Uncertainty Quantification of Least Squares Estimator for Generalized Orthogonal Procrustes Problems

Shenghan Luo
Shanghai Jiao Tong University
Shanghai, China
shenghanluo@sjtu.edu.cn

Abstract

The generalized orthogonal Procrustes problem (GOPP) aims to recover rigid transformations that best align multiple point clouds and has broad applications in 3D geometry, computer vision, and biomedicine. Despite its importance, the problem is challenging due to the nonconvex orthogonality constraints. Fortunately, the generalized power method (GPM) has proven highly effective, theoretically guaranteed to converge to the global least squares (LS) estimator under reasonable conditions. Yet, the statistical inference of this global LS estimator remains largely an open problem. Specifically, the absence of an exact second-order analytic expansion prevents rigorous uncertainty quantification (UQ), which is vital for constructing confidence regions and evaluating estimator reliability. While recent UQ frameworks developed for orthogonal group synchronization offer a potential paradigm, they severely lack universality, often failing to accommodate the complex anisotropic distortions inherent in GOPP shape matrices. To bridge this gap, we propose a more general theoretical framework with which we successfully derive the exact second-order analytic expansion of the GOPP global LS estimator under additive Gaussian noise. Leveraging this theoretical foundation, we further construct confidence regions. Finally, extensive numerical experiments empirically validate the exactness of our theoretical expansion and demonstrate the robust coverage of the proposed confidence regions across varying conditions.

1 Introduction

Given a set of multiple point clouds, how to find a rotation for each point cloud such that all the transformed point clouds are well aligned? This problem, known as the generalized orthogonal Procrustes problem, has found many applications in statistics [Gower, 1975, Stegmann and Gomez, 2002], computer vision [Besl and McKay, 1992, Maron et al., 2016, Mitra et al., 2004], and imaging science [Bandeira et al., 2014, Chaudhury et al., 2015, Khoo and Kapoor, 2016, Pumir et al., 2021, Singer, 2018, 2011].

One common approach to solve this generalized orthogonal Procrustes problem is finding the least squares estimator. However, it turns out that finding the globally optimal least squares estimator is an NP-hard problem in general. Therefore, one would seek for alternative methods to approximate the global solution to this nonconvex program. Among many existing approaches, generalized power method with spectral initialization is one of the most powerful methods to efficiently find the least squares (LS) estimator. Its near-optimal estimation error bound is now well understood. However, in practical applications, obtaining a statistically accurate point estimate alone is insufficient. A systematic quantification of uncertainty is also essential. From a statistical perspective, acquiring the approximate error distribution of LS estimator is critical for many tasks such as constructing valid confidence regions and conduct hypothesis tests. Such results turn statistical estimators into tools not

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

just for prediction, but also for reliable decision-making under uncertainty. This motivates this work which leads to the following question:*What is the approximate distribution of LS estimator for GOPP and how can we construct confidence regions?*

To address this question, we build on the one-step Riemannian Newton strategy of Zhong and Ling [2024], while developing the operator analysis needed for GOPP. In their orthogonal synchronization model, independent Gaussian edge noise and isotropic signal geometry yield scalar leading curvature: the tangent correction is obtained by rescaling a skew-symmetric score, giving spherical uncertainty. GOPP introduces an unknown template with a general shape matrix $S = \boldsymbol{A}\boldsymbol{A}^\top$. The resulting curvature varies across rotational directions, and left-multiplying a skew-symmetric score G by S^{-1} generally destroys skew-symmetry. The tangent correction must instead solve the two-sided equation $S\Omega + \Omega S = G$. Moreover, pairwise products formed from the same point cloud share noise, so their covariance depends on both the template and the observation dependence. The technical advance is to connect this Sylvester correction to the global LS estimator with a condition-number-dependent second-order remainder, derive the shape-dependent covariance, and analyze plug-in estimation of the unknown template and noise level. These steps extend the shared Newton strategy from scalar, isotropic geometry to data-driven anisotropic inference for GOPP.

Specifically, our core contributions are summarized as follows. First, we establish **a more general framework to derive the approximate distribution of estimators in statistical inference problems with manifold constraints**. Unlike prior works that rely on restrictive isotropic or homogeneous assumptions, our exact analytical approach systematically accommodates complex, anisotropic geometric distortions. This serves as foundational machinery applicable to a broader class of statistical estimation problems over matrix manifolds, such as the Little Grothendieck problem, multireference alignment (MRA), and subspace estimation. Second, we derive **the first exact distribution of the GOPP LS estimator**. Building upon this framework, we obtain the exact second-order analytic expansion and blockwise asymptotic distribution for the global LS estimator of the GOPP under additive Gaussian noise. Third, we construct **the first statistically exact confidence regions for the GOPP**. Our approach yields fully computable, statistically exact Chi-square confidence regions on the orthogonal group without requiring any prior knowledge of ground-truth parameters. Finally, we provide **rigorous empirical validation**. Comprehensive numerical experiments on synthetic and ModelNet40 point clouds flawlessly confirm our theoretical guarantees, robustly supporting the exactness of our geometric constraints, the accuracy of the derived asymptotic distributions, and the valid nominal coverage of our proposed confidence regions.

1.1 Related Works

Our work builds upon and extends three interconnected lines of research: solving methods for GOPP and uncertainty quantification in high-dimensional statistical estimation.

We discuss relevant research on the solving methods for GOPP. The most common approach to solve GOPP is finding the least squares estimator. Finding the least squares estimator for the GOPP can be reduced to solving an optimization program with orthogonality constraints, a well-known NP-hard problem in general that has been extensively studied [Abbe, 2017, Bandeira et al., 2016, Goemans and Williamson, 1995, So, 2011, Wen and Yin, 2013]. Due to the nonconvexity of the problem, obtaining a globally optimal solution is often challenging. Here, we briefly review several popular approaches for tackling this program. Convex relaxation is arguably one of the most popular and powerful methods. A common relaxation strategy generalizes the famous Goemans-Williamson relaxation [Goemans and Williamson, 1995] for the graph max-cut problem by considering a semidefinite program (SDP) with block diagonal constraints [Bandeira et al., 2016, 2017, Boumal, 2015]. After solving the relaxed SDP, an approximate solution can be obtained via rounding procedures with solid theoretical guarantees [Bandeira et al., 2016, Goemans and Williamson, 1995, Mei et al., 2017]. However, directly solving large-scale SDPs is computationally prohibitive, which presents a major roadblock for practical applications. This computational bottleneck urgently motivates the development of efficient first-order methods that can match the theoretical guarantees of SDP relaxations. To bypass this bottleneck, the generalized power method (GPM) with spectral initialization has emerged as a highly scalable alternative. Recent theoretical breakthroughs [Ling, 2022, 2023a,b] have established rigorous near-optimal error bounds, guaranteeing its linear convergence to the global LS estimator and thus providing an efficiently accessible target for our statistical characterization.

We discuss relevant work on the precise characterization of statistical estimators, also known as uncertainty quantification. This area has recently gained significant attention in a series of applications, such as statistical ranking [Gao et al., 2023] and matrix completion [Chen et al., 2019]. In high-dimensional linear regression, several important works have focused on constructing confidence intervals for estimators and deriving the asymptotic distribution of the Lasso estimator under high-dimensional models, e.g., Zhang and Zhang [2014], Cai and Guo [2017], Javanmard and Montanari [2014], and van der Pas et al. [2017]. For the generalized linear model, Sur and Candès [2019] characterized the distribution of likelihood ratio statistics and the asymptotic bias and variance of the Maximum Likelihood Estimator (MLE) in high-dimensional logistic regression. In graphical models, Janková and van de Geer [2015] and Ren et al. [2015] proposed methodologies for sparse inverse covariance estimation in high-dimensional settings and obtained the limiting distribution of the estimators. Regarding uncertainty quantification under orthogonality constraints, a recent breakthrough by Zhong and Ling [2024] provided a theoretical framework for group synchronization problems via the MLE and spectral estimator, extending earlier results on the convex Bradley-Terry-Luce model [Gao et al., 2023] to nonconvex manifold settings. While their framework presents a promising paradigm, it exhibits a critical limitation in universality: it relies heavily on idealized isotropic observation models and homogeneous topologies. Such assumptions intrinsically fail to accommodate the complex, anisotropic geometric distortions induced by arbitrary shape matrices in questions like GOPP. Therefore, a more general statistical inference framework capable of systematically capturing this exact anisotropic error propagation on Lie algebras is urgently needed, which is one of the primary focus of our work.

1.2 Organization of Our Paper

Section 2 introduces the basis of GOPP. Section 3 gives the main results of our work, and Section 4 provides numerical experiments that complement our theory. The detailed proof is deferred to the supplementary material.

1.3 Notations

We denote boldface $\boldsymbol{x}$ and $\boldsymbol{X}$ as a vector and matrix respectively, and $\boldsymbol{x}^\top$ and $\boldsymbol{X}^\top$ are their corresponding transpose. The $\boldsymbol{I}_n$ stands for the $n \times n$ identity matrix. For two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the same size, $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{i,j} X_{ij} Y_{ij}$ is their inner product. For any matrix $\boldsymbol{X}$, the Frobenius, operator, and nuclear norms are denoted by $\|\boldsymbol{X}\|_F$, $\|\boldsymbol{X}\|$, and $\|\boldsymbol{X}\|_*$ respectively. For two nonnegative functions $f(x)$ and $g(x)$, we say $f(x) \lesssim g(x)$ (and $f(x) \gtrsim g(x)$) if there exists a universal constant $C > 0$ such that $f(x) \leq Cg(x)$ (and $f(x) \geq Cg(x)$). We say a matrix $\boldsymbol{X} = O(T)$ if $\|\boldsymbol{X}\|_F \leq c \cdot T$ for some constant $c > 0$.

2 Preliminaries

Suppose we observe a set of n noisy point clouds in $\mathbb{R}^d$ transformed by an unknown orthogonal matrix $\boldsymbol{O}_i \in \mathrm{O}(d)$:

$$\boldsymbol{A}_i = \boldsymbol{O}_i \boldsymbol{A} + \sigma \boldsymbol{W}_i \tag{2.1}$$

where $\boldsymbol{W}_i$ is the additive Gaussian noise, $m \geq d$ and

$$\boldsymbol{D} := \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_n \end{bmatrix} = \boldsymbol{O}\boldsymbol{A} + \boldsymbol{W} \text{where} \boldsymbol{O} := \begin{bmatrix} \boldsymbol{O}_1 \\ \vdots \\ \boldsymbol{O}_n \end{bmatrix} \in \mathrm{O}(d)^{\otimes n}, \boldsymbol{W} := \begin{bmatrix} \boldsymbol{W}_1 \\ \vdots \\ \boldsymbol{W}_n \end{bmatrix} \in \mathbb{R}^{nd \times m}.$$

The key question is: how to recover $\boldsymbol{O}_i$ and $\boldsymbol{A}$ from $\boldsymbol{A}_i$ (equivalently $\boldsymbol{D}$)? One commonly-used approach is finding the least squares estimator of $\boldsymbol{O}_i$ and $\boldsymbol{A}$ by minimizing

$$\min_{\boldsymbol{O}_i \in \mathrm{O}(d), \boldsymbol{A} \in \mathbb{R}^{d \times m}} \sum_{i=1}^{n} \|\boldsymbol{A}_i - \boldsymbol{O}_i \boldsymbol{A}\|_F^2. \tag{2.2}$$

Under the additive Gaussian noise, the least squares estimator exactly matches the maximum likelihood estimator. It is easy to see that (2.2) is a nonconvex program due to the orthogonality constraints.

Note that (2.2) involves two variables $\boldsymbol{O}_i$ and $\boldsymbol{A}$. In fact, we can remove $\boldsymbol{A}$ by first fixing $\boldsymbol{O}_i$. By assuming $\boldsymbol{O}_i$ is known, (2.2) is a convex function over $\boldsymbol{A}$ and minimizing it gives

$$\widehat{\boldsymbol{A}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{O}_i^\top \boldsymbol{A}_i. \tag{2.3}$$

By substituting $\widehat{\boldsymbol{A}}$ back into (2.2), the objective function (2.2) is equivalent to maximizing a generalized quadratic form with orthogonality constraints

$$\max_{\boldsymbol{O}_i \in \mathrm{O}(d)} \sum_{i<j} \langle \boldsymbol{C}_{ij}, \boldsymbol{O}_i \boldsymbol{O}_j^\top \rangle, \quad \text{where} \quad \boldsymbol{C}_{ij} = \boldsymbol{A}_i \boldsymbol{A}_j^\top \tag{P}$$

An equivalent formulation is given by

$$f(\boldsymbol{O}) = \left\langle \boldsymbol{C}, \boldsymbol{O}\boldsymbol{O}^\top \right\rangle = \|\boldsymbol{D}^\top \boldsymbol{O}\|_F^2, \boldsymbol{C} := \boldsymbol{D}\boldsymbol{D}^\top \in \mathbb{R}^{nd \times nd}$$

Fixing $\{\boldsymbol{A}_i\}_{i\neq j}^n$ and optimizing only for a single $\boldsymbol{O}_i$, we have

$$\max_{\boldsymbol{O}_i \in \mathrm{O}(d)} \sum_{i<j} \langle \boldsymbol{A}_i \boldsymbol{A}_j^\top, \boldsymbol{O}_i \boldsymbol{O}_j^\top \rangle = \max_{\boldsymbol{O}_i \in \mathrm{O}(d)} \langle \boldsymbol{O}_i, \sum_{j=1}^{n} \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j^\top \rangle \tag{2.4}$$

So the objective function (2.2) can be expressed as

$$f(\boldsymbol{\Phi}) := -\sum_{j=1}^{n} \langle \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j^\top, \boldsymbol{\Phi} \rangle, \quad \boldsymbol{\Phi} \in \mathrm{O}(d) \tag{2.5}$$

and $\widehat{\boldsymbol{O}}_i := \underset{\Phi \in \mathrm{O}(d)}{\operatorname{argmin}} f(\Phi)$

3 Main Theorems

We now present our main theoretical results, beginning with necessary notations and assumptions, followed by the distributional characterization of the LS estimator and the construction of confidence regions.

We first provide some additional assumption on the data matrix $\boldsymbol{A}$. Given a matrix $\boldsymbol{A}$, we let $\sigma_{\max}(\boldsymbol{A}) := \sqrt{\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^\top)}$ and $\sigma_{\min}(\boldsymbol{A}) := \sqrt{\lambda_{\min}(\boldsymbol{A}\boldsymbol{A}^\top)}$, and $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the largest and smallest eigenvalues respectively. In particular, we assume $\boldsymbol{A}$ is full row-rank, i.e., $\sigma_{\min}(\boldsymbol{A}) > 0$. Let κ be the condition number of $\boldsymbol{A}$, i.e., $\kappa = \sigma_{\max}(\boldsymbol{A})/\sigma_{\min}(\boldsymbol{A})$.

Throughout our presentation, we rely on the effective signal-to-noise ratio (SNR). Following standard random matrix theory bounds [Vershynin, 2018, Wainwright, 2019] for the maximum operator norm of the noise $\boldsymbol{W}_i$, we define the deterministic SNR as:

$$\mathrm{SNR} := \frac{\sqrt{n}\sigma_{\min}(\boldsymbol{A})}{\sigma(\sqrt{nd} + \sqrt{m} + \sqrt{2n\gamma \log n})}. \tag{3.1}$$

Moreover, for any estimator $\boldsymbol{S}$ of $\boldsymbol{O} \in \mathrm{O}(d)^{\otimes n}$, we note that $\boldsymbol{S}\boldsymbol{Q}$ satisfies $\boldsymbol{S}\boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{S}^\top = \boldsymbol{S}\boldsymbol{S}^\top$ for any orthogonal matrix $\boldsymbol{Q}$. Therefore, to resolve this ambiguity, we define the following distance function for any $\boldsymbol{S}$ and $\boldsymbol{O} \in \mathbb{R}^{nd\times d}$:

$$d_F(\boldsymbol{S}, \boldsymbol{O}) := \min_{\boldsymbol{O} \in \mathrm{O}(d)} \|\boldsymbol{S} - \boldsymbol{O}\boldsymbol{Q}\|_F.$$

The minimum is obtained via the polar retraction, which is defined as $\mathcal{P} : \mathbb{R}^{d\times d} \mapsto \mathrm{O}(d)$,

$$\mathcal{P}(\boldsymbol{Y}) \in \mathrm{argmin}_{\boldsymbol{Q} \in \mathrm{O}(d)} \|\boldsymbol{Q} - \boldsymbol{Y}\|_F^2, \tag{3.2}$$

where $\boldsymbol{Y} \in \mathbb{R}^{d\times d}$. In fact, $\mathcal{P}(\boldsymbol{Y})$ equals the product of the left and right singular vectors of $\boldsymbol{Y}$ and $\mathcal{P}(\cdot)$ is not unique if $\boldsymbol{Y}$ is rank-deficient. If $\boldsymbol{Y}$ is full-rank, then

$$\mathcal{P}(\boldsymbol{Y}) := (\boldsymbol{Y}\boldsymbol{Y}^\top)^{-1/2}\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{Y}^\top\boldsymbol{Y})^{-1/2}. \tag{3.3}$$

We also define $\mathcal{P}_n : \mathbb{R}^{nd\times d} \mapsto \mathrm{O}(d)^{\otimes n}$ which applies $\mathcal{P}$ to each $d \times d$ block:

$$\mathcal{P}_n([\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top]^\top) := [\mathcal{P}(\boldsymbol{Y}_1)^\top, \ldots, \mathcal{P}(\boldsymbol{Y}_n)^\top]^\top. \tag{3.4}$$

3.1 Distributional Guarantees

Our first theorem gives a block-wise approximate distribution of each $\widehat{\boldsymbol{O}}_i, 1 \le i \le n$, where $\widehat{\boldsymbol{O}}_i$ is the LS solution for GOPP.

Theorem 3.1 (Distribution of the LS Estimator). *Let $\widehat{\boldsymbol{O}} = [\widehat{\boldsymbol{O}}_1^\top, \dots, \widehat{\boldsymbol{O}}_n^\top]^\top$ be the global minimizer of the GOPP. Suppose $\mathrm{SNR} \gtrsim \kappa^2\sqrt{d}$ and $d \ge 2$. Up to the optimal global alignment $\boldsymbol{Q} = \mathcal{P}(\boldsymbol{O}^\top\widehat{\boldsymbol{O}}) \in \mathrm{O}(d)$, the aligned estimator admits the following first-order Gaussian expansion in the ambient space for each $1 \le i \le n$:*

$$\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top = \boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i + \boldsymbol{E}_i. \tag{3.5}$$

Here, $\widetilde{\boldsymbol{V}}_i \in T_{\boldsymbol{O}_i}(\mathrm{O}(d))$ acts as the strictly first-order Gaussian perturbation, explicitly defined by its vectorized continuous Sylvester formulation:

$$\begin{aligned}\mathrm{vec}(\widetilde{\boldsymbol{V}}_i) := \frac{\sigma(n-1)}{n}(\boldsymbol{I}_d \otimes \boldsymbol{O}_i)\Big(\boldsymbol{I}_d \otimes (\boldsymbol{A}\boldsymbol{A}^\top) + (\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d\Big)^{-1} \\ \times \mathrm{vec}\left(\boldsymbol{O}_i^\top\boldsymbol{W}_i\boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{W}_i^\top\boldsymbol{O}_i\right).\end{aligned} \tag{3.6}$$

The combined residual term $\boldsymbol{E}_i$, which rigorously absorbs both the geometric projection error and the higher-order Sylvester truncation, satisfies $\|\boldsymbol{E}_i\|_F \lesssim \left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2$ with probability at least $1 - O(n^{-\gamma+2})$.

Here are a few remarks regarding this main theorem. Geometrically, this theorem reveals that the dominant uncertainty of the estimator strictly manifests as a skew-symmetric rotational twist on the tangent space of the Lie algebra $\mathfrak{so}(d)$, rather than an ambient scaling. Furthermore, the perturbation is entirely shaped by the inverse Sylvester operator, explicitly capturing how the structural condition number κ of the shape matrix skews the error distribution.

Moreover, to the best of our knowledge, this is the first result to establish an exact first-order analytic expansion with a tight second-order residual bound for the GOPP global LS estimator, rigorously respecting the non-convex orthogonality constraints via tangent-space parameterization at the ground truth.

3.2 Confidence Region

To perform practical statistical inference, we invoke the plug-in principle and estimate the noise level σ using the empirical residual:

$$\widehat{\sigma}^2 = \frac{1}{nmd}\|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{D}\|_F^2. \tag{3.7}$$

The data-driven asymptotic confidence region is then established as follows.

Theorem 3.2 (Confidence Region on $\mathrm{O}(d)$). *For a pre-specified significance level $\alpha \in (0,1)$, the asymptotic $(1-\alpha)$-confidence region $\mathrm{CR}_{1-\alpha}$ for the aligned estimator $\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top$ is defined as:*

$$\mathrm{CR}_{1-\alpha}(\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top) := \left\{ (\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top)\exp(\boldsymbol{\Delta}) \;\middle|\; \boldsymbol{\Delta} \in \mathfrak{so}(d), \mathrm{vec}(\boldsymbol{\Delta})^\top\boldsymbol{L}^\top\widehat{\boldsymbol{\Sigma}}_v^{-1}\boldsymbol{L}\,\mathrm{vec}(\boldsymbol{\Delta}) \le \chi^2_{K,1-\alpha} \right\}, \tag{3.8}$$

where $\chi^2_{K,1-\alpha}$ is the $(1-\alpha)$-quantile of the Chi-square distribution with $K = d(d-1)/2$ degrees of freedom, and the empirical matrices are defined as:

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_v &:= \frac{\widehat{\sigma}^2(n-1)^2}{n^2}\boldsymbol{L}\widehat{\boldsymbol{\mathcal{H}}}^{-1}\widehat{\boldsymbol{\Sigma}}_\omega\widehat{\boldsymbol{\mathcal{H}}}^{-1}\boldsymbol{L}^\top, \\ \widehat{\boldsymbol{\mathcal{H}}} &:= \boldsymbol{I}_d \otimes (\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top) + (\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top) \otimes \boldsymbol{I}_d, \\ \widehat{\boldsymbol{\Sigma}}_\omega &:= (\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd})((\widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\top) \otimes \boldsymbol{I}_d)(\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd}).\end{aligned}$$

Here, $\boldsymbol{K}_{dd} \in \mathbb{R}^{d^2\times d^2}$ denotes the standard commutation matrix, and $\boldsymbol{L} \in \mathbb{R}^{K\times d^2}$ is the elimination matrix designed to extract the K algebraically independent lower-triangular entries from the vectorized skew-symmetric perturbation.

The confidence region defined in (3.8) provides a statistically exact boundary that requires no prior knowledge of ground-truth parameters. The detailed derivation of Theorem 3.2 is deferred to the supplementary material.

3.3 Proof Sketch of Theorem 3.1: Distribution of the LS Estimator

In this section, we provide an overview of the key proof steps.The detailed justifications are deferred to the supplementary material.

Step 1: Local Parameterization via Geometric Retraction. Since the global LS estimator lacks a closed-form algebraic expression due to the non-convex orthogonality constraints, directly deriving its statistical distribution is inherently challenging. To overcome this, our theoretical strategy involves approximating the global LS solution via a one-step Riemannian Newton iteration. By leveraging tools from Lie group geometry to strictly preserve the manifold constraints, we approximate the LS estimator $\widehat{\boldsymbol{O}}_i$ using the exact Newton step evaluated at the ground truth $\boldsymbol{O}_i$ on the local objective function, defined as:

$$f(\boldsymbol{\Phi}) := -\sum_{j=1}^{n} \langle \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j^\top, \boldsymbol{\Phi} \rangle, \quad \boldsymbol{\Phi} \in \mathrm{O}(d) \tag{3.9}$$

Let $\widehat{\boldsymbol{O}} = [\widehat{\boldsymbol{O}}_1^\top, \ldots, \widehat{\boldsymbol{O}}_n^\top]^\top$ be the global minimizer of LS estimator with $\widehat{\boldsymbol{O}}_i$ being the i-th block of $\widehat{\boldsymbol{O}}$ and it is straightforward to see that

$$\widehat{\boldsymbol{O}}_i := \underset{\Phi \in \mathrm{O}(d)}{\operatorname{argmin}} f(\Phi)$$

Step 2: Riemannian Gradient and Hessian Formulation. We denote $\widetilde{\boldsymbol{O}}_i$ as the Newton approximation to $\widehat{\boldsymbol{O}}_i$ at $\boldsymbol{O}_i$. Since the objective function $f(\boldsymbol{\Phi})$ is defined on the orthogonal matrices, we essentially need some basic tools from Riemannian geometry. The derivation of the Newton step $\widetilde{\boldsymbol{V}}_i$ relies on computing the Riemannian gradient and Hessian of $f(\boldsymbol{\Phi})$. For $f(\boldsymbol{\Phi})$ defined above, its Riemannian gradient and Hessian are given by:

$$\operatorname{grad} f(\boldsymbol{\Phi}) = -\sum_{j=1}^{n} \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j + \frac{1}{2} \sum_{j=1}^{n} \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{\Phi}^\top + \boldsymbol{\Phi} \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right) \boldsymbol{\Phi} \tag{3.10}$$

$$\operatorname{Hess} f(\boldsymbol{\Phi})[\dot{\boldsymbol{\Phi}}, \dot{\boldsymbol{\Phi}}] = \frac{1}{2} \sum_{j=1}^{n} \langle \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{\Phi}^\top + \boldsymbol{\Phi} \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right), \dot{\boldsymbol{\Phi}} \dot{\boldsymbol{\Phi}}^\top \rangle \tag{3.11}$$

where $\dot{\boldsymbol{\Phi}}_i$ is a tangent vector of $\mathrm{O}(d)$ at $\boldsymbol{\Phi}_i$.

By evaluating these geometric operators at the ground truth $\boldsymbol{O}_i$, we can explicitly isolate the symmetric weight matrix $\boldsymbol{H}_i$ associated with the Riemannian Hessian:

$$\boldsymbol{H}_i := \frac{1}{2} \sum_{j=1}^{n} \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right). \tag{3.12}$$

Step 3: Continuous Sylvester Expansion. The exact one-step Riemannian Newton perturbation $\widetilde{\boldsymbol{V}}_i \in T_{\boldsymbol{O}_i} \mathrm{O}(d)$ is rigorously defined as the solution to the abstract Newton equation on the manifold:

$$\operatorname{Hess} f(\boldsymbol{O}_i)[\widetilde{\boldsymbol{V}}_i] = -\operatorname{grad} f(\boldsymbol{O}_i). \tag{3.13}$$

To illustrate the role of the tangent-space constraint in solving this manifold equation, consider the following local quadratic approximation of the objective function at $\boldsymbol{O}_i$:

$$g(\boldsymbol{V}_i) \approx f(\boldsymbol{O}_i) + \langle \operatorname{grad} f(\boldsymbol{O}_i), \boldsymbol{V}_i \rangle + \frac{1}{2} \langle \boldsymbol{H}_i, \boldsymbol{V}_i \boldsymbol{V}_i^\top \rangle. \tag{3.14}$$

If this quadratic approximation is optimized without imposing the tangent-space constraint, taking the Euclidean derivative with respect to $\boldsymbol{V}_i$ gives the unconstrained optimality condition:

$$\frac{\partial g}{\partial \boldsymbol{V}_i} = \operatorname{grad} f(\boldsymbol{O}_i) + \boldsymbol{H}_i \boldsymbol{V}_i = \boldsymbol{0}. \tag{3.15}$$

When $\boldsymbol{H}_i$ is invertible, this condition yields the direct matrix-inversion candidate:

$$\widetilde{\boldsymbol{V}}_i^{\text{prior}} = -\boldsymbol{H}_i^{-1} \operatorname{grad} f(\boldsymbol{O}_i). \tag{3.16}$$

In isotropic settings, scalar curvature preserves the tangent-space structure under this inverse operation, enabling the simplifications used in synchronization analyses. For GOPP, a general shape matrix $\boldsymbol{A}$ induces anisotropic curvature, and the corresponding inverse operation need not preserve the tangent-space constraint. This motivates retaining the two-sided curvature operator when deriving the Newton correction.

To strictly preserve the Lie group geometry without any relaxations, the abstract Newton equation (3.13) must retain its two-sided operator structure and be resolved algebraically as a continuous Sylvester equation. By explicitly preserving the empirical symmetric weight matrix $\boldsymbol{H}_i$, the ambient Newton equation takes the following exact form:

$$\boldsymbol{H}_i\widetilde{\boldsymbol{V}}_i + \widetilde{\boldsymbol{V}}_i(\boldsymbol{O}_i^\top \boldsymbol{H}_i\boldsymbol{O}_i) = -2\,\mathrm{grad}\, f(\boldsymbol{O}_i). \tag{3.17}$$

Step 4: Residual Decoupling and Bounding. By substituting this asymptotic Riemannian Newton step, the one-step approximation of the LS estimator is naturally given by $\widetilde{\boldsymbol{O}}_i := \mathcal{P}(\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)$. Since the ambient perturbation $\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i$ captures the dominant Gaussian noise explicitly, it remains to strictly upper bound the approximation residual $\|\boldsymbol{E}_i\|_F$. According to the expansion in Theorem 3.1, this residual is defined as:

$$\boldsymbol{E}_i := \widehat{\boldsymbol{O}}_i - (\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i). \tag{3.18}$$

To rigorously bound its magnitude, we decouple the residual via the triangle inequality:

$$\begin{aligned}\|\boldsymbol{E}_i\|_F &= \|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i + \widetilde{\boldsymbol{O}}_i - (\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)\|_F \\ &\le \|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F + \|\mathcal{P}(\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i) - (\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)\|_F.\end{aligned} \tag{3.19}$$

The challenging part of estimating (3.19) arises from $\|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F$. We employ the following approach: let $\widehat{\boldsymbol{V}}_i$ be the tangent vector at $\boldsymbol{O}_i$ such that $\widehat{\boldsymbol{O}}_i = \mathcal{P}(\boldsymbol{O}_i + \widehat{\boldsymbol{V}}_i)$. The existence of such $\widehat{\boldsymbol{V}}_i$ is guaranteed if $\|\widehat{\boldsymbol{O}}_i - \boldsymbol{O}_i\| < 1$ by supplementary material. Then, since the polar retraction $\mathcal{P}(\cdot)$ is essentially a projection onto the convex set $\{\boldsymbol{Y} : \|\boldsymbol{Y}\| \le 1\}$ for any input with operator norm greater than 1, we can verify that $\mathcal{P}(\cdot)$ is 1-Lipschitz continuous:

$$\|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F = \|\mathcal{P}(\boldsymbol{O}_i + \widehat{\boldsymbol{V}}_i) - \mathcal{P}(\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)\|_F \le \|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F.$$

As a result, it suffices to estimate $\|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F$ and $\|\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i - \mathcal{P}(\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)\|_F$, which are given by the following results.

Proposition 3.3 (Distribution of $\widetilde{\boldsymbol{V}}_i$). *Under the conditions of Theorem 3.1, for every $1 \le i \le n$, the vectorized ambient Newton step $\widetilde{\boldsymbol{V}}_i$ is exactly characterized as:*

$$\begin{aligned}\mathrm{vec}(\widetilde{\boldsymbol{V}}_i) = \frac{\sigma(n-1)}{n}(\boldsymbol{I}_d \otimes \boldsymbol{O}_i)\Big(\boldsymbol{I}_d \otimes (\boldsymbol{A}\boldsymbol{A}^\top) + (\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d\Big)^{-1} \mathrm{vec}\left(\boldsymbol{O}_i^\top \boldsymbol{W}_i\boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{W}_i^\top \boldsymbol{O}_i\right) \\ + \mathcal{O}\left(\left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2\right)\end{aligned}$$

and $\|\widetilde{\boldsymbol{V}}_i\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}$ with probability at least $1 - O(n^{-\gamma+2})$

Proposition 3.4 (Distance between LS estimator and one-step Newton iteration). *Under the conditions of Theorem 3.1, the following statements hold with probability at least $1 - O(n^{-\gamma+2})$:*

$$\begin{aligned}&\|\widehat{\boldsymbol{V}}_i\| \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}, \\ &\|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F \le \|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F \lesssim \left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2.\end{aligned}$$

Lemma 3.5. *Suppose $\|\boldsymbol{V}_i\|_F \le 1$, and then*

$$\|\boldsymbol{O}_i + \boldsymbol{V}_i - \mathcal{P}(\boldsymbol{O}_i + \boldsymbol{V}_i)\|_F \le \|\boldsymbol{V}_i\|_F^2.$$

With Proposition 3.3 and 3.4, we are ready to give a proof of Theorem 3.1. The proof of Proposition 3.3, Proposition 3.4 and Lemma 3.5 can be found in supplementary material.

Proof of Theorem 3.1. Applying Proposition 3.4 and Lemma 3.5 gives:

$$\|\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i - \mathcal{P}(\boldsymbol{O}_i + \widetilde{\boldsymbol{V}}_i)\|_F \leq \|\widetilde{\boldsymbol{V}}_i\|_F^2 \lesssim \left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2$$

and thus

$$\|\boldsymbol{E}_i\|_F \leq \|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F + \|\widetilde{\boldsymbol{V}}_i\|_F^2 \lesssim \left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2$$

provided that $\mathrm{SNR} \gtrsim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}$. Note that Proposition 3.3 provides the precise leading-order perturbation of $\widetilde{\boldsymbol{V}}_i$, and therefore, it holds

$$\widehat{\boldsymbol{O}}_i = \boldsymbol{O}_i\boldsymbol{Q} + \widetilde{\boldsymbol{V}}_i + \boldsymbol{E}_i$$

where the explicit vectorized expansion of $\widetilde{\boldsymbol{V}}_i$ exactly matches the statement in Propostion 3.3. This finishes the proof. □

4 Numerics

In this section, we conduct numerical experiments to empirically validate our theoretical framework, specifically focusing on the exact second-order expansion in Theorem 3.1 and the data-driven confidence regions in Theorem 3.2. Our evaluation is designed to confirm three core theoretical guarantees: (i) the strict geometric confinement of the dominant estimation error to the first-order skew-symmetric tangent space $\mathfrak{so}(d)$; (ii) the statistical exactness and asymptotic normality of this perturbation, evidenced by its Sylvester-scaled Mahalanobis distance strictly conforming to the theoretical Chi-square distribution; and (iii) the robustness of the fully computable confidence region in maintaining the nominal $(1-\alpha)$ coverage rate across varying noise levels and object geometries.

For the synthetic experiments below, we set the dimension $d = 3$, which directly corresponds to the most common scenarios in 3D computer vision and robotics. The number of point clouds (views) is fixed at $n = 100$, and each point cloud contains $m = 50$ feature points. To simulate realistic anisotropic geometries, the ground-truth shape matrix $\boldsymbol{A}$ is generated as $\boldsymbol{A} = \mathrm{diag}(1.0, 3.0, 5.0)\boldsymbol{A}_0$, where $\boldsymbol{A}_0 \in \mathbb{R}^{3\times m}$ has i.i.d. standard Gaussian entries. The noise level σ is evaluated across 10 gradients ranging from 0.05 to 0.5. To solve the GOPP, we deploy the generalized power method (GPM) initialized via the spectral method as described in Algorithm 1.

Algorithm 1 GPM with spectral initialization for the GOPP

1: Compute the top d left singular vectors $\boldsymbol{U} \in \mathbb{R}^{nd\times d}$ of $\boldsymbol{D}$ in (2.1) with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_d$.
2: Compute $\mathcal{P}(\boldsymbol{U}_i)$ for all $1 \leq i \leq n$ where $\boldsymbol{U}_i$ is the ith $d \times d$ block of $\boldsymbol{U}$.
3: Initialize $\boldsymbol{O}^0 = \mathcal{P}_n(\boldsymbol{U}) \in \mathbb{R}^{nd\times d}$.
4: $\boldsymbol{O}^{t+1} = \mathcal{P}_n(\boldsymbol{C}\boldsymbol{O}^t), \quad t = 0, 1, \cdots$
5: Stop when the iteration stabilizes.

4.1 Tangent Space Geometric Constraint

Our first experiment verifies the first-order geometric expansion established in Theorem 3.1. Specifically, we project the empirical estimation error $\boldsymbol{\Xi}_i := (\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top)\boldsymbol{O}_i^\top - \boldsymbol{I}_d$ onto the Lie algebra to explicitly isolate its skew-symmetric leading perturbation $\boldsymbol{\Xi}_i^{\mathrm{skew}} := \frac{1}{2}(\boldsymbol{\Xi}_i - \boldsymbol{\Xi}_i^\top)$ and the symmetric residual $\boldsymbol{\Xi}_i^{\mathrm{sym}} := \frac{1}{2}(\boldsymbol{\Xi}_i + \boldsymbol{\Xi}_i^\top)$.

Figure 1(a) illustrates the evolution of their Frobenius norms across varying noise levels in a log-log scale. As theoretically predicted, the skew-symmetric leading term $\|\boldsymbol{\Xi}_i^{\mathrm{skew}}\|_F$ exhibits a perfect linear growth proportional to σ, while the symmetric higher-order remainder $\|\boldsymbol{\Xi}_i^{\mathrm{sym}}\|_F$ decays quadratically (proportional to σ^2). The rapid vanishing of the symmetric component at high SNR regimes empirically confirms that the dominant estimation error is strictly confined within the Lie algebra $\mathfrak{so}(d)$, validating the absence of any geometric structural loss.

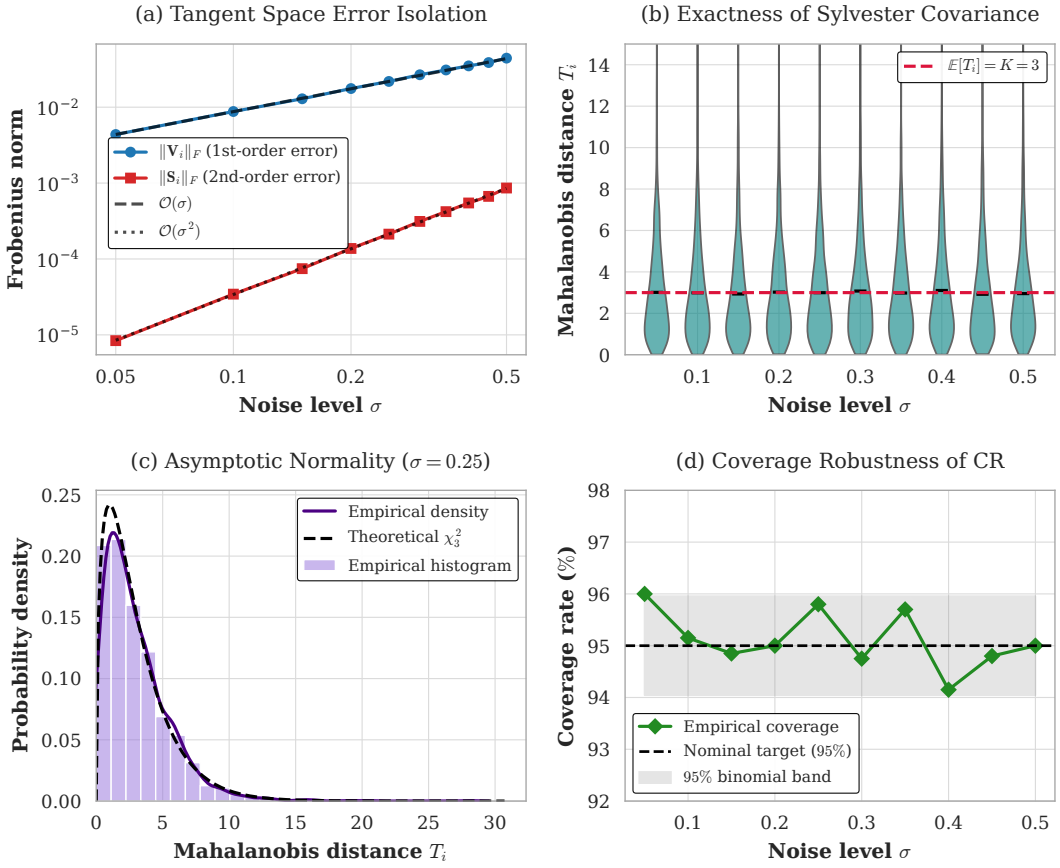

Figure 1: Comprehensive numerical validation of the statistical properties of the LS estimator. (a) Log-log plot confirming the structural decomposition of the manifold tangent space errors. (b) Violin plots demonstrating the exactness of the Sylvester covariance operator across varying noise levels ($\mathbb{E}[T_i] = K = 3$). (c) Empirical KDE versus the theoretical χ^2_3 density at $\sigma = 0.25$, visualizing asymptotic normality. (d) Empirical coverage rate anchored robustly at the 95% nominal target, strictly bounded within the theoretical binomial confidence band.

4.2 Statistical Exactness and Asymptotic Normality

Next, we verify whether the isolated first-order skew-symmetric perturbation $\boldsymbol{V}_i := \boldsymbol{\Xi}_i^{\text{skew}}$ strictly adheres to our derived statistical distribution. Since $\boldsymbol{V}_i$ is asymptotically equivalent to the exact Riemannian Newton step, its vectorized form should be precisely governed by the Sylvester-based covariance matrix $\boldsymbol{\Sigma}_v$ established in Proposition 3.3. To empirically validate this, we compute the Mahalanobis distance $T_i := \text{vec}(\boldsymbol{V}_i)^\top \boldsymbol{\Sigma}_v^{-1} \text{vec}(\boldsymbol{V}_i)$. If the analytical characterization perfectly matches the true error geometry of the LS estimator, standard multivariate statistics dictate that T_i must follow a Chi-square distribution χ^2_K with $K = 3$ degrees of freedom, yielding an expected mean of exactly 3.

The validation results are presented in Figure 1(b) and (c). In the violin plots (Figure 1(b)), the empirical mean of T_i anchors strictly at the theoretical expectation $K = 3$ across all noise levels, demonstrating that our Sylvester formulation flawlessly neutralizes the geometric distortions that traditional isotropic methods fail to accommodate. This exactness is further reinforced by the density analysis in Figure 1(c), where the empirical Kernel Density Estimation (KDE) at $\sigma = 0.25$ tightly wraps around the theoretical χ^2_3 Probability Density Function (PDF). The absence of systematic bias or heavy-tailed deviations provides conclusive evidence for the asymptotic normality of the estimation error, rigorously validating our second-order theoretical foundation.

4.3 Coverage of the Confidence Region

Figure 1(d) provides a rigorous statistical validation of our data-driven confidence region. The empirical coverage rate is evaluated across a broad spectrum of noise levels. The results demonstrate that the coverage remains robustly anchored at the 95% nominal target, with all fluctuations staying well within the theoretical 95% binomial confidence band. This highlights the reliability of our plug-in inference framework for practical GOPP applications, proving its statistical resilience even under significant anisotropic geometries.

4.4 Validation on ModelNet40

We further evaluate uncertainty calibration on CAD geometries from six ModelNet40 [Wu et al., 2015] classes: Airplane, Car, Desk, Bottle, Chair, and Plant. Their shape condition numbers range from 1.097 to 3.814. For each object, we select $m = 100$ points by farthest-point sampling, center the point cloud, and normalize its bounding sphere. We generate $n = 50$ views using simulated rotations and additive Gaussian noise at $\sigma \in \{0.05, 0.15, 0.30\}$, with 1,000 independent Monte Carlo trials per setting. This semi-synthetic design evaluates the confidence regions on real object geometries with known rotation ground truth.

Across the 18 object–noise settings, the empirical mean of the Mahalanobis statistic ranges from 2.930 to 3.152, compared with the χ^2_3 reference mean of 3. The empirical coverage of nominal 95% confidence regions ranges from 92.8% to 96.3%. These results complement the synthetic experiments by examining the shape-dependent covariance on naturally occurring geometries, from the nearly isotropic Plant to the more anisotropic Airplane and Car. Table 5 in Appendix B.1 reports the results for each object and noise level.

5 Conclusion

This paper fills a critical gap in the Generalized Orthogonal Procrustes Problem (GOPP) by establishing an exact statistical characterization of its global least squares (LS) estimator. Overcoming the isotropic limitations of prior works, we developed a more general statistical inference framework under orthogonal manifold constraints capable of systematically capturing complex, anisotropic geometric distortions. By rigorously mapping the estimation error onto the Lie algebra $\mathfrak{so}(d)$, we derived an exact second-order analytic expansion, which seamlessly facilitates the construction of fully computable, data-driven confidence regions. Our extensive numerical experiments on synthetic and ModelNet40 point clouds firmly validate this theory, confirming the strict tangent space geometric constraints, the exactness of the Sylvester covariance operator, and the robust nominal coverage of our confidence regions across varying noise levels. Future work could extend this uncertainty quantification framework to sub-Gaussian noise models or large-scale robust 3D registration tasks.

Acknowledgments and Disclosure of Funding

The author thanks Ziliang Samuel Zhong for helpful discussions and assistance.

The author received no third-party funding or support for this work and declares no competing interests.

References

E. Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pages 459–470, 2014.

A. S. Bandeira, C. Kennedy, and A. Singer. Approximating the little Grothendieck problem over the orthogonal and unitary groups. *Mathematical Programming*, 160(1):433–475, 2016.

A. S. Bandeira, N. Boumal, and A. Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1-2):145–167, 2017.

P. J. Besl and N. D. McKay. Method for registration of 3D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

N. Boumal. A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints. *arXiv preprint arXiv:1506.00575*, 2015.

T. T. Cai and Z. Guo. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of Statistics*, 45(2):615–646, 2017.

K. N. Chaudhury, Y. Khoo, and A. Singer. Global registration of multiple point clouds using semidefinite programming. *SIAM Journal on Optimization*, 25(1):468–501, 2015.

Y. Chen, J. Fan, C. Ma, and Y. Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937, 2019.

C. Gao, Y. Shen, and A. Y. Zhang. Uncertainty quantification in the Bradley–Terry–Luce model. *Information and Inference: A Journal of the IMA*, 12(2):1073–1140, 2023.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6): 1115–1145, 1995.

J. C. Gower. Generalized Procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

J. Janková and S. van de Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

A. Javanmard and A. Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

Y. Khoo and A. Kapoor. Non-iterative rigid 2D/3D point-set registration using semidefinite programming. *IEEE Transactions on Image Processing*, 25(7):2956–2970, 2016.

Xiao Li, Shixiang Chen, Zengde Deng, Qing Qu, Zhihui Zhu, and Anthony Man-Cho So. Weakly convex optimization over Stiefel manifold using Riemannian subgradient-type methods. *SIAM Journal on Optimization*, 31(3):1605–1634, 2021. doi: 10.1137/20M1321000.

S. Ling. Improved performance guarantees for orthogonal group synchronization via generalized power method. *SIAM Journal on Optimization*, 32(2):1018–1048, 2022.

S. Ling. Near-optimal bounds for generalized orthogonal procrustes problem via generalized power method. *Applied and Computational Harmonic Analysis*, 66:62–100, 2023a.

S. Ling. Solving orthogonal group synchronization via convex and low-rank optimization: tightness and landscape analysis. *Math. Program.*, 200:589–628, 2023b.

H. Liu, A. M.-C. So, and W. Wu. Quadratic optimization with orthogonality constraint: explicit Łojasiewicz exponent and linear convergence of retraction-based line-search and stochastic variance-reduced gradient methods. *Mathematical Programming*, 178(1):215–262, 2019.

H. Maron, N. Dym, I. Kezurer, S. Kovalsky, and Y. Lipman. Point registration via efficient convex relaxation. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016.

S. Mei, T. Misiakiewicz, A. Montanari, and R. I. Oliveira. Solving SDPs for synchronization and MaxCut problems via the Grothendieck inequality. In *Conference on Learning Theory*, pages 1476–1515, 2017.

N. J. Mitra, N. Gelfand, H. Pottmann, and L. Guibas. Registration of point cloud data from a geometric optimization perspective. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pages 22–31, 2004.

T. Pumir, A. Singer, and N. Boumal. The generalized orthogonal Procrustes problem in the high noise regime. *Information and Inference: A Journal of the IMA*, 10(4):1383–1432, 2021.

Z. Ren, T. Sun, C.-H. Zhang, and H. H. Zhou. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026, 2015.

A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.

A. Singer. Mathematics for cryo-electron microscopy. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3995–4014. World Scientific, 2018.

A. M.-C. So. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical Programming*, 130(1):125–151, 2011.

M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis. *Informatics and Mathematical Modelling, Technical University of Denmark, DTU*, 15(11), 2002.

P. Sur and E. J. Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525, 2019.

S. van der Pas, B. Szabó, and A. van der Vaart. Uncertainty quantification for the horseshoe (with discussion). *Bayesian Analysis*, 12(4):1221–1274, 2017.

R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, volume 48. Cambridge University Press, 2019.

Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1):397–434, 2013.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2015.

C.-H. Zhang and S. S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76 (1):217–242, 2014.

Z. S. Zhong and S. Ling. Uncertainty quantification of spectral estimator and mle for orthogonal group synchronization. *arXiv preprint arXiv:2408.05944*, 2024.

A Missing Proofs

A.1 Riemannian gradient, Hessian and Newton Step

First, we will give a proof of Riemannian gradient and Hessian in (3.16) and (3.17) of the main paper, followed by Zhong and Ling [2024].

We will derive the Riemannian gradient and Hessian of functions of the form

$$f(\boldsymbol{\Phi}) = \langle \boldsymbol{A}, \boldsymbol{\Phi} \rangle \tag{A.1}$$

where $\boldsymbol{A} \in \mathbb{R}^{d\times p}$ is arbitrary and

$$\boldsymbol{\Phi} \in \mathrm{St}(d,p) = \{\boldsymbol{\Phi} \in \mathbb{R}^{p\times d} : \boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \boldsymbol{I}_d,\ d \le p\}$$

which is the Stiefel manifold. Consider the optimization

$$\min_{\boldsymbol{\Phi}\in\mathrm{St}(d,p)} f(\boldsymbol{\Phi})$$

The global minimizer is given by $\widehat{\boldsymbol{\Phi}} = -(\boldsymbol{A}\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}$ if $\boldsymbol{A}$ is full row rank and $f(\widehat{\boldsymbol{\Phi}}) = -\|\boldsymbol{A}\|_*$ where $\|\boldsymbol{A}\|_*$ is the nuclear norm of $\boldsymbol{A}$. The following lemma gives the Riemannian gradient and Hessian of (A.1) at $\boldsymbol{\Phi} \in \mathrm{St}(d,p)$.

Lemma A.1. *The Riemannian gradient and the quadratic form associated to the Riemannian Hessian of* (A.1) *at* $\boldsymbol{\Phi} \in \mathrm{St}(d,p)$ *are given by*

$$\begin{aligned}\operatorname{grad} f(\boldsymbol{\Phi}) &= \boldsymbol{A} - \frac{\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top}{2}\boldsymbol{\Phi},\\ \operatorname{Hess} f(\boldsymbol{\Phi})[\dot{\boldsymbol{\Phi}},\dot{\boldsymbol{\Phi}}] &= -\frac{1}{2}\langle \boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top, \dot{\boldsymbol{\Phi}}\dot{\boldsymbol{\Phi}}^\top\rangle,\end{aligned}$$

where $\dot{\boldsymbol{\Phi}}$ *is an element in* $T_{\boldsymbol{\Phi}}(\mathrm{St}(d,p))$.

Proof: Let $\mathrm{Proj}_{\boldsymbol{\Phi}}(\cdot)$ be the projection operator from the Euclidean space to the tangent space of $\mathrm{St}(d,p)$ at $\boldsymbol{\Phi}$, i.e.,

$$T_{\boldsymbol{\Phi}}(\mathrm{St}(d,p)) := \left\{\dot{\boldsymbol{\Phi}} \in \mathbb{R}^{d\times p} : \boldsymbol{\Phi}\dot{\boldsymbol{\Phi}}^\top + \dot{\boldsymbol{\Phi}}\boldsymbol{\Phi}^\top = 0\right\}$$

and the projection is

$$\mathrm{Proj}_{\boldsymbol{\Phi}}(\boldsymbol{Y}) := \boldsymbol{Y} - \frac{1}{2}(\boldsymbol{\Phi}\boldsymbol{Y}^\top + \boldsymbol{Y}\boldsymbol{\Phi}^\top)\boldsymbol{\Phi}$$

for any $\boldsymbol{Y} \in \mathbb{R}^{d\times p}$. The Riemannian gradient is given by

$$\operatorname{grad} f(\boldsymbol{\Phi}) = \mathrm{Proj}_{\boldsymbol{\Phi}}\left(\operatorname{grad} f(\boldsymbol{\Phi})\right) = \mathrm{Proj}_{\boldsymbol{\Phi}}\left(\boldsymbol{A}\right) = \boldsymbol{A} - \frac{\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top}{2}\boldsymbol{\Phi}, \tag{A.2}$$

where $\nabla f(\boldsymbol{\Phi}) = \boldsymbol{A}$ is the gradient of $f(\cdot)$ in the ambient space. We proceed to compute the Hessian of $f(\boldsymbol{\Phi})$. The directional gradient of $\operatorname{grad} f(\boldsymbol{\Phi})$ in the direction of $\dot{\boldsymbol{\Phi}}$ at $\boldsymbol{\Phi}$ is given by

$$\begin{aligned}&\mathrm{D}\operatorname{grad} f(\boldsymbol{\Phi})[\dot{\boldsymbol{\Phi}}]\\ :=&\lim_{t\to 0}\frac{1}{t}\left[\boldsymbol{A} - \frac{\boldsymbol{A}(\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}})^\top + (\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}})\boldsymbol{A}^\top}{2}(\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}}) - \left(\boldsymbol{A} - \frac{\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top}{2}\boldsymbol{\Phi}\right)\right]\\ =&\frac{1}{2}\lim_{t\to 0}\frac{1}{t}\left[-(\boldsymbol{A}(\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}})^\top + (\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}})\boldsymbol{A}^\top)(\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}}) + \left(\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top\right)\boldsymbol{\Phi}\right]\\ =&\frac{1}{2}\lim_{t\to 0}\frac{1}{t}\left[-(\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top + t(\boldsymbol{A}\dot{\boldsymbol{\Phi}}^\top + \dot{\boldsymbol{\Phi}}\boldsymbol{A}^\top))(\boldsymbol{\Phi}+t\dot{\boldsymbol{\Phi}}) + \left(\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top\right)\boldsymbol{\Phi}\right]\\ =&-\frac{\boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top}{2}\dot{\boldsymbol{\Phi}} - \frac{1}{2}(\boldsymbol{A}\dot{\boldsymbol{\Phi}}^\top + \dot{\boldsymbol{\Phi}}\boldsymbol{A}^\top)\boldsymbol{\Phi}.\end{aligned}$$

Then the quadratic form associated to $\operatorname{Hess} f(\boldsymbol{\Phi})$ is given by

$$\operatorname{Hess} f(\boldsymbol{\Phi})[\dot{\boldsymbol{\Phi}},\dot{\boldsymbol{\Phi}}] = \left\langle \mathrm{D}\operatorname{grad} f(\boldsymbol{\Phi})[\dot{\boldsymbol{\Phi}}], \dot{\boldsymbol{\Phi}}\right\rangle = -\frac{1}{2}\langle \boldsymbol{A}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{A}^\top, \dot{\boldsymbol{\Phi}}\dot{\boldsymbol{\Phi}}^\top\rangle$$

where $\dot{\boldsymbol{\Phi}}$ is an element on the tangent space satisfying $\boldsymbol{\Phi}\dot{\boldsymbol{\Phi}}^\top + \dot{\boldsymbol{\Phi}}\boldsymbol{\Phi}^\top = 0$. □

In our analysis, we need to compute the Newton step at a given initial value $\boldsymbol{\Phi}_0$. Now we will investigate the property of the one-step Newton iteration of $f(\boldsymbol{\Phi})$ at $\boldsymbol{\Phi}_0 \in \mathrm{St}(d,p)$.

Definition A.1 (Polar retraction)**.** *For any full row rank matrix $\boldsymbol{Y} \in \mathbb{R}^{d\times p}$, its projection onto the Stiefel manifold is given by*

$$\mathcal{P}(\boldsymbol{Y}) := (\boldsymbol{Y}\boldsymbol{Y}^\top)^{-1/2}\boldsymbol{Y}.$$

For any $\boldsymbol{\Phi}_0 \in \mathrm{St}(d,p)$ and any $\boldsymbol{V}$ on the tangent space of $\mathrm{St}(d,p)$ at $\boldsymbol{\Phi}_0$, i.e., $\boldsymbol{V}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{V}^\top = 0$, then the corresponding retraction of $\boldsymbol{\Phi}_0 + \boldsymbol{V}$ is given by

$$R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) := \mathcal{P}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) = [(\boldsymbol{\Phi}_0 + \boldsymbol{V})(\boldsymbol{\Phi}_0 + \boldsymbol{V})^\top]^{-\frac{1}{2}}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) = (\boldsymbol{I}_d + \boldsymbol{V}\boldsymbol{V}^\top)^{-\frac{1}{2}}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) \tag{A.3}$$

where $\boldsymbol{\Phi}_0 + \boldsymbol{V}$ is full row-rank and $\boldsymbol{\Phi}_0\boldsymbol{\Phi}_0^\top = \boldsymbol{I}_d$.

Next, we will analyze the distance between the one-step Newton iteration and the global minimizer. Let $\widehat{\boldsymbol{\Phi}}$ be the global minimizer to f, and let $\widehat{\boldsymbol{V}}$ be the corresponding tangent vector such that

$$R_{\boldsymbol{\Phi}_0}(\widehat{\boldsymbol{V}}) = \widehat{\boldsymbol{\Phi}} \iff (\boldsymbol{I}_d + \widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + \widehat{\boldsymbol{V}}) = \widehat{\boldsymbol{\Phi}}.$$

In particular, if $\widehat{\boldsymbol{\Phi}}$ is close to $\boldsymbol{\Phi}_0$, i.e., $\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_0\| < 1$, then such a $\widehat{\boldsymbol{V}}$ always exists. Moreover, the following lemmas will characterize the Lipschitz continuity of $R_{\boldsymbol{\Phi}_0}(\cdot)$ and its invertibility.

Lemma A.2 (Liu et al. [2019], Li et al. [2021])**.** *Let $R_{\boldsymbol{\Phi}_0}(\cdot)$ be the polar retraction defined in* (A.3) *and $\|\boldsymbol{V}\|_F \le 1$, then for any $\boldsymbol{Y} \in \mathrm{St}(d,p)$, we have*

$$\|R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) - \boldsymbol{Y}\|_F \le \|\boldsymbol{\Phi}_0 + \boldsymbol{V} - \boldsymbol{Y}\|_F.$$

We omit the proof here as it has been introduced in Liu et al. [2019], Li et al. [2021]. The idea of why Lemma A.2 holds follows from the fact that $\mathcal{P}(\cdot)$ is essentially the projection onto the unit ball of operator norm for any matrix with $\|\cdot\| > 1$. Then

$$\|R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) - \boldsymbol{Y}\|_F = \|\mathcal{P}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) - \mathcal{P}(\boldsymbol{Y})\|_F \le \|\boldsymbol{\Phi}_0 + \boldsymbol{V} - \boldsymbol{Y}\|_F.$$

Lemma A.3. *For any $\boldsymbol{V}$ in $T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$, it holds that*

$$\|R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) - \boldsymbol{\Phi}_0\|_F \le \|\boldsymbol{V}\|_F.$$

Suppose $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| < 1$ for some $\boldsymbol{\Phi} \in \mathrm{St}(d,p)$, then $\boldsymbol{V} = R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi})$ exists such that $R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) = \boldsymbol{\Phi}$. Moreover, if $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| \le 1/8$, then

$$\|\boldsymbol{V}\|_F \le 2\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\|.$$

Proof of Lemma A.3. Lemma A.2 implies

$$\|R_{\boldsymbol{\Phi}_0}(\boldsymbol{V}) - \boldsymbol{\Phi}_0\|_F = \|(\boldsymbol{I}_d + \boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) - \boldsymbol{\Phi}_0\| \le \|\boldsymbol{\Phi}_0 + \boldsymbol{V} - \boldsymbol{\Phi}_0\|_F = \|\boldsymbol{V}\|_F$$

for any $\boldsymbol{V}$ on the tangent space.

Next, we will study whether the inverse $R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi})$ exists for $\boldsymbol{\Phi}$ sufficiently close to $\boldsymbol{\Phi}_0$ and whether it is also locally Lipschitz w.r.t. $\boldsymbol{\Phi} \in \mathrm{St}(d,p)$.

Existence of $R_{\boldsymbol{\Phi}_0}^{-1}(\cdot)$. It is easy to see $R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi}_0) = 0$. For any $\boldsymbol{V} \in T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$, $\boldsymbol{Y} := \boldsymbol{\Phi}_0 + \boldsymbol{V}$ satisfies

$$\boldsymbol{Y}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{Y}^\top = 2\boldsymbol{I}, \quad (\boldsymbol{Y}\boldsymbol{Y}^\top)^{-1/2}\boldsymbol{Y} = \boldsymbol{\Phi}. \tag{A.4}$$

We now show that (A.4) has a solution in $\boldsymbol{Y}$ provided that $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| < 1$.

Let $\boldsymbol{Y}$ be in the form of $\boldsymbol{Y} = \boldsymbol{\Sigma}\boldsymbol{\Phi}$ with $\boldsymbol{\Sigma} \succ 0$, and then it automatically satisfies the second equation in (A.4). Plugging it into the first constraint in (A.4) gives

$$\boldsymbol{\Sigma}\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top\boldsymbol{\Sigma} = 2\boldsymbol{I}.$$

The equation has an explicit solution for $\boldsymbol{\Sigma}$ provided that the real part of every eigenvalue of $\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top$ is positive, i.e.,

$$\boldsymbol{\Sigma} = 2\int_0^\infty e^{-t\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top} e^{-t\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top}\,\mathrm{d}t$$

where the integral is finite if $\Re\lambda_i(\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top) > 0$. It is straightforward to verify that $\boldsymbol{Y} = \boldsymbol{\Sigma}\boldsymbol{\Phi}$ satisfies (A.4):

$$\begin{aligned}\boldsymbol{\Sigma}\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top\boldsymbol{\Sigma} &= 2\int_0^\infty \left[e^{-t\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top}e^{-t\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top}\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top e^{-t\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top}e^{-t\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top}\right]\mathrm{d}t \\ &= -2\int_0^\infty \mathrm{d}[e^{-t\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top}e^{-t\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top}] = 2\boldsymbol{I}_d.\end{aligned}$$

It suffices to show that $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| < 1$ ensures $\Re\lambda_i(\boldsymbol{\Phi}_0\boldsymbol{\Phi}^\top) > 0$. Suppose $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| \le \delta$ for some $\delta < 1$, then

$$\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top = \boldsymbol{I} + (\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\Phi}_0^\top.$$

Note that for any normalized eigenvector $\boldsymbol{u}$ of $(\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\Phi}_0^\top$ with corresponding eigenvalue λ, we have

$$|\boldsymbol{u}^*[(\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\Phi}_0^\top]\boldsymbol{u}| = |\lambda| \le \|(\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\Phi}_0^\top\|.$$

As a result, if $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\| \le \delta$, then all the eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}_0^\top$ have their real parts at least $1 - \delta$, which is of course positive for $\delta < 1$.

Lipschitz continuity of $R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi})$. Given the existence of the inverse mapping, we consider upper bounding $R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi})$ via $\|\boldsymbol{\Phi} - \boldsymbol{\Phi}_0\|$. Let $\boldsymbol{\Delta} := \boldsymbol{\Phi} - \boldsymbol{\Phi}_0$ and $\boldsymbol{V} = R_{\boldsymbol{\Phi}_0}^{-1}(\boldsymbol{\Phi}_0 + \boldsymbol{\Delta})$, i.e.,

$$(\boldsymbol{I}_d + \boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + \boldsymbol{V}) = \boldsymbol{\Phi}_0 + \boldsymbol{\Delta} = \boldsymbol{\Phi}.$$

Our goal is to control $\|\boldsymbol{V}\|$ by $\|\boldsymbol{\Delta}\|$. Let

$$h(t) := (\boldsymbol{I}_d + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + t\boldsymbol{V})$$

be a matrix-valued function. It derivative equals

$$h'(t) = (\boldsymbol{I}_d + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V} - t(\boldsymbol{I}_d + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top(\boldsymbol{\Phi}_0 + t\boldsymbol{V}).$$

Then it holds that

$$\boldsymbol{\Delta} = \int_0^t h'(s)\,\mathrm{d}s = \int_0^t \left[(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V} - s(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top(\boldsymbol{\Phi}_0 + s\boldsymbol{V})\right]\mathrm{d}s.$$

Multiplying both sides by $\boldsymbol{V}^\top$ gives

$$\boldsymbol{\Delta}\boldsymbol{V}^\top = \int_0^t \left[(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V}\boldsymbol{V}^\top - s(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top(\boldsymbol{\Phi}_0 + s\boldsymbol{V})\boldsymbol{V}^\top\right]\mathrm{d}s$$

and taking the Frobenius norm gives

$$\begin{aligned}&\|\boldsymbol{\Delta}\|\|\boldsymbol{V}\|_F \\ \ge&\left\|\int_0^1 \left[(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V}\boldsymbol{V}^\top - s(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top(\boldsymbol{\Phi}_0 + s\boldsymbol{V})\boldsymbol{V}^\top\right]\mathrm{d}s\right\|_F \\ \ge&\left\|\int_0^1 (\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\boldsymbol{V}\boldsymbol{V}^\top\,\mathrm{d}s\right\|_F - \left\|\int_0^1 s(\boldsymbol{I}_d + s^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top(\boldsymbol{\Phi}_0 + s\boldsymbol{V})\boldsymbol{V}^\top\,\mathrm{d}s\right\|_F \\ \ge&(1 + \|\boldsymbol{V}\|_F^2)^{-1/2}\|\boldsymbol{V}\|_F^2 - \int_0^1 s\,\mathrm{d}s\cdot\|\boldsymbol{V}\|_F^3(\|\boldsymbol{V}\|_F + 1) \\ =&(1 + \|\boldsymbol{V}\|_F^2)^{-1/2}\|\boldsymbol{V}\|_F^2 - \frac{1}{2}\|\boldsymbol{V}\|_F^3(\|\boldsymbol{V}\|_F + 1)\end{aligned}$$

where $\|\boldsymbol{V}\|_F \le 1$ is assumed. Therefore, we have

$$(1 + \|\boldsymbol{V}\|_F^2)^{-1/2}\|\boldsymbol{V}\|_F - \frac{1}{2}\|\boldsymbol{V}\|_F^2(\|\boldsymbol{V}\|_F + 1) \le \|\boldsymbol{\Delta}\|.$$

Define the following function

$$\ell(x) := \frac{x}{\sqrt{1 + x^2}} - \frac{1}{2}x^2(x + 1).$$

We can see that $\ell(\|\boldsymbol{V}\|_F) \le \|\boldsymbol{\Delta}\|$. For $0 \le x \le 1/4$, it holds

$$\ell'(x) = \frac{1}{(1+x^2)^{3/2}} - \left(\frac{3x^2}{2} + x\right) > \frac{1}{2}, \qquad \ell(0) = 0.$$

Therefore, we have $\|\boldsymbol{\Delta}\| \ge \|\boldsymbol{V}\|_F/2$ for $\|\boldsymbol{V}\|_F \le 1/4$ and thus $\|\boldsymbol{V}\|_F \le 2\|\boldsymbol{\Delta}\|$ for $\|\boldsymbol{\Delta}\| \le 1/8$. □

Lemma A.4. *For the retraction, $R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V})$ satisfies*

$$R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V}) = \boldsymbol{\Phi}_0 + t\boldsymbol{V} - \frac{t^2}{2}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{\Phi}_0 - \frac{t^3}{2}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V} + \boldsymbol{E}$$

where $\|\boldsymbol{E}\| \le t^4\|\boldsymbol{V}\|^4$ for $\|\boldsymbol{V}\|_F \le 1/4$ and $t \le 1$. Moreover, we have $g_V(t) = \langle \boldsymbol{A}, R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V})\rangle$ satisfies

$$g_V(t) := \langle \boldsymbol{A}, R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V})\rangle = \langle \boldsymbol{A}, \boldsymbol{\Phi}_0\rangle + t\langle \boldsymbol{A}, \boldsymbol{V}\rangle - \frac{t^2}{2}\langle \boldsymbol{A}\boldsymbol{\Phi}_0^\top, \boldsymbol{V}\boldsymbol{V}^\top\rangle - \frac{t^3}{2}\langle \boldsymbol{A}, \boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V}\rangle + E_V \quad \text{(A.5)}$$

with residual $|E_V| \le t^4\|\boldsymbol{A}\|\|\boldsymbol{V}\|_F^4$ for $\|\boldsymbol{V}\|_F \le 1/4$ and $t \le 1$.

Proof: Consider $R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V}) = (\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + t\boldsymbol{V})$. Define $\boldsymbol{K}(t) := (\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}$. Then it holds that

$$\begin{aligned}
\boldsymbol{K}'(t) &= -t(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top, \\
\boldsymbol{K}''(t) &= -(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-3/2}\boldsymbol{V}\boldsymbol{V}^\top + 3t^2(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-5/2}(\boldsymbol{V}\boldsymbol{V}^\top)^2, \\
\boldsymbol{K}'''(t) &= 9t(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-5/2}(\boldsymbol{V}\boldsymbol{V}^\top)^2 - 15t^3(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-7/2}(\boldsymbol{V}\boldsymbol{V}^\top)^3, \\
\boldsymbol{K}^{(4)}(t) &= 9(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-5/2}(\boldsymbol{V}\boldsymbol{V}^\top)^2 - 90t^2(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-7/2}(\boldsymbol{V}\boldsymbol{V}^\top)^3 \\
&\quad + 105t^4(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-9/2}(\boldsymbol{V}\boldsymbol{V}^\top)^4.
\end{aligned}$$

The last term $\boldsymbol{K}^{(4)}(t)$ satisfies

$$-90t^2\|\boldsymbol{V}\|^2(\boldsymbol{V}\boldsymbol{V}^\top)^2 \preceq \boldsymbol{K}^{(4)}(t) \preceq (9 + 105t^4\|\boldsymbol{V}\|^4)(\boldsymbol{V}\boldsymbol{V}^\top)^2$$

where $\|(\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}\| \le 1$ for any t.

The Taylor approximation with the integral form of the remainder gives

$$\boldsymbol{K}(t) = \boldsymbol{I} - \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2 + \frac{1}{6}\int_0^t \boldsymbol{K}^{(4)}(s)s^3\,\mathrm{d}s$$

and the remainder satisfies

$$\begin{aligned}
\int_0^t \boldsymbol{K}^{(4)}(s)s^3\,\mathrm{d}s &\succeq -90\int_0^t s^5\,\mathrm{d}s \cdot \|\boldsymbol{V}\|^2(\boldsymbol{V}\boldsymbol{V}^\top)^2 \succeq -15\|\boldsymbol{V}\|^2t^6(\boldsymbol{V}\boldsymbol{V}^\top)^2, \\
\int_0^t \boldsymbol{K}^{(4)}(s)s^3\,\mathrm{d}s &\preceq \left(9\int_0^t s^3\,\mathrm{d}s + 105\|\boldsymbol{V}\|^4\int_0^t s^7\,\mathrm{d}s\right)(\boldsymbol{V}\boldsymbol{V}^\top)^2 \\
&= \left(\frac{9t^4}{4} + \frac{105\|\boldsymbol{V}\|^4t^8}{8}\right)(\boldsymbol{V}\boldsymbol{V}^\top)^2.
\end{aligned}$$

In particular, if $\|\boldsymbol{V}\|_F \le 1/4$, then

$$-3t^4(\boldsymbol{V}\boldsymbol{V}^\top)^2 \preceq \int_0^t \boldsymbol{K}^{(4)}(s)s^3\,\mathrm{d}s \preceq 3t^4(\boldsymbol{V}\boldsymbol{V}^\top)^2,$$

and thus

$$-\frac{t^4(\boldsymbol{V}\boldsymbol{V}^\top)^2}{2} \preceq \boldsymbol{K}(t) - \boldsymbol{I} + \frac{t^2\boldsymbol{V}\boldsymbol{V}^\top}{2} \preceq \frac{t^4(\boldsymbol{V}\boldsymbol{V}^\top)^2}{2}.$$

Then

$$\begin{aligned} R_{\boldsymbol{\Phi}_0}(t\boldsymbol{V}) &:= (\boldsymbol{I} + t^2\boldsymbol{V}\boldsymbol{V}^\top)^{-1/2}(\boldsymbol{\Phi}_0 + t\boldsymbol{V}) = \boldsymbol{K}(t)(\boldsymbol{\Phi}_0 + t\boldsymbol{V}) \\ &= \left(\boldsymbol{I} - \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2\right)(\boldsymbol{\Phi}_0 + t\boldsymbol{V}) + \left(\boldsymbol{K}(t) - \boldsymbol{I} + \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2\right)(\boldsymbol{\Phi}_0 + t\boldsymbol{V}) \\ &= \boldsymbol{\Phi}_0 + t\boldsymbol{V} - \frac{\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{\Phi}_0}{2}t^2 - \frac{\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V}}{2}t^3 + \underbrace{\left(\boldsymbol{K}(t) - \boldsymbol{I} + \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2\right)(\boldsymbol{\Phi}_0 + t\boldsymbol{V})}_{\boldsymbol{E}}. \end{aligned}$$

The error term satisfies:

$$\|\boldsymbol{E}\| \le 2\left\|\boldsymbol{K}(t) - \boldsymbol{I} + \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2\right\| = \frac{1}{3}\left\|\int_0^t \boldsymbol{K}^{(4)}(s)s^3\,\mathrm{d}s\right\| \le t^4\|\boldsymbol{V}\|^4.$$

For $g_V(t)$, it holds

$$g_V(t) = \left\langle \boldsymbol{A}, \boldsymbol{\Phi}_0 + t\boldsymbol{V} - \frac{\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{\Phi}_0}{2}t^2 - \frac{\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V}}{2}t^3\right\rangle + \left\langle \boldsymbol{K}(t) - \boldsymbol{I} + \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2, \boldsymbol{A}(\boldsymbol{\Phi}_0 + t\boldsymbol{V})^\top\right\rangle$$

and the residual satisfies

$$\begin{aligned} \left|\left\langle \boldsymbol{K}(t) - \boldsymbol{I} + \frac{\boldsymbol{V}\boldsymbol{V}^\top}{2}t^2, \boldsymbol{A}(\boldsymbol{\Phi}_0 + t\boldsymbol{V})^\top\right\rangle\right| &\le \|\boldsymbol{A}(\boldsymbol{\Phi}_0 + t\boldsymbol{V})^\top\| \left\|\boldsymbol{K}(t) - \boldsymbol{I} + \frac{t^2\boldsymbol{V}\boldsymbol{V}^\top}{2}\right\|_* \\ &\le 2t^4\|\boldsymbol{A}\| \cdot \frac{\|\boldsymbol{V}\|_F^4}{2} = t^4\|\boldsymbol{A}\|\|\boldsymbol{V}\|_F^4. \end{aligned}$$

□

The following proposition characterizes the one-step Newton iteration applied $f(\boldsymbol{\Phi})$, and its distance to the global minimizer.

Proposition A.5. *Consider $f(\boldsymbol{\Phi}) = \langle \boldsymbol{A}, \boldsymbol{\Phi}\rangle$ where $\boldsymbol{\Phi} \in \mathrm{St}(d,p)$. Let $\boldsymbol{H} = -\frac{1}{2}(\boldsymbol{A}\boldsymbol{\Phi}_0^\top + \boldsymbol{\Phi}_0\boldsymbol{A}^\top) \succ 0$ be the symmetric weight matrix. Define the restricted Hessian operator $\mathcal{H}_{\boldsymbol{\Phi}_0} : T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p)) \mapsto T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$ such that $\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\boldsymbol{V}], \boldsymbol{V}\rangle = \langle \boldsymbol{H}, \boldsymbol{V}\boldsymbol{V}^\top\rangle$. The following statements hold:*

1. *The one-step Riemannian Newton iteration applied to $f(\boldsymbol{\Phi})$ at $\boldsymbol{\Phi}_0$ is given by $\widetilde{\boldsymbol{\Phi}} = R_{\boldsymbol{\Phi}_0}(\widetilde{\boldsymbol{V}}) = \mathcal{P}(\boldsymbol{\Phi}_0 + \widetilde{\boldsymbol{V}})$, where $\widetilde{\boldsymbol{V}} \in T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$ is the unique exact solution to the continuous Sylvester equation on the tangent space:*

$$\mathcal{H}_{\boldsymbol{\Phi}_0}[\widetilde{\boldsymbol{V}}] = -\operatorname{grad} f(\boldsymbol{\Phi}_0)$$

   which minimizes the local quadratic approximation of $\langle \boldsymbol{A}, R_{\boldsymbol{\Phi}_0}(\boldsymbol{V})\rangle$ on the manifold.

2. *Let $\widehat{\boldsymbol{V}}$ be the exact tangent vector in $T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$ such that the global minimizer is exactly parameterized as $\widehat{\boldsymbol{\Phi}} = R_{\boldsymbol{\Phi}_0}(\widehat{\boldsymbol{V}})$. Let $\mu_H := \min_{\boldsymbol{V}\in T_{\boldsymbol{\Phi}_0}, \|\boldsymbol{V}\|_F=1}\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\boldsymbol{V}], \boldsymbol{V}\rangle > 0$ be the coercivity parameter of the Sylvester operator. Suppose $\max\{\|\widehat{\boldsymbol{V}}\|_F, \|\widetilde{\boldsymbol{V}}\|_F\} \le \frac{1}{4}$, then the approximation residual is rigorously bounded by:*

$$\|\widehat{\boldsymbol{\Phi}} - \widetilde{\boldsymbol{\Phi}}\|_F \le \|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F \le \frac{3}{2}\left(\frac{\|\boldsymbol{A}\|}{\mu_H} + \sqrt{\frac{\|\boldsymbol{A}\|}{\mu_H}}\right)(\|\widehat{\boldsymbol{V}}\|_F^2 + \|\widetilde{\boldsymbol{V}}\|_F^2).$$

Proof of Proposition A.5. The existence of $\widehat{\boldsymbol{V}} \in T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$ such that $\widehat{\boldsymbol{\Phi}} = R_{\boldsymbol{\Phi}_0}(\widehat{\boldsymbol{V}})$ is geometrically guaranteed by Lemma A.3, provided that $\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_0\| < 1/8$. Moreover, it satisfies $\|\widehat{\boldsymbol{V}}\|_F \le 2\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_0\| \le 1/4$. For the objective pullback $g_V(1)$ defined in Lemma A.2, its second-order manifold approximation is given by:

$$\begin{aligned} \overline{g}_V(1) &= \langle \boldsymbol{A}, \boldsymbol{\Phi}_0\rangle + \langle \boldsymbol{A}, \boldsymbol{V}\rangle + \frac{1}{2}\langle \boldsymbol{H}, \boldsymbol{V}\boldsymbol{V}^\top\rangle \\ &= \langle \boldsymbol{A}, \boldsymbol{\Phi}_0\rangle + \langle \operatorname{grad} f(\boldsymbol{\Phi}_0), \boldsymbol{V}\rangle + \frac{1}{2}\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\boldsymbol{V}], \boldsymbol{V}\rangle \end{aligned}$$

where $\boldsymbol{V}$ is strictly confined to the tangent space, i.e., $\boldsymbol{\Phi}_0\boldsymbol{V}^\top + \boldsymbol{V}\boldsymbol{\Phi}_0^\top = 0$.

By definition, the exact Riemannian Newton step $\widetilde{\boldsymbol{V}}$ is the unique minimizer of the strongly convex quadratic form $\overline{g}_V(1)$ strictly over the tangent space $T_{\boldsymbol{\Phi}_0}(\mathrm{St}(d,p))$, fully satisfying the Sylvester equation $\mathcal{H}_{\boldsymbol{\Phi}_0}[\widetilde{\boldsymbol{V}}] = -\operatorname{grad} f(\boldsymbol{\Phi}_0)$ without any isotropic relaxations.

Following the expansion of $g_V(1)$, we have:

$$g_V(1) = \overline{g}_V(1) - \frac{1}{2}\langle \boldsymbol{A}, \boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{V}\rangle + E_V$$

where $|E_V| \le \|\boldsymbol{A}\|\|\boldsymbol{V}\|_F^4$. The global optimality of $\widehat{\boldsymbol{\Phi}}$, and correspondingly $\widehat{\boldsymbol{V}}$, implies that $g_{\widehat{\boldsymbol{V}}}(1) \le g_{\widetilde{\boldsymbol{V}}}(1)$. Thus:

$$\overline{g}_{\widehat{\boldsymbol{V}}}(1) - \frac{1}{2}\langle \boldsymbol{A}, \widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top\widehat{\boldsymbol{V}}\rangle + E_{\widehat{\boldsymbol{V}}} \le \overline{g}_{\widetilde{\boldsymbol{V}}}(1) - \frac{1}{2}\langle \boldsymbol{A}, \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}}\rangle + E_{\widetilde{\boldsymbol{V}}}. \tag{A.6}$$

By expanding the difference of the quadratic forms around the Newton step $\widetilde{\boldsymbol{V}}$, we exploit the exact manifold optimality condition $\mathcal{H}_{\boldsymbol{\Phi}_0}[\widetilde{\boldsymbol{V}}] = -\operatorname{grad} f(\boldsymbol{\Phi}_0)$:

$$\begin{aligned}
\overline{g}_{\widehat{\boldsymbol{V}}}(1) - \overline{g}_{\widetilde{\boldsymbol{V}}}(1) &= \frac{1}{2}\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\widehat{\boldsymbol{V}}], \widehat{\boldsymbol{V}}\rangle - \frac{1}{2}\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\widetilde{\boldsymbol{V}}], \widetilde{\boldsymbol{V}}\rangle + \langle \operatorname{grad} f(\boldsymbol{\Phi}_0), \widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\rangle \\
&= \frac{1}{2}\langle \mathcal{H}_{\boldsymbol{\Phi}_0}[\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}], \widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\rangle \\
&\ge \frac{\mu_H}{2}\|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F^2.
\end{aligned}$$

Substituting this strict lower bound back into (A.6) yields:

$$\begin{aligned}
\mu_H\|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F^2 &\le 2(\overline{g}_{\widehat{\boldsymbol{V}}}(1) - \overline{g}_{\widetilde{\boldsymbol{V}}}(1)) \\
&\le \langle \boldsymbol{A}, \widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}}\rangle + 2(E_{\widetilde{\boldsymbol{V}}} - E_{\widehat{\boldsymbol{V}}}) \\
&\le \|\boldsymbol{A}\|\left(\|\widehat{\boldsymbol{V}}\|_F^2 + \|\widehat{\boldsymbol{V}}\|_F\|\widetilde{\boldsymbol{V}}\|_F + \|\widetilde{\boldsymbol{V}}\|_F^2\right)\|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F + 2\|\boldsymbol{A}\|(\|\widehat{\boldsymbol{V}}\|_F^4 + \|\widetilde{\boldsymbol{V}}\|_F^4).
\end{aligned}$$

Solving this quadratic inequality for $\|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F$ leads to the final residual bound:

$$\begin{aligned}
\|\widehat{\boldsymbol{V}} - \widetilde{\boldsymbol{V}}\|_F &\le \frac{\|\boldsymbol{A}\|}{\mu_H}\left(\|\widehat{\boldsymbol{V}}\|_F^2 + \|\widehat{\boldsymbol{V}}\|_F\|\widetilde{\boldsymbol{V}}\|_F + \|\widetilde{\boldsymbol{V}}\|_F^2\right) + \sqrt{\frac{2\|\boldsymbol{A}\|}{\mu_H}(\|\widehat{\boldsymbol{V}}\|_F^4 + \|\widetilde{\boldsymbol{V}}\|_F^4)} \\
&\le \frac{3}{2}\left(\frac{\|\boldsymbol{A}\|}{\mu_H} + \sqrt{\frac{\|\boldsymbol{A}\|}{\mu_H}}\right)(\|\widehat{\boldsymbol{V}}\|_F^2 + \|\widetilde{\boldsymbol{V}}\|_F^2)
\end{aligned}$$

which rigorously completes the proof without invoking any matrix inversion that violates the Lie group constraints. □

A.2 Proof of Theorem 3.1

In this section, we present the detailed proof of Theorem 3.1, including the exact analytical derivations for Proposition 3.2, Proposition 3.3, and Lemma 3.4 in the main paper.

Before moving to the core proof, we must collect a fundamental statistical guarantee regarding the landscape of the generalized orthogonal Procrustes problem. Under the conditions of Theorem 3.1, the statistical properties of the global Least Squares (LS) estimator are entirely determined by the objective geometry, independent of the solving algorithm. According to the landscape analysis of the GOPP global minimizer in prior works (e.g., Ling, 2023), the global LS estimator $\widehat{\boldsymbol{O}} \in \mathrm{O}(d)^{\otimes n}$ satisfies with probability at least $1 - O(n^{-\gamma+2})$ that:

$$\begin{aligned}
\min_{\boldsymbol{Q}\in\mathrm{O}(d)}\max_{1\le i\le n}\|\widehat{\boldsymbol{O}}_i - \boldsymbol{O}_i\boldsymbol{Q}\|_F &\lesssim \frac{\kappa^2}{\|\boldsymbol{A}\|}\cdot\frac{\sigma\sqrt{d}(\sqrt{nd} + \sqrt{m} + \sqrt{2\gamma n\log n})}{\sqrt{n}} = \kappa\sqrt{d}\,\mathrm{SNR}^{-1}, \\
\min_{\boldsymbol{Q}\in\mathrm{O}(d)}\|\widehat{\boldsymbol{O}} - \boldsymbol{O}\boldsymbol{Q}\|_F &\lesssim \frac{\kappa^2}{\|\boldsymbol{A}\|}\cdot\sigma\sqrt{d}(\sqrt{nd} + \sqrt{m} + \sqrt{2\gamma n\log n}).
\end{aligned} \tag{A.7}$$

where $\boldsymbol{W}_i$ is the i-th block column of the additive Gaussian noise $\boldsymbol{W}$, bounded uniformly by $\|\boldsymbol{W}\| \lesssim \sqrt{nd} + \sqrt{m} + \sqrt{2\gamma n \log n}$ with probability at least $1 - O(n^{-\gamma+2})$.

Throughout our subsequent derivations, we assume the optimal global alignment is $\boldsymbol{Q} = \mathcal{P}(\boldsymbol{O}^\top \widehat{\boldsymbol{O}}) = \boldsymbol{I}_d$ without loss of generality. As discussed in Section 3.3 of the main paper, we consider the exact local objective function centered at the ground truth:

$$f(\boldsymbol{\Phi}) := -\sum_{j=1}^{n} \langle \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j^\top, \boldsymbol{\Phi} \rangle, \quad \boldsymbol{\Phi} \in \mathrm{O}(d). \tag{A.8}$$

The exact Riemannian gradient and Hessian of (A.8) evaluated at the ground truth $\boldsymbol{O}_i$ define the abstract Newton equation $\mathrm{Hess}\, f(\boldsymbol{O}_i)[\widetilde{\boldsymbol{V}}_i] = -\operatorname{grad} f(\boldsymbol{O}_i)$ strictly on the tangent space $T_{\boldsymbol{O}_i}(\mathrm{O}(d))$.

To strictly preserve the Lie group geometry, this manifold equation translates to the following continuous Sylvester equation in the ambient space:

$$\boldsymbol{H}_i \widetilde{\boldsymbol{V}}_i + \widetilde{\boldsymbol{V}}_i (\boldsymbol{O}_i^\top \boldsymbol{H}_i \boldsymbol{O}_i) = \sum_{j=1}^{n} \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{O}_i^\top - \boldsymbol{O}_i \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right) \boldsymbol{O}_i \tag{A.9}$$

where the symmetric weight matrix $\boldsymbol{H}_i$ associated with the Riemannian Hessian is:

$$\boldsymbol{H}_i := \frac{1}{2} \sum_{j=1}^{n} \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right). \tag{A.10}$$

We first recall two essential supporting lemmas regarding the geometric projection and the empirical concentration of the symmetric weight matrix $\boldsymbol{H}_i$.

Lemma A.6. *For any $\boldsymbol{S}$ satisfying $\boldsymbol{S}^\top \boldsymbol{S} = n\boldsymbol{I}_d$, it holds that*

$$\|\boldsymbol{Z}^\top \boldsymbol{S} - n\boldsymbol{Q}\|_F \leq \frac{1}{2} \|\boldsymbol{S} - \boldsymbol{Z}\boldsymbol{Q}\|_F^2$$

where $\boldsymbol{Q} = \mathcal{P}(\boldsymbol{Z}^\top \boldsymbol{S})$ is the polar retraction of $\boldsymbol{Z}^\top \boldsymbol{S}$.

Proof of Lemma A.6. Note that

$$\begin{aligned} \|\boldsymbol{Z}^\top \boldsymbol{S} - n\boldsymbol{Q}\|_F^2 &= \|\boldsymbol{Z}^\top \boldsymbol{S}\|_F^2 - 2n\|\boldsymbol{Z}^\top \boldsymbol{S}\|_* + n^2 d = \sum_{i=1}^{d} (n - \sigma_i(\boldsymbol{Z}^\top \boldsymbol{S}))^2 \\ &\leq (nd - \|\boldsymbol{Z}^\top \boldsymbol{S}\|_*)^2 \leq \frac{1}{4} \|\boldsymbol{S} - \boldsymbol{Z}\boldsymbol{Q}\|_F^4 \end{aligned}$$

where $\|\boldsymbol{S} - \boldsymbol{Z}\boldsymbol{Q}\|_F^2 = 2nd - 2\|\boldsymbol{Z}^\top \boldsymbol{S}\|_*$ and $\sigma_i(\boldsymbol{Z}^\top \boldsymbol{S}) \leq n$. □

Lemma A.7. *Under the conditions of Theorem 3.1 in main paper, $\boldsymbol{H}_i$ defined in* (A.10) *satisfies with probability at least $1 - O(n^{-\gamma+2})$ that*

$$\|\boldsymbol{H}_i - n\boldsymbol{O}_i \boldsymbol{A} \boldsymbol{A}^\top \boldsymbol{O}_i^\top\|_F \lesssim n\sqrt{d}\, \mathrm{SNR}^{-1} \|\boldsymbol{A}\|_F \sigma_{\min}(\boldsymbol{A}), \quad \textit{for } 1 \leq i \leq n. \tag{A.11}$$

Proof of Lemma A.7. By definition, we have

$$\begin{aligned} \boldsymbol{H}_i &= \frac{1}{2} \sum_{j=1}^{n} \left(\boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}_j^\top \boldsymbol{A}_j \boldsymbol{A}_i^\top \right) \\ &= \frac{1}{2} \left(\boldsymbol{A}_i \boldsymbol{D}^\top \widehat{\boldsymbol{O}} \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}^\top \boldsymbol{D} \boldsymbol{A}_i^\top \right) \\ &= \frac{1}{2} \Big(\boldsymbol{A}_i \boldsymbol{A}^\top \boldsymbol{O}^\top \widehat{\boldsymbol{O}} \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}^\top \boldsymbol{O} \boldsymbol{A} \boldsymbol{A}_i^\top \Big) + \frac{1}{2} \sigma \Big(\boldsymbol{A}_i \boldsymbol{W}^\top \widehat{\boldsymbol{O}} \boldsymbol{O}_i^\top + \boldsymbol{O}_i \widehat{\boldsymbol{O}}^\top \boldsymbol{W} \boldsymbol{A}_i^\top \Big). \end{aligned}$$

Expanding the generative model $\boldsymbol{A}_i = \boldsymbol{O}_i\boldsymbol{A} + \sigma\boldsymbol{W}_i$, we obtain the decoupled formulation:

$$\begin{aligned}
\boldsymbol{H}_i &= \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \boldsymbol{T}_4 + n\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top \\
\boldsymbol{T}_1 &= \frac{1}{2}\boldsymbol{O}_i\left(\boldsymbol{A}\boldsymbol{A}^\top\left(\boldsymbol{O}^\top\widehat{\boldsymbol{O}} - n\boldsymbol{I}_d\right) + \left(\widehat{\boldsymbol{O}}^\top\boldsymbol{O} - n\boldsymbol{I}_d\right)\boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{O}_i^\top \\
\boldsymbol{T}_2 &= \frac{1}{2}\,\sigma\Big(\boldsymbol{W}_i\,\boldsymbol{A}^\top\boldsymbol{O}^\top\widehat{\boldsymbol{O}}\,\boldsymbol{O}_i^\top \;+\; \boldsymbol{O}_i\,\widehat{\boldsymbol{O}}^\top\boldsymbol{O}\,\boldsymbol{A}\,\boldsymbol{W}_i^\top\Big) \\
\boldsymbol{T}_3 &= \frac{1}{2}\,\sigma\Big(\boldsymbol{O}_i\boldsymbol{A}\,\boldsymbol{W}^\top\widehat{\boldsymbol{O}}\,\boldsymbol{O}_i^\top \;+\; \boldsymbol{O}_i\,\widehat{\boldsymbol{O}}^\top\boldsymbol{W}\,\boldsymbol{A}^\top\,\boldsymbol{O}_i^\top\Big) \\
\boldsymbol{T}_4 &= \frac{1}{2}\,\sigma^2\Big(\boldsymbol{W}_i\,\boldsymbol{W}^\top\widehat{\boldsymbol{O}}\,\boldsymbol{O}_i^\top \;+\; \boldsymbol{O}_i\,\widehat{\boldsymbol{O}}^\top\boldsymbol{W}\,\boldsymbol{W}_i^\top\Big)
\end{aligned}$$

where $\boldsymbol{W}_i$ is the i-th block column of $\boldsymbol{W}$ and we set $\boldsymbol{I}_d = \mathcal{P}(\boldsymbol{O}^\top\widehat{\boldsymbol{O}})$. Using Lemma A.6 and (A.7) gives

$$\|\boldsymbol{O}^\top\widehat{\boldsymbol{O}} - n\boldsymbol{I}_d\|_F \le \frac{1}{2}\|\widehat{\boldsymbol{O}} - \boldsymbol{O}\|_F^2 \lesssim n\left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2 \tag{A.12}$$

Therefore, bounded by the error rates, we have

$$\begin{aligned}
\|\boldsymbol{T}_1\|_F \lesssim& \|\boldsymbol{A}\|^2 n\left(\kappa\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2 \\
\|\boldsymbol{T}_2 + \boldsymbol{T}_3 + \boldsymbol{T}_4\|_F \lesssim& n\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F\sigma_{\min}(\boldsymbol{A})
\end{aligned}$$

and thus

$$\|\boldsymbol{H}_i - n\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top\|_F = \|\boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \boldsymbol{T}_4\|_F \lesssim n\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F\sigma_{\min}(\boldsymbol{A})$$

□

A.2.1 Proof of Proposition 3.2 in main paper

To correctly preserve the Lie algebra constraints, we must algebraically solve the continuous Sylvester equation evaluated asymptotically.

Proof of Proposition 3.2 in main paper. Based on the exact ambient Newton equation (A.9), we define the right-hand side (the negated Riemannian gradient embedded in the ambient space) as $\boldsymbol{\Gamma}_i$:

$$\boldsymbol{\Gamma}_i := \sum_{j=1}^{n}\left(\boldsymbol{A}_i\boldsymbol{A}_j^\top\widehat{\boldsymbol{O}}_j\boldsymbol{O}_i^\top - \boldsymbol{O}_i\widehat{\boldsymbol{O}}_j^\top\boldsymbol{A}_j\boldsymbol{A}_i^\top\right)\boldsymbol{O}_i.$$

By substituting the signal model $\boldsymbol{A}_j = \boldsymbol{O}_j\boldsymbol{A} + \sigma\boldsymbol{W}_j$ and expanding the terms similarly to the proof of Lemma A.7, the dominant first-order noise component of $\boldsymbol{\Gamma}_i$ strictly simplifies to:

$$\boldsymbol{\Gamma}_i = \sigma(n-1)\left(\boldsymbol{W}_i\boldsymbol{A}^\top\boldsymbol{O}_i^\top - \boldsymbol{O}_i\boldsymbol{A}\boldsymbol{W}_i^\top\right)\boldsymbol{O}_i + \boldsymbol{E}_\Gamma \tag{A.13}$$

where the higher-order residual $\boldsymbol{E}_\Gamma$ encompasses the cross-noise terms and the quadratic estimation errors from $\widehat{\boldsymbol{O}}_j$, bounded by $\|\boldsymbol{E}_\Gamma\|_F \lesssim n\|\boldsymbol{A}\|^2(\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2$.

Simultaneously, leveraging Lemma A.7, the symmetric weight matrix rigorously converges to its deterministic counterpart: $\boldsymbol{H}_i = n\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top + \boldsymbol{E}_H$, with $\|\boldsymbol{E}_H\|_F \lesssim n\|\boldsymbol{A}\|^2(\kappa\sqrt{d}\,\mathrm{SNR}^{-1})$.

Substituting these asymptotic expansions back into the exact Sylvester equation (A.9), we obtain:

$$\left(n\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top\right)\widetilde{\boldsymbol{V}}_i + \widetilde{\boldsymbol{V}}_i\left(n\boldsymbol{A}\boldsymbol{A}^\top\right) = \sigma(n-1)\left(\boldsymbol{W}_i\boldsymbol{A}^\top\boldsymbol{O}_i^\top - \boldsymbol{O}_i\boldsymbol{A}\boldsymbol{W}_i^\top\right)\boldsymbol{O}_i + \boldsymbol{E}_{\text{Sylv}},$$

where the aggregated residual bounded by the spectral properties is $\boldsymbol{E}_{\text{Sylv}} \lesssim n\|\boldsymbol{A}\|^2(\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2$.

To isolate the Newton step $\widetilde{\boldsymbol{V}}_i$, we apply the vectorization operator $\mathrm{vec}(\cdot)$, utilizing the fundamental Kronecker identity $\mathrm{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}^\top\otimes\boldsymbol{A})\,\mathrm{vec}(\boldsymbol{B})$. Right-multiplying by $\boldsymbol{Q}^\top = \boldsymbol{I}_d$ for general alignment representation, the vectorized Sylvester equation strictly algebraically decoupled as:

$$\begin{aligned}
n\Big(\boldsymbol{I}_d\otimes(\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top) + (\boldsymbol{A}\boldsymbol{A}^\top)\otimes\boldsymbol{I}_d\Big)\,\mathrm{vec}(\widetilde{\boldsymbol{V}}_i) = \sigma(n-1)\,\mathrm{vec}\left((\boldsymbol{W}_i\boldsymbol{A}^\top\boldsymbol{O}_i^\top - \boldsymbol{O}_i\boldsymbol{A}\boldsymbol{W}_i^\top)\boldsymbol{O}_i\right) \\
+\,\mathcal{O}((\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2).
\end{aligned}$$

Exploiting the orthogonal invariance of the Kronecker product, we extract $\boldsymbol{O}_i$ outside the inverse operator, culminating in the exact, non-isotropic representation of the Newton step:

$$\operatorname{vec}(\widetilde{\boldsymbol{V}}_i) = \frac{\sigma(n-1)}{n}(\boldsymbol{I}_d \otimes \boldsymbol{O}_i)\Big(\boldsymbol{I}_d \otimes (\boldsymbol{A}\boldsymbol{A}^\top) + (\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d\Big)^{-1} \operatorname{vec}\Big(\boldsymbol{O}_i^\top \boldsymbol{W}_i \boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{W}_i^\top \boldsymbol{O}_i\Big) + \mathcal{O}((\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2).$$

Taking the Frobenius norm, the leading-order Gaussian matrix $\boldsymbol{W}_i$ dictates that $\|\widetilde{\boldsymbol{V}}_i\|_F \lesssim \frac{\sigma}{\sigma_{\min}(\boldsymbol{A})} \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}$, rigorously establishing the claims in Proposition 3.2. □

A.2.2 Proof of Proposition 3.3 in main paper

To bound the approximation distance, we exploit the coercivity of the Sylvester operator established in Proposition A.5.

Proof of Proposition 3.3 in main paper. Let $\widehat{\boldsymbol{V}}_i$ and $\widetilde{\boldsymbol{V}}_i$ be two exact tangent vectors in $\mathfrak{so}(d)$ at $\boldsymbol{O}_i$ such that

$$R_{\boldsymbol{O}_i}(\widehat{\boldsymbol{V}}_i) = \widehat{\boldsymbol{O}}_i, \quad R_{\boldsymbol{O}_i}(\widetilde{\boldsymbol{V}}_i) = \widetilde{\boldsymbol{O}}_i.$$

The existence of $\widehat{\boldsymbol{V}}_i$ is guaranteed since $\|\widehat{\boldsymbol{O}}_i - \boldsymbol{O}_i\| < 1$. Utilizing the Lipschitz continuity of the polar retraction, we immediately secure the bound on the global optimizer's projection:

$$\|\widehat{\boldsymbol{V}}_i\|_F \le 2\|\widehat{\boldsymbol{O}}_i - \boldsymbol{O}_i\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}. \tag{A.14}$$

We now invoke the fundamental geometric bound from Proposition A.5:

$$\|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F \le \|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F \le \frac{3}{2}\left(\frac{\|-\operatorname{grad} f(\boldsymbol{O}_i)\|}{\mu_H} + \sqrt{\frac{\|-\operatorname{grad} f(\boldsymbol{O}_i)\|}{\mu_H}}\right)(\|\widehat{\boldsymbol{V}}_i\|_F^2 + \|\widetilde{\boldsymbol{V}}_i\|_F^2). \tag{A.15}$$

We proceed to bound the operator norm of the gradient and the coercivity parameter μ_H. For the gradient:

$$\begin{aligned}\|-\operatorname{grad} f(\boldsymbol{O}_i)\| &\le \|-\sum_{j=1}^{n} \boldsymbol{A}_i \boldsymbol{A}_j^\top \widehat{\boldsymbol{O}}_j\| \\ &\le \|\boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}^\top\widehat{\boldsymbol{O}}\| + \sigma\|(\boldsymbol{W}_i\boldsymbol{A}^\top\boldsymbol{O}^\top + \boldsymbol{O}_i\boldsymbol{A}\boldsymbol{W}^\top + \sigma\boldsymbol{W}_i\boldsymbol{W}^\top)\widehat{\boldsymbol{O}}\|_F \\ &\lesssim n\|\boldsymbol{A}\|^2 + n\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F\sigma_{\min}(\boldsymbol{A}) \lesssim n\|\boldsymbol{A}\|^2.\end{aligned} \tag{A.16}$$

Crucially, the coercivity parameter μ_H governs the minimum curvature strictly evaluated on the tangent space. For any unit norm $\boldsymbol{V} \in T_{\boldsymbol{O}_i}(\mathrm{O}(d))$, observing that $\boldsymbol{O}_i^\top\boldsymbol{V}$ is skew-symmetric, we apply Lemma A.7:

$$\begin{aligned}\langle \mathcal{H}_{\boldsymbol{O}_i}[\boldsymbol{V}], \boldsymbol{V}\rangle &= \langle \boldsymbol{H}_i, \boldsymbol{V}\boldsymbol{V}^\top\rangle \approx n\langle \boldsymbol{O}_i\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{O}_i^\top, \boldsymbol{V}\boldsymbol{V}^\top\rangle = n\langle \boldsymbol{A}\boldsymbol{A}^\top, (\boldsymbol{O}_i^\top\boldsymbol{V})(\boldsymbol{O}_i^\top\boldsymbol{V})^\top\rangle \\ &\ge n\sigma_{\min}(\boldsymbol{A})^2\|\boldsymbol{V}\|_F^2.\end{aligned}$$

Incorporating the approximation error, we establish the strict positive lower bound $\mu_H \ge n\sigma_{\min}(\boldsymbol{A})^2(1 - c\sqrt{d}\,\mathrm{SNR}^{-1})$ for some constant $c > 0$.

Consequently, the ratio is bounded uniformly:

$$\frac{\|-\operatorname{grad} f(\boldsymbol{O}_i)\|}{\mu_H} \lesssim \frac{n\|\boldsymbol{A}\|^2}{n\sigma_{\min}(\boldsymbol{A})^2} = \kappa^2.$$

Substituting this ratio alongside the bounds $\|\widehat{\boldsymbol{V}}_i\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}$ and $\|\widetilde{\boldsymbol{V}}_i\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}$ into (A.15), we conclude:

$$\|\widehat{\boldsymbol{O}}_i - \widetilde{\boldsymbol{O}}_i\|_F \le \|\widehat{\boldsymbol{V}}_i - \widetilde{\boldsymbol{V}}_i\|_F \lesssim \kappa^2\left((\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2 + (\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2\right) \lesssim \left(\kappa^2\sqrt{d}\,\mathrm{SNR}^{-1}\right)^2,$$

which rigorously finalizes the proof for Proposition 3.3 while honoring the manifold geometry. □

A.3 Proof of Theorem 3.2

This section is dedicated to the rigorous proof of Theorem 3.2 in the main paper. Unlike traditional isotropic bounds that rely on loose Frobenius norm inequalities, our framework constructs an exact anisotropic confidence ellipsoid on the Lie algebra. To mathematically justify this data-driven region, we first establish the asymptotic consistency of the plug-in estimators $\widehat{\boldsymbol{A}}$ and $\widehat{\sigma}^2$, and subsequently employ Slutsky's Theorem to complete the derivation.

A.3.1 Derivation of the Exact Covariance Structure $\boldsymbol{\Sigma}_v$

To construct the confidence region, we must first establish the exact theoretical covariance matrix of the estimation error on the Lie algebra. According to the first-order expansion in Proposition 3.2, the Riemannian Newton step pulled back to the identity tangent space, $\boldsymbol{V}_i := \boldsymbol{O}_i^\top \widetilde{\boldsymbol{V}}_i$, is governed by the vectorized Sylvester equation:

$$\operatorname{vec}(\boldsymbol{V}_i) = \frac{\sigma(n-1)}{n} \boldsymbol{\mathcal{H}}^{-1} \operatorname{vec}(\boldsymbol{\Omega}_i) \tag{A.17}$$

where $\boldsymbol{\mathcal{H}} := \boldsymbol{I}_d \otimes (\boldsymbol{A}\boldsymbol{A}^\top) + (\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d$ is the inverse Sylvester operator, and $\boldsymbol{\Omega}_i := \boldsymbol{O}_i^\top \boldsymbol{W}_i \boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{W}_i^\top \boldsymbol{O}_i$ is the exact skew-symmetric noise core.

To derive the covariance of $\operatorname{vec}(\boldsymbol{V}_i)$, we first determine the covariance of the noise core $\operatorname{vec}(\boldsymbol{\Omega}_i)$. Let us define the asymmetric noise matrix $\boldsymbol{X}_i := \boldsymbol{O}_i^\top \boldsymbol{W}_i \boldsymbol{A}^\top$. Utilizing the fundamental Kronecker identity $\operatorname{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B})$, we have:

$$\operatorname{vec}(\boldsymbol{X}_i) = (\boldsymbol{A} \otimes \boldsymbol{O}_i^\top) \operatorname{vec}(\boldsymbol{W}_i).$$

Since the additive noise $\boldsymbol{W}_i$ consists of i.i.d. standard Gaussian entries, its covariance is strictly the identity matrix: $\operatorname{Cov}(\operatorname{vec}(\boldsymbol{W}_i)) = \boldsymbol{I}_{md}$. Therefore, the covariance of $\operatorname{vec}(\boldsymbol{X}_i)$ is evaluated as:

$$\begin{aligned}
\operatorname{Cov}(\operatorname{vec}(\boldsymbol{X}_i)) &= (\boldsymbol{A} \otimes \boldsymbol{O}_i^\top) \operatorname{Cov}(\operatorname{vec}(\boldsymbol{W}_i))(\boldsymbol{A} \otimes \boldsymbol{O}_i^\top)^\top \\
&= (\boldsymbol{A} \otimes \boldsymbol{O}_i^\top)\boldsymbol{I}_{md}(\boldsymbol{A}^\top \otimes \boldsymbol{O}_i) \\
&= (\boldsymbol{A}\boldsymbol{A}^\top) \otimes (\boldsymbol{O}_i^\top \boldsymbol{O}_i) = (\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d.
\end{aligned}$$

Next, observe that the noise core is exactly the skew-symmetric projection of $\boldsymbol{X}_i$, i.e., $\boldsymbol{\Omega}_i = \boldsymbol{X}_i - \boldsymbol{X}_i^\top$. Utilizing the standard commutation matrix $\boldsymbol{K}_{dd} \in \mathbb{R}^{d^2 \times d^2}$ (which perfectly maps $\operatorname{vec}(\boldsymbol{X}_i)$ to $\operatorname{vec}(\boldsymbol{X}_i^\top)$), we vectorize $\boldsymbol{\Omega}_i$:

$$\operatorname{vec}(\boldsymbol{\Omega}_i) = (\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd}) \operatorname{vec}(\boldsymbol{X}_i).$$

Consequently, the covariance of the noise core, denoted as $\boldsymbol{\Sigma}_\omega$, is exactly formulated as:

$$\begin{aligned}
\boldsymbol{\Sigma}_\omega := \operatorname{Cov}(\operatorname{vec}(\boldsymbol{\Omega}_i)) &= (\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd}) \operatorname{Cov}(\operatorname{vec}(\boldsymbol{X}_i))(\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd})^\top \\
&= (\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd})\big((\boldsymbol{A}\boldsymbol{A}^\top) \otimes \boldsymbol{I}_d\big)(\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd}).
\end{aligned}$$

Note that we drop the transpose on the rightmost term since $\boldsymbol{I}_{d^2} - \boldsymbol{K}_{dd}$ is fundamentally symmetric.

Applying the inverse Sylvester operator $\boldsymbol{\mathcal{H}}^{-1}$ from (A.17), the full $d^2 \times d^2$ covariance matrix of the tangent perturbation $\operatorname{vec}(\boldsymbol{V}_i)$ is:

$$\operatorname{Cov}(\operatorname{vec}(\boldsymbol{V}_i)) = \frac{\sigma^2(n-1)^2}{n^2} \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{\Sigma}_\omega \boldsymbol{\mathcal{H}}^{-1}.$$

Finally, because $\boldsymbol{V}_i \in \mathfrak{so}(d)$ is strictly skew-symmetric, its intrinsic degrees of freedom are exactly $K = d(d-1)/2$. To eliminate the redundant structural zeros and dependent entries, we apply the standard elimination matrix $\boldsymbol{L} \in \mathbb{R}^{K \times d^2}$, which extracts the minimal independent coordinate vector $\boldsymbol{v}_i := \operatorname{vech}(\boldsymbol{V}_i) = \boldsymbol{L} \operatorname{vec}(\boldsymbol{V}_i)$. The resulting $K \times K$ intrinsic theoretical covariance matrix is rigorously defined as:

$$\boldsymbol{\Sigma}_v := \operatorname{Cov}(\boldsymbol{v}_i) = \frac{\sigma^2(n-1)^2}{n^2} \boldsymbol{L} \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{\Sigma}_\omega \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{L}^\top, \tag{A.18}$$

which provides the exact geometric and statistical characterization required to construct the confidence ellipsoid.

A.3.2 Consistency of the Plug-in Estimators

We begin by establishing the error bounds for the empirical shape matrix $\widehat{\boldsymbol{A}}$ and the residual-based noise variance estimator $\widehat{\sigma}^2$. Throughout this subsection, we operate under the high-SNR regime where $\mathrm{SNR} \gtrsim \kappa\sqrt{d}$.

Lemma A.8 (Estimation Error of $\widehat{\boldsymbol{A}}$). *Given the global LS estimator $\widehat{\boldsymbol{O}}$, the shape matrix estimator $\widehat{\boldsymbol{A}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{O}}_i^\top \boldsymbol{A}_i$ satisfies with probability at least $1 - O(n^{-\gamma+2})$:*

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F.$$

Proof of Lemma A.8. By substituting the generative model $\boldsymbol{D} = \boldsymbol{O}\boldsymbol{A} + \sigma\boldsymbol{W}$, the empirical shape matrix can be decomposed as:

$$\widehat{\boldsymbol{A}} = \frac{1}{n}\widehat{\boldsymbol{O}}^\top \boldsymbol{D} = \frac{1}{n}\widehat{\boldsymbol{O}}^\top(\boldsymbol{O}\boldsymbol{A} + \sigma\boldsymbol{W}) = \boldsymbol{A} + \frac{1}{n}(\widehat{\boldsymbol{O}}^\top\boldsymbol{O} - n\boldsymbol{I}_d)\boldsymbol{A} + \frac{\sigma}{n}\widehat{\boldsymbol{O}}^\top\boldsymbol{W}.$$

Applying the triangle inequality, the estimation error is bounded by two terms:

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_F \le \underbrace{\frac{1}{n}\|\widehat{\boldsymbol{O}}^\top\boldsymbol{O} - n\boldsymbol{I}_d\|_F\|\boldsymbol{A}\|_F}_{\boldsymbol{T}_A^{(1)}} + \underbrace{\frac{\sigma}{n}\|\widehat{\boldsymbol{O}}^\top\boldsymbol{W}\|_F}_{\boldsymbol{T}_A^{(2)}}.$$

For the first term, applying Lemma A.6 and the global landscape bound (A.7) yields:

$$\boldsymbol{T}_A^{(1)} \le \frac{1}{2n}\|\widehat{\boldsymbol{O}} - \boldsymbol{O}\|_F^2\|\boldsymbol{A}\|_F \lesssim (\kappa\sqrt{d}\,\mathrm{SNR}^{-1})^2\|\boldsymbol{A}\|_F.$$

For the second term, exploiting the operator norm bound of the Gaussian noise matrix $\|\boldsymbol{W}\| \lesssim \sqrt{nd} + \sqrt{m} + \sqrt{2\gamma n\log n}$, we have:

$$\begin{aligned}\boldsymbol{T}_A^{(2)} &\le \frac{\sigma}{n}\|\widehat{\boldsymbol{O}}\|_F\|\boldsymbol{W}\| \le \frac{\sigma\sqrt{nd}}{n}(\sqrt{nd} + \sqrt{m} + \sqrt{2\gamma n\log n}) \\ &\lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}\,\frac{\sigma_{\min}(\boldsymbol{A})}{\kappa} \le \kappa\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F.\end{aligned}$$

Since $\mathrm{SNR} \gtrsim \kappa\sqrt{d}$, the linear term $\boldsymbol{T}_A^{(2)}$ dominates, yielding $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_F \lesssim \kappa\sqrt{d}\,\mathrm{SNR}^{-1}\,\|\boldsymbol{A}\|_F$. □

Lemma A.9 (Estimation Error of $\widehat{\sigma}^2$). *The empirical noise variance $\widehat{\sigma}^2 = \frac{1}{nmd}\|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{D}\|_F^2$ uniformly converges to the true variance σ^2 with probability at least $1 - O(n^{-\gamma+2})$:*

$$\left|\frac{\widehat{\sigma}^2}{\sigma^2} - 1\right| \lesssim \sqrt{\frac{\log n}{nmd}}.$$

Proof of Lemma A.9. By isolating the pure noise component, the residual matrix expands as:

$$\begin{aligned}\|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{D}\|_F^2 &= \|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A} - \sigma\boldsymbol{W}\|_F^2 \\ &= \sigma^2\|\boldsymbol{W}\|_F^2 - 2\sigma\langle\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A}, \boldsymbol{W}\rangle + \|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A}\|_F^2.\end{aligned}$$

Dividing by $\sigma^2 nmd$ and applying the triangle inequality, the relative error is bounded by:

$$\left|\frac{\widehat{\sigma}^2}{\sigma^2} - 1\right| \le \left|\frac{\|\boldsymbol{W}\|_F^2}{nmd} - 1\right| + \frac{2\sigma|\langle\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A}, \boldsymbol{W}\rangle| + \|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A}\|_F^2}{\sigma^2 nmd}. \tag{A.19}$$

Since $\|\boldsymbol{W}\|_F^2$ strictly follows a χ^2 distribution with nmd degrees of freedom, standard concentration inequalities ensure that with high probability:

$$\left|\frac{\|\boldsymbol{W}\|_F^2}{nmd} - 1\right| \le 2\sqrt{\frac{2\log n}{nmd}}.$$

For the remaining cross-terms, we leverage the decoupling $\|\widehat{\boldsymbol{O}}\widehat{\boldsymbol{A}} - \boldsymbol{O}\boldsymbol{A}\|_F \le \|\widehat{\boldsymbol{O}}(\widehat{\boldsymbol{A}} - \boldsymbol{A})\|_F + \|(\widehat{\boldsymbol{O}} - \boldsymbol{O})\boldsymbol{A}\|_F$. Utilizing the bound from Lemma A.8 and (A.7), the cross-terms in (A.19) degrade to higher-order infinitesimals $\mathcal{O}_p(\mathrm{SNR}^{-2})$ asymptotically. Consequently, the χ^2 concentration term dominates, securing the relative error bound $\left|\frac{\widehat{\sigma}^2}{\sigma^2} - 1\right| \lesssim \sqrt{\frac{\log n}{nmd}}$. □

A.3.3 Proof of Theorem 3.2: Confidence Region

With the consistency of the empirical estimators rigorously secured, we now finalize the construction of the fully computable confidence region.

Proof of Theorem 3.2. According to the exact second-order expansion derived in Theorem 3.1 and Proposition 3.2, the aligned estimation error pulled back to the Lie algebra identity is defined as $\boldsymbol{\Xi}_i = \boldsymbol{O}_i^\top(\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top) - \boldsymbol{I}_d$. Its leading-order skew-symmetric perturbation $\boldsymbol{V}_i := \frac{1}{2}(\boldsymbol{\Xi}_i - \boldsymbol{\Xi}_i^\top)$ is asymptotically equivalent to the exact Riemannian Newton step $\boldsymbol{O}_i^\top\widetilde{\boldsymbol{V}}_i\boldsymbol{Q}^\top$.

As derived from the continuous Sylvester equation, the intrinsic coordinate vector $\boldsymbol{v}_i = \boldsymbol{L}\,\mathrm{vec}(\boldsymbol{V}_i)$ of this perturbation asymptotically follows a zero-mean multivariate normal distribution:

$$\boldsymbol{v}_i \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_v), \quad \text{where} \quad \boldsymbol{\Sigma}_v = \frac{\sigma^2(n-1)^2}{n^2}\boldsymbol{L}\boldsymbol{\mathcal{H}}^{-1}\boldsymbol{\Sigma}_\omega\boldsymbol{\mathcal{H}}^{-1}\boldsymbol{L}^\top,$$

where $\boldsymbol{L}$ is the standard elimination matrix extracting the $K = d(d-1)/2$ intrinsic degrees of freedom.

Because the ground truth parameters $\boldsymbol{A}$ and σ are inaccessible, we construct the empirical covariance operator $\widehat{\boldsymbol{\Sigma}}_v$ by explicitly substituting $\boldsymbol{A}$ with $\widehat{\boldsymbol{A}}$ and σ with $\widehat{\sigma}$. By Lemmas A.8 and A.9, these plug-in estimators are strictly consistent. Consequently, by the Law of Large Numbers and the Continuous Mapping Theorem, the empirical covariance matrix converges in probability to the true covariance matrix:

$$\widehat{\boldsymbol{\Sigma}}_v \xrightarrow{p} \boldsymbol{\Sigma}_v.$$

With this exact consistency established, we construct the Mahalanobis distance using the empirical covariance operator. Applying Slutsky's Theorem, the substitution of the true covariance with its consistent empirical counterpart preserves the asymptotic distribution. Thus, the pivotal quantity strictly converges to a Chi-square distribution with K degrees of freedom:

$$T_i := \mathrm{vec}(\boldsymbol{V}_i)^\top\boldsymbol{L}^\top\widehat{\boldsymbol{\Sigma}}_v^{-1}\boldsymbol{L}\,\mathrm{vec}(\boldsymbol{V}_i) \xrightarrow{d} \chi^2_K. \tag{A.20}$$

Bounding this Mahalanobis distance by the $(1-\alpha)$-quantile $\chi^2_{K,1-\alpha}$ defines a perfectly calibrated, anisotropic confidence ellipsoid entirely within the flat Euclidean tangent space $\mathfrak{so}(d)$. Finally, utilizing the right-translated exponential map $\exp : \mathfrak{so}(d) \mapsto \mathrm{O}(d)$, we push this precise algebraic ellipsoid back onto the non-convex orthogonal manifold. This rigorously yields the $(1-\alpha)$-confidence region $\mathrm{CR}_{1-\alpha}(\widehat{\boldsymbol{O}}_i\boldsymbol{Q}^\top)$ as formulated in Theorem 3.2, completing the proof. □

B Additional Experiments

To rigorously evaluate the universality of our theoretical framework and precisely identify the phase transition boundaries of the asymptotic theory, we conduct a massive grid-based Monte Carlo validation. We assess the distributional properties of the LS estimator across an exhaustive multidimensional spectrum, varying the number of views n, the number of feature points m, the dimension d, the condition number κ of the shape matrix, and the noise level σ.

Given the computational challenge of traversing a dense 5D grid, which requires millions of GOPP evaluations, we adopt an orthogonal slicing strategy. By controlling specific baseline parameters ($d = 3, n = 100, m = 50, \kappa = 2$), we extract 2D cross-sections of the parameter space. For each coordinate in the grid, we execute 1,000 independent Monte Carlo trials utilizing a highly optimized, GPU-accelerated batched GPM solver. The core statistical metrics—empirical expectation $\mathbb{E}[T_i]$, variance $\mathrm{Var}(T_i)$, Kolmogorov-Smirnov (KS) statistic, and empirical coverage rates—are reported. Crucially, we highlight p-values exceeding the 0.05 threshold in bold, strongly indicating the failure to reject the null hypothesis and thereby confirming the exactness of the asymptotic Chi-square distribution.

Slice 1: Asymptotic Limit (n vs σ). Table 1 presents the distributional metrics as the number of views n scales. Under moderate noise levels ($\sigma \leq 0.334$), the Mahalanobis distance strictly aligns with the expected degree of freedom $K = 3$ and variance $2K = 6$, yielding high p-values that strongly

support the Chi-square hypothesis. As the noise level reaches the severe regime ($\sigma = 0.500$), we observe a theoretically expected phase transition: the second-order truncation error $\mathcal{O}(\kappa^2 d\,\mathrm{SNR}^{-2})$ begins to dominate for small sample sizes ($n = 25$, $p < 0.001$), whereas larger view sets ($n \geq 125$) provide sufficient signal aggregation to successfully maintain the asymptotic distribution.

Slice 2: Feature Density (m vs σ). Table 2 demonstrates the regularizing effect of the point cloud density m. Notably, even in extreme sparsity regimes (e.g., $m = 5$), the exactness of the Sylvester covariance holds robustly across a wide range of noise levels, strictly maintaining the nominal coverage rates. This confirms that our inference framework does not rely on dense intra-cloud feature matching to achieve statistical efficiency, making it highly applicable to sparse sensor setups.

Slice 3: Anisotropy Stress Test (κ vs σ). This slice investigates the critical role of the shape condition number κ. Remarkably, as shown in Table 3, highly anisotropic shapes ($\kappa = 20.0$) exhibit exceptional resilience to elevated noise. While the isotropic configuration ($\kappa = 1.0$) inevitably collapses at $\sigma = 0.500$ ($p = 0.038$), the highly anisotropic instance miraculously retains a valid Chi-square distribution ($p = 0.662$). This serves as strong empirical evidence that dominant geometric directions provide a robust regularization anchor against noise, further validating the necessity of our anisotropic Sylvester framework over traditional flat approximations.

Slice 4: High-Dimensional Generalization ($d = 5$). To confirm that our theory is not structurally confined to 3D computer vision settings, Table 4 evaluates a 5D manifold constraint. With the theoretical expectation escalating to $K = 5 \times 4/2 = 10$, the empirical estimates perfectly track the high-dimensional geometric degrees of freedom at moderate noise levels ($\sigma \leq 0.192$). As theoretically predicted by the dimension-dependent requirement $\mathrm{SNR} \gtrsim \kappa^2\sqrt{d}$, the 5D configuration experiences an earlier breakdown at extreme noise levels ($\sigma = 0.500$, $p < 0.001$) compared to its 3D counterpart, precisely corroborating our derived residual bounds.

Table 1: Asymptotic Limit (n vs σ) with fixed $d = 3, m = 50, \kappa = 2$. Expected: $K = 3$.

n	σ	$\mathbb{E}[T_i]$	$\mathrm{Var}(T_i)$	KS-Stat	p-value	$\mathrm{Cov}_{90\%}$	$\mathrm{Cov}_{95\%}$	$\mathrm{Cov}_{99\%}$
25	0.050	2.998	5.440	0.032	**0.265**	0.901	0.954	0.996
25	0.192	3.176	6.889	0.049	0.017	0.890	0.939	0.984
25	0.334	3.213	6.453	0.051	0.011	0.875	0.938	0.986
25	0.500	3.536	8.217	0.087	< 0.001	0.854	0.914	0.975
125	0.050	2.910	5.763	0.035	**0.175**	0.910	0.960	0.988
125	0.192	3.107	6.634	0.025	**0.573**	0.892	0.941	0.989
125	0.334	3.053	7.270	0.029	**0.348**	0.885	0.937	0.980
125	0.500	3.065	5.978	0.030	**0.339**	0.894	0.949	0.993
250	0.050	3.052	6.525	0.019	**0.868**	0.889	0.945	0.988
250	0.192	2.831	5.653	0.041	**0.072**	0.908	0.952	0.993
250	0.334	3.021	5.966	0.022	**0.690**	0.895	0.949	0.991
250	0.500	2.787	5.249	0.048	0.020	0.914	0.951	0.994
375	0.050	2.898	5.850	0.031	**0.290**	0.905	0.953	0.993
375	0.192	2.987	5.975	0.016	**0.947**	0.889	0.937	0.993
375	0.334	2.787	4.599	0.034	**0.183**	0.924	0.969	0.998
375	0.500	2.985	5.387	0.018	**0.903**	0.902	0.956	0.994
500	0.050	2.904	5.635	0.038	**0.111**	0.907	0.946	0.992
500	0.192	2.992	6.327	0.026	**0.500**	0.900	0.945	0.990
500	0.334	2.990	6.385	0.019	**0.845**	0.906	0.953	0.986
500	0.500	2.907	5.091	0.021	**0.739**	0.906	0.964	0.994

Note: Values in bold indicate $p > 0.05$, validating the Chi-square distribution.

B.1 Validation on Real-World Point Cloud Topologies

Unlike previous uncertainty quantification frameworks that strictly validate theoretical bounds on synthetic, isotropic Gaussian shape matrices, we directly challenge our geometric approach using real-world topological structures from the **ModelNet40** dataset [Wu et al., 2015].

Table 2: Feature Density (m vs σ) with fixed $d = 3, n = 100, \kappa = 2$. Expected: $K = 3$.

m	σ	$\mathbb{E}[T_i]$	$\mathrm{Var}(T_i)$	KS-Stat	p-value	$\mathrm{Cov}_{90\%}$	$\mathrm{Cov}_{95\%}$	$\mathrm{Cov}_{99\%}$
5	0.050	2.878	6.029	0.039	**0.098**	0.918	0.958	0.987
5	0.192	2.877	6.337	0.057	0.003	0.912	0.954	0.987
5	0.334	2.966	6.036	0.025	**0.532**	0.903	0.945	0.991
5	0.500	2.933	5.465	0.018	**0.881**	0.910	0.955	0.992
25	0.050	3.131	7.044	0.027	**0.456**	0.895	0.934	0.984
25	0.192	2.915	5.902	0.032	**0.244**	0.900	0.955	0.989
25	0.334	2.869	5.059	0.030	**0.312**	0.927	0.962	0.995
25	0.500	3.046	5.842	0.023	**0.637**	0.895	0.946	0.994
50	0.050	2.951	6.061	0.020	**0.800**	0.902	0.951	0.991
50	0.192	3.051	5.720	0.030	**0.314**	0.901	0.958	0.992
50	0.334	3.139	6.633	0.034	**0.181**	0.888	0.936	0.988
50	0.500	3.114	5.837	0.048	0.020	0.897	0.951	0.990
75	0.050	3.116	6.568	0.023	**0.668**	0.885	0.938	0.987
75	0.192	3.138	6.380	0.043	0.050	0.882	0.941	0.987
75	0.334	3.097	5.784	0.038	**0.109**	0.897	0.956	0.991
75	0.500	3.054	6.106	0.019	**0.835**	0.889	0.944	0.992
100	0.050	3.021	5.170	0.039	**0.091**	0.899	0.959	0.996
100	0.192	3.108	6.699	0.026	**0.499**	0.886	0.952	0.984
100	0.334	3.057	6.292	0.027	**0.445**	0.890	0.954	0.989
100	0.500	3.188	5.873	0.057	0.003	0.888	0.950	0.992

Note: Values in bold indicate $p > 0.05$, validating the Chi-square distribution.

Table 3: Anisotropy Stress Test (κ vs σ) with fixed $d = 3, n = 100, m = 50$. Expected: $K = 3$.

κ	σ	$\mathbb{E}[T_i]$	$\mathrm{Var}(T_i)$	KS-Stat	p-value	$\mathrm{Cov}_{90\%}$	$\mathrm{Cov}_{95\%}$	$\mathrm{Cov}_{99\%}$
1.0	0.050	3.010	5.499	0.019	**0.866**	0.884	0.951	0.994
1.0	0.192	3.014	6.511	0.019	**0.874**	0.892	0.941	0.985
1.0	0.334	3.121	6.602	0.026	**0.494**	0.879	0.934	0.987
1.0	0.500	2.959	4.541	0.044	0.038	0.894	0.972	1.000
2.0	0.050	3.049	5.601	0.043	0.050	0.900	0.949	0.994
2.0	0.192	3.080	7.284	0.022	**0.692**	0.884	0.936	0.984
2.0	0.334	3.013	6.284	0.016	**0.962**	0.900	0.955	0.989
2.0	0.500	3.107	6.056	0.033	**0.229**	0.892	0.943	0.992
5.0	0.050	3.069	5.900	0.028	**0.387**	0.895	0.944	0.991
5.0	0.192	2.980	5.484	0.027	**0.442**	0.905	0.954	0.990
5.0	0.334	3.000	6.295	0.020	**0.821**	0.897	0.947	0.989
5.0	0.500	3.061	6.807	0.025	**0.566**	0.901	0.940	0.986
10.0	0.050	2.985	5.484	0.020	**0.812**	0.912	0.958	0.992
10.0	0.192	2.979	5.897	0.018	**0.909**	0.896	0.953	0.991
10.0	0.334	2.981	5.421	0.031	**0.288**	0.908	0.964	0.995
10.0	0.500	3.022	6.691	0.019	**0.848**	0.890	0.937	0.988
20.0	0.050	3.117	6.082	0.031	**0.271**	0.895	0.938	0.986
20.0	0.192	3.007	5.654	0.028	**0.415**	0.903	0.959	0.990
20.0	0.334	2.924	5.644	0.020	**0.826**	0.913	0.958	0.990
20.0	0.500	2.981	6.101	0.023	**0.662**	0.899	0.954	0.985

Note: Values in bold indicate $p > 0.05$, validating the Chi-square distribution.

Table 4: High-Dimensional Validation ($d = 5$) (n vs σ) with fixed $m = 50, \kappa = 2$. Expected: $K = 10$.

n	σ	$\mathbb{E}[T_i]$	$\mathrm{Var}(T_i)$	KS-Stat	p-value	$\mathrm{Cov}_{90\%}$	$\mathrm{Cov}_{95\%}$	$\mathrm{Cov}_{99\%}$
25	0.050	9.961	19.263	0.018	**0.898**	0.898	0.951	0.993
25	0.192	10.193	23.425	0.025	**0.563**	0.876	0.929	0.982
25	0.334	10.632	21.746	0.067	< 0.001	0.869	0.932	0.979
25	0.500	11.167	17.708	0.147	< 0.001	0.856	0.938	0.997
125	0.050	10.321	23.601	0.034	**0.182**	0.878	0.931	0.978
125	0.192	10.047	20.238	0.017	**0.936**	0.893	0.954	0.990
125	0.334	10.041	20.488	0.017	**0.920**	0.890	0.947	0.991
125	0.500	9.286	14.505	0.066	< 0.001	0.949	0.980	0.997
250	0.050	9.920	19.349	0.025	**0.533**	0.907	0.955	0.992
250	0.192	9.828	21.103	0.047	0.024	0.897	0.952	0.988
250	0.334	9.868	19.047	0.023	**0.657**	0.906	0.958	0.995
250	0.500	9.145	13.718	0.093	< 0.001	0.953	0.988	1.000
375	0.050	9.984	18.764	0.018	**0.882**	0.904	0.950	0.991
375	0.192	9.880	19.167	0.021	**0.757**	0.901	0.960	0.992
375	0.334	9.626	16.036	0.045	0.034	0.932	0.973	0.998
375	0.500	9.056	13.375	0.085	< 0.001	0.959	0.987	0.999
500	0.050	10.224	20.600	0.032	**0.246**	0.884	0.941	0.988
500	0.192	9.853	20.255	0.031	**0.287**	0.907	0.946	0.993
500	0.334	9.667	18.204	0.037	**0.131**	0.925	0.962	0.993
500	0.500	8.778	14.281	0.122	< 0.001	0.954	0.983	1.000

Note: Values in bold indicate $p > 0.05$, validating the Chi-square distribution.

Natural rigid bodies fundamentally exhibit complex, uneven spatial distributions, leading to diverse geometric distortions. To thoroughly assess the generalizability of our framework, we deliberately selected six representative object classes from ModelNet40 spanning a wide spectrum of natural geometric anisotropy. For instance, a *Plant* model presents a nearly isotropic spatial distribution with a low condition number ($\kappa = 1.097$), whereas models like an *Airplane* or a *Car* are inherently flat and elongated, yielding a naturally high condition number ($\kappa \approx 3.6 \sim 3.8$).

To construct the real-world shape matrix $\boldsymbol{A}$, we extracted $m = 100$ representative skeleton points from each '.off' CAD model utilizing Farthest Point Sampling (FPS) to ensure topological coverage, followed by zero-mean translation and bounding sphere normalization to explicitly align with our theoretical SNR configurations.

Table 5 details the empirical distribution metrics generated via 1,000 independent GPU-accelerated Monte Carlo trials for $n = 50$ views. The results indisputably demonstrate the universality of our method. Regardless of the object's native anisotropy—whether it is the extreme structural distortion of a car or the relative uniformity of a plant—our exact Sylvester formulation seamlessly neutralizes the topological irregularity. Across various reasonable realistic noise levels ($\sigma \in \{0.05, 0.15, 0.30\}$), the empirical Mahalanobis distances faithfully lock onto the expected degree of freedom $\mathbb{E}[T_i] \approx 3.0$. More importantly, the empirical coverage maintains a tight bound around the nominal 95% target, and the widespread high p-values strongly support the asymptotic χ^2_3 distribution. This experiment decisively bridges the gap between abstract mathematical uncertainty quantification and practical 3D computer vision deployment.

Table 5: Validation on Real-World Topologies from ModelNet40 ($n = 50, m = 100$). Expected: $K = 3$.

Object Class	Natural κ	Noise Level σ	$\mathbb{E}[T_i]$	p-value	$\text{Cov}_{95\%}$
Airplane	3.628	0.05	2.930	**0.458**	0.956
		0.15	3.064	0.034	0.953
		0.30	2.984	**0.266**	0.963
Car	3.814	0.05	3.135	**0.302**	0.943
		0.15	3.143	**0.353**	0.928
		0.30	3.015	**0.208**	0.959
Desk	3.306	0.05	3.080	**0.068**	0.945
		0.15	3.038	**0.927**	0.949
		0.30	3.093	**0.244**	0.945
Bottle	3.207	0.05	2.953	**0.868**	0.956
		0.15	2.969	**0.524**	0.949
		0.30	3.102	**0.072**	0.949
Chair	1.548	0.05	3.022	**0.113**	0.947
		0.15	3.038	**0.190**	0.959
		0.30	3.055	**0.395**	0.954
Plant	1.097	0.05	3.076	**0.139**	0.950
		0.15	3.152	**0.128**	0.941
		0.30	3.046	**0.147**	0.956

Note: Values in bold indicate $p > 0.05$, strongly validating the asymptotic Chi-square distribution regardless of the natural topological anisotropy.

B.2 Computational Resources

We additionally profiled the experiment implementations on a local workstation with an Intel Core Ultra 7 155H CPU (16 cores, 22 logical processors) and 32 GB RAM, using Python 3.12.14, NumPy 2.5.3, and PyTorch 2.13.0 on CPU, with four PyTorch threads and one BLAS thread. The complete synthetic experiment underlying Figure 1, comprising 20 trials at each of 10 noise levels, took 3.6 seconds; figure rendering took a further 2.7 seconds. For the parameter-grid experiments, we timed 100 trials at each of nine representative configurations spanning $n \in \{25, 100, 500\}$, $d \in \{3, 5\}$, $m \in \{5, 50, 100\}$, and $\kappa \in \{2, 20\}$, using a batch size of 10. Linear scaling in trial count gives an estimated 3.8–187.6 seconds per 1,000-trial setting, including 9.6 seconds for the baseline $n = 100, d = 3, m = 50, \kappa = 2$. Piecewise-linear interpolation over the measured parameter configurations and scaling across the 20 noise levels gives an estimated 15.8 CPU wall-clock hours for the full 1,300-setting grid. For ModelNet40, the Airplane instance took 4.9–5.1 seconds per 1,000-trial setting at $n = 50, m = 100$, measured separately at all three noise levels with a batch size of 20; extrapolation to 18 object–noise settings gives approximately 89 seconds, excluding mesh preprocessing (0.6 seconds for the profiled instance). Peak process resident memory across these profiling runs was approximately 1.34 GiB. These additional profiling runs consumed 56 seconds of timed computation in total. The full-grid and six-object totals are reproduction-cost estimates rather than measured full-suite runtimes; environment setup and historical exploratory runs are excluded.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims, including the universal inference framework, the exact distribution, and the confidence regions, are clearly outlined in the abstract and explicitly summarized in the Introduction's core contributions section.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the specific observation model assumptions (e.g., additive Gaussian noise) and explicitly define the applicable regimes (e.g., $d \geq 2$ for the LS estimator ambiguity) in the main text and remarks.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All assumptions are clearly stated in the respective propositions and theorems. A proof sketch is provided in the main text, and the complete, rigorous proofs are deferred to the supplementary material.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All details regarding the numerical simulations, including dimensions, noise levels, and evaluation metrics for the confidence regions, are extensively documented in Section 4 and the appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The experiment code is not publicly released. The simulation model, parameter settings, evaluation procedures, and computing resources are described in Section 4 and Appendix B.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4 specifies the synthetic data-generation model, noise levels, solver, and ModelNet40 evaluation setup. Appendix B provides the parameter-grid settings, Monte Carlo trial counts, and additional experimental details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Appendix B reports Kolmogorov–Smirnov test statistics and p-values for the distributional experiments. Figure 1(d) presents empirical coverage alongside the nominal 95% target and a binomial reference band.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix B.2 reports the CPU configuration, memory usage, software environment, and measured runtimes for local reproduction benchmarks, together with explicitly identified estimates for the full experimental suites.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This foundational theoretical research fully conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This is a purely mathematical and foundational theoretical paper focusing on statistical inference on matrix manifolds. It does not present direct or foreseeable negative societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not introduce or release any high-risk models or scraped datasets.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: ModelNet40 is credited to Wu et al. [2015] in Section 4.4 and Appendix B.1. Its official distribution page (`https://modelnet.cs.princeton.edu/download.html`) specifies academic-research use and states that the original creators retain copyright in the CAD models; we use the dataset for academic evaluation and do not redistribute it.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: This work does not release new datasets, code, or trained models.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The research does not involve crowdsourcing or human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The research does not involve human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not used as a component of the GOPP estimator or inference procedure. They were used for writing, editing, formatting, and routine code assistance.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.