Linear Algebra and its Applications 604 (2020) 21–51

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Local eigenvalue decomposition for embedded Riemannian manifolds

Javier Álvarez-Vizoso [a,*], Michael Kirby [b], Chris Peterson [b]

[a] *Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany*
[b] *Department of Mathematics, Colorado State University, 841 Oval Drive, Fort Collins, CO 80523, USA*

A R T I C L E   I N F O

Article history:
Received 13 November 2019
Accepted 8 June 2020
Available online 10 June 2020
Submitted by J.M. Landsberg

MSC:
53B99
55A07
62H25

Keywords:
Principal component analysis
Local eigenvalue decomposition
Curvature tensor
Manifold learning

A B S T R A C T

Local Principal Component Analysis can be performed over small domains of an embedded Riemannian manifold in order to relate the covariance analysis of the underlying point set with the local extrinsic and intrinsic curvature. We show that the volume of domains on a submanifold of general codimension, determined by the intersection with higher-dimensional cylinders and balls in the ambient space, have asymptotic expansions in terms of the mean and scalar curvatures. Moreover, we propose a generalization of the classical third fundamental form to general submanifolds and prove that the local eigenvalue decomposition (EVD) of the covariance matrices have asymptotic expansions that contain the curvature information encoded by the traces of this tensor. This proves the general correspondence between the local EVD integral invariants and differential-geometric curvature for arbitrary embedded Riemannian submanifolds, found so far for curves and hypersurfaces only. Thus, we establish a key theoretical bridge, via covariance matrices at scale, for potential applications in manifold learning relating

* Corresponding author.
E-mail addresses: javizoso@alumni.colostate.edu (J. Álvarez-Vizoso), kirby@math.colostate.edu (M. Kirby), peterson@math.colostate.edu (C. Peterson).

https://doi.org/10.1016/j.laa.2020.06.006

the statistics of point clouds sampled from Riemannian submanifolds to the underlying geometry.

1. Introduction

Integral invariants based on local principal component analysis have been introduced in the literature, [1], [2,3], [4,5], [6,7], as theoretical tools to perform manifold learning in computer graphics and geometry processing of low-dimensional submanifolds, like curves in the plane and surfaces in space. This approach aims to provide a theoretical link between the statistical covariance analysis of the underlying point-set of a domain and the differential-geometric invariants at a point of the domain inside the manifold. These local domains are usually defined by intersecting the submanifold with a ball in the ambient space to cut out a point-set whose covariance matrix has an eigenvalue decomposition that asymptotically expands with the scale of the ball. The relevance of this analysis lies in the fact that the local eigenvalue decomposition series encode information on the dimension, tangent and normal spaces and, hopefully, the curvature information of the submanifold at the center of the ball.

The integral invariant viewpoint has been developed theoretically and numerically especially in [8], [9], [4,5], [10], [6], [11], in order to process discrete samples of points to determine features and detect shapes at scale, or study descriptor stability with respect to noise [12], [13,14]. Voronoi-based covariance matrices have been also been of interest [15,16]. The discretization and numerical analysis to apply this approach to point clouds is a complementary development to the analytical establishment of this type of correspondences, e.g., see [17,18], [19,20] or [21]. The eigenvalue decomposition of covariance matrices of spherical intersection domains was introduced by [22,23] and [24,25] in order to obtain local adaptive Galerkin bases for the invariant manifold of large-dimensional dynamical systems. However those studies did not develop the second order structure of the local eigenvalue asymptotic series to relate covariance analysis to curvature. The present work precisely accomplishes that goal. We have shown, [26], that for regular curves in $\mathbb{R}^n$ the Frenet-Serret frame is recovered in the scale limit, and ratios of the covariance matrix eigenvalues provide descriptors at scale of the generalized curvatures. In the present work, first introduced in [27], we generalize to embedded Riemannian manifolds of general codimension our previous study of local eigenvalue decomposition for hypersurfaces [28], that followed the theoretical study of surfaces in [13]. We shall introduce a generalization to arbitrary codimension of the classical third fundamental form in order to encapsulate all the curvature information hidden in the local eigenvalues at second order. Our main results show how the eigenvalue decomposition of the covariance of cylindrical and spherical intersection domains has an asymptotic expansion with scale given in terms of the dimension, and the extrinsic and intrinsic curvature, as encoded in the traces of the third fundamental form. When the eigenvalues are different the limit eigenvectors are also shown to converge to a frame of generalized principal directions from these tensors.

The structure of the paper is as follows: in section 2 we define the integral invariants in the context of general Riemannian submanifolds, along with the two types of kernel domains on which we will perform the local *eigenvalue decomposition* (EVD) in the Euclidean case. In section 3, the study of the geometry of submanifolds via the second fundamental form is briefly reviewed and the little known third fundamental form is generalized to submanifolds of general codimension. In section 4 we compute the volume, barycenter and covariance matrix of a cylindrical domain inside an embedded submanifold. In particular, we show that the scaling of the eigenvalues of the covariance matrix singles out the tangent and normal spaces of the manifold at the point by the span of the corresponding limit eigenvectors, and how the next-to-leading order term in the asymptotic series of the eigenvalues is determined by those of traces of the third fundamental form. In particular this gives a closed formula for the integrals appearing in the eigenvalue expansions of [24,25]. In section 5 an analogous analysis is carried out for the domain determined by the intersection of a ball in ambient space with the manifold, which introduces considerable correction terms with respect to the previous case. This leads to an eigenvalue decomposition of the covariance matrix with tangent part given in terms of the Weingarten operator corresponding to the mean curvature normal vector; the normal part coincides with the cylindrical case. Finally, in section 6 we obtain the limit ratios of the eigenvalues in terms of this curvature information, and invert the asymptotic series to get descriptors at scale for the case of hypersurfaces. The cylindrical descriptors complement the spherical ones of [28] due to better error bounds.

These results show how local eigenvalue decomposition can be carried out on an embedded Riemannian submanifold to probe its local geometry. It establishes the most general relationship between the statistical covariance analysis of the underlying point-set of the manifold and the classical differential-geometric curvature tensors, furnishing a conceptual dictionary between covariance eigenvalues and eigenvectors and generalized principal curvatures and principal directions. Potential applications to manifold learning, optimization and geometry processing are promising, e.g. by providing an algorithm to characterize geometric descriptors at scale via this correspondence, cf. [28].

2. Integral invariants of Riemannian submanifolds

In our context, integral invariants are local integrals in ambient-space variables over domains of an n-dimensional submanifold $\mathcal{M} \subset \mathcal{N}$ determined by intersection with spheres or cylinders sitting in $\mathcal{N}$. Two such objects are the volume of the domain and the point in the ambient manifold that represents the center of mass of the region. A more interesting object is the covariance matrix obtained by integrating the relative covariance of the degrees of freedom of the points in the domain, i.e., the products of the coordinates of the points with respect to a chosen frame. In order to get a frame independent invariant, one takes the eigenvalue decomposition of the covariance matrix. Since the kernel domains have a natural scale, e.g., the radius of the sphere, it is useful to think of them as a matrix-valued function of scale at every point. Therefore, these invariants correspond to eigenvalues and eigenvectors that can be interpreted respectively as a set of scalar- and frame-valued functions of scale at every point. We define here these general intersection domains in terms of the exponential map $\exp_p : T_p\mathcal{N} \to \mathcal{N}$, i.e., by using the Riemann normal coordinates of $\mathcal{N}$ around $p \in \mathcal{M}$. Our computations however will be performed in the Euclidean case, $\mathcal{N} = \mathbb{R}^{n+k}$, but these definitions cover possible future extensions to matrix manifolds, e.g., to relate this work to information geometry. Here, r_p denotes the injectivity radius of $\mathcal{N}$, cf. [29].

Definition 2.1. The *spherical component* of radius $\varepsilon \leq r_p$, at a point p of a submanifold $\mathcal{M}$ of a Riemannian manifold $\mathcal{N}$ is the domain given by:

$$D_p(\varepsilon) := \mathcal{M} \cap \{q \in \mathcal{N} : \| \exp_p^{-1}(q)\| \leq \varepsilon \leq r_p\}. \tag{1}$$

In the Euclidean case this is just the set of points of $\mathcal{M}$ inside a Euclidean $(n+k)$-ball centered at p.

An element $\mathbb{V}$ in the Grassmannian $\mathrm{Gr}(m, n+k)$ is an m-dimensional linear subspace of $\mathbb{R}^{n+k}$. Fixing a point and m-dimensional ball inside $\mathbb{V}$, the standard three dimensional cylinder over the xy-plane can be generalized to an $\mathbb{V}$-cylinder by taking all points in the ambient space that project down onto the ball inside $\mathbb{V}$.

Definition 2.2. The *cylindrical component* of radius $\varepsilon \leq r_p$, at a point p of a submanifold $\mathcal{M}$ of a Riemannian manifold $\mathcal{N}$ over the m-plane $\mathbb{V} \in \mathrm{Gr}(m, n+k)$, upon a fixed identification of $T_p\mathcal{N} \cong \mathbb{R}^{n+k}$, is the $\mathbb{V}$-cylinder intersection:

$$\mathrm{Cyl}_p(\varepsilon, \mathbb{V}) := \mathcal{M} \cap \{q \in \mathcal{N} : \|\mathrm{proj}_{\mathbb{V}}(\exp_p^{-1}(q))\| \leq \varepsilon \leq r_p\}, \tag{2}$$

where $\mathrm{proj}_{\mathbb{V}}(\cdot)$ is the orthogonal projection onto $\mathbb{V}$ as a linear subspace of $T_p\mathcal{N}$. We shall write $\mathrm{Cyl}_p(\varepsilon)$ when $\mathbb{V} = T_p\mathcal{M}$ is assumed.

In the Euclidean case this is the set of points of $\mathcal{M}$ whose projection to the linear space $\mathbb{V}$ is inside the n-dimensional ball centered at p. For Riemannian manifolds embedded in Euclidean space, $\mathcal{M} \hookrightarrow \mathcal{N} = \mathbb{R}^{n+k}$, we have $\exp_p^{-1}(q) = \boldsymbol{q} - \boldsymbol{p}$ as usual vectors, and one recovers the common definition of local PCA invariants studied in the literature (e.g., [13]). Thus, we shall compute the following objects for the domains above, $D = D_p(\varepsilon)$ or $\mathrm{Cyl}_p(\varepsilon)$, asymptotically with the scale ε:

Definition 2.3. Let the domain $D \subset \mathbb{R}^{n+k}$ be a subset such that the induced measure dVol on D by restriction of the Euclidean volume form is well-defined, for example a compact subset of a Riemannian submanifold, then the *integral invariants* of D are to be defined as: the volume or mass

$$V(D) = \mathbb{E}[1 \cdot \chi_D(\boldsymbol{X})] = \int_D 1 \, \mathrm{dVol}, \tag{3}$$

the barycenter, or center of mass,

$$\boldsymbol{s}(D) = \mathbb{E}[\boldsymbol{X} \cdot \chi_D(\boldsymbol{X})] = \frac{1}{V(D)} \int_D \boldsymbol{X} \, \mathrm{dVol}, \tag{4}$$

and the eigenvalue decomposition $\{\lambda_\mu(D), \boldsymbol{e}_\mu(D)\}_{\mu=1}^{n+k}$ of the tensor or inertia, or covariance matrix centered at $q \in D$, (where usually $q = \boldsymbol{s}(D)$ is chosen):

$$C_q(D) = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{q}) \otimes (\boldsymbol{X} - \boldsymbol{q})^T \cdot \chi_D(\boldsymbol{X})] = \int_D (\boldsymbol{X} - \boldsymbol{q}) \otimes (\boldsymbol{X} - \boldsymbol{q})^T \, \mathrm{dVol}, \tag{5}$$

where the tensor product is to be understood as the outer product of the components in a chosen basis. $\mathbb{E}$ represents taking the expectation value over all possible $\boldsymbol{X}$ in the ambient space, and χ_D is the set-theoretic characteristic function of D (i.e., 1 if and only if $\boldsymbol{X} \in D$, zero otherwise).

By an *integral invariant descriptor* $F(D)$ of some geometric feature F of a measurable domain D, we mean any approximation of F given in terms of $V(D)$, $\boldsymbol{s}(D)$ and the eigenvalue decomposition $\{\lambda_\mu(D), \boldsymbol{e}_\mu(D)\}_{\mu=1}^{n+k}$ of $C_q(D)$. Our domain D of will possess a natural scale ε determined by the size of the ball or cylinder that defines it, we shall talk about *descriptors at scale*, cf. [28].

Our purpose is to establish the most general correspondence between the local eigenvalue decomposition, in this covariance analysis sense, and differential-geometric curvature. In particular this dictionary is very well represented by the asymptotic relationship of Corollary 6.1, which generalizes to any Riemannian submanifold the formula for curves [26] and hypersurfaces [28]. Thus, this work shall confirm the universality of the idea that ratios of local eigenvalues are proportional to curvature information, a potentially promising approach for manifold learning and geometry processing. We encourage the reader to check the detailed toy example of section 3 of [28], and section 5 of that paper for an introduction to the type of steps needed in the proofs of this type of integrals, which are of the same essence but more involved in the present paper.

3. Third fundamental form of a Riemannian submanifold

For a complete analysis of the geometry of Riemannian submanifolds see [29], [30], [31], [32]. Let $(\mathcal{M}, g)$ be an n-dimensional manifold isometrically embedded in an $(n+k)$-dimensional Riemannian manifold $(\mathcal{N}, \overline{g})$, and let $\nabla, \overline{\nabla}$ be the respective Levi-Civita connections. We shall write $g(\cdot, \cdot) = \langle \cdot, \cdot \rangle$, classically called the *first fundamental form* of $\mathcal{M}$ in $\mathcal{N}$. Then, at any point $p \in \mathcal{M}$ and for any vector $\boldsymbol{y} \in T_p\mathcal{M}$, and vector field

$\boldsymbol{X} \in \Gamma(T\mathcal{M})$, the metric connection of $\mathcal{M}$ is the projection of the metric connection of $\mathcal{N}$: $\nabla_{\boldsymbol{y}}\boldsymbol{X} = (\overline{\nabla}_{\boldsymbol{y}}\boldsymbol{X})^\top$, where $(\,\cdot\,)^\top : T_p\mathcal{N} \to T_p\mathcal{M}$ is the projection to the tangent space. The *second fundamental form* **II** *of* $\mathcal{M}$ *in* $\mathcal{N}$ is defined to be the normal projection of the ambient covariant derivative when acting on vectors fields tangent to $\mathcal{M}$, i.e., denoting $(\,\cdot\,)^\perp : T_p\mathcal{N} \to N_p\mathcal{M}$ for the normal space projection,

$$\mathbf{II}(\boldsymbol{x}, \boldsymbol{y}) = (\overline{\nabla}_{\boldsymbol{y}}\boldsymbol{X})^\perp, \quad \text{i.e.,} \quad \overline{\nabla}_{\boldsymbol{y}}\boldsymbol{X} = \nabla_{\boldsymbol{y}}\boldsymbol{X} + \mathbf{II}(\boldsymbol{x}, \boldsymbol{y}), \tag{6}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in T_p\mathcal{M}$, and $\boldsymbol{X} \in \Gamma(T\mathcal{M})$ such that $\boldsymbol{X}|_p = \boldsymbol{x}$. It is a symmetric bilinear form on the tangent space at every point taking values in the normal space, $\mathbf{II} : T_p\mathcal{M} \otimes T_p\mathcal{M} \to N_p\mathcal{M}$. Fixing a normal vector $\boldsymbol{n} \in N_p\mathcal{M}$, the scalar-valued bilinear form $\langle\, \mathbf{II}(\boldsymbol{x}, \boldsymbol{y}),\, \boldsymbol{n} \,\rangle$ has a corresponding self-adjoint map $\widehat{\boldsymbol{S}}_{\boldsymbol{n}} \in \mathrm{End}(T_p\mathcal{M})$, called the *Weingarten map at* $\boldsymbol{n}$, such that:

$$\langle\, \mathbf{II}(\boldsymbol{x}, \boldsymbol{y}),\, \boldsymbol{n} \,\rangle = \langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\, \boldsymbol{x},\, \boldsymbol{y} \,\rangle = \langle\, \boldsymbol{x},\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\, \boldsymbol{y} \,\rangle. \tag{7}$$

Fixing orthonormal bases $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ of $T_p\mathcal{M}$, and $\{\boldsymbol{n}_j\}_{j=1}^k$ of $N_p\mathcal{M}$, the components of the second fundamental form at point p are:

$$\mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^k \mathrm{II}^j(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\boldsymbol{n}_j = \sum_{j=1}^k \langle\, \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu),\, \boldsymbol{n}_j \,\rangle\, \boldsymbol{n}_j = \sum_{j=1}^k \langle\, \widehat{\boldsymbol{S}}_j\, \boldsymbol{e}_\mu,\, \boldsymbol{e}_\nu \,\rangle\, \boldsymbol{n}_j. \tag{8}$$

The geometric meaning of **II** lies in the fact that the Weingarten map measures the tangential rate of change of normal vectors to $\mathcal{M}$ when moving in tangent directions, cf. [29, Eq. II.2.4], $\widehat{\boldsymbol{S}}_{\boldsymbol{n}}\, \boldsymbol{x} = -(\overline{\nabla}_{\boldsymbol{x}}\boldsymbol{N})^\top$, for any $\boldsymbol{N} \in \Gamma(N\mathcal{M})$ such that $\boldsymbol{N}|_p = \boldsymbol{n}$. From this, [31, Ch. 4, Cor. 9, 10], $\mathbf{II}(\boldsymbol{x}, \boldsymbol{x})$ is to be interpreted as the curve acceleration in $\mathcal{N}$ of a geodesic inside $\mathcal{M}$ at p with tangent velocity $\boldsymbol{x}$. Therefore, **II** naturally measures the extrinsic curvature of the embedding since it represents the forced curving of the straightest lines inside $\mathcal{M}$ due to the bending of $\mathcal{M}$ itself inside $\mathcal{N}$.

The inverse function theorem and [30, Ch. VII, Ex. 3.3] establish the following lemma of fundamental importance for the computations of the present work.

Lemma 3.1. *Let $\mathcal{M}$ be an n-dimensional submanifold of an $(n+k)$-dimensional Riemannian manifold $(\mathcal{N}, g)$, with the induced metric $g|_{\mathcal{M}}$. For any point $p \in \mathcal{M}$ and orthonormal basis $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ of $T_p\mathcal{M}$, it is possible to choose normal coordinates $(y^1, \dots, y^{n+k})$ in $\mathcal{N}$ such that the coordinate tangent vectors at the origin $\mathbf{Y}^1, \dots, \mathbf{Y}^n$ coincide with $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$, and $\mathbf{Y}^{n+1}, \dots, \mathbf{Y}^{n+k}$ are an orthonormal basis $\{\boldsymbol{n}_j\}_{j=1}^k$ of $N_p\mathcal{M}$. Moreover, $\mathcal{M}$ is locally given by a graph manifold $y^1 = x^1, \dots, y^n = x^n, y^{n+1} = f^1(\boldsymbol{x}), \dots, y^{n+k} = f^k(\boldsymbol{x})$, such that the components of the second fundamental form at p can be expressed as:*

$$\mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^k \left[\frac{\partial^2 f^j}{\partial x^\mu \partial x^\nu}(0)\right] \boldsymbol{n}_j. \tag{9}$$

The invariance of the trace of $\mathbf{II}$ for any orthonormal tangent frame $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ leads to the definition of the *mean curvature* vector:

$$\boldsymbol{H} = \sum_{\mu=1}^{n} \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) = \sum_{j=1}^{k} H^j \boldsymbol{n}_j, \qquad \text{where } \; H^j = \sum_{\mu=1}^{n} \mathrm{II}^j(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu). \tag{10}$$

The study of the intrinsic geometry of $(\mathcal{M}, g)$ depends only on the metric and is given in terms of the Riemann curvature tensor: $\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{z} = (\nabla_{\boldsymbol{x}}\nabla_{\boldsymbol{y}} - \nabla_{\boldsymbol{y}}\nabla_{\boldsymbol{x}} - \nabla_{[\boldsymbol{x},\boldsymbol{y}]})\boldsymbol{Z}$, for any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in T_p\mathcal{M}$ and $\boldsymbol{Z} \in \Gamma(T\mathcal{M})$ such that $\boldsymbol{Z}|_p = \boldsymbol{z}$. This fundamental tensor equivalently measures the integrability of parallel transport, geodesic deviation and local flatness. Its traces yield the *Ricci tensor* $\boldsymbol{\mathcal{R}ic}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\mu=1}^{n} \langle \boldsymbol{R}(\boldsymbol{e}_\mu, \boldsymbol{x})\boldsymbol{y},\, \boldsymbol{e}_\mu \rangle = \langle \widehat{\boldsymbol{\mathcal{R}}}\, \boldsymbol{x},\, \boldsymbol{y} \rangle$, and the *scalar curvature*, $\mathcal{R} = \sum_\mu \boldsymbol{\mathcal{R}ic}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu)$. Here, $\widehat{\boldsymbol{\mathcal{R}}} \in \mathrm{End}(T_p\mathcal{M})$ is the *Ricci operator* associated to the Ricci tensor with respect to the metric.

The previous lemma and Gauß Theorema Egregium below establish an expression of the intrinsic curvature as particular combination of products of the local Hessians through the second fundamental form, which along with equation (9) yields a local expression for the Riemann curvature tensor in terms of the local Hessian matrices.

Theorem 3.2 *(Gauß equation). The Riemann curvature tensor of a submanifold $\mathcal{M}$ is related to the curvature $\overline{\boldsymbol{R}}$ of the ambient manifold $\mathcal{N}$ via*

$$\langle \boldsymbol{R}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{z},\, \boldsymbol{w} \rangle = \langle\, \overline{\boldsymbol{R}}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{z},\, \boldsymbol{w} \,\rangle + \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{w}),\, \mathbf{II}(\boldsymbol{y},\boldsymbol{z}) \,\rangle - \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{z}),\, \mathbf{II}(\boldsymbol{y},\boldsymbol{w}) \,\rangle \tag{11}$$

for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \in T_p\mathcal{M}$.

In classical differential geometry, [32, vol. 3], [33], the third fundamental form is a natural object to consider after the first fundamental form, $\mathrm{I}(\boldsymbol{x},\boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, and the second fundamental form $\mathrm{II}(\boldsymbol{x},\boldsymbol{y}) = \langle \widehat{\boldsymbol{S}}\, \boldsymbol{x}, \boldsymbol{y} \rangle$, so it is defined for hypersurfaces, e.g. in [34], as

$$\mathrm{III}(\boldsymbol{x},\boldsymbol{y}) = \langle\, \widehat{\boldsymbol{S}}\, \boldsymbol{x},\, \widehat{\boldsymbol{S}}\, \boldsymbol{y} \,\rangle = \langle \widehat{\boldsymbol{S}}^2 \boldsymbol{x}, \boldsymbol{y} \rangle.$$

However, it does not provide new information since it is completely determined by Gauß equation, [30, vol. 2, prop. 5.2], so inside ambient Euclidean space it is:

$$\langle\, \widehat{\boldsymbol{S}}^2 \boldsymbol{x},\, \boldsymbol{y} \,\rangle = H \langle\, \widehat{\boldsymbol{S}}\, \boldsymbol{x},\, \boldsymbol{y} \,\rangle - \boldsymbol{\mathcal{R}ic}(\boldsymbol{x},\boldsymbol{y}), \tag{12}$$

or, in terms of the Ricci operator, $\widehat{\boldsymbol{S}}^2 = H\widehat{\boldsymbol{S}} - \widehat{\boldsymbol{\mathcal{R}}}$. For a manifold $\mathcal{M}$ of higher codimension k, there are k linearly independent normal vectors at every point and, as mentioned before, the generalized second fundamental form takes values in the normal bundle precisely to reflect this structure in terms of the corresponding Weingarten operators at every normal vector. Therefore, the natural generalization of $\langle\, \widehat{\boldsymbol{S}}\, \boldsymbol{x},\, \widehat{\boldsymbol{S}}\, \boldsymbol{y} \,\rangle$ to this context is

Definition 3.3. The *third fundamental form* of a Riemannian submanifold $\mathcal{M} \subset \mathcal{N}$ is the fourth-rank tensor $\mathbf{III} \in (T_p\mathcal{M}^*)^2 \otimes N_p\mathcal{M}^* \otimes N_p\mathcal{M}$, given at every point $p \in \mathcal{M}$ by

$$\langle\, \mathbf{III}(\boldsymbol{x},\boldsymbol{y})\,\boldsymbol{n},\, \boldsymbol{m}\,\rangle = \langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{x}\,,\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{y}\,\rangle, \tag{13}$$

for any $\boldsymbol{x},\boldsymbol{y} \in T_p\mathcal{M}$, and $\boldsymbol{n},\boldsymbol{m} \in N_p\mathcal{M}$.

At any specific point, and because the Weingarten maps are self-adjoint, the linear operator $\mathbf{III}(\boldsymbol{x},\boldsymbol{y}) \in \mathrm{End}(N_p\mathcal{M})$ is written as the following linear combination, when a particular orthonormal basis $\{\boldsymbol{n}_j\}_{j=1}^k$ of the normal space is fixed and $\boldsymbol{\eta}^j = \overline{g}(\cdot,\boldsymbol{n}_j)$ is the dual basis:

$$\mathbf{III}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i,\,j=1}^{k} \langle\, \widehat{\boldsymbol{S}}_i\,\widehat{\boldsymbol{S}}_j\,\boldsymbol{x}\,,\,\boldsymbol{y}\,\rangle\, \boldsymbol{\eta}^i \otimes \boldsymbol{n}_j. \tag{14}$$

This is due to the linearity of the map $\boldsymbol{n} \mapsto \widehat{\boldsymbol{S}}_{\boldsymbol{n}} : N_p\mathcal{M} \to \mathrm{End}(T_p\mathcal{M})$; if $\boldsymbol{n} = \sum_j n^j \boldsymbol{n}_j$ then

$$\langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{x},\,\boldsymbol{y}\,\rangle = \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{y}),\,\boldsymbol{n}\,\rangle = \sum_{j=1}^{k} n^j \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{y}),\,\boldsymbol{n}_j\rangle = \langle\, \left(\sum_{j=1}^{k} n^j\,\widehat{\boldsymbol{S}}_j\right)\boldsymbol{x},\,\boldsymbol{y}\,\rangle,$$

for all $\boldsymbol{x},\boldsymbol{y} \in T_p\mathcal{M}$.

Let us define the *tangent trace* of a tensor $\boldsymbol{A} \in (T_p\mathcal{M}^*)^2 \otimes N_p\mathcal{M}^* \otimes N_p\mathcal{M}$ as the operator sum of the evaluations at an orthonormal basis $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ of $T_p\mathcal{M}$:

$$\mathrm{tr}_{\,\parallel}\boldsymbol{A} = \sum_{\mu=1}^{n} \boldsymbol{A}(\boldsymbol{e}_\mu,\boldsymbol{e}_\mu)\ \in \mathrm{End}(N_p\mathcal{M}); \tag{15}$$

and let the *normal trace* of such a tensor be the bilinear form

$$\mathrm{tr}_{\perp}\boldsymbol{A} = \sum_{j=1}^{k} \langle\, \mathbf{III}(\cdot,\cdot)\,\boldsymbol{n}_j,\,\boldsymbol{n}_j\,\rangle\ \in (T_p\mathcal{M}^*)^2, \tag{16}$$

for any orthonormal basis $\{\boldsymbol{n}_j\}_{j=1}^k$ of $N_p\mathcal{M}$. These tensors are well-defined since these sums can be easily shown to be independent of the orthonormal basis chosen.

Lemma 3.4. *At any point $p \in \mathcal{M}$, for any $\boldsymbol{x},\boldsymbol{y} \in T_p\mathcal{M}$, and $\boldsymbol{n},\boldsymbol{m} \in N_p\mathcal{M}$, the normal trace of the third fundamental form is*

$$\mathrm{tr}_{\perp}\mathbf{III}(\boldsymbol{x},\boldsymbol{y}) = \sum_{j=1}^{k} \langle\, \widehat{\boldsymbol{S}}_j^{\,2}\,\boldsymbol{x},\,\boldsymbol{y}\,\rangle = \langle\, (\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}} + \overline{\boldsymbol{\mathcal{R}}})\,\boldsymbol{x},\,\boldsymbol{y}\,\rangle, \tag{17}$$

where $\widehat{\boldsymbol{\mathcal{R}}}$ and $\overline{\boldsymbol{\mathcal{R}}}$ are the Ricci operators of $\mathcal{M}$ and $\mathcal{N}$ respectively. In particular, the sum of squares of the Weingarten operators $\widehat{\boldsymbol{S}}_j$, for an orthonormal basis $\{\boldsymbol{n}_j\}_{j=1}^k$ of $N_p\mathcal{M}$, is independent of the basis. The tangent trace of the third fundamental form is a linear operator on $N_p\mathcal{M}$ whose components with respect to the metric are the Frobenius inner products of the corresponding Weingarten operators:

$$\langle\, (\mathrm{tr}_{\,\|}\mathbf{III})\, \boldsymbol{n},\, \boldsymbol{m}\, \rangle = \mathrm{tr}\,(\widehat{\boldsymbol{S}}_{\boldsymbol{n}}\widehat{\boldsymbol{S}}_{\boldsymbol{m}}). \tag{18}$$

The total trace is

$$\mathrm{tr}\,\mathbf{III} = \mathrm{tr}_{\perp}\mathrm{tr}_{\,\|}\mathbf{III} = \|\boldsymbol{H}\|^2 - \mathcal{R} + \overline{\mathcal{R}}. \tag{19}$$

Proof. The normal trace bilinear form has components

$$\begin{aligned}\mathrm{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu) &= \sum_{j=1}^{k}\langle\, \widehat{\boldsymbol{S}}_j\ \boldsymbol{e}_\mu,\ \widehat{\boldsymbol{S}}_j\ \boldsymbol{e}_\nu\,\rangle = \sum_{j=1}^{k}\sum_{\alpha=1}^{n}\langle\, \widehat{\boldsymbol{S}}_j\boldsymbol{e}_\alpha,\ \boldsymbol{e}_\mu\,\rangle\langle\, \widehat{\boldsymbol{S}}_j\boldsymbol{e}_\alpha,\ \boldsymbol{e}_\nu\,\rangle \\ &= \sum_{j=1}^{k}\sum_{\alpha=1}^{n}\mathrm{II}^j(\boldsymbol{e}_\alpha,\boldsymbol{e}_\mu)\mathrm{II}^j(\boldsymbol{e}_\alpha,\boldsymbol{e}_\nu) = \sum_{\alpha=1}^{n}\langle\, \mathbf{II}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\mu),\ \mathbf{II}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\nu)\,\rangle,\end{aligned} \tag{20}$$

that using Gauß equation lead to the corresponding linear operator with respect to the metric:

$$\begin{aligned}\mathrm{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu) &= \sum_{\alpha=1}^{n}\langle\, \mathbf{II}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\alpha),\ \mathbf{II}(\boldsymbol{e}_\nu,\boldsymbol{e}_\mu)\,\rangle + \sum_{\alpha=1}^{n}\langle\, \overline{\boldsymbol{R}}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\nu)\boldsymbol{e}_\mu,\ \boldsymbol{e}_\alpha\,\rangle \\ &\quad - \sum_{\alpha=1}^{n}\langle\, \boldsymbol{R}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\nu)\boldsymbol{e}_\mu,\ \boldsymbol{e}_\alpha\,\rangle \\ &= \langle\, \mathbf{II}(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu),\ \boldsymbol{H}\,\rangle + \overline{\boldsymbol{\mathcal{R}ic}}(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu) - \boldsymbol{\mathcal{R}ic}(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu) \\ &= \langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{H}}\,\boldsymbol{e}_\mu,\ \boldsymbol{e}_\nu\,\rangle + \langle\, \overline{\boldsymbol{\mathcal{R}}}\,\boldsymbol{e}_\mu,\ \boldsymbol{e}_\nu\,\rangle - \langle\, \widehat{\boldsymbol{\mathcal{R}}}\,\boldsymbol{e}_\mu,\ \boldsymbol{e}_\nu\,\rangle.\end{aligned}$$

This is the generalization of the operator of the classical third fundamental form, equation (12). The tangent trace is trivial by definition of trace of a linear operator with respect to the metric and the self-adjointness of the Weingarten operators:

$$\langle\, (\mathrm{tr}_{\,\|}\mathbf{III})\ \boldsymbol{n},\ \boldsymbol{m}\,\rangle = \sum_{\mu=1}^{n}\langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{m}}\widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{e}_\mu,\ \boldsymbol{e}_\mu\,\rangle = (\widehat{\boldsymbol{S}}_{\boldsymbol{m}}, \widehat{\boldsymbol{S}}_{\boldsymbol{n}})_F.$$

In a fixed orthonormal basis this tensor is the linear combination

$$\mathrm{tr}_{\,\|}\mathbf{III} = \sum_{i,\,j=1}^{k}\sum_{\mu=1}^{n}\langle\, \widehat{\boldsymbol{S}}_i\,\widehat{\boldsymbol{S}}_j\ \boldsymbol{e}_\mu\,,\ \boldsymbol{e}_\mu\,\rangle\ \boldsymbol{\eta}^i\otimes\boldsymbol{n}_j = \sum_{i,\,j=1}^{k}\mathrm{tr}\,(\widehat{\boldsymbol{S}}_i\,\widehat{\boldsymbol{S}}_j)\ \boldsymbol{\eta}^i\otimes\boldsymbol{n}_j,$$

whose components can be expressed in terms of the second fundamental form as

$$\operatorname{tr}(\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_j) = \sum_{\mu,\,\nu=1}^{n} \langle\, \widehat{\boldsymbol{S}}_i\, \boldsymbol{e}_\mu,\, \boldsymbol{e}_\nu \,\rangle\langle\, \widehat{\boldsymbol{S}}_j\, \boldsymbol{e}_\mu,\, \boldsymbol{e}_\nu \,\rangle = \sum_{\mu,\,\nu=1}^{n} \mathrm{II}^i(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\mathrm{II}^j(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu). \tag{21}$$

Taking the total trace of **III** is analogous to the complete contraction of the Riemann curvature tensor indices to obtain the scalar curvature:

$$\begin{aligned}\operatorname{tr}\mathbf{III} = \operatorname{tr}_{\parallel}\operatorname{tr}_{\perp}\mathbf{III} &= \sum_{\mu=1}^{n}\langle\, (\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}} + \overline{\boldsymbol{\mathcal{R}}})\boldsymbol{e}_\mu,\, \boldsymbol{e}_\mu \,\rangle = \operatorname{tr}\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \operatorname{tr}\widehat{\boldsymbol{\mathcal{R}}} + \operatorname{tr}\overline{\boldsymbol{\mathcal{R}}} \\ &= \sum_{\mu=1}^{n} \operatorname{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) = \sum_{\alpha,\,\beta}^{n} \|\mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta)\|^2, \end{aligned} \tag{22}$$

where $\operatorname{tr}\widehat{\boldsymbol{S}}_{\boldsymbol{H}} = \sum_{\mu=1}^{n}\langle\, \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu),\, \boldsymbol{H} \,\rangle = \|\boldsymbol{H}\|^2$, and the traces of the Ricci operators are by definition the scalar curvatures. □

The asymmetry of the components of the third fundamental form operator $\mathbf{III}(\boldsymbol{x}, \boldsymbol{y})$ encodes the curvature information of the connection defined on the normal bundle $N\mathcal{M}$ by $(\overline{\nabla}_{\boldsymbol{x}}\boldsymbol{N})^{\perp}$, for any $\boldsymbol{x} \in T_p\mathcal{M},\ \boldsymbol{N} \in \Gamma(N\mathcal{M})$, where an analog to Gauß equation holds, so we can write [29, Ex. II.11] as:

Lemma 3.5 *(Ricci equation). The Riemann curvature of the induced normal connection, $\boldsymbol{R}_{\perp}$, satisfies:*

$$\langle \boldsymbol{R}_{\perp}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m} \,\rangle = \langle\, \overline{\boldsymbol{R}}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m} \,\rangle + \langle\, \mathbf{III}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m} \,\rangle - \langle\, \mathbf{III}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{m}, \boldsymbol{n} \,\rangle, \tag{23}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in T_p\mathcal{M}$, and $\boldsymbol{n}, \boldsymbol{m} \in N_p\mathcal{M}$, at any point $p \in \mathcal{M}$.

Equations (20) and (21) will be recognized inside the elements of the tangent and normal matrix blocks in our covariance matrices to express its eigenvalues in terms of the third fundamental form.

4. Cylindrical covariance analysis

In this section we compute the integral invariants of the cylindrical domain around a point on an n-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^{n+k}$ (i.e., $\overline{\boldsymbol{R}} = 0$). This serves as a warm-up exercise for the more involved spherical computations, whereas at the same time it provides descriptors with better error bounds (cf. sec. 6), and the normal block calculations here are valid in both cases. When the cylinder is not normal to the manifold at the point, we can only establish the leading order terms, but that is sufficient in the generic case to be able to detect the tangent space of the manifold by the scaling behavior of the eigenvalues of the covariance matrix. Once the cylinder is fixed to be normal to this tangent space, the local EVD can be computed to next-to-leading order in scale to see how it encodes the geometric information of **III**. We shall solve the normal eigenvalue integrals in [25], which requires to compute the covariance matrix with respect to p instead of the barycenter.

We shall always work in a neighborhood $U \subset \mathbb{R}^{n+k}$ of $p \in \mathcal{M}$, sufficiently small so that $U \cap \mathcal{M}$ is given by a graph representation $\boldsymbol{X} = [x^1, \ldots, x^n, f^1(\boldsymbol{x}), \ldots, f^k(\boldsymbol{x})]^T$ over its tangent space, i.e., $\mathbf{0}$ represents p, $\boldsymbol{x} = [x^1, \ldots, x^n]^T \in T_p\mathcal{M}$, and $\nabla f^j(\mathbf{0}) = \mathbf{0}$, so that the manifold is approximated at p by its osculating paraboloid, [32, vol. 3, p. 42]. The following local expressions of the metric and induced volume element for graph manifolds are an easy exercise in these coordinates.

Lemma 4.1. *The first fundamental form components of a graph manifold $\mathcal{M} \subset \mathbb{R}^{n+k}$, parametrized by $\boldsymbol{X} = [x^1, \ldots, x^n, f^1(\boldsymbol{x}), \ldots, f^k(\boldsymbol{x})]^T \in T_p\mathcal{M} \oplus N_p\mathcal{M} \cong \mathbb{R}^{n+k}$, are:*

$$g_{\mu\nu}(\boldsymbol{x}) = \langle \frac{\partial \boldsymbol{X}}{\partial x^\mu}, \frac{\partial \boldsymbol{X}}{\partial x^\mu} \rangle = \delta_{\mu\nu} + \sum_{j=1}^{k} \frac{\partial f^j}{\partial x^\mu} \frac{\partial f^j}{\partial x^\nu}. \tag{24}$$

Thus, the induced measure on $\mathcal{M}$ is given by the volume element (cf. [32, vol. 1, pp. 311–312]):

$$\mathrm{dVol} = \sqrt{\det g(\boldsymbol{x})}\, d^n\boldsymbol{x} = \left(1 + \frac{1}{2} \sum_{j=1}^{k} \sum_{\alpha=1}^{n} \left[\sum_{\beta=1}^{n} \left(\frac{\partial^2 f^j}{\partial x^\alpha \partial x^\beta}(0) \right) x^\beta \right]^2 + \mathcal{O}(x^3) \right) d^n\boldsymbol{x}. \tag{25}$$

Proof. Equation (24) follows from the definition of tangent space of a graph. Equation (25) is immediate from equation (24) by usual expansion of a determinant of the form $\det[I + H]$ in terms of $1 + trH + \ldots$, which then yields the leading order contribution in x^β after Taylor-expanding the functions $f^j(\boldsymbol{x})$. □

In the rest of this paper we shall abbreviate second derivatives at the origin by

$$\kappa^j_{\alpha\beta} = \kappa^j_{\beta\alpha} := \frac{\partial^2 f^j}{\partial x^\alpha \partial x^\beta}(0),$$

motivated by the notation of hypersurface principal curvatures, which are the eigenvalues of the local Hessian. We can now compute the Taylor expansion of the integral invariants in the chosen coordinates, and then relate the terms to the curvature differential invariants which are always combinations of second derivatives.

Theorem 4.2. *The n-dimensional volume of the cylindrical component for a generic $\mathbb{V} \in \mathrm{Gr}(n, n+k)$, such that $\mathbb{V}^\perp \cap T_p\mathcal{M} = \{\mathbf{0}\}$, is to leading order the volume of the ellipsoid of intersection between the $\mathbb{V}$-cylinder and $T_p\mathcal{M}$:*

$$V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = V_n(1) \prod_{\mu=1}^{n} \ell_\mu + \mathcal{O}(\varepsilon^{n+1}), \tag{26}$$

where ℓ_μ are the principal semi-axes *of the ellipsoid. When $\mathbb{V} = T_p\mathcal{M}$, the volume is*

$$V(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon) \left[1 + \frac{\varepsilon^2}{2(n+2)} \operatorname{tr} \mathbf{III} + \mathcal{O}(\varepsilon^4)\right] \tag{27}$$

where $\operatorname{tr}\mathbf{III} = \|\boldsymbol{H}\|^2 - \mathcal{R}$*, and $V_n(\varepsilon)$ is the volume of the n-dimensional ball of radius ε (cf. Appendix).*

Proof. To compute the leading term of $V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V}))$ we can approximate $\mathcal{M}$ near p by its tangent space, such that, fixing local coordinates with a basis for $T_p\mathcal{M} \oplus N_p\mathcal{M}$, a point is specified by $\boldsymbol{X} = [\boldsymbol{x}, \mathbf{0}]^T$, with $\boldsymbol{x} \in T_p\mathcal{M},\ \mathbf{0} \in N_p\mathcal{M}$. Since $\mathbb{V}^\perp \cap T_p\mathcal{M} = \{\mathbf{0}\}$, we have $T_p\mathcal{M} \oplus \mathbb{V}^\perp = \mathbb{R}^{n+k}$, and of course $\mathbb{V} \oplus \mathbb{V}^\perp = \mathbb{R}^{n+k}$. Let $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ be an orthonormal basis of $T_p\mathcal{M}$, and $\{\boldsymbol{u}_\alpha\}_{\alpha=1}^n \cup \{\boldsymbol{v}_j\}_{j=1}^k$ an orthonormal basis of $\mathbb{V} \oplus \mathbb{V}^\perp$, then the elements of the former are a linear combination of the latter, so there are matrices A, B such that:

$$\boldsymbol{e}_\mu = \sum_{\alpha=1}^{n} A_\mu^\alpha \boldsymbol{u}_\alpha + \sum_{j=1}^{k} B_\mu^j \boldsymbol{v}_j.$$

We need to find the region $\|\mathrm{proj}_{\mathbb{V}}(\boldsymbol{X})\| \le \varepsilon$, and since $\boldsymbol{X} = \sum_\mu x^\mu \boldsymbol{e}_\mu$ when $\boldsymbol{X} \in T_p\mathcal{M}$, the projection is

$$\mathrm{proj}_{\mathbb{V}}(\boldsymbol{X}) = \sum_{\alpha=1}^{n} \langle\, \boldsymbol{X},\, \boldsymbol{u}_\alpha \,\rangle\, \boldsymbol{u}_\alpha = \sum_{\alpha=1}^{n} \sum_{\mu=1}^{n} x^\mu A_\mu^\alpha \boldsymbol{u}_\alpha,$$

hence, the domain of integration in $\boldsymbol{x}$ in this approximation is

$$\|\mathrm{proj}_{\mathbb{V}}(\boldsymbol{X})\|^2 = \sum_{\alpha=1}^{n} \left(\sum_{\mu=1}^{n} x^\mu A_\mu^\alpha \right)^2 \le \varepsilon^2.$$

This is a quadratic equation that can be written as

$$\sum_{\mu,\,\nu}^{n} x^\mu \left[\sum_{\alpha=1}^{n} A_\mu^\alpha A_\nu^\alpha \right] x^\nu = \boldsymbol{x}^T [A \cdot A^T] \boldsymbol{x} = \boldsymbol{y}^T \cdot \boldsymbol{y} = \|\boldsymbol{y}\|^2 \le \varepsilon^2,$$

where $\boldsymbol{y} = A^T \boldsymbol{x}$. The matrix $[A \cdot A^T]$ is positive definite since it is clearly nonnegative from the last equation, and if $\boldsymbol{x} \in \ker A^T$ for nonzero $\boldsymbol{x}$, then $\mathrm{proj}_{\mathbb{V}}(\boldsymbol{X}) = \mathbf{0}$, thus $\boldsymbol{X} \in \mathbb{V}^\perp$, which contradicts $\boldsymbol{X} \in T_p\mathcal{M}$ under our assumption $\mathbb{V}^\perp \cap T_p\mathcal{M} = \{\mathbf{0}\}$. Therefore, the cylindrical domain is an n-dimensional ellipsoid in the tangent space at p, whose volume is given in terms of its principal semi-axes ℓ_μ:

$$V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \prod_{\mu=1}^{n} \ell_\mu + \mathcal{O}(\varepsilon^{n+1}).$$

When $\mathbb{V} = T_p\mathcal{M}$, the local graph approximation of $\mathcal{M}$ over $T_p\mathcal{M}$ yields

$$\mathrm{proj}_{T_p\mathcal{M}}(\boldsymbol{X}) = \|\mathrm{proj}_{T_p\mathcal{M}}([\boldsymbol{x}, f^1(\boldsymbol{x}), \dots, f^k(\boldsymbol{x})]^T)\| = \|\boldsymbol{x}\| \leq \varepsilon,$$

thus, we are integrating $\sqrt{\det g(\boldsymbol{x})}$ in equation (25) over the ball $B_p^{(n)}(\varepsilon) \subset T_p\mathcal{M}$, which can be computed in spherical coordinates using the integrals in the appendix, as it will be done from here onwards in the rest of the paper:

$$\begin{aligned} V(\mathrm{Cyl}_p(\varepsilon)) &= \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int\limits_0^{\varepsilon} \rho^{n-1} \left(1 + \frac{1}{2}\sum_{i=1}^{k}\sum_{\alpha=1}^{n}\left[\sum_{\beta=1}^{n} \kappa^i_{\alpha\beta}\,\rho\overline{x}^\beta\right]^2 + \mathcal{O}(x^3)\right) d\rho \\ &= V_n(\varepsilon) + \frac{\varepsilon^{n+2}}{2(n+2)} \sum_{i=1}^{k}\sum_{\alpha=1}^{n}\sum_{\beta,\gamma}^{n} \kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma} \int\limits_{\mathbb{S}^{n-1}} \overline{x}^\beta \overline{x}^\gamma \, d\mathbb{S} + \mathcal{O}(\varepsilon^{n+4}) \\ &= V_n(\varepsilon) + \frac{C_2\,\varepsilon^{n+2}}{2(n+2)} \sum_{i=1}^{k}\sum_{\alpha,\beta}^{n} (\kappa^i_{\alpha\beta})^2 + \mathcal{O}(\varepsilon^{n+4}). \end{aligned}$$

Here the spherical integral is only nonzero when $\beta = \gamma$, and the sums of the last equality are the component expression of equation (19). □

Proposition 4.3. *The barycenter of the cylindrical component, for* $\mathbb{V}$ *as in the previous theorem, is*

$$\boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \boldsymbol{0} + \mathcal{O}(\varepsilon^2). \tag{28}$$

In the case $\mathbb{V} = T_p\mathcal{M}$*, the barycenter is:*

$$\boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon)) = [\,\boldsymbol{0},\ \frac{\varepsilon^2}{2(n+2)}\ \boldsymbol{H}\,]^T + \mathcal{O}(\varepsilon^4). \tag{29}$$

Proof. For generic $\mathbb{V}$, approximating the manifold again by its tangent space, $\boldsymbol{X} = [\boldsymbol{x}, \boldsymbol{0} + \mathcal{O}(\varepsilon^2)]^T$, the normal component does not contribute until order two and the tangent component also vanishes at order 1 in ε. When $\mathbb{V} = T_p\mathcal{M}$, we saw that the integration domain reduces to a ball. The integrals of the tangent components x^μ weighed by $\sqrt{\det g}$ are of order $\mathcal{O}(\varepsilon^{n+4})$, since the first terms in the expansion have odd powers in the coordinates. Abbreviating $V = V(\mathrm{Cyl}_p(\varepsilon))$ and $\boldsymbol{s} = \boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon))$, cf. equation (4), the normal components integrate to leading order as:

$$
\begin{aligned}
V \cdot [\boldsymbol{s}]^j &= \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int\limits_0^{\varepsilon} f^j(\boldsymbol{x}) \sqrt{\det g}\, \rho^{n-1}\, d\rho \\
&= \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int\limits_0^{\varepsilon} \rho^{n-1} \left(\frac{1}{2} \sum_{\alpha,\,\beta=1}^{n} \kappa^j_{\alpha\,\beta} \rho^2 \overline{x}^{\alpha} \overline{x}^{\beta} + \mathcal{O}(x^3) \right) d\rho \\
&= \frac{\varepsilon^{n+2}}{2(n+2)} \sum_{\alpha,\beta=1}^{n} \kappa^j_{\alpha\beta} \int\limits_{\mathbb{S}^{n-1}} \overline{x}^{\alpha} \overline{x}^{\beta} d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+4}) = \frac{C_2\, \varepsilon^{n+2}}{2(n+2)}\, H^j + \mathcal{O}(\varepsilon^{n+4}).
\end{aligned}
$$

Dividing by V cancels $C_2 \varepsilon^n = V_n(\varepsilon)$ to leading order. $\square$

In order to study the eigenvalue decomposition of the covariance matrix we need to establish how to determine the limit eigenvectors and the first two terms of the series expansion of the eigenvalues, so that computing the integrals in an arbitrary orthonormal basis produces blocks identifiable in terms of the coordinate expressions of the second and third fundamental forms in that basis.

Lemma 4.4. *Let $C(\varepsilon)$ be an $(n+k) \times (n+k)$ real symmetric matrix depending on a real parameter ε with convergent series expansion in a neighborhood of 0 such that:*

$$
C(\varepsilon) = \varepsilon^2 \left(\begin{array}{c|c} a\,\mathrm{Id}_n & 0_{n\times k} \\ \hline 0_{k\times n} & 0_{k\times k} \end{array} \right) + \varepsilon^4 \left(\begin{array}{c|c} \mathrm{A}_{n\times n} & \mathrm{B}_{n\times k} \\ \hline \mathrm{B}_{k\times n} & \Gamma_{k\times k} \end{array} \right) + \mathcal{O}(\varepsilon^5),
$$

where $a \neq 0$, and the blocks $\mathrm{A}, \mathrm{B}, \Gamma$ are not completely zero. Let $[\boldsymbol{V}]_{\top}, [\boldsymbol{V}]_{\perp}$ denote the first n and last k components of a vector in $\mathbb{R}^{n+k}$. If $C(\varepsilon)$ has no repeated eigenvalues the series of eigenvectors of $C(\varepsilon)$ form a unique orthonormal basis of $\mathbb{R}^{n+k}$ that converges for $\varepsilon \to 0$. The first n eigenvalues are $\lambda_\mu(\varepsilon) = a\varepsilon^2 + \lambda^{(4)}_\mu \varepsilon^4 + \mathcal{O}(\varepsilon^5)$, where $\lambda^{(4)}_\mu$ and the corresponding limit eigenvectors $\{\boldsymbol{V}^{(0)}_\mu\}^n_{\mu=1}$ satisfy the eigenvalue decomposition of A:

$$
(\lambda^{(4)}_\mu\, \mathrm{Id}_n - \mathrm{A})\, [\boldsymbol{V}^{(0)}_\mu]_{\top} = 0_{n\times 1}, \qquad [\boldsymbol{V}^{(0)}_\mu]_{\perp} = 0_{k\times 1}.
$$

The last k eigenvalues are $\lambda_j(\varepsilon) = \lambda^{(4)}_j \varepsilon^4 + \mathcal{O}(\varepsilon^5)$, where $\lambda^{(4)}_j$ and the corresponding limit eigenvectors $\{\boldsymbol{V}^{(0)}_j\}^{n+k}_{j=n+1}$ satisfy the eigenvalue decomposition of Γ:

$$
(\lambda^{(4)}_j\, \mathrm{Id}_k - \Gamma)\, [\boldsymbol{V}^{(0)}_j]_{\perp} = 0_{n\times 1}, \qquad [\boldsymbol{V}^{(0)}_j]_{\top} = 0_{n\times 1}.
$$

Therefore, the fourth-order term of the eigenvalues is given by the eigenvalues of the blocks A and Γ, with the respective eigenvectors as the limit eigenvectors of $C(\varepsilon)$ for $\varepsilon \to 0$.

Proof. The eigenvalue decomposition $C(\varepsilon)\boldsymbol{V}(\varepsilon) = \lambda(\varepsilon)\boldsymbol{V}(\varepsilon)$ can be written as a convergent series expansion in ε within a neighborhood of 0 for all Hermitian matrices of converging power series elements [35]:

$$[\,\varepsilon^2 \left(\begin{array}{c|c} a\,\mathrm{Id}_n & 0_{n\times k} \\ \hline 0_{k\times n} & 0_{k\times k}\end{array}\right) + \varepsilon^4 \left(\begin{array}{c|c} \mathrm{A}_{n\times n} & \mathrm{B}_{n\times k} \\ \hline \mathrm{B}_{k\times n} & \Gamma_{k\times k}\end{array}\right) + \mathcal{O}(\varepsilon^5)\,] \cdot [\,\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}\varepsilon + \boldsymbol{V}^{(2)}\varepsilon^2 + \dots] =$$
$$= (\lambda^{(1)}\varepsilon^1 + \lambda^{(2)}\varepsilon^2 + \lambda^{(3)}\varepsilon^3 + \lambda^{(4)}\varepsilon^4 + \dots)[\,\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}\varepsilon + \boldsymbol{V}^{(2)}\varepsilon^2 + \dots].$$

The zero matrix $C(0)$ is the limit when $\varepsilon \to 0$, with $\lambda(0) = \lambda^{(0)} = 0$ as a totally degenerate eigenvalue of multiplicity $(n+k)$. By [35, ch. I, Th. 1], for $\varepsilon > 0$, this eigenvalue branches out into $(n+k)$ eigenvalues $\lambda_i(\varepsilon)$ with $(n+k)$ orthonormal eigenvectors $\boldsymbol{V}_i(\varepsilon)$, all convergent in a neighborhood of 0. Thus, if $C(\varepsilon)$ has no repeated eigenvalues, the vectors $\boldsymbol{V}_i^{(0)} = \lim_{\varepsilon\to 0} \boldsymbol{V}_i(\varepsilon)$ form an orthonormal basis of $\mathbb{R}^{n+k}$ that is uniquely determined by the perturbation matrix.

The eigenvalue difference between $C(\varepsilon)$ and its full diagonalization is bounded by the matrix norm difference between them, which implies $\lambda^{(1)} = \lambda^{(3)} = 0$, and also $\lambda_i^{(2)} = a$, for $i = 1, \dots, n$, and $\lambda_i^{(2)} = 0$, for $i = n+1, \dots, n+k$, since $C(\varepsilon)$ is already diagonal up to that order. One can obtain the relations satisfied by $\lambda^{(4)}$ and $\boldsymbol{V}^{(0)}$ equating order by order. At second order, $\lambda_i^{(2)} = a$ is nonzero for $i = 1, \dots, n$, hence

$$[\,\left(\begin{array}{c|c} a\,\mathrm{Id}_n & 0_{n\times k} \\ \hline 0_{k\times n} & 0_{k\times k}\end{array}\right) - \lambda_i^{(2)}\,\mathrm{Id}_{n+k}\,]\,\boldsymbol{V}_i^{(0)} = \left(\begin{array}{c|c} 0_{n\times n} & 0_{n\times k} \\ \hline 0_{k\times n} & -a\,\mathrm{Id}_k\end{array}\right)\boldsymbol{V}_i^{(0)} = 0$$

implies that $[\boldsymbol{V}_\mu^{(0)}]_\perp = 0_{k\times 1}$, for the limit of the first n eigenvectors. At fourth order we have

$$[\,\lambda_i^{(4)}\,\mathrm{Id}_{n+k} - \left(\begin{array}{c|c} \mathrm{A}_{n\times n} & \mathrm{B}_{n\times k} \\ \hline \mathrm{B}_{k\times n} & \Gamma_{k\times k}\end{array}\right)]\,\boldsymbol{V}_i^{(0)} = [\,\left(\begin{array}{c|c} a\,\mathrm{Id}_n & 0_{n\times k} \\ \hline 0_{k\times n} & 0_{k\times k}\end{array}\right) - \lambda_i^{(2)}\,\mathrm{Id}_{n+k}\,]\,\boldsymbol{V}_i^{(2)},$$

which in the present case, $i = 1, \dots, n$, makes the right-hand side become 0 for the first n rows. On the other hand, $[\boldsymbol{V}_i^{(0)}]_\perp = 0_{k\times 1}$ makes B not contribute in the left-hand side, hence the first n rows lead to the equation:

$$(\lambda_i^{(4)}\,\mathrm{Id}_n - \mathrm{A})\,[\boldsymbol{V}_i^{(0)}]_\top = 0_{n\times 1}.$$

When $i = n+1, \dots, n+k$, an analogous argument using $\lambda_i^{(2)} = 0$, leads to $[\boldsymbol{V}_i^{(0)}]_\top = 0_{n\times 1}$, and in turn to:

$$(\lambda_i^{(4)}\,\mathrm{Id}_n - \Gamma)\,[\boldsymbol{V}_i^{(0)}]_\perp = 0_{k\times 1}.$$

Since the limit eigenvectors are an orthonormal basis they cannot be zero and, therefore, the previous equations establish $\lambda_i^{(4)}$ and the nonzero components of $[\boldsymbol{V}_i^{(0)}]$ as the eigenvalue decomposition of A and Γ, which always has a solution due to being symmetric matrices. □

The previous lemma is a fundamental step to establish the main theorem of this and the next section.

Theorem 4.5. *For $\mathbb{V} \in \mathrm{Gr}(n, n+k)$ such that $\mathbb{V}^\perp \cap T_p\mathcal{M} = \{\mathbf{0}\}$, i.e. for non-normal transversality, and when $V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V}))$ is finite, its covariance matrix centered at p, $C_p(\mathrm{Cyl}_p(\varepsilon, \mathbb{V}))$, has as limit eigenvectors spanning $T_p\mathcal{M}$ those corresponding to the first n eigenvalues, which scale as ε^2. The other k eigenvalues scale at higher order and have limit eigenvectors that span $N_p\mathcal{M}$:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \frac{\varepsilon^2}{n+2}\ell_\mu^2\, V_n(1) \prod_{\alpha=1}^{n} \ell_\alpha + \mathcal{O}(\varepsilon^{n+3}), \qquad \mu = 1, \ldots, n, \tag{30}$$

$$\lambda_j(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = 0 + \mathcal{O}(\varepsilon^{n+3}), \qquad j = n+1, \ldots, n+k, \tag{31}$$

where ℓ_μ are the principal lengths *of the ellipsoid in 4.2. When $\mathbb{V} = T_p\mathcal{M}$, let $\lambda_l[\cdot]$ denote taking the l-th eigenvalue of a linear operator at p, or of its associated bilinear form with respect to the metric. Then the eigenvalues of the covariance matrix of the cylindrical component are:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{2(n+2)(n+4)}(\mathrm{tr}\,\mathbf{III} + 2\,\lambda_\mu[\,\mathrm{tr}_\perp\mathbf{III}\,]) + \mathcal{O}(\varepsilon^6)\right] \tag{32}$$

$$\lambda_j(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^4}{4(n+2)(n+4)}\lambda_j[\,\boldsymbol{H}\otimes\boldsymbol{H} + 2\,\mathrm{tr}_\parallel\mathbf{III}\,] + \mathcal{O}(\varepsilon^6)\right] \tag{33}$$

for all $\mu = 1, \ldots, n$, and $j = n+1, \ldots, n+k$. Moreover, the corresponding first n eigenvectors converge to the principal directions of the operator $\mathrm{tr}_\perp\mathbf{III} = \widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}}$, *and the last k eigenvectors to those of $\boldsymbol{H}\otimes\boldsymbol{H} + 2\,\mathrm{tr}_\parallel\mathbf{III}$.*

Proof. For generic $\mathbb{V}$ the manifold is again approximated by its tangent space as $\boldsymbol{X} = [\boldsymbol{x}, \mathbf{0}]^T$, which produces no contribution to the normal block at leading order $\mathcal{O}(\varepsilon^{n+2})$. Choosing the tangent orthonormal basis to be aligned with the principal axis of the ellipsoid, and changing variables so that $x^\mu = y^\mu \ell_\mu$, the tangent block becomes an integration over an n-dimensional ball:

$$\begin{aligned}[C_p(\mathrm{Cyl}_p(\varepsilon, \mathbb{V}))]^{\mu\nu} &= \int\limits_{\boldsymbol{x}^T A\cdot A^T \boldsymbol{x} \le \varepsilon^2} x^\mu x^\nu \, d^n\boldsymbol{x} = \int\limits_{\sum_\mu y_\mu^2 \le 1} y^\mu y^\nu \ell_\mu \ell_\nu \prod_{\alpha=1}^{n} \ell_\alpha \, d^n\boldsymbol{y} \\ &= \delta_{\mu\nu}\frac{\varepsilon^{n+2}}{n+2}\ell_\mu\ell_\nu V_n(1)\prod_{\alpha=1}^{n}\ell_\alpha + \mathcal{O}(\varepsilon^{n+3}).\end{aligned}$$

Thus, the covariance matrix leading term is proportional to $\mathrm{diag}(\ell_1^2, \ldots, \ell_n^2, 0, \ldots, 0)$, which has limit eigenvectors corresponding to the first n eigenvalues spanning $T_p\mathcal{M}$, and the other k eigenvectors spanning $N_p\mathcal{M}$, by a straightforward extension to Lemma 4.4 at order ε^2. For $\mathbb{V} = T_p\mathcal{M}$, we shall compute the integrals of the matrix blocks $[x^\mu x^\nu]_{\mu,\nu=1}^n$, and $[f^i f^j]_{i,j=1}^k$, so the next-to-leading order elements of those blocks will suffice to obtain the eigenvalues and limit eigenvectors by the results of the previous lemma. The tangent block is:

$$
\begin{aligned}
[C_p(\mathrm{Cyl}_p(\varepsilon))]^{\mu\nu} &= \int\limits_{B^{(n)}(\varepsilon)} x^\mu x^\nu \sqrt{\det g(\boldsymbol{x})}\, d^n\boldsymbol{x} \\
&= \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int\limits_0^\varepsilon \rho^{n+1}\overline{x}^\mu\overline{x}^\nu \left(1+\frac{1}{2}\sum_{i=1}^k\sum_{\alpha=1}^n\left[\sum_{\beta=1}^n \kappa^j_{\alpha\beta}\rho\,\overline{x}^\beta\right]^2 + \mathcal{O}(x^3)\right) d\rho \\
&= \frac{\varepsilon^{n+2}}{n+2}\int\limits_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu d\,\mathbb{S} + \frac{\varepsilon^{n+4}}{2(n+4)}\sum_{i=1}^k\sum_{\alpha,\beta,\gamma}^n \kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}\int\limits_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu\overline{x}^\beta\overline{x}^\gamma d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+6}),
\end{aligned}
$$

and the last integral is only nonzero for the following combination of indices using the notation explained in the appendix:

$$
\int\limits_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu\overline{x}^\beta\overline{x}^\gamma \; d\,\mathbb{S} = C_4(\overbracket{\mu\nu\beta\gamma}) + C_{22}\left[(\overbracket{\mu\nu}\underbracket{\beta\gamma}) + (\mu\nu\beta\gamma) + (\mu\nu\beta\gamma)\right]. \tag{34}
$$

This simplifies the sums using the relationship between C_4, C_{22} and C_2, and writing $(1-\delta_{\mu\nu})$ to enforce $\mu \neq \nu$ in the last two terms of C_{22}:

$$
\begin{aligned}
&\frac{\delta_{\mu\nu}C_2\varepsilon^{n+2}}{n+2} \\
&\quad + \frac{C_2\varepsilon^{n+4}}{2(n+2)(n+4)}\sum_{i=1}^k\left[3\delta_{\mu\nu}\sum_{\alpha=1}^n(\kappa^i_{\alpha\mu})^2 + \delta_{\mu\nu}\sum_{\substack{\alpha,\,\beta \\ \beta\neq\mu}}^n(\kappa^i_{\alpha\beta})^2 + 2(1-\delta_{\mu\nu})\sum_{\alpha=1}^n\kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu}\right] + \ldots \\
&= V_n(\varepsilon)\frac{\varepsilon^2}{n+2}\delta_{\mu\nu} + \frac{V_n(\varepsilon)\varepsilon^4}{2(n+2)(n+4)}\left[\delta_{\mu\nu}\sum_{i=1}^k\sum_{\alpha,\beta}^n(\kappa^i_{\alpha\beta})^2 + 2\sum_{i=1}^k\sum_{\alpha=1}^n\kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu}\right] + \mathcal{O}(\varepsilon^{n+6}).
\end{aligned}
$$

Using equation (9), the component expression of equations (20) and (22) identify this block matrix at order $\mathcal{O}(\varepsilon^{n+4})$ as the matrix elements of the operator $[(\mathrm{tr}_{\parallel}\mathrm{tr}_{\perp}\mathbf{III})\mathrm{Id}_n + 2\mathrm{tr}_{\perp}\mathbf{III}]$ in our chosen orthonormal basis, whose eigenvalues are then by Lemma 4.4 the next-to-leading order contribution to the first n eigenvalues of $C_p(\mathrm{Cyl}_p(\varepsilon))$, and whose eigenvectors are the limit eigenvectors of $C_p(\mathrm{Cyl}_p(\varepsilon))$.

We perform now the integration of the normal block, which truncated to leading order is $\int_{B^{(n)}(\varepsilon)} f^i(\boldsymbol{x})f^j(\boldsymbol{x})d^n\boldsymbol{x}$, therefore:

$$
[C_p(\mathrm{Cyl}_p(\varepsilon))]^{ij} = \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S}\int\limits_0^\varepsilon \frac{\rho^{n+3}}{4}d\rho\sum_{\alpha,\beta}^n\sum_{\gamma,\,\delta}^n \kappa^i_{\alpha\beta}\kappa^j_{\gamma\delta}\overline{x}^\alpha\overline{x}^\beta\overline{x}^\gamma\overline{x}^\delta + \mathcal{O}(\varepsilon^{n+6}),
$$

where the angular integral is only nonzero in the same cases as in equation (34) above, but with the indices relabeled accordingly. This again simplifies every summation by matching the combination of indices and using the relations among the constants:

$$[C_p(\mathrm{Cyl}_p(\varepsilon))]^{ij} = \frac{\varepsilon^{n+4}}{4(n+4)}\left[C_4\sum_{\alpha=1}^{n}\kappa^i_{\alpha\alpha}\kappa^j_{\alpha\alpha} + C_{22}\left(\sum_{\substack{\alpha,\gamma\\ \alpha\neq\gamma}}^{n}\kappa^i_{\alpha\alpha}\kappa^j_{\gamma\gamma} + 2\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}^{n}\kappa^i_{\alpha\beta}\kappa^j_{\alpha\beta}\right)\right] + \mathcal{O}(\varepsilon^{n+6}),$$

in which the first sum precisely completes the elements missing from the other two, which written in terms of the components of the second fundamental form yields:

$$= \frac{V_n(\varepsilon)\varepsilon^2}{4(n+2)(n+4)}\left[\left(\sum_{\alpha=1}^{n}\mathrm{II}^i(\boldsymbol{e}_\alpha,\boldsymbol{e}_\alpha)\right)\left(\sum_{\gamma=1}^{n}\mathrm{II}^j(\boldsymbol{e}_\gamma,\boldsymbol{e}_\gamma)\right) + 2\sum_{\alpha,\beta}^{n}\mathrm{II}^i(\boldsymbol{e}_\alpha,\boldsymbol{e}_\beta)\mathrm{II}^j(\boldsymbol{e}_\alpha,\boldsymbol{e}_\beta)\right] + \mathcal{O}(\varepsilon^{n+6}).$$

In this last expression we clearly identify the components in our orthonormal normal basis of $[\boldsymbol{H}\otimes\boldsymbol{H}]^{ij}$, and those of $2\,\mathrm{tr}_{\,\|}\mathbf{III}$, using the definition of $\boldsymbol{H}$ and equation (21). □

We shall see below that the spherical covariance matrix has the same normal eigenvalues, to leading order, as the cylindrical case above. In [24,25] these were expressed as an average of the squares of the curvatures of curves inside the manifold $\mathcal{M}$. Therefore, our previous computation provides an explicit formula for this interpretation of the normal eigenvalues.

Corollary 4.6. *Let $\mathcal{M}$ be an n-dimensional submanifold of Euclidean space $\mathbb{R}^{n+k}$, then the first generalized curvatures $\kappa(\gamma,\boldsymbol{x},\boldsymbol{n}_j)$ of curves $\gamma\subset\mathcal{M}$ passing through p with tangent vector $\boldsymbol{x}$ and principal normal vectors any of the eigenvectors $\boldsymbol{n}_j$, $j=1,\dots,k$, of $[\,\boldsymbol{H}\otimes\boldsymbol{H}+2\,\mathrm{tr}_{\,\|}\mathbf{III}\,]$, integrate to:*

$$\int_{B^{(n)}(\varepsilon)}\kappa^2(\gamma,\boldsymbol{x},\boldsymbol{n}_j)\,d^n\boldsymbol{x} = \frac{\varepsilon^4\,V_n(\varepsilon)}{(n+2)(n+4)}\lambda_j[\,\boldsymbol{H}\otimes\boldsymbol{H}+2\,\mathrm{tr}_{\,\|}\mathbf{III}\,]. \tag{35}$$

In particular:

$$\sum_{j=1}^{k}\int_{B^{(n)}(\varepsilon)}\kappa^2(\gamma,\boldsymbol{x},\boldsymbol{n}_j)\,d^n\boldsymbol{x} = \frac{3\|\boldsymbol{H}\|^2-2\mathcal{R}}{(n+2)(n+4)}\varepsilon^4\,V_n(\varepsilon). \tag{36}$$

5. Spherical covariance analysis

The difference between the cylindrical and spherical intersection domains for a graph manifold lies in the irregular projection onto the tangent space: by definition the cylinder is the extension in the normal directions of the ball $B_p^{(n)}(\varepsilon) \subset T_p\mathcal{M}$, so the points of the graph manifold satisfy $\|\mathrm{proj}_{T_p\mathcal{M}}([\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x})]^T)\| = \|\boldsymbol{x}\| \leq \varepsilon$, and thus the integration region is a perfect ball. However, in the spherical case the domain of integration is $\|\boldsymbol{x}\|^2 + \|\boldsymbol{f}(\boldsymbol{x})\|^2 \leq \varepsilon^2$, which is nontrivial and in general cannot be parametrized exactly. One can straightforwardly apply the same procedure as done originally in [13] and [28] to find the leading order corrections to the ball domain in the tangent space coordinates.

Lemma 5.1. *For $\varepsilon > 0$ small enough so that $\mathcal{M}$ is a graph manifold over $T_p\mathcal{M}$, using cylindrical coordinates, the radial parametric equation of a point $\boldsymbol{X} = [\rho\overline{x}^1, \ldots, \rho\overline{x}^n, f^1(\rho\overline{\boldsymbol{x}}), \ldots, f^k(\rho\overline{\boldsymbol{x}})]^T$ in $\partial D_p(\varepsilon) = \mathcal{M} \cap \mathbb{S}_p^n(\varepsilon)$ is*

$$r(\overline{\boldsymbol{x}}) := \rho(\overline{x}_1, \ldots, \overline{x}_n) = \varepsilon - \frac{K(\overline{\boldsymbol{x}})^2}{8}\varepsilon^3 + \mathcal{O}(\varepsilon^4), \tag{37}$$

where $\overline{\boldsymbol{x}} \in \mathbb{S}^{n-1} \subset T_p\mathcal{M}$, and

$$K(\overline{\boldsymbol{x}})^2 := \|\mathbf{II}(\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}})\|^2 = \sum_{i=1}^{k}\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n} \kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}\, \overline{x}^\alpha\overline{x}^\beta\overline{x}^\gamma\overline{x}^\delta \tag{38}$$

is the square of the ambient space acceleration of a geodesic curve of $\mathcal{M}$ with tangent vector $\overline{\boldsymbol{x}}$ at p (cf. [31, ch. 4, Cor. 10]).

This allows us to perform the same type of integrals over the radial coordinate as before but over an irregular tangent domain in order to obtain explicitly the higher-order contributions to the tangent eigenvalues, as studied to leading order in [23], [24,25].

Proposition 5.2. *The n-dimensional volume of the spherical component is*

$$V(D_p(\varepsilon)) = V_n(\varepsilon)\left[1 + \frac{\varepsilon^2}{8(n+2)}\left(2\,\mathrm{tr}\,\mathbf{III} - \|\boldsymbol{H}\|^2\right) + \mathcal{O}(\varepsilon^3)\right]. \tag{39}$$

Proof. In contrast to the proof of the cylindrical domain, the radial integration introduces new angular corrections due to $r(\overline{\boldsymbol{x}})$:

$$\begin{aligned} V(D_p(\varepsilon)) &= \int_{\mathbb{S}^{n-1}} d\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \rho^{n-1}\sqrt{\det g(\rho\overline{\boldsymbol{x}})}\, d\rho \\ &= \int_{\mathbb{S}^{n-1}} \frac{r(\overline{\boldsymbol{x}})^n}{n} d\mathbb{S} + \int_{\mathbb{S}^{n-1}} \frac{r(\overline{\boldsymbol{x}})^{n+2}}{2(n+2)} \sum_{i=1}^{k}\sum_{\alpha,\beta,\gamma}^{n} \kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}\overline{x}^\beta\overline{x}^\gamma\, d\mathbb{S} + \mathcal{O}(\varepsilon^{n+3}), \end{aligned}$$

the second integral is the same to leading order as in the cylindrical case, hence

$$= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S}\,\frac{\varepsilon^n}{n}\left[1 - n\frac{K(\overline{\boldsymbol{x}})^2}{8}\varepsilon^2 + \mathcal{O}(\varepsilon^3)\right] + \frac{V_n(\varepsilon)\,\varepsilon^2}{2(n+2)}\operatorname{tr}\mathbf{III} + \mathcal{O}(\varepsilon^{n+3})$$

$$= V_n(\varepsilon) - \frac{\varepsilon^{n+2}}{8}\sum_{i=1}^{k}\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}\int_{\mathbb{S}^{n-1}}\overline{x}^\alpha\overline{x}^\beta\overline{x}^\gamma\overline{x}^\delta\,d\,\mathbb{S} + \frac{V_n(\varepsilon)\,\varepsilon^2}{2(n+2)}\operatorname{tr}\mathbf{III} + \mathcal{O}(\varepsilon^{n+3}),$$

where the integral is only nonzero as in equation (34), thus

$$= V_n(\varepsilon) - \frac{C_2\,\varepsilon^{n+2}}{8(n+2)}\sum_{i=1}^{k}\left[3\sum_{a=1}^{n}(\kappa^i_{\alpha\alpha})^2 + \sum_{\substack{\alpha,\gamma\\ \alpha\neq\gamma}}^{n}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} + 2\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}^{n}(\kappa^i_{\alpha\beta})^2\right]$$
$$+ \frac{V_n(\varepsilon)\,\varepsilon^2}{2(n+2)}\operatorname{tr}\mathbf{III} + \mathcal{O}(\varepsilon^{n+3})$$

Collecting terms and using the first summation in the braces to complete the other two double sums, we can recognize a term as $\langle \sum_\alpha \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha), \sum_\gamma \mathbf{II}(\boldsymbol{e}_\gamma, \boldsymbol{e}_\gamma)\rangle = \|\boldsymbol{H}\|^2$, and the another term as $\operatorname{tr}\mathbf{III}$. □

Remark 5.3. Notice that it is not known the dependence of the error generated by the irregular radius $r(\overline{\boldsymbol{x}})$, $\mathcal{O}(\varepsilon^{n+3})$ in the previous proof, and whether it cancels at that order upon spherical integration, so the spherical component invariants may have error terms at lower order than the cylindrical ones.

Proposition 5.4. *The barycenter of the spherical component is to leading order the same as for the cylindrical component:*

$$\boldsymbol{s}(D_p(\varepsilon)) = [\,\mathbf{0},\ \frac{\varepsilon^2}{2(n+2)}\ \boldsymbol{H}\ ]^T + \mathcal{O}(\varepsilon^4). \tag{40}$$

Proof. The new contributions from $r(\overline{\boldsymbol{x}})$ to the cylindrical computations are at least of the same order as the overall error, $\mathcal{O}(\varepsilon^4)$, so they can be neglected. □

The covariance integral invariants for the spherical domain were obtained for hypersurfaces in [28] by performing the computations in the basis of principal and normal directions. In arbitrary codimension, the different osculating paraboloids of $f^i(\boldsymbol{x}), i = 1, \ldots, k$, cannot be diagonalized simultaneously to a common basis in general. The amount of terms and simplifications needed in this general case is of much higher complexity than for hypersurfaces but, nevertheless, an analogous result for the eigenvalue decomposition obtains.

Theorem 5.5. *Let $\lambda_l[\cdot]$ denote taking the l-th eigenvalue of a linear operator at p, or of its associated bilinear form with respect to the metric. Then the eigenvalues of the spherical component covariance matrix, $C_{\boldsymbol{s}}(D_p(\varepsilon))$, with respect to $\boldsymbol{s}(D_p(\varepsilon))$ are:*

$$\lambda_\mu(D_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{8(n+2)(n+4)}(2\,\mathrm{tr}\,\mathbf{III} - \|\boldsymbol{H}\|^2 - 4\,\lambda_\mu[\,\widehat{\boldsymbol{S}}_{\boldsymbol{H}}\,]) + \mathcal{O}(\varepsilon^5)\right] \tag{41}$$

$$\lambda_j(D_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^4}{2(n+2)(n+4)}\,\lambda_j[\,\mathrm{tr}_{\,\|}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}\,] + \mathcal{O}(\varepsilon^6)\right] \tag{42}$$

for all $\mu = 1,\ldots,n$, and $j = n+1,\ldots,n+k$. Moreover, the corresponding first n eigenvectors converge to the principal (tangent) directions of the Weingarten operator at $\boldsymbol{H}$, and the last k eigenvectors to those of $[\mathrm{tr}_{\,\|}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}]$ in the normal space.

Proof. From Lemma 4.4 again, only the tangent and normal blocks need to be computed. Now, however, the covariance matrix is taken with respect to the barycenter, so there is an extra matrix contribution from the tensor product,

$$C_{\boldsymbol{s}}(D_p(\varepsilon)) = \int\limits_{D_p(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{X}\ \mathrm{dVol} - \int\limits_{D_p(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{s}\ \mathrm{dVol},$$

because the other two products cancel each other upon integration. From the proof of the barycenter formula, this integral is to leading order:

$$\int\limits_{D_p(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{s}\ \mathrm{dVol} = V(D_p(\varepsilon))\boldsymbol{s}\otimes\boldsymbol{s} = \left(\begin{array}{c|c} \mathcal{O}(\varepsilon^{n+8})_{n\times n} & \mathcal{O}(\varepsilon^{n+6})_{n\times k} \\ \hline \mathcal{O}(\varepsilon^{n+6})_{k\times n} & \frac{V_n(\varepsilon)\varepsilon^4}{4(n+2)^2}\boldsymbol{H}\otimes\boldsymbol{H} \end{array}\right)$$

There is no difference in the normal block computations of this covariance matrix and the cylindrical case proved before, since the corrections coming from $r(\overline{\boldsymbol{x}})$ are $\mathcal{O}(\varepsilon^{n+6})$. Thus, subtracting the barycenter contribution:

$$\begin{aligned} &\frac{V_n(\varepsilon)\varepsilon^4}{4(n+2)(n+4)}(\boldsymbol{H}\otimes\boldsymbol{H} + 2\,\mathrm{tr}_{\,\|}\mathbf{III}) - \frac{V_n(\varepsilon)\varepsilon^4}{4(n+2)^2}\boldsymbol{H}\otimes\boldsymbol{H} \\ &= \frac{V_n(\varepsilon)\varepsilon^4}{2(n+2)(n+4)}(\,\mathrm{tr}_{\,\|}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}). \end{aligned}$$

For the tangent block, the number of correction terms due to the spherical domain irregularities in the coordinate boundary makes a substantial contribution at $\mathcal{O}(\varepsilon^{n+4})$ compared to the cylindrical case:

$$[C(D_p(\varepsilon))]^{\mu\nu} = \int\limits_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int\limits_0^{r(\overline{\boldsymbol{x}})} \rho^{n+1}\overline{x}^\mu\overline{x}^\nu \left(1 + \frac{1}{2}\sum_{i=1}^k\sum_{\alpha=1}^n\left[\sum_{\beta=1}^n \kappa^i_{\alpha\beta}\,\rho\,\overline{x}^\beta\right]^2 + \mathcal{O}(x^3)\right) d\rho$$

$$= \frac{\varepsilon^{n+2}}{n+2}\left[\delta_{\mu\nu}C_2 - (n+2)\int\limits_{\mathbb{S}^{n-1}} \overline{x}^{\mu}\overline{x}^{\nu}\frac{K(\overline{\boldsymbol{x}})^2\varepsilon^2}{8}d\,\mathbb{S} + \mathcal{O}(\varepsilon^3)\right]$$

$$+ \frac{\varepsilon^{n+4}}{2(n+4)}\sum_{i=1}^{k}\sum_{\alpha,\beta,\gamma}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}\int\limits_{\mathbb{S}^{n-1}}\overline{x}^{\mu}\overline{x}^{\nu}\overline{x}^{\beta}\overline{x}^{\gamma}d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+5})$$

$$= \delta_{\mu\nu}\frac{V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{\varepsilon^{n+4}}{2(n+4)}\sum_{i=1}^{k}\left[\sum_{\alpha,\beta,\gamma}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}C_{(\mu\nu\beta\gamma)} - \frac{n+4}{4}\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}C_{(\mu\nu\alpha\beta\gamma\delta)}\right]$$

$$+ \mathcal{O}(\varepsilon^{n+5}),$$

where we have made use of equation (38), and written $C_{(\alpha\beta\dots)}$ for the integral over $\mathbb{S}^{n-1}$ of the monomial product $\overline{x}^{\alpha}\overline{x}^{\beta}\dots$, (notice here the indices are not exponents but contravariant coordinate components). The first summation simplifies again with equation (34) to yield the cylindrical tangent block, but the other set of sums comprises the 31 spherical integrals of all possible monomials of degree six:

$$C_{(\mu\nu\alpha\beta\gamma\delta)} = \int\limits_{\mathbb{S}^{n-1}}\overline{x}^{\mu}\overline{x}^{\nu}\overline{x}^{\alpha}\overline{x}^{\beta}\overline{x}^{\gamma}\overline{x}^{\delta}d\,\mathbb{S} = C_6(\mu\nu\alpha\beta\gamma\delta) \ +$$

$$+ C_{24}\Big[\,(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) +$$
$$(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) +$$
$$(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta)\,\Big]$$

$$+ C_{222}\Big[\,(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) +$$
$$(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) +$$
$$(\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta)\Big]$$

Each of these contractions are only nonzero when the connected indices are equal, and at the same time different from the indices of the other connected groups, for instance:

$$\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}(\mu\nu\alpha\beta\gamma\delta) = \delta_{\mu\nu}\sum_{\alpha\neq\mu}^{n}\sum_{\substack{\gamma\neq\mu\\ \gamma\neq\alpha}}^{n}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma}.$$

Matching all the indices in this way for each of the terms just found, and taking into account the relation of C_6, C_{24} and C_{222} to C_2 in the appendix, we take out a common factor $\frac{C_2}{4(n+2)}$, and abbreviate the sum notation to produce all the terms of order $\mathcal{O}(\varepsilon^{n+4})$:

$$
\begin{aligned}
[C(D_p(\varepsilon))]_{\mu\nu} &= \frac{\delta_{\mu\nu} V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{C_2\, \varepsilon^{n+4}}{8(n+2)(n+4)} \sum_i \Bigg[4\delta_{\mu\nu} \sum_{\substack{\alpha,\beta \\ \beta\neq\mu}} (\kappa^i_{\alpha\beta})^2 + 8\not\delta_{\mu\nu} \sum_{\alpha} \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} \\
&+ 12\delta_{\mu\nu} \sum_{\alpha} (\kappa^i_{\alpha\nu})^2 - 15\delta_{\mu\nu}(\kappa^i_{\nu\nu})^2 - 3 \Bigg\{ \delta_{\mu\nu} \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\alpha})^2 + \not\delta_{\mu\nu}(\kappa^i_{\mu\nu}\kappa^i_{\nu\nu} + \kappa^i_{\nu\mu}\kappa_{\nu\nu} + \kappa^i_{\nu\nu}\kappa^i_{\mu\nu} \\
&+ \kappa^i_{\nu\nu}\kappa^i_{\nu\mu} + \kappa^i_{\nu\mu}\kappa^i_{\mu\mu} + \kappa^i_{\mu\nu}\kappa^i_{\mu\mu} + \kappa^i_{\mu\mu}\kappa^i_{\nu\mu} + \kappa^i_{\mu\mu}\kappa^i_{\mu\nu}) + \delta_{\mu\nu} \Bigg(\sum_{\alpha\neq\mu} \kappa^i_{\alpha\alpha}\kappa^i_{\nu\nu} + \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\nu})^2 \\
&+ \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\nu})^2 + \sum_{\beta\neq\mu} (\kappa^i_{\nu\beta})^2 + \sum_{\beta\neq\mu} (\kappa^i_{\nu\beta})^2 + \sum_{\gamma\neq\mu} \kappa^i_{\gamma\gamma}\kappa^i_{\nu\nu} \Bigg) \Bigg\} \\
&- \delta_{\mu\nu} \Bigg(\sum_{\alpha\neq\mu} \sum_{\gamma\neq\mu,\alpha} \kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} + \sum_{\alpha\neq\mu} \sum_{\beta\neq\mu,\alpha} (\kappa^i_{\alpha\beta})^2 + \sum_{\alpha\neq\mu} \sum_{\beta\neq\mu,\alpha} (\kappa^i_{\alpha\beta})^2 \Bigg) - \not\delta_{\mu\nu} \Bigg\{ \sum_{\gamma\neq\mu,\nu} \kappa^i_{\mu\nu}\kappa^i_{\gamma\gamma} \\
&+ \sum_{\beta\neq\mu,\nu} \kappa^i_{\mu\beta}\kappa^i_{\nu\beta} + \sum_{\beta\neq\mu,\nu} \kappa^i_{\mu\beta}\kappa^i_{\beta\nu} + \sum_{\gamma\neq\mu,\nu} \kappa^i_{\nu\mu}\kappa^i_{\gamma\gamma} + \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} + \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} \\
&+ \sum_{\beta\neq\mu,\nu} \kappa^i_{\nu\beta}\kappa^i_{\mu\beta} + \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\nu}\kappa^i_{\mu\alpha} + \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\alpha}\kappa^i_{\mu\nu} + \sum_{\beta\neq\mu,\nu} \kappa^i_{\nu\beta}\kappa^i_{\beta\mu} + \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\nu}\kappa^i_{\alpha\mu} \\
&+ \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\alpha}\kappa^i_{\nu\mu} \Bigg\} \Bigg] + \mathcal{O}(\varepsilon^{n+5})
\end{aligned}
$$

Many of the resulting summations are the same after relabeling indices and using $\kappa^i_{\alpha\beta} = \kappa^i_{\beta\alpha}$, so they can be gathered into common factors to yield:

$$
\begin{aligned}
&= \delta_{\mu\nu} \frac{V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{V_n(\varepsilon)\varepsilon^4}{8(n+2)(n+4)} \sum_i \Bigg[8 \sum_{\alpha} \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} - 12\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu} + \kappa^i_{\nu\nu}) - 4\kappa^i_{\mu\nu} \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\alpha} \\
&- 8 \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} + \delta_{\mu\nu} \Bigg\{ 4 \sum_{\alpha,\beta} (\kappa^i_{\alpha\beta})^2 - 3 \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\alpha})^2 + 21(\kappa^i_{\mu\mu})^2 - 2\kappa^i_{\mu\mu} \sum_{\alpha\neq\mu} \kappa^i_{\alpha\alpha} \\
&-12 \sum_{\alpha} (\kappa^i_{\alpha\mu})^2 - \sum_{\alpha\neq\mu} \sum_{\gamma\neq\alpha,\mu} \kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2 \sum_{\alpha\neq\mu} \sum_{\beta\neq\alpha,\mu} (\kappa^i_{\alpha\beta})^2 + 8 \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\mu})^2 \Bigg\} \Bigg] + \mathcal{O}(\varepsilon^{n+5}).
\end{aligned}
$$

Some terms inside the curly braces complement the missing elements of other summations:

$$
21(\kappa^i_{\mu\mu})^2 - 2\kappa^i_{\mu\mu} \sum_{\alpha\neq\mu} \kappa^i_{\alpha\alpha} - 12 \sum_{\alpha} (\kappa^i_{\alpha\mu})^2 + 8 \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\mu})^2
$$

$$= 15(\kappa^i_{\mu\mu})^2 - 2\kappa^i_{\mu\mu}\sum_{\alpha}\kappa^i_{\alpha\alpha} - 4\sum_{\alpha}(\kappa^i_{\alpha\mu})^2,$$

and

$$-3\sum_{\alpha\neq\mu}(\kappa^i_{\alpha\alpha})^2 - \sum_{\alpha\neq\mu}\sum_{\gamma\neq\alpha,\mu}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{\alpha\neq\mu}\sum_{\beta\neq\alpha,\mu}(\kappa^i_{\alpha\beta})^2 = -\sum_{\alpha,\gamma\neq\mu}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{\alpha,\beta\neq\mu}(\kappa^i_{\alpha\beta})^2.$$

Now, notice that this last type of double sum decomposes as follows

$$-\sum_{\alpha,\gamma\neq\mu}[\,\cdot\,]_{\alpha\gamma} = -\sum_{\alpha,\,\gamma}[\,\cdot\,]_{\alpha\gamma} + \sum_{\substack{\gamma\\ \alpha=\mu}}[\,\cdot\,]_{\alpha\gamma} + \sum_{\substack{\alpha\\ \gamma=\mu}}[\,\cdot\,]_{\alpha\gamma} - [\,\cdot\,]_{\mu\mu},$$

therefore, the right hand side of the previous two equations complement each other:

$$[C(D_p(\varepsilon))]^{\mu\nu} = \frac{\delta_{\mu\nu}V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{V_n(\varepsilon)\varepsilon^4}{8(n+2)(n+4)}\sum_i\Bigg[8\sum_{\alpha}\kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} - 12\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu}+\kappa^i_{\nu\nu})$$
$$-4\kappa^i_{\mu\nu}\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\alpha} - 8\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha}$$
$$+\delta_{\mu\nu}\left\{4\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2 + 12(\kappa^i_{\mu\mu})^2 - \sum_{\alpha,\,\gamma}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2\right\}\Bigg] + \ldots$$

To simplify further, use $12(\kappa^i_{\mu\mu})^2$ to complete the remaining sums and cancel terms:

$$8\sum_{\alpha}\kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} - 8\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu}+\kappa^i_{\nu\nu}) - 8\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} + 8(\kappa^i_{\mu\mu})^2\delta_{\mu\nu} = 0,$$

and

$$-4\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu}+\kappa^i_{\nu\nu}) - 4\kappa^i_{\mu\nu}\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\alpha} + 4(\kappa^i_{\mu\mu})^2\delta_{\mu\nu} = -4\kappa^i_{\mu\nu}\sum_{\alpha}\kappa^i_{\alpha\alpha}.$$

Finally, all these computations lead us to the simple expression:

$$[C(D_p(\varepsilon))]_{\mu\nu} = \frac{\delta_{\mu\nu}V_n(\varepsilon)\varepsilon^2}{n+2}$$
$$+\frac{V_n(\varepsilon)\varepsilon^4}{8(n+2)(n+4)}\sum_i\left[\delta_{\mu\nu}\left\{2\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2 - (H^i)^2\right\} - 4\kappa^i_{\mu\nu}H^i\right] + \mathcal{O}(\varepsilon^{n+5})$$

where $\sum_i \kappa^i_{\mu\nu}H^i = \langle\, \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu),\, \boldsymbol{H}\,\rangle = \langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{H}}\, \boldsymbol{e}_\mu,\, \boldsymbol{e}_\nu\,\rangle$, and $\sum_i(2\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2 - (H^i)^2) = 2\,\mathrm{tr}\,\mathbf{III} - \|\boldsymbol{H}\|^2$ identify the covariance tangent block to be the component matrix of the Weingarten operator at the mean curvature normal, plus a constant, in the orthonormal basis chosen. $\square$

Making $k = 1$ in the previous theorem recovers part of the results of [28]; setting $n = 2$ as well recovers the patch domain equations of [13]. When the generalized principal directions of the Weingarten operator and the $[\mathrm{tr}_{\parallel}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}]$ operator are different, the limit eigenvector obtained from the covariance analysis play the role of an adapted Galerkin basis in the sense of [23] and [25].

6. Correspondence between local EVD and curvature

Curvature descriptors in terms of the covariance eigenvalues were introduced for surfaces in [13] and in [28] for hypersurfaces. The formulas of the previous section recover those results for $n = 2$ and $k = 1$ respectively. A limit formula for the ratio of the eigenvalues was found for curves in [26], establishing a direct relationship between the local covariance analysis of a domain containing the point p and the Frenet-Serret curvature information at p, which in the case of curves completely determines the curve locally up to rigid motion [36, Th. 2.13]. This furnishes a reconstruction correspondence between local eigenvalues of covariance and curvature descriptors at scale. The two main theorems of the present work generalize this type of result to general submanifolds by directly taking the limits with ε of the covariance matrix eigenvalue expansions obtained.

Corollary 6.1. *Writing $\lambda_\mu(p,\varepsilon)$ for the tangent eigenvalues of the cylindrical covariance matrix $C(\mathrm{Cyl}_p(\varepsilon))$, they satisfy the asymptotic ratio*

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\lambda_\mu(p,\varepsilon)-\lambda_\nu(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{n+4}\left(\lambda_\mu[\mathrm{tr}_\perp\mathbf{III}] - \lambda_\nu[\mathrm{tr}_\perp\mathbf{III}]\right), \tag{43}$$

and the normal eigenvalues satisfy

$$\lim_{\varepsilon\to 0}\frac{V_n(\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)}\sum_{j=n+1}^{n+k}\lambda_j(p,\varepsilon) = \frac{n+2}{4(n+4)}\left(\|\boldsymbol{H}\|^2 + 2\,\mathrm{tr}\,\mathbf{III}\right), \tag{44}$$

for any $\mu,\nu = 1,\dots,n$. Let $\widetilde{\lambda}_\mu(p,\varepsilon)$ denote the eigenvalues in the case of the spherical domain covariance matrix, $C(D_p(\varepsilon))$, then the corresponding limits are

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\widetilde{\lambda}_\mu(p,\varepsilon)-\widetilde{\lambda}_\nu(p,\varepsilon)}{\widetilde{\lambda}_\mu(p,\varepsilon)\widetilde{\lambda}_\nu(p,\varepsilon)} = \frac{n+2}{2(n+4)}\left(\widetilde{\lambda}_\nu[\widehat{\boldsymbol{S}}_{\boldsymbol{H}}] - \widetilde{\lambda}_\mu[\widehat{\boldsymbol{S}}_{\boldsymbol{H}}]\right), \tag{45}$$

and

$$\lim_{\varepsilon\to 0}\frac{V_n(\varepsilon)}{\widetilde{\lambda}_\mu(p,\varepsilon)\widetilde{\lambda}_\nu(p,\varepsilon)}\sum_{j=n+1}^{n+k}\widetilde{\lambda}_j(p,\varepsilon) = \frac{n+2}{2(n+4)}\left(\mathrm{tr}\,\mathbf{III} - \frac{1}{n+2}\|\boldsymbol{H}\|^2\right). \tag{46}$$

The operators on the right-hand sides are understood to be evaluated at the point p.

Notice that normalizing the covariance matrix to the volume of the ball in equation (5) simplifies our results further, since $V_n(\varepsilon)$ would disappear from these asymptotic relations.

For hypersurfaces, the obtained terms of the series expansion of the eigenvalue decomposition can be inverted to extract the curvature descriptors upon truncations of the series. In the spherical case, the descriptors of the cited works are recovered.

Corollary 6.2. [28] *Let us write $\lambda(p,\varepsilon) \equiv \lambda(D_p(\varepsilon)), V_p(\varepsilon) \equiv V(D_p(\varepsilon))$ for the integral invariants of a spherical domain on a hypersurface $\mathcal{S}$, then the corresponding scalar, mean and principal curvature descriptors, at scale $\varepsilon > 0$ and point $p \in \mathcal{S}$, for any $\mu = 1, \ldots, n$, are:*

$$\mathcal{R}(D_p(\varepsilon)) = 2(n+2)^2(n+4)\frac{\lambda_{n+1}(p,\varepsilon)}{n\,\varepsilon^4\,V_n(\varepsilon)} - \frac{8(n+1)(n+2)}{n\,\varepsilon^2}\left(\frac{V_p(\varepsilon)}{V_n(\varepsilon)} - 1\right) \tag{47}$$

$$H(D_p(\varepsilon)) = (\pm)\sqrt{4(n+2)^2(n+4)\frac{\lambda_{n+1}(p,\varepsilon)}{n\,\varepsilon^4 V_n(\varepsilon)} + \frac{8(n+2)^2}{n\,\varepsilon^2}\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right)}, \tag{48}$$

$$\kappa_\mu(D_p(\varepsilon)) = \frac{2(n+2)}{\varepsilon^2 H(D_p^+(\varepsilon))}\left[\frac{V_p(\varepsilon)}{V_n(\varepsilon)} + \frac{n+4}{\varepsilon^2}\left(\frac{\varepsilon^2}{n+2} - \frac{\lambda_\mu(p,\varepsilon)}{V_n(\varepsilon)}\right) - 1\right], \tag{49}$$

where the overall sign can be chosen by fixing a normal orientation from

$$(\pm) = \operatorname{sgn}\langle\, \boldsymbol{e}_{n+1}(D_p(\varepsilon)),\ \boldsymbol{s}(D_p(\varepsilon))\,\rangle.$$

The eigenvectors $\boldsymbol{e}_\mu(D_p(\varepsilon))$ and $\boldsymbol{e}_{n+1}(D_p(\varepsilon))$ are descriptors of the principal and normal directions respectively. The errors are:

$$|H^2(p) - H^2(D_p(\varepsilon))| \leq \mathcal{O}(\varepsilon), \qquad |\mathcal{R}(p) - \mathcal{R}(D_p(\varepsilon))| \leq \mathcal{O}(\varepsilon),$$
$$|\kappa_\mu^2(p) - \kappa_\mu^2(D_p(\varepsilon))| \leq \mathcal{O}(\varepsilon).$$

However, the cylindrical domain descriptors of the present work may determine in general the squares of the principal curvatures with better truncation error than their spherical domain counterparts.

Corollary 6.3. *Denote $\lambda(p,\varepsilon) \equiv \lambda(\mathrm{Cyl}_p(\varepsilon)), V_p(\varepsilon) \equiv V(\mathrm{Cyl}_p(\varepsilon))$ the integral invariants of a cylindrical domain on a hypersurface $\mathcal{S}$, then the corresponding curvature descriptors at scale $\varepsilon > 0$ and point $p \in \mathcal{S}$, for any $\mu = 1, \ldots, n$, are:*

$$\mathcal{R}(\mathrm{Cyl}_p(\varepsilon)) = \frac{2(n+2)}{\varepsilon^2}\left[\frac{2(n+4)}{\varepsilon^2}\frac{\lambda_{n+1}(p,\varepsilon)}{V_n(\varepsilon)} + 3\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right)\right] \tag{50}$$

$$H(\mathrm{Cyl}_p(\varepsilon)) = (\pm)\sqrt{\frac{2(n+2)}{\varepsilon^2}\left[\frac{2(n+4)}{\varepsilon^2}\frac{\lambda_{n+1}(p,\varepsilon)}{V_n(\varepsilon)} + 2\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right)\right]}, \tag{51}$$

$$\kappa_\mu^2(\mathrm{Cyl}_p(\varepsilon)) = \frac{n+2}{\varepsilon^2}\left[\frac{n+4}{\varepsilon^2}\left(\frac{\lambda_\mu(p,\varepsilon)}{V_n(\varepsilon)} - \frac{\varepsilon^2}{n+2}\right) - \frac{V_p(\varepsilon)}{V_n(\varepsilon)} + 1\right], \tag{52}$$

where the overall sign can be chosen by fixing a normal orientation from

$$(\pm) = \mathrm{sgn}\langle\, \boldsymbol{e}_{n+1}(\mathrm{Cyl}_p(\varepsilon)),\, \boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon))\,\rangle.$$

The eigenvectors $\boldsymbol{e}_\mu(\mathrm{Cyl}_p(\varepsilon))$ *and* $\boldsymbol{e}_{n+1}(\mathrm{Cyl}_p(\varepsilon))$ *are descriptors of the principal and normal directions respectively. The truncation errors are:*

$$|H^2(p) - H^2(\mathrm{Cyl}_p(\varepsilon))| \le \mathcal{O}(\varepsilon^2), \qquad |\mathcal{R}(p) - \mathcal{R}(\mathrm{Cyl}_p(\varepsilon))| \le \mathcal{O}(\varepsilon^2),$$
$$|\kappa_\mu^2(p) - \kappa_\mu^2(\mathrm{Cyl}_p(\varepsilon))| \le \mathcal{O}(\varepsilon^2).$$

Proof. Solving for the next-to-leading order term in the volume formula (27), and for the normal eigenvalue in equation (33), we get a system of two equations $H^2 - \mathcal{R} = A(\varepsilon)$, $3H^2 - 2\mathcal{R} = B(\varepsilon)$, whose solution is $H^2 = B - 2A$ and $\mathcal{R} = B - 3A$, where

$$A(\varepsilon) = \frac{2(n+2)}{\varepsilon^2}\left(\frac{V_p(\varepsilon)}{V_n(\varepsilon)} - 1\right) + \mathcal{O}(\varepsilon^2), \qquad B(\varepsilon) = \frac{4(n+2)(n+4)}{\varepsilon^4}\frac{\lambda_{n+1}(p,\varepsilon)}{V_n(\varepsilon)} + \mathcal{O}(\varepsilon^2).$$

Finally, solving for κ_μ^2 from the tangent eigenvalue equation (32), and using $A(\varepsilon) = \sum_\alpha \kappa_\alpha^2$, the last formula obtains. □

A concrete algorithm using hypersurfaces to implement this procedure to estimate the Riemann curvature tensor of a Riemannian submanifold of arbitrary codimension is presented in [28].

7. Conclusions

We have studied kernel domains of local principal component analysis determined by the intersection of embedded Riemannian submanifolds with higher-dimensional cylinders and balls in the ambient space. We have introduced a generalization of the classical third fundamental form to submanifolds of any codimension and showed how it relates to the Weingarten and Ricci operators. Then the covariance analysis of these domains was found to have local eigenvalues encoding curvature in terms of the third fundamental form. For cylindrical domains the first n eigenvalues are related to those of the normal trace of the third fundamental form operator and the corresponding eigenvectors converge to its principal directions, whereas the last k eigenvalues and eigenvectors are related to the tangent trace of this tensor. In the case of the spherical domain the tangent eigenvalues and eigenvectors of the covariance matrix are related to the Weingarten operator at the mean curvature normal vector, and the normal eigenvalue decomposition is the same as in the previous case. These results show how integral invariants in terms of the local eigenvalue decomposition at scale relate to curvature information traditionally characterized by differential-geometric invariants. We have expressed the most general form of this correspondence as an asymptotic equality between the ratio of the difference and product of local eigenvalues and the difference of generalized principal curvatures. These results represent a fundamental step towards understanding the underlying connection between the statistics of point clouds sampled from Riemannian submanifolds and their geometry, a core goal of manifold learning, e.g., for optimization inside higher-dimensional matrix manifolds.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Louis Scharf for very helpful discussions during the writing of this paper. J.Á.-V. would like to thank Miguel Dovale for helpful insights. This paper is based on research partially supported by the National Science Foundation under Grants No. DMS-1513633, and DMS-1322508.

Appendix A. Integration of monomials over spheres

Let $\boldsymbol{x} = [x^1, \dots, x^n]^T \in \mathbb{R}^n$, and denote unit the sphere and ball of radius ε in $\mathbb{R}^n$ by:

$$\mathbb{S}^{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}, \quad B^n(\varepsilon) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \leq \varepsilon\}.$$

All the integrals in the text are separated into radial and angular parts using general spherical coordinates, $(\rho, \phi_1, \dots, \phi_{n-1})$, where $\rho = \|\boldsymbol{x}\|$. However, to integrate monomials over the unit sphere, it is sufficient to work with the direction cosines $\overline{x}^\mu := x^\mu/\rho \in \mathbb{S}^{n-1}$, instead of the angles, so that the formulas below apply straightforwardly.

Definition A.1. For any integers $p_1, \dots, p_n \in \{0, 1, 2, \dots\}$, the integrals of the monomials $(x^1)^{p_1} \cdots (x^n)^{p_n}$ over the unit sphere are denoted by:

$$C^{(n)}_{p_1 \cdots p_n} = \int\limits_{\mathbb{S}^{n-1}} (\overline{x}^1)^{p_1} \cdots (\overline{x}^n)^{p_n} \, d\,\mathbb{S}^{n-1}, \tag{A.1}$$

where $d\,\mathbb{S}^{n-1} = \sin^{n-2}(\phi_1)\sin^{n-3}(\phi_2)\cdots\sin(\phi_{n-2})d\phi_1\cdots d\phi_{n-1}$, is the induced Euclidean measure on the sphere, abbreviated to $d\,\mathbb{S}$ in the text since our dimension n is arbitrary but fixed throughout.

The following formula is crucial to the computations of the present paper.

Theorem A.2. [37] *Let $b_i = \frac{1}{2}(p_i + 1)$, then the values of the integrals (A.1) over spheres are*

$$C^{(n)}_{p_1 \ldots p_n} = \begin{cases} 0, & \textit{if some } p_i \textit{ is odd,} \\ 2\dfrac{\Gamma(b_1)\Gamma(b_2)\cdots\Gamma(b_n)}{\Gamma(b_1+b_2+\cdots+b_n)}, & \textit{if all } p_i \textit{ are even.} \end{cases} \tag{A.2}$$

Example A.3. We shall need the relations among integrals of monomials of even powers:

$$C_2 = \int\limits_{\mathbb{S}^{n-1}} (\overline{x}^1)^2 \, d\,\mathbb{S} = 2\frac{\Gamma(\frac{3}{2})\Gamma(\frac{1}{2})^{n-1}}{\Gamma(\frac{3}{2}+\frac{n-1}{2})} = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}, \quad C_{22} = \int\limits_{\mathbb{S}^{n-1}} (x^1)^2(x^2)^2 \, d\,\mathbb{S} = \frac{C_2}{n+2},$$

$$C_4 = \int\limits_{\mathbb{S}^{n-1}} (\overline{x}^1)^4 \, d\,\mathbb{S} = \frac{3\,C_2}{n+2} = 3\,C_{22}, \quad C_{222} = \int\limits_{\mathbb{S}^{n-1}} (x^1)^2(x^2)^2(x^3)^2 \, d\,\mathbb{S} = \frac{C_2}{(n+2)(n+4)},$$

$$C_{24} = \int\limits_{\mathbb{S}^{n-1}} (\overline{x}^1)^2(\overline{x}^2)^4 \, d\,\mathbb{S} = \frac{3\,C_2}{(n+2)(n+4)} = 3\,C_{222}$$

$$C_6 = \int\limits_{\mathbb{S}^{n-1}} (\overline{x}^1)^6 \, d\,\mathbb{S} = \frac{15\,C_2}{(n+2)(n+4)} = 15\,C_{222}.$$

The volume of a ball of radius ε, and the area of the unit sphere satisfy $V_n(\varepsilon) = \mathrm{Vol}(B^n(\varepsilon)) = \varepsilon^n C_2$, $S_{n-1} = \mathrm{Area}(\mathbb{S}^{n-1}) = n\,C_2$.

The integral of a general combination of coordinates depends on the superindices involved, which must not be confused with exponents. For instance

$$\int\limits_{\mathbb{S}^{n-1}} \overline{x}^\mu \overline{x}^\nu \overline{x}^\beta \overline{x}^\gamma \, d\,\mathbb{S} = C_4(\mu\nu\beta\gamma) + C_{22}\left[(\mu\nu\beta\gamma) + (\mu\nu\beta\gamma) + (\mu\nu\beta\gamma)\right]$$

is the general value of the integral of any product of 4 coordinates, that can be all equal to produce C_4, or be a couple of different pairs to result in C_{22}. We introduce the following notation:

$$(\mu\nu\beta\gamma) = \delta_{\mu\nu}\,\delta_{\beta\gamma}\not{\delta}_{\mu\beta},$$

so that the symbol is 1 only when the connected indices are equal and the nonconnected indices are different, and 0 otherwise, and where $\not{\delta}_{\mu\beta} := (1 - \delta_{\mu\beta})$ is the negation of the Kronecker delta, i.e., nonzero only if $\mu \neq \beta$. An example of order 6 is

$$(\mu\nu\alpha\beta\gamma\delta) = \delta_{\mu\gamma}\,\delta_{\nu\delta}\,\delta_{\alpha\beta}\not{\delta}_{\mu\nu}\not{\delta}_{\mu\alpha}\not{\delta}_{\nu\alpha}.$$

References

[1] M. Connolly, Measurement of protein surface shape by solid angles, J. Mol. Graph. 4 (1) (1986) 3–6.
[2] F. Cazals, M. Pouget, Estimating differential quantities using polynomial fitting of osculating jets, in: Eurographics Symposium on Geometry Processing, 2003, pp. 177–178.
[3] F. Cazals, M.p. Pouget, Molecular shape analysis based upon the Morse-Smale complex and the Connolly function, in: Proc. Symp. Comp. Geometry, 2003.
[4] U. Clarenz, M. Rumpf, A. Telea, Robust feature detection and local classification for surfaces based on moment analysis, IEEE Trans. Vis. Comput. Graph. 10 (2003) 516–524.
[5] U. Clarenz, M. Griebel, M. Rumpf, M. Schweitzer, A. Telea, Feature sensitive multiscale editing on surfaces, Vis. Comput. 20 (5) (2004) 329–343.
[6] S. Manay, B.-W. Hong, A.J. Yezzi, S. Soatto, Integral invariant signatures, in: T. Pajdla, J. Matas (Eds.), Computer Vision - ECCV 2004, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 87–99.
[7] S. Manay, D. Cremers, B.-W. Hong, A.J. Yezzi, Integral invariants for shape matching, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1602–1618.
[8] P. Alliez, D. Cohen-Steiner, Y. Tong, M. Desbrun, Voronoi-based variational reconstruction of unoriented point sets, in: Proceedings of the Fifth Eurographics Symposium on Geometry Processing, SGP '07, Eurographics Association, Aire-la-Ville, Switzerland, 2007, pp. 39–48.
[9] J. Berkmann, T. Caelli, Computation of surface geometry and segmentation using covariance techniques, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1994) 1114–1116.
[10] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, W. Stuetzle, Surface reconstruction from unorganized points, SIGGRAPH Comput. Graph. 26 (2) (1992) 71–78.
[11] Y.-L. Yang, Y.-K.-K. Lai, S.-M. Hu, H. Pottmann, Robust principal curvatures on multiple scales, in: Proc. Eurographics Symposium on Geometry Processing, 2006, pp. 223–226.
[12] Y.-K. Lai, S.-M. Hu, T. Fang, Robust principal curvatures using feature adapted integral invariants, in: Proceedings - SPM 2009: SIAM/ACM Joint Conference on Geometric and Physical Modeling, 2009, pp. 325–330.
[13] H. Pottmann, J. Wallner, Y.-L. Yang, Y.-K. Lai, S.-M. Hu, Principal curvatures from the integral invariant viewpoint, Comput. Aided Geom. Des. 24 (2007) 428–442.
[14] H. Pottmann, J. Wallner, Q.-X. Huang, Y.-L. Yang, Integral invariants for robust geometry processing, Comput. Aided Geom. Des. 26 (1) (2008).
[15] Q. Mérigot, Geometric Structure Detection in Point Clouds, PhD thesis, Université Nice Sophia Antipolis, 2009.
[16] Q. Mérigot, M. Ovsjanikov, L. Guibas, Voronoi-based curvature and feature estimation from point clouds, IEEE Trans. Vis. Comput. Graph. 17 (2011) 743–756.
[17] A.V. Little, Estimating the Intrinsic Dimension of High-Dimensional Data Sets: A Multiscale, Geometric Approach, PhD thesis, Duke University, 2011.
[18] A.V. Little, J. Lee, Y.-M. Jung, M. Maggioni, Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD, in: IEEE/SP 15th Workshop on Statistical Signal Processing, SSP'09, IEEE, 2009, pp. 85–88.
[19] G. Chen, A.V. Little, M. Maggioni, L. Rosasco, Wavelets and Multiscale Analysis, Springer-Verlag, 2011 (Ch. 10).
[20] A.V. Little, Y.-M. Jung, M. Maggioni, Multiscale estimation of intrinsic dimensionality of data sets, in: Proc. A.A.A.I., 2009.
[21] A.V. Little, M. Maggioni, L. Rosasco, Multiscale Geometric Methods for Data Sets I: Multiscale SVD, Noise and Curvature, Tech. rep., MIT, 2012.
[22] D. Broomhead, R. Jones, G. King, Topological dimension and local coordinates from time series data, J. Phys. A, Math. Gen. 20 (9) (1987) L563.
[23] D. Broomhead, R. Indik, A. Newell, D. Rand, Local adaptive Galerkin bases for large-dimensional dynamical systems, Nonlinearity 4 (2) (1991) 159.
[24] F. Solis, Geometric Aspects of Local Adaptive Galerkin Bases, PhD thesis, University of Arizona, 1993.
[25] F. Solis, Geometry of local adaptive Galerkin bases, Appl. Math. Optim. 41 (2000) 331–342.
[26] J. Álvarez-Vizoso, R. Arn, M. Kirby, C. Peterson, B. Draper, Geometry of curves in $\mathbb{R}^n$ from the local singular value decomposition, Linear Algebra Appl. 571 (2019) 180–202.
[27] J. Álvarez-Vizoso, M. Kirby, C. Peterson, Manifold curvature from covariance analysis, in: 2018 IEEE Statistical Signal Processing Workshop (SSP), 2018, pp. 353–357.

[28] J. Álvarez-Vizoso, M. Kirby, C. Peterson, Manifold curvature learning from hypersurface integral invariants, Linear Algebra Appl. 602 (2020) 179–205.
[29] I. Chavel, Riemannian Geometry, 2nd edition, Cambridge University Press, 2006.
[30] S. Kobayashi, K. Nomizu, Foundations of Differential Geometry, John Wiley & Sons, Inc., 1969.
[31] B. O'Neill, Semi-Riemannian Geometry, Academic Press, 1983.
[32] M. Spivak, A Comprehensive Introduction to Differential Geometry, 3rd edition, Publish or Perish, Inc., 1999.
[33] V. Toponogov, Differential Geometry of Curves and Surfaces. A Concise Guide, Birkhäuser Boston, 2006.
[34] H. Liu, U. Simon, L. Verstraelen, C. Wang, The third fundamental form metric for hypersurfaces in nonflat space forms, J. Geom. 65 (1999) 130–142.
[35] F. Rellich, Perturbation Theory of Eigenvalue Problems, Gordon and Breach, 1969.
[36] W. Kühnel, Differential Geometry: Curves-Surfaces-Manifolds, vol. 16, American Mathematical Society, 2006.
[37] G. Folland, How to integrate a polynomial over a sphere, Am. Math. Mon. 108 (5) (2001) 446–448.