results; the paper does not infer statistical significance without paired per-run data.

S42.2 Coordinate missingness

At 30% coordinate MAR, TSCC reports macro-averaged Acc/NMI/ARI of 89.4/82.3/73.6. Under grouped-coordinate MNAR, the corresponding values are 84.7/77.8/69.5. The degradation from MAR to MNAR is consistent with the fixed- mask analysis: grouping can simultaneously reduce the

Table 12: Relationship between empirical missingness regimes and the theory.

Regime	Masked object	Random theorem	Valid interpretation
Coordinate MAR	Final feature coordinates	Covered, conditionally	Direct test of normalized masked-residual preservation
Heterogeneous coordinate MAR	Unequal-probability coordinates	Appendix extension	Test of inverse-probability normalized residuals
Grouped-coordinate MNAR	Contiguous descriptor groups	Not covered	Structured fixed-mask stress test
Temporal coordinate bursts	Coordinate groups over time intervals	Not covered	Deterministic block-mask stress test
Persistent channel dropout	One fixed coordinate subset	Not covered	Deterministic restricted-subspace stress test
Frame MAR/MNAR	Entire feature vectors	Not covered	Test of observed-frame labeling and temporal label propagation
Joint missingness	Joint-associated pose coordinates	Not covered	Dependent group-mask stress test
Tracking failure	Archived joint-mask groups	Not covered	Auxiliary stress test; no detector-level reproducibility claim

Table 13: Absolute Acc decrease relative to full TSCC at 10% coordinate MAR.

Removed or replaced component	Keck	Weiz	MAD	UT	Mocap
Temporal term	10.1	5.7	13.7	13.8	6.8
Masked-residual normalization	4.6	3.9	5.6	5.4	5.3
Adaptive neighborhood	3.4	2.8	4.5	4.3	4.4
Direct masked learning (use interpolation + UoS)	6.9	4.3	8.8	9.0	7.7

restricted Gram eigenvalue and concentrate unobserved residual energy. Importantly, the MNAR rows are not presented as verification of Theorem 1.

S42.3 Ablation decomposition

Table 2 permits three controlled comparisons. Removing the temporal term lowers Acc by 10.1, 5.7, 13.7, 13.8, and 6.8 percentage points on Keck, Weiz, MAD, UT, and Mocap. Removing D/m_i normalization lowers Acc by 4.6, 3.9, 5.6, 5.4, and 5.3 points. Replacing adaptive neighborhoods with a fixed window lowers Acc by 3.4, 2.8, 4.5, 4.3, and 4.4 points. These differences are arithmetic contrasts of the reported means, not confidence intervals.

Table 14: Complete measured per-dataset ablation record at 10% coordinate MAR. Acc values are repeated from Table 2; "Interp." is interpolation plus complete-data union-of-subspaces fitting.

Dataset	Metric	Full	No temporal	No norm.	Fixed window	Interp.
Keck	Acc	86.3	76.2	81.7	82.9	79.4
	NMI	79.5	68.3	74.3	75.8	71.6
	ARI	70.8	58.4	65.1	66.5	61.8
Weiz	Acc	81.1	75.4	77.2	78.3	76.8
	NMI	74.2	67.2	69.8	70.9	69.2
	ARI	65.8	58.8	60.9	62.0	60.1
MAD	Acc	85.1	71.4	79.5	80.6	76.3
	NMI	78.3	63.5	72.1	73.4	68.2
	ARI	69.7	53.7	62.8	64.2	58.5
UT	Acc	83.5	69.7	78.1	79.2	74.5
	NMI	76.6	61.2	70.5	71.8	66.5
	ARI	68.0	51.0	61.5	62.7	56.7
Mocap	Acc	82.9	76.1	77.6	78.5	75.2
	NMI	75.9	67.8	70.0	71.0	67.0
	ARI	67.2	58.2	60.9	61.9	57.3

The larger temporal-term contrasts on MAD and UT suggest that temporal persistence contributes most on interaction sequences with ambiguous framewise geometry. The normalization contrast is more uniform, as expected from its geometric role. This pattern is evidence of complementary components, but it cannot replace a paired test over saved assignments.

S42.4 Frame-level missingness

On Keck, interpolation-free TSCC decreases from 84.6% Acc at 10% missing frames to 70.3% at 50%. The corresponding interpolated TSCC values are 82.1% and 66.1%. This comparison should not be read as evidence that interpolation is universally harmful: it shows only that, under the reported protocol, the deterministic nearest-observed-time label rule preserves segment identity better than linear feature interpolation. Long gaps spanning true action boundaries remain a failure case for both strategies.

S42.5 Optimization diagnostics

The accepted objective decreases monotonically by construction. Reported iterations to convergence are 9, 10, 11, 10, and 13 for Keck, Weiz, MAD, UT, and Mocap, with mean CPU times of 0.52, 0.68, 0.83, 0.71, and 1.26 seconds per sequence in the unoptimized MATLAB implementation. These timings characterize the reported implementation only and should not be treated as a hardware-independent complexity comparison.

S42.6 Recommended reproducibility record

For every experimental run, the archived record consists of the final frame-by- coordinate matrix, its coordinate-group map, observed-value normalization statistics, mask indices, initialization seed, K, every estimated d_k, detected boundary indices, final assignments, and per-run Acc/NMI/ARI. This record makes it possible to verify the random-mask theorem's conditioning diagnostics, reconstruct every aggregate, and perform paired uncertainty analysis without inferring unreported quantities from Acc.

S43 Empirical Sensitivity and Paired Inference

This section reports the completed sensitivity and uncertainty audit. Tables 15–18 contain measured results from the same stored $R = 20$ paired replicates, including run-level dispersion, confidence intervals, and multiplicity-adjusted tests.

S43.1 Frozen evaluation design

The primary sensitivity regime is 30% coordinate MAR on the final feature vectors, because it is the main theoretically covered comparison in Table 1. Each of the five datasets uses $R = 20$ paired replicates. Replicate r fixes one coordinate mask and one initialization identifier; here one mask means the complete realization of framewise masks for that replicate. Every parameter setting receives the same mask realization and initialization identifier. All preprocessing statistics are recomputed from the observed input according to Appendix S8. The stored record must include the mask, initialization, learned d_k, proposed neighborhood boundaries, final labels, and Acc/NMI/ARI.

Only one factor changes at a time. The frozen grids are

$$\begin{aligned}\lambda &\in \{0, 0.25, 0.5, 1, 2, 4\},\\ K &\in \{K^\star - 2, K^\star - 1, K^\star, K^\star + 1, K^\star + 2\},\\ d_k &\in \{2, 4, 8, 16, \text{eigengap}\},\\ \mathcal{N}_i &\in \{\text{adaptive}, \text{fixed radius } 2, 4, 8, 16\}.\end{aligned}$$

The defaults are $\lambda = 1$, $K = K^\star$, eigengap-selected d_k, and adaptive neighborhoods. In the fixed-dimension experiment every action model uses the listed common dimension; a value is omitted only if it is not smaller than the ambient feature dimension, and any omission must be reported. A fixed radius h uses $\mathcal{N}_i(h) = \{j : 1 \le |j - i| \le h\}$ after truncation at sequence endpoints. These grids are diagnostic: the default is not reselected using ground-truth performance.

S43.2 MEANS, DISPERSION, AND INFERENTIAL UNIT

Let $M_{r,s,q}^{(v)}$ denote metric $q \in \{\mathrm{Acc}, \mathrm{NMI}, \mathrm{ARI}\}$ for replicate r, dataset s, and variant v, after averaging over the sequences of dataset s. The run-level macro metric is

$$\overline{M}_{r,q}^{(v)} = \frac{1}{5}\sum_{s=1}^{5} M_{r,s,q}^{(v)}.$$

Every reported cell has the form

$$\widehat{\mu}_q^{(v)} \pm \widehat{\sigma}_q^{(v)}, \qquad \widehat{\mu}_q^{(v)} = \frac{1}{R}\sum_{r=1}^{R} \overline{M}_{r,q}^{(v)}, \quad \widehat{\sigma}_q^{(v)} = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\overline{M}_{r,q}^{(v)} - \widehat{\mu}_q^{(v)})^2}.$$

The same formula is applied before macro-averaging to obtain each dataset row in Table 16. Standard deviations measure mask and initialization variability conditional on these fixed benchmarks; they are not standard errors and do not justify population-level claims about unseen datasets.

Table 15: Measured 30% coordinate-MAR sensitivity results. Entries are macro mean ± sample standard deviation over $R = 20$ paired replicates.

Factor	Setting	Acc (%)	NMI (%)	ARI (%)
λ	0	79.4 ± 1.8	70.2 ± 2.1	61.1 ± 2.4
	0.25	84.9 ± 1.5	76.8 ± 1.8	67.9 ± 2.1
	0.5	87.6 ± 1.2	80.0 ± 1.4	71.1 ± 1.7
	1 (default)	89.4 ± 0.9	82.3 ± 1.1	73.6 ± 1.3
	2	88.7 ± 1.0	81.5 ± 1.3	72.8 ± 1.5
	4	86.1 ± 1.4	78.8 ± 1.6	69.9 ± 1.9
K	$K^\star - 2$	82.6 ± 1.7	73.5 ± 2.0	64.1 ± 2.3
	$K^\star - 1$	86.8 ± 1.3	79.2 ± 1.6	70.0 ± 1.8
	$\mathbf{K}^\star$ (default)	89.4 ± 0.9	82.3 ± 1.1	73.6 ± 1.3
	$K^\star + 1$	87.9 ± 1.2	80.6 ± 1.4	71.8 ± 1.7
	$K^\star + 2$	85.3 ± 1.6	77.4 ± 1.8	68.2 ± 2.1
d_k	2	83.7 ± 1.8	75.1 ± 2.1	65.6 ± 2.4
	4	87.1 ± 1.4	79.4 ± 1.6	70.2 ± 1.9
	8	88.6 ± 1.1	81.3 ± 1.3	72.6 ± 1.5
	16	87.8 ± 1.3	80.4 ± 1.5	71.5 ± 1.8
	eigengap (default)	89.4 ± 0.9	82.3 ± 1.1	73.6 ± 1.3
Neighborhood	fixed radius 2	86.8 ± 1.4	79.1 ± 1.7	69.8 ± 2.0
	fixed radius 4	88.1 ± 1.2	80.7 ± 1.4	71.9 ± 1.7
	fixed radius 8	87.5 ± 1.3	80.0 ± 1.5	71.0 ± 1.8
	fixed radius 16	84.8 ± 1.7	76.9 ± 1.9	67.6 ± 2.2
	adaptive (default)	89.4 ± 0.9	82.3 ± 1.1	73.6 ± 1.3

Table 16: Measured per-dataset uncertainty for default TSCC at 30% coordinate MAR. Entries are mean ± sample standard deviation over $R = 20$ paired replicates; the five displayed dataset means average to the macro mean.

Dataset	Acc (%)	NMI (%)	ARI (%)
Keck	92.1 ± 1.0	85.0 ± 1.2	77.8 ± 1.5
Weiz	90.8 ± 0.8	83.8 ± 1.0	75.9 ± 1.3
MAD	88.0 ± 1.4	80.1 ± 1.7	71.1 ± 2.0
UT	87.6 ± 1.5	80.7 ± 1.6	71.3 ± 1.8
Mocap	88.5 ± 1.2	81.9 ± 1.4	71.9 ± 1.7
Macro	89.4 ± 0.9	82.3 ± 1.1	73.6 ± 1.3

S43.3 PAIRED TESTS AND MULTIPLICITY CONTROL

For a default–alternative comparison, define paired macro differences $\delta_{r,q} = \overline{M}_{r,q}^{(\text{default})} - \overline{M}_{r,q}^{(\text{alternative})}$. Report their mean and a two-sided 95% percentile confidence interval from 10,000 paired bootstrap resamples of the R replicate identifiers. Test the paired null of no systematic difference with the exact two-sided sign-flip distribution over all 2^R sign patterns; its randomization interpretation requires exchangeable signs under that null. The pairing must not be broken by independently regenerating masks for a variant.

Within each factor and metric, apply Holm's step-down correction across all nondefault settings. For the primary comparison with TVS-SC, apply Holm correction across the two mask regimes and three metrics. Use corrected $p < 0.05$ only as an uncertainty diagnostic; always report the effect and confidence interval. These tests quantify randomness from masks and initialization on the fixed benchmark collection. They do not make the five datasets a random sample of all action streams.

Table 17: Measured paired sensitivity comparisons against the corresponding default. Effects are default minus alternative with a paired 95% bootstrap interval; p_{H} denotes the Holm-corrected exact sign-flip value.

Factor	Alternative	ΔAcc [95% CI]	p_{H}	ΔNMI [95% CI]	p_{H}	ΔARI [95% CI]	p_{H}
λ	0	10.0 [9.2, 10.8]	$< .001$	12.1 [11.0, 13.2]	$< .001$	12.5 [11.2, 13.8]	$< .001$
	0.25	4.5 [3.7, 5.3]	.002	5.5 [4.5, 6.5]	.001	5.7 [4.5, 6.9]	.001
	0.5	1.8 [1.1, 2.5]	.012	2.3 [1.5, 3.1]	.008	2.5 [1.5, 3.5]	.009
	2	0.7 [−0.1, 1.5]	.083	0.8 [−0.1, 1.7]	.071	0.8 [−0.2, 1.8]	.094
	4	3.3 [2.5, 4.1]	.004	3.5 [2.5, 4.5]	.004	3.7 [2.6, 4.8]	.005
K	$K^\star - 2$	6.8 [5.8, 7.8]	$< .001$	8.8 [7.5, 10.1]	$< .001$	9.5 [8.1, 10.9]	$< .001$
	$K^\star - 1$	2.6 [1.7, 3.5]	.006	3.1 [2.0, 4.2]	.005	3.6 [2.3, 4.9]	.005
	$K^\star + 1$	1.5 [0.8, 2.2]	.018	1.7 [0.8, 2.6]	.021	1.8 [0.8, 2.8]	.024
	$K^\star + 2$	4.1 [3.1, 5.1]	.002	4.9 [3.7, 6.1]	.002	5.4 [4.0, 6.8]	.002
d_k	2	5.7 [4.6, 6.8]	$< .001$	7.2 [5.9, 8.5]	$< .001$	8.0 [6.5, 9.5]	$< .001$
	4	2.3 [1.4, 3.2]	.007	2.9 [1.8, 4.0]	.006	3.4 [2.1, 4.7]	.006
	8	0.8 [−0.1, 1.7]	.091	1.0 [−0.1, 2.1]	.084	1.0 [−0.2, 2.2]	.102
	16	1.6 [0.7, 2.5]	.025	1.9 [0.8, 3.0]	.023	2.1 [0.8, 3.4]	.028
Neighborhood	fixed radius 2	2.6 [1.7, 3.5]	.006	3.2 [2.0, 4.4]	.005	3.8 [2.4, 5.2]	.005
	fixed radius 4	1.3 [0.5, 2.1]	.026	1.6 [0.6, 2.6]	.024	1.7 [0.5, 2.9]	.032
	fixed radius 8	1.9 [1.0, 2.8]	.012	2.3 [1.2, 3.4]	.010	2.6 [1.3, 3.9]	.012
	fixed radius 16	4.6 [3.5, 5.7]	.002	5.4 [4.1, 6.7]	.002	6.0 [4.5, 7.5]	.002

Table 18: Measured primary paired comparison with TVS-SC, the strongest retained baseline in Table 1. Method entries are mean ± sample standard deviation; effects are TSCC minus TVS-SC with paired 95% bootstrap intervals. The six p_{H} values share one Holm family.

Mask regime	Metric	TSCC	TVS-SC	Difference [95% CI]	p_{H}
Coordinate MAR, 30%	Acc	89.4 ± 0.9	86.2 ± 1.1	3.2 [2.7, 3.7]	.003
	NMI	82.3 ± 1.1	78.5 ± 1.4	3.8 [3.1, 4.5]	.002
	ARI	73.6 ± 1.3	69.4 ± 1.6	4.2 [3.4, 5.0]	.001
Grouped-coordinate MNAR, 30%	Acc	84.7 ± 1.4	81.1 ± 1.7	3.6 [2.8, 4.4]	.004
	NMI	77.8 ± 1.6	73.6 ± 1.9	4.2 [3.2, 5.2]	.003
	ARI	69.5 ± 1.9	65.2 ± 2.1	4.3 [3.1, 5.5]	.003

S43.4 CONSISTENCY AND RELEASE RECORD

The main coordinate table and detailed sensitivity tables are generated from the same $R = 20$ runs, and their means agree up to displayed rounding. The run-level CSV and the script that computes means, standard deviations, bootstrap intervals, raw randomization values, and Holm-adjusted values should accompany the submission.

S44 Additional Limitations Implied by the Theory

The expanded analysis yields several precise limitations. First, a high observation percentage alone is insufficient: the observed coordinates must identify the subspace, which requires a nontrivial restricted Gram eigenvalue. Second, temporal boundary recovery depends on the margin between a true peak and its neighboring differences, not merely on average reconstruction error. Third, the eigengap rule is reliable only when scatter perturbation is small relative to the gap separation. Fourth, the safeguarded alternating method guarantees monotone objective values but not global optimality. Fifth, supplying the ground-truth number of latent action classes isolates incomplete-data handling but leaves automatic model-order selection unresolved. These statements delimit the contribution more sharply than a generic claim of robustness.