Supplementary Information for
Proteus: Adaptive Scale-Space Analysis for Learning Non-Linear Fuzzy Manifold Memberships

S1. Initial Dimensionality Reduction

Randomized SVD: Sample up to 20 000 rows (or all if small), compute a randomized SVD with sketch width $k_{\text{target}} \approx 256$ and oversampling 20. Determine rank by cumulative energy $\geq 99.9\%$ and/or Levina–Bickel intrinsic dimension; keep $r = \max(r_{0.999}, 2\lceil D_{\text{LB}}\rceil)$ if it yields a $\geq 5\%$ reduction; otherwise skip projection. Optionally append 32 JL vectors to tighten distance preservation.

$$P = V_{1:r}^{\top} \;\Rightarrow\; Z = X\,P\,, \quad Z \in \mathbb{R}^{N\times r}.$$

Rationale: Cuts ANN costs and per-sample updates by $\times 5$–$\times 40$ on high-d data while retaining $\geq 99.9\%$ variance and local distances.

S2. First-Principles Derivations

S2.1 Geometric Grid Spacing

Normalized responses in scale-space obey $\Delta_{\text{FWHM}}(\log\sigma) \approx 0.586$. If two peaks are separated by less than this, they are not reliably distinguishable. Therefore, choose adjacent scales to differ by $\log\sigma_{j+1} - \log\sigma_j = \Delta \approx \frac{1}{2}\,\Delta_{\text{FWHM}}$, giving $\sigma_{j+1}/\sigma_j = e^{\Delta} \approx \sqrt{2}$. Since $\tau = \sigma^2$, the geometric ratio in τ becomes $r = \tau_{j+1}/\tau_j = (\sigma_{j+1}/\sigma_j)^2 \approx 1/2$. In practice we span both directions, so we use $r = 1/\sqrt{2} \approx 0.71$ in σ (7–9 points/decade) and map consistently into τ.

S2.2 EWMA Weight from Neighborhood Size

EWMA recursion $m_{t+1} = (1-\alpha)m_t + \alpha x_{t+1}$ has half-life $T_{1/2} = \ln(0.5)/\ln(1-\alpha) \approx \ln 2/\alpha$ for small α. Set $T_{1/2}{=}k$ (one neighborhood of updates) $\Rightarrow \alpha = \ln 2/k$. This ties estimator memory to the actual neighborhood support.

S2.3 Dual-Rate Motion from Variance Correction

Under isotropy, variance $\sigma_i^2 \approx \mathbb{E}\|x{-}w_i\|^2$. A radial shift δr changes variance by $\Delta\sigma_i^2 \approx 2\sqrt{\sigma_i^2}\,\delta r$. Spreading over $H{=}k$ effective updates gives

$$\eta_{\text{GNG},i} \approx \frac{\Delta\sigma_i^2}{2H\,\sigma_i^2} = \frac{1}{2k}\Big(1 - \frac{\sigma_i^2}{\tau}\Big)\,; \qquad \eta_{\text{GNG},i} = \frac{\ln 2}{2k}\Big(1 - \frac{\sigma_i^2}{\tau}\Big) \text{ with half-life normalization.}$$

For centering, with $\delta_{\min} = \kappa(1{-}r)\sqrt{\tau}$ and expected drift per update $\approx \sqrt{\tau}$, set $k\,\eta_{\text{cent}}\sqrt{\tau} = \delta_{\min}$ $\Rightarrow \eta_{\text{cent}} = \kappa(1{-}r)/k$.

S2.4 Control-to-Threshold Mapping

Constrain growth to local intrinsic dimension by mapping optimizer control to thresholds: $\tau_{\text{global}} = -D_{\text{subspace}}\,\log(1{-}s_{\text{control}})$ and $\tau_{\text{local},i} = -d_{\text{final},i}\,\log(1{-}s_{\text{control}})$.

S3. Statistical Gauntlets

S3.1 Link Pruning

Power shield: protect a new link until $n_{\text{eff}} \geq n_{\text{req}}$ from an analytic power formula for proportions. *Relative weakness:* compute Wilson one-sided upper bound ν^+ for link proportion p with count n:

$$\nu^+ = \frac{p + z^2/(2n) + z\sqrt{p(1-p)/n + z^2/(4n^2)}}{1 + z^2/n}, \ z{=}1.2816.$$

Vote to prune if $\nu^+ <$ median peer significance; require bilateral agreement from both endpoints.

S3.2 Node Pruning

Trigger: require neighborhood stability (Welch t on $\tilde{\rho}$). *Fair chance:* Poisson arrival-rate test $\lambda_{\text{est}} = h_i/\text{lifetime} \geq 2/\delta_{\min}$. *Relative:* prune if hits$<0.5\times$ neighbor mean *and* variance $\sigma_i^2 < 0.5\,\tau_{\text{local},i}$. This avoids deleting low-mass but still-curved support.

S4. Torsion and Shape Quality

Let $\bar{\mathbf{w}}$ and $\bar{\mathbf{m}}$ be barycenters. With $E = [\mathbf{w}_{v_1}-\bar{\mathbf{w}} \ \cdots \ \mathbf{w}_{v_d}-\bar{\mathbf{w}}]$ and $M = [\mathbf{m}_{v_1}-\bar{\mathbf{m}} \ \cdots \ \mathbf{m}_{v_d}-\bar{\mathbf{m}}]$, define

$$\Omega_S = M^\top E - E^\top M\,, \qquad \kappa_S = \|\Omega_S\|_F\,; \qquad R_S = \kappa_S/\tau^*.$$

Split rule: if $0.30 \leq R_S{<}0.60$ and budget permits, insert along the dominant torsion axis; if any child has $Q_S{<}0.25$, redo via centroid star split.

S5. Dual Flow Mechanisms

Accumulate facet pressures for winning simplex S via $\Delta p_f \propto \max(0, (\mathbf{x}-\bar{\mathbf{w}}_S)^\top \mathbf{n}_f)$. On the dual graph, solve

$$\min_{\{p_f\}} \lambda \sum_f (p_f - \hat{p}_f)^2 + \mu \sum_S \|A_S\,\mathbf{p}_S\|_2^2\,,$$

with Gauss–Seidel iterations $p_f \leftarrow (\lambda\hat{p}_f + \mu \sum_{S\ni f}(A_S^\top A_S\,\mathbf{p}_S)_f)/(\lambda + \mu\ \deg(f))$, or equivalently by loopy BP messages until convergence.

S6. Masking for Missing Modalities

For input $\mathbf{x} \in \mathbb{R}^d$ with structured sparsity, build boolean mask $\mathbf{m} = (x \neq 0)$. Compute distances and residuals only on active dimensions: $\|\mathbf{x}-\mathbf{w}\|^2 = \sum_{j:m_j}(x_j - w_j)^2$; residual $\mathbf{e} = \mathbf{x}-\mathbf{w}$ with $e_j{=}0$ if $m_j{=}0$. Apply the mask to all vector updates (moments, nudges, Oja), and accumulate variance only on active dimensions. Optionally normalize updates by the active count to equalize per-sample influence. This preserves geometry and statistics in multi-modal fusion where zeros indicate missingness rather than small values.

S7. Non-Linear Warp Strategies

When geometric refinement (node moves and splits) is insufficient to resolve high residual curvature, Proteus can apply a non-linear warp. The core strategy is a hybrid approach that adapts to the nature of the curvature, measured by the torsion coverage P_κ (the fraction of simplices with torsion ratio $R_S \geq 0.30$).

Global vs. Patch-wise Strategy:

- If curvature is **widespread** ($P_\kappa > 50\%$), a shallow *global Glow* is trained. Its invertible 1×1 convolutions mix dimensions efficiently, providing a good baseline correction for system-wide bends.
- If curvature is **patchy** ($P_\kappa \leq 25\%$), compact and expressive *mini-Normalizing Spline Flows (NSFs)* are attached only to the highest-torsion patches. This avoids wasting model capacity on regions that are already linear.
- If curvature is **mixed** ($25\% < P_\kappa \leq 50\%$), a moderate global Glow is combined with mini-NSFs on the worst ∼10% of patches.

Patch Identification and Mini-NSF Training: A facet-adjacency graph is built over all "hot" simplices ($R_S \geq 0.60$). Leiden community detection finds contiguous patches of high torsion, with resolution auto-tuned to ensure no patch exceeds a size cap $P_{\max}$ (e.g., 64 simplices). Each mini-NSF is a compact model (2-3 layers of rational-quadratic spline couplings, 64-128 hidden units, 8-12 bins) trained for a small number of epochs on its routed data, with the global model frozen.

S8. Numerical Stability and Implementation

Conditioning for Ω_S: Use QR (preferred) or thin SVD to stabilize E; scale residuals M by a local norm to avoid magnitude blow-up; compute $\Omega_S = M^T E - E^T M$ from stabilized factors; cache per-simplex factors and recompute only after geometric moves/splits.

Small constants and clipping: Set ε in ρ_i to $10^{-8} - 10^{-6}$ depending on data scale; clip η_{GNG} to $|\eta| \leq 0.3$ (Stage 1) and ≤ 0.05 (Stage 2); constrain spline slopes/bins to avoid near-singular Jacobians; bound $\|\mathbf{a}_i\|$ growth between apply steps.

ANN configuration (HNSW): Typical defaults: M=16 − 32, efConstruction=100 − 400, efSearch=50 − 200; increase efSearch modestly near convergence to improve neighbor stability. Persist indices per recursion level; rebuild only after significant mesh mutation.

S9. Robustness and Edge Cases

Outliers: Use Huber-like weighting on residuals in moment updates or winsorize top 0.5% residual norms; optionally cap per-sample hit contributions.

Boundaries and orientation: For open meshes, treat boundary faces with Neumann-like conditions in the dual flow (no flux across exterior). In non-orientable regions, operate on local orientation patches to keep face normals consistent.

Sparse patches: Require a minimum routed sample count $n_{\min} = \max(10\,d, 1000)$ before fitting a mini-NSF; otherwise defer to geometric refinement.

S10. Evaluation Protocols

Defaults: Stage 1 k=8, α= $\ln 2/8$, grid r=$1/\sqrt{2}$; Stage 2 $k \approx d$, α= $\ln 2/k$. Stopping: CV< 0.01 and R_S<0.30. PH: VR complex, max dim 2, filtration up to $1.5\,\sigma_\star$. MMD: RBF kernel, bandwidth median heuristic, 1 000 samples/1 000 tests. Ablations: remove Stage 1 scaffold, Torsion Ladder, or Dual Flow.

Persistent Homology: Vietoris–Rips (max dim 2), filtration to $1.5\,\sigma_\star$; report bottleneck/2-Wasserstein. **MMD:** RBF kernel (median heuristic), 1000 samples/1000 tests, 5 seeds. **Log-likelihood:** Sum of face-pressure potentials within simplex plus normalization; use held-out data.

S11. Algorithmic Details

Algorithm 1 Stage 1 at Fixed τ

procedure RunToEquilibrium(X, τ)
  Initialize mesh $\mathcal{M} \leftarrow \varnothing$
  repeat
    Sample $\mathbf{x}$ from X
    $\mathcal{N}_k \leftarrow$ QueryANN($\mathbf{x}, k$) ▷ Find k nearest nodes
    ApplyRankWeightedUpdates($\mathbf{x}$, $\mathcal{M}$, $\mathcal{N}_k$)
    for each updated node i **do**
      if $\|\mathbf{node}_i.\mathbf{a}\| \geq \delta_{min}$ **then** ▷ Deferred move
        $\mathbf{node}_i.\mathbf{w}$ += $\mathbf{node}_i.\mathbf{a}$; $\mathbf{node}_i.\mathbf{a} \leftarrow 0$; RefreshANN($\mathbf{node}_i$)
      end if
      if $\mathbf{node}_i.\sigma^2 > \tau_{local,i}$ **then** ▷ Split rule
        SplitNode($\mathbf{node}_i$) ▷ Flush nudge, create child, shrink-inherit stats
      end if
    end for
    ApplyPruningGauntlets($\mathcal{M}$) ▷ Per SI S3
  until ComputeCV($\mathcal{M}$) < 0.01
  return $\mathcal{M}$
end procedure

Algorithm 2 Rank-weighted neighbor updates

procedure ApplyRankWeightedUpdates($\mathbf{x}, \mathcal{M}, \mathcal{N}_k$)
  for $j = 1 \ldots k$ **in** $\mathcal{N}_k$ (rank-ordered) **do**
    $w_j \leftarrow 2^{-(j-1)}$
    $\mathbf{node} \leftarrow \mathcal{N}_k[j]$
    $\mathbf{e} \leftarrow \mathbf{x} - \mathbf{node.w}$
    $\mathbf{node.m} \leftarrow (1 - \alpha w_j)\mathbf{node.m} + \alpha w_j\, \mathbf{e}$; $\mathbf{node.s} \leftarrow (1 - \alpha w_j)\mathbf{node.s} + \alpha w_j\, \mathbf{e}^{\circ 2}$
    $\rho \leftarrow \|\mathbf{node.m}\|/(\mathbf{node}.\sigma + \varepsilon)$; $\mathbf{node.a} \leftarrow \mathbf{node.a} + \eta_{cent}\, \rho\, \mathbf{node.m}$
    UpdateOja($\mathbf{node.u}, \mathbf{e}$)
    $\mathbf{node.h} \leftarrow \mathbf{node.h} + w_j$ ▷ Fractional hits and link counters
  end for
end procedure

S12. Theoretical Guarantees

S12.1 Expressivity and Approximation

Let p be a probability density on a compact set $\Omega \subset \mathbb{R}^d$ with a (piecewise) Lipschitz continuous log-density and bounded curvature of its principal level sets. For any $\varepsilon>0$, there exist finite budgets for recursion depth, mesh resolution, and non-linear warps (mini-NSFs) such that the model density q learned by the Proteus procedure satisfies $\|p - q\|_{L_1} < \varepsilon$.

The proof is constructive, following the algorithmic stages. We show that the total approximation error can be bounded by the sum of errors from a hierarchical decomposition, a piecewise-linear approximation of each component, and a non-linear residual correction. By making the resolution and capacity of each stage sufficient, the total error can be made arbitrarily small.

S12.2 Correspondence: Simplex Equilibrium Implies Effective Node Equilibrium

Setting and assumptions. Consider a simplicial complex with nondegenerate simplex stars around each node, locally stationary sampling, and a fixed characteristic scale τ^*. Let $P(V_i)$ denote Voronoi mass at node i and $P(C)$ the mass of simplex C.

Claim. Driving the system toward simplex equilibrium ($P(C) \approx 1/M$ for all simplexes) induces effective node equilibrium ($P(V_i) \approx 1/N$ up to junction effects and sampling noise).

Sketch. Suppose a node i has excess mass $P(V_i) \gg 1/N$. Samples concentrate near w_i, biasing residuals in the adjacent simplex star $\mathcal{S}(i)$. The barycentric residual means in $\mathcal{S}(i)$ acquire outward components, raising torsional stress and face pressures that cannot be canceled by PL translation alone. The Torsion Ladder therefore either (i) performs a torsion-aligned split creating a child j in the outward direction, or (ii) fits a local mini-NSF that unbends the region and reduces the asymmetric flux. In both cases the effective domain of V_i contracts in favor of V_j or of neighboring stars, strictly decreasing $P(V_i)$ while increasing mass in adjacent cells. Repeating this argument shows that any persistent Voronoi over-mass produces simplex-level stresses whose resolution decreases the imbalance, establishing a negative feedback toward node equilibrium. Deviations persist only at structural junctions (e.g., different intrinsic dimensions), where systematic bimodality and asymmetric links form an informative signature rather than a failure.

Remarks. The argument uses only local stationarity and nondegenerate stars; concentration bounds convert it to a finite-sample statement. Empirically, the correspondence manifests around overloaded nodes as: sharply bimodal simplex activity within the node's star at dimensionality junctions; a clean partition of link statistics into high-significance backbone links versus low-significance bridge links; and bridge links that exhibit persistent directional hit asymmetry together with low net face pressure, distinguishing stable junctions from active frontiers.