TANGENT-ALIGNED QUADRATIC REGRESSION FOR TANGENT SPACE ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Local quadratic fitting estimates tangent spaces by learning an orientation and polynomial coefficients. We study a tangent-aligned normal graph with an intercept and no normal linear term, so the optimized frame is the fitted tangent. Profiling the coefficients separates coefficient conditioning, residualized linear separation, and full local orientation curvature. We construct an exact quadratic saddle with nonsingular coefficient and linear designs but a singular profiled Hessian: predictor rotation cancels linear tilt information. A deformation gives curvature proportional to δ^2, and a perturbation bound preserves weak curvature without exact symmetry. Conversely, symmetric isotropic graphs receive a nonnegative curvature contribution. We also establish conditional noisy-anchor recovery and local optimization bounds. Experiments with observed-data safeguards cover synthetic and image neighborhoods. Matched ablations show improved noiseless quadratic reproduction from full optimization, but no uniform accuracy benefit under noise or cubic model mismatch. The results explain which orientation information survives coefficient refitting and distinguish local objective geometry from practical recovery guarantees.

1 INTRODUCTION

Tangent spaces describe local directions of variation in manifold data and support geometric learning and reconstruction (Zhang & Zha, 2004; Cheng & Wu, 2013; Hoppe et al., 1992). Local polynomial fitting represents curvature beyond the affine model of local PCA. However, fitting a polynomial and identifying its orientation are different tasks: when a frame changes, its coefficients can be refitted. A well-conditioned coefficient design therefore need not imply that nearby frames can be distinguished. We ask: *which orientation information survives coefficient refitting, and how does it govern recovery and local optimization?*

Joint tangent and polynomial estimation is not new. Aamari & Levrard (2019) jointly optimize a projector and polynomial tensors; jets and moving least squares infer differential structure through local polynomial coordinates (Cazals & Pouget, 2005; Alexa et al., 2003; Sober & Levin, 2020). SQMF jointly learns orthogonal subspaces, quadratic structure, and sample coordinates (Zhai & Li, 2024). We study a projected-coordinate quadratic normal graph with a free intercept and no normal linear term. The intercept accommodates normal anchor displacement, and the optimized frame is the fitted tangent. With zero intercept, the criterion has the quadratic normal-output form of Aamari–Levrard. Our focus is the orientation information exposed by profiling the coefficients, rather than a new statistical rate or a generally more accurate nonlinear estimator.

The analysis separates three objects. The quadratic-feature Gram matrix controls coefficient stability. The residualized linear design measures the part of a tangent tilt that an intercept and quadratic terms cannot absorb. The full residualized frame Jacobian determines the exact-reference Hessian. The last object also includes predictor rotation. On a twelve-point quadratic saddle, this term cancels a linear tilt even though the coefficient and residualized linear designs are nonsingular. Nearby designs retain these two properties while their full orientation curvature approaches zero. Small asymmetric perturbations preserve weak curvature, so the distinction is not confined to an exactly symmetric singular sample. Conversely, on symmetric isotropic quadratic graphs predictor rotation adds nonnegative curvature. These contrasting mechanisms explain why the full derivative is needed to characterize local orientation geometry.

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.

Our contributions are threefold:

1. **Computable tangent-aligned fitting.** We formulate constrained coefficient profiling and implement local-chart and ambient Grassmann solvers. The representation makes the frame itself the fitted tangent and incorporates normal anchor displacement through an intercept.
2. **Recovery under coefficient refitting.** We establish localized bounds for feasible returned outputs and sufficient sampling conditions for design and PCA localization after noisy observed-radius selection. The conditions distinguish coefficient feasibility from orientation separation.
3. **Local geometry and empirical scope.** We derive the full exact-reference Hessian, an anisotropic cancellation family, and a bound on weak curvature under asymmetric perturbations. Observed-setting factorial comparisons and matched frame-correction ablations test empirical scope. These results do not prove convergence of the implemented L-BFGS method from a general PCA initialization.

2 Problem Setup and Notations

Let $M \subset \mathbb{R}^D$ be a compact embedded C^2 manifold without boundary of dimension $1 \le d < D$ and codimension $m = D - d$. Assume $\mathrm{reach}(M) = \tau_M \ge \tau_{\min} > 0$. The reach is the supremum of radii within which every point has a unique nearest point on M (Federer, 1959). This condition bounds the curvature and prevents near self-intersections of the manifold. Following the normal-noise model of Aamari & Levrard (2019), let $(x_j^*, \xi_j)_{j=0}^{n_{\mathrm{tot}}-1}$ be i.i.d. with

$$x_j = x_j^* + \xi_j, \quad x_j^* \in M, \quad \xi_j \in N_{x_j^*}M, \quad \mathbb{E}[\xi_j \mid x_j^*] = 0, \quad \|\xi_j\| \le \sigma < \tau_M. \tag{1}$$

These conditions include the prescribed anchor $j = 0$. Our target is $T_{x_0^*}M = T_{\pi_M(x_0)}M$, not a tangent at the generally off-manifold observation x_0.

For a prescribed radius $r > 0$, select indices $\mathcal{I}_r = \{j \ne 0 : \|x_j - x_0\| \le r\}$, and set $N = |\mathcal{I}_r|$. Relabel these observations as $x_1, \dots, x_N$ and write $u_i = x_i - x_0$. The construction below requires $N \ge 1$; otherwise it reports insufficient data. Independence and conditional mean zero in equation 1 describe the original sampling law, not an unconditional model for the selected rows. In particular, mean zero need not survive neighborhood selection.

Write $P = [P_1 \mid P_2] \in O(D)$, with $P_1 \in \mathbb{R}^{D\times d}$, for a candidate tangent–normal frame, and let P^* be a reference frame with $\mathrm{im}(P_1^*) = T_{x_0^*}M$. Define $\Pi(P) = P_1P_1^\top$ and $\Pi^* = P_1^*(P_1^*)^\top$. We report projector error $\|\widetilde{\Pi} - \Pi^*\|_F$.

3 Profiled regression for tangent estimation

3.1 Quadratic regression with a learned frame

For anchor-centered observations $u_i = x_i - x_0$, write $s_i = P_1^\top u_i$ and $n_i = P_2^\top u_i$. For symmetric matrices $Q_k \in \mathrm{Sym}(d)$, define $\mathcal{Q}_Q(s) = \frac{1}{2}(s^\top Q_1 s, \dots, s^\top Q_m s)^\top$. We minimize

$$L_N(P, c, Q) = \frac{1}{N}\sum_i \|n_i - c - \mathcal{Q}_Q(s_i)\|^2. \tag{2}$$

The fitted graph is $\Gamma(s) = x_0 + P_1 s + P_2\{c + \mathcal{Q}_Q(s)\}$. Since $D\Gamma(0) = P_1$, its tangent at $x_0 + P_2 c$ is $\mathrm{im}(P_1)$. A free normal linear term Bs would instead give tangent $\mathrm{im}(P_1 + P_2 B)$. Setting $B = 0$ therefore aligns the frame, the fitted tangent, and the variable used in the recovery and optimization analysis. We denote the projector returned numerically by $\widetilde{\Pi}$.

Anchor displacement and the oracle fit. In a true tangent–normal frame a local graph satisfies

$$x^*(t) = x_0^* + P_1^* t + P_2^* g(t), \quad g(t) = \mathcal{Q}_{Q^*}(t) + R(t), \quad g(0) = Dg(0) = 0. \tag{3}$$

A Lipschitz Hessian gives $\|R(t)\| \le L_3\|t\|^3/6$ (Appendix A). Write $a_0 = (P_2^*)^\top \xi_0$, $\eta_i = (P_1^*)^\top \xi_i$, and $\zeta_i = (P_2^*)^\top \xi_i$. Then

$$s_i(P^*) = t_i + \eta_i, \qquad n_i(P^*) = g(t_i) + \zeta_i - a_0. \tag{4}$$

The oracle intercept $c^* = -a_0$ cancels the shared normal offset exactly. Predictor noise, response noise, and selection bias remain. We exclude the self-anchor row $u_0 = 0$, which would penalize this intercept.

Relation to existing local polynomials. With $c = 0$, multiplying the residual by P_2 gives the normal-output parameterization of the quadratic criterion of Aamari & Levrard (2019); see Appendix A. Our analysis concerns the intercept-bearing profile, the separation of orientation from coefficient refitting, and localized numerical outputs. The intercept and feasible set distinguish the estimator specified below from the zero-intercept criterion.

3.2 Constraints and coefficient profiling

Let ϕ_h contain a constant and Frobenius-orthonormal quadratic features scaled by h^{-2}, and let $\widehat{\Sigma}_{Q,h}(P) = N^{-1}\sum_i \phi_h(s_i)\phi_h(s_i)^\top$. There are $p = 1+d(d+1)/2$ regression features and $q = p+d$ augmented features. Here h is a supplied regression scale. In the recovery statements it also bounds the latent coordinates; the implementation scale is distinguished in Section 4.1. Given the anchor-centered PCA projector Π_0, define

$$\mathcal{F} = \{P : \lambda_{\min}(\widehat{\Sigma}_{Q,h}(P)) \geq \tau,\ \|\Pi(P) - \Pi_0\|_F \leq r_0\}, \tag{5}$$

$$\mathcal{Q}_K = \{Q : \textstyle\sum_k \|Q_k\|_F^2 \leq K^2\}. \tag{6}$$

For $\tau > 0$ and finite K, the analyzed estimator is

$$(\widehat{P}, \widehat{c}, \widehat{Q}) \in \operatorname*{argmin}_{P\in\mathcal{F},\ c,\ Q\in\mathcal{Q}_K} L_N(P, c, Q), \qquad \widehat{\Pi} = \widehat{P}_1\widehat{P}_1^\top. \tag{7}$$

Changing bases to P_1R_1, P_2R_2 sends c to $R_2^\top c$ and Q_a to $R_1^\top\sum_k (R_2)_{ka}Q_kR_1$. These transformations preserve the loss, curvature norm, Gram eigenvalues, and projector constraints, so the profile is a function on $\mathrm{Gr}(d, D)$ (Theorem C.14). The Gram constraint stabilizes coefficient estimation; it alone does not certify tangent identifiability. For each frame, the inner problem $J_{N,K}(P) = \min_{c,Q\in\mathcal{Q}_K} L_N(P, c, Q)$ is convex with a unique solution on $\mathcal{F}$. Unconstrained least squares applies only when its solution satisfies the curvature bound. Nonempty compact feasibility yields existence; it does not imply solver convergence. This coefficient elimination follows variable projection (Golub & Pereyra, 1973) and exposes the design matrices used below. Because both predictors and responses contain noise, ordinary fixed-design regression guarantees cannot simply be transferred to the tangent estimator.

3.3 Numerical optimization and ambient implementation

We initialize with the top d eigenvectors of $N^{-1}\sum_i u_iu_i^\top$ and complete them to $P_0 = [P_{01} \mid P_{02}]$. Local-chart L-BFGS optimizes

$$P(A) = P_0 \exp\begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix}, \qquad A \in \mathbb{R}^{(D-d)\times d}, \quad \|A\|_{\mathrm{op}} < a_{\max} < \pi/2. \tag{8}$$

Each trial refits the coefficients. Feasible Armijo backtracking uses the actual directional derivative; a non-descent direction triggers a memory reset. Optimization uses normalized displacements u_i/h and curvature bound Kh. Gram failure, gradient tolerance, relative-loss tolerance, and line-search failure are distinct diagnostics. Profile derivatives and local convergence assumptions appear in Appendix C.

For large D, we store only P_1 and regress the ambient normal residual $(I - P_1P_1^\top)u_i$ on the same features. Ambient coefficients lie in the normal space, preserving residual norms and the curvature norm. Grassmann exponential steps avoid a full $D \times D$ matrix exponential. This implementation re-centers the chart and transports L-BFGS memory by projection: it has the same fitting criterion at matched settings, but a different iteration path. Small-dimensional numerical checks compare loss, gradients, active curvature fits, and returned projectors.

3.4 From regression structure to recovery and numerical bounds

Let Φ_h have rows $\phi_h(t_i)^\top$ at the latent coordinates, let $L_h = [t_i^\top/h]_i$, and let P_Φ project onto $\mathrm{col}(\Phi_h)$. Define

$$S_{\mathrm{lin}} = N^{-1}L_h^\top(I - \mathsf{P}_\Phi)L_h. \tag{9}$$

Then $\inf_{c,Q} N^{-1} \sum_i \|At_i - c - \mathcal{Q}_Q(t_i)\|^2 = h^2 \operatorname{tr}(AS_{\mathrm{lin}}A^\top)$: separation measures the part of a linear tilt that coefficient refitting cannot absorb. Write $\Psi_h = [\Phi_h \mid L_h]$. The augmented-design condition is $N^{-1}\Psi_h^\top \Psi_h \succeq \mu I$, with $\mu > 0$; it implies $S_{\mathrm{lin}} \succeq \mu I$.

Theorem 3.1 (Localized recovery and numerical residuals)**.** *Assume $\|t_i\| \le h \le H$ on a supplied graph ball of finite radius H, $N^{-1}\Psi_h^\top \Psi_h \succeq \mu I$, $Q^* \in \mathcal{Q}_K$, and true-frame feasibility. Suppose every feasible projector lies in the true-frame graph chart* $\operatorname{im}\begin{pmatrix} I_d \\ A \end{pmatrix}$, *$\|A\|_{\mathrm{op}} \le 1/2$. Assume a Lipschitz Hessian with constant L_3, and normal noises $\|\xi_i\| \le \sigma \le h$, including the anchor, under the model of Section 2. Every feasible output satisfies*

$$\|\widetilde{\Pi} - \Pi^*\|_F \le C_H \left(h^2 + \frac{\sigma}{h} + \frac{\sqrt{L_N(\widetilde{P}, \widetilde{c}, \widetilde{Q})}}{h} \right). \tag{10}$$

Here C_H depends on μ, K, L_3, d, m, H, not on N, h, σ. For an output with objective gap at most Δ_{opt}, the last term can be replaced by $\sqrt{\Delta_{\mathrm{opt}}}/h$, with a different constant. For $C^{2,\alpha}$, $0 < \alpha \le 1$, replace the geometric term by $h^{1+\alpha}$.

Corollary C.39 proves this statement. The loss in equation 10 is computed at the returned output: it requires neither the unknown optimum value nor a certified objective gap. Geometry, separation, and localization remain premises. Direct separation $S_{\mathrm{lin}} \succeq \mu_{\mathrm{sep}} I$ can replace augmented conditioning if observed coefficient feasibility and localization are retained separately (Proposition C.25). The bound describes fixed-scale robustness. The observed-radius sampling corollary verifies design and PCA localization under explicit sample-size and signal conditions; the clean designated-anchor corollary gives a sample-size-dependent rate (Corollaries C.42 and C.44).

Exact reproduction and a local landscape. For exact quadratic neighbors and normal anchor offset a_0, the true frame has zero loss at $c = -a_0$. Exact recovery of every localized minimizer additionally follows under $c_h > 0$ in Theorem C.40 for noiseless data; zero oracle loss alone is insufficient. For the landscape result, assume full-rank design, strict Gram and trust-region margins, and $K_* = (\sum_k \|Q_k^*\|_F^2)^{1/2} < K$. The constrained profile then agrees locally with the smooth unconstrained profile. In the true-centered *exponential* chart, its Hessian is

$$D^2 J(0)[H, H] = 2\|(I - \mathsf{P}_\Phi)[Ht + Dg(t)H^\top (g(t) - a_0)]\|_N^2. \tag{11}$$

Here $H \in \mathbb{R}^{m \times d}$ has the Frobenius inner product and $\|v\|_N^2 = N^{-1} \sum_i \|v_i\|^2$. If $S_{\mathrm{lin}} \succeq \mu_{\mathrm{sep}} I$ and $\gamma_{\mathrm{loc}} = \sqrt{\mu_{\mathrm{sep}}} - K_2(K_2 h^2/2 + \|a_0\|) > 0$, with $K_2^2 = \sum_k \|Q_k^*\|_{\mathrm{op}}^2$, this Hessian is bounded below by $D^2 J(0)[H, H] \ge 2h^2 \gamma_{\mathrm{loc}}^2 \|H\|_F^2$. Thus the exact solution is locally isolated. The predictor-rotation term in equation 11 is essential. Appendix C.11 proves this formula and explicit small-perturbation conditions for a strongly convex basin and a $\|\nabla J\|/h^2$ numerical-error term. The graph and exponential charts share their first variation at zero; away from zero their coordinates differ. These bounds apply to outputs inside the specified basin. Appendix C.12 analyzes a gradient-descent iteration under additional basin assumptions; PCA-to-L-BFGS basin entry remains unproved. For the same chart and fixed h, the normalized loss $\bar{J} = J/h^2$ satisfies $\nabla \bar{J} = \nabla J/h^2$; gradients from different charts require coordinate conversion.

When predictor rotation adds curvature. For a clean anchor and an exact graph $g(t) = \kappa\|t\|^2/2$, $\kappa \in \mathbb{R}^m$, suppose the empirical coordinates are centrally symmetric and the strict reference margins above hold. Corollary C.51 gives

$$D^2 J(0)[H, H] = 2\|Ht\|_N^2 + \frac{1}{N} \sum_i \left(2\|t_i\|^2 + \frac{\|\kappa\|^2}{2} \|t_i\|^4 \right) (\kappa^\top H t_i)^2.$$

The added term is nonnegative. If the tangent second moment is positive definite, local curvature is positive without requiring $\gamma_{\mathrm{loc}} > 0$. This is a specialization of the exact Hessian, not an extension of the general noisy recovery theorem. In higher codimension, additional directional curvature need not improve the condition number.

When predictor rotation cancels linear information. For the saddle $g(x, y) = (x^2 - y^2)/2$, consider the twelve points $(\pm 1, \pm\sqrt{3-\delta})$, $(\pm 2, \pm\sqrt{6-\delta})$, $(0, \pm 1)$, and $(0, \pm 2)$, with all sign combinations and $0 \le \delta \le 1$. Using $h = 1$, the intercept-bearing coefficient design stays full rank and $S_{\text{lin}} = \text{diag}(5/3, 23/6 - 2\delta/3)$, yet

$$\lambda_{\min}(\nabla^2 J(0)) = \frac{5}{6}\delta^2, \qquad 2e_1^\top S_{\text{lin}} e_1 = \frac{10}{3}.$$

Indeed, the full directional residual is $x[1 + (x^2 - y^2)/2] = \delta x/2$ on the off-axis points and zero on the axis points. It vanishes at $\delta = 0$. Proposition C.52 proves this identity with strict local feasibility. Corollary C.53 further shows that arbitrary sufficiently small latent perturbations preserve positive coefficient and linear designs while full curvature is at most $884\varepsilon^2$, where ε is their RMS magnitude. Thus the two design conditions alone do not provide a positive lower bound on local orientation curvature. This is a landscape statement, not a claim of non-identifiability or inevitable numerical failure.

Three levels of guarantee. Theorem 3.1 is conditional on realized geometry, separation, and localization. Corollary C.42 supplies sufficient sampling conditions for design and PCA localization. Local gradient bounds require additional strict margins and basin membership. None makes an observed residual or Gram eigenvalue an unconditional accuracy certificate. Statistical PCA localization does not establish L-BFGS basin entry or retention.

4 Experiments

We test the distinction between coefficient conditioning, linear separation, and full orientation curvature, then assess the estimator under observed settings. A matched frame-correction ablation separates exact quadratic reproduction from accuracy under model mismatch.

4.1 Settings and comparison protocol

The primary practical estimator takes observed displacements U and supplied intrinsic dimension d. Anchor-centered PCA gives P_0; an unconstrained intercept-plus-quadratic pilot at this frame, without a normal linear term, gives $\widehat{Q}_{\text{pilot}}$. We fix

$$h_{\text{obs}} = \max_i \|u_i\|, \qquad K_{\text{obs}} = 2\|\widehat{Q}_{\text{pilot}}\|_F, \qquad \tau_{\text{obs}} = .05\lambda_{\min}(\widehat{\Sigma}_{Q,h_{\text{obs}}}(P_0)), \qquad r_0 = .4.$$

These are empirical rules, not confidence bounds. The four matched arms cross observed versus supplied $K = 2\|Q^*\|_F$ with Gram safeguard on versus off. All use the same cloud, initialization, observed scale, and ambient L-BFGS solver. Reported on-arm thresholds are positive. The relative threshold ensures initial Gram feasibility on these samples, not an absolute conditioning standard or inclusion of the true tangent.

Synthetic studies use ten seeds per cell; radius tuning uses ten calibration and ten disjoint evaluation seeds. Images use 20 base images per family and ten draws per image. These counts are separate from the supplementary supplied-geometry studies. All methods receive the same selected neighbors. Internal Jet fits $c + Bs + \mathcal{Q}_Q(s)$ at PCA and reads out its combined derivative. External MMLS uses the authors' `manapprox` implementation, degree two, SVD initialization, and native weighting/base-point refinement (Aizenbud & Sober, 2021; Sober & Levin, 2020); its tangent is read from the polynomial's linear coefficient. CGAL Jet is restricted to supported surfaces in $\mathbb{R}^3$ (Cazals & Pouget, 2025). Native fitting points and weights differ across these baselines, as detailed in Appendix B.7.

4.2 Cancellation invisible to linear separation

On the twelve-point family of Proposition C.52, all six values $\delta \in \{0, .05, .1, .25, .5, 1\}$ give feature rank four. At zero, S_{lin} has eigenvalues $5/3, 23/6$, whereas the full Hessian has eigenvalues $0, 36.75$. Directional curvature agrees with $5\delta^2/6$ to absolute error 1.34×10^{-15}. Finite differences of the intercept-refitted loss have maximum absolute error 3.54×10^{-6} at step 10^{-3}. Absolute rather than relative error is used for the zero prediction.

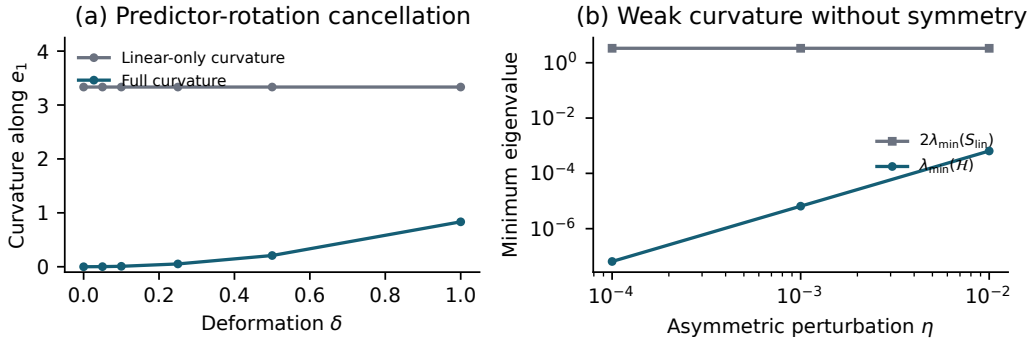


Figure 1: Exact quadratic saddle, clean anchor, intercept-bearing profile. (a) Linear-only directional curvature stays at $10/3$, while the full curvature is $5\delta^2/6$. The coefficient design remains full rank. (b) For the prescribed asymmetric perturbation of the $\delta = 0$ sample, linear separation remains stable but full minimum curvature is small. These are reference-landscape diagnostics, not noisy recovery certificates.

Table 1: Mean projector error with observed settings and matched references. Both Obs. and Oracle use the positive Gram threshold; their Gram-off arms have identical recorded errors on every instance. Jet is internal. The radius row uses independently selected radii and 160 held-out instances per displayed method. Other rows pool their stated grids descriptively; no pooled significance or cross-domain accuracy comparison is intended.

Study	n	Obs.	Oracle	Jet	MMLS	PCA
Radius (held out)	160	0.01677	0.01596	0.01583	0.01439	0.03031
Ambient dimension	4200	0.07090	0.07086	0.07213	0.06344	0.08842
Intrinsic dimension	480	0.01924	0.01912	0.01928	0.02270	0.03690
Images	4800	0.11170	0.11084	0.09774	0.06875	0.14536

Asymmetric perturbations of magnitudes $\eta = 10^{-4}, 10^{-3}, 10^{-2}$ retain feature rank four and minimum linear separation approximately $1.667, 1.666, 1.663$. Full minimum curvature is respectively $6.52 \times 10^{-8}, 6.51 \times 10^{-6}, 6.45 \times 10^{-4}$ (Figure 1). This one deterministic path illustrates the perturbation bound, not a prevalence claim for arbitrary data. Conversely, symmetric isotropic-graph checks verify the nonnegative contribution of Corollary C.51 (Appendix B.8).

4.3 Observed settings across the principal domains

The factorial study contains 2,240 radius-study instances (including calibration), 4,200 ambient-dimension instances through $(d, D) = (5, 100)$, 480 intrinsic-dimension instances through $d = 15, D = 100$, and 4,800 image instances. All four proposed arms return on all 11,720 instances; this is not a convergence or recovery rate. Observed and oracle means are close in several domains, but not interchangeable in every condition (Table 1). In the intrinsic study, external MMLS has higher pooled error than the observed fit; in the other displayed studies it has lower error. Full condition tables are in Appendix B.9.

For radius selection we retain a fixed latent sampling window and absolute noise, select each method's radius by reference-tangent error on calibration seeds, and evaluate on disjoint seeds. This is *supervised benchmark calibration*; the pilot rule does not turn it into unlabeled radius selection. The selected-radius mean is 0.01677 with observed curvature and 0.01596 with supplied curvature; external MMLS gives 0.01439. Candidate radii and selection are distinct from same-radius factorial comparisons.

What the safeguard does and does not establish. Gram on/off gives identical recorded errors in all 11,720 matched instances, for both curvature settings. At observed-arm outputs the mini-

mum Gram margin is substantial: $\lambda_{\min}(\widehat{\Sigma})/\tau \geq 13.3$. The saved outputs satisfy the curvature and Gram constraints to numerical tolerance. Thus positive-threshold fitting is evaluated, but no accuracy benefit from the Gram safeguard is observed in this grid. Endpoint margins do not reconstruct every rejected trial along the trajectory. Pilot curvature undercoverage occurs in 129/2,240 radius instances and 25/4,200 ambient-dimension instances. Truth lies outside the trust region in 92 and 182 instances, respectively. These are distinct from the more conservative sufficient chart-localization check. Neither low loss nor observed feasibility certifies the latent-design or basin premises.

Image transfer and localization. Gaussian-blob and smoothed MNIST transformations give $d = 1, 2, 3$ in $D = 1024, 784$. Pixel Gaussian noise is outside the bounded-normal-noise model. Observed fitting has lower mean error than external MMLS in seven of 24 conditions, and higher error in seventeen. In the noisiest blob translation-plus-rotation condition its error is 0.24346, versus 0.23022 for the oracle arm. For the corresponding MNIST condition the means are 0.48354 (observed), 0.32929 (PCA), and 0.18272 (MMLS). Across image instances, truth is outside the trust region in 1,349/4,800 and the observed output reaches its boundary in 1,225/4,800. These tests establish large-ambient-dimension execution and expose a failure regime, not general superiority or theorem coverage. Condition-level paired uncertainty uses 20 anchor clusters; instance-level pooled image intervals are not used.

4.4 Is one frame correction enough?

On 120 clean-anchor neighborhoods we compare one fixed-frame Jet fit (A), a second Jet fit in A's returned tangent frame (B), internal Iterative QF limited to four iterations (C), and full observed-setting optimization (D). All share data, PCA initialization, scale, and uniform weights. A/B retain a free linear term, C moves its base point, and D profiles a no-linear model; this is a matched method comparison, not an iteration-only ablation of one objective. No trust clipping is added to A/B/C. Their returned frames satisfy D's trust/Gram orientation constraints in the main grid. A common constrained no-linear refit supplies an evaluation loss without changing their outputs.

In the noiseless quadratic $b = .5$ condition, mean errors for A/B/C/D are $1.32 \times 10^{-2}, 1.33 \times 10^{-3}, 1.22 \times 10^{-4}, 2.39 \times 10^{-9}$. Full optimization improves noiseless reproduction beyond a single correction. However, all three noisy-quadratic D-minus-B pointwise intervals include zero; this does not establish equivalence. In all six cubic conditions, mean errors satisfy A<B<D and the B-minus-A and D-minus-B intervals are positive. More complete quadratic fitting therefore need not reduce tangent error under model mismatch. D also costs more than the one- and two-fit alternatives. C's iteration limit precludes comparison with a converged Iterative QF solution.

Six separate oracle-assisted saddle diagnostics start at true-frame rotations $\theta_0 = \pm.05$, for $\delta = 0, .1, .5$. At $\delta = 0$, D terminates on gradient tolerance with losses 1.57×10^{-16} and 2.36×10^{-13}, yet errors 5.48×10^{-4} and 3.41×10^{-3}. At $\delta = .1$, both errors are about 2.5×10^{-11}. This connects weak reference curvature with the interpretation of stopping diagnostics; it is not a controlled estimate of solver convergence rates or proof that singularity causes failure.

5 Related Work

Local geometry and learning. Local plane fitting supports reconstruction (Hoppe et al., 1992), including robust variants (Zhan & Yin, 2011; Li et al., 2010). PCA analyses quantify curvature and noise effects (Tyagi et al., 2013; Kaslovsky & Meyer, 2014); multiscale SVD separates local spectral regimes (Little et al., 2017). LTSA aligns tangent approximations into global coordinates (Zhang & Zha, 2004), and MALLER estimates functions and gradients on local-PCA planes (Cheng & Wu, 2013). Here normal-coordinate regression defines the tangent estimation objective itself.

Polynomial and joint fitting. Jets infer differential quantities from fitted derivatives (Cazals & Pouget, 2005); MLS and MMLS construct local coordinates and polynomial approximants (Levin, 1998; Alexa et al., 2003; Sober & Levin, 2020). Aamari & Levrard (2019) jointly estimate projectors and polynomial tensors with nonasymptotic and minimax results. Our zero-intercept criterion has their quadratic normal-output form. The localized dependence $h^2 + \sigma/h$ matches its bias–noise scaling; the conditional noisy bound here is not a new sample-size-dependent statistical rate. SQMF (Zhai & Li, 2024) jointly fits orthogonal linear/quadratic subspaces and per-sample coordinates,

with convergence analysis for its alternating procedure. Our projected predictors remove those free coordinates and expose a query-centered orientation profile (Table 2; Appendix A.2).

Noisy recovery and learned fitting. Aizenbud & Sober (2025) iteratively refine a projected point and tangent through polynomial regression, obtaining sample-size-dependent rates under uniform tubular sampling. Our bounded-normal-noise analysis concerns the query-centered profile and conditional feasible-output bounds. Whole-manifold recovery instead controls set distance and geometry (Genovese et al., 2012; Fefferman et al., 2018). PCPNet predicts normals/curvature (Guerrero et al., 2018); DeepFit and AdaFit learn regression weights or offsets (Ben-Shabat & Gould, 2020; Zhu et al., 2021). Second-order geometry is also the target of Weingarten-map estimation (Cao et al., 2021); GMLS uses local Monge coordinates for vector-valued PDEs (Li et al., 2024). Our target here is tangent orientation, with quadratic coefficients supporting the fit rather than a separately evaluated curvature estimator. Our comparisons concern untrained degree-two estimators across codimensions.

6 Discussion and limitations

Coefficient stability, residualized linear separation, and full orientation curvature are distinct. The cancellation family makes the distinction explicit: nondegenerate designs can coexist with zero or arbitrarily small local curvature, even after small asymmetric perturbations. The symmetric isotropic result shows the opposite, nonnegative contribution. These mechanisms motivate the full residualized derivative as a local landscape object; they do not establish general non-identifiability or solver failure. Their predictive value for practical estimation remains limited by the available evidence. A small follow-up initialization check produced essentially unchanged outputs under the tested perturbations, providing no evidence of initialization sensitivity in that setting. This observation does not contradict the cancellation construction, which concerns local objective curvature rather than sensitivity of every solver trajectory.

Observed-data safeguards are evaluated across the principal domains, but their empirical feasibility does not certify curvature coverage, latent scale, localization, or entry into a positive-curvature basin. The Gram threshold has no observed accuracy effect in the factorial grid. PCA-to-L-BFGS basin entry remains open, as does active-constraint optimization analysis. Fixed-scale bounded-noise error need not decrease with sample count, and an intercept leaves predictor noise and selection bias.

Full optimization improves the tested noiseless quadratic reproduction but has no uniform advantage under noise or cubic mismatch; a single correction can capture much of the useful change at lower cost. External MMLS is often more accurate, with different weighting and fitting-location conventions. The image experiments are out of model and reveal substantial trust-region exclusion. In a follow-up study on selected image conditions, increasing the trust radius from .4 to .8 reduced tangent error, particularly at the higher noise level, whereas reducing it to .2 generally increased exclusion, boundary incidence, and error. Thus the fixed trust radius contributes to the interpretation of the reported image performance. This limited sensitivity study does not establish an optimal radius or an observed-data selection rule. The intrinsic dimension is supplied throughout. These limitations place the contribution in tangent-aligned modeling and conditional analysis of the orientation information retained after coefficient refitting.

Future work. Useful next steps are observed-data control of the full frame Jacobian, initialization and radius selection with verifiable coverage, and downstream tests that connect the fitted quadratic model to a concrete geometric task.

AI USE STATEMENT

Claude Code was used to assist with implementing and executing experiment code. Claude and OpenAI ChatGPT/Codex were used for research discussion, literature comparison, analysis of experimental results, mathematical derivations and proof drafts, manuscript drafting and revision, and preparation of plots and tables from computed data. In particular, AI assistance contributed to the local-landscape extension and to auditing baseline provenance, experiment settings, and aggregation code. Reported measurements originate from executed experiments rather than language-model-generated numerical observations. Checks included source inspection, raw-data reaggregation, derivative sanity checks, and LaTeX compilation; these checks do not replace author review. The authors are responsible for the mathematical claims, experimental validity, citations, and final text.

REPRODUCIBILITY STATEMENT

Appendix C states assumptions and proofs; Appendix B specifies protocols, baseline versions, and uncertainty calculations. Experiment archives contain per-run data, actual configurations, aggregation and plotting scripts, and file manifests. The follow-up diagnostics additionally preserve sampled point clouds, stable seeds, and an executed analysis notebook. The radius study uses disjoint calibration and evaluation seeds. Supplied geometry, observed-data heuristics, and experiments outside the noise model are identified explicitly. The primary comparisons use an observed-setting factorial grid with 10 seeds per synthetic cell and 20 anchors with 10 draws per image cell. Source snapshots, input hashes, raw rows, and supplied aggregate tables are archived. Primary means and radius selections were checked against raw rows. Mechanism and matched-ablation archives retain their executable protocols and numerical outputs; manuscript figures are rendered from those outputs. The mechanism strict-margin comparison with half the reference Gram threshold is a post-processing check of saved fits, not an additional solver run. Some supplied-geometry results in the supplementary record retain historical reproducibility limitations. The unidentified external MMLS fallback warnings are disclosed in Appendix B.9.

REFERENCES

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

Yariv Aizenbud and Barak Sober. manapprox: Manifold approximation software, 2021. URL `https://github.com/aizeny/manapprox`. Commit c517a8ed48fc5924b98a849bf7830d8590fe3a88; ManApprox class.

Yariv Aizenbud and Barak Sober. Estimation of local geometric structure on manifolds from noisy data. *Journal of Machine Learning Research*, 26(64):1–89, 2025. URL `https://jmlr.org/papers/v26/25-0183.html`.

Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Cláudio T Silva. Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(1):3–15, 2003.

Yizhak Ben-Shabat and Stephen Gould. DeepFit: 3D surface fitting via neural network weighted least squares. In *Computer Vision – ECCV 2020*, pp. 20–34, 2020.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023. URL `https://www.nicolasboumal.net/book/`.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015.

Yueqi Cao, Didong Li, Huafei Sun, Amir H. Assadi, and Shiqiang Zhang. Efficient weingarten map and curvature estimation on manifolds. *Machine Learning*, 2021. doi: 10.1007/s10994-021-05953-4.

Frederic Cazals and Marc Pouget. Estimating differential quantities using polynomial fitting of osculating jets. *Computer Aided Geometric Design*, 22(2):121–146, 2005.

Frédéric Cazals and Marc Pouget. Estimation of local differential properties of point-sampled surfaces. CGAL User and Reference Manual, Jet_fitting_3, 2025. URL https://doc.cgal.org/latest/Jet_fitting_3/index.html. Experiments use CGAL 6.1.2.

Ming-Yen Cheng and Hau-Tieng Wu. Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association*, 108(504):1421–1434, 2013. doi: 10.1080/01621459.2013.827984.

John M. Danskin. *The Theory of Max-Min and its Application to Weapons Allocation Problems*. Springer, 1967. doi: 10.1007/978-3-642-46092-0.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959. ISSN 00029947. URL http://www.jstor.org/stable/1993504.

Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Fitting a putative manifold to noisy data. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 688–720. PMLR, 2018. URL https://proceedings.mlr.press/v75/fefferman18a.html.

Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13(43):1263–1291, 2012. URL https://jmlr.org/papers/v13/genovese12a.html.

Gene H. Golub and Victor Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973. doi: 10.1137/0710036.

Paul Guerrero, Yanir Kleiman, Maks Ovsjanikov, and Niloy J. Mitra. PCPNet: Learning local shape properties from raw point clouds. *Computer Graphics Forum*, 37(2):75–85, 2018. doi: 10.1111/cgf.13343.

Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 71–78, 1992. doi: 10.1145/133994.134011.

Daniel N. Kaslovsky and François G. Meyer. Non-asymptotic analysis of tangent space perturbation. *Information and Inference: A Journal of the IMA*, 3(2):134–187, 2014. doi: 10.1093/imaiai/iau004.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

David Levin. The approximation power of moving least-squares. *Mathematics of Computation*, 67 (224):1517–1531, 1998.

Bao Li, Ruwen Schnabel, Reinhard Klein, Zhiquan Cheng, Gang Dang, and Shiyao Jin. Robust normal estimation for point clouds with sharp features. *Computers & Graphics*, 34(2):94–106, 2010. doi: 10.1016/j.cag.2010.01.004.

Rongji Li, Qile Yan, and Shixiao W. Jiang. Generalized moving least-squares methods for solving vector-valued pdes on unknown manifolds. *arXiv preprint arXiv:2406.12210*, 2024. URL https://arxiv.org/abs/2406.12210.

Anna V. Little, Mauro Maggioni, and Lorenzo Rosasco. Multiscale geometric methods for data sets I: Multiscale SVD, noise and curvature. *Applied and Computational Harmonic Analysis*, 43(3): 504–567, 2017.

Colin McDiarmid. On the method of bounded differences. In J. Siemons (ed.), *Surveys in Combinatorics, 1989*, pp. 148–188. Cambridge University Press, 1989. doi: 10.1017/CBO9781107359949.008.

Barak Sober and David Levin. Manifold approximation by moving least-squares projection (MMLS). *Constructive Approximation*, 52:433–478, 2020. doi: 10.1007/s00365-019-09489-8.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12:389–434, 2012.

Hemant Tyagi, Elif Vural, and Pascal Frossard. Tangent space estimation for smooth embeddings of Riemannian manifolds. *Information and Inference: A Journal of the IMA*, 2(1):69–114, 2013. doi: 10.1093/imaiai/iat003.

Zheng Zhai and Xiaohui Li. Subspace-constrained quadratic matrix factorization: Algorithm and applications. *arXiv preprint arXiv:2411.04717*, 2024. URL https://arxiv.org/abs/2411.04717.

Yubin Zhan and Jianping Yin. Robust local tangent space alignment via iterative weighted PCA. *Neurocomputing*, 74(11):1985–1993, 2011. doi: 10.1016/j.neucom.2010.06.029.

Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004. doi: 10.1137/S1064827502419154.

Runsong Zhu, Yuan Liu, Zhen Dong, Yuan Wang, Tengping Jiang, Wenping Wang, and Bisheng Yang. AdaFit: Rethinking learning-based normal estimation on point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6118–6127, 2021.

Table 2: Variables and tangent readouts in local polynomial formulations. SQMF allows a reduced quadratic subspace. Details and the matched-coordinate algebra appear in Appendix A.2.

Method	Coordinates / orientation	Tangent readout
AL	Projected predictors; joint projector/tensors	Estimated projector
SQMF	Free sample coordinates; joint orthogonal subspaces	Derivative of fitted map
Jet	Fixed preliminary frame; polynomial regression	Fitted derivative
MMLS	Refined point/plane; polynomial regression	Fitted derivative at query coordinates
Ours	Projected predictors; profiled frame optimization	Optimized frame at fitted reference point

A Guide to the formulation and its guarantees

The detailed definitions and proofs are collected once in Appendix C. The following references connect the computational formulation in Section 3 to those results.

Geometry and anchor displacement. Proposition C.8 constructs the tangent-aligned graph, identifies its quadratic coefficients with the second fundamental form, and proves the remainder bound. Proposition C.10 gives the observed-coordinate decomposition. At the oracle frame the intercept cancels the common normal anchor offset, while predictor noise and selection effects remain. Adding the self-anchor as a zero regression row would add an intercept penalty to the unnormalized loss.

Criterion and regression subproblem. Proposition C.11 proves the zero-intercept correspondence with the normal-output quadratic criterion of Aamari & Levrard (2019). Section C.4 defines the normalized polynomial features, coefficient and orientation constraints, and profiled estimator. Coefficient conditioning and residualized-linear separation have distinct roles; feasibility alone does not establish recovery.

Reading the guarantees. The returned-loss bound (Corollary C.39) applies to any feasible output under geometric separation and localization. Its loss is observable; its model constants and premises are not certified by that loss. The alternative objective-gap bound requires an externally bounded gap. The gradient bound (Theorem C.56) applies inside a specified strongly convex basin with inactive coefficient constraints. None requires the implemented L-BFGS method to be globally convergent. The separation argument supports recovery, while the full residualized frame derivative supports the landscape result; neither proof is a consequence of the other. The sampling route is: selected latent density, design concentration, observed-design transfer, PCA localization, then recovery.

Numerical outputs. Section C.10 derives the implementation formulas. Section C.11 gives the exact quadratic reference Hessian and a conditional local gradient bound. Section C.12 states the additional basin and iteration assumptions for gradient-descent convergence; these are not an unconditional guarantee for the implemented L-BFGS method.

A.1 Tangent-aligned representation

Local PCA motivates the model as an affine starting point. Allowing a quadratic normal component models curvature while preserving an explicit first-order orientation. Indeed, with a free intercept, setting all quadratic coefficients to zero and removing the orientation constraints recovers mean-centered local PCA at the objective level (Proposition C.12). The anchor-centered PCA used for initialization is a separate choice.

In tangent-aligned coordinates the normal graph has zero first derivative; excluding a free normal slope makes the frame its tangent readout. Figure 2 illustrates this for a rotated quadratic. General smooth manifolds introduce a higher-order remainder.

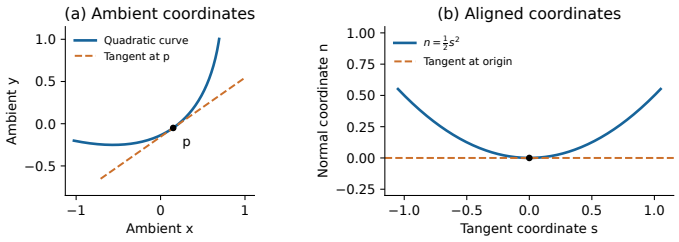


Figure 2: A rotated quadratic graph, shown in ambient coordinates (left) and its tangent–normal coordinates at the clean reference point p (right). Here $\gamma(t) = p + tv_\theta + \frac{1}{2}t^2 w_\theta$ and $\theta = 35°$; in the aligned frame, $n = \frac{1}{2}s^2$ exactly. This illustrates latent geometry, not noiseless observations. For a general smooth manifold, the aligned quadratic representation includes a remainder.

A.2 Relation to SQMF and the coordinate choice

SQMF (Zhai & Li, 2024) represents samples by $a + Uz_i + V\mathcal{A}(z_i, z_i)$, with orthogonal linear and quadratic subspaces and optimized per-sample coordinates z_i. Its quadratic subspace may have dimension smaller than $D-d$. Alternating updates include regression, subspace fitting, and coordinate projection; its convergence results concern that procedure. The optional regularizer acts on sample quadratic displacements.

The following algebra isolates the coordinate distinction. Match the representation to a full normal frame, set $a = x_0 + P_2 c$, and use the same quadratic convention as equation 2. Orthogonality gives

$$\|x_i - [a + P_1 z_i + P_2 \mathcal{Q}_Q(z_i)]\|^2 = \|s_i - z_i\|^2 + \|n_i - c - \mathcal{Q}_Q(z_i)\|^2.$$

Restricting $z_i = s_i = P_1^\top (x_i - x_0)$ recovers our residual. This identity assumes matched centers, sufficient normal span, and no SQMF regularizer; the actual optimization variables and feasible sets differ. Our analysis addresses this projected-coordinate, query-centered profile: coefficient conditioning, orientation separation, and the predictor-rotation contribution to its Hessian. At a general SQMF coordinate, the tangent uses the derivative of its map; at zero it is its linear subspace.

Guarantee scope. AL supplies nonasymptotic statistical and minimax analysis of joint local polynomial estimation. Jet and MMLS supply approximation/convergence theory for their respective fitting constructions. SQMF analyzes its alternating optimization. Here Theorem 3.1 bounds any feasible returned output under localized geometric/design premises; the gradient bound needs an additional strongly convex basin. The implemented L-BFGS trajectory is not covered by the conditional gradient-descent convergence theorem.

B Additional experimental details and results

Reading the experimental record. The primary observed-setting study, cancellation mechanism, and matched frame-correction ablation are specified in Appendix B.9. The earlier studies below preserve their own protocols and seed counts. Supplied-geometry settings are diagnostic comparisons, not evidence that an observed pilot satisfies the recovery premises. In particular, older $\tau = 0$ results do not evaluate the positive-threshold safeguard used in the primary comparison.

B.1 Geometries, sample sizes, and implementations

The external comparison uses paraboloid and saddle exact quadratics, signed cubic surfaces, and a cubic curve. Exact-reproduction radii are $0.03, 0.056, 0.105, 0.196, 0.366, 0.5$; the cubic bias experiment uses eight log-spaced radii from 0.03 to 0.5. These noiseless sweeps scale the latent window with radius, targeting about 200 selected neighbors, and measure fixed-count bias rather than a fixed-noise bias–variance trade-off. The SNR sweep fixes selection radius 0.2 and uses $\sigma/(r + 2\sigma) \in \{0, .05, .10, .15, .20, .30\}$.

Higher-dimensional graphs use $(d, D) \in \{(2, 6), (3, 10), (5, 20), (5, 50), (5, 100)\}$, quadratic or signed cubic terms, and symmetric/asymmetric sampling. Clifford tori with $d \in \{2, 3, 5\}$ are ro-

tated into $D = 20, 100$. The selection-radius axis is $\kappa r \in \{.1, .2, .4, .7, 1\}$, with $\kappa = \|Q^*\|$, and noise levels are $\sigma/(r+2\sigma) = 0, .1$. The target count is $\max\{200, 20(1+d(d+1)/2)\}$, with 30 seeds per cell. Geometry identity, including the sign of the cubic term, is retained in bias and reproduction calculations. Internal Iterative QF is excluded from these performance tables.

Table 3: Higher-dimensional quadratic reproduction and cubic bias. Success counts pool five radii, two sampling designs, and 30 seeds (300 per row). Slopes use noiseless positive-cubic graphs under asymmetric sampling. Jet is internal; MMLS is external. Gram safeguard is disabled.

d	D	Exact successes	Prop. slope	Jet slope	MMLS slope	PCA slope
2	6	300/300	1.944	1.960	1.984	1.278
3	10	297/300	1.971	1.975	1.999	1.219
5	20	295/300	1.980	1.981	1.995	1.454
5	50	298/300	1.970	1.970	1.986	1.620
5	100	295/300	1.946	1.945	1.971	1.709

External code provenance. CGAL `Monge_via_jet_fitting` is called with fitting and Monge degrees both two (CGAL 6.1.2 in the reported environment). Its normal gives $I - nn^\top$; the fitting frame uses mean-centered PCA. The query is passed as the first of $N + 1$ rows, and the returned tangent is at the fitted Monge origin. It is supported only for $(d, D) = (2, 3)$. Internal Jet is never silently substituted in a CGAL result.

External MMLS uses `manapprox` commit `c517a8e` (full commit in the provenance record), `ManApprox.projectPointsGetPoly`, degree two, and zero bias-removal iterations. The main comparison uses `initSVD=True`; the random native initializer is retained as a separate diagnostic, not counted as an additional method in headline ranks. Its bandwidth parameter is explicitly set to the regression scale h; its compact weight has support $1.5h$, and its local-coordinate step can move the base point and change the effective subset. The tangent is the orthonormalized degree-one coefficient block. These procedural differences are not isolated by the comparison.

Numerical settings and supplied constants. The external-baseline, ambient-dimension, and image studies set $\tau = 0$, $r_0 = 0.4$, and $K = 2\|Q^*\|$. The ambient-dimension and image studies use ambient Riemannian L-BFGS; the external comparison uses the original fixed-chart solver. The nominal gradient tolerance is 10^{-9}, relative-loss tolerance 10^{-12}, memory ten, and Armijo constant 10^{-4}. Limits are 300 iterations for the main high-dimensional and multi-start runs, and 200 for images and the original external-comparison solver. Of 12,600 Proposed high-dimensional runs, 4,077 stopped on gradient and 8,523 on relative-loss tolerance. Of 10,800 image runs, the corresponding counts are 1,559 and 9,232, with nine stalled line searches. All returned estimates; this is not a claim of stationarity.

B.2 Radius selection and uncertainty

The external radius grid is $\{.05, .075, .10, .15, .20, .30, .40\}$; the latent sampling window is fixed at 0.52 and absolute noise at $0, .005, .01, .02$. Calibration and evaluation use disjoint 50-seed sets. Each geometry–sampling–noise–method cell selects the lowest calibration mean among radii returning on all 50 seeds. Shared insufficient-data rows are expanded to all applicable methods before counting returns. The clean asymmetric surface cell selects radius 0.075. There are 104 method-specific selections and 5,200 selected evaluation rows.

Paired comparisons match seed IDs within a cell and use 2,000 bootstrap resamples. Intervals are pointwise and conditional on the selected radii; we make no equivalence or simultaneous-significance claim. Surface and curve comparisons use their respective supported method sets. With the external MMLS variants counted once, Proposed has mean rank 4.375 of six on surface cells and 3.50 of five on curve cells. The external MMLS values are 1.50 and 1.25. Ranks supplement, rather than replace, absolute errors.

B.3 IMAGE CONSTRUCTION AND ALL MAIN CELLS

Gaussian-blob families render sums of three to five anisotropic Gaussians on 32×32 grids. Twenty MNIST base images (two per digit) are normalized, blurred at 1.5 pixels, and resampled by cubic interpolation on 28×28 grids (LeCun et al., 1998). Rotation parameters are scaled as arc length at the grid corner. The tangent reference uses central differences with step 10^{-3}; halving that step changes the projector by at most approximately 10^{-7} in the recorded checks. Curvature uses second differences with step 10^{-2} and was not step-validated. It determines K, so uncertainty in that calibration is a limitation.

Pixel noise has per-coordinate standard deviation $\rho/\sqrt{D}$, hence $\mathbb{E}\|\xi\|^2 = \rho^2$. It is neither bounded nor tubular; the reference is the tangent at the latent transformation, not necessarily the nearest projection of the noisy image. Main cells use radius neighborhoods with about 300 points. Additional translation cells compare radius and KNN selection at roughly 150 or 300 points; changing the count also changes the neighborhood extent. No formal equivalence of selection rules is inferred.

Table 4 contains all 24 asymmetric main cells. For each comparison the bootstrap samples the 20 base images, keeping their ten seed replicates together. This quantifies variation over that empirical image set; the selected two-per-digit MNIST set is not a random sample of all possible natural images. Per-method intervals in the accompanying audited CSV use the same clustering. The positive-noise cells and KNN variants are outside the statistical model, and the full image study is a practical transfer test.

B.4 STRUCTURAL ABLATIONS AND NOISE INTERPRETATION

The positive-threshold ablations use 50 seeds per cell and supplied $K = 2\|Q^*\|$, with $\tau = 0.25\lambda_{\mathrm{ref}}$ from an independently seeded reference design. They are distinct from the $\tau = 0$ comparisons. On the one-dimensional cubic examples, the small-tilt expansion $c_{3,\mathrm{eff}} \approx c_3 - \kappa^2\varphi/2$ explains how the preliminary frame can cancel or reinforce cubic bias. The reported correlation of the predicted and observed error ratios is 0.9989 on that diagnostic design; this is a local expansion, not a general formula for high-dimensional jet error.

Free-linear ablations report both the frame and combined tangent readouts. Their small but positive profiled Hessian eigenvalues and start dependence indicate weak conditioning, not exact non-identifiability. Conversely, the combined readout sometimes improves noisy accuracy; choosing $B = 0$ is a geometric modeling choice, not an empirically dominant regularizer.

The internal unweighted Iterative QF approximately agrees with Proposed on some noiseless cells. Its stationarity equation and the joint-frame equation differ in an errors-in-variables term. In the one-dimensional weak-noise heuristic, the tilt difference has leading sign $\Delta\varphi \approx +\kappa\sigma_\zeta^2\mathbb{E}[t]/V_{\mathrm{res}}$, where V_{res} is the residualized linear variance. The prediction agrees in sign on 89.5% of 4,200 runs and 93.2% of 3,600 runs excluding the symmetric population design. The correlation is 0.919 over 28 asymmetry–noise cells pooling radius (0.892 over 84 cells retaining radius). The through-origin regression is observed approximately 0.662 times predicted, so it does not imply global underprediction. The symmetric population leading term vanishes, but its finite-sample estimate need not. This heuristic has not been verified for the external MMLS implementation and is not a theorem about it.

Safeguards, weighting, and computational scope. Noise-free calibration of τ is scale-fragile: normalized quadratic features can shrink as noise increases $h = r + 2\sigma$, with Gram eigenvalues scaling like $(r/h)^4$ when tangential spread remains limited. These rejection percentages therefore cannot be interpreted as intrinsic geometric failure rates. The $\tau = 0$ runs avoid this rejection mechanism but provide no positive-threshold safeguard validation. Trust-region and intercept ablations show both benefits and costs depending on geometry and noise. Gaussian observation weighting is an out-of-theory practical variant, not part of the analyzed unweighted estimator. We make no causal attribution of external MMLS's advantage to weighting alone, and no uniform speedup claim from a single implementation timing.

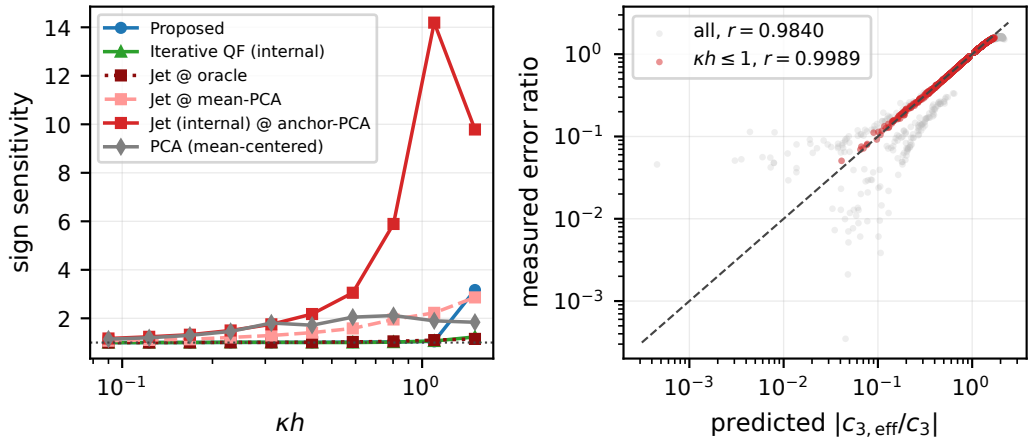


Figure 3: One-dimensional structural diagnostic: cubic-sign sensitivity and a tilted-frame expansion. Jet curves are internal implementations. Error insensitivity to a sign change is not initialization invariance, and the curvature-axis range does not verify the theoretical basin conditions.

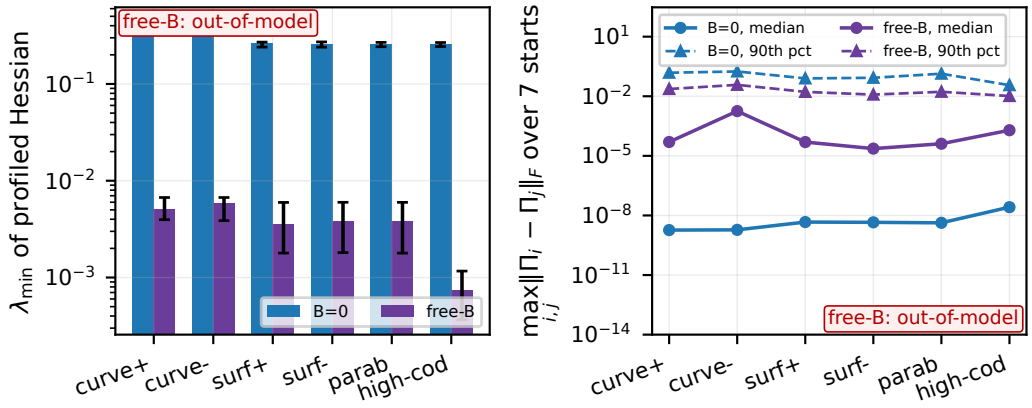


Figure 4: Free-linear ablation for the structural study. Measured small Hessian eigenvalues indicate weak conditioning, not a proof of exact non-identifiability. The frame-only and combined tangent readouts are different outputs.

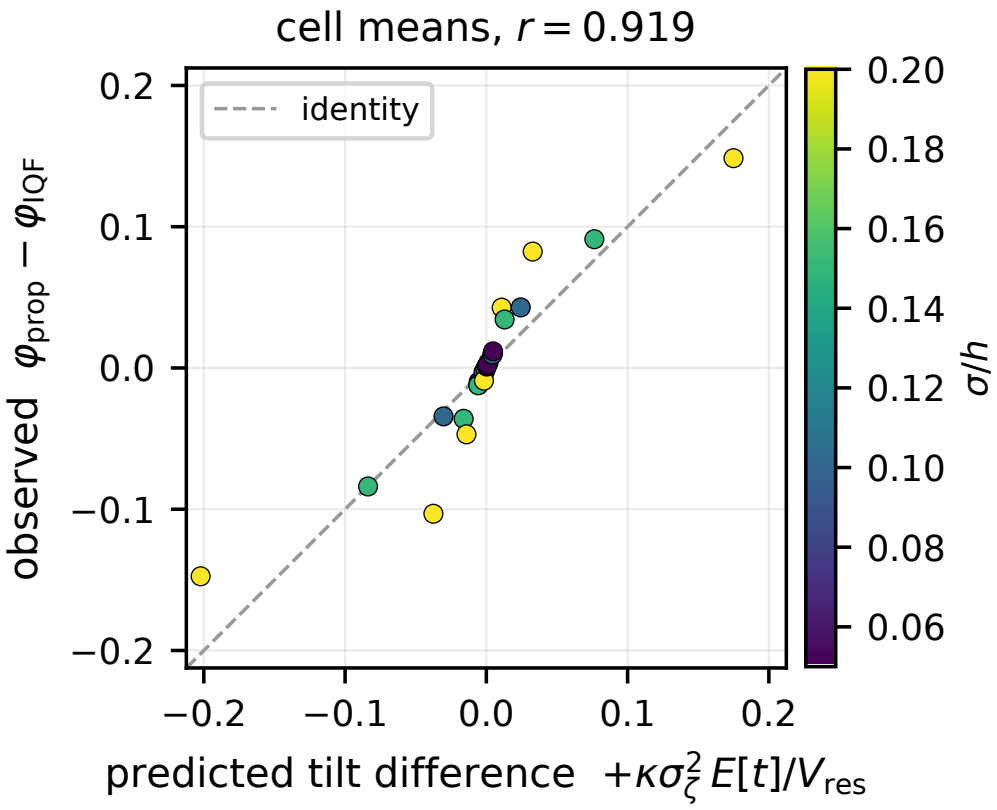


Figure 5: Corrected-sign EIV diagnostic for Proposed versus the internal Iterative QF routine. This is not a comparison with the external MMLS code. The fitted slope below one describes observed versus predicted tilt difference; it is not evidence that the prediction globally underestimates magnitude.

Table 4: All main asymmetric image cells. Mean projector errors and anchor-cluster 95% interval for Proposed minus external MMLS. Jet is internal. Lower is better; negative differences favor Proposed. Radius neighborhoods target 300 points.

Family	Transform	ρ/r	Prop.	Jet	MMLS	PCA	Paired interval
Blob	Rot.	0.00	0.0025	0.0025	0.0016	0.0191	[0.0007, 0.0011]
Blob	Rot.	0.05	0.0094	0.0095	0.0116	0.0211	[-0.0024, -0.0021]
Blob	Rot.	0.10	0.0184	0.0188	0.0231	0.0259	[-0.0049, -0.0045]
Blob	Rot.	0.20	0.0383	0.0414	0.0464	0.0405	[-0.0084, -0.0079]
Blob	Trans.	0.00	0.0364	0.0384	0.0259	0.1020	[0.0094, 0.0116]
Blob	Trans.	0.05	0.0391	0.0417	0.0321	0.1027	[0.0057, 0.0085]
Blob	Trans.	0.10	0.0465	0.0513	0.0455	0.1042	[-0.0004, 0.0025]
Blob	Trans.	0.20	0.0744	0.0906	0.0797	0.1126	[-0.0077, -0.0029]
Blob	Trans.+rot.	0.00	0.0796	0.0860	0.0637	0.1824	[0.0146, 0.0171]
Blob	Trans.+rot.	0.05	0.0828	0.0904	0.0699	0.1823	[0.0113, 0.0146]
Blob	Trans.+rot.	0.10	0.0923	0.1042	0.0852	0.1841	[0.0043, 0.0104]
Blob	Trans.+rot.	0.20	0.2302	0.1949	0.1295	0.1928	[0.0436, 0.1915]
MNIST	Rot.	0.00	0.0032	0.0032	0.0021	0.0221	[0.0010, 0.0014]
MNIST	Rot.	0.05	0.0097	0.0097	0.0117	0.0238	[-0.0023, -0.0018]
MNIST	Rot.	0.10	0.0185	0.0189	0.0231	0.0280	[-0.0049, -0.0043]
MNIST	Rot.	0.20	0.0384	0.0417	0.0464	0.0419	[-0.0083, -0.0077]
MNIST	Trans.	0.00	0.1008	0.1107	0.0756	0.1941	[0.0239, 0.0265]
MNIST	Trans.	0.05	0.1014	0.1129	0.0776	0.1936	[0.0225, 0.0254]
MNIST	Trans.	0.10	0.1037	0.1201	0.0835	0.1935	[0.0185, 0.0220]
MNIST	Trans.	0.20	0.1449	0.1611	0.1048	0.1932	[0.0336, 0.0471]
MNIST	Trans.+rot.	0.00	0.2767	0.2001	0.1374	0.3339	[0.0476, 0.2645]
MNIST	Trans.+rot.	0.05	0.2902	0.2059	0.1407	0.3336	[0.0501, 0.2744]
MNIST	Trans.+rot.	0.10	0.3411	0.2289	0.1499	0.3322	[0.0561, 0.3519]
MNIST	Trans.+rot.	0.20	0.4815	0.3628	0.1827	0.3293	[0.1567, 0.4611]

B.5 Derivative, safeguard, and localization diagnostics

These fixed-neighborhood studies complement the radius-selection experiments. They use the maximum observed displacement as h_{solver}; the nominal radius 0.35 is a noise-generation reference, not an observed selection cutoff. The follow-up uses SHA-256-derived seeds and saved point-cloud realizations. The initial pilot-K study uses its recorded raw results and is not paired with the new realizations. The latter avoids reliance on a process-dependent hash used in that initial study's seed mapping.

Analytic derivative check. At $d = 2$ and codimension one or two, exact quadratic graphs use two curvature scales, well-conditioned or asymmetric anisotropic sampling, and anchor offsets either zero or with normal components between 0.06 and 0.10. Each of 16 configurations has 180 rows and three finite-difference steps $10^{-4}, 3 \times 10^{-5}, 10^{-5}$. The analytic map $(I - \mathsf{P}_\Phi)\mathcal{D}/\sqrt{N}$ is constructed directly from Q^*, t, and a_0. Independently, central differences refit coefficients at each perturbed frame to obtain the residual Jacobian and loss Hessian. The residual derivative has sign $-\mathcal{D}$, consistently in all 48 checks. At step 10^{-5}, relative Frobenius errors are at most 1.093×10^{-9} for the Jacobian and 9.349×10^{-10} for the Hessian. Omitting the predictor-rotation term changes the analytic Jacobian by 0.0149–0.2798 relative to its full norm. The fits are unconstrained; this checks the exact-reference expression, not second derivatives across active curvature constraints or a noisy nonzero-residual identity.

Positive Gram thresholds. Two geometries (quadratic and cubic surfaces), two designs, two noise levels, and 12 seeds give 96 instances with $N = 180$, $r_0 = 0.4$, and $K = 2\|Q^*\|_F$. Relative thresholds use $\alpha = 0, 0.5, 0.75, 0.9$ times the initial observed Gram eigenvalue. All instances return at every relative threshold with unchanged paired errors. An absolute threshold is fixed to the lower quartile of initial Gram values on a separate 32-instance calibration set covering both designs and noise levels. Table 5 reports its behavior. The noise coordinate is $\rho/0.35$; the generator's stated bound is $\sigma = \sqrt{3}\rho$, and both ρ/h_{solver} and $\sigma/h_{\mathrm{solver}}$ are retained in raw records. The relative rule shrinks with a poorly conditioned initializer, while the absolute rule need not distinguish noise-inflated predictor variation from useful tangent information. No accuracy certificate follows from these checks.

Table 5: Absolute-threshold diagnostic, 24 instances per row. The clean near-degenerate cases have median error 0.0143 with $\tau = 0$. Rejection is not counted as successful estimation.

Design	Noise coordinate	Returned	Median error
Near-degenerate	0	0/24	—
Near-degenerate	0.08	24/24	1.4035
Well-conditioned	0	24/24	0.0252
Well-conditioned	0.08	24/24	0.0452

Curvature-bound and localization diagnostics. The pilot fits an intercept and quadratic terms, without a linear term or regularization, in the observed PCA frame, and sets $K_{\mathrm{pilot}} = 2\|\widehat{Q}_{\mathrm{pilot}}\|_F$. On 96 pairs with $\tau = 0.05\lambda_{\min}(\widehat{\Sigma}(P_{\mathrm{PCA}}))$, its median error difference from $K = 2\|Q^*\|_F$ is zero and mean difference is 1.39×10^{-4}. In 72 pairs whose truth is inside the trust region, both returned curvature constraints are inactive and the median difference is zero. The pilot bound falls below the true curvature in 3/96 cases (one inside the trust region). This is a heuristic, not a validated upper bound.

A separate follow-up compares PCA and perturbed starts, and oracle starts when feasible, on four conditions with six seeds each. The projector trust region stays PCA-centered, but the supplied start also centers the numerical chart. Clean quadratic cases recover to numerical precision; cubic/noisy cases have median error 0.0664 from all tested starts. In the near-degenerate noisy graph and torus conditions, truth remains outside both $r_0 = 0.4$ and $r_0 = 1.2$, with PCA-start median errors 1.398 and 1.414. Thus these stress cases do not isolate behavior after the truth becomes admissible. Final loss, coefficient norm, Gram and feasibility are recomputed; recorded gradient norms are history diagnostics, not independently certified final-point gradients. Noisy observed coordinates, including projections at an oracle frame, are distinguished from latent coordinates in the theory.

Small-dimensional fitting cost. With $d = 2, D = 3$, eight data seeds per sample count and repeated timings after warm-up, Table 6 includes PCA initialization. Data generation and neighborhood search are excluded. Counts of timing repeats are not counts of independent data instances. Jet and Iterative QF are internal implementations; external-library timing is not inferred from them. The supplied K and threshold are fixed before timing, so this does not include a pilot-based tuning pipeline.

Table 6: Median initialization-including fitting time (ms), small-dimensional quadratic/noisy diagnostic. Post-initialization timings are retained separately in the experiment archive.

N	Anchor PCA	Jet (int.)	Iterative QF (int.)	Proposed
80	0.0431	0.1310	0.6648	2.2111
160	0.0436	0.1395	0.7230	2.0467
320	0.0443	0.1622	0.8567	2.2435

B.6 Intrinsic dimension at fixed ambient dimension

At $D = 100$, $d \in \{5, 10, 15\}$, define $p = 1 + d(d+1)/2$, $q = p + d$. The two protocols use $N = 200$ and $N = 10q$, respectively (210, 660, and 1360). Twenty seeds per cell, sampling asymmetry $\beta \in \{0, 0.6\}$, relative cubic magnitude $0, 0.1$, and noise ratio $\sigma/h_{\text{latent}} \in \{0, 0.05\}$ give 960 instances and 4800 main method evaluations. Each seed draws a random geometry and ambient rotation; methods share each instance. Total curvature is fixed at $K_F = 2$, $h_{\text{latent}} = 0.3$, with $K = 2K_F$, $\tau = 0$, and $r_0 = 0.4$. Cubic normalization uses 1500 sphere probes, not an exact supremum. Bounded noise is normal at each latent point. As normal dimension falls from 95 to 85, fixed noise norm increases the per-direction variance by the factor $95/85$, and standard deviation by $\sqrt{95/85}$. Sample graph displacement and gradient also vary with d; this design does not isolate intrinsic dimension as a sole cause.

Table 7: Mean projector error in the asymmetric cubic/noisy cell ($\beta = 0.6$, relative cubic magnitude 0.1, $\sigma/h_{\text{latent}} = 0.05$), 20 seeds per entry. Jet is internal, MMLS is external; Gram safeguard is off.

Protocol	d	Proposed	Jet (int.)	MMLS (ext.)	PCA (mean)
$N = 200$	5	.022291	.022311	.027723	.046946
	10	.040746	.040769	.051512	.050803
	15	.084106	.086080	.094870	.063426
$N = 10q$	5	.022020	.022034	.027423	.046474
	10	.020141	.020150	.026679	.029865
	15	.019735	.019743	.025998	.024772

All methods return on all instances; Proposed has one maximum-iteration termination at $d = 15$, with the other 959 runs ending on gradient or relative-loss tolerances. Truth lies inside the trust region on all 960 instances, and no return is on its boundary. The largest PCA projector error over these instances is 0.129. At the displayed fixed-N cell, five seeds per dimension compare PCA and oracle starts with the same PCA-centered trust region; paired error differences are at most 3.19×10^{-8}. Oracle initialization provides no measured benefit in those 15 cases, rather than excluding initialization effects in general.

At $\beta = 0.6$, noiseless quadratic threshold counts $e < 10^{-8}$ are 20/20, 18/20, 0/20 for fixed N, and 20/20, 20/20, 18/20 for $N = 10q$. The corresponding mean errors at $d = 15$ are 4.86×10^{-8} and 6.55×10^{-9}; threshold failure therefore need not indicate a large estimation error. In the displayed noisy cell, the median residualized-linear eigenvalue at the returned observed coordinates falls from 0.0952 to 0.0076 at fixed N, versus 0.0933 to 0.0420 with $N = 10q$. These are observed-design associations, not verification of the latent separation assumption or proof of causation.

Instrumented high-dimensional cost. The high-dimensional wrappers time fitting together with method-specific diagnostics and error evaluation under `tracemalloc`; Proposed and internal Jet receive a precomputed PCA frame. These are not the same timing boundary as Table 6. At the

displayed $N = 10q$ cell, Proposed median wrapper times are 0.030, 0.270, and 1.354 s, and tracked peak allocations average 1.7, 11.9, and 65.5 MB. Tracked allocations are not whole-process peak resident memory. External MMLS uses the unchanged `manapprox` implementation, whose basis-construction enumeration is reported in the run provenance as costly at $d = 15$; library costs are not an algorithmic lower bound. The dimension-15 pass uses five workers rather than twenty. These measurements document this implementation and environment, not a universal complexity or speed advantage.

B.7 Audit of baseline evaluation locations

All external quadratic-reproduction comparisons score returned projectors against the tangent at the same latent anchor $t = 0$. Proposed returns $\mathrm{im}(P_1)$, the tangent at its fitted point $x_0 + P_2 c$. CGAL receives the query first and returns the normal at its fitted Monge origin. The MMLS wrapper extracts the linear block of the polynomial map fitted around its refined affine base q; the associated surface point is $q + p(0)$, not q. Its affine update makes the query's coordinates relative to q zero, so this is a query-associated readout. The recorded scalar reference-point shift measures $\|q\|$, not the surface-point displacement. Internal Jet evaluates its derivative at query coordinates zero.

The reproduction archive contains 600 instances (two surfaces, six radii, 50 seeds): Proposed meets the 10^{-8} threshold in all 600, and CGAL and external MMLS in none; all return estimates. These outcomes compare native query-to-tangent procedures. Fitted-point vectors, polynomial coefficients, and full projectors were not retained in this archive, so it cannot retrospectively isolate evaluation-location effects or compare all fitted derivatives at an identical spatial point. Source inspection explains the wrapper semantics; it does not establish byte-identical provenance of every historical executable.

B.8 Revision diagnostics: geometry, observed settings, and fitted locations

These new runs use the supplied solver and external baselines without changing their numerical algorithms. Removing an eager import of the optional memory profiler allows the numerical package to load without `psutil`; no memory measurements are reported here. The execution environment is Python 3.12.14, NumPy 2.3.5, and SciPy 1.17.0, with a single OpenBLAS thread. This differs from the environment recorded for historical runs. Seeds are obtained from SHA-256 hashes of explicit experiment identifiers. Per-run CSVs and NPZ clouds preserve the data and fitted outputs.

Independent checks of existing results. The external radius table was reaggregated from its original raw rows. Shared `insufficient_data` events are expanded to all methods supported for that geometry before checking whether a calibration radius returns 50/50 estimates. Selecting the eligible radius with the lowest calibration mean produces 104 method-specific selections and 5,200 evaluation rows, all returned; the displayed means in Table 15 are recovered. This is raw-data reaggregation, not a new simulation of those instances. The 48 recorded analytic-derivative configurations were rerun using the original numerical functions and seed mapping, isolating them from optional notebook-reporting imports. The maxima at step 10^{-5} remain 1.093×10^{-9} and 9.349×10^{-10} for the Jacobian and Hessian. The largest change in the reference Hessian's minimum eigenvalue from the archived value is below 10^{-16}.

Symmetric-graph mechanism. For $(d, m) \in \{(1,1),(1,3),(2,1),(2,3)\}$, draw 60 directions from normalized Gaussian vectors and radial factors uniformly from $[.1, 1]$. Rescale their maximum norm to $h = .4$ and append their negatives, giving 120 exactly paired coordinates. Choose a fixed random unit normal direction and $\|\kappa\|h \in \{.1, .25, .5, 1, 2, 4\}$, with $g(t) = \kappa\|t\|^2/2$. Compare the moment expression of Corollary C.51, the full residualized-Jacobian Gram matrix, and the central finite-difference Hessian of the refitted loss at step 10^{-5}. These are unconstrained reference fits; the same expression applies locally to the constrained profile when its strict margins hold. This experiment checks the structural identity, not the perturbation conditions or the numerical solver's basin. All 24 full Hessians are positive; the sufficient margin is nonpositive in 10 configurations. The maximum finite-difference relative Frobenius error is 4.49×10^{-9}.

Observed-data rule. Use $d = 1, D = 2$ with $Q^* = [1.2]$, or $d = 2, D = 3$ with $Q^* = \mathrm{diag}(1.2, -.8)$, and $g(t) = t^\top Q^* t/2 + .3\sum_j t_j^3$. Generate 900 non-anchor latent coordinates in

$[-.52, .52]^d$. The symmetric design is uniform on this box. The asymmetric design replaces the first coordinate by a uniform magnitude on $[0, .52]$ and a positive sign with probability .8. A third, thin design in $d = 2$ scales the second coordinate by .08 and retains symmetric sampling. Draw independent uniform signed normal noise on $[-\sigma, \sigma]$, with the unit normal computed at each latent point; the anchor at $t = 0$ has independent noise of the same law. Use $\sigma \in \{0, .005, .02\}$ and 20 seeds per cell, for 300 instances in total.

Select the 180 nearest observed non-anchor points, equivalently using the 180th distance as an adaptive observed radius. Set $h_{\mathrm{solver}} = \max_i \|u_i\|$, compute anchor-centered PCA P_0, and obtain an unconstrained intercept-plus-quadratic pilot in that frame. The observed rule uses $K = 2\|\widehat{Q}_{\mathrm{pilot}}\|_F$, $\tau = .05\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0))$, and $r_0 = .4$; the dimension is supplied. These rules and the sample-count choice are fixed for this study and are not calibrated using evaluation errors. The paired oracle variant uses the same observed neighborhoods, $K = 2\|Q^*\|_F$, and $\tau = 0$. Internal Jet uses the same anchor-PCA frame and its combined derivative readout; mean-centered PCA and external MMLS use their usual readouts. External MMLS uses SVD initialization and bandwidth h_{solver}. All methods return on all 300 instances.

Table 8: New observed-data diagnostic: mean projector errors, 20 seeds per row. "Oracle" supplies curvature but uses the same adaptive neighborhood. Jet is internal; MMLS is external. All 1,500 method–instance runs return.

d	Design	σ	Observed	Oracle	Jet	MMLS	PCA
1	asymmetric	0	0.00295	0.00295	0.00275	0.00180	0.01976
1	asymmetric	0.005	0.00547	0.00465	0.00470	0.00727	0.02091
1	asymmetric	0.02	0.02022	0.02025	0.02123	0.01848	0.02920
1	symmetric	0	0.00270	0.00270	0.00271	0.00175	0.00293
1	symmetric	0.005	0.00480	0.00480	0.00479	0.00563	0.00669
1	symmetric	0.02	0.01496	0.01496	0.01514	0.01814	0.01580
2	asymmetric	0	0.02036	0.02036	0.01948	0.01343	0.07667
2	asymmetric	0.005	0.02096	0.02096	0.02006	0.01514	0.06580
2	asymmetric	0.02	0.02401	0.02401	0.02283	0.02248	0.07717
2	symmetric	0	0.02050	0.02050	0.02034	0.01414	0.02756
2	symmetric	0.005	0.02115	0.02115	0.02110	0.01641	0.02565
2	symmetric	0.02	0.02101	0.02101	0.02092	0.01998	0.02315
2	thin	0	0.00275	0.00275	0.00275	0.00183	0.01095
2	thin	0.005	0.01169	0.01169	0.01171	0.01143	0.01564
2	thin	0.02	0.09880	0.09658	0.09950	0.07672	0.06491

Returns, feasibility, and uncertainty. The observed solver stops on gradient tolerance in 193 runs and relative-loss tolerance in 107. The oracle variant has 193 gradient stops, 106 relative-loss stops, and one stalled line search. A returned estimate is not necessarily a stationary point. Final normalized PCA-chart gradients are independently recomputed and retained in the CSVs. The curvature constraint is active in 8 observed and 13 oracle runs. The observed pilot underbounds true curvature in 12 instances; five true tangents are outside the trust region. The observed true-frame curvature, Gram, and trust-region conditions hold jointly in 283/300, but this is not the complete recovery or local-basin certificate.

The conservative localization check $\|\Pi_0 - \Pi^*\|_F + r_0 < 1/\sqrt{5}$ holds in 189/300. Its failure does not prove that the feasible region leaves the required chart. The implemented scale bounds the realized latent-coordinate norms in 249/300. An initializer-relative Gram threshold cannot certify orientation separation, and the adaptive radius is outside the prescribed-radius concentration statement. Pointwise paired bootstrap intervals (2,000 seed resamples) are provided for observed-minus-oracle errors in each cell, without an equivalence claim or multiple-comparison correction. In the noisy thin cell at $\sigma = .02$, the mean difference is .00222, with interval $[.000022, .00491]$; the practical rule is not uniformly interchangeable with supplied curvature.

Linear separation versus the full reference derivative. For each selected cloud, the saved latent coordinates define $h_{\mathrm{latent}} = \max_i \|t_i\|$ and a quadratic reference with the true Q^* and the same anchor offset. The reference-Jacobian diagnostics use this latent scale, separately from h_{solver}.

They are not the Hessian of the noisy, cubic empirical loss. The minimum-eigenvalue ratio of full to linear-only reference Hessians lies between .962 and 1.070 on these 300 instances. Their Spearman correlation is .999. Thus these modest-curvature cases do not demonstrate an additional failure-prediction advantage for the full derivative; Figure 6 isolates its structural contribution in a controlled regime. No nonzero-residual Gauss–Newton matrix is claimed to be an exact Hessian.

Recorded fitted locations. A separate 60-instance experiment uses the clean paraboloid $g(t) = t^\top \operatorname{diag}(1.5, 1)t/2$, 240 samples, latent-window half-widths $b \in \{.05, .2, .5\}$, and 20 seeds per scale. The second coordinate is uniform on $[-b, b]$; the first has a uniform magnitude on $[0, b]$ and a positive sign with probability .8. There is no radius selection in this diagnostic, so b is not the selection radius used in Figure 7. The solver uses $h_{\text{solver}} = \max_i \|u_i\|$, $K = 2\|Q^*\|_F$, $\tau = 0$, and $r_0 = .4$.

For each method, retain the fitted point $\widehat{x}$, the tangent projector, and available polynomial coefficients. For MMLS the fitted point is $q + p(0)$, not the affine base q; both are saved. CGAL supplies its fitted Monge origin. The primary target remains the tangent at $t = 0$. For a secondary location diagnostic, associate $\widehat{x}$ with the true graph point having tangent coordinates $\widehat{x}_{1:2}$. This is an explicit graph-coordinate association, not a nearest-point projection or a common physical evaluation point for every method.

Table 9: Recorded-location diagnostic. Mean primary anchor error and secondary graph-associated error. The last column is the true tangent shift between the two targets. Twenty instances per scale and method.

b	Method	Anchor error	Associated error	Target shift
0.05	CGAL	2.67e-05	2.66e-05	4e-09
0.05	Jet	7.98e-05	7.98e-05	3.39e-08
0.05	MMLS	1.25e-05	1.25e-05	9.83e-10
0.05	Proposed	5.71e-10	5.71e-10	1.93e-21
0.2	CGAL	0.00149	0.00149	3.3e-06
0.2	Jet	0.00484	0.00481	3.18e-05
0.2	MMLS	0.000705	0.000704	7.64e-07
0.2	Proposed	1.72e-09	1.72e-09	7.42e-20
0.5	CGAL	0.0259	0.0256	0.000384
0.5	Jet	0.0791	0.0749	0.0042
0.5	MMLS	0.0136	0.0135	0.000104
0.5	Proposed	0.00214	0.00214	2.45e-06

Changing the target in this way does not remove the external methods' error on the new samples. Oracle-frame Jet has error below 10^{-8} in all 60 instances, showing that the coordinate choice is material. Proposed meets that threshold in 58/60. Its two failures occur at $b = .5$, with PCA errors .407 and .432 exceeding $r_0 = .4$; both terminate at the trust boundary. All CGAL, MMLS, and anchor-PCA Jet outputs exceed the reproduction threshold. These results narrow the location ambiguity for this family but do not isolate weighting, refinement, and polynomial parameterization separately, or retroactively reconstruct the historical reproduction archive.

B.9 Observed-setting factorial comparisons

The primary results use the completed ten-seed observed-safeguard archive (recorded on 26 September 2026). It contains 93,760 method rows: 17,920 for the radius study, 33,600 for ambient dimension, 3,840 for intrinsic dimension, and 38,400 for images. Eight method labels are retained per instance, including unsupported CGAL rows. There are no duplicate instance–method pairs. A/B/C/D denote observed curvature with Gram on/off and oracle curvature with Gram on/off, respectively. Baselines are internal Jet, external MMLS, mean-centered PCA, and supported CGAL Jet. The A/B/C/D input hashes and recorded setup agree; the common initializer is confirmed by the runner's shared code path, not a stored initializer hash.

Estimator settings and numerical interpretation. The pilot is an unconstrained intercept-plus-quadratic fit at anchor PCA, without a linear term. With the orthonormal quadratic basis, its actual curvature norm is $\|C_q\|_F/h_{\text{obs}}^2$, not $\|C_q\|_F$. All factorial arms use $h_{\text{obs}} = \max_i \|u_i\|$, $r_0 = .4$,

Table 10: Observed-scale held-out radius comparison. Both displayed proposed arms have Gram on. Jet is internal in both geometries. Radius selection is supervised and method-specific; values are means over 20 evaluation instances per row. CGAL's supported surface rows are retained in the data archive.

Geometry	σ	Obs.	Oracle	Jet	MMLS	PCA
Curve	0	0.00382	0.00243	0.00230	0.00158	0.01400
Curve	0.005	0.00903	0.00783	0.00692	0.00943	0.02033
Curve	0.01	0.01539	0.01342	0.01138	0.01095	0.02148
Curve	0.02	0.01930	0.01744	0.02179	0.01835	0.03464
Surface	0	0.00261	0.00261	0.00260	0.00232	0.02536
Surface	0.005	0.01801	0.01801	0.01799	0.01612	0.03237
Surface	0.01	0.02428	0.02428	0.02383	0.02382	0.03966
Surface	0.02	0.04170	0.04170	0.03981	0.03251	0.05469

and the same ambient L-BFGS implementation. The observed arm takes $K = 2\|\widehat{Q}_{\text{pilot}}\|_F$ and $\tau = .05\lambda_{\min}(\widehat{\Sigma}(P_0))$; the oracle arm replaces only K by $2\|Q^*\|_F$. Gram-off sets $\tau = 0$. Dimension is supplied, but true geometry and noise are not inputs to the observed settings function. The radius benchmark still uses supervised calibration. Zero pilot curvature gives $K = 0$, and a nonpositive initial Gram is clipped to zero in the supplied implementation, yielding an inactive relative threshold. These cases must not be interpreted as positive-threshold certificates; all reported observed-arm thresholds in this grid are positive. Invalid data or insufficient rows are not silently replaced by oracle fits.

Domains and selection. The C4 grid uses cubic curves/surfaces, symmetric/asymmetric latent sampling, absolute noise $\sigma \in \{0, .005, .01, .02\}$, latent half-window .52, and radii $\{.05, .075, .1, .15, .2, .3, .4\}$. Ten calibration seeds and ten disjoint evaluation seeds are used per condition/radius. The manuscript reaggregation requires ten returned calibration estimates for radius eligibility; all supported method/radius combinations in this archive meet that requirement. It selects the lowest calibration mean, breaking ties by smaller radius. The 120 selected supported method–condition combinations give 1,200 held-out method rows. Each displayed method in Table 10 contributes 160 held-out evaluations. The table averages two sampling designs with equal seed counts. Original generator scales $r + 2\sigma$ are stored but not supplied to the matched observed-scale solvers. A matched oracle arm is consequently not claimed to reproduce the older implementation bit for bit.

The ambient study retains the 420 condition grid of polynomial graphs and rotated Clifford tori, ten seeds per condition. Polynomial settings include $(d, D) = (2, 6), (3, 10), (5, 20), (5, 50), (5, 100)$, quadratic and both cubic signs; tori cover $d = 2, 3, 5$ in $D = 20, 100$. Both sampling designs, five radii indexed by curvature, and two noise levels are retained. Geometry determines the synthetic benchmark grid, but does not enter the observed estimator. The intrinsic study retains 48 conditions, ten seeds each, through $d = 15, D = 100$, with fixed and dimension-scaled sample counts. Summaries below pool conditions descriptively; condition CSVs retain each noise, curvature, sampling, and sample-size combination separately. Images retain all 24 asymmetric main conditions: two families, three transformations, and four pixel-noise levels, with 20 anchors and ten draws per anchor. No r_0-sensitivity experiment is included.

Feasibility, termination, and uncertainty. All four proposed arms return on 11,720 instances, while CGAL is unsupported on 10,600. Unsupported cases are not counted as numerical failures. For the observed Gram-on arm, radius-study terminations comprise 1,302 gradient stops, 936 relative-loss stops, one iteration limit, and one line-search stall; ambient dimension has 1,342 gradient and 2,858 relative-loss stops; intrinsic dimension has 198 and 282; images have 1,031 gradient, 3,766 relative-loss, and three stalled stops. A returned or relative-loss-stopped output is not certified optimal. Supplied gradient fields based on the last history entry are not treated as independently recomputed final KKT certificates. At active boundaries, an unconstrained gradient alone is not an appropriate constrained optimality test.

Table 11: Ambient-dimension means under observed-scale matched settings. Polynomial rows average the three cubic-coefficient settings, both sampling designs, five radii, and two noise levels; torus rows average the latter three factors. Each condition has ten seeds. These are descriptive aggregates.

Family	d	D	Obs.	Oracle	Jet	MMLS	PCA
Torus	2	20	0.01189	0.01189	0.01623	0.01595	0.04286
Torus	2	100	0.01204	0.01204	0.01656	0.01610	0.04519
Torus	3	20	0.01610	0.01610	0.01795	0.02184	0.04271
Torus	3	100	0.01636	0.01636	0.01823	0.02181	0.04347
Torus	5	20	0.01920	0.01910	0.02002	0.02648	0.03630
Torus	5	100	0.01923	0.01915	0.02002	0.02692	0.03654
Polynomial	2	6	0.07436	0.07439	0.07684	0.05925	0.09399
Polynomial	3	10	0.06203	0.06203	0.06280	0.05334	0.08168
Polynomial	5	20	0.07137	0.07127	0.07162	0.06561	0.08232
Polynomial	5	50	0.10909	0.10902	0.10927	0.09522	0.12005
Polynomial	5	100	0.14786	0.14778	0.14802	0.12766	0.15858

Table 12: Intrinsic-dimension means, pooled within dimension and sample-count protocol. Each constituent condition has ten seeds; means do not establish performance uniformity across the pooled geometry/noise conditions.

d	Protocol	Obs.	Oracle	Jet	MMLS	PCA
5	N10q	0.01288	0.01288	0.01294	0.01541	0.03954
5	fixed_N	0.01343	0.01343	0.01349	0.01592	0.04120
10	N10q	0.01162	0.01162	0.01163	0.01482	0.02542
10	fixed_N	0.02188	0.02188	0.02189	0.02710	0.04396
15	N10q	0.01107	0.01107	0.01107	0.01417	0.01989
15	fixed_N	0.04458	0.04384	0.04464	0.04875	0.05141

In radius/ambient/intrinsic/image order, observed pilot undercoverage counts are 129, 25, 0, 0; true-tangent trust exclusion counts are 92, 182, 0, 1,349; returned trust-boundary counts are 40, 1, 0, 1,225. The same observed-pilot audit attached to an oracle row does not describe oracle curvature coverage. The stored sufficient chart check is not a solver-basin test, and its failure does not prove chart exclusion. Latent scale and true-frame feasibility remain separate from feasibility of the returned output. Image geometry is used for diagnostics only; its noise model remains outside the theorem.

Condition-specific image paired uncertainty resamples anchor-level mean differences over 20 anchors. The supplied pooled image intervals instead resample instance IDs and are not used in the manuscript. Synthetic condition-specific intervals have ten seed units and are pointwise, not simultaneous guarantees. Pooling repeated seed labels across radii or conditions does not create additional independent seed units; descriptive pooled means are not assigned such confidence intervals here.

External-library and archival qualifications. Two external MMLS weighted-least-squares warning lines were recorded during the radius stage. Source inspection identifies a least-squares fallback, not an exception, but the warning was not captured per row; the affected row identities and exact call count cannot be reconstructed. All 11,720 MMLS errors are finite, which alone does not certify ordinary well-conditioned solves. MMLS comparisons retain this disclosed limitation. A reported 7.612-second C4 file-modification span is not estimator runtime and is not used for timing claims. No raw rows from the earlier partial run are mixed into the completed primary archive.

B.10 Cancellation mechanism and asymmetric perturbations

The standalone experiment uses exactly the intercept-bearing feature basis of Proposition C.52, with $h = 1$, clean anchor, and $Q = \mathrm{diag}(1, -1)$. The scale is a normalization, not a bound on the latent radius. At each tilt angle the intercept and all quadratic coefficients are refit. Directional finite differences use steps $10^{-1}, .05, .02, .01, .005, .002, .001, .0005$. The primary singularity claim follows from the analytic identity; finite differences are numerical checks. The minimum coefficient

Table 13: All 24 observed-setting main image conditions. Each row averages 20 anchors with ten draws each. Both proposed columns use Gram on; Jet is internal and MMLS is external. These are out-of-model transfer tests.

Family	Transform	ρ/r	Obs.	Oracle	Jet	MMLS	PCA
Blob	Rotation	0	0.00247	0.00247	0.00247	0.00159	0.01913
Blob	Rotation	0.05	0.00942	0.00942	0.00947	0.01165	0.02105
Blob	Rotation	0.1	0.01844	0.01844	0.01876	0.02312	0.02586
Blob	Rotation	0.2	0.03830	0.03830	0.04137	0.04645	0.04052
Blob	Translation	0	0.03640	0.03640	0.03844	0.02593	0.10201
Blob	Translation	0.05	0.03913	0.03913	0.04173	0.03205	0.10267
Blob	Translation	0.1	0.04652	0.04652	0.05126	0.04545	0.10423
Blob	Translation	0.2	0.07436	0.07436	0.09057	0.07972	0.11261
Blob	Both	0	0.07959	0.07959	0.08599	0.06372	0.18236
Blob	Both	0.05	0.08277	0.08277	0.09040	0.06992	0.18229
Blob	Both	0.1	0.09231	0.09231	0.10425	0.08522	0.18414
Blob	Both	0.2	0.24346	0.23022	0.19490	0.12954	0.19276
MNIST	Rotation	0	0.00323	0.00323	0.00324	0.00209	0.02207
MNIST	Rotation	0.05	0.00966	0.00966	0.00973	0.01174	0.02379
MNIST	Rotation	0.1	0.01853	0.01853	0.01891	0.02314	0.02796
MNIST	Rotation	0.2	0.03840	0.03840	0.04172	0.04642	0.04193
MNIST	Translation	0	0.10077	0.10077	0.11073	0.07556	0.19411
MNIST	Translation	0.05	0.10142	0.10142	0.11288	0.07761	0.19358
MNIST	Translation	0.1	0.10368	0.10368	0.12011	0.08350	0.19347
MNIST	Translation	0.2	0.14490	0.14490	0.16107	0.10483	0.19320
MNIST	Both	0	0.27818	0.27668	0.20006	0.13741	0.33387
MNIST	Both	0.05	0.29125	0.29018	0.20590	0.14067	0.33360
MNIST	Both	0.1	0.34411	0.34113	0.22888	0.14986	0.33219
MNIST	Both	0.2	0.48354	0.48153	0.36282	0.18272	0.32929

Gram eigenvalue at $\delta = 0$ is approximately .05976 and its condition number is about 108. Thus we claim nondegeneracy, not exceptionally good numerical conditioning.

The source archive's margin table used $\tau = \lambda_{\min}(\Sigma_{\text{truth}})$, putting truth on the Gram boundary. For the strict-interior interpretation we instead evaluate the *saved unconstrained fits* at $\tau = \lambda_{\min}(\Sigma_{\text{truth}})/2$, $K = 2$, and $r_0 = .2$ centered at truth. Across the six δ values and $\theta \in \{-.05, 0, .05\}$, the minimum Gram, curvature, and trust margins are .02988, .57666, and .12932. This is post-processing, not a new constrained solver run. Continuity and strict reference margins give local coincidence of profiles; the three numerical points do not verify an entire interval of frames.

For $i = 1, \ldots, 12$, the asymmetric displacement is

$$e_i = \eta\big(\sin(1.7i) + .37\cos(.4i),\ \cos(2.3i) + .21\sin(.9i)\big).$$

We recompute $g(t_i + e_i)$ from the exact graph. The scale parameter η is not identical to RMS displacement ε in Corollary C.53. Full Σ, residualized S_{lin}, and both Hessian eigenvalues are saved. Directional curvature positivity alone is not used to infer positive definiteness.

B.11 Matched frame-correction ablation

The main grid uses $d = 2, D = 3$, a clean anchor at zero, and either $g(x, y) = (1.5x^2 - y^2)/2$ or that graph plus $.3(x^3 + y^3)$. For each half-width $b \in \{.1, .3, .5\}$, draw 240 latent points: $|x|$ is uniform on $[0, b]$, with positive-sign probability .8, and y is uniform on $[-b, b]$. Independent scalar noise uniform on $[-\sigma, \sigma]$, $\sigma \in \{0, .01b\}$, is added along each point's true unit normal. Ten fixed seeds give 120 instances. No additional observed-radius selection is performed, so this is not a test of the radius-selection theorem. Geometry enters generation and scoring, not the observed settings. All main outputs return and satisfy the common trust/Gram orientation tests.

A fits a free-linear quadratic graph at common anchor PCA and returns the combined derivative. B repeats this fit once in A's tangent frame, using the original observations. C is the project's internal `fit_mmls` routine with uniform weights, moving base point, `shift_base=True`, `n_iter=4`,

Table 14: Matched-ablation mean projector errors, ten seeds per row. A: one Jet fit; B: two Jet fits; C: internal Iterative QF, four iterations; D: full observed-setting optimization. The noise column is σ/b. All 480 method outputs returned.

Graph	b	σ/b	A	B	C (4 it.)	D
Cubic	0.1	0	0.00343	0.00350	0.00350	0.00351
Cubic	0.1	0.01	0.00325	0.00331	0.00331	0.00332
Cubic	0.3	0	0.02857	0.03071	0.03070	0.03097
Cubic	0.3	0.01	0.02953	0.03156	0.03153	0.03179
Cubic	0.5	0	0.07690	0.08700	0.08638	0.08753
Cubic	0.5	0.01	0.07426	0.08447	0.08374	0.08524
Quadratic	0.1	0	0.00010	4.35×10^{-7}	3.47×10^{-8}	8.65×10^{-10}
Quadratic	0.1	0.01	0.00144	0.00144	0.00144	0.00144
Quadratic	0.3	0	0.00283	0.00010	9.35×10^{-6}	1.14×10^{-9}
Quadratic	0.3	0.01	0.00370	0.00193	0.00195	0.00192
Quadratic	0.5	0	0.01320	0.00133	0.00012	2.39×10^{-9}
Quadratic	0.5	0.01	0.01371	0.00202	0.00183	0.00178

and `tol=1e-12`. It is not published MMLS and all its outputs reach the four-iteration budget. D uses the observed safeguards and ambient L-BFGS, with at most 300 iterations. A/B/C are not clipped into D's feasible set. At each returned frame a separate constrained intercept-plus-quadratic, no-linear fit gives the common evaluation loss. This evaluation neither changes the returned tangent nor enters timing. Native free-linear curvature coefficients are not equated with this refitted curvature. Differences between A/B/C/D include their models and readouts; they cannot be attributed solely to the number of iterations.

Pairwise intervals resample the ten seed differences within each condition. All six cubic B-minus-A and D-minus-B intervals are positive. D improves noiseless quadratic mean error over B at all three widths; all three noisy-quadratic D-minus-B intervals include zero. These intervals are pointwise and small-sample, not equivalence or multiplicity-adjusted tests. The report's headline median is the median of twelve condition medians, not the pooled median of 120 errors, and is not used as a primary accuracy summary. Method-runtime medians aggregated in that manner are approximately .149, .297, .474, and 1.828 milliseconds for A/B/C/D, excluding separately recorded initialization/settings costs. We use this only as descriptive cost information, not a precise hardware-independent speedup. End-to-end timings in the archive add the initialization and settings needed by each method. D's main terminations are 102 gradient and 18 relative-loss stops.

The six mechanism runs use the twelve-point family at $\delta = 0, .1, .5$ and initial true-frame rotations $\pm.05$, with $h = 1, K = 2, \tau = \lambda_{\min}(\Sigma_{\text{truth}})/2$, and trust radius .2 centered at the supplied initial frame. This is an oracle-assisted local diagnostic, not a practical initialization benchmark. At $\delta = 0$, A/B return errors about .071 and C about .072; D lowers the error but stops with the nonzero values stated in Section 4.4. At $\delta = .1$, B's error is about .064 versus D's 2.5×10^{-11}. C remains a four-step diagnostic throughout; none of these comparisons establishes superiority to a converged C solution. Loss and error behavior are consistent with weak curvature, but six trajectories do not isolate all numerical or algorithmic causes.

B.12 Supplementary comparisons and structural diagnostics

The supplied-geometry comparisons in this subsection are retained as supplementary diagnostics; they are not the observed-setting primary results in Section 4. Unless an observed pilot rule is explicitly specified, external-baseline, ambient-dimension, and image comparisons below use $K = 2\|Q^*\|$, $\tau = 0$, and $r_0 = .4$. Their larger synthetic seed counts are distinct from the primary ten-seed grid. The synthetic radius generator supplies $h_{\text{out}} = r + 2\sigma$; image experiments use maximum observed displacement. These settings must not be conflated with the common observed scale used in the primary factorial study.

We evaluate whether the profiled formulation can be computed across the tested settings, whether its structural predictions are observed, and how its tangent accuracy compares with established nonlin-

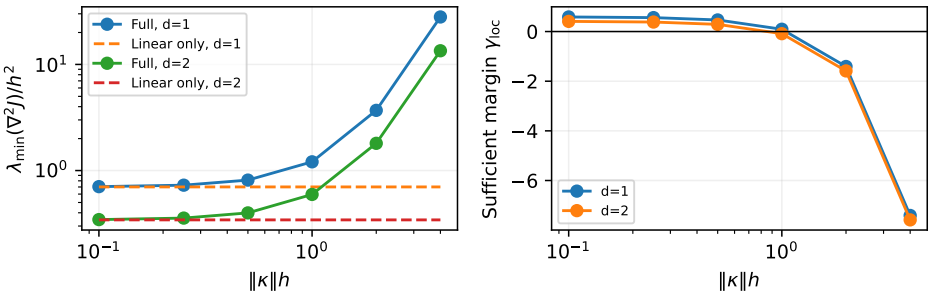


Figure 6: Centrally symmetric isotropic quadratic graphs, clean anchor, $m = 1$. The full Hessian remains positive as curvature increases, while γ_{loc} crosses zero. Linear-only curvature omits a positive contribution here. This is an exact-reference mechanism test, not a noisy accuracy certificate.

ear estimators. Quadratic reproduction and orientation ablations probe the model; external-baseline and image comparisons assess its practical scope.

B.12.1 Settings and comparison protocol

We measure $e = \|\widetilde{\Pi} - \Pi^*\|_F$. Synthetic data include quadratic and cubic curves/surfaces, higher-codimension polynomial graphs, and rotated Clifford tori. The external-baseline comparison uses 50 seeds per cell; the ambient-dimension study uses 30, and the intrinsic-dimension study uses 20. Image experiments use 20 base images per family and 10 neighborhood/noise draws per image. Appendix B gives the geometry, settings, and uncertainty calculations.

External and internal baselines. CGAL's degree-two jet fitter is used only for surfaces in $\mathbb{R}^3$ (Cazals & Pouget, 2025). The original authors' (Aizenbud and Sober) `manapprox` implementation of MMLS uses degree two and its native weighting and base-point refinement (Aizenbud & Sober, 2021; Sober & Levin, 2020). We read its tangent from the linear coefficient block of the polynomial map, not its intermediate affine frame. CGAL uses mean-centered PCA and receives the query as an extra first row; our fit excludes that row. The same selected neighbors are supplied to the methods. Their native query-associated outputs are scored against the same latent-anchor tangent; weights, fitting rows, and fitted reference points vary (Appendix B.7). In other dimensions Jet is our internal fixed-frame implementation, with $c + Bs + \mathcal{Q}_Q(s)$ and readout $\mathrm{im}(P_1 + P_2 B)$. The internal iterative fitter is called *Iterative QF*; it is not a reproduction of published MMLS. Oracle-frame Jet is diagnostic only.

Implementation and theory coverage. The external-baseline, ambient-dimension, and image studies use $K = 2\|Q^*\|$, $r_0 = 0.4$, and $\tau = 0$: they disable the Gram safeguard and do not validate the full safeguarded estimator. Curvature is supplied from synthetic geometry or estimated from the image model, not inferred blindly from the observed cloud. The structural ablations use an independently calibrated positive threshold. In the supplied synthetic radius code the scale is $r + 2\sigma$, while the image scale is the maximum observed displacement. This differs from the observed scale used for all primary factorial arms. The radius-selection noise axis uses the nominal ratio $q_\sigma = \sigma/(r + 2\sigma)$, not $\sigma/h_{\mathrm{solver}}$. The fixed-neighborhood diagnostics in Appendix B.5 do not test the observed-radius selection theorem. The large-dimensional studies use the ambient implementation. Returns and termination reasons are reported separately; most runs stop on relative-loss tolerance.

B.12.2 Orientation information and an observed-data implementation

We test Corollary C.51 on 24 configurations with $d \in \{1, 2\}$, $m \in \{1, 3\}$, 120 sign-paired samples, and six values of $\|\kappa\|h$. The moment formula and full residualized Jacobian agree to numerical precision; the maximum relative error against a finite-difference Hessian is 4.49×10^{-9}. Figure 6 illustrates why failure of a sufficient margin is not evidence of a flat objective.

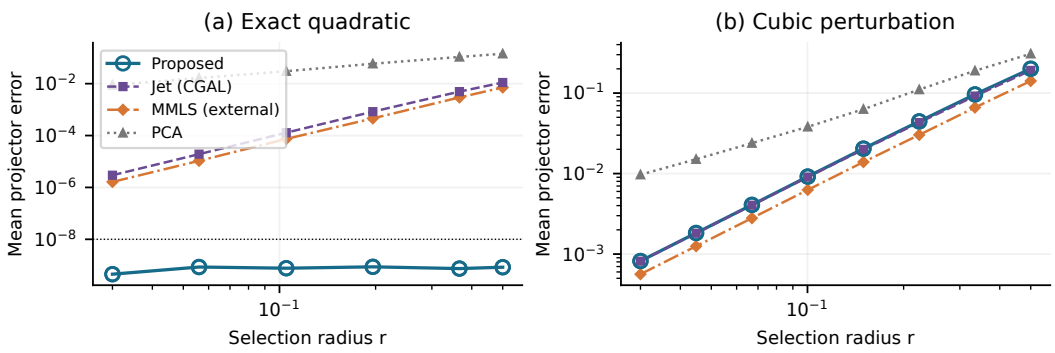


Figure 7: External-baseline comparison, 50 seeds per radius, asymmetric sampling, no noise, Gram safeguard disabled. (a) Mean projector error on a paraboloid; the dotted line is the 10^{-8} reproduction threshold. (b) Cubic-surface bias. All quadratic methods have approximately second-order bias. Panel (a) evaluates native query outputs against the latent-anchor tangent.

A separate 300-instance cubic-graph study selects the neighborhood radius as the distance to the 180th of 900 observed non-anchor points. It sets $h = \max_i \|u_i\|$, $K = 2\|\widehat{Q}_{\text{pilot}}\|_F$, $\tau = .05\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0))$, and $r_0 = .4$. These fixed rules use no latent geometry. All estimates return; mean error is 0.01949, versus 0.01929 with supplied curvature and $\tau = 0$ on the same neighborhoods (Jet: 0.01933; external MMLS: 0.01631; mean-centered PCA: 0.03219). The pilot underbounds true curvature in 12/300 instances, and truth is outside the trust region in five. The sufficient chart-localization inequality holds in 189/300. Thus the practical rule is computable but does not certify the theory's premises. The adaptive radius is not covered by the prescribed-radius sampling theorem. All cells and further diagnostics appear in Appendix B.8.

B.12.3 Curvature bias and quadratic reproduction

On paraboloid and saddle neighborhoods, Proposed attains $e < 10^{-8}$ in all 50 trials at each of six radii from 0.03 to 0.5; CGAL Jet and external MMLS attain this threshold in none. These are native query-output projector errors: the archives do not retain the fitted points needed to separate location effects from fitting errors (Appendix B.7). On a cubic surface, log–log slopes over the radius sweep are 1.963 for Proposed, 1.945 for CGAL Jet, 1.968 for external MMLS, and 1.243 for PCA (Figure 7). Thus second-order bias is shared with existing nonlinear methods. Across five higher-dimensional cubic designs, Proposed slopes range from 1.944 to 1.980; exact-quadratic threshold success ranges from 98.3% to 100%, including $(d, D) = (5, 100)$ (Table 3). These experiments do not individually certify all theoretical margin conditions. A separate 60-instance paraboloid study stores fitted points and coefficients. At latent-window scale 0.2, evaluating at a graph-associated fitted location changes external MMLS mean error from 0.0007046 to 0.0007038 and CGAL from 0.0014886 to 0.0014853. Location changes alone do not remove the gap on these samples. Oracle-frame Jet reproduces all 60 quadratics; Proposed does so in 58, with both failures having truth outside its trust region (Appendix B.8).

B.12.4 Orientation, initialization, and the linear term

The one-dimensional cubic ablation separates a fixed fitting frame from a jointly fitted frame. Reversing the cubic coefficient gives error ratios about 1.00–1.02 for Proposed and 3.05 for internal Jet initialized by anchor-PCA on the reported designs. The tilted-frame heuristic $c_{3,\text{eff}} \approx c_3 - \kappa^2\varphi/2$ explains cancellation in this example (Appendix B.4). The sign sensitivity is specific to these designs and is less prominent in the higher-codimension tests.

Allowing a free linear term weakens the measured curvature of the profiled frame objective and can produce differing frames across starts. However, $\mathrm{im}(P_1)$ is then not the tangent of the fitted graph; the combined readout $\mathrm{im}(P_1 + P_2 B)$ can improve accuracy, especially with noise. This ablation distinguishes frame orientation from the fitted derivative. Separately, 6,300 noiseless high-

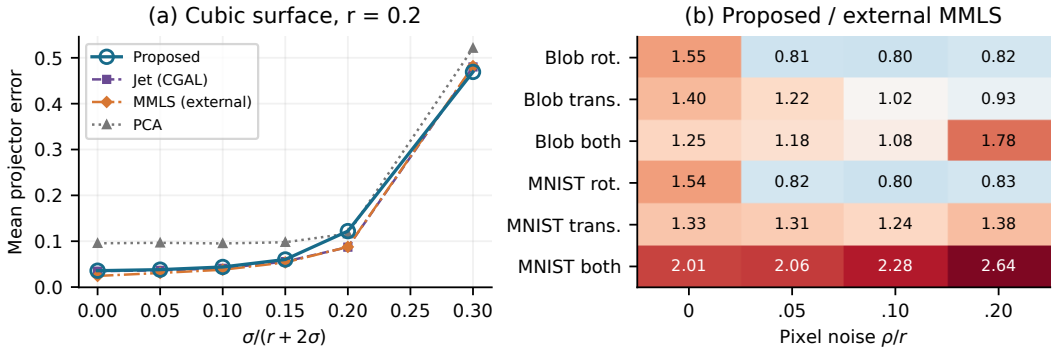


Figure 8: Accuracy depends on the regime. (a) External comparison on an asymmetric cubic surface with tubular noise, 50 seeds per level. (b) Mean-error ratio Proposed/external MMLS for all 24 main asymmetric image cells, radius selection, approximately 300 neighbors; below one favors Proposed. Image pixel noise is outside the bounded-normal-noise model. No panel isolates weighting, base-point refinement, or trust-region effects.

Table 15: Held-out mean projector error after per-cell radius tuning. Surface: CGAL Jet; curve: internal Jet (CGAL unsupported). Each row averages two sampling designs and 50 evaluation seeds per design. External MMLS uses SVD initialization. All 5,200 selected method–seed evaluations returned estimates.

Geometry	σ	Proposed	Jet	MMLS (ext.)	PCA
Curve	0.000	0.0023	0.0022	0.0016	0.0180
Curve	0.005	0.0099	0.0096	0.0092	0.0246
Curve	0.010	0.0157	0.0138	0.0147	0.0284
Curve	0.020	0.0247	0.0213	0.0208	0.0402
Surface	0.000	0.0032	0.0032	0.0025	0.0254
Surface	0.005	0.0188	0.0186	0.0173	0.0313
Surface	0.010	0.0247	0.0245	0.0231	0.0373
Surface	0.020	0.0391	0.0390	0.0353	0.0544

dimensional instances were run from PCA and five perturbed starts. The maximum pairwise output projector distance was 2.05×10^{-6}. This measures local output stability across the tested starts.

B.12.5 External MMLS, noise, and radius selection

At radius 0.2, external MMLS has lower mean error than Proposed at five of six noise levels; at the largest nominal noise ratio $q_\sigma = 0.3$, the means are 0.4819 and 0.4693, respectively (Figure 8a). The near-agreement between Proposed and unweighted Iterative QF in noiseless data is not an identity with published MMLS. Likewise, the EIV heuristic concerns that internal routine only (Appendix B.4).

For radius tuning we keep the latent sampling window and absolute noise fixed, choose a radius separately for each geometry, sampling design, noise, and method on 50 calibration seeds, and evaluate on 50 disjoint seeds. Only radii returning estimates for all 50 calibration seeds are eligible. Shared insufficient-data events are counted for every applicable method. Table 15 averages the symmetric and asymmetric evaluation cells. External MMLS has lower mean error than Proposed in 15 of 16 individual cells; eight paired 95% bootstrap intervals exclude zero in that direction. These are pointwise intervals conditional on selected radii, not simultaneous or selection-adjusted guarantees. The result rules out a general claim that radius tuning makes the external methods indistinguishable.

B.12.6 HIGHER DIMENSIONS AND IMAGE MANIFOLDS

The ambient-dimension study reaches $(d, D) = (5, 100)$; rankings vary across quadratic, cubic, and torus families. A separate study fixes $D = 100$ and tests $d = 5, 10, 15$, using either $N = 200$ or $N = 10q$, $q = 1 + d(d+1)/2 + d$. In the asymmetric cubic/noisy cell, Proposed mean error rises from 0.0223 to 0.0841 at fixed N, versus 0.0220 to 0.0197 with $N = 10q$. At $d = 15, N = 200$, mean-centered PCA is better (0.0634), whereas external MMLS is worse (0.0949). At $N = 1360$, Proposed and internal Jet are close (0.019735 and 0.019743). Table 7 reports the full dimension comparison. The clean quadratic 10^{-8} threshold is met in 0/20 trials at $d = 15, N = 200$, but mean error is only 4.86×10^{-8}; at $N = 1360$, 18/20 meet it. These results show the importance of sample count relative to feature count in the tested designs.

We also use transformations of Gaussian-blob images in $\mathbb{R}^{1024}$ and smoothed MNIST images in $\mathbb{R}^{784}$: rotation, translation, and both give $d = 1, 2, 3$. The reference tangent is the span of a numerically computed transformation Jacobian, validated by step halving. Pixel Gaussian noise is ambient and unbounded; these are out-of-model transfer tests, not validation of the tubular-noise theorem.

Across the 24 main asymmetric image cells, external MMLS has lower mean error in 17 and Proposed in seven. Anchor-cluster paired intervals exclude zero in 16 and seven cells, respectively. Proposed improves over mean-centered PCA in 21 cells, but on noisy MNIST translation-plus-rotation its error is 0.4815 versus 0.3293 for PCA and 0.1827 for external MMLS. The image results therefore demonstrate feasibility at large ambient dimension and a substantial failure regime, rather than general superiority. Full cells and uncertainty estimates appear in Table 4.

B.12.7 PROFILE DERIVATIVES, SAFEGUARDS, AND COST

Exact quadratic checks cover codimension one and two, two curvature scales, two sampling designs, and zero/nonzero normal anchor offsets. The analytic frame Jacobian and Hessian match finite differences with maximum relative errors 1.10×10^{-9} and 9.35×10^{-10} at step 10^{-5}. Omitting predictor rotation changes the Jacobian by 1.49–27.98% relative to the full expression. These checks use the unconstrained inner profile at the quadratic reference.

In fixed-neighborhood stress tests, relative thresholds $\tau = \alpha\lambda_{\min}(\widehat{\Sigma}(P_{\mathrm{PCA}}))$, $\alpha \le 0.9$, leave all paired errors unchanged. An independently calibrated absolute threshold rejects all 24 clean near-degenerate cases, yet returns all 24 noisy near-degenerate cases with median error 1.4035. Their true tangents are outside the trust region. Thus coefficient feasibility is distinct from accurate orientation and localization. With $K_{\mathrm{pilot}} = 2\|\widehat{Q}_{\mathrm{pilot}}\|_F$ and $\tau = 0.05\lambda_{\min}(\widehat{\Sigma}(P_{\mathrm{PCA}}))$, 96 pairs give mean error difference 1.39×10^{-4}; the pilot underbounds true curvature in 3/96. The 72 pairs with truth inside the trust region have median difference zero and inactive curvature constraints. Small-dimensional timings including PCA initialization give Proposed 2.05–2.24 ms, versus 0.131–0.162 ms for internal Jet at $N = 80, 160, 320$. Protocols and qualified diagnostics are in Appendix B.5.

C THEORETICAL ANALYSIS AND PROOFS

C.1 ORGANIZATION AND ASSUMPTIONS

The proofs proceed from graph geometry and regression conditioning to orientation separation and recovery. Sampling and PCA results verify local premises under noisy observed-radius selection; the final sections derive profile derivatives, local curvature, and conditional numerical bounds. Deterministic statements require a realized sample with the stated geometric, design, and feasibility conditions. Probability statements add sampling, chart coverage, density, count, and initialization assumptions. The global Aamari–Levrard model is used only where explicitly invoked.

Sources and proof dependencies. The manifold class, parameterization lemma, and minimax benchmark are adopted from Aamari & Levrard (2019) at the explicitly cited statements. Coefficient elimination follows the variable-projection principle of Golub & Pereyra (1973); the constrained envelope argument is a unique-optimizer application of Danskin (1967). Concentration and spectral perturbation use Tropp (2012), McDiarmid (1989), and Davis & Kahan (1970). For background on manifold derivatives and first-order optimization, see Boumal (2023, Chapters 3–4) and Bubeck (2015, Chapter 3). We state the required specializations and verify their hypotheses below; standard general proofs are left to these sources. The selected-sample law, anchor and predictor perturbations, residualized orientation separation, and full frame derivative are derived for the present estimator.

C.2 NOISY-ANCHOR LOCAL PROBLEM AND GEOMETRIC SETUP

Let $M^d \subset \mathbb{R}^D$ be a closed embedded C^2 manifold without boundary, of intrinsic dimension $1 \le d < D$ and codimension $m = D - d$. For a vector-valued multilinear map, $\|\cdot\|_{\mathrm{op}}$ denotes the supremum of its output Euclidean norm over unit input vectors. Write

$$\tau_M = \mathrm{reach}(M) = \sup\left\{ r > 0 : \begin{array}{l} \text{every } x \text{ with } \mathrm{dist}(x, M) < r \\ \text{has a unique nearest point in } M \end{array} \right\},$$

and assume $\tau_M > 0$. Here closed means closed as a subset of $\mathbb{R}^D$, not necessarily compact. Let $x_i^* \in M$ be latent clean points and let

$$x_i = x_i^* + \xi_i$$

be the observations. We begin with the tubular-noise conditions

$$\xi_i \in N_{x_i^*}M, \qquad \mathbb{E}[\xi_i \mid x_i^*] = 0, \qquad \|\xi_i\| \le \sigma < \tau_M \tag{12}$$

almost surely, for every observation, including $i = 0$. The normal-tube projection property gives $x_i^* = \pi_M(x_i)$. Strict inequality is essential: uniqueness can fail at distance τ_M. This identifies the geometric target relative to the unknown manifold; it does not by itself establish statistical recovery. Later we additionally control the noise relative to the neighborhood scale when proving recovery. Section C.6 shows why such an additional condition is necessary for the present least-squares criterion.

For the deterministic local theorem, σ is only a uniform perturbation radius and need not be random. For the observed-radius statistical specialization we add the i.i.d. sampling and local density/chart assumptions below. The global model in Assumption C.6 is optional and is used only when explicitly invoked for calibration or benchmark comparisons.

The observed anchor and its latent clean counterpart are

$$x_0 = x_0^* + \xi_0, \qquad x_0^* \in M.$$

Our target is $T_{x_0^*}M$, not a tangent space at the generally off-manifold point x_0. Equivalently, the target is $T_{\pi_M(x_0)}M$.

Default neighborhood and sampling convention. For the statistical implementation, the original pairs $(X_i^*, \xi_i)_{i=0}^{n-1}$ are i.i.d. under the observation model above. The anchor index 0 and radius $r > 0$ are prescribed before examining the sample. Set

$$\mathcal{I}_r(X_0) = \{i = 1, \dots, n-1 : \|X_i - X_0\| \le r\}, \qquad N_r = |\mathcal{I}_r(X_0)|\,.$$

List the selected points in increasing original-index order and relabel them $x_1, \dots, x_N$, where $N = N_r$, for local regression. The global sample size is n; the selected sample size is N. Recovery requires enough selected points and the stated design conditions. Let $\mathcal{F}_0 = \sigma(X_0^*, \xi_0)$ be the anchor sigma-field.

Lemma C.1 (Conditional law of a fixed-radius sample)**.** *Conditional on $\mathcal{F}_0$ and $N_r = q \geq 1$, the selected pairs in original-index order are independent with common law*

$$\mathsf{P}_{r,0} = \mathcal{L}\big((X^*, \xi) \mid \mathcal{F}_0,\ \|X^* + \xi - X_0\| \leq r\big).$$

All conditional laws are regular conditional laws: the assertion holds for almost every anchor realization with positive selection probability, and for counts having positive conditional probability.

Proof. Fix the anchor variables. Each selection event depends only on its own non-anchor pair. Conditional on any particular selected index set I, independence gives the selected pairs product law $\mathsf{P}_{r,0}^{\otimes |I|}$. For every set of size q, this is the same ordered product law. Mixing over those sets conditional on $N_r = q$ leaves it unchanged. Subtracting the fixed X_0 preserves the conditional independence. □

Local concentration statements for this procedure condition on $\mathcal{F}_0$ and the selected count. Selection can change the noise mean: $\mathbb{E}[\xi \mid X^*, \text{selected}, \mathcal{F}_0]$ need not vanish. The design-transfer and PCA perturbation arguments use bounded noise and do not require mean-zero noise after selection. After this conditioning, lowercase letters denote a realized selected local sample; the deterministic statements below are samplewise statements about that realization.

Definition C.2 (Fixed noisy-anchor local instance)**.** Fix $N \geq 1$ and a latent chart scale $h > 0$, with the closed coordinate ball contained in the chart domain. Relative to a tangent-aligned C^2 chart based at x_0^*, a local instance consists of an observed anchor $x_0 = x_0^* + \xi_0$ and non-anchor observations $x_i = x_i^* + \xi_i$, $i = 1, \dots, N$, with $\xi_i \in N_{x_i^*}M$, $\|\xi_i\| \leq \sigma < \tau_M$, and $x_i^* = F(t_i)$ for coordinates satisfying $\|t_i\| \leq h$. Thus h bounds latent tangent coordinates; it is not, in general, an observed-radius bound on $\|x_i - x_0\|$. The estimator observes only the displacements $u_i = x_i - x_0$; the variables x_i^*, ξ_i, t_i, and x_0^* are used only for analysis. This definition is a deterministic description of a realized local sample. For the default procedure, the radius sandwich below gives the conservative latent/design scale $h = h_{\text{out}} = r + 2\sigma$, provided the selected latent points lie in the stated chart. This scale is induced by the observed radius; it is not the observed radius itself. All deterministic results concern such fixed local instances. The conditional mean-zero part of equation 12 belongs to the statistical model and is not required for the deterministic comparison or the bounded-perturbation design and PCA arguments.

Remark. The displacements share the noisy anchor and need not be independent unconditionally. Lemma C.1 supplies the exact conditional law for the default procedure. Deterministic statements require no independence.

Choose a true tangent–normal frame

$$P^* = [P_1^* \mid P_2^*] \in O(D), \qquad \text{im}(P_1^*) = T_{x_0^*}M, \qquad \text{im}(P_2^*) = N_{x_0^*}M,$$

and define the true tangent projector

$$\Pi^* = P_1^*(P_1^*)^\top.$$

The frame P^* is not unique, but the projector Π^* is unique.

For a candidate $P = [P_1 \mid P_2] \in O(D)$, write

$$\Pi(P) = P_1 P_1^\top.$$

The primary error is $\|\Pi(P) - \Pi^*\|_F$. If $\theta_1, \dots, \theta_d$ are the principal angles, then

$$\|\Pi(P) - \Pi^*\|_F^2 = 2\sum_{j=1}^{d} \sin^2\theta_j. \tag{13}$$

C.2.1 REGULARITY HIERARCHY

We separate the smoothness needed to define the method from the smoothness needed to obtain a polynomial recovery rate. The local graph, the identity $Dg(0) = 0$, and the curvature matrices $Q_k^* = D^2 g_k(0)$ require only that M be C^2. For a quantitative rate, define the local Hessian modulus

$$\omega_2(r) = \sup_{\substack{t \in W \\ \|t\| \le r}} \left\| D^2 g(t) - D^2 g(0) \right\|_{\mathrm{op}}.$$

For a fixed C^2 graph, $\omega_2(r) \to 0$ as $r \downarrow 0$, but the definition of a bare C^2 class supplies no uniform power law for this convergence. When, for some $0 < \alpha \le 1$,

$$\left\| D^2 g(t) - D^2 g(0) \right\|_{\mathrm{op}} \le L_{2,\alpha} \|t\|^{\alpha}, \tag{14}$$

we say that the Hessian is α-Hölder at the anchor. This pointwise condition is sufficient for all remainder and risk bounds below. The standard class $C^{2,\alpha}$ requires the stronger pairwise bound $\|D^2 g(t) - D^2 g(s)\|_{\mathrm{op}} \le L_{2,\alpha} \|t - s\|^{\alpha}$ throughout the chart, and therefore implies equation 14. In particular, $C^{2,1}$ implies its $\alpha = 1$ case. A quantitative C^3 assumption with a bounded third derivative implies equation 14 with $\alpha = 1$, but is stronger than what the deterministic h^2 bias calculation itself requires.

C.2.2 ADOPTED AAMARI–LEVRARD MODEL

We adopt the classes of (Aamari & Levrard, 2019, Definitions 1–2), using d, D in place of their intrinsic and ambient dimensions. The following geometric result motivates the construction.

Lemma C.3 (Aamari–Levrard, Lemma 1)**.** *For a compact connected C^2 embedded manifold without boundary with $\tau_M \ge \tau_{\min} > 0$, the exponential map at p is injective on $B_{T_pM}(0, \tau_{\min}/4)$ and has the representation*

$$\exp_p(v) = p + v + N_p(v), \qquad N_p(0) = 0, \quad d_0 N_p = 0, \quad \|d_v N_p\|_{\mathrm{op}} \le \frac{5}{4\tau_{\min}} \|v\|.$$

Also, for $p, y \in M$,

$$\left\| (I - \Pi_{T_pM})(y - p) \right\| \le \frac{\|y - p\|^2}{2\tau_{\min}}.$$

Source. This is (Aamari & Levrard, 2019, Lemma 1), restated in our notation. □

Definition C.4 (Aamari–Levrard manifold class)**.** For $k \ge 3$, let $\mathcal{C}^k_{\tau_{\min},\mathbf{L}}$, where $\mathbf{L} = (L_\perp, L_3^{\mathrm{AL}}, \dots, L_k^{\mathrm{AL}})$, contain the compact connected embedded C^k manifolds without boundary with reach at least $\tau_{\min}$ admitting, at every p, an injective parameterization

$$\Psi_p(v) = p + v + N_p(v), \quad N_p : B_{T_pM}(0, r_p) \to \mathbb{R}^D, \quad r_p \ge r_* = (4L_\perp)^{-1},$$

where N_p is C^k, $N_p(0) = d_0 N_p = 0$, and on $\|v\| \le r_*$,

$$\left\| d_v^2 N_p \right\|_{\mathrm{op}} \le L_\perp, \qquad \left\| d_v^j N_p \right\|_{\mathrm{op}} \le L_j^{\mathrm{AL}} \quad (3 \le j \le k).$$

Definition C.5 (Aamari–Levrard distribution classes)**.** Let $\mathcal{P}^k$ denote the class with parameters $\tau_{\min}, \mathbf{L}, f_{\min}, f_{\max}$, consisting of laws supported on $M \in \mathcal{C}^k_{\tau_{\min},\mathbf{L}}$ with volume density $f_{\min} \le f_M \le f_{\max}$, where $0 < f_{\min} \le f_{\max} < \infty$. For $0 \le \sigma < \tau_{\min}$, define

$$\mathcal{P}^k(\sigma) = \{\mathrm{Law}(X^* + \xi) : \mathrm{Law}(X^*) \in \mathcal{P}^k,\ \xi \in N_{X^*}M,\ \|\xi\| \le \sigma,\ \mathbb{E}[\xi \mid X^*] = 0\}.$$

Assumption C.6 (Optional benchmark sampling model)**.** *When the global statistical benchmark is invoked, the pairs $(X_i^*, \xi_i)_{i=0}^{n-1}$ are i.i.d. from a construction in Definition C.5 with $k = 3$. In particular, $P_X \in \mathcal{P}^3(\sigma)$ and $\sigma < \tau_{\min} \le \tau_M$, including for the observed anchor.*

This optional global C^3 model is used only for the benchmark comparison and for one conservative geometric calibration route. The deterministic local recovery theorems below are stated directly in terms of their local graph assumptions.

Distinction from the regression chart. Neither the adopted definition nor Lemma C.3 requires $N_p(v) \in N_pM$. Thus Ψ_p is not automatically a graph in tangent-projection coordinates. Moreover, Lemma C.3 alone supplies no third-derivative bound. The graph below is obtained by inverting the tangent projection, not by identifying it with $\exp_p$. We use L_3^{AL} for the adopted parameterization bound and L_3 for the graph remainder constant; these constants must not be equated. The deterministic theorems state their graph bounds explicitly.

Positive reach and $\sigma < \tau_M$ already fix the projection target in the basic model. Compactness, connectedness, uniform parameterization bounds, and global density bounds are additional benchmark conditions. Adopting this model does not transfer an estimator-specific upper bound to our criterion. The minimax comparison uses exactly Definition C.5, not a smaller graph-constrained class.

C.2.3 Tangent-aligned local graph

Lemma C.7 (From the adopted parameterization to a normal graph)**.** *Suppose $M \in \mathcal{C}^3_{\tau_{\min},\mathbf{L}}$. At each $p \in M$, the image of its adopted parameterization contains a normal graph $p+t+G_p(t)$, with $G_p(t) \in N_pM$, on $\|t\| < r_*/2$, where $r_* = (4L_\perp)^{-1}$. It satisfies $G_p(0) = d_0G_p = 0$ and*

$$\left\|D^2G_p\right\|_{\mathrm{op}} \le 3L_\perp, \qquad \left\|D^3G_p\right\|_{\mathrm{op}} \le 4L_3^{\mathrm{AL}} + 16L_\perp^2.$$

Consequently, in orthonormal coordinates the graph remainder constant may be chosen as $L_3 = 4L_3^{\mathrm{AL}} + 16L_\perp^2$. This asserts existence of a uniform graph patch, not that it contains every point of an arbitrarily prescribed ambient neighborhood.

Proof. Write $\Pi = \Pi_{T_pM}$, $N = N_p$, and $H(v) = v + \Pi N(v)$. On the ball of radius r_*, the derivative bound and $DN(0) = 0$ give $\|DN(v)\|_{\mathrm{op}} \le L_\perp r_* = 1/4$. For $\|t\| < r_*/2$, the map $v \mapsto t - \Pi N(v)$ is a contraction of the closed radius-$3r_*/4$ ball into itself, since $\|N(v)\| \le L_\perp\|v\|^2/2 \le 9r_*/128$ there and $r_*/2 + 9r_*/128 < 3r_*/4$. This closed ball is strictly inside the parameterization domain, even when its stated radius equals r_*. Its unique fixed point $V(t)$ satisfies $H(V(t)) = t$ and lies in the interior. The inverse function theorem gives a C^3 inverse there. Put $a = 4/3$. Differentiating $H(V(t)) = t$ yields

$$\|DV\|_{\mathrm{op}} \le a, \quad \left\|D^2V\right\|_{\mathrm{op}} \le a^3L_\perp, \quad \left\|D^3V\right\|_{\mathrm{op}} \le a^4L_3^{\mathrm{AL}} + 3a^5L_\perp^2.$$

Define $G_p(t) = (I - \Pi)N(V(t))$. Then $\Psi_p(V(t)) = p + t + G_p(t)$. The chain rule, using $\|DN\| \le 1/4$, gives

$$\left\|D^2G_p\right\|_{\mathrm{op}} \le (a^2 + a^3/4)L_\perp < 3L_\perp,$$

and

$$\left\|D^3G_p\right\|_{\mathrm{op}} \le (a^3 + a^4/4)L_3^{\mathrm{AL}} + (3a^4 + 3a^5/4)L_\perp^2 \le 4L_3^{\mathrm{AL}} + 16L_\perp^2.$$

The identities at zero follow from $N(0) = DN(0) = 0$. Orthogonal changes of coordinates preserve these multilinear operator bounds. Taylor's integral formula then gives the cubic remainder bound. □

Proposition C.8 (Tangent-aligned local quadratic representation)**.** *Suppose that M is C^2 without boundary and $x_0^* \in M$. There are neighborhoods $0 \in W \subset \mathbb{R}^d$, with W chosen to be a ball, $x_0^* \in U \subset M$, and a C^2 map $g : W \to \mathbb{R}^m$ such that*

$$F(t) = x_0^* + P_1^*t + P_2^*g(t)$$

parametrizes U, and

$$g(0) = 0, \qquad Dg(0) = 0.$$

Let

$$Q_k^* = D^2g_k(0) \in \mathrm{Sym}(d), \qquad k = 1, \dots, m.$$

Then

$$g(t) = \mathcal{Q}_{Q^*}(t) + R(t), \qquad \mathcal{Q}_{Q^*}(t) = \frac{1}{2}\begin{pmatrix} t^\top Q_1^*t \\ \vdots \\ t^\top Q_m^*t \end{pmatrix}, \tag{15}$$

where $\|R(t)\| = o(\|t\|^2)$. *If* $D^2 g$ *is* L_3*-Lipschitz, then*

$$\|R(t)\| \le \frac{L_3}{6} \|t\|^3 . \tag{16}$$

Moreover, for $a, b \in \mathbb{R}^d$,

$$\mathrm{II}_{x_0^*}(P_1^* a, P_1^* b) = P_2^* \begin{pmatrix} a^\top Q_1^* b \\ \vdots \\ a^\top Q_m^* b \end{pmatrix}.$$

Proof. The tangent projection $\varphi(x) = (P_1^*)^\top (x - x_0^*)$ has derivative $D\varphi(x_0^*)P_1^* = I_d$. The inverse function theorem therefore gives $F(t) = x_0^* + P_1^* t + P_2^* g(t)$, with $g(0) = Dg(0) = 0$. Taylor's integral formula gives

$$R(t) = \int_0^1 (1-s)\{D^2 g(st) - D^2 g(0)\}[t,t]\, ds.$$

Continuity yields $R(t) = o(\|t\|^2)$; Lipschitz continuity gives $\|R(t)\| \le L_3 \|t\|^3 \int_0^1 s(1-s) ds = L_3 \|t\|^3 / 6$. Finally $D^2 F(0) = P_2^* D^2 g(0)$ is normal, yielding the stated second fundamental form. □

Remark (Latent geometry and observed regression). Proposition C.8 describes the latent manifold, even when all observations are noisy. It does not assert that observed points lie on that graph. Proposition C.10 transfers the geometry to observed coordinates, and Proposition C.13 bounds the resulting fitting error. Probabilistic results for the default procedure use the conditional selected law of Lemma C.1.

Proposition C.9 (Regularity-dependent quadratic remainder)**.** *Suppose the line segment* $\{st : 0 \le s \le 1\}$ *is contained in* W. *Under the* C^2 *assumption of Proposition C.8,*

$$\|R(t)\| \le \frac{1}{2}\omega_2(\|t\|) \|t\|^2 = o(\|t\|^2). \tag{17}$$

If the Hölder-Hessian condition equation 14 holds, then

$$\|R(t)\| \le \frac{L_{2,\alpha}}{(1+\alpha)(2+\alpha)} \|t\|^{2+\alpha} . \tag{18}$$

For $\alpha = 1$, *this becomes* $\|R(t)\| \le L_{2,1} \|t\|^3 / 6$, *which is exactly equation 16 with* $L_3 = L_{2,1}$.

Proof. Use the integral remainder in Proposition C.8. Bounding its Hessian difference by $\omega_2(\|t\|)$ gives equation 17. Under equation 14, use $L_{2,\alpha} s^\alpha \|t\|^\alpha$ and $\int_0^1 (1-s) s^\alpha ds = ((1+\alpha)(2+\alpha))^{-1}$. □

C.2.4 Observed coordinates

For a non-anchor observation, define

$$u_i = x_i - x_0, \qquad i = 1, \dots, N.$$

We deliberately do not use $u_0 = x_0 - x_0 = 0$ as a regression row. The reason is that x_0 is noisy and generally does not lie on the latent graph through x_0^*. A remark after the estimator definition explains what happens if the self-anchor row is retained.

For $P = [P_1 \mid P_2]$, define

$$s_i(P) = P_1^\top u_i \in \mathbb{R}^d, \qquad n_i(P) = P_2^\top u_i \in \mathbb{R}^m.$$

Proposition C.10 (Observed-coordinate decomposition)**.** *Suppose* $x_i^* = F(t_i)$ *lies in the chart of Proposition C.8. Write*

$$\xi_0 = P_2^* a_0 \iff a_0 = (P_2^*)^\top \xi_0.$$

Then

$$u_i = P_1^* t_i + P_2^* g(t_i) + (\xi_i - \xi_0),$$

and, at the true frame,

$$s_i(P^*) = t_i + \eta_i, \qquad \eta_i = (P_1^*)^\top \xi_i, \tag{19}$$

$$n_i(P^*) = g(t_i) + \zeta_i - a_0, \qquad \zeta_i = (P_2^*)^\top \xi_i. \tag{20}$$

Proof. Use $x_i = x_i^* + \xi_i$, $x_0 = x_0^* + \xi_0$, and $x_i^* - x_0^* = P_1^* t_i + P_2^* g(t_i)$. This gives the formula for u_i. Multiplication by $(P_1^*)^\top$ and $(P_2^*)^\top$ gives equation 19–equation 20; here $(P_1^*)^\top \xi_0 = 0$ because $\xi_0 \in N_{x_0^*} M$. □

Remark. Although $\xi_i \in N_{x_i^*} M$, it need not be normal to $T_{x_0^*} M$. Therefore η_i need not vanish. The noisy regression is an errors-in-variables problem: both the predictor and the response contain noise.

Remark (Why normal noise matters at the anchor). Normality is a structural assumption, not general ambient additive noise. It gives $(P_1^*)^\top \xi_0 = 0$, allowing the common anchor displacement to be removed by an intercept at the target tangent. With an anchor tangent component $\eta_0 \neq 0$, the predictors would contain $-\eta_0$, and expanding a quadratic in the shifted predictor would generate a linear term. The present no-linear-term model and guarantees do not cover that setting unchanged. Replacing a latent point by $\pi_M(x_i)$ does not automatically make the resulting normal noise conditionally mean zero.

C.3 The estimator and its structural justification

For $Q = (Q_1, \ldots, Q_m)$, $Q_k \in \mathrm{Sym}(d)$, define

$$\mathcal{Q}_Q(s) = \frac{1}{2} \begin{pmatrix} s^\top Q_1 s \\ \vdots \\ s^\top Q_m s \end{pmatrix}.$$

The empirical loss is

$$L_N(P, c, Q) = \frac{1}{N} \sum_{i=1}^{N} \left\| n_i(P) - c - \mathcal{Q}_Q(s_i(P)) \right\|^2. \tag{21}$$

Proposition C.11 (Correspondence with the order-two local polynomial criterion)**.** *Let* $\Pi = P_1 P_1^\top$*, and define the ambient quadratic tensor by*

$$T_Q\big((\Pi u)^{\otimes 2}\big) = P_2 \mathcal{Q}_Q(P_1^\top u).$$

Then

$$\begin{aligned} &\left\| u - \Pi u - P_2 c - T_Q((\Pi u)^{\otimes 2}) \right\|^2 \\ &\quad = \left\| P_2^\top u - c - \mathcal{Q}_Q(P_1^\top u) \right\|^2. \end{aligned}$$

In particular, when $c = 0$*, the present criterion is the normal-output parameterization of the* $k = 3$ *local polynomial criterion of Aamari and Levrard Aamari & Levrard (2019).*

Proof. Since $I_D - \Pi = P_2 P_2^\top$, the ambient residual equals

$$P_2\{P_2^\top u - c - \mathcal{Q}_Q(P_1^\top u)\}.$$

The columns of P_2 are orthonormal, so multiplication by P_2 preserves the Euclidean norm. This proves the identity. □

The proposition is an algebraic correspondence, not an assertion that the two estimators are identical under every constraint. The present admissible design set, curvature norm, intercept, and local trust region must still be handled in the estimator-specific upper-bound proof.

C.3.1 Tangent readout and affine restriction

With a free normal slope B, the fitted tangent is $\mathrm{im}(P_1 + P_2 B)$, rather than $\mathrm{im}(P_1)$. Setting $B = 0$ retains the ambient linear term $P_1 s$ and makes the frame the tangent readout. At the oracle coefficients $P = P^*, c = -a_0$, the fitted base point is x_0^*; no separate recovery theorem for the estimated base point is claimed. This parameterization alone does not establish identifiability. Both predictor and response coordinates are noisy, so the model does not assume a correctly specified conditional regression mean.

Proposition C.12 (Affine restriction and local PCA)**.** *For any observed local sample with $N \geq 1$, including a noisy anchor, let $\bar{u} = N^{-1}\sum_i u_i$ and $S_c = N^{-1}\sum_i (u_i - \bar{u})(u_i - \bar{u})^\top$. If $Q = 0$, then*

$$\min_c L_N(P, c, 0) = \mathrm{tr}\{(I - \Pi(P))S_c\}.$$

Consequently, minimizing this affine restriction over all rank-d projectors yields a top-d spectral projector of S_c.

Proof. The optimal intercept is $c = P_2^\top \bar{u}$. Substitution yields

$$N^{-1}\sum_i \|P_2^\top (u_i - \bar{u})\|^2,$$

which equals the trace above. The spectral variational principle maximizes $\mathrm{tr}(\Pi S_c)$ at a top-d spectral projector. □

Thus the unrestricted affine model recovers mean-centered PCA, distinct from the anchor-centered initialization. Keeping the orientation constraints instead gives constrained affine fitting. Adding quadratic coefficients can lower fitting loss without lowering tangent error.

C.3.2 The oracle parameter and its residual

Define the aggregate curvature size

$$K_2 = \left(\sum_{k=1}^m \|Q_k^*\|_{\mathrm{op}}^2\right)^{1/2}.$$

Proposition C.13 (Oracle approximation at the true frame)**.** *Assume $\|t_i\| \leq h$ and the cubic remainder bound equation 16. At*

$$P = P^*, \qquad c^* = -a_0, \qquad Q = Q^*,$$

the residual is

$$r_i^* = R(t_i) + \zeta_i - \big(\mathcal{Q}_{Q^*}(t_i + \eta_i) - \mathcal{Q}_{Q^*}(t_i)\big). \tag{22}$$

In particular,

$$\|r_i^*\| \leq \frac{L_3}{6}h^3 + \sigma + K_2\left(h\sigma + \frac{\sigma^2}{2}\right). \tag{23}$$

Proof. At the true frame, Proposition C.10 and equation 15 give

$$n_i(P^*) = \mathcal{Q}_{Q^*}(t_i) + R(t_i) + \zeta_i - a_0, \qquad s_i(P^*) = t_i + \eta_i.$$

Subtracting $c^* + \mathcal{Q}_{Q^*}(s_i(P^*))$ proves equation 22. Notice that the anchor displacement $-a_0$ cancels exactly.

For each k,

$$\frac{1}{2}(t+\eta)^\top Q_k^*(t+\eta) - \frac{1}{2}t^\top Q_k^* t = t^\top Q_k^*\eta + \frac{1}{2}\eta^\top Q_k^*\eta.$$

Cauchy–Schwarz, $\|\eta_i\| \leq \sigma$, and $\|\zeta_i\| \leq \sigma$ therefore give

$$\|\mathcal{Q}_{Q^*}(t_i + \eta_i) - \mathcal{Q}_{Q^*}(t_i)\| \leq K_2\left(h\sigma + \frac{\sigma^2}{2}\right).$$

Combining this with equation 16 proves the claim. □

Remark (Why the self-anchor row is omitted). If the artificial row $u_0 = 0$ is included, its residual is $-c$. At the oracle $c^* = -a_0$, this row contributes $\|a_0\|^2/(N+1)$ to the loss and shrinks the intercept toward zero, even though the latent clean anchor is located at $x_0 - \xi_0$. One may retain this row, but the resulting extra term must be carried through the proof. Omitting it gives the cleanest noisy formulation. In the clean case $a_0 = 0$, the added row has zero oracle residual. It can still change the finite-sample fitted intercept and hence the estimator; only the oracle penalty vanishes.

C.3.3 Block invariance and the Grassmann parameter

Theorem C.14 (Block rotational invariance)**.** *Let $R_1 \in O(d)$, $R_2 \in O(m)$, and*

$$\widetilde{P} = P \operatorname{diag}(R_1, R_2).$$

Define

$$\widetilde{c} = R_2^\top c, \qquad \widetilde{Q}_a = R_1^\top \left(\sum_{k=1}^{m} (R_2)_{ka} Q_k \right) R_1.$$

Then

$$L_N(\widetilde{P}, \widetilde{c}, \widetilde{Q}) = L_N(P, c, Q).$$

Hence the profiled loss depends on P only through $\Pi(P) = P_1 P_1^\top$, and its natural parameter space is

$$O(D)/(O(d) \times O(m)) \simeq \mathrm{Gr}(d, D).$$

Proof. The transformed coordinates are

$$\widetilde{s}_i = R_1^\top s_i, \qquad \widetilde{n}_i = R_2^\top n_i.$$

Direct substitution into the definition of $\widetilde{Q}_a$ gives

$$\mathcal{Q}_{\widetilde{Q}}(\widetilde{s}_i) = R_2^\top \mathcal{Q}_Q(s_i).$$

Therefore

$$\widetilde{n}_i - \widetilde{c} - \mathcal{Q}_{\widetilde{Q}}(\widetilde{s}_i) = R_2^\top \big(n_i - c - \mathcal{Q}_Q(s_i) \big).$$

Orthogonal transformations preserve Euclidean norms, so the losses are equal. Two frames have the same tangent projector exactly when they differ by such a block rotation. This proves the Grassmann statement. □

Remark. The theorem removes basis nonuniqueness only. It does not show that the true tangent projector minimizes the loss. That is the separate orientation identifiability problem studied in Section C.5.

C.4 Design conditioning and well-posedness

The fitting criterion is built from noisy observed coordinates. We now show why its design conditions follow from the default observed-radius sampling procedure. After defining the regression features, we first derive the law of the latent coordinates attached to the selected noisy observations. This law gives the design separation needed to detect tangent tilts. A perturbation argument then transfers conditioning to the observed predictors actually used by the estimator. The clean case is obtained by setting $\sigma = 0$ in this same procedure; it is not used as a surrogate sampling model for the noisy proof. Finally, a population perturbation bound and sample concentration place anchor-centered PCA near the true tangent. An explicit Gram margin then makes this computed initializer admissible. These are separate steps: information about latent tangent tilts, stability of the observed regression, and feasibility of its numerical starting point.

Let

$$q_d = \frac{d(d+1)}{2}, \qquad p_d = 1 + q_d, \qquad r_d = 1 + d + q_d.$$

Choose a Frobenius-orthonormal basis $E_1, \ldots, E_{q_d}$ of $\mathrm{Sym}(d)$.

C.4.1 Two different feature vectors

For a scale $h > 0$, define the constant-plus-quadratic feature

$$\phi_h(s) = \begin{pmatrix} 1 \\ \frac{1}{2h^2} s^\top E_1 s \\ \vdots \\ \frac{1}{2h^2} s^\top E_{q_d} s \end{pmatrix} \in \mathbb{R}^{p_d}$$

and the augmented degree-two feature

$$\psi_h(s) = \begin{pmatrix} 1 \\ s/h \\ \frac{1}{2h^2} s^\top E_1 s \\ \vdots \\ \frac{1}{2h^2} s^\top E_{q_d} s \end{pmatrix} \in \mathbb{R}^{r_d}.$$

The empirical quadratic and augmented Gram matrices are

$$\widehat{\Sigma}_{Q,h}(P) = \frac{1}{N} \sum_{i=1}^{N} \phi_h(s_i(P))\phi_h(s_i(P))^\top, \tag{24}$$

$$\widehat{\Sigma}_{2,h}(P) = \frac{1}{N} \sum_{i=1}^{N} \psi_h(s_i(P))\psi_h(s_i(P))^\top. \tag{25}$$

To state the latter distinction directly, let $\Phi_h(z) \in \mathbb{R}^{N \times p_d}$ have rows $\phi_h(z_i)^\top$, and let $L_h(z) \in \mathbb{R}^{N \times d}$ have rows $(z_i/h)^\top$. When $\Phi_h(z)$ has full column rank, write

$$\mathsf{P}_\Phi(z) = \Phi_h(z)\big(\Phi_h(z)^\top \Phi_h(z)\big)^{-1} \Phi_h(z)^\top$$

for the orthogonal projector in sample-index space onto the vectors generated by constant-plus-quadratic functions. The residualized-linear Gram matrix is

$$\frac{1}{N} L_h(z)^\top \big(I_N - \mathsf{P}_\Phi(z)\big) L_h(z).$$

Its positive definiteness means that no nonzero linear sample vector can be represented by a constant-plus-quadratic sample vector. The augmented conditioning inequality in Lemma C.24 is a convenient quantitative sufficient condition for this property and yields the lower bound used in the tangent-risk proof.

Proposition C.15 (Fixed-orientation profiling)**.** *Fix P such that $\widehat{\Sigma}_{Q,h}(P) \succ 0$. Let $X_h(P) \in \mathbb{R}^{N \times p_d}$ have rows $\phi_h(s_i(P))^\top$, and let $Y(P) \in \mathbb{R}^{N \times m}$ have rows $n_i(P)^\top$. In the unconstrained inner problem, the unique coefficient matrix is*

$$\widehat{B}(P) = \big(X_h(P)^\top X_h(P)\big)^{-1} X_h(P)^\top Y(P),$$

where the first row represents c and the remaining rows represent the scaled quadratic coefficients. Moreover, the profiled loss is

$$J_N(P) = \frac{1}{N} \left\| \big(I_N - X_h(P)(X_h(P)^\top X_h(P))^{-1} X_h(P)^\top\big) Y(P) \right\|_F^2.$$

Proof. For fixed P, the model is a multi-output linear least-squares regression in the coefficient matrix B, with loss $N^{-1}\|Y(P) - X_h(P)B\|_F^2$. The positive definiteness assumption makes $X_h(P)^\top X_h(P)$ invertible, so the normal equations give the displayed unique minimizer. Substitution gives the orthogonal-residual formula for $J_N(P)$. □

Remark. The proposition is the computational reduction used throughout the paper: only the orientation remains in the outer nonconvex search. When the curvature constraint $Q \in \mathfrak{Q}_K$ is active, the inner problem is instead a finite convex quadratic program; coefficient uniqueness still follows from $\widehat{\Sigma}_{Q,h}(P) \succ 0$.

The scaling by h changes neither rank nor fitted values. For example, the coefficient vector corresponding to $c_k + \frac{1}{2} s^\top Q_k s$ is

$$\big(c_k, h^2 \langle E_1, Q_k \rangle_F, \ldots, h^2 \langle E_{q_d}, Q_k \rangle_F\big).$$

Changing the Frobenius-orthonormal basis of $\mathrm{Sym}(d)$ right-multiplies the feature matrices by an orthogonal matrix and preserves their Gram eigenvalues.

C.4.2 DENSITY INDUCED BY NOISY OBSERVED-RADIUS SELECTION

Use the sampling convention of Section C.2. The radius r is fixed before sampling, and we assume $r > 2\sigma$ to obtain a nonempty latent region that is always selected. The first step explicitly accounts for the two noise contributions in each observed distance.

Lemma C.16 (Observed-radius sandwich and conditional sampling)**.** *For every non-anchor point,*

$$\|x_i - x_0\| \le r \quad \Longrightarrow \quad \|x_i^* - x_0^*\| \le r + 2\sigma,$$

and

$$\|x_i^* - x_0^*\| \le r - 2\sigma \quad \Longrightarrow \quad i \in \mathcal{I}_r(x_0).$$

Conditional on the anchor variables and on $N_r = q$, the q selected latent/noise pairs, in increasing original-index order, are independent copies of the original pair conditional on the event $\|X - x_0\| \le r$.

Proof. The first implication follows from

$$\|x_i^* - x_0^*\| \le \|x_i - x_0\| + \|\xi_i\| + \|\xi_0\|.$$

The second follows by reversing the same triangle inequality. The conditional sampling assertion is Lemma C.1. □

Lemma C.17 (Slope control from graph regularity)**.** *Let g be a C^2 normal graph on a neighborhood of $\overline{B}_d(0, H)$, with $g(0) = Dg(0) = 0$, and suppose $\sup_{\|t\| \le H} \left\|D^2 g(t)\right\|_{\mathrm{op}} \le B$. For $0 < h \le H$,*

$$\sup_{\|t\| \le h} \|Dg(t)\|_{\mathrm{op}} \le Bh, \qquad \|g(t)\| \le \frac{B}{2}\|t\|^2.$$

In particular, $Bh \le 1$ implies the slope bound used below.

Proof. Integrate $D^2 g(st)[t, \cdot]$ along the segment from zero to t to bound $Dg(t)$, and integrate $Dg(st)t$ once more to bound $g(t)$. The segment is in the graph domain. □

Under Lemma C.7, one may take $B = 3L_\perp$ on any closed ball of radius $H < 1/(8L_\perp)$, giving $BH < 3/8 < 1$. Thus the slope restriction follows from that adopted graph model and its radius; it is not a separate small-curvature assumption. Coverage of all potentially selected latent points remains a distinct requirement. More generally, $\|Dg\|_{\mathrm{op}} \le L$ would replace the constants $\sqrt{2}$ and $2^{d/2}$ below by $\sqrt{1+L^2}$ and $(1+L^2)^{d/2}$. The choice $L = 1$ keeps the constants simple. For noisy radius selection these conditions concern $h_{\mathrm{out}} = r + 2\sigma$, not r alone: a supplied valid local radius H must permit $2\sigma < r \le H - 2\sigma$. The requirement $4\sigma < H$ for such an interval to exist is a compatibility condition for these sufficient conditions, not a universal necessary condition for tangent estimation.

Proposition C.18 (Latent density after observed-radius selection)**.** *Use the i.i.d. noisy model, designated anchor, and prescribed radius $r > 2\sigma$ of Section C.2, with $0 \le \sigma < \tau_M$. Suppose the graph chart in Proposition C.8 is defined on $B_d(0, h_{\mathrm{out}})$, where*

$$h_{\mathrm{out}} = r + 2\sigma,$$

and contains every clean point satisfying $\|x^* - x_0^*\| \le h_{\mathrm{out}}$. Assume the slope bound

$$\sup_{\|t\| \le h_{\mathrm{out}}} \|Dg(t)\|_{\mathrm{op}} \le 1,$$

as guaranteed for example by Lemma C.17 when $Bh_{\mathrm{out}} \le 1$, and the manifold density bound $0 < f_{\min} \le f_M \le f_{\max} < \infty$ on this patch. Conditional on the anchor variables and selection $E_r = \{\|X - X_0\| \le r\}$, the latent tangent coordinate T is supported on $B_d(0, h_{\mathrm{out}})$ and has density

$$f_{T|E_r, \mathcal{F}_0}(t) \ge \frac{\rho_r}{v_d h_{\mathrm{out}}^d} \quad \textit{for almost every } t \in B_d(0, \alpha_r h_{\mathrm{out}}),$$

where $v_d = \mathrm{Vol}(B_d(0,1))$ and

$$\alpha_r = \frac{r - 2\sigma}{\sqrt{2}(r + 2\sigma)}, \qquad \rho_r = \frac{f_{\min}}{2^{d/2} f_{\max}}.$$

Moreover, conditional on the anchor variables,

$$\mathbb{P}\left\{N_r < \frac{n-1}{2} p_-(r,\sigma)\right\} \le \exp\left\{-\frac{n-1}{8} p_-(r,\sigma)\right\},$$

where

$$p_-(r,\sigma) = f_{\min} v_d \left(\frac{r-2\sigma}{\sqrt{2}}\right)^d.$$

Proof. The first implication of Lemma C.16 and the chart assumption show that every selected clean point has a coordinate $\|T\| \le h_{\mathrm{out}}$. Conversely, if $\|t\| \le (r-2\sigma)/\sqrt{2}$, then

$$\|F(t) - x_0^*\| \le \sqrt{\|t\|^2 + \|g(t)\|^2} \le \sqrt{2}\,\|t\| \le r - 2\sigma,$$

so selection occurs for every admissible noise realization.

Define the graph volume factor

$$j_g(t) = \sqrt{\det(I_d + Dg(t)^\top Dg(t))}.$$

The slope bound gives $1 \le j_g(t) \le 2^{d/2}$. By the area formula, the unnormalized latent coordinate density on the patch is $f_M(F(t))j_g(t)$. The graph image covers every latent point that can be selected, so it accounts for the entire selection probability. Let $p_r = \mathbb{P}\{\|X - x_0\| \le r \mid x_0^*, \xi_0\}$. The selected coordinate density is

$$f_{T|\mathrm{sel}}(t) = \frac{f_M(F(t)) j_g(t) \mathbb{P}\{\|X - x_0\| \le r \mid X^* = F(t), x_0^*, \xi_0\}}{p_r}.$$

On the inner ball, the conditional selection probability equals one. Also, $1 \le j_g(t) \le 2^{d/2}$, and selection implies $\|T\| \le h_{\mathrm{out}}$. Hence

$$p_r \le f_{\max} 2^{d/2} v_d h_{\mathrm{out}}^d,$$

which yields

$$f_{T|\mathrm{sel}}(t) \ge \frac{\rho_r}{v_d h_{\mathrm{out}}^d} \qquad (\|t\| \le \alpha_r h_{\mathrm{out}}).$$

Finally, the inner-ball inclusion gives $p_r \ge p_-(r,\sigma)$. Since N_r is conditionally binomial with parameters $n-1$ and p_r, the standard binomial Chernoff bound proves the last display. Conditional on any possible count $N_r = q \ge 1$, the selected coordinates have this same marginal density and are independent by Lemma C.1. □

The variable T is latent, but the selection event E_r is observed and includes noise in both the anchor and the non-anchor point. Thus this proposition is a result for the noisy sampling procedure. It assumes local manifold-density bounds and chart coverage; the selected-coordinate density bound is a conclusion, not an additional sampling assumption. No density or mean-zero condition on the selected noise is used. The remaining question is how this derived density controls polynomial features and, ultimately, observed regression.

C.4.3 From the selected density to latent polynomial conditioning

To reuse the same polynomial argument, we name the density property just proved. For the default procedure, its constants are fixed by Proposition C.18; it is not an independent assumption that must be verified later.

Definition C.19 (Local density condition)**.** Conditional on the anchor variables and the local sample size, we say a coordinate law satisfies the local density condition with parameters (h, ρ, α) if its random vector $Z \in \mathbb{R}^d$ has a density f_h, is supported in $B_d(0,h)$, and satisfies

$$f_h(z) \ge \frac{\rho}{v_d h^d} \qquad \text{for almost every } z \in B_d(0, \alpha h),$$

where $\alpha \in (0,1]$, $\rho > 0$, and $v_d = \mathrm{Vol}(B_d(0,1))$.

Normalization implies $\rho\alpha^d \le 1$. This condition on latent coordinates is used to verify the residualized-linear separation condition. It is not an assumption on the observed regression predictor $s_i(P^*) = t_i + \eta_i$. For the latter, the only design property required by the estimator is the observed quadratic Gram condition, which is transferred separately in Lemma C.26. In every probability statement below, ρ and α are assumed deterministic and uniform over the anchor variables being conditioned on.

For $a \in \mathbb{R}$, $b \in \mathbb{R}^d$, and $H \in \mathrm{Sym}(d)$, define

$$\gamma_d^{(2)} = \min_{a^2+\|b\|^2+\|H\|_F^2=1} \frac{1}{v_d} \int_{B_d(0,1)} \left(a + b^\top y + \frac{1}{2} y^\top H y\right)^2 dy.$$

Lemma C.20 (Degree-two population conditioning)**.** *We have $\gamma_d^{(2)} > 0$. Under the density condition in Definition C.19,*

$$\mathbb{E}[\psi_h(Z)\psi_h(Z)^\top] \succeq \kappa_2 I_{r_d}, \qquad \kappa_2 = \rho\alpha^{d+4}\gamma_d^{(2)}.$$

The same lower bound, restricted to the constant-plus-quadratic coordinates, holds for $\mathbb{E}[\phi_h(Z)\phi_h(Z)^\top]$.

Proof. First consider $\gamma_d^{(2)}$. If its value were zero, compactness of the unit sphere would give a nonzero triple (a, b, H) such that

$$a + b^\top y + \frac{1}{2} y^\top H y = 0$$

for almost every y in the unit ball. A polynomial that vanishes almost everywhere on an open ball vanishes everywhere on that ball. Its constant, linear, and quadratic coefficients must all be zero, a contradiction.

Now let $\theta = (a, b, H)$. By the density lower bound and the substitution $z = \alpha h y$,

$$\begin{aligned}
&\mathbb{E}\left[\left(a + b^\top \frac{Z}{h} + \frac{1}{2h^2} Z^\top H Z\right)^2\right] \\
&\quad \ge \rho\alpha^d \frac{1}{v_d} \int_{B_d(0,1)} \left(a + \alpha b^\top y + \frac{\alpha^2}{2} y^\top H y\right)^2 dy \\
&\quad \ge \rho\alpha^d \gamma_d^{(2)} \left(a^2 + \alpha^2 \|b\|^2 + \alpha^4 \|H\|_F^2\right) \\
&\quad \ge \rho\alpha^{d+4} \gamma_d^{(2)} \left(a^2 + \|b\|^2 + \|H\|_F^2\right).
\end{aligned}$$

This is precisely the Rayleigh-quotient lower bound for the augmented Gram matrix. Setting $b = 0$ gives the quadratic-feature statement. □

C.4.4 Finite-sample conditioning

The next two lemmas are generic polynomial-design facts. Their variable Z deliberately does not designate an observed predictor: the results can be reused for any law satisfying their hypotheses. In the present model we apply them, conditional on the noisy anchor and the selected count, to the latent coordinates T attached to observations selected by observed distance. The following corollary makes that application explicit. Almost-sure rank is a qualitative statement; only the quantitative concentration bound supplies the uniform separation used in the risk proof.

Lemma C.21 (Almost-sure full rank)**.** *Fix $h > 0$. Suppose $Z_1, \dots, Z_N$ are independent and each has an absolutely continuous distribution on $\mathbb{R}^d$. If $N \ge r_d$, then the matrix with rows $\psi_h(Z_i)^\top$ has full column rank almost surely. Consequently, its constant-plus-quadratic submatrix also has full column rank.*

Proof. The coordinates of $\psi_h(z)$ form a basis of the real polynomials of total degree at most two. Suppose the first $j < r_d$ sampled feature vectors span a proper subspace. There is then a nonzero

coefficient vector orthogonal to that span, hence a nonzero degree-two polynomial p_j vanishing at the first j points. The event that the next feature vector remains in the same span is contained in

$$\{z : p_j(z) = 0\}.$$

The zero set of a nonzero polynomial has Lebesgue measure zero, so absolute continuity makes this conditional event have probability zero. Induction shows that the rank increases until it reaches r_d. □

Lemma C.22 (High-probability empirical conditioning)**.** *Under the density condition in Definition C.19, let $Z_1, \dots, Z_N$ be conditionally independent copies of Z. Then*

$$\mathbb{P}\left\{\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N}\psi_h(Z_i)\psi_h(Z_i)^\top\right) < \frac{\kappa_2}{2}\right\} \le r_d \exp\left(-\frac{N\kappa_2}{20}\right).$$

In particular, the augmented and quadratic empirical designs are both well-conditioned with probability at least $1-\delta$ whenever

$$N \ge \frac{20}{\kappa_2}\log\frac{r_d}{\delta}.$$

Proof. Frobenius orthonormality gives $\sum_j (z^\top E_j z)^2 = \|z\|^4$, hence $\|\psi_h(z)\|^2 \le 1+1+1/4 = 9/4$ for $\|z\| \le h$. Thus $Y_i = \psi_h(Z_i)\psi_h(Z_i)^\top$ satisfies $0 \preceq Y_i \preceq (9/4)I$, and Lemma C.20 gives $\lambda_{\min}(\sum_i \mathbb{E}Y_i) \ge N\kappa_2$. Apply the matrix Chernoff lower tail (Tropp, 2012, Corollary 5.2 and Remark 5.3) with deviation $1/2$. Its exponent $-N\kappa_2/18$ implies the stated relaxed constant 20. The quadratic design is a principal submatrix of the augmented design. □

Corollary C.23 (Latent separation for samples selected by noisy distance)**.** *Under Proposition C.18, condition on the anchor variables and $N_r = q$, and order the selected observations by their original indices. Let $t_1, \dots, t_q$ be their latent coordinates. Set $h = h_{\text{out}}$ and $\kappa_{2,r} = \rho_r \alpha_r^{d+4}\gamma_d^{(2)}$. If $q \ge r_d$, their augmented feature matrix has full column rank almost surely. If*

$$q \ge \frac{20}{\kappa_{2,r}}\log\frac{r_d}{\delta},$$

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

$$\frac{1}{q}\sum_{i=1}^{q}\left(a + b^\top\frac{t_i}{h} + \frac{1}{2h^2}t_i^\top H t_i\right)^2 \ge \frac{\kappa_{2,r}}{2}\left(a^2 + \|b\|^2 + \|H\|_F^2\right) \tag{26}$$

simultaneously for all $a \in \mathbb{R}$, $b \in \mathbb{R}^d$, and $H \in \mathrm{Sym}(d)$. Thus item 1 of Theorem C.37 holds with $\mu = \kappa_{2,r}/2$.

Proof. Lemma C.1 gives conditional independence of the selected pairs. Proposition C.18 gives the density of their latent coordinates with parameters $(h_{\text{out}}, \rho_r, \alpha_r)$. Apply Lemma C.21 for rank and Lemma C.22 for the quantitative bound. Expanding the augmented Gram Rayleigh quotient and using the Frobenius-orthonormal basis of $\mathrm{Sym}(d)$ gives equation 26. □

Remark (What this application does and does not require). The selected observations and the anchor remain noisy throughout this corollary. The t_i's are proof variables, not inputs supplied to the algorithm. No absolute continuity of $s_i(P^*) = t_i + \eta_i$ has been established or assumed: bounded, location-dependent noise alone does not justify that inference. Consequently, the generic full-rank and concentration lemmas are not applied directly to these observed predictors. Their observed quadratic Gram is controlled by Lemma C.26.

C.4.5 From latent conditioning to tangent-tilt separation

Before transferring the quadratic Gram to observed predictors, we extract the geometric content of the augmented Gram. The next lemma is deterministic: its z_i's may be any points satisfying the displayed inequality. Keeping this general form separates an algebraic implication from the sampling argument that verifies its premise.

Lemma C.24 (Residualized linear separation)**.** *Suppose $z_1, \dots, z_N \in \mathbb{R}^d$ satisfy*

$$\frac{1}{N}\sum_{i=1}^{N}\left(a + b^\top \frac{z_i}{h} + \frac{1}{2h^2} z_i^\top H z_i\right)^2 \geq \mu\left(a^2 + \|b\|^2 + \|H\|_F^2\right) \tag{27}$$

for every a, b, H. Then, for every $A \in \mathbb{R}^{m\times d}$,

$$\inf_{c,H_1,\dots,H_m} \frac{1}{N}\sum_{i=1}^{N}\left\| Az_i - c - \frac{1}{2}\begin{pmatrix} z_i^\top H_1 z_i \\ \vdots \\ z_i^\top H_m z_i \end{pmatrix}\right\|^2 \geq \mu h^2 \|A\|_F^2 . \tag{28}$$

Proof. Write $a_k^\top$ for the k-th row of A. Apply equation 27 to the k-th output with linear coefficient $b = ha_k$. After minimizing over its constant and quadratic coefficients, the right-hand side is at least $\mu h^2 \|a_k\|^2$. Summing over $k = 1, \dots, m$ gives equation 28. □

Remark (Application to the noisy-data estimator). In Corollary C.23, set $z_i = t_i$, $N = q$, $h = h_{\text{out}}$, and $\mu = \kappa_{2,r}/2$. Equation equation 28 then applies to the selected latent coordinates. Lemma C.36 transfers this separation to the observed residual with explicit Taylor and noise errors; no density of the observed predictors is assumed.

Proposition C.25 (Direct residualized separation)**.** *For any coordinates z_i, let P_Φ be the orthogonal projector onto the column space of the constant-plus-quadratic design (using a pseudoinverse if necessary), and put*

$$S_{\text{lin}} = N^{-1} L_h^\top (I - P_\Phi) L_h.$$

For every $A \in \mathbb{R}^{m\times d}$,

$$\inf_{c,H} \frac{1}{N}\sum_i \|Az_i - c - \mathcal{Q}_H(z_i)\|^2 = h^2 \operatorname{tr}(AS_{\text{lin}}A^\top).$$

Thus $S_{\text{lin}} \succeq \mu_{\text{sep}} I$ is the exact separation condition required for the tilt argument. If the augmented Gram is bounded below by μI, then $S_{\text{lin}} \succeq \mu I$.

Proof. Least-squares projection of each column of $hL_hA^\top$ onto the column space of Φ_h leaves $(I - P_\Phi)hL_hA^\top$. Its squared Frobenius norm divided by N is the displayed trace. The last assertion is Lemma C.24 applied to each row. □

Remark. In the deterministic risk results, augmented conditioning can therefore be replaced by this direct separation condition with $\mu_{\text{sep}} > 0$, provided observed-design feasibility and localization are retained separately. The density-based corollaries use the augmented Gram only as a convenient way to verify all these premises simultaneously. A small quadratic coefficient eigenvalue need not imply equally weak orientation separation.

C.4.6 Transferring conditioning to the observed predictors

So far the design calculation concerns t_i, the latent coordinates of the points selected by noisy distance. The actual regression uses $s_i(P^*) = t_i + \eta_i$. Rather than posit a density for these errors-in-variables predictors, we control their finite-sample Gram matrix directly. The anchor is still noisy: its tangent projection vanishes at the true frame because its perturbation is normal at x_0^*, while its normal offset is handled by the intercept.

Lemma C.26 (Transfer from latent to observed quadratic design)**.** *Suppose $\|t_i\| \leq h$, $s_i(P^*) = t_i + \eta_i$, and $\|\eta_i\| \leq \sigma \leq h$. Define*

$$\widehat{\Sigma}_{Q,h}^{\text{lat}} = \frac{1}{N}\sum_{i=1}^{N} \phi_h(t_i)\phi_h(t_i)^\top.$$

Then

$$\left\|\widehat{\Sigma}_{Q,h}(P^*) - \widehat{\Sigma}_{Q,h}^{\text{lat}}\right\|_{\text{op}} \leq 6\frac{\sigma}{h}.$$

Consequently,

$$\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P^*)) \geq \lambda_{\min}(\widehat{\Sigma}_{Q,h}^{\text{lat}}) - 6\frac{\sigma}{h}.$$

In particular, if $\lambda_{\min}(\widehat{\Sigma}_{Q,h}^{\text{lat}}) \geq \kappa$, $\sigma/h \leq \kappa/12$, and $\tau \leq \kappa/2$, then $\Pi^* \in \mathcal{A}_{\tau,h}$.

Proof. Frobenius orthonormality of the basis of $\mathrm{Sym}(d)$ gives

$$\begin{aligned}\|\phi_h(t+\eta)-\phi_h(t)\| &= \frac{1}{2h^2}\left\|(t+\eta)(t+\eta)^\top - tt^\top\right\|_F \\ &\le \frac{\|t\|\,\|\eta\|}{h^2} + \frac{\|\eta\|^2}{2h^2} \\ &\le \frac{\sigma}{h} + \frac{1}{2}\left(\frac{\sigma}{h}\right)^2 \le \frac{3}{2}\frac{\sigma}{h}.\end{aligned}$$

Moreover,

$$\|\phi_h(t)\| \le \frac{\sqrt{5}}{2}, \qquad \|\phi_h(t+\eta)\| \le \sqrt{5}.$$

Using

$$\left\|aa^\top - bb^\top\right\|_{\mathrm{op}} \le (\|a\| + \|b\|)\,\|a-b\|$$

and averaging gives the displayed bound. Weyl's inequality yields the eigenvalue statement. □

Lemma C.20 and this transfer lemma have different roles. The former is a population statement proved from a density lower bound; empirical concentration is needed before applying it to a sample. The present lemma is a deterministic, samplewise perturbation bound and requires neither independence nor a density for observed predictors. In particular, it remains valid after conditioning on the noisy anchor and radius selection. Together they control the observed quadratic Gram that determines feasibility of Π^*.

Lemma C.27 (Normal-noise refinement of predictor perturbation)**.** *In a fixed local instance, suppose* $\sup_{\|t\|\le h}\|D^2 g(t)\|_{\mathrm{op}} \le B$. *Set*

$$\beta_h = \min\{\sigma/h, B\sigma\}.$$

If $\beta_h \le 1$, *then*

$$\|\widehat{\Sigma}_{Q,h}(P^*) - \widehat{\Sigma}^{\mathrm{lat}}_{Q,h}\|_{\mathrm{op}} \le 6\beta_h.$$

Without a supplied Hessian bound one may use $\beta_h = \sigma/h$. *This statement is samplewise and survives noisy-radius selection.*

Proof. Write $(P^*)^\top \xi_i = (\eta_i, \zeta_i)$. Orthogonality to every $(v, Dg(t_i)v)$ gives $\eta_i = -Dg(t_i)^\top \zeta_i$. Since $Dg(0) = 0$, $\|\eta_i\| \le Bh\sigma$, as well as $\|\eta_i\| \le \sigma$. The anchor contributes no tangent component at the true frame. Apply the proof of Lemma C.26 with perturbation bound $h\beta_h$. No mean-zero or independence property is used. □

Remark (Separate noise requirements). With a latent Gram lower bound $\kappa/2$, the condition $\beta_h \le \kappa/24$ gives an observed Gram lower bound $\kappa/4$. It can replace the stronger $\sigma/h \le \kappa/24$ in design transfer. It does not remove the separate response-noise, radius-coverage, PCA or risk conditions. The conservative corollaries below remain valid as stated; the refined alternative is recorded explicitly where used.

C.4.7 Local stability of the quadratic design

We now vary the orientation while keeping the selected observed sample fixed. Both Gram matrices in the next proposition are computed from $s_i(P) = P_1^\top u_i$, where $u_i = x_i - x_0$; even the Gram at P^* is an observed Gram. This is a deterministic stability result, not a second clean-to-noisy transfer.

Proposition C.28 (Observed conditioning persists near the true tangent)**.** *Assume* $\|u_i\| \le bh$ *for all* i, *for some* $b \ge 1$, *and*

$$\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P^*)) \ge \kappa.$$

Then

$$\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P)) \ge \kappa - C_b\,\|\Pi(P) - \Pi^*\|_F, \qquad C_b = 2b^2\sqrt{1 + b^4/4}.$$

In particular, if $\|\Pi(P) - \Pi^*\|_F \le \kappa/(2C_b)$, *then*

$$\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P)) \ge \frac{\kappa}{2}.$$

Proof. Let $\delta = \|\Pi(P) - \Pi^*\|_F$. Orthogonal Procrustes alignment gives an $R \in O(d)$ such that

$$\|P_1 R - P_1^*\|_F \le \delta.$$

Rotating P_1 by R only rotates the quadratic feature coordinates and does not change Gram eigenvalues. Put

$$a_i = (P_1 R)^\top u_i / h, \qquad b_i = (P_1^*)^\top u_i / h.$$

Then $\|a_i\|, \|b_i\| \le b$ and $\|a_i - b_i\| \le b\delta$.

For

$$\phi(v) = \left(1, \frac{1}{2} v^\top E_1 v, \ldots, \frac{1}{2} v^\top E_{q_d} v\right)^\top,$$

we have

$$\|\phi(a_i) - \phi(b_i)\| = \frac{1}{2} \left\| a_i a_i^\top - b_i b_i^\top \right\|_F \le b^2 \delta$$

and $\|\phi(a_i)\|, \|\phi(b_i)\| \le \sqrt{1 + b^4/4}$. Hence

$$\left\| \phi(a_i)\phi(a_i)^\top - \phi(b_i)\phi(b_i)^\top \right\|_{\mathrm{op}} \le C_b \delta.$$

Averaging and applying Weyl's inequality proves the result. □

Remark. The form with $b = 1$ applies to observed-radius selection with $h = r + 2\sigma$, since selected rows satisfy $\|u_i\| \le r \le h$. For a fixed noisy-anchor instance, a latent bound $\|t_i\| \le h$ does not by itself imply $\|u_i\| \le h$. Under the curvature and remainder bounds of Theorem C.37, it instead gives

$$\|u_i\| \le h + \frac{K}{2} h^2 + \frac{L_3}{6} h^3 + 2\sigma,$$

so the proposition may be invoked with the corresponding b. This is why latent and observed neighborhood scales are kept distinct.

C.4.8 An admissible compact optimization problem

The preceding bounds motivate two restrictions on the estimator: an observed Gram threshold makes coefficient fitting stable, and a local trust region confines the orientation search to the tangent chart used in the risk analysis. We define the statistical estimator with these restrictions first; the numerical methods in Appendix C.10 approximate this constrained problem. Compactness establishes existence, not numerical convergence.

For $\tau > 0$, define

$$\mathcal{A}_{\tau,h} = \left\{ \Pi(P) \in \mathrm{Gr}(d, D) : \lambda_{\min}(\widehat{\Sigma}_{Q,h}(P)) \ge \tau \right\}. \tag{29}$$

This definition is basis-independent by block rotational invariance.

For the risk theorem, it is convenient to use the geometrically natural curvature constraint

$$\mathfrak{Q}_K = \left\{ Q : \sum_{k=1}^{m} \|Q_k\|_F^2 \le K^2 \right\}.$$

It prevents a nearly singular orientation from being compensated by unbounded curvature. The constraint is invariant under tangent and normal basis changes. In the deterministic theory, K is a supplied upper bound that must contain the true coordinate curvature. In an implementation it is a regularization parameter and must be fixed or tuned independently of the oracle quantity Q^*; the unconstrained fit is the $K = \infty$ baseline.

The bound need not be oracle when a quantitative geometric class is supplied. For example, if a known graph-Hessian bound $\sup_{\|t\| \le h} \|D^2 g(t)\|_{\mathrm{op}} \le B_2$ is available, then $\|Q_k^*\|_F \le \sqrt{d}\, B_2$ for every normal coordinate, so the conservative choice $K = \sqrt{dm}\, B_2$ ensures $Q^* \in \mathfrak{Q}_K$. Under the adopted quantitative C^3 class, Lemma C.7 gives the usable bound $B_2 = 3L_\perp$. Alternatively, $\sum_k \|Q_k^*\|_F^2 = \sum_{a,b} \|D^2 g(0)[e_a, e_b]\|^2 \le d^2 B_2^2$, so $K = dB_2$ is also valid and avoids codimension dependence. A $C^{2,1}$ remainder bound controls variation of the Hessian but does not by itself bound

$D^2 g(0)$; it must be combined with a known reach or curvature bound to calibrate K. For a bare C^2 model with no such quantitative information, K remains a supplied regularization parameter.

Let $\Pi_0 \in \mathrm{Gr}(d, D)$ be a preliminary estimator and let $r_0 > 0$. For the default implementation, Π_0 is the anchor-centered PCA projector specified in Proposition C.32. Define

$$\mathcal{C}_{\tau,h}(\Pi_0, r_0) = \mathcal{A}_{\tau,h} \cap \{\Pi : \|\Pi - \Pi_0\|_F \le r_0\}.$$

Define

$$\widehat{\Pi} \in \operatorname*{arg\,min}_{\Pi(P) \in \mathcal{C}_{\tau,h}(\Pi_0, r_0)} \min_{c,\, Q \in \mathfrak{Q}_K} L_N(P, c, Q). \tag{30}$$

Two feasibility questions must be distinguished. Nonemptiness of this set, for example because it contains Π^*, suffices to define the exact estimator. A feasible numerical method initialized at Π_0 also needs $\Pi_0 \in \mathcal{A}_{\tau,h}$. Proximity to Π^* alone does not imply this if its Gram eigenvalue is exactly at the threshold. Corollary C.34 supplies the additional margin and its high-probability guarantee for the actual PCA initializer.

Failure convention. For $N = 0$, or an empty localized feasible set, the procedure returns a failure flag rather than an undefined projector. A constrained numerical solve additionally checks initializer feasibility. Probability guarantees below are success-event statements. If a total measurable estimator is required, fix measurable tie breaking and a deterministic rank-d fallback projector on failure; no small-error bound is claimed on failure.

Remark (Supplied tuning and observable diagnostics). The finite-sample guarantees treat K, τ, and r_0 as fixed supplied parameters. If quantitative geometry, density, noise, and initialization-error bounds are known, they may be set conservatively before seeing the local sample: the preceding paragraph gives one route for K, Corollary C.34 permits $\tau < \kappa_{2,r}/4$, and Corollary C.42 gives the sufficient window $r_{\mathrm{init}} \le r_0$ and $r_{\mathrm{init}} + r_0 < 1/\sqrt{5}$. In applications where these constants are unavailable, they are regularization parameters rather than certified data-only choices.

The test $\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0)) \ge \tau$ is an implementable diagnostic for feasibility of the computed initializer. It does not by itself certify that Π^* belongs to the localized feasible set. That stronger inclusion is a model-based conclusion under the stated density, geometry, noise, and tuning conditions. Tuning K, τ, or r_0 from the same local sample requires additional selection analysis if the present probability guarantees are to be retained.

Theorem C.29 (Existence and coefficient uniqueness)**.** *Suppose $\mathcal{C}_{\tau,h}(\Pi_0, r_0) \neq \varnothing$. Then a minimizer in equation 30 exists. For every fixed admissible P, the minimizing coefficient pair (c, Q) is unique.*

Proof. The map $P \mapsto \widehat{\Sigma}_{Q,h}(P)$ is continuous, as is its smallest eigenvalue. Hence $\mathcal{A}_{\tau,h}$ is closed. The Grassmann manifold is compact, so $\mathcal{C}_{\tau,h}(\Pi_0, r_0)$ is compact.

For fixed P, the loss is a quadratic function of the coefficients. The condition $\lambda_{\min} \ge \tau > 0$ makes this quadratic strictly convex. It is therefore coercive in the intercept and strictly convex on the convex set $\mathbb{R}^m \times \mathfrak{Q}_K$. Thus it has a unique minimizer.

In any continuous local choice of frame, the fitted coefficients depend continuously on P. Indeed the minimizing intercept is $N^{-1} \sum_i (n_i(P) - \mathcal{Q}_Q(s_i(P)))$, uniformly bounded on the compact frame/coefficient domain. Every sequence of minimizers thus has convergent subsequences, and continuity of the loss and uniqueness identify each limit. Block invariance makes the resulting value function globally well-defined. Consequently, the profiled loss is continuous on the compact set $\mathcal{C}_{\tau,h}(\Pi_0, r_0)$. The extreme-value theorem now gives an empirical minimizer. □

Remark. The set defined only by $\lambda_{\min}(\widehat{\Sigma}_{Q,h}) > 0$ is open and need not contain a minimizer. The quantitative threshold $\lambda_{\min} \ge \tau$ is what makes the admissible set closed. Global uniqueness is neither asserted nor needed. Because of block rotations, the frame itself could never be unique; the meaningful object is the locally identified projector.

C.4.9 PUTTING THE NOISY DESIGN GUARANTEES TOGETHER

We now combine the derived selected density, empirical concentration, and observed-design transfer. The first conclusion supplies the latent separation used in the risk proof; the second makes the true projector feasible for the regression on noisy observations.

Corollary C.30 (Observed-radius verification of the local design)**.** *Under Proposition C.18, set*

$$\kappa_{2,r} = \rho_r \alpha_r^{d+4} \gamma_d^{(2)}.$$

Conditional on the anchor variables and on $N_r = q$, if

$$q \geq \frac{20}{\kappa_{2,r}} \log \frac{r_d}{\delta},$$

then the latent augmented-design condition in item 1 of Theorem C.37 holds with probability at least $1 - \delta$, with $\mu = \kappa_{2,r}/2$. If in addition

$$\frac{\sigma}{h_{\mathrm{out}}} \leq \frac{\kappa_{2,r}}{24}, \qquad 0 < \tau \leq \frac{\kappa_{2,r}}{4},$$

then $\Pi^* \in \mathcal{A}_{\tau, h_{\mathrm{out}}}$ on the same event.

Proof. Proposition C.18 verifies the density condition in Definition C.19 with $h = h_{\mathrm{out}}$, $\rho = \rho_r$, and $\alpha = \alpha_r$. Apply Corollary C.23. Its quadratic principal submatrix has minimum eigenvalue at least $\kappa_{2,r}/2$. Since $\kappa_{2,r} \leq 1$, the displayed noise condition implies $\sigma \leq h_{\mathrm{out}}$, so Lemma C.26 yields

$$\lambda_{\min}(\widehat{\Sigma}_{Q, h_{\mathrm{out}}}(P^*)) \geq \frac{\kappa_{2,r}}{4} \geq \tau.$$

□

Remark (Refined normal-noise route). If a supplied graph-Hessian bound B is available, the noise premise in Corollary C.30 may be replaced by $\beta_h = \min\{\sigma/h, B\sigma\} \leq \kappa_{2,r}/24$, using Lemma C.27. The same replacement is valid in Corollaries C.34 and C.42 provided $\sigma \leq h$ is retained for the risk result and all separate PCA and radius conditions remain in force.

C.4.10 THE NOISELESS CASE OF THE SAME RADIUS PROCEDURE

Corollary C.31 (Clean observed-radius design as a special case)**.** *Under the geometric and density conditions of Proposition C.18, set $\sigma = 0$. Then $h = r$, $\alpha_0 = 1/\sqrt{2}$, and $\rho_0 = f_{\min}/(2^{d/2} f_{\max})$. The selected coordinate law satisfies the local density condition with these constants. Let $\kappa_{2,0} = \rho_0 \alpha_0^{d+4} \gamma_d^{(2)}$. Conditional on the anchor and $N_r = q$, if*

$$q \geq \frac{20}{\kappa_{2,0}} \log \frac{r_d}{\delta},$$

then, with probability at least $1 - \delta$, latent augmented separation holds with $\mu = \kappa_{2,0}/2$ and

$$\lambda_{\min}(\widehat{\Sigma}_{Q,r}(P^*)) \geq \frac{\kappa_{2,0}}{2}.$$

In particular, $\Pi^* \in \mathcal{A}_{\tau,r}$ for $0 < \tau \leq \kappa_{2,0}/2$.

Proof. Setting $\sigma = 0$ in the noisy selection proposition gives the stated density constants. Apply Lemma C.22 to the selected latent coordinates. Now $s_i(P^*) = t_i$, so the observed quadratic Gram is the quadratic principal submatrix of the same augmented Gram, with no perturbation loss. □

This specialization uses the same ambient radius event, now with $X = X^*$ and $X_0 = X_0^*$. Conditioning on an independently specified latent graph patch would be a different sampling rule and is not needed to justify the default estimator.

C.4.11 INITIALIZATION AND RECOVERY FOR THE DEFAULT RADIUS PROCEDURE

The preceding results establish design separation and observed feasibility from the noisy sampling model. To use the local risk theorem, we must also place its feasible region near the true tangent. We next analyze anchor-centered PCA for initialization and combine this with the design event. The argument has three parts: concentration around the selected population scatter, geometric control of its population eigenspace, and a Gram margin that places the computed PCA projector inside the admissible set. The last step is needed to start the feasible numerical method. We then combine initialization and design to obtain local tangent recovery.

Proposition C.32 (PCA initialization under a covariance certificate)**.** *Condition on the anchor and on $N_r = q \geq 1$. Let $U = X - x_0$ have the selected observed-radius law, define*

$$\Sigma_{u,r} = \mathbb{E}[UU^\top \mid x_0, \|X - x_0\| \leq r],$$

and let Π_{pop} be its top-d spectral projector. Suppose

$$\lambda_d(\Sigma_{u,r}) - \lambda_{d+1}(\Sigma_{u,r}) \geq \gamma > 0, \qquad \|\Pi_{\mathrm{pop}} - \Pi^*\|_F \leq b_{\mathrm{pop}}.$$

Let Π_0 be the top-d projector of

$$\widehat{\Sigma}_{u,r} = \frac{1}{q} \sum_{i \in \mathcal{I}_r(x_0)} u_i u_i^\top.$$

Then there is a universal constant $C > 0$ such that, with conditional probability at least $1 - \delta$,

$$\|\Pi_0 - \Pi^*\|_F \leq b_{\mathrm{pop}} + C \frac{\sqrt{d}\, r^2}{\gamma} \sqrt{\frac{\log(2/\delta)}{q}},$$

provided

$$C r^2 \sqrt{\frac{\log(2/\delta)}{q}} \leq \frac{\gamma}{4}.$$

Proof. All expectations and probabilities in this proof condition on the anchor and selected count. The selected rows are independent and $\|u_i\| \leq r$. Put $Y_i = u_i u_i^\top$, $S = q^{-1} \sum_i (Y_i - \mathbb{E}Y_i)$, and $f = \|S\|_F$. Independence and the Hilbert-space variance identity give

$$\mathbb{E}f^2 = \frac{1}{q}\mathbb{E}\|Y_1 - \mathbb{E}Y_1\|_F^2 \leq \frac{1}{q}\mathbb{E}\|Y_1\|_F^2 \leq \frac{r^4}{q}.$$

Changing one row changes f by at most $2r^2/q$. The scalar bounded-differences inequality (McDiarmid, 1989) gives, with probability at least $1 - \delta$,

$$\|S\|_{\mathrm{op}} \leq f \leq \frac{r^2}{\sqrt{q}} + r^2 \sqrt{\frac{2\log(1/\delta)}{q}} \leq C r^2 \sqrt{\frac{\log(2/\delta)}{q}}.$$

The concentration step is dimension-free. Under the displayed smallness condition, Davis–Kahan Davis & Kahan (1970) gives

$$\|\Pi_0 - \Pi_{\mathrm{pop}}\|_F \leq C \frac{\sqrt{d}}{\gamma} \|S\|_{\mathrm{op}}.$$

The triangle inequality proves the assertion. □

Proposition C.33 (Population PCA alignment from local signal dominance)**.** *Under the assumptions of Proposition C.18, retain the curvature and cubic-remainder assumptions of Theorem C.37, and write $h = h_{\mathrm{out}}$ and $\lambda_{\mathrm{lin},r} = \rho_r \alpha_r^{d+2}/(d+2)$. Define*

$$G_h = \frac{K}{2} + \frac{L_3 h}{6}, \qquad \Delta_h = G_h h^2 + 2\sigma, \qquad \varepsilon_{\mathrm{pop}} = 2h\Delta_h + \Delta_h^2.$$

If

$$\varepsilon_{\mathrm{pop}} \leq \frac{\lambda_{\mathrm{lin},r} h^2}{4},$$

then the selected-population scatter matrix $\Sigma_{u,r}$ has an d-th spectral gap at least

$$\gamma = \frac{\lambda_{\mathrm{lin},r} h^2}{2},$$

and its top-d projector satisfies

$$\|\Pi_{\mathrm{pop}} - \Pi^*\|_F \le C \frac{\sqrt{d}\,\varepsilon_{\mathrm{pop}}}{\lambda_{\mathrm{lin},r} h^2}.$$

This conclusion is conditional on the anchor variables and does not require the selected noise to remain mean-zero.

Proof. Work in the true tangent–normal frame. For a selected point, write

$$V = (P^*)^\top U, \qquad V_0 = \begin{pmatrix} T \\ 0 \end{pmatrix}, \qquad W = V - V_0.$$

The Taylor representation and the curvature constraint give

$$\|g(T)\| \le \frac{K}{2}\|T\|^2 + \frac{L_3}{6}\|T\|^3 \le G_h h^2.$$

Since $U = (F(T) - x_0^*) + (\xi - \xi_0)$, $\|W\| \le \Delta_h$ and $\|V_0\| \le h$. Thus, with $S_0 = \mathbb{E}[V_0 V_0^\top \mid x_0, \mathrm{sel}]$,

$$\left\|\mathbb{E}[VV^\top \mid x_0, \mathrm{sel}] - S_0\right\|_{\mathrm{op}} \le 2h\Delta_h + \Delta_h^2 = \varepsilon_{\mathrm{pop}}.$$

For every unit vector v, the selected density lower bound and the second moment of a uniform unit ball give

$$\mathbb{E}[(v^\top T)^2]/h^2 \ge \rho_r \alpha_r^{d+2} \frac{1}{v_d} \int_{B_d(0,1)} (v^\top y)^2\, dy = \frac{\rho_r \alpha_r^{d+2}}{d+2} = \lambda_{\mathrm{lin},r}.$$

Consequently,

$$\lambda_d(S_0) \ge \lambda_{\mathrm{lin},r} h^2, \qquad \lambda_{d+1}(S_0) = 0.$$

Weyl's inequality and the assumed perturbation bound therefore give an d-th eigengap of at least $\lambda_{\mathrm{lin},r} h^2 - 2\varepsilon_{\mathrm{pop}} \ge \gamma$. Davis–Kahan then yields the projector bound, after using the Frobenius–operator comparison for a perturbation of two rank-d projectors (the factor $\sqrt{2}$ is absorbed into C). □

Remark (Initializer convention and perturbation tools). The Frobenius bounded-differences argument and Davis–Kahan perturbation bound provide the PCA sampling guarantee. Applying them to the conditional selected law explicitly accounts for the noisy anchor and observed-radius selection; no preservation of the pre-selection noise mean is required. Proposition C.32 analyzes the anchor-centered scatter matrix $q^{-1}\sum u_i u_i^\top$. A PCA implementation that additionally subtracts the selected sample mean has a different population covariance and requires a separate mean-concentration argument; it should not be claimed to be covered by this proposition without that argument.

Remark. Proposition C.32 proves the sampling part of an initializer guarantee. Proposition C.33 derives its population-alignment premise in the local signal-dominance regime. It is not a statement for fixed noise at arbitrarily small radius or for an unrestricted global neighborhood; those settings can fail through the population phenomenon discussed in Section C.6. If the combined displayed bounds are at most r_{init} and $r_{\mathrm{init}} + r_0 < 1/\sqrt{5}$, then the initialization and chart clauses in item 3 of Theorem C.37 hold.

Corollary C.34 (High-probability admissibility of the PCA initializer)**.** *Under Proposition C.18, set $h = h_{\mathrm{out}}$, $\kappa = \kappa_{2,r}$, and condition on the anchor variables and $N_r = q$. Suppose the population certificate $(\gamma, b_{\mathrm{pop}})$ in Proposition C.32 holds. Let $\delta_{\mathrm{des}}, \delta_{\mathrm{pca}} \in (0,1)$, assume*

$$q \ge \frac{20}{\kappa}\log\frac{r_d}{\delta_{\mathrm{des}}}, \qquad \frac{\sigma}{h} \le \frac{\kappa}{24}, \qquad Cr^2\sqrt{\frac{\log(2/\delta_{\mathrm{pca}})}{q}} \le \frac{\gamma}{4},$$

and define

$$\varepsilon_{\mathrm{init}}(q) = b_{\mathrm{pop}} + C\frac{\sqrt{d}\,r^2}{\gamma}\sqrt{\frac{\log(2/\delta_{\mathrm{pca}})}{q}}.$$

For the anchor-centered PCA projector Π_0, with conditional probability at least $1-\delta_{\mathrm{des}}-\delta_{\mathrm{pca}}$,

$$\|\Pi_0-\Pi^*\|_F \le \varepsilon_{\mathrm{init}}(q), \qquad \lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0)) \ge \frac{\kappa}{4}-\sqrt{5}\,\varepsilon_{\mathrm{init}}(q),$$

where P_0 is any orthogonal completion of a basis of $\operatorname{im}\Pi_0$. In particular, for $0<\tau<\kappa/4$ and $\nu\ge 0$, if

$$\sqrt{5}\,\varepsilon_{\mathrm{init}}(q)+\tau+\nu \le \frac{\kappa}{4}, \tag{31}$$

then $\Pi_0\in\mathcal{C}_{\tau,h}(\Pi_0,r_0)$ for any $r_0>0$, with observed Gram eigenvalue at least $\tau+\nu$. If $\nu>0$, this starting point is strictly feasible for both orientation constraints.

Proof. The proof of Corollary C.30 gives $\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P^*))\ge\kappa/4$ outside an event of conditional probability δ_{des}. Proposition C.32 gives the projector error bound outside an event of conditional probability δ_{pca}. Intersect the events by a union bound; independence between the PCA and design events is not needed, even though they use the same sample. Since $\|u_i\|\le r\le h$, Proposition C.28 applies with $b=1$ and $C_1=\sqrt{5}$, giving the displayed eigenvalue bound. The margin condition proves Gram feasibility, and $\|\Pi_0-\Pi_0\|_F=0<r_0$ proves strict trust-region feasibility. □

Remark (How the margin can be met). Proposition C.33 derives the population certificate from the same noisy sampling model. For example, $\tau\le\kappa/8$ and $\varepsilon_{\mathrm{init}}(q)\le\kappa/(16\sqrt{5})$ leave a strict Gram margin $\nu=\kappa/16$. Increasing q reduces only the sampling part of $\varepsilon_{\mathrm{init}}(q)$; curvature and noise must also make b_{pop} small enough. More generally, for a target error $e>b_{\mathrm{pop}}$, it suffices, in addition to the design and concentration requirements above, to take

$$q \ge \frac{Cdr^4}{\gamma^2(e-b_{\mathrm{pop}})^2}\log\frac{2}{\delta_{\mathrm{pca}}}.$$

These are sufficient local conditions, not a claim that arbitrary PCA initializations are admissible. Strict orientation feasibility does not imply an inactive curvature constraint in the profiled coefficient problem or positive curvature of the profiled orientation loss.

Remark (Combined calibration, not only design SNR). The quadratic-design constant $\kappa=\kappa_{2,r}$ and the tangent second-moment constant $\lambda=\lambda_{\mathrm{lin},r}$ have different roles. Write $w=\Delta_h/h=Kh/2+L_3h^2/6+2\sigma/h$. The population certificate is $b_{\mathrm{pop}}\le C\sqrt{d}(2w+w^2)/\lambda$, whereas strict initializer feasibility typically requires error of order κ. For fixed fractions of the Gram margin, sufficient calibration has the form

$$w\le c\min\{\lambda,\lambda\kappa/\sqrt{d}\}, \qquad q\ge C\frac{d(r/h)^4}{\lambda^2\kappa^2}\log\frac{2}{\delta_{\mathrm{pca}}},$$

in addition to count, design, and trust-region requirements. Thus a small design-transfer perturbation alone is not an initialization guarantee. These are conservative sufficient conditions, not sample-complexity lower bounds. In particular, the density-derived κ can be very small in moderate dimension; observable Gram diagnostics and empirical success should not be confused with satisfaction of this full certificate.

Corollary C.42 below combines these sampling and initialization events with deterministic recovery.

C.5 Tangent identifiability and risk

C.5.1 Local coordinates on the Grassmann manifold

Work in the true frame, so a clean centered point has coordinates

$$\begin{pmatrix} t \\ g(t)\end{pmatrix}.$$

Every sufficiently close d-dimensional subspace can be represented by a matrix $A\in\mathbb{R}^{m\times d}$:

$$P_1(A)=\begin{pmatrix} I_d \\ A\end{pmatrix}(I_d+A^\top A)^{-1/2}, \tag{32}$$

$$P_2(A)=\begin{pmatrix} -A^\top \\ I_m\end{pmatrix}(I_m+AA^\top)^{-1/2}. \tag{33}$$

The corresponding projector is denoted by $\Pi(A)$.

Here A is a local coordinate for a *candidate* tangent subspace, not an additional regression coefficient. In the true frame, $\mathrm{im}(P_1(A)) = \{(v, Av) : v \in \mathbb{R}^d\}$; hence $A = 0$ is the true tangent space and At is the normal displacement predicted by a tilted candidate plane at tangent coordinate t. The singular values of A are the tangents of the principal angles to the true tangent space.

Lemma C.35 (Tilt size and projector distance)**.** *If $\|A\|_{\mathrm{op}} \le 1/2$, then*

$$c_A \|A\|_F \le \|\Pi(A) - \Pi^*\|_F \le \sqrt{2}\,\|A\|_F,$$

where one may take $c_A = \sqrt{8/5}$.

Proof. The singular values a_j of A are the tangents of the principal angles: $a_j = \tan\theta_j$. Therefore

$$\|\Pi(A) - \Pi^*\|_F^2 = 2\sum_j \frac{a_j^2}{1 + a_j^2}.$$

When $a_j \le 1/2$,

$$\frac{4}{5}a_j^2 \le \frac{a_j^2}{1 + a_j^2} \le a_j^2.$$

Summing proves the inequalities. □

C.5.2 Why a tangent tilt leaves a linear residual

Let

$$B_d(A) = (I_d + A^\top A)^{-1/2}, \qquad B_m(A) = (I_m + AA^\top)^{-1/2}.$$

For a clean graph point, direct multiplication gives

$$s_A(t) = B_d(A)\big(t + A^\top g(t)\big), \tag{34}$$

$$n_A(t) = B_m(A)\big(g(t) - At\big). \tag{35}$$

These coordinates explain why finding any regression graph is insufficient. At the latent origin, $Ds_A(0) = B_d(A)$ is invertible, so the same clean manifold admits a local graph $n = f_A(s)$ in every such candidate frame. Its derivative is

$$Df_A(0) = -B_m(A)AB_d(A)^{-1} = -A,$$

where the last identity follows from $B_m(A)A = AB_d(A)$, as seen by a singular value decomposition of A. Thus the zero-slope condition selects $A = 0$ geometrically within this chart. This is a geometry statement at the latent origin. For finite noisy observations, a quadratic fit can partly mimic a slope; the residualized design condition and the following perturbation lemma are needed to turn this observation into an error bound.

Lemma C.36 (Tilted-frame decomposition at a general radius)**.** *Assume $\|t\| \le h$, $h > 0$, $\|A\|_{\mathrm{op}} \le 1/2$, $\sum_k \|Q_k^*\|_F^2 \le K^2$, and $\|R(t)\| \le R_h$ throughout the graph ball. Let $Q \in \mathfrak{Q}_K$, and define*

$$D_h = \frac{1}{2}\left(\frac{K}{2}h^2 + R_h\right).$$

For every c, Q, the clean residual can be written as

$$n_A(t) - c - \mathcal{Q}_Q(s_A(t)) = -B_m(A)At - a - \mathcal{Q}_H(t) + e_A(t) \tag{36}$$

for some $a \in \mathbb{R}^m$ and symmetric $H_1, \dots, H_m$, with

$$\|e_A(t)\| \le R_h + KhD_h + \frac{K}{2}D_h^2. \tag{37}$$

For observed anchored displacements, including a noisy anchor, the same representation holds with

$$\|e_A(t)\| \le \mathcal{E}(h, \sigma; R_h) := R_h + KhD_h + \frac{K}{2}D_h^2 + 2\sigma + 2K(h + D_h)\sigma + 2K\sigma^2. \tag{38}$$

A sharper version retains the tilt. With $G_h = Kh^2/2 + R_h = 2D_h$, $a = \|A\|_{\mathrm{op}}$, and

$$V_h = 2\sigma + 2K(h + D_h)\sigma + 2K\sigma^2,$$

the observed remainder also satisfies

$$\|e_A(t)\| \le R_h + KhG_h a + \frac{K}{2} G_h^2 a^2 + V_h. \tag{39}$$

For the cubic model take $R_h = L_3 h^3/6$. No numerical upper bound on h is imposed.

Proof. Substituting $g(t) = \mathcal{Q}_{Q^*}(t) + R(t)$ gives

$$n_A(t) = -B_m(A)At + B_m(A)\mathcal{Q}_{Q^*}(t) + B_m(A)R(t).$$

The middle term is quadratic in t. Next write

$$s_A(t) = B_d(A)t + d_A(t), \qquad d_A(t) = B_d(A)A^\top g(t).$$

Since $\|B_d(A)\|_{\mathrm{op}} \le 1$,

$$\|d_A(t)\| \le \frac{1}{2}\|g(t)\| \le \frac{1}{2}\left(\frac{K}{2}h^2 + R_h\right) = D_h.$$

For the aggregate coefficient norm defining $\mathfrak{Q}_K$,

$$\|\mathcal{Q}_Q(v + w) - \mathcal{Q}_Q(v)\| \le K\|v\|\|w\| + \frac{K}{2}\|w\|^2.$$

Apply this with $v = B_d(A)t$ and $w = d_A(t)$. Collect the constant and both quadratic polynomials in a, H; the remaining error has norm at most $R_h + KhD_h + KD_h^2/2$.

For the noisy displacement, $u_i = (x_i^* - x_0^*) + (\xi_i - \xi_0)$, where $\|\xi_i - \xi_0\| \le 2\sigma$. Both candidate coordinates change by at most 2σ. The clean predictor norm is at most $h + D_h$, so the additional response and quadratic errors are at most $2\sigma + 2K(h + D_h)\sigma + 2K\sigma^2$. This proves the asserted uniform bound without dropping any powers of h. For equation 39, retain $\|d_A(t)\| \le \|A\|_{\mathrm{op}} G_h$ in the clean calculation; the same uniform noisy perturbation V_h remains valid. □

C.5.3 Finite-sample local tangent bound

For the cubic graph model, set

$$\varepsilon(h, \sigma) = \mathcal{E}(h, \sigma; L_3 h^3/6),$$

with $\mathcal{E}$ defined in Lemma C.36. The general bound below retains its dependence on the physical radius. For the default sampling procedure, Lemma C.16 and orthogonal projection give, for every selected observation,

$$\|t_i\| = \left\|(P_1^*)^\top (x_i^* - x_0^*)\right\| \le \|x_i^* - x_0^*\| \le r + 2\sigma.$$

Thus $h = h_{\mathrm{out}}$ guarantees the coordinate bound, including noise at the anchor. Chart coverage must still hold. There is no requirement that this radius be at most one.

Theorem C.37 (Deterministic noisy-anchor local tangent bound)**.** *Consider a fixed noisy-anchor local instance in the sense of Definition C.2. Assume the following.*

1. *The latent coordinates satisfy $\|t_i\| \le h$, with $h > 0$ inside the supplied graph domain, and the augmented conditioning inequality equation 27 with constant $\mu > 0$. Under Definition C.19 and conditional independence given the anchor and local count, Lemma C.22 verifies this item with $\mu = \kappa_2/2$ by its Rayleigh-quotient bound. Corollary C.23 gives its explicit high-probability application to the default noisy radius procedure.*
2. *The true curvature belongs to $\mathfrak{Q}_K$, and the cubic remainder bound equation 16 holds.*
3. *The admissible trust region in equation 30 contains Π^* and is contained in the chart $\|A\|_{\mathrm{op}} \le 1/2$.*
4. *For $i = 0, \dots, N$, $\xi_i \in N_{x_i^*}M$ and $\|\xi_i\| \le \sigma < \tau_M$, and $\sigma \le h$.*

Then every minimizer $\widehat{\Pi}$ in equation 30 satisfies the general-radius bound

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le \frac{C}{\sqrt{\mu}} \frac{\varepsilon(h,\sigma)}{h}, \tag{40}$$

where C is a numerical constant. For any fixed finite graph radius $H > 0$, uniformly over $0 < h \le H$, this implies

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_H \left(h^2 + \frac{\sigma}{h}\right). \tag{41}$$

This is a finite-scale robustness statement and permits a fixed positive noise bound σ, provided the local signal-to-noise condition required elsewhere in the analysis is met. In particular, the theorem does not require $\sigma/h \to 0$. If σ/h is merely bounded by a sufficiently small constant, the displayed bound gives a nonvanishing term in the upper bound rather than an asymptotic consistency claim. The conclusion is local and oracle-localized: item 3 is an explicit feasibility and initialization assumption, not a consequence of the other items and not directly verifiable from the observed data alone. The constant C_H may depend on μ, K, L_3, d, m, H, but not on N, h, σ. Thus, for a fixed local instance, N does not shrink the deterministic error radius in equation 41; its role in the subsequent observed-radius results is to make the design-conditioning and initialization events hold with high probability. This theorem should therefore be read as a high-probability design-conditioned local robustness guarantee after those events are verified, rather than as an N-dependent stochastic estimation rate. The fixed radius H may exceed one; it is a supplied geometric scale, not a universal numerical cutoff.

Remark (Scale and dimensionless local conditions). The general bound retains the units carried by K, L_3, h, σ. Under a dilation $x \mapsto ax$, these quantities become $K/a, L_3/a^2, ah, a\sigma$; accordingly $\mathcal{E}$ is multiplied by a, while $\mathcal{E}/h$ and projector error are unchanged. If $Kh \le c$, $L_3 h^2 \le c$, and $\sigma \le h$, the general bound also gives

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le \frac{C_c}{\sqrt{\mu}} \left\{(L_3 + K^2)h^2 + \frac{\sigma}{h}\right\}.$$

Indeed $D_h/h = Kh/4 + L_3 h^2/12$ is bounded by a constant depending only on c. The terms KD_h and KD_h^2/h are then bounded by $C_c(K^2 + L_3)h^2$, and the noise terms divided by h by $C_c\sigma/h$. These dimensionless sufficient conditions express the relative size of curvature and radius; they do not select a preferred measurement unit. They are an optional simplification, not additional premises of equation 40.

Remark. Conditional mean-zero noise is deliberately absent from this deterministic statement. The residual comparison uses the rowwise normality and norm bounds; the reach condition identifies x_0^* with $\pi_M(x_0)$. The design and PCA concentration arguments require the stated conditional sampling assumptions, but not mean-zero noise after selection. Mean-zero is part of the adopted statistical model, not a premise of this comparison.

Proof. Write $\widehat{\Pi} = \Pi(\widehat{A})$. By item 2, $Q^* \in \mathfrak{Q}_K$, and the intercept $c^* = -a_0$ is unrestricted. Together with the feasibility of Π^* in item 3, the oracle triple (P^*, c^*, Q^*) is feasible. The definition of the estimator therefore gives

$$L_N(\widehat{P}, \widehat{c}, \widehat{Q}) \le L_N(P^*, c^*, Q^*).$$

Proposition C.13 implies

$$\sqrt{L_N(P^*, c^*, Q^*)} \le C\varepsilon(h,\sigma). \tag{42}$$

We next lower-bound the left-hand side. Lemma C.36 uses the graph-frame representative $P(\widehat{A})$ of $\widehat{\Pi}$. By Theorem C.14, we may replace $(\widehat{P}, \widehat{c}, \widehat{Q})$ by the corresponding block-rotated representative without changing the loss or the curvature constraint. Applying the lemma rowwise, the vector of residuals is the sum of

$$-B_m(\widehat{A})\widehat{A}t_i - \widehat{a} - \mathcal{Q}_{\widehat{H}}(t_i)$$

and an error whose empirical root-mean-square norm is at most $C\varepsilon(h,\sigma)$. The reverse triangle inequality and Lemma C.24 give

$$\sqrt{L_N(\widehat{P}, \widehat{c}, \widehat{Q})} \ge \sqrt{\mu} h \left\|B_m(\widehat{A})\widehat{A}\right\|_F - C\varepsilon(h,\sigma). \tag{43}$$

Because $\left\|\widehat{A}\right\|_{\mathrm{op}} \le 1/2$, the smallest eigenvalue of $B_m(\widehat{A})$ is at least $2/\sqrt{5}$. Thus

$$\left\|B_m(\widehat{A})\widehat{A}\right\|_F \ge \frac{2}{\sqrt{5}}\left\|\widehat{A}\right\|_F.$$

Combining equation 42 and equation 43 yields

$$\sqrt{\mu}\,h\left\|\widehat{A}\right\|_F \le C\varepsilon(h,\sigma).$$

Using Lemma C.35 gives equation 40. To obtain the simplified rate on a fixed graph ball, put $a_H = K/4 + L_3 H/12$. For $h \le H$, $D_h \le a_H h^2$, so

$$R_h + KhD_h + \frac{K}{2}D_h^2 \le \left(\frac{L_3}{6} + Ka_H + \frac{K}{2}a_H^2 H\right)h^3.$$

Since $\sigma \le h \le H$, the remaining terms are at most

$$\left[2 + 2K(H + a_H H^2) + 2KH\right]\sigma.$$

Dividing by h proves equation 41, with the stated dependence on H. No step requires $h \le 1$. □

Theorem C.38 (Regularity-dependent noisy-anchor tangent bound)**.** *Retain items 1, 3, and 4 of Theorem C.37 for the same fixed noisy-anchor instance, and suppose $Q^* \in \mathfrak{Q}_K$. Under the C^2 remainder bound equation 17, put $R_h = h^2\omega_2(h)/2$. Then*

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le \frac{C}{\sqrt{\mu}}\frac{\mathcal{E}(h,\sigma;R_h)}{h}.$$

On any fixed graph ball of finite radius $H > 0$, with $\omega_2(H) < \infty$, uniformly for $0 < h \le H$,

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_H\left\{h^2 + h\omega_2(h) + \frac{\sigma}{h}\right\}. \tag{44}$$

Here C_H may depend on $\mu, K, d, m, H, \omega_2(H)$, but not on N, h, σ. If equation 14 holds for $0 < \alpha \le 1$, then

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_{\alpha,H}\left(h^{1+\alpha} + \frac{\sigma}{h}\right), \tag{45}$$

where the constant may additionally depend on $L_{2,\alpha}$. For a fixed C^2 graph, the geometric term is $o(h)$ as $h \to 0$, but a bare C^2 model class supplies no uniform polynomial rate. For $C^{2,1}$, or quantitative C^3 with bounded third derivative, the rate is $C_H(h^2+\sigma/h)$. The bounds are informative when their right-hand sides are small; the chart containment required in item 3 is a separate premise.

Proof. Proposition C.9 gives $\|R(t)\| \le h^2\omega_2(h)/2$. Lemma C.36 applies with this R_h; the oracle residual is also bounded by $\mathcal{E}(h,\sigma;R_h)$. The same residualized-linear comparison therefore proves the first bound. For $h \le H$, let $a_H = (K + \omega_2(H))/4$. Then $D_h \le a_H h^2$, and

$$R_h + KhD_h + \frac{K}{2}D_h^2 \le \frac{1}{2}h^2\omega_2(h) + \left(Ka_H + \frac{K}{2}a_H^2 H\right)h^3.$$

The noisy terms are at most $[2 + 2K(H + a_H H^2) + 2KH]\sigma$. Divide by h to obtain equation 44. Under the Hölder condition, $h\omega_2(h) \le L_{2,\alpha}h^{1+\alpha}$, and

$$h^2 \le H^{1-\alpha}h^{1+\alpha} \qquad (0 < h \le H).$$

This gives equation 45 without requiring $h \le 1$. Bounded third derivative implies the Lipschitz-Hessian condition by integration along segments. □

Corollary C.39 (Residual certificate and approximate minimization)**.** *Assume the geometric, curvature, separation and localization premises of Theorem C.38, using any valid uniform remainder bound R_h. Write $E_h = \mathcal{E}(h,\sigma;R_h)$. Every feasible triple (P, c, Q), not necessarily stationary or minimizing, satisfies*

$$\|\Pi(P) - \Pi^*\|_F \le \frac{C}{\sqrt{\mu}\,h}\{\sqrt{L_N(P,c,Q)} + E_h\}. \tag{46}$$

In particular, if a feasible numerical output satisfies

$$L_N(\widetilde{P},\widetilde{c},\widetilde{Q}) \le \inf_{\Pi(P)\in\mathcal{C}_{\tau,h},\ c,\ Q\in\mathfrak{Q}_K} L_N(P,c,Q) + \Delta_{\mathrm{opt}},$$

then

$$\|\widetilde{\Pi} - \Pi^*\|_F \le \frac{C}{\sqrt{\mu}\,h}\{E_h + \sqrt{\Delta_{\mathrm{opt}}}\}. \tag{47}$$

Neither statement assumes positive profiled Hessian or convergence to a specified minimizer. The loss in equation 46 is observable, but its geometric constants and localization premise are model-based. A small gradient alone does not certify Δ_{opt}.

Proof. Block invariance allows the graph-frame representative $P(A)$. Lemma C.36, separation, and the reverse triangle inequality give

$$\sqrt{L_N(P,c,Q)} \ge \frac{2\sqrt{\mu}}{\sqrt{5}}h\|A\|_F - E_h.$$

Combine this with $\|\Pi(A)-\Pi^*\|_F \le \sqrt{2}\|A\|_F$. The oracle is feasible and its root loss is at most E_h, so an approximate minimizer has root loss at most $E_h + \sqrt{\Delta_{\mathrm{opt}}}$. This proves both claims, including inexact inner solves whenever their reported coefficients are feasible and their total objective gap is bounded. □

Theorem C.40 (Tilt-sensitive recovery and exact quadratic reproduction)**.** *Under the same premises as Corollary C.39, define*

$$G_h = Kh^2/2 + R_h, \qquad c_h = \frac{2\sqrt{\mu}}{\sqrt{5}}h - KhG_h - \frac{K}{4}G_h^2,$$

$$O_h = R_h + \sigma + K(h\sigma + \sigma^2/2), \qquad V_h = 2\sigma + 2K(h + G_h/2)\sigma + 2K\sigma^2.$$

If $c_h > 0$*, every feasible triple satisfies*

$$\|\Pi(P) - \Pi^*\|_F \le \frac{\sqrt{2}}{c_h}\{\sqrt{L_N(P,c,Q)} + R_h + V_h\}.$$

For a feasible Δ_{opt}*-approximate minimizer, this gives*

$$\|\widetilde{\Pi} - \Pi^*\|_F \le \frac{\sqrt{2}}{c_h}\{O_h + R_h + V_h + \sqrt{\Delta_{\mathrm{opt}}}\}.$$

In particular, if $\sigma = 0$*,* $R_h = 0$*, and* $c_h > 0$*, every exact localized minimizer recovers* Π^* *exactly.*

Proof. Put $a = \|A\|_{\mathrm{op}} \le 1/2$. Then $a \le \|A\|_F$ and $a^2 \le \|A\|_F/2$. Use equation 39 in the same root-loss comparison to obtain

$$\sqrt{L_N(P,c,Q)} \ge c_h\|A\|_F - R_h - V_h.$$

The projector comparison proves the first bound. The oracle residual bound of Proposition C.13, with its Taylor bound replaced by R_h and $K_2 \le K$, gives root loss at most O_h. The approximate-minimization inequality proves the second bound. The last assertion follows by setting all three error sources to zero. □

Remark. The additional condition $c_h > 0$ is explicit, not a consequence of localization alone. When $c_h \ge c\sqrt{\mu}h$, the refined bound has geometric term proportional to R_h/h, without the uniform proof's extra curvature term. For fixed geometry and separation it holds at sufficiently small radius. The original general-radius bounds remain valid without this additional absorption condition.

Corollary C.41 (Conditional noisy-anchor recovery and vanishing-noise consistency)**.** *Suppose that the latent coordinates satisfy the density condition in Definition C.19 and form a local sample of size* N *that is conditionally independent given the anchor variables and the count. All probabilities in this statement are conditional on these quantities. For the default radius rule, this is the selected law of Lemma C.1, with its density verified in Proposition C.18. Suppose that the* C^2 *remainder assumptions and the remaining geometric assumptions of Theorem C.38 hold uniformly on a fixed*

graph ball of radius $H > 0$, with $\omega_2(H) < \infty$. All selected scales satisfy $0 < h \le H$, with no restriction $H \le 1$. Assume

$$N \ge \frac{20}{\kappa_2} \log \frac{r_d}{\delta}, \qquad \frac{\sigma}{h} \le \frac{\kappa_2}{24}, \qquad 0 < \tau \le \frac{\kappa_2}{4}.$$

Let $\mathcal{E}_{\mathrm{init}}$ be the event that

$$\|\Pi_0 - \Pi^*\|_F \le r_0$$

and that the localized feasible set is contained in the chart $\|A\|_{\mathrm{op}} \le 1/2$. Then, on an event of probability at least $1 - \delta - \mathbb{P}(\mathcal{E}_{\mathrm{init}}^c)$,

$$\Pi^* \in \mathcal{C}_{\tau,h}(\Pi_0, r_0)$$

and every localized empirical minimizer satisfies

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_H \left\{h^2 + h\omega_2(h) + \frac{\sigma}{h}\right\}.$$

The preceding finite-sample statement allows a fixed positive σ. Its condition $\sigma/h \le \kappa_2/24$ only requires that the chosen local scale be sufficiently large relative to that fixed noise level; it does not assert that the right-hand side vanishes. The following paragraph is a separate asymptotic specialization for consistency.

The sample-size lower bound in this corollary controls the probability of the conditioning event (and, through $\mathcal{E}_{\mathrm{init}}$, of successful local initialization). Conditional on those events, the displayed recovery radius is the deterministic C^2 radius of Theorem C.38 and has no additional $N^{-1/2}$-type term. Consequently, at fixed h and fixed σ, this result does not by itself predict that increasing the number of selected observations decreases tangent error; it predicts a smaller failure probability for the stated local guarantee.

For a triangular array of local problems, take fixed d, m, K, H, a common bounded Hessian modulus on the supplied graph balls, and $\inf_N \kappa_{2,N} > 0$. These uniformity conditions hold, for example, for a fixed graph and uniform density constants. Choose $\delta_N \downarrow 0$ such that

$$\frac{N\kappa_{2,N}}{20} - \log \frac{r_d}{\delta_N} \longrightarrow \infty.$$

If, in addition,

$$h_N \longrightarrow 0, \qquad \frac{\sigma_N}{h_N} \longrightarrow 0, \qquad \mathbb{P}(\mathcal{E}_{\mathrm{init}}^c) \longrightarrow 0,$$

then $\left\|\widehat{\Pi} - \Pi^*\right\|_F \to 0$ in probability. If design or geometry constants vary instead, replace these simplified uniformity requirements by the explicit condition

$$\frac{\mathcal{E}_N(h_N, \sigma_N; R_{h_N})}{h_N \sqrt{\kappa_{2,N}}} \longrightarrow 0,$$

together with vanishing design and initialization failure probabilities. For approximate minimizers also require $\sqrt{\Delta_{\mathrm{opt},N}}/(h_N\sqrt{\kappa_{2,N}}) \to 0$.

If, in addition, the Hölder-Hessian condition equation 14 holds for some $0 < \alpha \le 1$, then the finite-sample bound above sharpens on the same event to

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_{\alpha,H} \left(h^{1+\alpha} + \frac{\sigma}{h}\right).$$

Thus the Hölder condition is used to express the geometric term by the explicit power $h^{1+\alpha}$, rather than to obtain the bare C^2 consistency conclusion.

Proof. The Rayleigh-quotient form of Lemma C.22 gives item 1 of Theorem C.38, with $\mu = \kappa_2/2$, with probability at least $1 - \delta$. Its constant-plus-quadratic principal submatrix has minimum eigenvalue at least $\kappa_2/2$. Since the constant coordinate is present, $\kappa_2 \le 1$; hence $\sigma/h \le \kappa_2/24$ implies $\sigma \le h$, as required by Lemma C.26. That lemma and $\sigma/h \le \kappa_2/24$ imply

$$\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P^*)) \ge \frac{\kappa_2}{4} \ge \tau.$$

Together with $\mathcal{E}_{\mathrm{init}}$, this proves feasibility of Π^* and the local-chart condition. Theorem C.38 then gives the stated C^2 bound. The uniform geometry bounds keep C_H bounded, and the common bounded modulus gives $h_N\omega_{2,N}(h_N) \to 0$. For a fixed C^2 graph, continuity additionally gives $\omega_2(h_N) \to 0$; the consistency conclusion follows by letting the displayed deterministic bound and the two failure probabilities tend to zero. Under the additional Hölder condition, the same theorem gives the stated polynomial-rate refinement. □

Corollary C.42 (Observed-radius recovery with certified PCA initialization)**.** *The preceding corollary is a conditional local-sample statement. We now specialize it to the default observed-radius procedure, using the selected density, design, and PCA results of Section C.4. Fix a finite valid graph radius $H > 0$, which need not be at most one. Assume the i.i.d. normal-noise sampling, chart coverage, and density conditions of Proposition C.18, together with $Q^* \in \mathfrak{Q}_K$ and $\|R(t)\| \le L_3\|t\|^3/6$ on the supplied graph ball. Design and localization are conclusions below, not assumptions imported from Theorem C.37. Set $h = h_{\mathrm{out}} = r + 2\sigma$ and $\kappa_{2,r} = \rho_r\alpha_r^{d+4}\gamma_d^{(2)}$. Let $\delta_{\mathrm{des}}, \delta_{\mathrm{pca}} \in (0,1)$, set $\lambda_{\mathrm{lin},r} = \rho_r\alpha_r^{d+2}/(d+2)$, and define*

$$\gamma = \frac{\lambda_{\mathrm{lin},r}h^2}{2}, \qquad b_{\mathrm{pop}} = C\frac{\sqrt{d}\,\varepsilon_{\mathrm{pop}}}{\lambda_{\mathrm{lin},r}h^2},$$

where $\varepsilon_{\mathrm{pop}}$ is defined in Proposition C.33. Choose a deterministic lower count threshold satisfying

$$q_0 \ge \max\left\{\frac{20}{\kappa_{2,r}}\log\frac{r_d}{\delta_{\mathrm{des}}}, \quad C\frac{r^4}{\gamma^2}\log\frac{2}{\delta_{\mathrm{pca}}}\right\}.$$

Assume

$$0 < h \le H, \qquad q_0 \le \frac{n-1}{2}p_-(r,\sigma), \qquad \frac{\sigma}{h} \le \frac{\kappa_{2,r}}{24}, \qquad 0 < \tau < \frac{\kappa_{2,r}}{4},$$

that

$$\varepsilon_{\mathrm{pop}} \le \frac{\lambda_{\mathrm{lin},r}h^2}{4},$$

and that

$$b_{\mathrm{pop}} + C\frac{\sqrt{d}\,r^2}{\gamma}\sqrt{\frac{\log(2/\delta_{\mathrm{pca}})}{q_0}} \le r_{\mathrm{init}} \le r_0, \qquad r_{\mathrm{init}} + r_0 < \frac{1}{\sqrt{5}}.$$

Finally, assume that a fixed margin $\nu > 0$ satisfies

$$\sqrt{5}\,r_{\mathrm{init}} + \tau + \nu \le \frac{\kappa_{2,r}}{4}.$$

The lower count threshold q_0 may be increased to meet the displayed PCA error bound, subject to the count condition above. Then, conditional on the anchor variables, for the anchor-centered PCA initializer and the observed-radius neighborhood, with conditional probability at least

$$1 - \exp\left\{-\frac{n-1}{8}p_-(r,\sigma)\right\} - \delta_{\mathrm{des}} - \delta_{\mathrm{pca}},$$

both Π^* and Π_0 belong to $\mathcal{C}_{\tau,h}(\Pi_0, r_0)$, the PCA initializer has observed Gram eigenvalue at least $\tau + \nu$, and every localized minimizer obeys

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \le C_H\left(h^2 + \frac{\sigma}{h}\right).$$

The constant may depend on H as specified in Theorem C.37. Without simplifying the radius dependence, the same event gives equation 40. This certifies the statistical localized problem and its feasible starting point, not convergence of a specified numerical solver to its minimum.

Proof. The count bound in Proposition C.18 gives $N_r \ge q_0$ outside the first failure event. Conditional on this event, condition on $N_r = q \ge q_0$ and apply Corollary C.30 to verify item 1 of Theorem C.37 and feasibility of Π^*. Proposition C.33 supplies the spectral gap and population-alignment bounds required by Proposition C.32. The latter then gives $\|\Pi_0 - \Pi^*\|_F \le r_{\mathrm{init}}$ outside the second failure event. On the intersection of these same events, Corollary C.34 and the margin

assumption give $\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0)) \geq \tau + \nu$; no additional failure probability is incurred. Hence Π^* lies in the trust region because $r_{\text{init}} \leq r_0$. For every Π in that region,

$$\|\Pi - \Pi^*\|_{\text{op}} \leq \|\Pi - \Pi_0\|_F + \|\Pi_0 - \Pi^*\|_F \leq r_0 + r_{\text{init}} < \frac{1}{\sqrt{5}}.$$

Thus the associated graph coordinate satisfies $\|A\|_{\text{op}} < 1/2$, so item 3 also holds. Theorem C.37 completes the proof. The conditioning is essential: observed-radius selection changes the non-anchor noise law, so its conditional mean need not remain zero after selection. □

C.5.4 Comparison with the Aamari–Levrard benchmark

The local form $h^2 \vee \sigma/h$ agrees with the quadratic benchmark of Aamari & Levrard (2019, Theorem 2); our high-probability statements require the estimator-specific selection, design, and initialization events proved above. The next corollaries give conditional bandwidth consequences.

Corollary C.43 (Bandwidth-optimized $C^{2,\alpha}$ deterministic robustness)**.** *Suppose local-radius samples satisfy $N_h \asymp nh^d$ on a count event of probability at least $1 - \delta_{\text{count}}$, over an admissible range of radii with fixed geometry and $\kappa_{2,h} \geq \underline{\kappa} > 0$. Assume the design-transfer and initialization conditions hold at the chosen radius, with initialization failure probability δ_{init}. For a target design failure probability δ, define*

$$h_{\text{samp}} = \left(\frac{C \log(r_d/\delta)}{n}\right)^{1/d}.$$

The constant C includes $\underline{\kappa}^{-1}$ and the lower sample-count constant. Subject to the chosen radius remaining in the admissible range of Corollary C.41, choose

$$h_n \asymp \max\left\{\sigma_n^{1/(2+\alpha)}, h_{\text{samp}}\right\}.$$

Then, with probability at least $1 - \delta_{\text{count}} - \delta - \delta_{\text{init}}$,

$$\left\|\widehat{\Pi} - \Pi^*\right\|_F \leq C \max\left\{\sigma_n^{(1+\alpha)/(2+\alpha)}, h_{\text{samp}}^{1+\alpha}\right\}.$$

For $\alpha = 1$, corresponding to $C^{2,1}$ or quantitative C^3 regularity, this becomes

$$h_n \asymp \max\{\sigma_n^{1/3}, h_{\text{samp}}\}, \qquad \left\|\widehat{\Pi} - \Pi^*\right\|_F \leq C \max\{\sigma_n^{2/3}, h_{\text{samp}}^2\}.$$

Thus fixed sufficiently small $\sigma > 0$ gives the familiar $O(\sigma^{2/3})$ upper-bound plateau in the Lipschitz-Hessian case.

Proof. The sample-size condition requires $h \gtrsim h_{\text{samp}}$. Without this constraint, balancing $h^{1+\alpha}$ and σ/h gives $h \asymp \sigma^{1/(2+\alpha)}$ and value $O\{\sigma^{(1+\alpha)/(2+\alpha)}\}$. Taking the larger of the noise-optimal and sample-feasible scales gives the result. □

This is an optimized upper-bound statement, not a minimax lower bound and not a proof that the actual error must remain above the displayed noise-dependent plateau.

Corollary C.44 (Clean designated-anchor rate)**.** *Consider i.i.d. clean observations and a designated anchor index. Suppose the chart and density assumptions of Proposition C.18 hold uniformly at that anchor, $Q^* \in \mathfrak{Q}_K$, and equation 14 holds with fixed constants and $0 < \alpha \leq 1$. Take $0 < \tau \leq \kappa_{2,0}/2$, and let the localization event be $\mathcal{E}_{\text{init}}$ of Corollary C.41. For any fixed $a > 0$, choose the deterministic observed radius*

$$h_n = r_n = C_a(\log n/n)^{1/d},$$

with C_a large enough for count and design concentration. If the localization event has conditional failure probability $\delta_{\text{init},n}$, then, for all sufficiently large n, with conditional probability at least $1 - 2n^{-a} - \delta_{\text{init},n}$,

$$\|\widehat{\Pi} - \Pi^*\|_F \leq C(\log n/n)^{(1+\alpha)/d}.$$

For $\alpha = 1$ this is $C(\log n/n)^{2/d}$.

Proof. The inner-ball count bound gives $N \geq cnh_n^d$ except with probability n^{-a} after increasing C_a. At $\sigma = 0$, the density constants are independent of n. Conditional on each sufficiently large count, Lemma C.22 with $\delta = n^{-a}$ gives separation and observed feasibility. Intersect with the localization event and apply Theorem C.38. □

Remark. This statement uses the same observed-radius rule, not an oracle neighbor set. Localization failure remains explicit. The PCA certificate can verify it when its curvature, count and margin conditions hold (with a suitable remainder bound); the rate alone does not establish those conditions.

C.6 Why bounded mean-zero tubular noise is not enough

This section concerns population failure of the present uncorrected local criterion. Here "uncorrected" means that the least-squares criterion does not use a separate noise-variance, deconvolution, or latent-variable correction; it does *not* mean that the noisy-anchor intercept is omitted. The point is not a failure of concentration. Even if the non-anchor observations are conditionally i.i.d., their empirical loss converges to a population loss that can itself prefer an incorrect orientation.

This is not a minimax impossibility theorem for the global class $\mathcal{P}^3(\sigma)$ in Definition C.5. The two claims use different levels of global geometric and sampling structure. It is also not a radius-selection counterexample: conditioning on an observed-radius event can change the joint law of the latent coordinate and noise. Rather, the example isolates the local signal-to-noise obstruction already present in the anchored regression criterion. The comparison concerns the unrestricted population criterion, or any admissible set containing both orientations in the example. It does not by itself contradict a recovery theorem that assumes prior localization around the true tangent; rather, it explains why that localization, or another population-identifiability mechanism, cannot follow from bounded mean-zero noise alone.

Example C.45 (A wrong orientation can have smaller population loss)**.** *Let $d = m = 1$, let the clean manifold be the x-axis, and fix a latent anchor $x_0^* = (0, 0)$. Let*

$$x_0 = (0, \varepsilon_0), \qquad X = (T, \varepsilon), \qquad U = X - x_0 = (T, \varepsilon - \varepsilon_0),$$

where ε_0, T, and ε are independent, T and ε are centered continuous random variables supported on $[-h, h]$ and $[-\sigma, \sigma]$, and ε_0 has the same bounded centered normal-noise law as ε. Thus every observation, including the anchor, obeys the bounded mean-zero tubular-noise model. We condition on the realized noisy anchor ε_0.

At the true tangent, the predictor is T and the normal response is $\varepsilon - \varepsilon_0$. Since ε is independent of T, the best constant-plus-quadratic prediction has intercept $-\varepsilon_0$ and zero quadratic coefficient. Its conditional population loss is

$$\mathrm{Var}(\varepsilon).$$

At the orthogonal, incorrect tangent, the predictor is $\varepsilon - \varepsilon_0$ and the normal response is T. Conditional on ε_0, these variables are independent and centered in the response, so again the best constant-plus-quadratic prediction is zero and the loss is

$$\mathrm{Var}(T).$$

If $\mathrm{Var}(\varepsilon) > \mathrm{Var}(T)$, the incorrect tangent has smaller population loss. Both constant-plus-quadratic designs are nondegenerate when the two variables have nondegenerate continuous laws.

C.7 Comparison with local linear fitting

Proposition C.12 identifies the affine restriction of our criterion with mean-centered local PCA when orientation constraints are removed. The following comparison uses the same feasible basin as the quadratic theorem, so it does not assume that unrestricted PCA enters it.

Corollary C.46 (Affine and quadratic bounds in the same local basin)**.** *Under the assumptions of Theorem C.37, define*

$$\widehat{\Pi}_{\mathrm{aff}} \in \operatorname*{arg\,min}_{\Pi(P) \in \mathcal{C}_{\tau,h}(\Pi_0, r_0)} \min_{c} L_N(P, c, 0).$$

Every affine minimizer satisfies

$$\left\| \widehat{\Pi}_{\mathrm{aff}} - \Pi^* \right\|_F \leq \frac{C}{\sqrt{\mu}} \left(Kh + L_3 h^2 + \frac{\sigma}{h} \right).$$

On a fixed valid graph ball $0 < h \le H < \infty$, this is $C_H(h + \sigma/h)$, whereas the quadratic bound is $C_H(h^2+\sigma/h)$. The constants may differ and depend on H. These are upper bounds, not a pointwise dominance claim or a lower bound for PCA.

Proof. Continuity and compactness give an affine minimizer. At P^*, the choice $c = -a_0, Q = 0$ leaves residual $g(t_i) + \zeta_i$, bounded by $Kh^2/2 + L_3h^3/6 + \sigma$. For a tilted affine candidate, collect the quadratic part of $B_m(A)g(t)$ with the freely adjustable quadratic polynomial in the separation argument. Its remainder is bounded by $L_3h^3/6$; the noisy anchored response perturbation is at most 2σ. There is no fitted quadratic predictor perturbation since $Q = 0$. Thus its root loss is at least

$$\frac{2\sqrt{\mu}}{\sqrt{5}} h \|A\|_F - L_3h^3/6 - 2\sigma.$$

Compare with the affine oracle and divide by h, then apply Lemma C.35. For $h \le H$, use $h^2 \le Hh$ to obtain the simplified affine rate. □

The natural linear baseline fits only an intercept in the rotated normal coordinates. Equivalently, it chooses a local affine subspace. Its leading orientation error comes from treating quadratic curvature as an unmodeled response.

Remark. The improvement in geometric bias is purchased by estimating $q_d = d(d+1)/2$ quadratic coefficients per normal component. The sample-size condition must therefore reflect $r_d = 1+d+q_d$, not merely d. This is the relevant bias–variance tradeoff when comparing the methods.

Remark (Regularity-dependent comparison). Under $C^{2,\alpha}$ regularity, the same calculation replaces the quadratic fit's $O(h^2)$ bias by $O(h^{1+\alpha})$, while the generic affine bias remains $O(h)$. Thus every $\alpha > 0$ yields an asymptotic order improvement over the generic asymmetric affine fit. At $\alpha = 1$, corresponding to a Lipschitz Hessian, the quadratic bias is $O(h^2)$. Under bare C^2 regularity one obtains only $o(h)$ for a fixed graph, with no uniform polynomial improvement over a C^2 class unless a common Hessian modulus is imposed.

Remark (Asymmetry and the sampling regime). The example $T = hZ$ with a fixed asymmetric law for Z is an allowed local-design example, not a universal description of shrinking neighborhoods under a fixed smooth manifold density. At an interior anchor, leading-order symmetry can suppress third moments and improve the affine bias order. Experiments should distinguish bounded asymmetric local laws from fixed smooth densities with shrinking observed-radius neighborhoods. Comparing two upper bounds does not establish a matching lower bound for PCA.

C.8 Minimax benchmark under tubular noise

The deterministic bound is estimator-specific, whereas a minimax lower bound is a property of the statistical model. We therefore use the quantitative class in Definition C.5 and translate the tangent-space lower bound of Aamari and Levrard Aamari & Levrard (2019) to the projector loss used in this paper.

Corollary C.47 (Aamari–Levrard lower bound in projector loss)**.** *Suppose the constants defining the quantitative C^3 manifold class satisfy the conditions of the tangent-space lower bound in Aamari and Levrard (Aamari & Levrard, 2019, Theorem 3). Then*

$$\inf_{\widetilde{\Pi}} \sup_{P \in \mathcal{P}^3(\sigma)} \mathbb{E}_{P^{\otimes n}} \left\| \widetilde{\Pi} - \Pi_{T_{\pi_M(X_1)}M} \right\|_F^2 \ge c \left\{ (n-1)^{-4/d} \vee \left(\frac{\sigma}{n-1} \right)^{4/(d+3)} \right\}.$$

The same order holds for the fixed-point version of their model.

Here the infimum is over measurable estimators taking values in the set of rank-d orthogonal projectors. The target projector is evaluated at the latent projection $\pi_M(X_1)$, not at the generally off-manifold observed anchor X_1.

Proof. Aamari and Levrard prove the root tangent lower bound

$$\inf_{\widetilde{T}} \sup_{P \in \mathcal{P}^3(\sigma)} \mathbb{E} \angle(T_{\pi_M(X_1)}M, \widetilde{T}) \ge c \left\{ (n-1)^{-2/d} \vee \left(\frac{\sigma}{n-1} \right)^{2/(d+3)} \right\}.$$

The largest-principal-angle loss is bounded above and below, up to universal constants, by the operator norm of the difference of the corresponding orthogonal projectors. Moreover,

$$\left\|\widetilde{\Pi} - \Pi^*\right\|_F^2 \geq \left\|\widetilde{\Pi} - \Pi^*\right\|_{\mathrm{op}}^2.$$

Jensen's inequality therefore squares the cited lower bound and gives the displayed result. □

For $k = 3$, the Aamari–Levrard upper bound has the root-error form $h^2 \vee \sigma/h$ at $h \asymp (\log n/n)^{1/d}$, whereas Corollary C.47 has root lower bound

$$n^{-2/d} \vee \left(\frac{\sigma}{n}\right)^{2/(d+3)}.$$

Thus the cited local-polynomial estimator is near-minimax, up to logarithmic factors, in the clean-dominated regime $\sigma \lesssim n^{-3/d}$. The present estimator has a matching deterministic local bias–noise form, but no end-to-end minimax claim is made without the additional estimator-specific global analysis described above.

C.9 Scope of the recovery guarantees

The local bounds have geometric term $h\omega_2(h)$ under C^2 regularity, $h^{1+\alpha}$ under $C^{2,\alpha}$, and h^2 under anchorwise Lipschitz Hessian control. The robust noise term is σ/h. At fixed h, σ, increasing the selected count improves the design and initialization events; the deterministic error radius does not acquire an $N^{-1/2}$ factor. Vanishing-error consistency requires the corresponding geometric term and σ/h to vanish. Fixed positive noise and shrinking radius are not covered by that conclusion.

The observed-radius theorem uses a fixed anchor, fixed radius, supplied K, τ, r_0, chart coverage, and explicit signal and initialization margins. The observed Gram test certifies initializer feasibility, not inclusion of Π^*. KNN selection, adaptive anchors or radii, and tuning on the same sample require additional analysis. The cited minimax benchmark is not a minimax theorem for this estimator.

PCA localization and Gram feasibility do not establish entry into a positive-Hessian basin. Corollary C.39 bounds feasible outputs using their residual or a certified objective gap; gradient tolerance alone does not certify that gap. The local gradient certificate below needs its own strict-margin and perturbation assumptions. Neither result proves unrestricted global recovery or convergence of L-BFGS from every feasible PCA initialization.

C.10 Optimization formulas

We give the profile derivatives used by the numerical formulation.

Let

$$J_d = \begin{pmatrix} I_d \\ 0_{m\times d} \end{pmatrix}, \qquad J_m = \begin{pmatrix} 0_{d\times m} \\ I_m \end{pmatrix}.$$

Then $P_1 = PJ_d$, $P_2 = PJ_m$,

$$s_i = J_d^\top P^\top u_i, \qquad n_i = J_m^\top P^\top u_i.$$

For a residual

$$r_i = n_i - c - \mathcal{Q}_Q(s_i),$$

define

$$G_i = [Q_1 s_i \ \cdots \ Q_m s_i] \in \mathbb{R}^{d\times m}, \qquad A_i = J_m - J_d G_i \in \mathbb{R}^{D\times m}.$$

Proposition C.48 (Fixed-coefficient gradient)**.** *For fixed c, Q,*

$$\nabla_P L_N(P, c, Q) = \frac{2}{N}\sum_{i=1}^N u_i r_i^\top A_i^\top.$$

The Riemannian gradient on $O(D)$, under the Frobenius metric, is

$$\mathrm{grad}_{O(D)} L_N = P\,\mathrm{Skew}\big(P^\top \nabla_P L_N\big), \qquad \mathrm{Skew}(B) = \frac{1}{2}(B - B^\top).$$

Proof. For an ambient perturbation V, $Ds_i[V] = J_d^\top V^\top u_i$, $Dn_i[V] = J_m^\top V^\top u_i$, and symmetry of Q_k gives $Dr_i[V] = A_i^\top V^\top u_i$. Therefore $DL_N[V] = \langle (2/N)\sum_i u_i r_i^\top A_i^\top, V\rangle_F$. Projection onto $T_P O(D)$ gives $P\,\mathrm{Skew}(P^\top \nabla_P L_N)$, the standard embedded-manifold gradient construction (Boumal, 2023, Chapter 3). □

C.10.1 Envelope differentiability and the profiled Hessian

Let $\beta = (c, Q)$ collect coefficients in fixed bases and let $P(A)$ be a smooth local frame. Write $\ell(A, \beta) = L_N(P(A), \beta)$.

Proposition C.49 (Envelope gradient and Schur-complement Hessian)**.** *On an open full-rank region, the unique constrained coefficient solution $\widehat{\beta}(A) \in \mathbb{R}^m \times \mathfrak{Q}_K$ is continuous and $F_{N,K}(A) = \ell(A, \widehat{\beta}(A))$ is continuously differentiable, with*

$$DF_{N,K}(A)[V] = D_A \ell(A, \widehat{\beta}(A))[V].$$

This first-derivative identity remains valid at transitions of the inner curvature active set. If that constraint is inactive throughout a neighborhood, the profiled Hessian in these same local coordinates is

$$D^2 F_{N,K} = \ell_{AA} - \ell_{A\beta}\ell_{\beta\beta}^{-1}\ell_{\beta A}.$$

At an active coefficient constraint, a second-derivative formula requires appropriate differentiability and regularity of its KKT system; smoothness across active-set transitions is not asserted.

Proof. Eliminate the intercept and minimize over the compact curvature ball. Full rank gives a unique optimizer, continuous by the argument of Theorem C.29. The unique-optimizer envelope theorem (Danskin, 1967) then gives $DF_{N,K} = \ell_A$: the coefficient feasible set is fixed, so differentiability of its optimizer across active-set transitions is unnecessary. When the constraint is inactive, the implicit function theorem applied to $\ell_\beta = 0$ yields $D\widehat{\beta} = -\ell_{\beta\beta}^{-1}\ell_{\beta A}$, giving the Schur complement. This is the local smooth coefficient-elimination principle underlying variable projection (Golub & Pereyra, 1973). □

Remark (Coordinate and Riemannian second derivatives). Here ℓ_{AA} differentiates the composed map $L_N(P(A), \beta)$, including the derivatives of the chart. No additional embedding-curvature term should be added to this coordinate Hessian. If one instead starts with ambient derivatives in P and constructs a Riemannian Hessian or a geodesic second derivative, the usual tangent-projection/acceleration terms must be included exactly once. The residualized-linear condition controls local loss separation; it does not by itself prove positivity of the full profiled Hessian in the noisy, curved problem.

C.11 Local curvature from residualized orientation separation

This subsection supplements, rather than replaces, the deterministic risk and residual certificates. It derives local curvature in an exact-quadratic reference model and gives explicit sufficient conditions for its persistence under data perturbations. Neighborhood indices are fixed throughout; no radius-selection indicator is differentiated. All constants involving the true frame or latent coordinates are model-based, not automatically observable.

Reference model and coordinates. Work in the true frame and put

$$u_i^0 = (t_i, g(t_i) - a_0), \qquad g_k(t) = \tfrac{1}{2} t^\top Q_k^* t, \qquad \|t_i\| \le h.$$

Thus the neighbors are exact quadratic samples and the anchor may have a normal displacement a_0. The artificial self-anchor row is omitted. Write $K_2 = (\sum_k \|Q_k^*\|_{\mathrm{op}}^2)^{1/2}$ and $K_* = (\sum_k \|Q_k^*\|_F^2)^{1/2} < K$. Use the true-frame exponential chart

$$P(A) = P^* \exp \Omega(A), \qquad \Omega(A) = \begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix}.$$

This differs from the graph chart away from zero, but has the same first variation there. Let S_{lin} and P_Φ be formed from t_i, as in Proposition C.25. Define normalized matrices and their unconstrained least-squares profile by

$$X(A) = N^{-1/2}\Phi_h(s(A)), \qquad Y(A) = (h\sqrt{N})^{-1}[n_i(A)^\top]_{i=1}^N,$$
$$\beta = (X^\top X)^{-1}X^\top Y, \qquad \mathcal{R} = Y - X\beta.$$

The intercept block of β is c/h, and its quadratic block represents hQ in the orthonormal symmetric bases. Where the curvature constraint is inactive, the original profiled loss is $F(A) = h^2\|\mathcal{R}(A)\|_F^2$.

Proposition C.50 (Exact residualized Hessian and a curvature lower bound)**.** *Suppose the reference true frame has full-rank coefficient design and strict design and trust-region margins, and $K_* < K$. Then its constrained profile is smooth near zero. For $H \in \mathbb{R}^{m\times d}$, set*

$$\mathcal{D}(H)_i = Ht_i + Dg(t_i)H^\top(g(t_i) - a_0).$$

With $\|V\|_N^2 = N^{-1}\sum_i \|V_i\|^2$,

$$D^2F(0)[H,H] = 2\|(I - \mathsf{P}_\Phi)\mathcal{D}(H)\|_N^2. \tag{48}$$

If $S_{\mathrm{lin}} \succeq \mu_{\mathrm{sep}} I$ and

$$\gamma_{\mathrm{loc}} = \sqrt{\mu_{\mathrm{sep}}} - K_2(K_2h^2/2 + \|a_0\|) > 0,$$

then

$$D^2F(0)[H,H] \geq 2h^2\gamma_{\mathrm{loc}}^2\|H\|_F^2.$$

Consequently the true tangent is the unique minimizer in a sufficiently small convex chart ball on which the loss is strongly convex.

Proof. At zero the coefficients $(-a_0, Q^*)$ give zero residual. Full rank and strict coefficient interior imply a smooth unconstrained profile locally, which equals the constrained profile. The coordinate derivatives are $\dot{s}_i = H^\top(g(t_i) - a_0)$ and $\dot{n}_i = -Ht_i$. The fixed-coefficient residual derivative is therefore $-\mathcal{D}(H)$. Differentiating the normal equations at a zero residual projects this derivative onto the orthogonal complement of the coefficient design. The standard zero-residual least-squares identity gives equation 48; combined with coefficient elimination (Golub & Pereyra, 1973), this is the specialization of Proposition C.49. By Proposition C.25, $\|(I - \mathsf{P}_\Phi)Ht\|_N \geq h\sqrt{\mu_{\mathrm{sep}}}\|H\|_F$. Also $\|Dg(t)\|_{\mathrm{op}} \leq K_2 h$, $\|g(t)\| \leq K_2h^2/2$, and orthogonal projection is a contraction. The reverse triangle inequality proves the lower bound. Continuity of the Hessian and strict margins give a small feasible strongly convex ball. Its stationary point at zero is its unique minimizer. □

Remark (Separation is a sufficient condition, not a converse). The exact curvature object is the entire projected map $(I - \mathsf{P}_\Phi)\mathcal{D}$, not just S_{lin}. The predictor-rotation term $Dg(t)H^\top(g(t) - a_0)$ cannot generally be omitted. Singularity of S_{lin} alone does not imply singularity of this Hessian; a singular Hessian alone does not imply nonuniqueness. The condition $\gamma_{\mathrm{loc}} > 0$ is sufficient and may be conservative. Anchor displacement cancels from the oracle residual but enters this stability bound and can also affect the selected design and feasibility.

Remark (A short consequence of the existing tilt bound). When the anchor is also clean, the proof of Theorem C.40, with $R_h = \sigma = 0$, already gives $F(A) \geq c_h^2\|A\|_F^2$ in the graph chart. If the true solution has strict margins and $c_h > 0$, Taylor expansion yields $D^2F(0)[H,H] \geq 2c_h^2\|H\|_F^2$. Proposition C.50 additionally identifies the exact Hessian and treats an anchor-only normal displacement.

Corollary C.51 (Nonnegative predictor-rotation contribution on symmetric graphs)**.** *Retain the full-rank and strict reference margins of Proposition C.50. Suppose the anchor is clean, $g(t) = \kappa\|t\|^2/2$ for $\kappa \in \mathbb{R}^m$, and the empirical coordinate distribution is invariant under $t \mapsto -t$. Then*

$$D^2F(0)[H,H] = 2\|Ht\|_N^2 + \frac{1}{N}\sum_i \left(2\|t_i\|^2 + \frac{\|\kappa\|^2}{2}\|t_i\|^4\right)(\kappa^\top Ht_i)^2$$
$$\geq 2\lambda_{\min}\left(N^{-1}\sum_i t_i t_i^\top\right)\|H\|_F^2.$$

Consequently a positive tangent second moment gives a strictly positive Hessian even if the sufficient margin γ_{loc} is nonpositive. For $d = 1$, writing $M_j = N^{-1}\sum_i t_i^j$, the Hessian matrix is

$$2M_2 I_m + \left(2M_4 + \frac{M_6}{2}\|\kappa\|^2\right)\kappa\kappa^\top.$$

Proof. The full frame derivative is $\mathcal{D}(H)_i = Ht_i + \|t_i\|^2 \kappa\kappa^\top H t_i/2$. Each component is odd, whereas the coefficient regressors are constant or quadratic and hence even. Empirical central symmetry therefore gives $\mathsf{P}_\Phi \mathcal{D}(H) = 0$. Expanding $2\|\mathcal{D}(H)\|_N^2$ proves the equality. The remaining terms are nonnegative, and the Rayleigh bound for the tangent second moment proves the inequality. In one dimension the expansion gives the stated matrix. □

Remark (Scope of the symmetric specialization). Symmetry is required of the realized empirical distribution, not merely the population. The result concerns an exact quadratic graph with a clean anchor and strict feasible margins. It proves local Hessian positivity, not global identifiability or basin entry. For $m > 1$, directions satisfying $\kappa^\top H = 0$ receive no extra curvature; the condition number can therefore increase. These qualifications also apply to the corresponding mechanism experiment.

C.11.1 Anisotropic cancellation and stability under asymmetry

We now show that positivity of the coefficient and residualized linear designs alone does not imply positive local orientation curvature. This concerns the exact-reference profile, not the sufficient small-curvature hypotheses in the recovery or basin results.

Proposition C.52 (Nondegenerate designs with vanishing orientation curvature)**.** *Let $d = 2, m = 1$, the anchor be clean, and $g(x, y) = (x^2 - y^2)/2$. For $0 \le \delta \le 1$, take the twelve latent points*

$$T_\delta = \{(\pm 1, \pm\sqrt{3-\delta}), (\pm 2, \pm\sqrt{6-\delta}), (0, \pm 1), (0, \pm 2)\},$$

with all four sign combinations in each of the first two groups. Use $h = 1$ as feature normalization and $\phi = (1, x^2/2, xy/\sqrt{2}, y^2/2)^\top$. Let $\Sigma_\delta = \Phi^\top\Phi/12$. The matrix Σ_δ is positive definite throughout $[0, 1]$, with a uniformly positive minimum eigenvalue, and

$$S_{\mathrm{lin}} = \mathrm{diag}\left(\frac{5}{3}, \frac{23}{6} - \frac{2\delta}{3}\right).$$

For the intercept-bearing unconstrained profile in the true-centered exponential chart, the Hessian is

$$\mathcal{H}_\delta = \mathrm{diag}\left(\frac{5}{6}\delta^2, 2(2 - \delta/2)^2(3 - 2\delta/3) + \frac{51}{4}\right).$$

In particular, $\lambda_{\min}(\mathcal{H}_\delta) = 5\delta^2/6$, whereas $2e_1^\top S_{\mathrm{lin}} e_1 = 10/3$. The same Hessian holds locally for the constrained profile with $K = 2$, any $0 < \tau < \lambda_{\min}(\Sigma_\delta)$, and a positive trust radius centered at the true projector.

Proof. A linear combination of the four features that vanishes on the sample must have zero mixed coefficient, by comparing opposite signs of xy. The points $(0, \pm 1), (0, \pm 2)$ then force the constant and y^2 coefficients to vanish; the remaining points force the x^2 coefficient to vanish. Thus Φ has full column rank. Continuity in δ and compactness of $[0, 1]$ give the uniform lower bound. Central symmetry makes T orthogonal to the even features, so $S_{\mathrm{lin}} = T^\top T/12$, with the displayed moments.

Write $q = x^2 - y^2$ and $v = (x(1 + q/2), y(1 - q/2))^\top$. The full derivative is $\mathcal{D}(H)_i = Hv_i$. It is odd, hence orthogonal to every column of Φ. Proposition C.50 therefore gives $\mathcal{H} = 2\sum_i v_i v_i^\top/12$. On the first eight points $q = -2 + \delta$, so $v_1 = \delta x/2$; on the last four, $v_1 = 0$. Consequently $\mathcal{H}_{11} = 5\delta^2/6$. Independent coordinate sign symmetry gives $\mathcal{H}_{12} = 0$. The first eight points contribute $2(2 - \delta/2)^2(3 - 2\delta/3)$ to $\mathcal{H}_{22}$; the axis points contribute $51/4$. The latter eigenvalue is larger throughout the interval. Finally $\|Q\|_F = \sqrt{2} < 2$, the reference Gram and trust margins are strict, and the unique unconstrained coefficient fit varies continuously. It remains feasible in a neighborhood, where the two profiles coincide. □

Corollary C.53 (Weak curvature persists under arbitrary small perturbations)**.** *Let t_i^0 be the points of Proposition C.52 at $\delta = 0$. There exists $\rho_0 \in (0, 1/2]$ such that for every $t_i = t_i^0 + e_i$ with $\max_i \|e_i\| \le \rho_0$, retaining the exact graph g, clean anchor, and $h = 1$,*

$$\Sigma \succeq \tfrac{1}{2}\lambda_{\min}(\Sigma_0) I, \qquad S_{\mathrm{lin}} \succeq \tfrac{5}{6} I, \qquad 0 \le \lambda_{\min}(\mathcal{H}) \le 884\varepsilon^2, \quad \varepsilon^2 = \frac{1}{12}\sum_i \|e_i\|^2.$$

No symmetry is required of the perturbations. Strict reference margins for $K = 2$, $\tau = \lambda_{\min}(\Sigma_0)/2$, and a fixed positive trust radius centered at truth can be retained by decreasing ρ_0.

Proof. Full column rank makes Σ, its inverse, and $S_{\mathrm{lin}} = T^\top(I - \mathsf{P}_\Phi)T/12$ continuous in the finite collection of points. At the reference their minimum eigenvalues are positive and $5/3$, respectively. Choose a sufficiently small neighborhood so that the stated bounds hold, and, for strict feasibility, so that the minimum Gram eigenvalue is greater than half its reference value.

For $f(x, y) = x(1 + (x^2 - y^2)/2)$, all reference values are zero. Within radius $1/2$ of the reference points, $|x|, |y| \le 3$, and

$$\nabla f = (1 + (3x^2 - y^2)/2, -xy), \qquad \|\nabla f\|^2 \le 19^2 + 9^2 = 442.$$

The mean value inequality on each connecting segment gives $12^{-1}\sum_i f(t_i)^2 \le 442\varepsilon^2$. Orthogonal projection is contractive, hence

$$\lambda_{\min}(\mathcal{H}) \le e_1^\top \mathcal{H} e_1 = \frac{2}{12}\|(I - \mathsf{P}_\Phi)[f(t_i)]_i\|^2 \le 884\varepsilon^2.$$

Positive semidefiniteness follows from the Gram representation of the Hessian. Curvature and trust margins remain strict at truth. □

Quantifying coefficient absorption without symmetry. More generally for a scalar exact quadratic graph $g(t) = t^\top Q t/2$, put $v_i = t_i + (t_i^\top Q t_i)Q t_i/2$, let V have rows $v_i^\top$, and define $C = \Phi^\top V/N$, $\Sigma = \Phi^\top\Phi/N$. Under the same local interiority and full-rank assumptions,

$$\mathcal{H} = 2\left(V^\top V/N - C^\top \Sigma^{-1} C\right), \qquad \lambda_{\min}(\mathcal{H}) \ge 2\lambda_{\min}(V^\top V/N) - 2\|\Sigma^{-1/2}C\|_{\mathrm{op}}^2.$$

This follows by expanding $2V^\top(I - \mathsf{P}_\Phi)V/N$. Central symmetry sets $C = 0$; otherwise the second term measures the orientation information absorbed by coefficient refitting. It is distinct from the predictor-rotation contribution within V.

Remark (Interpretation). The perturbation bound is a deterministic local statement with a conservative explicit upper-bound constant; it is not a frequency claim for random data. Hessian singularity does not by itself imply multiple minimizers or solver failure. The value $h = 1$ is a feature normalization and does not bound the sample radius here. These examples therefore do not verify every premise of the noisy recovery theorem or contradict its additional localization and smallness requirements.

C.11.2 Explicit derivative and feasibility bounds

Consider perturbed rows $u_i(z) = u_i^0 + hzv_i$, with $\|v_i\| \le 1$ and $0 \le z \le \delta \le 1$. All matrices above now depend on (A, z). For joint derivatives use the parameter norm $\|(H, b)\|_\times = \|H\|_F + |b|$; output norms are matrix Frobenius norms, except that design and Gram matrices may use operator norms. Let

$$R = \max_i \|u_i^0\|/h, \quad \bar{R} = R + 1, \quad w_j = \bar{R} + j \ (0 \le j \le 3), \quad y_j = w_j,$$

$$x_0 = \sqrt{1 + \bar{R}^4/4}, \qquad x_j = \tfrac{1}{2}\sum_{k=0}^{j}\binom{j}{k} w_k w_{j-k} \quad (1 \le j \le 3).$$

These bounds are deliberately conservative.

Lemma C.54 (Derivative bounds for the normalized profile)**.** *On a region where $X^\top X \succeq \lambda I$, the quantities x_j, y_j bound the joint derivatives of X, Y through order three. Define, for $1 \le j \le 3$,*

$$G_j = \sum_{k=0}^{j}\binom{j}{k} x_k x_{j-k}, \qquad b_j = \sum_{k=0}^{j}\binom{j}{k} x_k y_{j-k}, \qquad B_0 = y_0/\sqrt{\lambda},$$

$$B_j = \lambda^{-1}\left(b_j + \sum_{k=1}^{j}\binom{j}{k} G_k B_{j-k}\right), \qquad C_0 = y_0, \quad C_j = y_j + \sum_{k=0}^{j}\binom{j}{k} x_k B_{j-k}.$$

Then $\|D^j\beta\| \le B_j$ and $\|D^j\mathcal{R}\| \le C_j$, including $\|\mathcal{R}\| \le C_0$.

Proof. The ordered-integral formula for derivatives of a matrix exponential, together with orthogonality of exponentials of skew matrices, bounds its j-th derivative by the product of the direction operator norms. Moreover $\|\Omega(H)\|_{\mathrm{op}} = \|H\|_{\mathrm{op}} \le \|H\|_F$. For the normalized rotated row, differentiating its affine data factor adds at most j terms; its order-j bound is therefore w_j. The quadratic features are the orthonormal coordinates of $ss^\top/(2h^2)$; the product rule gives x_j. Averaging the squared row bounds gives the stated normalized matrix bounds. Write $G = X^\top X$, $b = X^\top Y$. The product rule bounds their order-j derivatives by G_j, b_j. The pseudoinverse bound gives $\|\beta\|_F \le y_0/\sqrt{\lambda}$. Differentiate $G\beta = b$ repeatedly, move all terms except $GD^j\beta$ to the right, and use $\|G^{-1}\|_{\mathrm{op}} \le 1/\lambda$. This gives the recurrence for B_j, including mixed directions. Differentiate $\mathcal{R} = Y - X\beta$ for C_j; at order zero, least-squares orthogonal projection gives the sharper bound $C_0 = y_0$. □

Lemma C.55 (Margins along the perturbation path)**.** *Let $\lambda_* = \lambda_{\min}(X(0,0)^\top X(0,0)) > \tau$, $d_0 = \|\Pi^* - \Pi_0\|_F < r_0$, and $K_* < K$. Evaluate Lemma C.54 at $\lambda = \lambda_*/2$. If*

$$r + \delta \le \min\left\{ \frac{\lambda_* - \tau}{2G_1}, \frac{h(K - K_*)}{2B_1} \right\}, \qquad r \le \frac{r_0 - d_0}{4}, \tag{49}$$

then for $\|A\|_F \le r$, $0 \le z \le \delta$, the design and trust-region inequalities have strict margins, and the unconstrained profile is interior to $\mathfrak{Q}_K$. It equals the constrained profile.

Proof. The Gram derivative bound, which does not require invertibility, gives $\|G(A,z) - G(0,0)\|_{\mathrm{op}} \le G_1(r+\delta)$. Thus $\lambda_{\min} G(A,z) \ge (\lambda_* + \tau)/2 > \tau$, and in particular it is at least $\lambda_*/2$. The coefficient derivative bound is now valid along the straight segment from $(0,0)$, giving $\|\beta(A,z) - \beta(0,0)\|_F \le B_1(r+\delta)$. Its quadratic block represents hQ, so the curvature norm is at most $K_* + (K - K_*)/2 < K$. Finally, the exponential-chart projector has $\|D\Pi(A)[H]\|_F \le 2\|H\|_F$. For example, express this derivative as a commutator with the integral of orthogonal conjugates of $\Omega(H)$; its off-diagonal block has Frobenius norm at most $\|H\|_F$. Integration gives $\|\Pi(A) - \Pi^*\|_F \le 2r$, and hence strict trust-region feasibility. □

C.11.3 Perturbation stability and a gradient certificate

Theorem C.56 (A quantitative local basin and noisy gradient certificate)**.** *Assume Proposition C.50, with $\gamma_{\mathrm{loc}} > 0$. For fixed perturbed rows $u_i = u_i^0 + w_i$, put $\delta = \max_i \|w_i\|/h \le 1$; use the path above (constant if $\delta = 0$). Use the constants from Lemma C.54 with $\lambda = \lambda_*/2$, and let*

$$M_0 = 1 + K_2(K_2 h^2/2 + \|a_0\|), \qquad C_{\mathrm{pert}} = 6C_1C_2 + 2C_0C_3.$$

Choose $0 < r \le 1/4$ satisfying equation 49 and

$$6M_0C_2 r + 3C_2^2 r^2 \le \gamma_{\mathrm{loc}}^2, \qquad C_{\mathrm{pert}}\delta \le \gamma_{\mathrm{loc}}^2/2, \qquad \delta < \frac{\gamma_{\mathrm{loc}}^2 r}{8C_1^2}. \tag{50}$$

Then the actual loss F_δ is $m_{\mathrm{loc}} = h^2\gamma_{\mathrm{loc}}^2/2$-strongly convex on the radius-r ball. Its minimizer over that ball is unique, lies in its interior, and satisfies

$$\|A^\dagger\|_F \le 4C_1^2\delta/\gamma_{\mathrm{loc}}^2 < r/2.$$

For every output $\widetilde{A}$ in this ball,

$$\|\Pi(\widetilde{A}) - \Pi^*\|_F \le C_{\mathrm{ch}} \left\{ \frac{4C_1^2}{\gamma_{\mathrm{loc}}^2}\delta + \frac{2\|\nabla F_\delta(\widetilde{A})\|_F}{h^2\gamma_{\mathrm{loc}}^2} \right\}, \tag{51}$$

where one may take $C_{\mathrm{ch}} = 2$ in this chart. The minimizer is local; no identification with every global constrained minimizer and no claim of L-BFGS basin entry are required.

Proof. At zero in the reference problem, $\mathcal{R}(0,0) = 0$, $\|D_A\mathcal{R}(0,0)\| \le M_0$, and Proposition C.50 gives $\nabla^2 F_0(0) \succeq 2h^2\gamma_{\mathrm{loc}}^2 I$. For $\|A\|_F \le r$, Taylor's theorem gives

$$\|D_A\mathcal{R}(A,0) - D_A\mathcal{R}(0,0)\| \le C_2 r, \qquad \|\mathcal{R}(A,0)\|_F \le M_0 r + C_2 r^2/2.$$

Differentiate the squared residual twice to obtain

$$D_A^2 F_z[H,H] = 2h^2\{\|D_A\mathcal{R}[H]\|_F^2 + \langle \mathcal{R}, D_A^2\mathcal{R}[H,H]\rangle_F\}.$$

The two Taylor bounds imply $\|\nabla^2 F_0(A) - \nabla^2 F_0(0)\|_{\mathrm{op}} \le h^2(6M_0C_2r + 3C_2^2r^2)$. Next differentiate the displayed Hessian in z. Its three terms are $2\langle D_A\mathcal{R}, D_zD_A\mathcal{R}\rangle$, $\langle D_z\mathcal{R}, D_A^2\mathcal{R}\rangle$, and $\langle \mathcal{R}, D_zD_A^2\mathcal{R}\rangle$, multiplied by $2h^2$. Lemma C.54 therefore yields

$$\|\nabla^2 F_\delta(A) - \nabla^2 F_0(A)\|_{\mathrm{op}} \le h^2 C_{\mathrm{pert}}\delta.$$

The smallness assumptions give the claimed strong convexity. Also $\|\mathcal{R}(0,\delta)\|_F \le C_1\delta$, whence $\|\nabla F_\delta(0)\|_F \le 2h^2C_1^2\delta < m_{\mathrm{loc}}r/2$. A minimizer exists on the compact ball. At a boundary point, strong monotonicity gives $\langle \nabla F_\delta(A), A\rangle \ge m_{\mathrm{loc}}r^2 - \|\nabla F_\delta(0)\|_F r > 0$, so an inward step decreases the loss. Thus the minimizer is interior and stationary. Strong monotonicity gives $\|A^\dagger\|_F \le \|\nabla F_\delta(0)\|_F/m_{\mathrm{loc}}$ and $\|\widetilde{A} - A^\dagger\|_F \le \|\nabla F_\delta(\widetilde{A})\|_F/m_{\mathrm{loc}}$. The projector Lipschitz bound proves equation 51. □

Remark (Statistical interpretation and limitations). For actual graph rows, write $g(t) = \mathcal{Q}_{Q^*}(t) + R(t)$, $\sup_{\|t\|\le h}\|R(t)\| \le R_h$, and bound the neighbor noise by σ_{nbr}. Keeping the same selected indices and the same normal anchor displacement gives $\delta \le (R_h + \sigma_{\mathrm{nbr}})/h$. For a cubic remainder this is $L_3h^2/6 + \sigma_{\mathrm{nbr}}/h$, and for $C^{2,\alpha}$ it is of order $h^{1+\alpha} + \sigma_{\mathrm{nbr}}/h$. No extra manifold differentiability is needed for the third derivatives above: they differentiate the finite quadratic regression in its parameters and fixed-data interpolation, not the manifold itself.

The constants and basin membership remain model-based. In particular, $r + \delta \le h(K - K_*)/(2B_1)$ may be much stronger than the noise conditions of the statistical theorem; if B_1 and the margin are held fixed, this sufficient condition can require neighbor noise of order $h^2(K - K_*)$. It is not asserted to be a necessary noise threshold. The general residual certificate in Corollary C.39 continues to apply without these landscape premises. If the implemented objective is F/h^2, its gradient absorbs the h^{-2} factor in equation 51. A gradient from a PCA-centered chart additionally requires the corresponding coordinate Jacobian bound; it must not be substituted unchanged into this formula. Approximate inner solves require separate control of their residual and gradient errors. Active coefficient boundaries are not covered here.

Remark (What is and is not proved about PCA and iteration). If an existing PCA result supplies $\|\Pi_0 - \Pi^*\|_F \le \varepsilon_{\mathrm{PCA}}$, the principal-angle representative in the true-centered exponential chart satisfies $\|A_0\|_F \le \pi\varepsilon_{\mathrm{PCA}}/(2\sqrt{2})$. Thus $\pi\varepsilon_{\mathrm{PCA}}/(2\sqrt{2}) < r$ is an additional sufficient entry test; d_0 in equation 49 can be replaced by its upper bound $\varepsilon_{\mathrm{PCA}}$. This comparison is not automatic, and entry alone does not establish that the implemented L-BFGS iterates remain in the ball. Theorem C.59 retains its basin and iteration premises. The present results do not assert convergence of the default solver from every feasible PCA start, nor a fully observable end-to-end certificate.

C.11.4 Local-chart implementation

Default initializer and feasibility check. For the observed-radius neighborhood, form the anchor-centered scatter

$$\widehat{\Sigma}_{u,r} = \frac{1}{N_r}\sum_{i\in\mathcal{I}_r(x_0)}(x_i - x_0)(x_i - x_0)^\top$$

and take its top-d eigenvectors as $P_{0,1}$, with any orthogonal completion P_0. No additional sample-mean subtraction is used in this initializer. Before starting a constrained solver, check $\lambda_{\min}(\widehat{\Sigma}_{Q,h}(P_0)) \ge \tau$; for a strictly feasible method require $> \tau$, and take $r_0 > 0$. If this check fails, report an initialization failure for the specified constrained problem. Corollary C.34 bounds the probability of such failure under its stated conditions. This observable check requires no knowledge of Π^*. The high-probability theorem supplies its theoretical justification rather than replacing the check.

The local-chart solver uses constrained coefficient profiling. Given a frame $P = [P_1 \mid P_2]$, define

$$\mathrm{Profile}(P) = \big(J_{N,K}(P), \widehat{c}(P), \widehat{Q}(P), Z(P), \widehat{\Sigma}_{Q,h}(P)\big),$$

where $(\widehat{c}(P), \widehat{Q}(P))$ solves the constrained inner problem and

$$Z(P) = \nabla_P L_N\big(P, \widehat{c}(P), \widehat{Q}(P)\big)$$

Algorithm 1 Local-chart profiled L-BFGS

Input: Strictly feasible P_0, memory M_{mem}, iteration limit T, tolerance $\varepsilon_{\mathrm{opt}}$
Output: Projector $\widetilde{\Pi}$ and profiled coefficients

$A \leftarrow 0$; clear the L-BFGS history
Define $F_N(A) = J_{N,K}(P(A))$
for $k = 0, \dots, T-1$ **do**
  Evaluate $F_N(A)$, $g_k = \nabla F_N(A)$, c_k, and Q_k
  if $\|g_k\|_F \le \varepsilon_{\mathrm{opt}}$ **then**
    break
  end if
  $D_k \leftarrow -\mathrm{LBFGSTwoLoop}(g_k, \{S_j, Y_j\})$
  if $\langle g_k, D_k \rangle_F \ge 0$ **then**
    Clear the history and set $D_k \leftarrow -g_k$
  end if
  Choose η_k by feasible Armijo backtracking
  $A^+ \leftarrow A + \eta_k D_k$
  $S_k \leftarrow A^+ - A$
  $Y_k \leftarrow \nabla F_N(A^+) - g_k$
  if $\langle S_k, Y_k \rangle_F > 0$ **then**
    Store (S_k, Y_k); retain at most M_{mem} pairs
  end if
  $A \leftarrow A^+$
end for
Refit (c_k, Q_k) at the final $P(A)$
return $\widetilde{\Pi} = \Pi(A)$, c_k, and Q_k

is the envelope gradient. The subroutine also reports the feasibility flag

$$\mathbf{1}_{\mathrm{feas}}(P) = \mathbf{1}\{\Pi(P) \in \mathcal{C}_{\tau,h}(\Pi_0, r_0)\}.$$

The curvature constraint $\widehat{Q}(P) \in \mathfrak{Q}_K$ is enforced inside every call to $\mathrm{Profile}$.

On the full-rank region, first-order envelope gradients remain valid even at transitions of the inner coefficient active set. Smooth second-order methods require an inactive constraint or a sufficiently regular KKT sensitivity analysis.

For Algorithm 1, define explicitly

$$\Omega(A) = \begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix}, \qquad P(A) = P_0 \exp(\Omega(A)), \qquad A \in \mathbb{R}^{m \times d}.$$

Use only its local injectivity domain, for example $\|A\|_{\mathrm{op}} < \pi/2$, and check this in trial steps in addition to both orientation constraints. This exponential coordinate is centered at the computed initializer; it is not the graph coordinate centered at P^* in the risk proof. If a feasible line search stalls at an active boundary, report that failure or switch to a constraint-aware method, rather than declaring stationarity.

Algorithm 1 is the default implementation. Its optimization variable has dimension md, without within-block rotational redundancy. We use feasible Armijo backtracking by default; feasible Wolfe search is an optional alternative on a smooth interior region. Its strict orientation feasibility at initialization is guaranteed by Corollary C.34 with $\nu > 0$; the coefficient-smoothness requirements stated above remain separate.

At an active design or trust-region boundary, a rejected line search is not a stationarity certificate. A constraint-aware step or feasibility restoration may be necessary. The convergence statements below concern local gradient descent under explicit interior-basin assumptions; they do not supply a convergence theorem for the L-BFGS pseudocode.

C.12 Conditional numerical convergence and optimization error

Let $\widehat{\Pi}_N$ be a minimizer of equation 30 and $\widetilde{\Pi}_T$ a numerical output. Then

$$\|\widetilde{\Pi}_T - \Pi^*\|_F \le \|\widetilde{\Pi}_T - \widehat{\Pi}_N\|_F + \|\widehat{\Pi}_N - \Pi^*\|_F. \tag{52}$$

Corollary C.39 controls feasible outputs through loss without strong convexity. Here we instead analyze local gradient descent under stronger basin assumptions, not convergence of the default L-BFGS solver from every feasible PCA start. Use the chart $P(A) = P_0 \exp \Omega(A)$ defined above, write $\Pi(A) = P_1(A)P_1(A)^\top$, and set $F_N(A) = J_{N,K}(P(A))$. Let $\widehat{A}_N$ represent a selected minimizer in this chart, so $\widehat{\Pi}_N = \Pi(\widehat{A}_N)$. No uniqueness outside the chart is assumed.

Assumption C.57 (Strictly feasible smooth basin)**.** *There is an open convex local chart $\mathcal{U} \subset \mathbb{R}^{m\times d}$, containing the initialization A_0 and the selected empirical minimizer $\widehat{A}_N$, with the following properties.*

1. *The chart is contained in the strictly feasible interior: for some $\delta_{\mathrm{des}}, \delta_{\mathrm{rad}} > 0$, every $A \in \mathcal{U}$ under consideration satisfies*
$$\lambda_{\min}\big(\widehat{\Sigma}_{Q,h}(P(A))\big) \ge \tau + \delta_{\mathrm{des}}, \qquad \|\Pi(A) - \Pi_0\|_F \le r_0 - \delta_{\mathrm{rad}}.$$
2. *The fitted coefficient pair is interior to the curvature constraint, so that the profiled objective is smooth on $\mathcal{U}$. In particular, the envelope formula in Proposition C.49 applies.*
3. *The gradient of F_N is L-Lipschitz on $\mathcal{U}$:*
$$\|\nabla F_N(A) - \nabla F_N(B)\|_F \le L\,\|A - B\|_F\,.$$
4. *The initial sublevel set*
$$\mathcal{S}_0 = \{A \in \mathcal{U} : F_N(A) \le F_N(A_0)\}$$
is compact and lies strictly inside $\mathcal{U}$.

The strict margins in Assumption C.57 have a simple role. They make the design and trust-region constraints safeguards rather than active boundaries near the desired solution. Consequently, a sufficiently small gradient step remains feasible. If a proposed step leaves the chart or violates a margin, backtracking may be used until it returns to the feasible neighborhood.

Theorem C.58 (First-order convergence)**.** *Suppose Assumption C.57 holds. Consider the local profiled gradient iteration*

$$A_{k+1} = A_k - \eta \nabla F_N(A_k), \qquad 0 < \eta \le L^{-1}, \tag{53}$$

where the fixed step is small enough that the update segments remain in $\mathcal{U}$. Such a uniform step exists: if $d_0 = \mathrm{dist}(\mathcal{S}_0, \mathcal{U}^c) > 0$ and $G_0 = \max_{\mathcal{S}_0} \|\nabla F_N\|_F > 0$, take additionally $\eta < d_0/G_0$. Descent then keeps every iterate in $\mathcal{S}_0$ by induction. If $G_0 = 0$, the starting point is stationary. Let

$$F_{N,\inf} = \inf_{A\in\mathcal{S}_0} F_N(A).$$

Then

$$\min_{0\le k<T} \|\nabla F_N(A_k)\|_F^2 \le \frac{2\{F_N(A_0) - F_{N,\inf}\}}{\eta T}. \tag{54}$$

In particular, the smallest gradient norm among the first T iterates is of order $T^{-1/2}$. Every accumulation point of the iterates is a first-order stationary point of the local profiled objective.

Proof. Apply the standard smooth descent inequality (Boumal, 2023, Chapter 4) along the stipulated update segments:

$$F_N(A_{k+1}) \le F_N(A_k) - \frac{\eta}{2}\|\nabla F_N(A_k)\|_F^2. \tag{55}$$

The step restriction and compact sublevel set keep these segments in the smooth feasible chart. Summing gives $\frac{\eta}{2}\sum_{k<T} \|\nabla F_N(A_k)\|_F^2 \le F_N(A_0) - F_{N,\inf}$, proving equation 54. The same bound implies summability of the squared gradient norms; compactness and continuity yield stationarity of accumulation points. □

The preceding theorem does not claim that the stationary point is the desired empirical minimizer. For that stronger conclusion, one needs local curvature of the *profiled* loss. The Hessian relevant to this condition is the Schur-complement Hessian in Proposition C.49, not the Hessian obtained while incorrectly holding (c, Q) fixed.

Theorem C.59 (Local convergence to a selected empirical minimizer)**.** *Here μ_{opt} is an optimization strong-convexity constant, distinct from the statistical separation constant in Section C.5. In addition to Assumption C.57, suppose that F_N is μ_{opt}-strongly convex on the convex chart $\mathcal{U}$, containing both the iterates and $\widehat{A}_N$:*

$$F_N(B) \geq F_N(A) + \langle \nabla F_N(A), B - A \rangle_F + \frac{\mu_{\mathrm{opt}}}{2} \|B - A\|_F^2 . \tag{56}$$

Then $\widehat{A}_N$ is the unique minimizer in that chart, and the iteration equation 53 satisfies

$$F_N(A_k) - F_N(\widehat{A}_N) \leq (1 - \mu_{\mathrm{opt}}\eta)^k \{F_N(A_0) - F_N(\widehat{A}_N)\}. \tag{57}$$

Moreover, because $A \mapsto \Pi(A)$ is smooth on the compact local basin, there is a chart constant $C_{\mathrm{ch}} > 0$ such that

$$\left\| \Pi(A_k) - \widehat{\Pi}_N \right\|_F^2 \leq \frac{2C_{\mathrm{ch}}^2}{\mu_{\mathrm{opt}}} (1 - \mu_{\mathrm{opt}}\eta)^k \{F_N(A_0) - F_N(\widehat{A}_N)\}. \tag{58}$$

Proof. The standard strong-convexity inequalities (Bubeck, 2015, Chapter 3) give uniqueness and

$$\|\nabla F_N(A)\|_F^2 \geq 2\mu_{\mathrm{opt}} \{F_N(A) - F_N(\widehat{A}_N)\}, \qquad \|A - \widehat{A}_N\|_F^2 \leq \frac{2}{\mu_{\mathrm{opt}}} \{F_N(A) - F_N(\widehat{A}_N)\}.$$

Combine the first with equation 55 and iterate to obtain equation 57. Smoothness of the projector chart on the compact basin gives $\|\Pi(A) - \Pi(B)\|_F \leq C_{\mathrm{ch}} \|A - B\|_F$. Combining this with the second inequality proves equation 58. □

Corollary C.60 (Optimization error in the tangent risk)**.** *Under the assumptions of Theorem C.59, suppose an iterate $\widetilde{A}_T$ satisfies*

$$\left\| \nabla F_N(\widetilde{A}_T) \right\|_F \leq \varepsilon_{\mathrm{opt}}.$$

Then

$$\left\| \Pi(\widetilde{A}_T) - \widehat{\Pi}_N \right\|_F \leq \frac{C_{\mathrm{ch}}}{\mu_{\mathrm{opt}}} \varepsilon_{\mathrm{opt}}, \tag{59}$$

and consequently

$$\left\| \Pi(\widetilde{A}_T) - \Pi^* \right\|_F^2 \leq 2 \left\| \widehat{\Pi}_N - \Pi^* \right\|_F^2 + \frac{2C_{\mathrm{ch}}^2}{\mu_{\mathrm{opt}}^2} \varepsilon_{\mathrm{opt}}^2. \tag{60}$$

If instead one knows the objective gap

$$F_N(\widetilde{A}_T) - F_N(\widehat{A}_N) \leq \Delta_{\mathrm{opt}},$$

then the final term in equation 60 may be replaced by $4C_{\mathrm{ch}}^2 \Delta_{\mathrm{opt}} / \mu_{\mathrm{opt}}$.

Proof. Strong monotonicity and $\nabla F_N(\widehat{A}_N) = 0$ imply $\mu_{\mathrm{opt}} \|\widetilde{A}_T - \widehat{A}_N\|_F \leq \varepsilon_{\mathrm{opt}}$. The projector chart bound proves equation 59; equation 52 and $(a + b)^2 \leq 2a^2 + 2b^2$ give equation 60. For a known objective gap, use $\|\widetilde{A}_T - \widehat{A}_N\|_F^2 \leq 2\Delta_{\mathrm{opt}} / \mu_{\mathrm{opt}}$. □

The coefficient estimates converge as well. On the strictly feasible smooth basin, the implicit-function argument used in Proposition C.49 makes $A \mapsto \widehat{\beta}(A)$ locally Lipschitz. Hence, for some $C_\beta > 0$,

$$\left\| \widehat{\beta}(\widetilde{A}_T) - \widehat{\beta}(\widehat{A}_N) \right\| \leq C_\beta \left\| \widetilde{A}_T - \widehat{A}_N \right\|_F \leq \frac{C_\beta}{\mu_{\mathrm{opt}}} \varepsilon_{\mathrm{opt}}.$$

This is a coordinate-level statement in the chosen local frame. Globally, the frame and the coefficient tuple are not unique because tangent and normal bases may be rotated; the projector is the intrinsic tangent parameter.

Remark (What is and is not proved). The convergence results are local and conditional on the stated basin and iteration premises. They do not establish global recovery on $\mathrm{Gr}(d, D)$ or convergence of the implemented L-BFGS method from every PCA start. Active constraints require a separate constrained-optimization analysis.