DISSERTATION

COVARIANCE INTEGRAL INVARIANTS OF EMBEDDED RIEMANNIAN MANIFOLDS FOR MANIFOLD LEARNING

Submitted by

Javier Álvarez Vizoso

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Advisor: Christopher Peterson
Co-Advisor: Michael Kirby

Dan Bates
Renzo Cavalieri
Richard Eykholt

Copyright by Javier Álvarez Vizoso 2018

All Rights Reserved

ABSTRACT

COVARIANCE INTEGRAL INVARIANTS OF EMBEDDED RIEMANNIAN MANIFOLDS FOR MANIFOLD LEARNING

This thesis develops an effective theoretical foundation for the integral invariant approach to study submanifold geometry via the statistics of the underlying point-set, i.e., Manifold Learning from covariance analysis. We perform Principal Component Analysis over a domain determined by the intersection of an embedded Riemannian manifold with spheres or cylinders of varying scale in ambient space, in order to generalize to arbitrary dimension the relationship between curvature and the eigenvalue decomposition of covariance matrices. In the case of regular curves in general dimension, the covariance eigenvectors converge to the Frenet-Serret frame and the corresponding eigenvalues have ratios that asymptotically determine the generalized curvatures completely, up to a constant that we determine by proving a recursion relation for a certain sequence of Hankel determinants. For hypersurfaces, the eigenvalue decomposition has series expansion given in terms of the dimension and the principal curvatures, where the eigenvectors converge to the Darboux frame of principal and normal directions. In the most general case of embedded Riemannian manifolds, the eigenvalues and limit eigenvectors of the covariance matrices are found to have asymptotic behavior given in terms of the curvature information encoded by the third fundamental form of the manifold, a classical tensor that we generalize to arbitrary dimension, and which is related to the Weingarten map and Ricci operator. These results provide descriptors at scale for the principal curvatures and, in turn, for the second fundamental form and the Riemann curvature tensor of a submanifold, which can serve to perform multi-scale Geometry Processing and Manifold Learning, making use of the advantages of the integral invariant viewpoint when only a discrete sample of points is available.

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Chris Peterson and Dr. Michael Kirby for their patience and support over the last two years. I am enormously indebted to them for their permanent advice, understanding and encouragement to make this project successful.

I am most grateful to Dr. Louis Scharf for many illuminating and enjoyable discussions.

I am very thankful to Dr. Renzo Cavalieri for his guidance during my first year here, and to everybody at the Department of Mathematics at CSU for making this a hardly forgettable experience.

Conversations with Nand Sharma have been an endless source of philosophical debate, for which I am really grateful. I also thank Kwancheol Shin and Zhuoran Wang for our mutual support ever since our arrival.

I would also like to thank my professors at Universidade de Santiago de Compostela and Universidad de Oviedo for their encouragement and help to continue my studies after difficult times.

Finally, I deeply want to thank my family and friends for their permanent support and patience. My early passion for physics and mathematics was heavily influenced by Moisés L. Caeiro. Great moments with Daniel Barja, David Barral, Daniel Lowe, Xabier Palacios and Rodolfo Pardo also shaped the person I am now. Unconditional support from my mother has been crucial to be able to make my dreams come true. And the transcendental role played by Miguel Dovale to climb the glass staircase of life has been and shall be fundamental until each of us fulfils his true purpose in life.

And to you, the only one that defines what was, is and shall be important.

DEDICATION

Ás derradeiras lembranzas laranxas.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

The purpose of Manifold Learning is the reconstruction of local geometry from the analysis of a subset of its points, usually a finite sample and possibly with noise. The present dissertation aims to describe an effective theoretical generalization to arbitrary dimension of a line of research whose ultimate goal and application is the recovery of submanifold curvature via the study of the statistical information of the underlying point-set.

A set of data points in some configuration space, typically a Riemannian manifold, often belongs to a lower-dimensional submanifold due to correlations among its degrees of freedom, e.g. a constrained dynamical system in phase space. From the geometric perspective, arbitrary samples of points can be generated if one knows the submanifold *a priori* from implicit equations or by given local chart parametrizations. From a data analysis point of view, only the point-set of a sample from the submanifold is known and, ideally in the limit of the number of points, one would like to characterize as uniquely as possible the geometric properties of the manifold from which those samples arise by studying the statistical properties of the set. The core of our results shows that the classical statistical concept of covariance matrix is essentially a purely geometrical one: the eigenvalue decomposition of these matrices encodes the curvature information of the third fundamental form, and its principal directions furnish an adapted frame for the tangent and normal spaces of the submanifold.

In classical differential geometry manifolds are defined intrinsically from an atlas of coordinate charts that cover the point-set with smooth transition functions between them. This definition was established historically through the abstraction of embedded submanifolds in Euclidean space: smooth subsets of ambient space which require fewer degrees of freedom to be described analytically, e.g. by local parametrizations. The classical differential geometry of curves and surfaces in space built the foundation for the types of definitions, questions and structures studied in Riemannian geometry for general manifolds of arbitrary dimension.

Regular curves in space [26], [33], [58] have as natural differential invariants the velocity and acceleration vectors with respect to arc-length, which can be completed to an orthonormal basis of the ambient space by the Gram-Schmidt method with higher derivative vectors. This provides a comoving frame, the Frenet-Serret frame, which measures how the tangent line to the curve, and its osculating planes move and rotate from point to point, providing a natural definition of curvature that generalizes the inverse of the radius of the osculating circle, tangent to the curve at every point. The fundamental theorem of regular curves states that the Frenet-Serret curvature functions completely determine the parametric curve up to rigid motion, since the curve is locally given by the solution of a system of ordinary differential equations whose coefficients are the curvature functions. Therefore, curve parametrizations and their curvature functions can be thought of as dual descriptions of the same local embedded geometric object. The case of curves in any dimension is special with respect to higher-dimensional submanifolds because one-dimensional objects do not have intrinsic curvature, which is defined by parallel transport in different tangent directions. Indeed, the mathematical tools employed in our work are very different for each scenario.

The case of surfaces in space [26], [33], [58], very similar to hypersurfaces in any dimension, was the first type of manifold with intrinsic geometry thoroughly studied by Gauß, Darboux, Weingarten and others, paving the way to the abstract generalizations of Riemann, Levi-Civita, Ricci, Cartan, and many other great mathematicians since then. A parametrized smooth surface has a tangent plane at every point with an induced metric, or first fundamental form, given by the Euclidean scalar product restricted to tangent vectors. Integration over curves inside the surface provides an intrinsic metric distance between points. The twisting and torsion of the tangent plane from point to point measures how the surface bends in ambient space. Equivalently, the change in different tangent directions of the unit vector normal to the surface encodes this curving. Since Euclidean space has a canonical notion of directional derivative, playing the role of global covariant derivative, it is natural to define the derivative of the normal vector in a tangent direction as a measure of the extrinsic curvature of the surface at the point. Indeed, moving tangentially to the surface in a certain direction provides a canonical measurement of curvature given by the acceleration of the curve inside the surface associated to that direction and point. Since there is a tangent plane worth of possible directions, the object to encode all this curvature information is a linear map, the Weingarten operator, that associates to every tangent vector the directional derivative of the normal vector. This map turns out to be self-adjoint with respect to the metric and its components in an orthonormal tangent basis represent the second fundamental form of the surface. The eigenvalues and eigenvectors of the Weingarten map are called the principal curvatures and principal directions of the surface at the point. The principal curvatures are in fact the minimum and maximum curvatures of curves inside the surface cut out by a normal plane, i.e., a plane spanned by the normal vector and a tangent direction. The corresponding eigenvectors point in these directions. The sum of the principal curvatures is the mean curvature, and their product is the Gaußian curvature; these correspond to the trace and determinant of the Weingarten operator. Gauß Theorema Egregium shows how the Gaußian curvature is an intrinsic invariant independent of the embedding that is determined by combinations of the second fundamental form components.

Hypersurfaces [18], [35], [49] share a very similar differential-geometric structure, where the main difference is that the tangent space is now higher-dimensional so the Weingarten operator has as many principal curvatures and directions as the dimension of the hypersurface. The frame given by the normal vector and principal directions is called the Darboux frame. A theorem by Bonnet, similar to the Frenet-Serret characterization of curves, is available for hypersurfaces: if smooth parametrization functions for the first and second fundamental form are given and satisfy the Gauß-Codazzi-Mainardi-Peterson equations, then a local hypersurface exists with those forms, unique up to rigid motion. Notice that, in comparison, this is now a system of partial differential equations. We can however regard the Darboux frame and the principal curvatures as all the information needed to characterize the differential geometry of a hypersurface. This information is completely encoded in the osculating quadrics that approximate the hypersurface at every point, since this is determined by the Hessian of the functions that locally parametrize the hypersurface as a graph manifold.

Embedded Riemannian manifolds [16], [32], [35], [49], [46] can be studied in a similar way by means of a second fundamental form that now takes values in the normal bundle of the submanifold. Since there are now more than one independent normal vectors, the generalized second fundamental form keeps track of how they change tangentially by means of a Weingarten operator per normal vector. Hence, principal directions and curvatures are only defined with respect to a given normal direction. In order to measure the intrinsic curvature, Riemann introduced a tensor which is essentially the only one constructible from the metric and its first and second derivatives, and linear in the latter. The Riemann tensor is zero if and only if the manifold is locally flat, i.e., if there is a chart where the metric is Euclidean. It equivalently measures how initially parallel geodesics, the straightest intrinsic lines or curves of shortest length, deviate because of the manifold curvature, the essential feature of non-Euclidean geometries. It can be defined as a differential-invariant by the non-commutativity of the second covariant derivative. This reflects the non-integrability of parallel transport around an infinitesimal closed loop. The Gauß equation again expresses the Riemann tensor as a product of components of the second fundamental form, thus relating intrinsic and extrinsic curvature. Taking traces of this tensor generates other objects like the sectional curvature, which for every plane in the tangent space measures the Gaußian curvature of the geodesic surface tangent to that plane. The Ricci tensor and the scalar curvature are further contractions over the degrees of freedom of this tensor.

The purpose of our work [4–7] is to show the relationship between these classical differential invariants and local integral invariants in general dimension. In particular, we shall see that integration over small domains on a submanifold encodes the same curvature information as the differential-geometric tensors. Performing integration can be computationally better behaved than differentiation since it reduces to sums in the discrete case, which have a naturally averaging nature (e.g. for noise concerns), in contrast to the finite-differences and quotients in differential approximations. Since integrals over regions have a natural scale, the curvature information that can be obtained provides multi-scale descriptors of geometric features.

Integral invariants from Principal Component Analysis were introduced in [21], [13, 14], [19, 20], [41, 42] as theoretical tools to perform Manifold Learning and Geometry Processing of low-dimensional submanifolds, like curves and surfaces in space. They have been used for shape and feature detection at scale as geometric low-pass filters, [3], [11], [19, 20], [29], [42], [60]. The focus in these settings has been on curves, surfaces and the study of stability with respect to noise [34], [50, 51]. Voronoi-based covariance matrices have also been of interest, [44, 45], where a relationship to the derivative of the normal vector is found for hypersurfaces.

Descriptors can be interpreted as approximations of certain characteristic variables of a system given in terms of other relevant information. The seminal work of [51] developed the idea of performing covariance matrix analysis over domains on surfaces determined by balls in space in order to recover curvature information at scale. The series expansion of the eigenvalue decomposition was shown to reproduce the principal curvatures at second order, and the limit eigenvectors were shown to converge to the principal curvatures and normal direction. In [60], [50], [34] the stability and robustness of this viewpoint was studied both theoretically and computationally. The theoretical results of [51] are generalized in this thesis to hypersurfaces and Riemannian manifolds of arbitrary codimension. The higher-dimensional integrals and approximations involved become much more complicated and a unifying approach and notation is taken in our proofs, using spherical coordinates and integration of monomials over spheres [23].

The other side of this theory is the study of finite point clouds and how their discrete PCA covariance matrices converge with the number of points to the exact analytical result. Multi-scale SVD methods, using geometric measure theory and harmonic analysis, have been developed [36, 38], [17, 37] in order to study noisy samples from probability distributions supported on submanifolds of a high-dimensional Euclidean space [39]. In these works, ranges of scales are determined, taking into account curvature, for the covariance matrices to be most informative and close to the noisy empirical matrices. In particular, the leading order term of the eigenvalues is obtained and it is seen that tangent and normal eigenvalues scale differently, with the specific expression of the normal eigenvalue to leading order in terms of the principal curvatures in the hypersurface case. In the case of general codimension, the authors obtain a similar conclusion, following [12] and [55, 56] discussed below. In this sense, the noisy point cloud approach of [39] is complemented by the approach of the present dissertation that computes the covariance analysis of the smooth point-set for different types of kernel domains. We obtain in particular the next to leading order terms of the tangent eigenvalues for the complete smooth data set, and the normal eigenvalues leading term in general codimension, providing the direct theoretical link between curvature and covariance, i.e. between differential and integral invariants. Since [39] develops an explicit algorithm for the estimation of the dimension of the manifold, a natural next step would be to expand these multiscale SVD methods in order to apply them to our main theorems and thus to estimate curvature from noisy point clouds. Our curvature descriptors at scale aim to fulfil this task.

Adapted frames from the eigenvalue decomposition of covariance matrices of spherical intersection domains were introduced in [12], [55, 56], in order to obtain local adaptive Galerkin bases for dynamical systems. This was motivated by the study of the long-term behavior of dynamical systems confined to an invariant manifold. Choosing a general operator in an optimal way, in order to reflect natural nonlinear structures of the system, leads as well to the covariance matrix integral invariant. This allowed the authors to obtain the dimension of the submanifold and approximations at scale of its tangent and normal spaces, even for manifolds with general measures. However, the approximations made in [55, 56] for submanifolds of Euclidean space reduce to the leading order terms of the cylindrical case studied in the present thesis, which do not single out the geometrically natural frame specified by the limit eigenvectors, nor the curvature information hidden in the eigenvalues. We complete the analysis of the covariance matrix series to the next order to obtain explicitly this information in terms of the traces of the third fundamental form tensor.

For this, we generalize to arbitrary Riemannian submanifolds the notions of integral invariants based on the volume, barycenter and covariance matrix of a point-set, weighted by the induced measure of the submanifold. This can be done via the exponential map [16], [46] of the ambient manifold, which uses the lengths of geodesics tangent to an orthonormal frame at a point as the

Riemannian generalization of Cartesian coordinates. Measuring the geodesic normal coordinates of the points of a submanifold domain provides a general definition for these integral invariants. Normal coordinates are naturally used to make geometric measurements needed to perform probability and statistics inside Riemannian manifolds, e.g. [47, 48]. Optimization on Riemannian manifolds [1, 2] has been studied assuming the underlying geometry is known, for which any characterization and reconstruction techniques through Manifold Learning would be the first step needed to perform optimization.

These integrals perform Principal Component Analysis on domains determined by the submanifold, so they feature a scale dependent behavior. This type of analysis can be understood in two different ways: from a physics perspective, the integral invariants measure the total mass of the domain, the center of mass point, and an analogue of the moments of inertia with the corresponding principal directions; from a statistical perspective, these integrals compute the total volume of the data set, its average point, and the covariance matrix of its degrees of freedom. The volume and barycenter can be easily defined as the integral over the domain of the identity and the position vector respectively, weighed by the induced measure on the submanifold. The covariance matrix of the domain with respect to a fixed point is constructed by choosing an orthonormal frame at the point and measuring the coordinates of the other points in the region; the pairwise products of these coordinates determine a matrix function, whose integration over the domain yields a matrix dependent only on the scale of the region. We are interested in regions determined by the intersection of the submanifold with balls and cylinders in ambient space, which have a natural radius as scale. Since the matrix so constructed is frame dependent but symmetric, the covariance integral invariants of interest are to be defined as its eigenvalues and eigenvectors.

This covariance analysis can be thought of as the study of a matrix-valued function of scale at every point of the manifold, which can be given a Taylor series expansion by the classical perturbation theory of Hermitian matrices [52]. One of the main results of our analysis is that the eigenvectors converge, when the scale tends to zero, to a special orthonormal frame of the tangent and normal spaces of the manifold, specifying an adapted frame which turns out to have geometric meaning as generalized principal directions. Precisely, the scaling behavior of the eigenvalues permits the detection of which eigenvectors span either the tangent or normal spaces in the limit. Moreover, our main result is the computation of the asymptotic expansion with scale of the eigenvalues, to second order, in order to find curvature information in the Taylor coefficients. For this we need to introduce the generalization to arbitrary codimension of the classical third fundamental form.

We also study the volumes of these regions. Geodesic balls inside manifolds have intrinsic volume with asymptotic series given as corrections to the Euclidean ball volume, completely determined by intrinsic scalar curvature invariants [27]. These invariants also appear in the volume of tubes generated by the normal flow of an embedded manifold [25]. In our case, the domains of integration depend on the embedding of the submanifold, so the extrinsic curvature will play a crucial role in the volume corrections, as found in [30].

The structure of the dissertation is as follows.

In chapter 2 we study the notion of PCA integral invariants within the context of Riemannian Geometry. In section §2.1 we give an explicit coordinate expression for the first fundamental form and the induced measure on graph embedded manifolds. Then we overview the geometry of Riemannian submanifolds, where curvature is classically defined as a differential invariant via the the second fundamental form, whose local coordinate expression will be crucial for our computations. The particular case of hypersurfaces is reviewed. In section §2.2 PCA integral invariants are defined in a general setting using the exponential map and given by the volume, barycenter and covariance matrix of a domain determined by the submanifold. In particular, we shall study PCA kernel domains delimited by the intersection with balls and higher-dimensional cylinders in ambient space. Geometric descriptors are introduced to show how the study of hypersurfaces is sufficient to build descriptors in any codimension, by applying the analysis on the manifold hypersurface projections, since principal curvatures and principal directions determine the local Hessians and, therefore, the second fundamental form and Riemann tensor via the Gauß equation.

In chapter 3 we deal with the case of regular curves in Euclidean space of any dimension, which requires completely different tools since curves do not have intrinsic curvature. In section §3.1 we recall the Frenet-Serret apparatus and how the Frenet curvatures completely determine the curve up to rigid motion. We state previous results known to obtain the Frenet-Serret frame and curvatures from covariance analysis, and arrive at an asymptotic formula that relates the ratio of the eigenvalues of the covariance matrix to the Frenet curvatures by the recursion relation of certain Hankel determinants. In order to prove this formula, the theory of moments and orthogonal polynomials is reviewed in section §3.2, from which the recursion relation of a general family of Hankel determinants is obtained.

In chapter 4 three different domains for hypersurfaces are studied. In section §4.1 the integral invariants are computed for a volume region delimited by a hypersurface inside a ball centered at a point of the hypersurface. We need to prove a fundamental lemma to approximate the corresponding integrals to high enough order. The asymptotic expansion of the invariants with respect to the scale of the ball are shown to be given in terms of the principal curvatures and the dimension, and the eigenvectors of the covariance matrix are shown to converge in the scale limit to the principal and normal directions. In section §4.2 the analogous analysis is carried out for the integral invariants of the hypersurface patch cut out by the ball and a higher-dimensional cylinder. The patch covariance eigenvalues reproduce the principal curvatures as well but either squared or multiplied by the mean curvature; the corresponding eigenvectors converge again to the principal and normal directions when all curvatures are different.

In chapter 5 we study the most general setting of the present work: embedded Riemannian manifolds of arbitrary dimension. In section §5.1 the classical third fundamental form is generalized to submanifolds of general codimension by means of the metric product of any two Weingarten maps, measuring the curvature of the induced normal connection on the manifold via the Ricci equation. Its different traces are shown to relate to the Weingarten map at the mean curvature vector and the Ricci operator. In section §5.2 we compute the volume, barycenter and covariance matrix of a cylindrical domain inside an embedded submanifold. In particular, for generic cylinders, 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 via the span of the corresponding limit eigenvectors. Moreover, for normal cylinders, the next-to-leading order term in the asymptotic series of the covariance eigenvalues is determined by the eigenvalues of the tangent and normal traces of the third fundamental form. The limit eigenvectors then converge to the principal directions determined by these tensors in the tangent and normal spaces. In section §5.3 an analogous analysis is carried out for the domain given by the intersection of a ball in ambient space with the manifold, which introduces a considerable number of correction terms with respect to the previous case. This leads to an eigenvalue decomposition of the covariance matrix with essentially the same normal part as the cylindrical case, and with tangent part given in terms of the Weingarten operator at the mean curvature normal.

Finally, in chapter 6 all previous results are used to produce estimators and get the most general asymptotic ratio between eigenvalues and curvature. In section §6.1 we see how the volume and eigenvalue asymptotic formulas can be inverted and truncated to yield geometric descriptors of the principal curvatures and principal directions of hypersurfaces, thus establishing concrete formulas to use in the general method outlined for Riemannian submanifold. In section §6.2 we obtain the limit ratios of the covariance eigenvalues, in the cylindrical and spherical domain cases, in terms of those of the third fundamental form, generalizing the asymptotic ratios found for regular curves. The descriptors that these domains provide also recover the principal curvatures and directions, where the cylindrical estimators have a better truncation error than the general spherical case.

In appendix A we set the notation for the spherical coordinates used, and review the formula for the integrals of monomials over spheres and balls. We also define specific symbols to encapsulate the possible values of these integrals under arbitrary products of coordinates that depend on the indices involved.

These results show how Principal Component Analysis can be carried out on a general embedded Riemannian submanifold to probe its local geometry. From a theoretical point of view our work establishes the generalization of the relationship between the statistical covariance analysis of the underlying point-set of a submanifold and the classical differential-geometric curvature using the third fundamental form. From the applied and computational point of view, the integral invariant approach used in the literature to perform Geometry Processing of low-dimensional manifolds can be employed with embedded manifolds of arbitrary dimension via the study of the hypersurface descriptors obtained here. This opens the way for computational implementations of Manifold Learning with big data sets, and the potential detection and classification of features of this data via the curvature profile of its embedded geometric representation.

Chapter 2

PCA Integral Invariants in Differential Geometry

In this chapter we review the differential geometry of Riemannian submanifolds [16], [18], [32], [33], [35], [46], [49], [57]. We introduce the first and second fundamental forms, give local expressions for them for graph submanifolds, and define the Riemann curvature tensor. The case of hypersurfaces and their principal curvatures and directions is covered as well. Then we generalize the definition of PCA integral invariants and descriptors to this general setting using geodesic normal coordinates, and define the cylindrical and spherical intersection domains via the exponential map. Finally, we see how the covariance analysis of hypersurfaces is enough to obtain descriptors for submanifolds of general codimension via the local expression of the second fundamental form and the Gauß equation.

2.1 Geometry of Riemannian Submanifolds

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 $\mathcal{N}$.

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 $[x^1, \ldots, x^n, f^1(\boldsymbol{x}), \ldots, f^k(\boldsymbol{x})]^T$ 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 quadric.

Lemma 2.1.1. *The first fundamental form components of a graph manifold $\mathcal{M} \subset \mathbb{R}^{n+k}$, parametrized by $[x^1, \ldots, x^n, f^1(\boldsymbol{x}), \ldots, f^k(\boldsymbol{x})]^T \in T_p\mathcal{M} \oplus N_p\mathcal{M} \cong \mathbb{R}^{n+k}$, are:*

$$g_{\mu\nu}(\boldsymbol{x}) = \delta_{\mu\nu} + \sum_{j=1}^{k} \frac{\partial f^j}{\partial x^\mu}\frac{\partial f^j}{\partial x^\nu}. \tag{2.1}$$

The induced measure on $\mathcal{M}$ in these coordinates is given by

$$\mathrm{dVol} = \sqrt{\det g(\boldsymbol{x})}\, d^n\boldsymbol{x} = \left(1 + \frac{1}{2}\sum_{j=1}^{k}\sum_{\alpha=1}^{n}\left[\sum_{\beta=1}^{n}\left(\frac{\partial^2 f^j}{\partial x^\alpha \partial x^\beta}(0)\right)x^\beta\right]^2 + \mathcal{O}(x^3)\right) d^n\boldsymbol{x}. \qquad (2.2)$$

Proof. The tangent space in these coordinates is spanned by the vectors

$$\boldsymbol{X}_\mu = \frac{\partial}{\partial x^\mu}[x^1, \ldots, x^n, f^1(\boldsymbol{x}), \ldots, f^k(\boldsymbol{x})]^T = [0, \ldots, 1, \ldots, 0, \frac{\partial f^1}{\partial x^\mu}, \ldots, \frac{\partial f^k}{\partial x^\mu}]^T,$$

for $\mu = 1, \ldots, n$, which yields the canonical orthonormal basis at p since $\nabla f^j(\mathbf{0}) = \mathbf{0}$. The induced metric tensor is then

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

From this, recalling that the $f^j(\boldsymbol{x})$ have Taylor expansions starting at order 2 in these coordinates, the matrix of the metric components is of the form $[g] = \mathrm{Id}_n + [h]$, where the correction matrix $[h] = [\sum_j \partial_\mu f^j \partial_\nu f^j]$ is small because we are in a neighborhood of $\mathbf{0}$ with $\nabla f^j(\mathbf{0}) = \mathbf{0}$. Let

$$f^j(\boldsymbol{x}) = \frac{1}{2}\sum_{\alpha,\beta=1}^{n}\left(\frac{\partial^2 f^j}{\partial x^\alpha \partial x^\beta}(0)\right)x^\alpha x^\beta + \mathcal{O}(x^3),$$

for every $j = 1, \ldots, k$, then

$$\frac{\partial f^j}{\partial x^\mu} = \sum_{\beta=1}^{n}\left(\frac{\partial^2 f^j}{\partial x^\beta \partial x^\mu}(0)\right)x^\beta + \mathcal{O}(x^2).$$

The natural volume form of a Riemannian manifold is given by $\sqrt{\det g}\; dx^1 \wedge \cdots \wedge dx^n$, [46, Ch. 7, Lem. 19], whose lowest order approximation is $\det g \approx 1 + \mathrm{tr}\, h$, so $\sqrt{\det g} \approx 1 + \frac{1}{2}\mathrm{tr}\, h$, i.e.,

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

$\square$

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 $(\,\cdot\,)^{\top} : T_p\mathcal{N} \to T_p\mathcal{M}$. The *second fundamental form* $\mathbf{II}$ *of* $\mathcal{M}$ *in* $\mathcal{N}$ is defined to be the normal projection of the ambient covariant derivative when acting on vector fields tangent to $\mathcal{M}$, i.e., denoting $(\,\cdot\,)^{\perp} : T_p\mathcal{N} \to N_p\mathcal{M}$,

$$\mathbf{II}(\boldsymbol{x}, \boldsymbol{y}) = (\overline{\nabla}_{\boldsymbol{y}}\boldsymbol{X})^{\perp}, \quad \text{i.e.,} \quad \overline{\nabla}_{\boldsymbol{y}}\boldsymbol{X} = \nabla_{\boldsymbol{y}}\boldsymbol{X} + \mathbf{II}(\boldsymbol{x}, \boldsymbol{y}), \tag{2.3}$$

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:

$$\langle\, \mathbf{II}(\boldsymbol{x}, \boldsymbol{y}),\, \boldsymbol{n} \,\rangle = \langle\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\, \boldsymbol{x},\, \boldsymbol{y} \,\rangle = \langle\, \boldsymbol{x},\, \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\, \boldsymbol{y} \,\rangle. \tag{2.4}$$

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:

$$\mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^{k} \mathrm{II}^j(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\boldsymbol{n}_j = \sum_{j=1}^{k} \langle\, \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu),\, \boldsymbol{n}_j \,\rangle\, \boldsymbol{n}_j = \sum_{j=1}^{k} \langle\, \widehat{\boldsymbol{S}}_j\, \boldsymbol{e}_\mu,\, \boldsymbol{e}_\nu \,\rangle\, \boldsymbol{n}_j. \tag{2.5}$$

The geometric meaning of $\mathbf{II}$ 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. [16, 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, [46, 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, $\mathbf{II}$ naturally measures the extrinsic curvature of the embedding since it represents the forced curving of the straightest lines in $\mathcal{M}$ due to the curving of $\mathcal{M}$ itself in $\mathcal{N}$.

The inverse function theorem and [32, Ch. VII, Ex. 3.3] establish the following lemma, of fundamental importance for the computations in the proofs of this dissertation.

Lemma 2.1.2. *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 $(y^1, \ldots, y^{n+k})$ in $\mathcal{N}$ such that the coordinate tangent vectors at the origin $\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^n$ coincide with $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$, and $\boldsymbol{Y}^{n+1}, \ldots, \boldsymbol{Y}^{n+k}$ 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 $y^1 = x^1, \ldots, y^n = x^n, y^{n+1} = f^1(\boldsymbol{x}), \ldots, y^{n+k} = f^k(\boldsymbol{x})$, such that the components of the second fundamental form at p can be written as:*

$$\mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^{k} \left[\frac{\partial^2 f^j}{\partial x^\mu \partial x^\nu}(0) \right] \boldsymbol{n}_j. \tag{2.6}$$

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:

$$\boldsymbol{H} = \sum_{\mu=1}^{n} \mathbf{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) = \sum_{j=1}^{k} H^j \boldsymbol{n}_j, \qquad \text{where } H^j = \sum_{\mu=1}^{n} \mathrm{II}^j(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu). \tag{2.7}$$

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*

$$\boldsymbol{\mathcal{R}ic}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\mu=1}^{n} \langle \boldsymbol{R}(\boldsymbol{e}_\mu, \boldsymbol{x})\boldsymbol{y},\, \boldsymbol{e}_\mu \rangle = \langle \hat{\boldsymbol{\mathcal{R}}}\, \boldsymbol{x},\, \boldsymbol{y} \rangle,$$

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 bilinear form with respect to the metric.

Gauß Theorema Egregium establishes that the intrinsic curvature of surfaces is a particular combination of products of the components of the second fundamental form. This generalizes to higher dimension in

Theorem 2.1.3 (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*

$$\langle \boldsymbol{R}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{z},\, \boldsymbol{w}\rangle = \langle\, \overline{\boldsymbol{R}}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{z},\, \boldsymbol{w}\,\rangle + \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{w}),\, \mathbf{II}(\boldsymbol{y},\boldsymbol{z})\,\rangle - \langle\, \mathbf{II}(\boldsymbol{x},\boldsymbol{z}),\, \mathbf{II}(\boldsymbol{y},\boldsymbol{w})\,\rangle \tag{2.8}$$

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

A hypersurface $\mathcal{S}$ is an embedded manifold of codimension 1, many of whose properties generalize those of surfaces in $\mathbb{R}^3$. Its second fundamental form can also be introduced via the Weingarten map, or shape operator $\widehat{\boldsymbol{S}}$ defined as follows: given a choice of unit normal vector field $\boldsymbol{N}$ around $p \in \mathcal{S}$, there is a linear endomorphism of $T_p\mathcal{S}$ given by

$$\widehat{\boldsymbol{S}}(\boldsymbol{x}) = -\overline{\nabla}_{\boldsymbol{x}}\boldsymbol{N}, \quad \forall \boldsymbol{x} \in T_p\mathcal{M}$$

such that the classical second fundamental form is related to the one defined above by:

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

The Weingarten map encodes how the hypersurface normal vector varies in the ambient space when moving in a direction tangent to the hypersurface, thus measuring curvature. Moreover, $\widehat{\boldsymbol{S}}$ is self-adjoint with respect to the metric so there is an orthonormal basis of $T_p\mathcal{S}$ given by its eigenvectors called the *principal directions* of $\mathcal{S}$ at p. The corresponding eigenvalues are called *principal curvatures*, $\{\kappa_\mu(p)\}_{\mu=1}^n$, because $\langle\, \widehat{\boldsymbol{S}}(\boldsymbol{u}),\, \boldsymbol{u}\,\rangle$ measures the normal acceleration of a curve inside $\mathcal{S}$ with unit tangent $\boldsymbol{u}$. The 2-plane spanned by a tangent vector $\boldsymbol{u} \in T_p\mathcal{S}$ and the normal vector $\boldsymbol{n} \in N_p\mathcal{S}$ intersects the hypersurface in a normal section curve whose first Frenet-Serret curvature is precisely the normal curvature given by $\langle\, \widehat{\boldsymbol{S}}(\boldsymbol{u}),\, \boldsymbol{u}\,\rangle$. For any tangent vector the normal section curvature is

$$\kappa(\boldsymbol{x}) = \frac{\mathrm{II}(\boldsymbol{x}, \boldsymbol{x})}{\mathrm{I}(\boldsymbol{x}, \boldsymbol{x})}, \quad \forall \boldsymbol{x} \in T_p\mathcal{S}. \tag{2.9}$$

Furthermore, one can define elementary curvature scalars $K_1(p), \ldots, K_n(p)$ as the elementary symmetric polynomials on the $\{\kappa_\mu(p)\}_{\mu=1}^n$. In particular the mean curvature of a hypersurface is

$$H(p) = K_1(p) = \operatorname{tr} \widehat{\boldsymbol{S}} = \sum_{\mu=1}^{n} \kappa_\mu(p) = \sum_{\mu=1}^{n} \mathrm{II}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu), \tag{2.10}$$

the scalar curvature is $\mathcal{R}(p) = 2K_2(p)$, and the *Gaußian curvature* is

$$K_n(p) = \det \widehat{\boldsymbol{S}} = \prod_{\mu=1}^{n} \kappa_\mu(p) \tag{2.11}$$

Remark 2.1.4. To simplify notation we shall write $\kappa_\mu, H, \mathcal{R}$ instead of $\kappa_\mu(p), H(p), \mathcal{R}(p)$, etc. if the point is understood from the context. The point itself may be denoted p if interpreted set-theoretically in $\mathcal{S}$, or $\boldsymbol{p}$ if considered as a vector when it appears in linear operations of $\mathbb{R}^{n+1}$.

Notice the most elementary Newton relation between the power sum function of order 2 and the elementary symmetric polynomials yields the useful expression:

$$\operatorname{tr} \widehat{\boldsymbol{S}}^2 = \sum_{\mu=1}^{n} \kappa_\mu^2 = K_1^2 - 2K_2 = H^2 - \mathcal{R}. \tag{2.12}$$

In fact, more is true since the Gauß equation applied to the Ricci tensor of a hypersurface leads to

$$\boldsymbol{\mathcal{R}ic}(\boldsymbol{x}, \boldsymbol{y}) = H\langle\, \widehat{\boldsymbol{S}}(\boldsymbol{x}),\, \boldsymbol{y}\,\rangle - \langle\, \widehat{\boldsymbol{S}}^2(\boldsymbol{x}),\, \boldsymbol{y}\,\rangle. \tag{2.13}$$

Using the lemmas introduced above and Gauß equation in codimension 1, we get the following crucial lemma for the approximations made in chapter 4.

Lemma 2.1.5. *There is an open neighborhood U_p around any point $p \in \mathcal{S}$ such that the smooth hypersurface $\mathcal{S}$ is locally given by a graph $z : U_p \subset T_p\mathcal{S} \cong \mathbb{R}^n \to T_p\mathcal{S} \oplus \langle \boldsymbol{n}_p \rangle \cong \mathbb{R}^{n+1}$, with $\boldsymbol{p} = \boldsymbol{0}$, and $\nabla z(\boldsymbol{0}) = \boldsymbol{0}$. Thus, it is defined to leading order by an osculating quadric which in the basis of principal directions becomes:*

$$z(\boldsymbol{x}) = \frac{1}{2}\sum_{\mu=1}^{n} \kappa_\mu x_\mu^2 + \mathcal{O}(x^3). \tag{2.14}$$

In this neighborhood the area element is

$$\mathrm{dVol}|_{U_p} = \sqrt{\det g}\ d^n\boldsymbol{x} = \sqrt{1 + \sum_{\mu=1}^{n}\left(\frac{\partial z}{\partial x_\mu}\right)^2}\, dx_1 \cdots dx_n. \tag{2.15}$$

The second fundamental form at p corresponds to the Hessian matrix of $z(\boldsymbol{x})$ at p:

$$\mathrm{II}_p(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \left[\frac{\partial^2 z}{\partial x_\mu \partial x_\nu}\right]_p \tag{2.16}$$

where $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ are the principal basis vectors. In this basis the Riemann tensor reduces to

$$\langle \boldsymbol{R}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\boldsymbol{e}_\alpha,\ \boldsymbol{e}_\beta\rangle = \kappa_\mu(p)\kappa_\nu(p)(\delta_{\alpha\nu}\delta_{\mu\beta} - \delta_{\alpha\mu}\delta_{\beta\nu}), \tag{2.17}$$

the diagonal components of the Ricci tensor are

$$\boldsymbol{\mathcal{R}ic}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) = R_{\mu\mu}(p) = \sum_{\alpha\neq\mu}^{n} \kappa_\alpha(p)\kappa_\mu(p), \tag{2.18}$$

and the scalar curvature is

$$\mathcal{R}(p) = 2K_2(p) = 2\sum_{\mu<\nu}^{n} \kappa_\mu(p)\kappa_\nu(p). \tag{2.19}$$

2.2 Integral Invariants and Descriptors

In our context, integral invariants are local integrals over domains of a submanifold determined by intersection with objects in the ambient space, like spheres. Two such integrals 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 integral 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 integral 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. The study of covariance matrices in order to obtain adapted frames of general submanifols was introduced in [12] and [55, 56], whereas the integral invariant approach was developed in detail to extract the curvature information of surfaces in space, e.g. [20], [51].

In order to do this type of Principal Component Analysis on a general Riemannian submanifold and generalize local integral invariants, definitions using Cartesian coordinates must be naturally promoted to Riemann normal coordinates [16], [46]. If the n-dimensional submanifold $\mathcal{M}^{(n)}$ sits inside an ambient Riemannian manifold $(\mathcal{N}^{(n+k)}, g)$, the curves in $\mathcal{N}$ that generalize the axes used in $\mathbb{R}^{n+k}$ are the geodesic curves $\gamma_{\boldsymbol{v}}(t)$, and these always exist uniquely, locally at any point $p \in \mathcal{N}$ and direction $\boldsymbol{v}$. Given an orthonormal frame of $T_p\mathcal{M} \oplus N_p\mathcal{M}$, the geodesics tangent to each of the basis vectors will trace out generalized coordinate axes in $\mathcal{N}$ that, through the exponential map, will uniquely specify any point in a local neighborhood around p. Assuming $\mathcal{N}$ is geodesically complete to simplify the exposition, the exponential map collects all geodesics starting at p by mapping straight lines through the origin in $T_p\mathcal{N} \cong \mathbb{R}^{n+k}$ to geodesics through p:

$$\exp_p : T_p\mathcal{M} \to \mathcal{N} \quad \text{given by} \quad \exp_p(t\boldsymbol{v}) = \gamma_{t\boldsymbol{v}}(1) = \gamma_{\boldsymbol{v}}(t).$$

At any point p there is a neighborhood $\widetilde{\mathcal{U}}$ of $\mathbf{0}$ in $T_p\mathcal{N}$ where $\exp$ is a diffeomorphism onto a neighborhood $\mathcal{U}$ of p in $\mathcal{N}$. From this, for star-shaped $\widetilde{\mathcal{U}}$, there is also a unique geodesic $\gamma(t)$ connecting p and any other point $q \in \mathcal{U}$ such that the tangent $\gamma'(0) = \exp_p^{-1}(q)$. Moreover, the arclength of γ between the two points, i.e. the distance $d(p,q)$ between them determined by the metric g, is the length of the tangent vector representation through this map, $d(p,q) = \| \exp_p^{-1}(q) \|$, cf. [46, Ch. 5, Lem. 13]. These normal neighborhoods allow the parametrization of points using the geodesic distances tangent to a given frame $\{\boldsymbol{e}_\mu\}_{\mu=1}^{n+k}$ at p. The injectivity radius r_p is the radius of the largest ball $B_{\mathbf{0}}(\varepsilon)$ in $T_p\mathcal{N}$ where $\exp$ is a diffeomorphism, so $\mathcal{B}_p(r_p) = \exp_p(B_{\mathbf{0}}(r_p))$ is the largest ball in $\mathcal{N}$ created by radial geodesics of the same length around p where normal coordinates are well-defined. In fact, $r_p > 0$ always. Since our main theorems 5.2.4 and 5.3.5 are asymptotic results in the scale limit, in a general Riemannian manifold one could always use normal coordinates to study domains of submanifolds small enough so that they can be mapped to Euclidean space, thus, we propose the following general definition of PCA integral invariants.

Definition 2.2.1. Let D be a measurable domain in a Riemannian manifold $(\mathcal{N}, g)$ such that $D \subset \mathcal{B}_p(r_p)$ for some point $p \in \mathcal{N}$, The *integral invariants* associated to the moments of order 0, 1 and 2 of the geodesic coordinate functions of the points of D with respect to p are:
the volume

$$V(D) = \int_D 1 \text{ dVol}, \tag{2.20}$$

the barycenter

$$\boldsymbol{s}(D) = \frac{1}{V(D)} \int_D [\exp_p^{-1}(q)] \text{ dVol}, \tag{2.21}$$

and the eigenvalue decomposition of the covariance matrix:

$$C(D) = \int_D [\exp_p^{-1}(q)] \otimes [\exp_p^{-1}(q)] \text{ dVol}. \tag{2.22}$$

Here dVol is the measure on D, restriction of the measure on $\mathcal{N}$ induced by the metric g, and the tensor product is to be understood as the outer product of the components of the $\exp^{-1}$ map in a chosen orthonormal basis of $T_p\mathcal{N}$. The reference point of the covariance matrix is often chosen to be the barycenter, $\exp_p(\boldsymbol{s})$, instead of p.

These can be interpreted as statistical characterization measurements of a continuous distribution: the volume measures the size or mass of the set; the barycenter measures the centralization of the domain as a mean or average point, i.e., a center of mass; finally, the covariance matrix is a measure of the dispersion of the points in D around its center of mass. From this statistical point of view, we could have defined the covariance matrix normalized by $V(D)$ as well, so that $\frac{\mathrm{dVol}}{V}$ is a density, but this will not affect our results in any significant way (essentially, the second-to-leading order term in the volume equations would get added to the eigenvalues at that order).

The two types of domains that we shall study are regions in a submanifold $\mathcal{M} \subset \mathcal{N}$ determined by the intersection with a ball and a cylinder, cf. Figure 2.1. Using the exponential map one can define such intersections by mapping Euclidean balls and higher-dimensional cylinders in $T_p\mathcal{N}$ to their geodesic generalizations in the ambient manifold $\mathcal{N}$.

Definition 2.2.2. The *spherical component* of radius $\varepsilon \leqslant r_p$, at a point p of a submanifold $\mathcal{M}$ of a Riemannian manifold $\mathcal{N}$, is the domain given by:

$$D_p(\varepsilon) := \mathcal{M} \cap \{q \in \mathcal{N} : \|\exp_p^{-1}(q)\| \leqslant \varepsilon \leqslant r_p\}. \tag{2.23}$$

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 a $\mathbb{V}$-cylinder by taking all points in the ambient space that project down onto the ball inside $\mathbb{V}$.

Definition 2.2.3. The *cylindrical component* of radius $\varepsilon \leqslant r_p$, at a point p of a submanifold $\mathcal{M}$ of a Riemannian manifold $\mathcal{N}$ over the m-plane $\mathbb{V} \in \mathrm{Gr}(m, n+k)$, is the $\mathbb{V}$-cylinder intersection:

$$\mathrm{Cyl}_p(\varepsilon, \mathbb{V}) := \mathcal{M} \cap \{q \in \mathcal{N} : \|\mathrm{proj}_{\mathbb{V}}(\exp_p^{-1}(q))\| \leqslant \varepsilon \leqslant r_p\}, \tag{2.24}$$

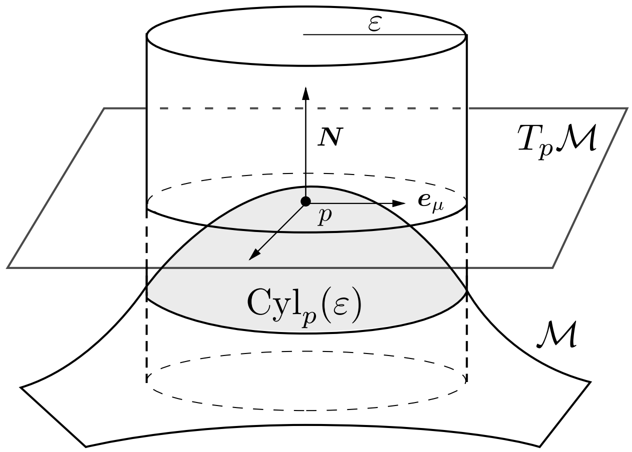

Figure 2.1: A normal cylinder to the surface at a point cuts out a patch domain (greyed region) whose covariance matrix EVD shall encode generalized principal curvatures and directions at p.

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 following chapters, we shall compute the integral invariants defined above for these domains on embedded submanifolds of Euclidean space, $\mathcal{N} = \mathbb{R}^{n+k}$, where $\exp_p^{-1}(q) = \boldsymbol{q} - \boldsymbol{p}$ as vectors and the tensor product recovers the common definition of PCA integral invariants studied in the literature. The points $q \in D$ are then parametrized by a vector $\boldsymbol{X}$ such that the barycenter is

$$\boldsymbol{s}(D) = \frac{1}{V(D)} \int_D \boldsymbol{X} \ \mathrm{dVol}, \tag{2.25}$$

and the the covariance matrix can be interpreted as analogous to a moment of inertia matrix, which for the cylindrical component shall be taken with respect to the center p, following the convention and motivation of [56],

$$C(\mathrm{Cyl}_p(\varepsilon)) = \int_{\mathrm{Cyl}_p(\varepsilon)} (\boldsymbol{X} - \boldsymbol{p}) \otimes (\boldsymbol{X} - \boldsymbol{p}) \ \mathrm{dVol}, \tag{2.26}$$

whereas for the spherical component the covariance matrix shall be taken with respect to the barycenter, following [51],

$$C(D_p(\varepsilon)) = \int_{D_p(\varepsilon)} (\boldsymbol{X} - \boldsymbol{s}(D_p(\varepsilon))) \otimes (\boldsymbol{X} - \boldsymbol{s}(D_p(\varepsilon))) \, \mathrm{dVol}. \tag{2.27}$$

An *integral invariant descriptor* $F(D)$ of some feature F of a measurable domain D is any expression for F completely given in terms of $V(D)$, $\boldsymbol{s}(D)$, the eigenvalue decomposition of $C(D)$ or other integral invariants. If the domain D is determined by a region of a hypersurface $\mathcal{S}$, the main geometric descriptors are any principal curvature estimators $\kappa_\mu(D)$ of $\kappa_\mu(p)$, and principal and normal direction estimators $\boldsymbol{e}_\mu(D)$, $\boldsymbol{n}(D)$ of $\boldsymbol{e}_\mu(p), \boldsymbol{n}(p)$, for some known point $p \in \mathcal{S}$. If the domain D is determined by a region of an embedded manifold $\mathcal{M}$, the main geometric descriptor is any second fundamental form estimator, $\mathbf{II}(D)$ of $\mathbf{II}_p$, for some known point $p \in \mathcal{M}$. Since our domain D of interest 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*. Moreover, we consider ε to be small enough so that we can approximate the submanifold by the local graph representation of its osculating paraboloids at p, which is sufficient to obtain the first terms of the asymptotic expansions of the integral invariants with respect to scale.

When the asymptotic expansions with respect to scale of hypersurface integral invariants are available to high enough order, curvature information can be extracted by truncating the series and inverting the relations in order to obtain a computable multi-scale estimator of the actual curvatures. In particular, the eigenvalues of the covariance matrix will provide such a descriptor for the principal curvatures of a smooth hypersurface, $\kappa_\mu(D)$, and its eigenvectors $\{\boldsymbol{e}_\mu(D)\}_{\mu=1}^n$, and $\boldsymbol{e}_{n+1}(D)$, will do the same for the principal and normal directions. In order to produce analogous descriptors for an embedded Riemannian manifold of higher codimension, we just need to apply the procedure to the k hypersurfaces created by projecting the manifold down to $(n+1)$-linear subspaces.

Lemma 2.2.4. *Let $\mathcal{M} \subset \mathbb{R}^{n+k}$ be an n-dimensional embedded Riemannian manifold, and let an orthonormal tangent basis $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$ of the tangent space $T_p\mathcal{M}$, and an orthonormal basis $\{\boldsymbol{n}_j\}_{j=1}^k$ of the normal space $N_p\mathcal{M}$ be fixed at $p \in \mathcal{M}$. Consider a ball $B_p^{(n+k)}(\varepsilon)$ for small enough $\varepsilon > 0$, such that the projections of $\mathcal{M} \cap B_p^{(n+k)}(\varepsilon)$ onto the linear subspaces $T_p\mathcal{M} \oplus \langle \boldsymbol{n}_j \rangle$, for all*

$j = 1, \ldots, k$, are smooth hypersurfaces $\mathcal{S}_j$. Then, if $\kappa_\mu^{(j)}(D)$, $\{\boldsymbol{e}_\mu^{(j)}(D)\}_{\mu=1}^n$ are descriptors of the principal curvatures and principal directions at p for each of the hypersurfaces $\mathcal{S}_j$, then the second fundamental form of $\mathcal{M}$ at p has a descriptor:

$$\mathbf{II}_p(D)(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^{k} [V_j(D) K_j(D) V(D)_j^T]_{\mu\nu} \, \boldsymbol{n}_j \,, \qquad \mu, \nu = 1, \ldots, n, \tag{2.28}$$

where $[V_j(D)]$ are the matrices whose columns are the components of $\{\boldsymbol{e}_\mu^{(j)}(D)\}_{\mu=1}^n$ in the chosen basis $\{\boldsymbol{e}_\mu\}_{\mu=1}^n$, and $[K_j(D)]$ is the diagonal matrix of principal curvature estimators. In turn, the Riemann curvature tensor of $\mathcal{M}$ at p acquires a descriptor:

$$\langle \boldsymbol{R}(D)(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\boldsymbol{e}_\alpha, \, \boldsymbol{e}_\beta \rangle = \sum_{j=1}^{k} \left([V_j K_j V_j^T]_{\mu\beta} [V_j K_j V_j^T]_{\nu\alpha} - [V_j K_j V_j^T]_{\mu\alpha} [V_j K_j V_j^T]_{\nu\beta} \right). \tag{2.29}$$

Proof. From lemma 2.1.2, there is a neighborhood of $U_p \subset T_p\mathcal{M}$ such that the manifold can be locally given by a graph $\boldsymbol{x} \mapsto (\boldsymbol{x}, f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))$, where $\boldsymbol{x} \in U_p$, p corresponds to 0, and $\nabla f_j(\mathbf{0}) = \mathbf{0}$. From this, the projection hypersurfaces $\mathcal{S}_j$ are just $(\boldsymbol{x}, f_j(\boldsymbol{x}))$ for $j = 1, \ldots, k$. The second fundamental form of $\mathcal{M}$ at p is precisely the linear combination of the second fundamental forms of each of the hypersurface projections weighed by the corresponding normal vector, i.e.,

$$\mathbf{II}_p(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \sum_{j=1}^{k} \left[\frac{\partial^2 f_j}{\partial x_\mu \partial x_\nu}(p) \right] \boldsymbol{n}_j$$

Analyzing each of those hypersurfaces in $T_p\mathcal{M} \oplus \langle \boldsymbol{n}_j \rangle \cong \mathbb{R}^{n+1}$, to obtain descriptors $\kappa_\mu^{(j)}(D)$, $\{\boldsymbol{e}_\mu^{(j)}(D)\}_{\mu=1}^n$ for every j, we obtain precisely a descriptor of the eigenvalue decomposition of each Hessian, i.e., Hess $f_j|_p(D) = [V_j(D) K_j(D) V(D)_j^T]$ is an estimator of the second fundamental form of $\mathcal{S}_j$ at p in the original basis. Applying Gauß equation 2.1.3 yields a corresponding descriptor for the Riemann tensor. □

These descriptors become valuable tools to perform Manifold Learning, feature detection and shape estimation when only partial knowledge of the complete set of points is known or when noise is present. In this regard, [50, 51, 60] carried out experimental and theoretical analysis of the stability of these and other descriptors in the case of curves and especially surfaces in $\mathbb{R}^3$, reporting for example that the invariants of the spherical component domain are more robust with respect to noise than the patch region ones. It is to be expected that the same stability behavior holds in the hypersurface case studied in chapter 4, due to the sensitivity to small changes of an n-dimensional patch compared to an $(n + 1)$-dimensional volume which has the perturbed patch as part of its boundary.

Chapter 3

Regular Curves and Hankel Determinants

The covariance analysis of regular curves in a Euclidean space of arbitrary dimension was already studied in [55,56], where the eigenvectors were shown to converge with scale to the Frenet-Serret frame, and the eigenvalues series expansion was found to be proportional to leading order to products of the Frenet curvatures. A formula for the coefficients was not explicitly known however since they depend on the value of certain Hankel determinants. These results provide an asymptotic relationship between the squares of the Frenet-Serret curvatures and ratios of successive eigenvalues of the covariance matrix. Since curves are locally determined by the Frenet curvature functions, up to rigid motion, the covariance integral invariants fully characterize the curve in the limit, providing descriptors at scale for these curvatures. In order to find the explicit value of the ratio coefficient, we obtain the recursion relation of a certain family of Hankel determinants by using the theory of orthogonal polynomials and its relation to the moment problem.

3.1 Frenet-Serret Apparatus from Covariance Matrices

Smooth curves in Euclidean space $\mathbb{R}^n$, for any dimension $n > 1$, have essentially the same structure [26], [33], [35], [57], [58] as the classical cases of plane curves and three-dimensional space curves: generically, they possess a comoving orthonormal frame constructed from the velocity and acceleration vectors, and their orthonormal completion, along with generalized curvature functions at every point. The latter were originally defined for plane curves in terms of the inverse radius of the osculating circle.

A regular parametrized curve $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ is a continuously differentiable immersion, so that $\gamma'(t) = \dfrac{d\gamma}{dt} \neq 0$ for all $t \in I$. Two such curves are considered equivalent if they are related by a bijective, continuously differentiable reparametrization of I that preserve orientation. The length of the curve from point $t = a$ to $t = b$ is given by the ambient space metric integration of its tangent vector:

$$s(a,b) = \int_a^b \sqrt{\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \rangle}\, dt = \int_a^b \left\| \frac{d\gamma}{dt} \right\| dt \tag{3.1}$$

Every regular curve can be parametrized by its arc length $s(a,t)$ so that its velocity is a unit vector, $\|\gamma'(s)\| = 1$. From now on, we shall consider only regular curves n-times continuously differentiable and parametrized by their arc length. In order to use a system of reference adapted to the curve, one introduces the Frenet-Serret frame. When $\gamma(s)$ is a regular curve in $\mathbb{R}^n$, we say it is a Frenet curve when all the derivatives $\gamma', \gamma'', ..., \gamma^{(n)}$ form a set of n linearly independent vectors in $\mathbb{R}^n$. The *Frenet-Serret frame* $\{\boldsymbol{e}_j\}_{j=1}^n$ is the positively oriented comoving orthonormal basis of $\mathbb{R}^n$ obtained from the Gram-Schmidt orthogonalization procedure applied to $\gamma', \gamma'', ..., \gamma^{(n)}$:

$$e_j(s) = \frac{\widetilde{e}_j(s)}{\|\widetilde{e}_j(s)\|} \text{ with } \widetilde{e}_j(s) = \gamma^{(j)}(s) - \sum_{i=1}^{j-1} \langle \gamma^{(j)}(s), e_i(s) \rangle\, e_i(t) \text{ for } 1 \leqslant j \leqslant n.$$

In $\mathbb{R}^3$, $\boldsymbol{e}_1$ is the tangent vector, $\boldsymbol{e}_2$ is the principal normal vector, and $\boldsymbol{e}_3 = \boldsymbol{e}_1 \times \boldsymbol{e}_2$ is the binormal vector, which specify the tangent line and osculating plane at every point of the curve. From the classical curvature functions of plane and space curves, one arrives at a definition of *Frenet curvatures* at a point s:

$$\kappa_j(s) = \langle\, \boldsymbol{e}_j'(s),\, \boldsymbol{e}_{j+1}(s) \,\rangle \text{ for } 1 \leqslant j \leqslant n-1. \tag{3.2}$$

They satisfy the Frenet equations [33, Th. 2.13] in any dimension.

Theorem 3.1.1. *Let γ be a Frenet curve in $\mathbb{R}^n$ with Frenet-Serret frame $\{\boldsymbol{e}_j\}_{j=1}^n$. Then the curvatures $\{\kappa_j\}_{j=1}^{n-1}$ satisfy $\kappa_1, \ldots, \kappa_{n-2} > 0$, and every κ_j is $(n-1-j)$-continuously differentiable such that*

$$\frac{d}{ds}\begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix} = \begin{pmatrix} 0 & \kappa_1(s) & 0 & 0 & 0 \\ -\kappa_1(s) & 0 & \kappa_2(s) & 0 & 0 \\ 0 & -\kappa_2(s) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \kappa_{n-1}(s) \\ 0 & \dots & 0 & -\kappa_{n-1}(s) & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix}. \tag{3.3}$$

And the Frenet-Serret frame and curvatures are invariant under Euclidean motions.

From this result the local characterization of regular curves is obtained. This establishes the existence and uniqueness of a Frenet curve when a point, initial Frenet-Serret frame and generalized curvature functions are given, cf. [33, Th. 2.15].

Theorem 3.1.2. *Let $\kappa_1, \dots, \kappa_{n-1} : (a, b) \to \mathbb{R}$ be given functions such that each κ_j is a $(n-1-j)$-continuously differentiable function with $\kappa_1, \dots, \kappa_{n-2} > 0$. Let $s_0 \in (a, b)$, and let a point $p_0 \in \mathbb{R}^n$ and a frame $\{\boldsymbol{e}_j^{(0)}\}_{j=1}^n$ of $\mathbb{R}^n$ be fixed. Then there is a unique n times continuously differentiable Frenet curve $\gamma : (a, b) \to \mathbb{R}^n$, parametrized by arc length satisfying $\gamma(s_0) = p_0$, $\{\boldsymbol{e}_j^{(0)}\}_{j=1}^n$ is the Frenet-Serret frame of γ at the point p_0, and where $\{\kappa_j(s)\}_{j=1}^n$ are the Frenet curvature functions of γ.*

The integral invariant approach to Frenet curves thus aims to find a relationship between the Frenet-Serret apparatus and the eigenvalue dcomposition of the covariance matrix. The spherical and cylindrical covariance matrix applied to a regular curve γ in $\mathbb{R}^n$, at point s and scale ε, reduces to leading order to

$$C(s, \varepsilon) = \frac{1}{2\epsilon}\int_{s-\epsilon}^{s+\epsilon} (\gamma(t) - \gamma(s)) \cdot (\gamma(t) - \gamma(s))^T dt. \tag{3.4}$$

This expression must be understood as the outer product of the position vectors of the curve with respect to the center point $\gamma(s)$. Notice that here we are normalizing by the first order approximation of the arc length of the intersection domain in order to use the results of [56]. Indeed, it was proven [55], [4] that the eigenvectors of the covariance matrix converge to the Frenet-Serret directions.

Theorem 3.1.3. *Let $\gamma : I \to \mathbb{R}^n$ be a parametric curve of class C^{n+1}, regular of order n. Let $\boldsymbol{e}_1(s), \ldots, \boldsymbol{e}_n(s)$ denote the Frenet-Serret frame at $\gamma(s)$. Let $\boldsymbol{V}_1(s), \ldots, \boldsymbol{V}_n(s)$ denote the limit eigenvector of $C(s,\varepsilon)$ at $\gamma(s)$ for $\varepsilon \to 0$. Then the covariance eigenvectors converge to the Frenet-Serret frame, i.e., $\boldsymbol{e}_j(s) = \pm \boldsymbol{V}_j(s)$, for $j = 1, \ldots, n$.*

The following key result by F.J. Solis [55,56] expresses the series expansion of the eigenvalues to leading order in terms of the Frenet curvatures:

Lemma 3.1.4. *Let $\gamma(s)$ be a regular curve in $\mathbb{R}^n$, and let p_0 be a point on the curve, then the eigenvalues associated with $C(p_0, \epsilon)$ are given by*

$$\lambda_1(\epsilon) = P_1 \epsilon^2 + O(\epsilon^4), \tag{3.5}$$

$$\lambda_j(\epsilon) = \frac{(\kappa_1 \cdots \kappa_{j-1})^2}{(j!)^2} P_j \epsilon^{2j} + O(\epsilon^{2j+2}), \qquad j = 2, \ldots, n \tag{3.6}$$

and the eigenvectors are given by the Frenet-Serret frame at p_0. The κ_i's are the Frenet curvatures of the γ and P_k is the k-th $(k = 1, \ldots, n)$ pivot of the $n \times n$ matrix A_n defined by

$$A_{ij} = \begin{cases} \frac{1}{i+j+1}, & \textit{if } i+j \textit{ is even}; \\ 0 & \textit{otherwise}. \end{cases} \tag{3.7}$$

From the proof of this lemma, a typo is corrected for the denominator of $\lambda_j(\epsilon)$ in the final statement. With this result we can express the curvatures κ_j in terms of the eigenvalues by writing the pivots as quotients of the determinants B_j of A_j, that is $P_j = B_j/B_{j-1}$, so that:

$$\lim_{\epsilon \to 0} \frac{\lambda_{j+1}(\varepsilon)}{\lambda_1(\varepsilon)\lambda_j(\varepsilon)} = \kappa_j^2 \frac{B_{j+1}B_{j-1}}{(j+1)^2 B_1 B_j^2}. \tag{3.8}$$

In fact, a different route towards the covariance analysis of curves was developed in [4]: by solving the characterization theorem 3.1.2, given curvature constants at a point, the system of ODEs can be solved to obtain canonical helix-type curves with those curvatures that approximate any smooth curve at that point as an osculating helix:

$$\gamma_e(s) = \begin{bmatrix} a_1 \cos(\alpha_1 s) \\ a_1 \sin(\alpha_1 s) \\ \vdots \\ a_k \cos(\alpha_k s) \\ a_k \sin(\alpha_k s)) \end{bmatrix} \qquad \text{or} \qquad \gamma_o(s) = \begin{bmatrix} a_1 \cos(\alpha_1 s) \\ a_1 \sin(\alpha_1 s) \\ \vdots \\ a_k \cos(\alpha_k s) \\ a_k \sin(\alpha_k s)) \\ bs \end{bmatrix} \tag{3.9}$$

where the first equation is for the case when n is even, such that $k = n/2$, and the second equation is for the case when n is odd, such that $k = (n-1)/2$. By directly computing the covariance matrix for this type of curves, one can relate explicitly the eigenvalues to the parameters a_μ, α_μ, b of those solutions and establish a relationship with the Frenet curvatures that leads to a conjectured formula, whose proof constitutes the present author contribution: determining the coefficient of equation 3.8 by using the theory of Hankel determinants.

Theorem 3.1.5. *Let $\gamma : I \to \mathbb{R}^n$ be a parametric curve of class C^{n+1}, regular of order n. Let $\kappa_j(s)$ denote the j^{th} curvature function of γ evaluated at s and let $\lambda_j(\varepsilon)$ denote the j^{th} local eigenvalue of the covariance matrix $C(s, \varepsilon)$. For each $s \in I$ and each $j = 1, ..., n-1$,*

$$\kappa_j^2(s) = a_j \lim_{\varepsilon \to 0} \frac{\lambda_{j+1}(\varepsilon)}{\lambda_1(\varepsilon)\lambda_j(\varepsilon)}, \quad a_{j-1} = \left(\frac{j}{j + (-1)^j} \right)^2 \frac{4j^2 - 1}{3}. \tag{3.10}$$

For fixed $\varepsilon > 0$, these eigenvalue ratios furnish descriptors at scale of the generalized curvatures, with their respective eigenvectors becoming descriptors of the Frenet-Serret frame. This permits the local characterization of the curve within the given approximation. In particular, if we are given a big sample of points belonging to a regular curve in some Euclidean space, the covariance integral can be computed for small balls around every point to obtain a set of curvature descriptor functions. This curvature profile can be used as a classifier of point sets belonging to different curves when explicit parametrization functions of the curves that generate those point are unknown.

3.2 Hankel Matrices and Orthogonal Polynomials

The determinants B_j are of Hankel type for the sequence $\{\mu_n\}_{n=0}^{\infty} = \{\frac{1}{3}, 0, \frac{1}{5}, 0, \frac{1}{7}, ...\}$, i.e.,

$$B_1 = \frac{1}{3},\ B_2 = \begin{vmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{5} \end{vmatrix},\ B_3 = \begin{vmatrix} \frac{1}{3} & 0 & \frac{1}{5} \\ 0 & \frac{1}{5} & 0 \\ \frac{1}{5} & 0 & \frac{1}{7} \end{vmatrix},\ B_j = \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_{j-1} \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_j \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{j+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mu_{j-1} & \mu_j & \mu_{j+1} & \cdots & \mu_{2j-2} \end{vmatrix}.$$

Then to get our coefficient in theorem 3.1.5 amounts to showing that the aforementioned Hankel determinants satisfy the following recurrence relation:

$$\frac{B_j B_{j-2}}{(B_{j-1})^2} = \frac{(j + (-1)^j)^2}{4j^2 - 1}. \tag{3.11}$$

This is indeed the case after we realize that such a recurrence relation appears in the theory of monic orthogonal polynomials generated from $\{x^n\}_{n=0}^{\infty}$ by Gram-Schmidt orthogonalization with respect to a measure generating our sequence μ_n as the integral moments. Indeed, choose a nondecreasing function $\lambda(x)$ on $\mathbb{R}$ having finite limits at $\pm\infty$ such that it induces a positive measure $d\lambda$ with finite moments to all orders

$$\mu_n(d\lambda) = \int_{\mathbb{R}} x^n d\lambda(x), \quad n = 0, 1, 2, ...$$

then apply the Gram-Schmidt orthogonalization procedure to $\{x^n\}_{n=0}^{\infty}$ using the scalar product

$$\langle p(x), q(x) \rangle = \int_{\mathbb{R}} p(x)q(x)d\lambda(x)$$

to obtain a sequence of monic orthogonal polynomials $P_n(x)$ (without normalization). If the given scalar product is positive-definite, such a sequence is infinite and unique, and this is the case if $B_n > 0$ for all $n \in \mathbb{N}$, see Gautschi [24, Th. 1.2, 1.6]. Moreover, in this case, the infinite sequence of monic orthogonal polynomials obtained in this manner obeys the recursion relation [24, Th. 1.27]:

$$P_{-1}(x) = 0,\ P_0(x) = 1, \qquad P_{n+1}(x) = (x - \alpha_n)P_n(x) - \beta_n P_{n-1}(x) \tag{3.12}$$

where

$$\alpha_n = \frac{\langle P_n, xP_n \rangle}{\langle P_n, P_n \rangle}, \quad \beta_n = \frac{\langle P_n, P_n \rangle}{\langle P_{n-1}, P_{n-1} \rangle} = \frac{||P_n(x)||^2}{||P_{n-1}(x)||^2}, \text{ for } n = 1, 2, \ldots$$

The importance of this result is that the recursion coefficients β_n are precisely the recursion coefficients of the Hankel determinants B_n for the sequence μ_n, as it is proved in [24, eq. 2.1.5]

$$\beta_{j-1} = \frac{B_j B_{j-2}}{(B_{j-1})^2}, \text{ for } n = 2, 3, \ldots \tag{3.13}$$

so finding a measure to reproduce our sequence as its moments and a way to compute the norms of the corresponding polynomials will yield our coefficient formula. There is a fundamental determinantal representation of the monic orthogonal polynomials generated in the previous way [24, Th. 2.1]

$$P_n(x) = \frac{1}{B_n} \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ 1 & x & \vdots & x^n \end{vmatrix}, \quad ||P_n(x)||^2 = \frac{B_{n+1}}{B_n},$$

that yields Heine's integral representation formula [28, p. 288] by essentially pulling the integrals of each moment out of the determinant and expanding i:

$$P_n(x) = \frac{1}{n!B_n} \int \cdots \int_{\mathbb{R}^n} \prod_{i=1}^{n} (x - x_i) \prod_{1 \leqslant l < k \leqslant n} (x_k - x_l)^2 d\lambda(x_1) \cdots d\lambda(x_n).$$

Since the polynomials are monic, B_n can be solved equating to 1 the leading coefficient of the previous equation

$$B_n = \frac{1}{n!} \int_{\mathbb{R}^n} \prod_{1 \leqslant l < k \leqslant n} (x_k - x_l)^2 d\lambda(x_1) \cdots d\lambda(x_n), \tag{3.14}$$

which is a closed formula for all Hankel determinants of any sequence as long as this can be written as moments of a positive measure.

Using the theory above for Hankel determinants of a particular type we arrive at the following key result.

Theorem 3.2.1. *For any inverse arithmetic sequence* $\left\{ \frac{1}{\alpha n + \beta} \right\}_{n=0}^{\infty}$*, where* $\alpha, \beta \in \mathbb{R}_{>0}$*, the corresponding Hankel determinants*

$$F_n(\alpha, \beta) = \begin{vmatrix} \frac{1}{\beta} & \frac{1}{\alpha+\beta} & \frac{1}{2\alpha+\beta} & \cdots & \frac{1}{(n-1)\alpha+\beta} \\ \frac{1}{\alpha+\beta} & \frac{1}{2\alpha+\beta} & \frac{1}{3\alpha+\beta} & \cdots & \frac{1}{n\alpha+\beta} \\ \frac{1}{2\alpha+\beta} & \frac{1}{3\alpha+\beta} & \frac{1}{4\alpha+\beta} & \cdots & \frac{1}{(n+1)\alpha+\beta} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{(n-1)\alpha+\beta} & \frac{1}{n\alpha+\beta} & \frac{1}{(n+1)\alpha+\beta} & \cdots & \frac{1}{(2n-2)\alpha+\beta} \end{vmatrix} \tag{3.15}$$

are given by

$$F_n(\alpha, \beta) = \frac{1}{\alpha^n} \prod_{k=0}^{n-1} \frac{\Gamma(\beta/\alpha + k)(k!)^2}{\Gamma(\beta/\alpha + n + k)} = \frac{1}{\alpha^n} \prod_{k=0}^{n-1} (k!)^2 \prod_{j=0}^{n-1} \frac{\alpha}{\alpha(k+j) + \beta}, \tag{3.16}$$

and satisfy the recursion relation

$$\frac{F_n F_{n-2}}{F_{n-1}^2} = \frac{\alpha^2 \left(\alpha(n-2) + \beta\right)^2 (n-1)^2}{\left(\alpha(2n-2) + \beta\right) \left(\alpha(2n-3) + \beta\right)^2 \left(\alpha(2n-4) + \beta\right)}, \tag{3.17}$$

starting with $F_1 = \frac{1}{\beta},\ F_2 = \frac{\alpha^2}{\beta(2\alpha + \beta)(\alpha + \beta)^2}$.

Proof. Choose the function $\lambda(x) = x^{\beta/\alpha}/\beta$ which is always nondecreasing in the interval $[0,1]$ for $\beta/\alpha > 0$, then the corresponding positive measure

$$d\lambda(x) = \chi_{[0,1]}\frac{x^{\beta/\alpha-1}}{\alpha}dx,$$

where χ_I is the characteristic function of a measurable set $I \subset \mathbb{R}$, yields moments

$$\mu_n = \int_{\mathbb{R}} x^n d\lambda(x) = \frac{1}{\alpha}\int_0^1 x^{n+\frac{\beta}{\alpha}-1}dx = \frac{1}{\alpha}\left[\frac{x^{n+\frac{\beta}{\alpha}}}{n+\frac{\beta}{\alpha}}\right]_0^1 = \frac{1}{\alpha n+\beta}.$$

Notice that this solves the Stieltjes moment problem uniquely for these sequences because our measure is infinitely supported on $[0,\infty)$, and its moments satisfy Carleman's condition [53, th. 1.10]. From this, the necessary condition $F_n > 0$ is guaranteed to hold for any dimension n [53, th. 1.2], so the induced inner product is positive definite and thus the sequence of monic orthogonal polynomials $P_n(x)$ is infinite and unique. Thus their recurrence relations (3.12) hold for any $n \in \mathbb{N}$, so we can compute the determinants $F_n(\alpha,\beta)$ of any dimension. This is done by computing equation (3.14)

$$F_n(\alpha,\beta) = \frac{1}{n!}\int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} \frac{x_i^{\frac{\beta}{\alpha}-1}}{\alpha} \prod_{1\leqslant l<k\leqslant n}(x_k - x_l)^2 dx_1 \cdots dx_n$$

by means of Selberg's integral formula [10, 8.1.1], an extension of Euler's Beta function which has applications in different fields within mathematics and physics:

$$\int_{[0,1]^n} \prod_{i=0}^{n} x_i^{a-1}(1-x_i)^{b-1} \prod_{1\leqslant l<k\leqslant j} |x_k - x_l|^{2g}\, d^n\boldsymbol{x} = \prod_{k=0}^{n-1} \frac{\Gamma(a+kg)\Gamma(b+kg)\Gamma(1+(k+1)g)}{\Gamma(a+b+(n+k-1)g)\Gamma(1+g)},$$

when $\Re e(a) > 0, \Re e(b) > 0$ and $\Re e(g) > -\min\{1/n, \Re e(a)/(n-1), \Re e(b)/(n-1)\}$. These conditions are satisfied for our case $a = \beta/\alpha > 0$, and $b = g = 1$. Therefore by substitution of these values

$$F_n(\alpha,\beta) = \frac{1}{n!\alpha^n}\prod_{k=0}^{n-1}\frac{\Gamma(\beta/\alpha+k)\Gamma(1+k)\Gamma(2+k)}{\Gamma(\beta/\alpha+n+k)\Gamma(2)} = \frac{1}{\alpha^n}\prod_{k=0}^{n-1}\frac{\Gamma(\beta/\alpha+k)(k!)^2}{\Gamma(\beta/\alpha+n+k)},$$

where the Gamma functions can be simplified by the factorial property $\Gamma(z+1) = z\Gamma(z)$ to get a closed formula:

$$F_n(\alpha,\beta) = \frac{1}{\alpha^n}\prod_{k=0}^{n-1}(k!)^2\prod_{j=0}^{n-1}\frac{\alpha}{\alpha(k+j)+\beta}.$$

Finally, the recursion equation 3.13 can be worked out by telescoping the products of Gamma functions:

$$\frac{F_nF_{n-2}}{F_{n-1}^2} = \frac{1}{\alpha^n}\prod_{k=0}^{n-1}\frac{\Gamma(\frac{\beta}{\alpha}+k)(k!)^2}{\Gamma(\frac{\beta}{\alpha}+n+k)}\cdot\alpha^{n-1}\prod_{k=0}^{n-2}\frac{\Gamma(\frac{\beta}{\alpha}+n-1+k)}{\Gamma(\frac{\beta}{\alpha}+k)(k!)^2}\cdot$$

$$\alpha^{n-1}\prod_{k=0}^{n-2}\frac{\Gamma(\frac{\beta}{\alpha}+n-1+k)}{\Gamma(\frac{\beta}{\alpha}+k)(k!)^2}\cdot\frac{1}{\alpha^{n-2}}\prod_{k=0}^{n-3}\frac{\Gamma(\frac{\beta}{\alpha}+k)(k!)^2}{\Gamma(\frac{\beta}{\alpha}+n-2+k)} =$$

$$= \frac{\Gamma(\frac{\beta}{\alpha}+n-1)(n-1)!^2}{\Gamma(\frac{\beta}{\alpha}+n-2)(n-2)!^2}\prod_{k=0}^{n-1}\frac{1}{(\frac{\beta}{\alpha}+n-1+k)\Gamma(\frac{\beta}{\alpha}+n-1+k)}\prod_{k=0}^{n-2}\Gamma\left(\frac{\beta}{\alpha}+n-1+k\right)\cdot$$

$$\prod_{k=0}^{n-2}\left(\frac{\beta}{\alpha}+n-2+k\right)\Gamma\left(\frac{\beta}{\alpha}+n-2+k\right)\prod_{k=0}^{n-3}\frac{1}{\Gamma\left(\frac{\beta}{\alpha}+n-2+k\right)} =$$

$$= \frac{(\frac{\beta}{\alpha}+n-2)(n-1)^2\Gamma(\frac{\beta}{\alpha}+2n-4)}{\Gamma(\frac{\beta}{\alpha}+2n-2)}\prod_{k=0}^{n-1}\frac{1}{\left(\frac{\beta}{\alpha}+n-1+k\right)}\prod_{k=0}^{n-2}\left(\frac{\beta}{\alpha}+n-2+k\right)$$

$$= \frac{(\frac{\beta}{\alpha}+n-2)^2(n-1)^2}{(\frac{\beta}{\alpha}+2n-2)(\frac{\beta}{\alpha}+2n-3)^2(\frac{\beta}{\alpha}+2n-4)}$$

which yields the stated formula upon multiplying numerator and denominator by α^4. □

Remarkably, this means that our polynomial recursion coefficients satisfy $\beta_n = \frac{1}{4}\beta_n^J$, where β_n^J are those of the classical monic Jacobi polynomials of type $(\frac{\beta}{\alpha}-1, 0)$. These are generated by the measure $\chi_{[-1,1]}(1-x)^{\frac{\beta}{\alpha}-1}dx$, which induces a completely different moment sequence and set of orthogonal polynomials.

Our actual determinants B_n have alternating 0's in the even positions of the moment sequence, so a block decomposition is needed to get them into the form of the theorem.

Corollary 3.2.2. *For any sequence of type* $\left\{ \frac{1}{\alpha n + \beta}, 0 \right\}_{n=0}^{\infty}$ *with* $\alpha, \beta \in \mathbb{R}_{>0}$*, where zeros alternate every other position, the corresponding Hankel determinants* B_n *are given by the following block decomposition for even* $n = 2m$ *or odd* $n = 2m-1$ *dimension,* $m \in \mathbb{N}$*:*

$$B_{2m} = F_m(\alpha, \beta) F_m(\alpha, \beta + \alpha), \quad B_{2m-1} = F_m(\alpha, \beta) F_{m-1}(\alpha, \beta + \alpha), \tag{3.18}$$

and obey the recurrence relations:

$$\frac{B_{2m} B_{2m-2}}{(B_{2m-1})^2} = \frac{(\alpha(m-1) + \beta)^2}{(\alpha(2m-1) + \beta)(\alpha(2m-2) + \beta)}, \tag{3.19}$$

$$\frac{B_{2m-1} B_{2m-3}}{(B_{2m-2})^2} = \frac{\alpha^2 (m-1)^2}{(\alpha(2m-2) + \beta)(\alpha(2m-3) + \beta)}, \tag{3.20}$$

starting with $B_1 = \frac{1}{\beta}$*,* $B_2 = \frac{1}{\beta(\alpha + \beta)}$.

Proof. The Hankel determinants with 0's at every even position of the first row can be decomposed into blocks by a procedure of moving rows and columns without altering the overall sign. Notice that the second block has as Hankel sequence the original one but shifted in index by $+1$, so the blocks are $F_m := F_m(\alpha, \beta)$ and $E_m := F_m(\alpha, \beta + \alpha)$. Analogously for $n = 2m - 1$, but in this case the number of 0's is now $m - 1$, so the size of the second block is $(m-1)^2$ whereas the first is still m^2. Thus

$$B_{2m} = F_m E_m, \quad B_{2m-1} = F_m E_{m-1}.$$

Whence the recursion coefficients for the induced polynomials are, for even n,

$$\beta_{n-1} = \beta_{2m-1} = \frac{B_{2m} B_{2(m-1)}}{B_{2m-1}^2} = \frac{E_m}{E_{m-1}} \frac{F_{m-1}}{F_m},$$

and for odd n:

$$\beta_{n-1} = \beta_{2m-2} = \frac{B_{2m-1}B_{2(m-1)-1}}{B_{2(m-1)}^2} = \frac{E_{m-2}}{E_{m-1}}\frac{F_m}{F_{m-1}}.$$

Therefore using equation 3.16, that the corresponding β/α for the E_m blocks is $\beta/\alpha + 1$ and the factorial property of the Gamma function, the products can be simplified in the same way as in our previous proof:

$$\frac{B_{2m}B_{2(m-1)}}{B_{2m-1}^2} = \frac{1}{\alpha^m}\prod_{k=0}^{m-1}\frac{\Gamma(\beta/\alpha+1+k)(k!)^2}{\Gamma(\beta/\alpha+1+m+k)} \cdot \alpha^{m-1}\prod_{k=0}^{m-2}\frac{\Gamma(\beta/\alpha+m+k)}{\Gamma(\beta/\alpha+1+k)(k!)^2}$$

$$\frac{1}{\alpha^{m-1}}\prod_{k=0}^{m-2}\frac{\Gamma(\beta/\alpha+k)(k!)^2}{\Gamma(\beta/\alpha+m-1+k)} \cdot \alpha^m\prod_{k=0}^{m-1}\frac{\Gamma(\beta/\alpha+m+k)}{\Gamma(\beta/\alpha+k)(k!)^2} =$$

$$= (\beta/\alpha+m-1)\prod_{k=0}^{m-1}\frac{1}{(\beta/\alpha+m+k)} \cdot \prod_{k=0}^{m-2}(\beta/\alpha+m+k-1) =$$

$$= \frac{(\beta/\alpha+m-1)^2}{(\beta/\alpha+2m-1)(\beta/\alpha+2m-2)}.$$

Similarly,

$$\frac{B_{2m-1}B_{2(m-1)-1}}{B_{2(m-1)}^2} = \frac{1}{\alpha^{m-2}}\prod_{k=0}^{m-3}\frac{\Gamma(\beta/\alpha+1+k)(k!)^2}{\Gamma(\beta/\alpha+m-1+k)} \cdot \alpha^{m-1}\prod_{k=0}^{m-2}\frac{\Gamma(\beta/\alpha+m+k)}{\Gamma(\beta/\alpha+1+k)(k!)^2}$$

$$\frac{1}{\alpha^m}\prod_{k=0}^{m-1}\frac{\Gamma(\beta/\alpha+k)(k!)^2}{\Gamma(\beta/\alpha+m+k)} \cdot \alpha^{m-1}\prod_{k=0}^{m-2}\frac{\Gamma(\beta/\alpha+m-1+k)}{\Gamma(\beta/\alpha+k)(k!)^2} =$$

$$= \frac{(m-1)!^2\Gamma(\beta/\alpha+m-1)\Gamma(\beta/\alpha+2m-3)}{(m-2)!^2\Gamma(\beta/\alpha+m-1)\Gamma(\beta/\alpha+2m-1)} =$$

$$= \frac{(m-1)^2}{(\beta/\alpha+2m-2)(\beta/\alpha+2m-3)}.$$

□

Finally the coefficient formula of theorem 3.1.5 is obtained from this using equation 3.8.

Corollary 3.2.3. *The Hankel determinants of size* $n \times n$

$$B_n = \det(A_n), \quad (A_n)_{ij} = \begin{cases} \frac{1}{i+j+1}, & \textit{if } i+j \textit{ is even}; \\ 0 & \textit{otherwise}. \end{cases}$$

satisfy the recurrence relation

$$\frac{B_n B_{n-2}}{(B_{n-1})^2} = \frac{(n+(-1)^n)^2}{4n^2-1}. \tag{3.21}$$

Proof. Notice the matrix entry at $(A_n)_{ij}$ is precisely the element of the sequence $\left\{\frac{1}{2n+3}, 0\right\}_{n=0}^{\infty}$ where $n = i+j-2$. Thus substituting $\alpha = 2$ and $\beta = 3$ into the equations 3.19 and 3.20 above, the result follows straightforwardly when simplifying the theorem formulas after indices are written in terms of the dimension, $m = n/2$ or $m = (n+1)/2$ for the even and odd cases respectively. □

The manifold reconstruction problem is thus solved for regular curves in Euclidean space in terms of covariance integral invariants.

Chapter 4

Covariance Analysis of Smooth Hypersurfaces

We generalize to hypersurfaces in any dimension major results known about the covariance analysis of surfaces in space [20], [51], whose descriptors shall yield a method to estimate the extrinsic and intrinsic curvature of an embedded Riemannian submanifold of general codimension. We obtain the asymptotic expansion of the PCA integral invariants for a spherical volume component delimited by a hypersurface and a ball in ambient space, and for the hypersurface patch created by ball and cylinder intersections. The domain volumes have asymptotic expansion with scale that correct the volume of a ball by the extrinsic curvature of the hypersurface at the center point. The EVD of the covariance matrix of the spherical volume component has eigenvalues with series expansion in terms of the principal curvatures and the mean curvature at the center, and eigenvectors that converge to the respective principal and normal directions. In the case of the patch invariants, the results are analogous but in terms of the squares and products of principal curvatures.

4.1 Spherical Component Integral Invariants

The following domain was introduced in [30] to study the relation between the mean curvature of hypersurfaces and the volume of ball sections (we reserve their notation $B_p^+(\varepsilon)$ for the half-ball).

Definition 4.1.1. Let $\mathcal{S}$ be a smooth hypersurface in $\mathbb{R}^{n+1}$ with a locally chosen normal vector field $\boldsymbol{n} : \mathcal{S} \to \mathbb{R}^{n+1}$. Let $B_p^{(n+1)}(\varepsilon)$ be a ball of radius $\varepsilon > 0$ centered at a point $p \in \mathcal{S}$, for small enough ε the hypersurface always separates this ball into two connected components. Consider the region $V_p^+(\varepsilon)$ to be that spherical component such that $\boldsymbol{n}(p)$ points towards inside the region $V_p^+(\varepsilon)$.

All the methods and results of [51] for surfaces using this domain generalize because to approximate integrals of functions over this type of region in $\mathbb{R}^3$, the formula developed in their work makes use of the hypersurface approximations of [30], valid in any dimension.

Lemma 4.1.2. *Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a function of order $\mathcal{O}(\rho^k z^l)$ in cylindrical coordinates $\boldsymbol{X} = (\boldsymbol{x}, z) = (\rho\overline{\boldsymbol{x}}, z)$, $\overline{\boldsymbol{x}} \in \mathbb{S}^{n-1}$, let $\mathcal{S}$ be a graph hypersurface given by the function $z(\boldsymbol{x})$ whose normal at the origin points in the positive z-axis, and $V_p^+(\varepsilon)$ the spherical component delimited by this $\mathcal{S}$, then*

$$\int_{V_p^+(\varepsilon)} f(\boldsymbol{X})\mathrm{dVol} = \int_{B_p^+(\varepsilon)} f(\boldsymbol{X})\mathrm{dVol} - \int_{B_p^n(\varepsilon)} \left[\int_{z=0}^{z=\frac{1}{2}\sum_{\mu=1}^n \kappa_\mu x_\mu^2} f(\boldsymbol{x}, z)\, dz \right] d^n\boldsymbol{x} + \mathcal{O}(\varepsilon^{k+2l+n+3}) \tag{4.1}$$

where the half ball $B_p^+(\varepsilon)$ consists of the points of $B_p^{n+1}(\varepsilon)$ such that $z \geqslant 0$.

Proof. We are going to approximate $z(\boldsymbol{x})$ by its osculating quadric at the origin, $\frac{1}{2}\sum_{\mu=1}^n \kappa_\mu x_\mu^2$, and remove from the complete half-ball integral of $f(\boldsymbol{X})$ its contribution from below the quadric approximation but, since what we know is the function over the tangent space at p, what can be computed is the contribution below the quadric over a domain in the tangent space that is explicitly integrable. The exact domain is determined by the sphere intersection with the hypersurface, $\{\|\boldsymbol{x}\|^2 + z(\boldsymbol{x})^2 \leqslant \varepsilon^2\}$, and what we can compute exactly is the integral over the cylinder $\{\rho \leqslant \varepsilon\}$, so that for every $\boldsymbol{x} \in B_p^{(n)}(\varepsilon) \subset T_p\mathcal{S}$, we can remove the contribution of $\int_0^z f(\boldsymbol{x}, z)dz$. This results in the approximation:

$$\int_{V_p^+(\varepsilon)} f(\boldsymbol{X})\mathrm{dVol} \approx \int_{B_p^+(\varepsilon)} f(\boldsymbol{X})\mathrm{dVol} - \int_{B_p^n(\varepsilon)} \left[\int_{z=0}^{z(\boldsymbol{x})} f(\boldsymbol{x}, z(\boldsymbol{x}))\, dz \right] d^n\boldsymbol{x}.$$

What we need to find is the order of the error in this expression. The volume in the second integral extends outside the ball that defines $V_p^+(\varepsilon)$, which is inscribed in the cylinder, and thus the integral below the hypersurface is subtracting an extra contribution from the region Ω, that lies outside the sphere but inside the cylinder and is bounded by the hypersurface. Then

$$\int_\Omega f(\boldsymbol{X})\mathrm{dVol} \quad \leqslant \quad \max_{\boldsymbol{X}\in\Omega} |f(\boldsymbol{X})| \cdot \mathrm{Vol}(\Omega).$$

Since $z(\rho\overline{\boldsymbol{x}}) \sim \mathcal{O}(\rho^2)$, we have $\max_{\boldsymbol{X}\in\Omega} |f(\boldsymbol{X})| \sim \mathcal{O}(\rho^k(\rho^2)^l)$. To bound the volume of Ω, notice ρ is bounded by ε from the cylinder and by approximately $\varepsilon - C\varepsilon^3$ from the intersection of the sphere with the hypersurface, for some constant C (cf. lemma 5.3.1 below or the estimation in [30]). This maximum thickness $\mathcal{O}(\varepsilon^3)$ is added up for every point of the base sphere, whose area is $\sim \mathcal{O}(\varepsilon^{n-1})$. Now, the maximum height in the z direction of Ω is of order $\mathcal{O}(\varepsilon^2)$ because it is given by the intersection of the cylinder with the hypersurface. Therefore, $\mathrm{Vol}(\Omega) \sim \mathcal{O}(\varepsilon^2\varepsilon^{n-1}\varepsilon^3) \sim \mathcal{O}(\varepsilon^{n+4})$. The total error of this approximation is then $\mathcal{O}(\varepsilon^{k+2l+n+4})$. Finally, the graph function $z(\boldsymbol{x})$ is to be approximated by its osculating quadric, truncating the terms $\mathcal{O}(\rho^3)$ from its Taylor series. This makes a new error in the second integral of our formula, given by the integral over the region inbetween the quadric approximation and the actual hypersurface, which has height given by the $\mathcal{O}(\rho^3)$ difference between the full series of z and the quadratic terms. Therefore, the integral we are neglecting by this truncation makes an error

$$\int_{\mathbb{S}^{n-1}} \int_{\rho=0}^{\rho=\varepsilon} \mathcal{O}(\rho^k(\rho^2)^l)\mathcal{O}(\rho^3)\rho^{n-1} d\rho\, d\mathbb{S} \sim \mathcal{O}(\varepsilon^{k+2l+n+3})$$

which is the leading order of the two errors for small $\varepsilon > 0$. □

The key idea of the approximations carried out in the previous lemma were developed in [30] precisely to obtain the first integral invariant.

Proposition 4.1.3 (Hulin and Troyanov)**.** *The volume of the spherical component cut by a hypersurface has the asymptotic expansion*

$$V(V_p^+(\varepsilon)) = \frac{V_{n+1}(\varepsilon)}{2} - \frac{\varepsilon^2\, V_n(\varepsilon)}{2(n+2)} H + \mathcal{O}(\varepsilon^{n+3}). \tag{4.2}$$

Proof. Using lemma 4.1.2 the computation is immediate since $\int_{B_p^+(\varepsilon)} \mathrm{dVol} = \frac{V_{n+1}(\varepsilon)}{2}$, and

$$\int_{B_p^n(\varepsilon)} \left[\int_{z=0}^{z=\frac{1}{2}\sum_{\mu=1}^n \kappa_\mu x_\mu^2} dz \right] d^n\boldsymbol{x} = \frac{1}{2}\sum_{\mu=1}^{n} \kappa_\mu \left[\int_{B_p^n(\varepsilon)} x_\mu^2\, d^n\boldsymbol{x} \right] = \frac{D_2}{2}\sum_{\mu=1}^{n} \kappa_\mu.$$

□

Proposition 4.1.4. *The barycenter of the spherical component is of the form:*

$$\boldsymbol{s}(V_p^+(\varepsilon)) = [\ 0, \ldots, 0,\ 2\frac{V_n(\varepsilon)}{V_{n+1}(\varepsilon)}\frac{\varepsilon^2}{n+2}\left(1 + \frac{V_n(\varepsilon)}{V_{n+1}(\varepsilon)}\frac{\varepsilon^2}{n+2}H\right)\ ]^T\ + \mathcal{O}(\varepsilon^3). \tag{4.3}$$

Proof. Notice that $\int_{V_p^+(\varepsilon)} \boldsymbol{x}\, \mathrm{dVol} = \mathcal{O}(\varepsilon^{n+4})$ because applying lemma 4.1.2, $\int_{B_p^+(\varepsilon)} \boldsymbol{x}\, d^n\boldsymbol{x}\, dz = 0$, and the second integral is also of monomials of odd degree. We get right away the normal component

$$[V(V_p^+(\varepsilon))\boldsymbol{s}]_z = \int_{B_p^+(\varepsilon)} z\, d^n\boldsymbol{x}\, dz - \int_{B_p^n(\varepsilon)} \frac{1}{2}\left[\sum_{\mu=1}^{n} \kappa_\mu x_\mu^2\right]^2 d^n\boldsymbol{x} + \mathcal{O}(\varepsilon^{n+4}) = D_1^{(n+1)} + \mathcal{O}(\varepsilon^{n+4}),$$

where we have discarded the second integral since its order is $\mathcal{O}(D_4^{(n)}) = \mathcal{O}(D_{22}^{(n)}) \sim \mathcal{O}(\varepsilon^{n+4})$, which leaves the same $\mathcal{O}(\varepsilon^3)$ as the error after dividing by the volume. The final expression follows from inverting the volume formula from the previous proposition and using the value of $D_1^{(n+1)}$ from the appendix. $\square$

Theorem 4.1.5. *The covariance matrix $C(V_p^+(\varepsilon))$ has eigenvalues with the following series expansion, for all $\mu = 1, \ldots, n$:*

$$\lambda_\mu(V_p^+(\varepsilon)) = V_{n+1}(\varepsilon)\frac{\varepsilon^2}{2(n+3)} - V_n(\varepsilon)\frac{\varepsilon^4}{2(n+2)(n+4)}(2\kappa_\mu + H) + \mathcal{O}(\varepsilon^{n+5}), \tag{4.4}$$

$$\lambda_{n+1}(V_p^+(\varepsilon)) = V_{n+1}(\varepsilon)\frac{\varepsilon^2}{2(n+3)} - 2\frac{V_n(\varepsilon)^2}{V_{n+1}(\varepsilon)}\frac{\varepsilon^4}{(n+2)^2}\left(1 + \frac{V_n(\varepsilon)}{V_{n+1}(\varepsilon)}\frac{\varepsilon^2}{n+2}H\right) + \mathcal{O}(\varepsilon^{n+5}). \tag{4.5}$$

Moreover, in the limit $\varepsilon \to 0^+$, when the principal curvatures are different, the corresponding eigenvectors $\boldsymbol{e}_\mu(V_p^+(\varepsilon))$ converge linearly to the principal directions of $\mathcal{S}$ at p, and $\boldsymbol{e}_{n+1}(V_p^+(\varepsilon))$ converges quadratically to the hypersurface normal vector $\boldsymbol{n}$ at p.

Proof. Working in the basis formed by the principal directions and the normal vector of the hypersurface at the fixed point p, we shall compute the entries of the covariance matrix and see that it is diagonal to all orders smaller than $\mathcal{O}(\varepsilon^{n+5})$, precisely the error we get in the diagonal elements, therefore the eigenvalues coincide with those diagonal terms up to that error since differences between eigenvalues of symmetric matrices are bounded by the matrix norm distance. The covariance matrix with respect to the barycenter is

$$C(V_p^+(\varepsilon)) = \int_{V_p^+(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{X}^T \text{ dVol} - \int_{V_p^+(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{s}^T \text{ dVol} - \int_{V_p^+(\varepsilon)} \boldsymbol{s}\otimes\boldsymbol{X}^T \text{ dVol} + \int_{V_p^+(\varepsilon)} \boldsymbol{s}\otimes\boldsymbol{s}^T \text{ dVol},$$

and the last two terms cancel each other upon integration. To compute the second term we can use the expression for $V\boldsymbol{s}$ from the proof of the barycenter formula to get:

$$\int_{V_p^+(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{s}^T \text{ dVol} = V(V_p^+(\varepsilon))\boldsymbol{s}\otimes\boldsymbol{s}^T = \left(\begin{array}{c|c} \mathcal{O}(\varepsilon^{n+7})_{n\times n} & \mathcal{O}(\varepsilon^{n+5})_{n\times 1} \\ \hline \mathcal{O}(\varepsilon^{n+5})_{1\times n} & V(V_p^+(\varepsilon))s_z^2 \end{array}\right)$$

where

$$V(V_p^+(\varepsilon))s_z^2 = \frac{[D_1^{(n+1)}]^2}{V(V_p^+(\varepsilon))} + \mathcal{O}(\varepsilon^{n+5}).$$

The other contribution to the last entry of the covariance matrix is

$$\int_{V_p^+(\varepsilon)} z^2 \text{ dVol} = \int_{B_p^+(\varepsilon)} z^2\, d^n\boldsymbol{x}\, dz - \frac{1}{24}\int_{B_p^n(\varepsilon)} \left[\sum_{\mu=1}^{n} \kappa_\mu x_\mu^2\right]^3 d^n\boldsymbol{x} + \mathcal{O}(\varepsilon^{n+7}) = \frac{D_2^{(n+1)}}{2} + \mathcal{O}(\varepsilon^{n+6}),$$

in which we have neglected the second integral for being of higher order than the barycenter matrix error, whose subtraction yields the stated result for the normal eigenvalue. Notice that the other elements in the last column and row of the complete covariance matrix are $\mathcal{O}(\varepsilon^{n+5})$ since the remaining contributions come from $\int_{V_p^+(\varepsilon)} x_\mu z \text{ dVol} \sim \mathcal{O}(\varepsilon^{n+6})$, and its approximation formula has all monomials with odd powers in x.

Now, we compute the tangent coordinates block. This can be done at once for any $\mu, \nu = 1, \ldots, n$, noticing that when $\mu \neq \nu$, the integrals of lemma 4.1.2 are of monomials of odd degree in tangent coordinates so the off-diagonal elements are $\mathcal{O}(\varepsilon^{n+5})$:

$$\int_{V_p^+(\varepsilon)} x_\mu^2 \,\mathrm{dVol} = \int_{B_p^+(\varepsilon)} x_\mu^2 \, d^n\boldsymbol{x}\, dz - \int_{B_p^n(\varepsilon)} x_\mu^2 \left(\frac{1}{2}\sum_{\alpha=1}^{n} \kappa_\alpha x_\alpha^2 \right) d^n\boldsymbol{x} + \mathcal{O}(\varepsilon^{n+5})$$

$$= \frac{D_2^{(n+1)}}{2} - \frac{1}{2}\int_{B_p^n(\varepsilon)} \left(\kappa_\mu x_\mu^4 + \sum_{\alpha\neq\mu} \kappa_\alpha x_\alpha^2 x_\mu^2 \right) + \mathcal{O}(\varepsilon^{n+5})$$

$$= \frac{D_2^{(n+1)}}{2} - \frac{D_4^{(n)}}{2}\kappa_\mu - \frac{D_{22}^{(n)}}{2}\sum_{\alpha\neq\mu}\kappa_\alpha + \mathcal{O}(\varepsilon^{n+5}) = \frac{D_2^{(n+1)}}{2} - \frac{D_{22}^{(n)}}{2}(2\kappa_\mu + H) + \mathcal{O}(\varepsilon^{n+5})$$

Here we have completed the last sum and used the fact that $D_4 = 3D_{22}$.

The perturbation theory of Hermitian matrices [22], [31] shows the convergence of the eigenvectors to the principal directions in the case of no multiplicity: truncating $C(V_p^+(\varepsilon))$ to order lower than $\mathcal{O}(\varepsilon^{n+5})$, that is precisely the order of the perturbation with respect to the exact diagonalized matrix. Fixing an eigenvalue $\lambda_\mu(V_p^+(\varepsilon))$ with $\mu \neq n+1$, the minimum difference to the other eigenvalues is of order $\sim \varepsilon^{n+4}(\kappa_\mu - \kappa_\nu)$, whereas for the last eigenvalue its distance to all the others is already at leading order $\sim \varepsilon^{n+3}$. Therefore, from the $\sin\theta$ theorem [22], the perturbation $\mathcal{O}(\varepsilon^{n+5})$ changes the eigenvectors $\{\boldsymbol{e}_\mu(V_p^+(\varepsilon))\}_{\mu=1}^n$ with respect to the principal directions as $\mathcal{O}(\varepsilon^{n+5})/\mathcal{O}(\varepsilon^{n+4}(\kappa_\mu - \kappa_\nu)) \sim \frac{\varepsilon}{\kappa_\mu - \kappa_\nu}$, and changes the eigenvector $\boldsymbol{e}_{n+1}(V_p^+(\varepsilon))$ with respect to the normal as $\mathcal{O}(\varepsilon^{n+5})/\mathcal{O}(\varepsilon^{n+3}) \sim \varepsilon^2$, i.e., in the limit $\varepsilon \to 0^+$ the eigevectors of $C(V_p^+(\varepsilon))$ get a vanishing correction with respect to the principal and normal directions. $\square$

Therefore, since the Weingarten operator $\widehat{\boldsymbol{S}}$ at p is $\mathrm{diag}(\kappa_1(p), \ldots, \kappa_n(p))$ in our basis, we may write the covariance matrix as:

$$C(V_p^+(\varepsilon)) = \frac{V_{n+1}(\varepsilon)\,\varepsilon^2}{2(n+3)}\,\mathrm{Id}_{n+1} \;-\; \frac{V_n(\varepsilon)\,\varepsilon^4}{(n+2)(n+4)} \left(\begin{array}{c|c} \widehat{\boldsymbol{S}} + \frac{H}{2}\mathrm{Id}_n & 0_{n\times 1} \\ \hline 0_{1\times n} & 2\frac{V_n(\varepsilon)(n+4)}{V_{n+1}(\varepsilon)(n+2)} \end{array} \right) \;+\; \mathcal{O}(\varepsilon^{n+5}).$$

In [51], following [21], the spherical shell $V_p^+(\varepsilon) \cap \mathbb{S}_p^n(\varepsilon)$ is also considered for surfaces in $\mathbb{R}^3$, and its invariants are shown to be just the derivative with respect to scale of those obtained for the ball region. This is due to the fact that the integral of a function over a region delimited by a ball is the radial integration of the corresponding result over spheres. The same property holds in our case, therefore the derivatives with respect to ε of the invariants in this section are the corresponding integral invariants of the n-dimensional spherical shell.

4.2 Patch Integral Invariants

Now, we shall state the results of the integral invariants of the hypersurface patch domain given by a cylinder intersection as corollaries to the main theorems in the next chapter. Cf. 5.2.1 for

Proposition 4.2.1. *The n-dimensional volume of the hypersurface 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 $T_p\mathcal{S}$:*

$$V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = V_n(1) \prod_{\mu=1}^{n} \ell_\mu + \mathcal{O}(\varepsilon^{n+1}), \tag{4.6}$$

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

$$V(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon) \left[1 + \frac{\varepsilon^2}{2(n+2)} \sum_{\mu=1}^{n} \kappa_\mu^2 + \mathcal{O}(\varepsilon^4) \right]. \tag{4.7}$$

The barycenter for the cylindrical domain is the same as for the spherical domain computed below in proposition 4.2.5. Finally, the covariance matrix analysis yields a direct relation between its eigenvalues and the squares of the principal curvatures.

Theorem 4.2.2. *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 $\mathrm{Cyl}_p(\varepsilon, \mathbb{V})$ is finite, the covariance matrix $C_p(\varepsilon, \mathbb{V})$ of a hypersurface $\mathcal{S}$ has n limit eigenvectors that form an orthonormal basis of $T_p\mathcal{S}$, corresponding to the first n eigenvalues that scale as ε^2. The other eigenvalue scales at higher order and has limit eigenvector converging to the normal of $\mathcal{S}$ at p:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \frac{\varepsilon^2}{n+2} \ell_\mu^2 V_n(1) \prod_{\alpha=1}^{n} \ell_\alpha + \mathcal{O}(\varepsilon^{n+3}), \qquad \mu = 1, \ldots, n, \tag{4.8}$$

$$\lambda_{n+1}(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = 0 + \mathcal{O}(\varepsilon^{n+3}), \tag{4.9}$$

where ℓ_μ are the principal lengths *of the ellipsoid in 4.2.1. When $\mathbb{V} = T_p\mathcal{M}$ the eigenvalues of the covariance matrix of the cylindrical component are:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{2(n+2)(n+4)}\left(2\kappa_\mu^2 + \sum_{\alpha=1}^{n}\kappa_\alpha^2\right) + \ \mathcal{O}(\varepsilon^6)\right] \tag{4.10}$$

$$\lambda_{n+1}(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^4}{4(n+2)(n+4)}(3H^2 - 2\mathcal{R}) + \mathcal{O}(\varepsilon^6)\right] \tag{4.11}$$

for all $\mu = 1, \ldots, n$. Moreover, if the principal curvatures are different, the first n eigenvectors converge to the principal directions, and the last eigenvector to the normal direction at p.

However, as a warm up exercise for the more involved computations of the general codimension case, we shall explicitly compute below the asymptotic expansions of the integral invariants of the hypersurface spherical patch. We are integrating over the domain $D_p(\varepsilon) = \mathcal{S} \cap B_p^{n+1}(\varepsilon)$, using again the local graph representation in a small neighborhood around the point. What follows served as a toy model for the general case and was obtained first during our research.

Since a parametrization of the region is needed to perform the integrals locally, we need to find local parametric equations of the boundary $\partial(\mathcal{S} \cap B_p^{n+1}(\varepsilon))$, which is no longer a sphere (cf. Figure 4.1), to high enough order in ε so that we can expand asymptotically the integral invariants in terms of the geometric information of the hypersurface at the point. The strategy of [51], hinted in [30], obtaining a cylindrical coordinate approximation for the boundary radius of the patch, works in general dimension as follows. The result is general for higher codimension so the proof is given in lemma 5.3.1.

Lemma 4.2.3. *In cylindrical coordinates $(\rho, \phi_1, \ldots, \phi_{n-1}, z)$ over the tangent space $T_p\mathcal{S}$, fixing the basis to the principal directions and the normal vector of $\mathcal{S}$ at p, the parametric equations of a point $\boldsymbol{X} = (\rho\overline{x}_1, \ldots, \rho\overline{x}_n, z)^T$ in $\partial D_p(\varepsilon) = \mathcal{S} \cap \mathbb{S}_p^n(\varepsilon)$, are*

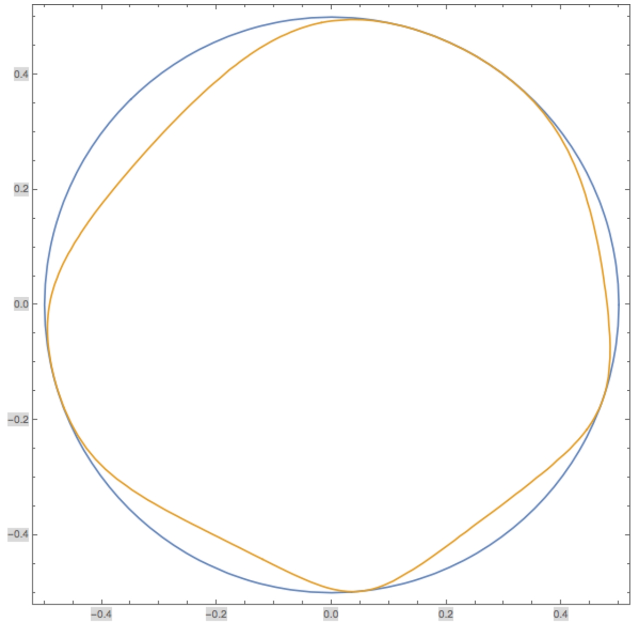

Figure 4.1: The intersection of a sphere with a hypersurface no longer projects as a ball onto the tangent space, which renders local integration more difficult. Lemma 5.3.1 provides a way to tackle the correction terms due to the irregularities of the boundary. Here, the example of the integration domain of the graph hypersurface $z = x^5 - 2x^3y + 3x^2y + \frac{x^2}{2} + 3xy^3 - 5xy^2 + \frac{y^2}{3}$, for $\varepsilon = 0.5$.

$$r(\overline{\boldsymbol{x}}) := \rho(\overline{x}_1, \dots, \overline{x}_n) = \varepsilon - \frac{1}{8}\kappa^2(\overline{\boldsymbol{x}})\varepsilon^3 + \mathcal{O}(\varepsilon^4), \qquad z(\overline{x}_1, \dots, \overline{x}_n) = \frac{1}{2}\kappa^2(\overline{\boldsymbol{x}})\varepsilon^2 + \mathcal{O}(\varepsilon^3), \quad (4.12)$$

where $\overline{x}_1, \dots, \overline{x}_n$ are the coordinates of points on $\mathbb{S}^{n-1} \subset T_p\mathcal{S}$, and

$$\kappa(\overline{\boldsymbol{x}}) = \kappa(\overline{x}_1, \dots, \overline{x}_n) = \sum_{\mu=1}^{n} \kappa_\mu \overline{x}_\mu^2 \tag{4.13}$$

is the normal curvature of $\mathcal{S}$ at p in the direction of $\overline{\boldsymbol{x}}$.

For this type of domain the previous parametric expansions are enough to asymptotically expand both the integrand and the measure, collect terms and solve the integrals using the appendix formulas. The area or mass of the domain can be expressed as a correction to the volume of the n-ball in terms of the extrinsic and intrinsic curvature of $\mathcal{S}$ at the point.

Proposition 4.2.4. *The n-dimensional area of the hypersurface patch has the asymptotic expansion*

$$V(D_p(\varepsilon)) = V_n(\varepsilon)\left[1 + \frac{\varepsilon^2}{8(n+2)}(H^2 - 2\mathcal{R}) + \mathcal{O}(\varepsilon^3)\right]. \tag{4.14}$$

Proof. Using lemma 5.3.1 in equation 2.15, we have that

$$\mathrm{dVol}|_{D_p(\varepsilon)} = \sqrt{\det g(\boldsymbol{x})}\, dx_1 \cdots dx_n = \left[1 + \frac{1}{2}\sum_{\mu=1}^{n} \kappa_\mu^2 x_\mu^2 + \mathcal{O}(x^3)\right] dx_1 \cdots dx_n,$$

since $\sum_{\mu=1}^{n}\left(\frac{\partial z}{\partial x_\mu}\right)^2 = \|\nabla z(\boldsymbol{x})\|^2$ can be considered small for small enough $\varepsilon > 0$, because in our coordinates $\nabla z(\mathbf{0}) = 0$. With this and the cylindrical measure, eq. A.1, the integration becomes

$$V(D_p(\varepsilon)) = \int_{\mathcal{S}\cap B_p^{n+1}(\varepsilon)} \mathrm{dVol} = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \left[1 + \frac{1}{2}\sum_{\mu=1}^{n} \kappa_\mu^2 \rho^2 \overline{x}_\mu^2 + \mathcal{O}(\rho^3)\right] \rho^{n-1}\, d\rho$$

$$= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \left[\frac{1}{n}(\varepsilon - \frac{\kappa(\overline{\boldsymbol{x}})^2\varepsilon^3}{8} + \mathcal{O}(\varepsilon^4))^n + \frac{1}{2}\sum_{\mu=1}^{n} \frac{\kappa_\mu^2 \overline{x}_\mu^2}{n+2}(\varepsilon - \frac{\kappa(\overline{\boldsymbol{x}})^2\varepsilon^3}{8} + \mathcal{O}(\varepsilon^4))^{n+2} + \mathcal{O}(\varepsilon^{n+4})\right]$$

after integrating over ρ up to the boundary radius. Expanding the binomial series and the square of the normal curvature, all the remaining integrals are in example A.0.4, leading to

$$V(D_p(\varepsilon)) = \frac{\varepsilon^n}{n} S_{n-1} - \frac{\varepsilon^{n+2}}{8}\int_{\mathbb{S}^{n-1}} \kappa(\overline{\boldsymbol{x}})^2\, d\,\mathbb{S} + \frac{\varepsilon^{n+2}}{2(n+2)}\sum_{\mu=1}^{n} \kappa_\mu^2 \int_{\mathbb{S}^{n-1}} \overline{x}_\mu^2\, d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+3})$$

$$= V_n(\varepsilon) - \frac{\varepsilon^{n+2}}{8}\int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \left(\sum_{\mu=1}^{n} \kappa_\mu^2 \overline{x}_\mu^4 + 2\sum_{\mu<\nu}^{n} \kappa_\mu\kappa_\nu \overline{x}_\mu^2 \overline{x}_\nu^2\right) + \frac{C_2\,\varepsilon^{n+2}}{2(n+2)}\sum_{\mu=1}^{n} \kappa_\mu^2 + \mathcal{O}(\varepsilon^{n+3})$$

$$= V_n(\varepsilon) + \frac{\varepsilon^{n+2}}{n+2}\left[\left(\frac{C_2}{2} - \frac{n+2}{8}C_4\right)\sum_{\mu=1}^{n} \kappa_\mu^2 - C_{22}\frac{n+2}{8} 2\sum_{\mu<\nu}^{n} \kappa_\mu\kappa_\nu\right] + \mathcal{O}(\varepsilon^{n+3})$$

$$= V_n(\varepsilon) + \frac{\varepsilon^{n+2}}{n+2}\left[\frac{C_2}{8}(H^2 - \mathcal{R}) - \frac{C_2}{8}\mathcal{R}\right] + \mathcal{O}(\varepsilon^{n+3}),$$

where we use equation 2.12 and the relations among the coefficients from the appendix. □

It is natural to expect the extrinsic curvature H to be present in the second order correction since the domain depends on how $\mathcal{S}$ is embedded, in contrast to an intrinsically defined geodesic ball where the correction only depends on $\mathcal{R}$. Now, the center of mass in this case turns out to deviate from the center of the ball, to leading order in ε, only in the normal direction.

Proposition 4.2.5. *The barycenter of the patch region has coordinates in the principal basis with respect to* p *given by*

$$\boldsymbol{s}(D_p(\varepsilon)) = [\,\mathcal{O}(\varepsilon^4), \dots, \mathcal{O}(\varepsilon^4),\ \frac{\varepsilon^2}{2(n+2)}H + \mathcal{O}(\varepsilon^3)\,]^T. \tag{4.15}$$

Proof. When integrating any tangent component x_α of $\boldsymbol{X}$, only factors with an odd power in some component are produced because the known terms (see previous proof) now contain products $\overline{x}_\alpha \overline{x}_\mu^2,\ \overline{x}_\alpha \overline{x}_\mu^4$ and $\overline{x}_\alpha \overline{x}_\mu^2 \overline{x}_\nu^2$, which always have an odd power factor regardless of the subindices combination. Therefore the first n components of $V(D_p(\varepsilon))\boldsymbol{s}(D_p(\varepsilon))$ are of order $\mathcal{O}(\varepsilon^{n+4})$, coming from the error inside $r(\overline{\boldsymbol{x}})^{n+1}$ after integrating radially the first term $x_\alpha \rho^{n-1} d\rho$. The normal component of $\boldsymbol{X}$ integrates as

$$\begin{aligned}
\int_{\mathcal{S}\cap B_p^{n+1}(\varepsilon)} z \,\mathrm{dVol} &= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \left[\frac{1}{2}\kappa(\overline{\boldsymbol{x}})\rho^2 + \mathcal{O}(\rho^3)\right] \left[1 + \frac{1}{2}\sum_{\mu=1}^{n} \kappa_\mu^2 \rho^2 \overline{x}_\mu^2 + \mathcal{O}(\rho^3)\right] \rho^{n-1}\, d\rho \\
&= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \left[\frac{\kappa(\overline{\boldsymbol{x}})}{2(n+2)}(\varepsilon - \frac{\kappa(\overline{\boldsymbol{x}})^2}{8}\varepsilon^3 + \mathcal{O}(\varepsilon^4))^{n+2} + \mathcal{O}(\varepsilon^{n+3})\right] \\
&= \frac{\varepsilon^{n+2}}{2(n+2)} \sum_{\mu=1}^{n} \kappa_\mu \int_{\mathbb{S}^{n-1}} \overline{x}_\mu^2 \, d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+3}) = C_2 \frac{\varepsilon^{n+2}}{2(n+2)} H + \mathcal{O}(\varepsilon^{n+3}).
\end{aligned}$$

Then normalizing by the volume to lowest order cancels the coefficient $C_2\varepsilon^n$. □

Finally, the study of the covariance matrix of the patch domain shows a behavior similar to the spherical component, but where the next-to-leading order contribution to the eigenvalues includes only products of principal curvatures and no linear terms.

Theorem 4.2.6. *The covariance matrix* $C(D_p(\varepsilon))$ *has* n *eigenvalues that scale like* ε^{n+2} *as*

$$\lambda_\mu(D_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{8(n+2)(n+4)}(H^2 - 2\mathcal{R} - 4H\kappa_\mu)\right] + \mathcal{O}(\varepsilon^{n+5}), \tag{4.16}$$

for all $\mu = 1, \dots, n$*, and one eigenvalue scaling as* ε^{n+4} *with leading term*

$$\lambda_{n+1}(D_p(\varepsilon)) = V_n(\varepsilon)\frac{\varepsilon^4}{2(n+2)(n+4)}\left(\frac{n+1}{n+2}H^2 - \mathcal{R}\right) + \mathcal{O}(\varepsilon^{n+5}). \tag{4.17}$$

Moreover, in the limit $\varepsilon \to 0^+$, if the principal curvatures at p are all different, the eigenvectors $\boldsymbol{e}_\mu(D_p(\varepsilon))$ corresponding to the first n eigenvalues converge to the principal directions of $\mathcal{S}$ at p, and the last eigenvector $\boldsymbol{e}_{n+1}(D_p(\varepsilon))$ converges to the hypersurface normal vector $\boldsymbol{n}(p)$.

Proof. We need to evaluate $\int_{D_p(\varepsilon)} \boldsymbol{X}(\boldsymbol{x})\otimes\boldsymbol{X}(\boldsymbol{x})^T \sqrt{\det g}\, d^n\boldsymbol{x}$ and $V(D_p(\varepsilon))\boldsymbol{s}(D_p(\varepsilon))\otimes\boldsymbol{s}(D_p(\varepsilon))^T$. The latter can be obtained from the previous proof:

$$[\,\mathcal{O}(\varepsilon^{n+4}),\ldots,\mathcal{O}(\varepsilon^{n+4}),\, \frac{C_2\varepsilon^{n+2}}{2(n+2)}H + \mathcal{O}(\varepsilon^{n+3})\,]^T \otimes [\,\mathcal{O}(\varepsilon^4),\ldots,\mathcal{O}(\varepsilon^4),\, \frac{\varepsilon^2}{2(n+2)}H + \mathcal{O}(\varepsilon^3)\,],$$

resulting in all entries of the $n \times n$ block being $\mathcal{O}(\varepsilon^{n+8})$, the first n elements of the last column and last row being $\mathcal{O}(\varepsilon^{n+6})$, and the last element of the matrix becoming

$$[V(D_p(\varepsilon))\boldsymbol{s}(D_p(\varepsilon)) \otimes \boldsymbol{s}(D_p(\varepsilon))^T]_{(n+1),(n+1)} = \frac{V_n(\varepsilon)\,\varepsilon^4}{4(n+2)^2}H^2 + \mathcal{O}(\varepsilon^{n+5}),$$

(we already disregarded the term of $\mathcal{O}(\varepsilon^{n+6})$ that can be computed for this matrix entry because we shall see below that the other contributing term in that position has error at $\mathcal{O}(\varepsilon^{n+5})$).

Now, the rest of the covariance matrix requires the longest computations so far. The entries of $\boldsymbol{X}(\boldsymbol{x}) \otimes \boldsymbol{X}(\boldsymbol{x})^T$ are of three types: $x_\mu x_\nu$, $x_\mu z(\boldsymbol{x})$ and $z(\boldsymbol{x})^2$. The first n entries of the last column and last row, $x_\mu z(\boldsymbol{x})$, contribute at order $\mathcal{O}(\varepsilon^{n+4})$. This implies that the matrix may not decompose at order $\mathcal{O}(\varepsilon^{n+4})$ as the direct sum of a "tangent" $n\times n$ block, the integrals of $[x_\mu x_\nu]$, and a "normal" 1×1 block, the integral of $z(\boldsymbol{x})^2$. Then the argument in the proof of theorem 4.1.5 to equate the diagonal elements of this expansion with that of the actual eigenvalues cannot be made here, since there are off-diagonal error elements at the same order as the diagonal approximation. However, this will not affect the eigenvalue decomposition as we shall see later in the crucial lemma 5.2.3, and the eigenvalues will be given to order $\mathcal{O}(\varepsilon^{n+4})$ by the diagonals of these blocks.

The normal block entry is:

$$\int_{\mathcal{S}\cap B_p^{n+1}(\varepsilon)} z^2 \,\mathrm{dVol} = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \left[\frac{1}{4}\kappa(\overline{\boldsymbol{x}})^2\rho^4 + \mathcal{O}(\rho^5)\right]\left[1 + \frac{1}{2}\sum_{\mu=1}^{n}\kappa_\mu^2\rho^2\overline{x}_\mu^2 + \mathcal{O}(\rho^3)\right]\rho^{n-1}\,d\rho$$

$$= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \left[\frac{1}{4}\sum_{\alpha,\beta=1}^{n}\kappa_\alpha\kappa_\beta\overline{x}_\alpha^2\overline{x}_\beta^2\,\rho^{n+3} + \mathcal{O}(\rho^{n+4})\right]d\rho$$

$$= \frac{1}{4}\left[\sum_{\alpha=1}^{n}\kappa_\alpha^2\int_{\mathbb{S}^{n-1}}\overline{x}_\alpha^4\,d\,\mathbb{S} + 2\sum_{\alpha<\beta}^{n}\kappa_\alpha\kappa_\beta\int_{\mathbb{S}^{n-1}}\overline{x}_\alpha^2\overline{x}_\beta^2\,d\,\mathbb{S}\right]\frac{\varepsilon^{n+4}}{n+4} + \mathcal{O}(\varepsilon^{n+5})$$

$$= \frac{\varepsilon^{n+4}}{4(n+4)}\left[C_4(H^2-\mathcal{R}) + C_{22}\mathcal{R}\right] + \mathcal{O}(\varepsilon^{n+5});$$

subtracting the contribution from the barycenter matrix term, the last eigenvalue becomes

$$\lambda_{n+1}(p,\varepsilon) = \frac{C_2\,\varepsilon^{n+4}}{4(n+2)(n+4)}\left[3H^2 - 2\mathcal{R}\right] - \frac{C_2\,\varepsilon^{n+4}}{4(n+2)^2}H^2 + \mathcal{O}(\varepsilon^{n+5})$$

which simplifies to the stated result.

The tangent block entries can be computed simultaneously considering arbitrary $\mu,\nu = 1,...,n$:

$$\int_{\mathcal{S}\cap B_p^{n+1}(\varepsilon)} x_\mu x_\nu \,\mathrm{dVol} = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S}\int_0^{r(\overline{\boldsymbol{x}})} \rho^2\overline{x}_\mu\overline{x}_\nu\rho^{n-1}\left[1 + \frac{1}{2}\sum_{\alpha=1}^{n}\kappa_\alpha^2\rho^2\overline{x}_\alpha^2 + \mathcal{O}(\rho^3)\right]d\rho$$

$$= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S}\frac{\overline{x}_\mu\overline{x}_\nu}{n+2}\Big(\varepsilon - \frac{\kappa(\overline{\boldsymbol{x}})^2}{8}\varepsilon^3 + \mathcal{O}(\varepsilon^4)\Big)^{n+2} + \frac{1}{2}\sum_{\alpha=1}^{n}\kappa_\alpha^2\int_{\mathbb{S}^{n-1}}\overline{x}_\alpha^2\overline{x}_\mu\overline{x}_\nu\,d\,\mathbb{S}\frac{\varepsilon^{n+4}}{n+4} + \mathcal{O}(\varepsilon^{n+5})$$

$$= \frac{\varepsilon^{n+2}}{n+2}\left[\delta_{\mu\nu}C_2 - \frac{\varepsilon^2(n+2)}{8}\int_{\mathbb{S}^{n-1}} d\,\mathbb{S}\,\overline{x}_\mu\overline{x}_\nu\left(\sum_{\alpha=1}^{n}\kappa_\alpha^2\overline{x}_\alpha^4 + 2\sum_{\alpha<\beta}^{n}\kappa_\alpha\kappa_\beta\overline{x}_\alpha^2\overline{x}_\beta^2\right)\right] +$$

$$+ \frac{\varepsilon^{n+4}}{n+4}\frac{\delta_{\mu\nu}}{2}\left(\kappa_\mu^2\int_{\mathbb{S}^{n-1}}\overline{x}_\mu^4\,d\,\mathbb{S} + \sum_{\alpha\neq\mu}^{n}\kappa_\alpha^2\int_{\mathbb{S}^{n-1}}\overline{x}_\alpha^2\overline{x}_\mu^2\,d\,\mathbb{S}\right) + \mathcal{O}(\varepsilon^{n+5}),$$

where the $\delta_{\mu\nu}$ appears because the monomials get an odd power if $\mu \neq \nu$. Now, the different integrals inside the indexed sums result in different constants depending on the different monomials that the terms $\overline{x}_\mu^2\overline{x}_\alpha^4$ and $\overline{x}_\mu^2\overline{x}_\alpha^2\overline{x}_\beta^2$ can combine into, thus

$$= C_2 \frac{\varepsilon^{n+2}}{n+2}\delta_{\mu\nu} + \frac{\varepsilon^{n+4}}{n+4}\delta_{\mu\nu}\left[-\frac{n+4}{8}\left(C_6\kappa_\mu^2 + C_{24}\sum_{\alpha\neq\mu}^n \kappa_\alpha^2 + \sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}^n \kappa_\alpha\kappa_\beta(\delta_{\alpha\mu}+\delta_{\beta\mu})C_{24} + \right.\right.$$

$$\left.\left. + \sum_{\substack{\alpha\neq\beta\\ \alpha,\beta\neq\mu}}^n \kappa_\alpha\kappa_\beta C_{222}\right) + \frac{C_4}{2}\kappa_\mu^2 + \frac{C_{22}}{2}\sum_{\alpha\neq\mu}^n \kappa_\alpha^2\right] + \mathcal{O}(\varepsilon^{n+5})$$

$$= C_2 \frac{\varepsilon^{n+2}}{n+2}\delta_{\mu\nu} + \frac{\varepsilon^{n+4}}{n+4}\delta_{\mu\nu}\left[\left(\frac{C_4}{2} - \frac{n+4}{8}C_6\right)\kappa_\mu^2 + \left(\frac{C_{22}}{2} - \frac{n+4}{8}C_{24}\right)\sum_{\alpha\neq\mu}^n \kappa_\alpha^2\right.$$

$$\left. -\frac{n+4}{8}\left(2C_{24}\sum_{\alpha\neq\mu}^n \kappa_\mu\kappa_\alpha + C_{222}\sum_{\substack{\alpha\neq\beta\\ \alpha,\beta\neq\mu}}^n \kappa_\alpha\kappa_\beta\right)\right] + \mathcal{O}(\varepsilon^{n+5}).$$

Notice that the summations in the last equation are all over indices that must be different from μ, so we can add and subtract the corresponding missing terms to those sums as long as we subtract them in the correct place. Doing this, and using the crucial relationships between the constants from the appendix, each of the different terms under the big braces simplify to:

$$\left(\frac{C_4}{2} - \frac{n+4}{8}C_6 - \frac{C_{22}}{2} + \frac{n+4}{8}C_{24}\right)\kappa_\mu^2 = -\frac{C_2}{2(n+2)}\kappa_\mu^2,$$

$$\left(\frac{C_{22}}{2} - \frac{n+4}{8}C_{24}\right)\sum_{\alpha=1}^n \kappa_\alpha^2 = \frac{C_2}{8(n+2)}(H^2 - \mathcal{R}),$$

$$-\frac{n+4}{8}\left((2C_{24} - 2C_{222})\sum_{\alpha\neq\mu}^n \kappa_\mu\kappa_\alpha + 2C_{222}\sum_{\alpha<\beta}^n \kappa_\alpha\kappa_\beta\right) = -\frac{C_2}{2(n+2)}R_{\mu\mu} + \frac{C_2}{8(n+2)}\mathcal{R}.$$

Finally, these lead to the expression

$$\int_{D_p(\varepsilon)} x_\mu x_\nu \mathrm{dVol} = \delta_{\mu\nu}V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{8(n+2)(n+4)}(H^2 - 2\mathcal{R} - 4\kappa_\mu^2 - 4R_{\mu\mu})\right] + \mathcal{O}(\varepsilon^{n+5}),$$

and since $\kappa_\mu^2 + R_{\mu\mu} = \kappa_\mu H$, from equation 2.13, the stated formula for the tangent eigenvalues follows from the diagonal of this block. Therefore, we can write $C(D_p^+(\varepsilon)) =$

$$\frac{V_n(\varepsilon)\,\varepsilon^2}{n+2}\left(\begin{array}{c|c} \mathrm{Id}_n & 0_{n\times 1} \\ \hline 0_{1\times n} & 0 \end{array}\right) + \frac{V_n(\varepsilon)\,\varepsilon^4}{2(n+2)(n+4)}\left(\begin{array}{c|c} \frac{H^2-2\mathcal{R}}{4}\mathrm{Id}_n - H\,\widehat{\boldsymbol{S}} & A_{n\times 1} \\ \hline A_{1\times n} & \frac{n+1}{n+2}H^2 - \mathcal{R} \end{array}\right) + \mathcal{O}(\varepsilon^{n+5}),$$

so the Weingarten operator appears inside the covariance matrix in this case as well but multiplied by the mean curvature, which is a term in equation 2.13. □

These covariance matrix eigenvalues will be inverted in chapter 6 to extract the principal curvatures and obtain descriptors at scale of them by truncating the series. The eigenvectors at fixed $\varepsilon > 0$ also coverge to the principal and normal directions, so they serve as multi-scale estimators of these as well. The spherical component invariants provide a direct relationship to the Weingarten operator, thus the principal curvatures will be estimated without the need for sign choices. In the cylindrical and spherical cases, the principal curvatures appear in products which leads to sign choices that can be made using the barycenter.

Chapter 5

Covariance Analysis of Embedded Riemannian Manifolds

It is shown in this chapter 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 [26], [58] to general submanifolds and prove that the eigenvalue decomposition of the covariance matrices of the domains have asymptotic expansions with scale that contain the curvature information encoded by the traces of this tensor, where the limit eigenvectors converge to its generalized principal directions. Theorems 5.2.4 and 5.3.5 represent the most important contributions of this thesis, proving for embedded submanifolds of arbitrary dimension the direct relationship between PCA covariance analysis and the generalized principal curvatures and directions that can be defined from the third fundamental form operators. This achieves a major development with respect to the leading order approximations of [56], and the expansions for surfaces in $\mathbb{R}^3$ of [51].

5.1 Third Fundamental Form of a Riemannian Submanifold

In classical differential geometry, [26], [58], the third fundamental form is the natural object to construct out of scalar products 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. [40], 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 2.1.3, i.e., in Euclidean space [32, Ch. VII, Prop. 5.2]:

$$\langle\,\widehat{\boldsymbol{S}}^2\boldsymbol{x},\,\boldsymbol{y}\,\rangle = H\langle\,\widehat{\boldsymbol{S}}\,\boldsymbol{x},\,\boldsymbol{y}\,\rangle - \boldsymbol{\mathcal{R}ic}(\boldsymbol{x},\boldsymbol{y}), \tag{5.1}$$

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 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. See §2.1. Therefore, the natural generalization of $\langle\,\widehat{\boldsymbol{S}}\,\boldsymbol{x},\,\widehat{\boldsymbol{S}}\,\boldsymbol{y}\,\rangle$ to this context is

Definition 5.1.1. 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

$$\langle\,\mathbf{III}(\boldsymbol{x},\boldsymbol{y})\,\boldsymbol{n},\,\boldsymbol{m}\,\rangle := \langle\,\widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{x}\,,\,\widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{y}\,\rangle. \tag{5.2}$$

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 $\boldsymbol{\eta}^j = g(\cdot,\boldsymbol{n}_j)$ is the dual basis:

$$\mathbf{III}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i,\,j=1}^{k}\langle\,\widehat{\boldsymbol{S}}_i\,\widehat{\boldsymbol{S}}_j\,\boldsymbol{x}\,,\,\boldsymbol{y}\,\rangle\,\boldsymbol{\eta}^i \otimes \boldsymbol{n}_j. \tag{5.3}$$

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 $T_p\mathcal{M}$:

$$\mathrm{tr}_{\parallel}\boldsymbol{A} := \sum_{\mu=1}^{n} \boldsymbol{A}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) \ \in \mathrm{End}(N_p\mathcal{M}), \tag{5.4}$$

And let the *normal trace* of such a tensor be

$$\mathrm{tr}_{\perp}\boldsymbol{A} := \sum_{j=1}^{k} \big\langle\, \mathbf{III}(\cdot,\cdot)\,\boldsymbol{n}_j,\, \boldsymbol{n}_j \,\big\rangle \ \in (T_p\mathcal{M}^*)^2, \tag{5.5}$$

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

Lemma 5.1.2. *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*

$$\mathrm{tr}_{\perp}\mathbf{III}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{k} \big\langle\, \widehat{\boldsymbol{S}}_j^2\, \boldsymbol{x},\, \boldsymbol{y} \,\big\rangle = \big\langle\, (\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}} + \overline{\boldsymbol{\mathcal{R}}})\,\boldsymbol{x},\, \boldsymbol{y} \,\big\rangle, \tag{5.6}$$

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:

$$\big\langle\, (\mathrm{tr}_{\parallel}\mathbf{III})\, \boldsymbol{n},\, \boldsymbol{m} \,\big\rangle = \mathrm{tr}\,(\widehat{\boldsymbol{S}}_{\boldsymbol{n}}\widehat{\boldsymbol{S}}_{\boldsymbol{m}}). \tag{5.7}$$

The total trace is

$$\mathrm{tr}\,\mathbf{III} = \mathrm{tr}_{\perp}\mathrm{tr}_{\parallel}\mathbf{III} = \|\boldsymbol{H}\|^2 - \mathcal{R} + \overline{\mathcal{R}}. \tag{5.8}$$

Proof. The normal trace bilinear form has components

$$\begin{aligned}
\mathrm{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) &= \sum_{j=1}^{k} \big\langle\, \widehat{\boldsymbol{S}}_j\, \boldsymbol{e}_\mu,\, \widehat{\boldsymbol{S}}_j\, \boldsymbol{e}_\nu \,\big\rangle = \sum_{j=1}^{k}\sum_{\alpha=1}^{n} \big\langle\, \widehat{\boldsymbol{S}}_j \boldsymbol{e}_\alpha,\, \boldsymbol{e}_\mu \,\big\rangle \big\langle\, \widehat{\boldsymbol{S}}_j \boldsymbol{e}_\alpha,\, \boldsymbol{e}_\nu \,\big\rangle \\
&= \sum_{j=1}^{k}\sum_{\alpha=1}^{n} \mathrm{II}^j(\boldsymbol{e}_\alpha, \boldsymbol{e}_\mu)\mathrm{II}^j(\boldsymbol{e}_\alpha, \boldsymbol{e}_\nu) = \sum_{\alpha=1}^{n} \big\langle\, \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\mu),\, \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\nu) \,\big\rangle,
\end{aligned} \tag{5.9}$$

that using Gauß equation, th. 2.1.3, lead to the corresponding linear operator with respect to the metric:

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

This is the generalization of the operator of the classical third fundamental form, equation 5.1:

$$
\sum_{j=1}^{k}\widehat{\boldsymbol{S}}_j^2 = \widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}} + \overline{\boldsymbol{\mathcal{R}}}.
$$

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:

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

In a fixed orthonormal basis this tensor is the linear combination

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

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

$$
\operatorname{tr}(\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_j) = \sum_{\mu,\,\nu=1}^{n}\langle\,\widehat{\boldsymbol{S}}_i\;\boldsymbol{e}_\mu,\,\boldsymbol{e}_\nu\,\rangle\langle\,\widehat{\boldsymbol{S}}_j\;\boldsymbol{e}_\mu,\,\boldsymbol{e}_\nu\,\rangle = \sum_{\mu,\,\nu=1}^{n}\mathrm{II}^i(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu)\mathrm{II}^j(\boldsymbol{e}_\mu,\boldsymbol{e}_\nu). \tag{5.10}
$$

Taking the total trace of $\mathbf{III}$ is analogous to the complete contraction of the Riemann curvature tensor indices to obtain the scalar curvature:

$$\operatorname{tr}\mathbf{III} = \operatorname{tr}_{\|}\operatorname{tr}_{\perp}\mathbf{III} = \sum_{\mu=1}^{n}\langle\, (\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \widehat{\boldsymbol{\mathcal{R}}} + \overline{\boldsymbol{\mathcal{R}}})\boldsymbol{e}_\mu,\ \boldsymbol{e}_\mu\,\rangle = \operatorname{tr}\widehat{\boldsymbol{S}}_{\boldsymbol{H}} - \operatorname{tr}\widehat{\boldsymbol{\mathcal{R}}} + \operatorname{tr}\overline{\boldsymbol{\mathcal{R}}}$$

$$= \sum_{\mu=1}^{n}\operatorname{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu, \boldsymbol{e}_\mu) = \sum_{\alpha,\,\beta}^{n}\|\mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta)\|^2, \tag{5.11}$$

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. □

Equations 5.9 and 5.10 shall 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.

Example 5.1.3. For a smooth hypersurface $\mathcal{S}$, there is only one unit normal vector $\boldsymbol{n}$ at every point $p \in \mathcal{S}$, up to orientation. Choosing $\{\boldsymbol{e}_\mu\}_{\mu=1}^{n}$ as the orthonormal basis of the tangent space given by the principal directions at p, the components of the third fundamental form are:

$$\langle\,\mathbf{III}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)\boldsymbol{n},\ \boldsymbol{n}\,\rangle = \langle\,\widehat{\boldsymbol{S}}\,\boldsymbol{e}_\mu,\ \widehat{\boldsymbol{S}}\,\boldsymbol{e}_\nu\,\rangle = \langle\,\widehat{\boldsymbol{S}}^2\,\boldsymbol{e}_\mu,\ \boldsymbol{e}_\nu\,\rangle = \kappa_\mu^2\delta_{\mu\nu} = \operatorname{tr}_{\perp}\mathbf{III}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu). \tag{5.12}$$

The tangent trace component coincides with the total trace:

$$\langle\,\operatorname{tr}_{\|}\mathbf{III}\,\boldsymbol{n},\ \boldsymbol{n}\,\rangle = \operatorname{tr}(\widehat{\boldsymbol{S}}^2) = \sum_{\mu=1}^{n}\kappa_\mu^2 = \left(\sum_{\mu=1}^{n}\kappa_\mu\right)^2 - 2\sum_{\mu<\nu}^{n}\kappa_\mu\kappa_\nu = H^2 - \mathcal{R} = \operatorname{tr}\mathbf{III}. \tag{5.13}$$

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.

Lemma 5.1.4 (Ricci equation)**.** *The Riemann curvature of the induced normal connection, $\boldsymbol{R}_{\perp}$, satisfies:*

$$\langle\boldsymbol{R}_{\perp}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m}\,\rangle = \langle\,\overline{\boldsymbol{R}}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m}\,\rangle + \langle\,\mathbf{III}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{n}, \boldsymbol{m}\,\rangle - \langle\,\mathbf{III}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{m}, \boldsymbol{n}\,\rangle, \tag{5.14}$$

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

Proof. Writing the classical equation [16, Ex. II.11] in terms of Weingarten maps leads to

$$\langle \overline{\boldsymbol{R}}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{n},\boldsymbol{m}\rangle - \langle \boldsymbol{R}_{\perp}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{n},\boldsymbol{m}\rangle = \sum_{\mu=1}^{n} [\,\langle \mathbf{II}(\boldsymbol{e}_\mu,\boldsymbol{x}),\,\boldsymbol{n}\rangle\langle \mathbf{II}(\boldsymbol{e}_\mu,\boldsymbol{y}),\,\boldsymbol{m}\rangle +$$

$$-\langle \mathbf{II}(\boldsymbol{e}_\mu,\boldsymbol{y}),\,\boldsymbol{n}\rangle\langle \mathbf{II}(\boldsymbol{e}_\mu,\boldsymbol{x}),\,\boldsymbol{m}\rangle\,]$$

$$= \sum_{\mu=1}^{n}\langle \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{x},\,\boldsymbol{e}_\mu\rangle\langle \widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{y},\,\boldsymbol{e}_\mu\rangle - \sum_{\mu=1}^{n}\langle \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{y},\,\boldsymbol{e}_\mu\rangle\langle \widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{x},\,\boldsymbol{e}_\mu\rangle$$

$$= \langle \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{x},\,\widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{y}\rangle - \langle \widehat{\boldsymbol{S}}_{\boldsymbol{n}}\,\boldsymbol{y},\,\widehat{\boldsymbol{S}}_{\boldsymbol{m}}\,\boldsymbol{x}\rangle.$$

for any orthonormal basis $\{\boldsymbol{e}_\mu\}_{\mu=1}^{n}$ of $T_p\mathcal{M}$. □

5.2 Cylindrical Domains

In this section we compute the integral invariants of the cylindrical domain around a point on an n-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^{n+k}$. In the case 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 behaviour of the eigenvalues of the covariance matrix. Once the cylinder is fixed to be normal to this tangent space, the integral invariants can be computed to next-to-leading order to see how they encode the geometric information of the third fundamental form.

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 of the defining function. 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 5.2.1. *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 $T_p\mathcal{M}$:*

$$V(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = V_n(1) \prod_{\mu=1}^{n} \ell_\mu + \mathcal{O}(\varepsilon^{n+1}), \tag{5.15}$$

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

$$V(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon) \left[1 + \frac{\varepsilon^2}{2(n+2)} \operatorname{tr} \mathbf{III} + \mathcal{O}(\varepsilon^4) \right] \tag{5.16}$$

where $\operatorname{tr} \mathbf{III} = \|\boldsymbol{H}\|^2 - \mathcal{R}$.

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 of $T_p\mathcal{M} \oplus N_p\mathcal{M}$, 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 orthornomal 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 $\|\mathrm{proj}_{\mathbb{V}}(\boldsymbol{X})\| \leqslant \varepsilon$, 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

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

This is a quadratic equation that can be written as

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

where $\boldsymbol{y} = A^{T}\boldsymbol{x}$. The matrix $[A \cdot A^{T}]$ is positive definite since it is clearly nonnegative, 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:

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

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

$$\mathrm{proj}_{T_{p}\mathcal{M}}(\boldsymbol{X}) = \|\mathrm{proj}_{T_{p}\mathcal{M}}([\boldsymbol{x}, f^{1}(\boldsymbol{x}), \ldots, f^{k}(\boldsymbol{x})]^{T})\| = \|\boldsymbol{x}\| \leqslant \varepsilon,$$

thus, we are integrating $\sqrt{\det g(\boldsymbol{x})}$ over the ball $B_{p}^{(n)}(\varepsilon) \subset T_{p}\mathcal{M}$, which can be computed using the integrals from the appendix A:

$$\begin{aligned}
V(\mathrm{Cyl}_{p}(\varepsilon)) &= \int_{\mathbb{S}^{n-1}} d\mathbb{S} \int_{0}^{\varepsilon} \rho^{n-1}\left(1 + \frac{1}{2}\sum_{i=1}^{k}\sum_{\alpha=1}^{n}\left[\sum_{\beta=1}^{n} \kappa_{\alpha\beta}^{i} \rho\, \overline{x}^{\beta}\right]^{2} + \mathcal{O}(x^{3})\right) d\rho \\
&= V_{n}(\varepsilon) + \frac{\varepsilon^{n+2}}{2(n+2)} \sum_{i=1}^{k}\sum_{\alpha=1}^{n}\sum_{\beta,\gamma}^{n} \kappa_{\alpha\beta}^{i}\kappa_{\alpha\gamma}^{i} \int_{\mathbb{S}^{n-1}} \overline{x}^{\beta}\overline{x}^{\gamma}\, d\mathbb{S} + \mathcal{O}(\varepsilon^{n+4}) \\
&= V_{n}(\varepsilon) + \frac{C_{2}\,\varepsilon^{n+2}}{2(n+2)} \sum_{i=1}^{k}\sum_{\alpha,\beta}^{n} (\kappa_{\alpha\beta}^{i})^{2} + \mathcal{O}(\varepsilon^{n+4}) \\
&= V_{n}(\varepsilon) + \frac{V_{n}(\varepsilon)\,\varepsilon^{2}}{2(n+2)} \sum_{\alpha,\beta}^{n} \langle \mathbf{II}(\boldsymbol{e}_{\alpha}, \boldsymbol{e}_{\beta}), \mathbf{II}(\boldsymbol{e}_{\alpha}, \boldsymbol{e}_{\beta}) \rangle + \mathcal{O}(\varepsilon^{n+4}).
\end{aligned}$$

Here the spherical integral is only nonzero when $\beta = \gamma$, and the last term is the component expression of equation 5.8. $\square$

Proposition 5.2.2. *The barycenter of the cylindrical component, for $\mathbb{V}$ as in the previous theorem, is*

$$\boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \mathbf{0} + \mathcal{O}(\varepsilon^2). \tag{5.17}$$

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

$$\boldsymbol{s}(\mathrm{Cyl}_p(\varepsilon)) = [\ \mathbf{0},\ \frac{\varepsilon^2}{2(n+2)}\ \boldsymbol{H}\ ]^T + \mathcal{O}(\varepsilon^4). \tag{5.18}$$

Proof. For generic $\mathbb{V}$, approximating the manifold again by its tangent space, $\boldsymbol{X} = [\boldsymbol{x}, \mathbf{0} + \mathcal{O}(\varepsilon^2)]^T$, the normal component does not contribute until order two and the tangent component also vanishes at order one 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. On the other hand the normal components integrate as:

$$V[\boldsymbol{s}]^j = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^\varepsilon f^j \sqrt{\det g}\rho^{n-1}\, d\rho = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^\varepsilon \rho^{n-1} \left(\frac{1}{2} \sum_{\alpha,\,\beta=1}^{n} \kappa^j_{\alpha\,\beta} \rho^2 \overline{x}^\alpha \overline{x}^\beta + \mathcal{O}(x^3) \right) d\rho$$

$$= \frac{\varepsilon^{n+2}}{2(n+2)} \sum_{\alpha,\beta=1}^{n} \kappa^j_{\alpha\beta} \int_{\mathbb{S}^{n-1}} \overline{x}^\alpha \overline{x}^\beta d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+4}) = \frac{C_2\,\varepsilon^{n+2}}{2(n+2)}\, H^j + \mathcal{O}(\varepsilon^{n+4}),$$

Dividing by $V = V(\mathrm{Cyl}_p(\varepsilon))$ 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 5.2.3. *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 $\mathbb{R}^{n+k}$. Then the series of eigenvectors of $C(\varepsilon)$ form an orthonormal basis of $\mathbb{R}^{n+k}$ that converges for $\varepsilon \to 0$. The first n eigenvalues are $\lambda_\mu(\varepsilon) = a\varepsilon^2 + \lambda_\mu^{(4)}\varepsilon^4 + \mathcal{O}(\varepsilon^5)$, where $\lambda_\mu^{(4)}$ and the corresponding limit eigenvectors $\{\boldsymbol{V}_\mu^{(0)}\}_{\mu=1}^n$ satisfy the eigenvalue decomposition of A:

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

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

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

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 [52]:

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

$$= (\lambda^{(1)}\varepsilon^1 + \lambda^{(2)}\varepsilon^2 + \lambda^{(3)}\varepsilon^3 + \lambda^{(4)}\varepsilon^4 + \ldots)[\,\boldsymbol{V}^{(0)} + \boldsymbol{V}^{(1)}\varepsilon + \boldsymbol{V}^{(2)}\varepsilon^2 + \ldots].$$

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 [52, 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, the vectors $\boldsymbol{V}_i^{(0)} = \lim_{\varepsilon\to 0} \boldsymbol{V}_i(\varepsilon)$ are a unique orthonormal basis of $\mathbb{R}^{n+k}$ that is completely 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 $i = 1, \ldots, n$, and $\lambda_i^{(2)} = 0$, for $i = n+1, \ldots, n+k$, 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 $i = 1, \ldots, n$, 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, $i = 1, \ldots, n$, 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 $i = n+1, \ldots, n+k$, 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. $\square$

The previous lemma is a fundamental step to establish the main theorem of this and the next section, along with the special case of hypersurfaces in §4.2.

Theorem 5.2.4. *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 $\mathrm{Cyl}_p(\varepsilon, \mathbb{V})$ is finite, the covariance matrix $C_p(\varepsilon, \mathbb{V})$ has limit eigenvectors that span $T_p\mathcal{M}$ those corresponding to the first n eigenvalues, which scale as ε^2. The other k eigenvalues scaling at higher order have limit eigenvectors that span $N_p\mathcal{M}$:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = \frac{\varepsilon^2}{n+2}\ell_\mu^2\, V_n(1) \prod_{\alpha=1}^{n} \ell_\alpha + \mathcal{O}(\varepsilon^{n+3}), \qquad \mu = 1, \dots, n, \tag{5.19}$$

$$\lambda_j(\mathrm{Cyl}_p(\varepsilon, \mathbb{V})) = 0 + \mathcal{O}(\varepsilon^{n+3}), \qquad j = n+1, \dots, n+k, \tag{5.20}$$

where ℓ_μ are the principal lengths *of the ellipsoid in 5.2.1. 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 bilinar form with respect to the metric. Then the eigenvalues of the covariance matrix of the cylindrical component are:*

$$\lambda_\mu(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{2(n+2)(n+4)}\lambda_\mu\big[\,(\mathrm{tr}\,\mathbf{III})\,\mathrm{Id}_n + 2\,\mathrm{tr}_\perp\mathbf{III}\,\big] + \mathcal{O}(\varepsilon^6)\right] \tag{5.21}$$

$$\lambda_j(\mathrm{Cyl}_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^4}{4(n+2)(n+4)}\lambda_j\big[\,\boldsymbol{H}\otimes\boldsymbol{H} + 2\,\mathrm{tr}_\parallel\mathbf{III}\,\big] + \mathcal{O}(\varepsilon^6)\right] \tag{5.22}$$

for all $\mu = 1, \dots, n$, and $j = n+1, \dots, n+k$. 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 a ball:

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

Thus, the covariance matrix leading term is proportional to $\mathrm{diag}(\ell_1^2,\dots,\ell_n^2,0,\dots,0)$, 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 an straightforward extension to lemma 5.2.3 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:

$$
\begin{aligned}
[C(\mathrm{Cyl}_p(\varepsilon))]^{\mu\nu} &= \int_{B^{(n)}(\varepsilon)} x^\mu x^\nu \sqrt{\det g(\boldsymbol{x})}\, d^n\boldsymbol{x} \\
&= \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^\varepsilon \rho^{n+1}\overline{x}^\mu\overline{x}^\nu \left(1 + \frac{1}{2}\sum_{i=1}^k\sum_{\alpha=1}^n\left[\sum_{\beta=1}^n \kappa^j_{\alpha\beta}\rho\,\overline{x}^\beta\right]^2 + \mathcal{O}(x^3)\right) d\rho \\
&= \frac{\varepsilon^{n+2}}{n+2}\int_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu d\,\mathbb{S} + \frac{\varepsilon^{n+4}}{2(n+4)}\sum_{i=1}^k\sum_{\alpha=1}^n\sum_{\beta,\gamma}^n \kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}\int_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu\overline{x}^\beta\overline{x}^\gamma d\,\mathbb{S} + \mathcal{O}(\varepsilon^{n+6}),
\end{aligned}
$$

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

$$
\int_{\mathbb{S}^{n-1}} \overline{x}^\mu\overline{x}^\nu\overline{x}^\beta\overline{x}^\gamma \, d\,\mathbb{S} = C_4(\mu\nu\beta\gamma) + C_{22}\left[(\mu\nu\beta\gamma) + (\mu\nu\beta\gamma) + (\mu\nu\beta\gamma)\right]. \tag{5.23}
$$

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}:

$$\frac{\delta_{\mu\nu} C_2 \varepsilon^{n+2}}{n+2} + \frac{C_2 \varepsilon^{n+4}}{2(n+2)(n+4)} \sum_{i=1}^{k} \left[3\delta_{\mu\nu} \sum_{\alpha=1}^{n} (\kappa^i_{\alpha\mu})^2 + \delta_{\mu\nu} \sum_{\substack{\alpha,\beta \\ \beta\neq\mu}}^{n} (\kappa^i_{\alpha\beta})^2 + 2(1-\delta_{\mu\nu}) \sum_{\alpha=1}^{n} \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} \right] + \ldots$$

$$= \frac{V_n(\varepsilon)\varepsilon^2}{n+2}\delta_{\mu\nu} + \frac{V_n(\varepsilon)\varepsilon^4}{2(n+2)(n+4)} \left[\delta_{\mu\nu} \sum_{i=1}^{k} \sum_{\alpha,\beta}^{n} (\kappa^i_{\alpha\beta})^2 + 2 \sum_{i=1}^{k} \sum_{\alpha=1}^{n} \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} \right] + \mathcal{O}(\varepsilon^{n+6})$$

$$= \frac{V_n(\varepsilon)\varepsilon^2}{n+2}\delta_{\mu\nu} + \frac{V_n(\varepsilon)\varepsilon^4}{2(n+2)(n+4)} \left[\delta_{\mu\nu} \sum_{\alpha,\beta}^{n} \|\mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta)\|^2 + 2 \sum_{\alpha=1}^{n} \langle\, \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\mu),\, \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\nu)\, \rangle \right] + \ldots$$

The component expression of equations 5.9 and 5.11 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 5.2.3 the next-to-leading order contribution to the first n eigenvalues of $C(\mathrm{Cyl}_p(\varepsilon))$, and whose eigenvectors are the limit eigenvectors of $C(\mathrm{Cyl}_p(\varepsilon))$.

We perform now the integration of the normal block, which truncated to leading order yields:

$$[C(\mathrm{Cyl}_p(\varepsilon))]^{ij} \approx \int_{B^{(n)}(\varepsilon)} f^i f^j d^n\boldsymbol{x} = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{\varepsilon} \frac{\rho^{n+3}}{4} d\rho \sum_{\alpha,\beta}^{n} \sum_{\gamma,\delta}^{n} \kappa^i_{\alpha\beta}\kappa^j_{\gamma\delta}\overline{x}^\alpha\overline{x}^\beta\overline{x}^\gamma\overline{x}^\delta + \mathcal{O}(\varepsilon^{n+6}),$$

where the angular integral is only nonzero in the same cases as in equation 5.23 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:

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

$$= \frac{C_2\, \varepsilon^{n+4}}{4(n+2)(n+4)} \left[3 \sum_{\alpha=1}^{n} \mathrm{II}^i(\boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha)\mathrm{II}^j(\boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha) + \sum_{\substack{\alpha,\gamma \\ \alpha\neq\gamma}}^{n} \mathrm{II}^i(\boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha)\mathrm{II}^j(\boldsymbol{e}_\gamma, \boldsymbol{e}_\gamma) + \right.$$

$$\left. +2 \sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}}^{n} \mathrm{II}^i(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta)\mathrm{II}^j(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) \right] + \mathcal{O}(\varepsilon^{n+6}),$$

in which the first sum precisely completes the elements missing from the other two

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

In this last expression we clearly identify the components $[\boldsymbol{H}\otimes\boldsymbol{H}]^{ij}$ and those of $2\,\mathrm{tr}_{\parallel}\mathbf{III}$ using the definition of $\boldsymbol{H}$ and equation 5.10. □

We shall see below that the spherical covariance matrix has the same normal eigenvalues, to leading order, as the cylindrical case above. In [55, 56] 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 5.2.5. *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 $[\,\boldsymbol{H}\otimes\boldsymbol{H}+2\,\mathrm{tr}_{\parallel}\mathbf{III}\,]$, integrate to:*

$$\frac{1}{V_n(\varepsilon)}\int_{B^{(n)}(\varepsilon)}\kappa^2(\gamma,\boldsymbol{x},\boldsymbol{n}_j)\,d^n\boldsymbol{x}=\frac{\varepsilon^4}{(n+2)(n+4)}\lambda_j[\,\boldsymbol{H}\otimes\boldsymbol{H}+2\,\mathrm{tr}_{\parallel}\mathbf{III}\,]. \tag{5.24}$$

In particular:

$$\sum_{j=1}^{k}\frac{1}{V_n(\varepsilon)}\int_{B^{(n)}(\varepsilon)}\kappa^2(\gamma,\boldsymbol{x},\boldsymbol{n}_j)\,d^n\boldsymbol{x}=\frac{3\|\boldsymbol{H}\|^2-2\mathcal{R}}{(n+2)(n+4)}\varepsilon^4. \tag{5.25}$$

5.3 Spherical Domains

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 $\|\mathrm{proj}_{T_p\mathcal{M}}([\boldsymbol{x},\boldsymbol{f}(\boldsymbol{x})]^T)\|=\|\boldsymbol{x}\|\leqslant\varepsilon$, and thus the integration region is a perfect ball. However, in the spherical case the domain of integration is $\|\boldsymbol{x}\|^2+\|\boldsymbol{f}(\boldsymbol{x})\|^2\leqslant\varepsilon^2$, which is nontrivial and in general cannot be parametrized exactly. One can nevertheless generalize the same procedure done originally for surfaces in space [51] to find the leading order corrections to the ball domain.

Lemma 5.3.1. *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 $\partial D_p(\varepsilon) = \mathcal{M} \cap \mathbb{S}^n_p(\varepsilon)$, is*

$$r(\overline{\boldsymbol{x}}) := \rho(\overline{x}_1, \ldots, \overline{x}_n) = \varepsilon - \frac{K(\overline{\boldsymbol{x}})^2}{8}\varepsilon^3 + \mathcal{O}(\varepsilon^4), \tag{5.26}$$

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

$$K(\overline{\boldsymbol{x}})^2 := \|\mathbf{II}(\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}})\|^2 = \sum_{i=1}^{k}\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n} \kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}\, \overline{x}^\alpha\overline{x}^\beta\overline{x}^\gamma\overline{x}^\delta \tag{5.27}$$

is the square of the ambient space acceleration of a geodesic curve of $\mathcal{M}$ with tangent $\overline{\boldsymbol{x}}$ at p.

Proof. A point of the spherical boundary satisfies $\|\boldsymbol{x}\|^2 + \sum_{i=1}^k (f^i(\boldsymbol{x}))^2 = \varepsilon^2$. Since $\|\boldsymbol{x}\|^2 = \rho^2$, and $f^i(\boldsymbol{x}) = \frac{1}{2}\sum_{\alpha,\beta}^n \kappa^i_{\alpha\beta}x^\alpha x^\beta + \mathcal{O}(x^3)$, it is immediate that

$$\rho^2 + \frac{1}{4}\rho^4 \sum_{i=1}^{k}\left(\sum_{\alpha,\beta}^{n}\kappa^i_{\alpha\beta}\overline{x}^\alpha\overline{x}^\beta\right)^2 - \varepsilon^2 = \mathcal{O}(\rho^5).$$

Defining $K(\overline{\boldsymbol{x}})^2$ as the coefficient of $\frac{\rho^4}{4}$, we can solve the equation to order four to get

$$\rho^2 = \frac{2}{K(\overline{\boldsymbol{x}})^2}\left(-1 + \sqrt{1 + K(\overline{\boldsymbol{x}})^2\varepsilon^2}\right) = \varepsilon^2 - \frac{1}{4}K(\overline{\boldsymbol{x}})^2\varepsilon^4 + \mathcal{O}(\varepsilon^6),$$

whose square root yields the result. Note that the actual error may be of order four because this could contribute at order five upon squaring the expression, which is the order neglected in the original equation. In our chosen orthonormal basis at p, we have that $\mathbf{II}(\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}}) = \sum_i \sum_{\alpha,\beta} \kappa^i_{\alpha\beta}\overline{x}^\alpha\overline{x}^\beta\boldsymbol{n}_i$, and this is precisely the ambient space acceleration of a geodesic of $\mathcal{M}$, cf. [46, ch. 4, Cor. 10]. □

Proposition 5.3.2. *The n-dimensional volume of the spherical component is*

$$V(D_p(\varepsilon)) = V_n(\varepsilon)\left[1 + \frac{\varepsilon^2}{8(n+2)}\left(2\operatorname{tr}\mathbf{III} - \|\boldsymbol{H}\|^2\right) + \mathcal{O}(\varepsilon^3)\right] \tag{5.28}$$

where $2\operatorname{tr}\mathbf{III} - \|\boldsymbol{H}\|^2 = \|\boldsymbol{H}\|^2 - 2\mathcal{R}$*.*

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

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

where the integral is only nonzero as in equation 5.23, so

$$\begin{aligned} &= V_n(\varepsilon) - \frac{C_2\,\varepsilon^{n+2}}{8(n+2)}\sum_{i=1}^{k}\left[3\sum_{a=1}^{n}(\kappa^i_{\alpha\alpha})^2 + \sum_{\substack{\alpha,\gamma\\ \alpha\neq\gamma}}^{n}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} + 2\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}^{n}(\kappa^i_{\alpha\beta})^2\right] + \frac{V_n(\varepsilon)\,\varepsilon^2}{2(n+2)}\operatorname{tr}\mathbf{III} + \mathcal{O}(\varepsilon^{n+3}) \\ &= V_n(\varepsilon)\left[1 + \frac{\varepsilon^2}{8(n+2)}\left(4\operatorname{tr}\mathbf{III} - \sum_{i=1}^{k}\sum_{\alpha,\gamma}^{n}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{i=1}^{k}\sum_{\alpha,\beta}^{n}(\kappa^i_{\alpha\beta})^2\right) + \mathcal{O}(\varepsilon^3)\right] \end{aligned}$$

Now, the first set of sums in the braces is $\langle \sum_\alpha \mathbf{II}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha),\, \sum_\gamma \mathbf{II}(\boldsymbol{e}_\gamma, \boldsymbol{e}_\gamma)\rangle = \|\boldsymbol{H}\|^2$, and the second set is $\operatorname{tr}\mathbf{III}$. □

Remark 5.3.3. Notice that the dependence of the error generated by the irregular radius $r(\overline{\boldsymbol{x}})$ is not known, leaving $\mathcal{O}(\varepsilon^{n+3})$ in the previous proof, or 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 5.3.4. *The barycenter of the spherical component is to leading order the same as for the cylindrical component:*

$$\boldsymbol{s}(D_p(\varepsilon)) = [\,\mathbf{0},\ \frac{\varepsilon^2}{2(n+2)}\ \boldsymbol{H}\ ]^T + \mathcal{O}(\varepsilon^4). \tag{5.29}$$

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

In contrast to the hypersurface case, for arbitrary codimension the different osculating quadrics of $f^i(\boldsymbol{x})$, $i = 1, \ldots, k$, cannot be diagonalized simultaneously to a common basis in general. The number 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 is obtained.

Theorem 5.3.5. *Let $\lambda_l[\cdot]$ denote taking the l-th eigenvalue of a linear operator at p, or of its associated bilinar form with respect to the metric. Then the eigenvalues of the covariance matrix of the spherical component are:*

$$\lambda_\mu(D_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^2}{n+2} + \frac{\varepsilon^4}{8(n+2)(n+4)}\lambda_\mu[\,(2\,\mathrm{tr}\,\mathbf{III} - \|\boldsymbol{H}\|^2)\mathrm{Id}_n - 4\widehat{\boldsymbol{S}}_{\boldsymbol{H}}\,] + \ \mathcal{O}(\varepsilon^5)\right] \tag{5.30}$$

$$\lambda_j(D_p(\varepsilon)) = V_n(\varepsilon)\left[\frac{\varepsilon^4}{2(n+2)(n+4)}\lambda_j[\,\mathrm{tr}_{\,\|}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}\,] + \mathcal{O}(\varepsilon^6)\right] \tag{5.31}$$

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 Weingarten operator at $\boldsymbol{H}$, i.e., $\widehat{\boldsymbol{S}}_{\boldsymbol{H}}$, and the last k eigenvectors to those of $[\mathrm{tr}_{\,\|}\mathbf{III} - \frac{1}{n+2}\boldsymbol{H}\otimes\boldsymbol{H}]$.

Proof. From lemma 5.2.3 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,

$$C(D_p(\varepsilon)) = \int_{D_p(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{X}\ \mathrm{dVol} - \int_{D_p(\varepsilon)} \boldsymbol{X}\otimes\boldsymbol{s}\ \mathrm{dVol},$$

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

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

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:

$$\frac{V_n(\varepsilon)\varepsilon^4}{4(n+2)(n+4)}(\boldsymbol{H}\otimes\boldsymbol{H}+2\,\mathrm{tr}_{\parallel}\mathbf{III}) - \frac{V_n(\varepsilon)\varepsilon^4}{4(n+2)^2}\boldsymbol{H}\otimes\boldsymbol{H} = \frac{V_n(\varepsilon)\varepsilon^4}{2(n+2)(n+4)}[\mathrm{tr}_{\parallel}\mathbf{III} - \frac{\boldsymbol{H}\otimes\boldsymbol{H}}{n+2}].$$

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

$$[C(D_p(\varepsilon))]^{\mu\nu} = \int_{\mathbb{S}^{n-1}} d\,\mathbb{S} \int_0^{r(\overline{\boldsymbol{x}})} \rho^{n+1}\overline{x}^{\mu}\overline{x}^{\nu}\left(1+\frac{1}{2}\sum_{i=1}^{k}\sum_{\alpha=1}^{n}\left[\sum_{\beta=1}^{n}\kappa^i_{\alpha\beta}\rho\,\overline{x}^{\beta}\right]^2 + \mathcal{O}(x^3)\right)d\rho$$

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

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

$$= \delta_{\mu\nu}\frac{V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{\varepsilon^{n+4}}{2(n+4)}\sum_{i=1}^{k}\left[\sum_{\alpha,\beta,\gamma}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\alpha\gamma}C_{(\mu\nu\beta\gamma)} - \frac{n+4}{4}\sum_{\alpha,\beta}^{n}\sum_{\gamma,\delta}^{n}\kappa^i_{\alpha\beta}\kappa^i_{\gamma\delta}C_{(\mu\nu\alpha\beta\gamma\delta)}\right] + \mathcal{O}(\varepsilon^{n+5}),$$

where we have made use of equation 5.27, 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 coordinate components). The first summation simplifies again with equation 5.23 to yield the cylindrical tangent block, but the other set of sums comprises the 31 spherical integrals of all possible monomials of degree six:

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

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

$$+ (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta)$$

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

$$+ (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta)$$

$$+ (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta) + (\mu\nu\alpha\beta\gamma\delta)\Big]$$

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:

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

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(\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 and using $\kappa^i_{\alpha\beta} = \kappa^i_{\beta\alpha}$, so they can be gathered into common factors:

$$
\begin{aligned}
[C(D_p(\varepsilon))]^{\mu\nu} = \delta_{\mu\nu}\frac{V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{V_n(\varepsilon)\varepsilon^4}{8(n+2)(n+4)} \sum_i \Bigg[& 4\delta_{\mu\nu} \sum_{\alpha,\beta} (\kappa^i_{\alpha\beta})^2 + 8\sum_\alpha \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} - 15\delta_{\mu\nu}(\kappa^i_{\nu\nu})^2 \\
& - 3\delta_{\mu\nu} \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\alpha})^2 - 12(1-\delta_{\mu\nu})\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu} + \kappa^i_{\nu\nu}) - 6\delta_{\mu\nu} \sum_{\alpha\neq\mu} \kappa^i_{\alpha\alpha}\kappa^i_{\nu\nu} - 12\delta_{\mu\nu} \sum_{\alpha\neq\mu} (\kappa^i_{\alpha\nu})^2 \\
& -\delta_{\mu\nu} \sum_{\alpha\neq\mu} \sum_{\gamma\neq\alpha,\mu} \kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\delta_{\mu\nu} \sum_{\alpha\neq\mu} \sum_{\beta\neq\alpha,\mu} (\kappa^i_{\alpha\beta})^2 - (1-\delta_{\mu\nu}) \Bigg(4\kappa^i_{\mu\nu} \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\alpha} + 8 \sum_{\alpha\neq\mu,\nu} \kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} \Bigg) \Bigg] + \ldots
\end{aligned}
$$

for which regrouping terms and completing some sums will clarify the simplifications below,

$$= \delta_{\mu\nu}\frac{V_n(\varepsilon)\varepsilon^2}{n+2} + \frac{V_n(\varepsilon)\varepsilon^4}{8(n+2)(n+4)}\sum_i\Bigg[8\sum_\alpha \kappa^i_{\alpha\mu}\kappa^i_{\alpha\nu} - 12\kappa^i_{\mu\nu}(\kappa^i_{\mu\mu}+\kappa^i_{\nu\nu}) - 4\kappa^i_{\mu\nu}\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\alpha}$$

$$- 8\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} + \delta_{\mu\nu}\Bigg\{4\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2 - 3\sum_{\alpha\neq\mu}(\kappa^i_{\alpha\alpha})^2 + 21(\kappa^i_{\mu\mu})^2 - 2\kappa^i_{\mu\mu}\sum_{\alpha\neq\mu}\kappa^i_{\alpha\alpha} - 12\sum_\alpha(\kappa^i_{\alpha\mu})^2$$

$$- \sum_{\alpha\neq\mu}\sum_{\gamma\neq\alpha,\mu}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{\alpha\neq\mu}\sum_{\beta\neq\alpha,\mu}(\kappa^i_{\alpha\beta})^2 + 8\sum_{\alpha\neq\mu}(\kappa^i_{\alpha\mu})^2\Bigg\}\Bigg] + \mathcal{O}(\varepsilon^{n+5}).$$

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:

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

$$-4\kappa^i_{\mu\nu}\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\alpha} - 8\sum_{\alpha\neq\mu,\nu}\kappa^i_{\alpha\mu}\kappa^i_{\nu\alpha} + \delta_{\mu\nu}\Bigg\{4\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2 + 12(\kappa^i_{\mu\mu})^2 - \sum_{\alpha,\,\gamma}\kappa^i_{\alpha\alpha}\kappa^i_{\gamma\gamma} - 2\sum_{\alpha,\,\beta}(\kappa^i_{\alpha\beta})^2\Bigg\}\Bigg] + \ldots$$

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:

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

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 matrix of the Weingarten operator at the mean curvature, plus a constant, in the orthonormal basis chosen. The error is $\mathcal{O}(\varepsilon^{n+5})$. □

The theorems of this chapter provide the most general relationship known between PCA covariance matrices and the local curvature of submanifolds. The geometric role played by the generalized third fundamental form is thus uncovered via its appearance in the asymptotic series of the eigenvalue decomposition, which completely justifies the importance of this tensor as an independent object of interest to describe curvature from an integral invariant point of view.

Chapter 6

Descriptors at Scale for Manifold Learning

By solving the second order term from the series expansion of our integral invariants, we can extract the curvature information they encode and write it in terms of the volume and eigenvalues at a fixed scale. The integral invariants can be computed without a priori knowledge of the manifold geometry, which implies that these local statistical measurements of the underlying point set provide descriptors of the local differential geometry of the manifold, e.g., approximated from a cloud of points.

The limit formula for the ratio of the eigenvalues in the case of curves was the major result of chapter 3, establishing a direct relationship between the local covariance analysis and the Frenet-Serret curvature information. The two main theorems 5.2.4 and 5.3.5 generalize this type of result to general submanifolds by directly taking the limits of the covariance matrix eigenvalues.

Corollary 6.0.1. *Writing $\lambda_\mu(p,\varepsilon)$ for the tangent eigenvalues of the cylindrical covariance matrix $C(\mathrm{Cyl}_p(\varepsilon))$, they satisfy the asymptotic ratio*

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\lambda_\mu(p,\varepsilon)-\lambda_\nu(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{n+4}\left(\,\lambda_\mu[\mathrm{tr}_\perp\mathbf{III}]-\lambda_\nu[\mathrm{tr}_\perp\mathbf{III}]\,\right), \tag{6.1}$$

and the normal eigenvalues satisfy

$$\lim_{\varepsilon\to 0}\frac{V_n(\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)}\sum_{j=n+1}^{n+k}\lambda_j(p,\varepsilon) = \frac{n+2}{4(n+4)}\left(\,\|\boldsymbol{H}\|^2+2\,\mathrm{tr}\,\mathbf{III}\right), \tag{6.2}$$

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, $C_p(D_p(\varepsilon))$, then the corresponding limits are

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\widetilde{\lambda}_\mu(p,\varepsilon)-\widetilde{\lambda}_\nu(p,\varepsilon)}{\widetilde{\lambda}_\mu(p,\varepsilon)\widetilde{\lambda}_\nu(p,\varepsilon)} = \frac{n+2}{2(n+4)}\left(\widetilde{\lambda}_\nu[\widehat{\boldsymbol{S}}_{\boldsymbol{H}}]-\widetilde{\lambda}_\mu[\widehat{\boldsymbol{S}}_{\boldsymbol{H}}]\right), \tag{6.3}$$

and

$$\lim_{\varepsilon\to 0} \frac{V_n(\varepsilon)}{\widetilde{\lambda}_\mu(p,\varepsilon)\widetilde{\lambda}_\nu(p,\varepsilon)} \sum_{j=n+1}^{n+k} \widetilde{\lambda}_j(p,\varepsilon) = \frac{n+2}{2(n+4)} \left(\operatorname{tr} \mathbf{III} - \frac{1}{n+2} \|\boldsymbol{H}\|^2 \right). \tag{6.4}$$

These ratios can be used at fixed $\varepsilon > 0$ to obtain estimators of the eigenvalues and eigenvectors of the third fundamental form and the Weingarten map at the mean curvature.

6.1 Spherical Component Descriptors

Now we focus on smooth hypersurfaces in $\mathbb{R}^{n+1}$ since our integral invariants furnish descriptors at scale of the principal curvatures, and the principal and normal directions which, by lemma 2.2.4, are sufficient to construct descriptors for an embedded Riemannian manifold of general codimension. Employing the asymptotic series of section §4.1, we solve for the principal curvatures in terms of the eigenvalues. In this case no sign choices are needed. The eigenvectors generically converge to the principal and normal directions, so at fixed $\varepsilon > 0$ they provide approximations to the Darboux frame at every point.

Corollary 6.1.1. *Abbreviating the integral invariants of the spherical component as* $\lambda_\mu(p,\varepsilon) \equiv \lambda_\mu(V_p^+(\varepsilon)), V_p(\varepsilon) \equiv V(V_p^+(\varepsilon))$*, then the corresponding descriptors of the principal curvatures, at scale* $\varepsilon > 0$ *and point* $p \in \mathcal{S}$*, are given by*

$$\kappa_\mu(V_p^+(\varepsilon)) = \frac{n+4}{\varepsilon^4 V_n(\varepsilon)} \left[\frac{\varepsilon^2 V_{n+1}(\varepsilon)}{n+3} - (n+1)\lambda_\mu(p,\varepsilon) + \sum_{\alpha\neq\mu}^{n} \lambda_\alpha(p,\varepsilon) \right], \tag{6.5}$$

or equivalently by

$$H(V_p^+(\varepsilon)) = \frac{(n+2)V_{n+1}(\varepsilon)}{\varepsilon^2 V_n(\varepsilon)} \left(1 - 2\frac{V_p(\varepsilon)}{V_{n+1}(\varepsilon)} \right), \tag{6.6}$$

$$\kappa_\mu(V_p^+(\varepsilon)) = \frac{(n+2)(n+4)}{\varepsilon^4 V_n(\varepsilon)} \left(\frac{\varepsilon^2 V_{n+1}(\varepsilon)}{2(n+3)} - \lambda_\mu(p,\varepsilon) \right) + \frac{1}{2} H(V_p^+(\varepsilon)). \tag{6.7}$$

The truncation errors are $|H(p) - H(V_p^+(\varepsilon))| \leqslant \mathcal{O}(\varepsilon)$, *and* $|\kappa_\mu(p) - \kappa_\mu(V_p^+(\varepsilon))| \leqslant \mathcal{O}(\varepsilon)$, *for any* $\mu = 1, \ldots, n$. *The eigenvectors* $\boldsymbol{e}_\mu(V_p^+(\varepsilon))$ *and* $\boldsymbol{e}_{n+1}(V_p^+(\varepsilon))$ *are descriptors of the principal and normal directions respectively.*

Proof. Let us define the coefficients at scale

$$a = \frac{\varepsilon^2 \, V_{n+1}(\varepsilon)}{2(n+3)}, \qquad b = -\frac{\varepsilon^4 V_n(\varepsilon)}{2(n+2)(n+4)},$$

then the tangent eigenvalues from equation 4.4 solve the principal curvatures

$$\kappa_\mu = \frac{\lambda_\mu - a}{2b} - \frac{1}{2} H + \mathcal{O}(\varepsilon).$$

Fixing one $\mu = 1, \ldots, n$, and subtracting any two such equations with $\mu \neq \alpha$ results in

$$\kappa_\alpha = \frac{\lambda_\alpha - \lambda_\mu}{2b} + \kappa_\mu + \mathcal{O}(\varepsilon),$$

inserting this into the definition of H one gets

$$H = n\kappa_\mu + \sum_{\alpha \neq \mu}^{n} \frac{\lambda_\alpha - \lambda_\mu}{2b} + \mathcal{O}(\varepsilon),$$

which substituting back leaves

$$\kappa_\mu(V_p^+(\varepsilon)) = \frac{\lambda_\mu - a}{b(n+2)} - \sum_{\alpha \neq \mu}^{n} \frac{\lambda_\alpha - \lambda_\mu}{2b(n+2)} = \frac{1}{2b(n+2)} \left(-2a + (n+1)\lambda_\mu - \sum_{\alpha \neq \mu}^{n} \lambda_\alpha \right).$$

The truncation error is given by the order of $\mathcal{O}(\varepsilon^{n+5})/b \sim \mathcal{O}(\varepsilon)$. Alternatively, one can solve the Hulin-Troyanov relation 4.2 to obtain a descriptor of H, and then use this in the expression of κ_μ in terms of λ_μ and H above. $\square$

The asymptotic relations of corollary 6.0.1 reduce to very simple formulas in the case of hypersurfaces, relating the ratios of differences and products of eigenvalues to the principal curvatures.

Corollary 6.1.2. *Let $p \in \mathcal{S}$ and consider the spherical component invariants. Then for any $\mu, \nu = 1, \ldots, n$, the first n eigenvalues $\lambda_\mu(p, \varepsilon) \equiv \lambda_\mu(V_p^+(\varepsilon))$ of the covariance matrix $C(V_p^+(\varepsilon))$ satisfy the following limit ratio:*

$$\lim_{\varepsilon \to 0^+} \frac{V_{n+1}^2(\varepsilon)}{V_n(\varepsilon)} \frac{\lambda_\mu(p,\varepsilon) - \lambda_\nu(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{4(n+3)^2}{(n+2)(n+4)} [\kappa_\nu(p) - \kappa_\mu(p)]. \tag{6.8}$$

These eigenvalue ratios can be used along with the volume formula to obtain other expressions for the descriptors of the principal curvatures.

6.2 Cylindrical and Spherical Patch Descriptors

An analogous inversion process can be carried out with the series expansions of section §4.2. However, the relation to the principal curvatures is now quadratic so sign choices are needed when taking roots, and thus truncation errors are worse than in the spherical component volume above, as expected from the explanation at the end of chapter 2.

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

Corollary 6.2.1. *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:*

$$\mathcal{R}(\mathrm{Cyl}_p(\varepsilon)) = \frac{2(n+2)}{\varepsilon^2} \left[\frac{2(n+4)}{\varepsilon^2} \frac{\lambda_{n+1}(p,\varepsilon)}{V_n(\varepsilon)} + 3\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right) \right] \tag{6.9}$$

$$H(\mathrm{Cyl}_p(\varepsilon)) = (\pm)\sqrt{\frac{2(n+2)}{\varepsilon^2} \left[\frac{2(n+4)}{\varepsilon^2} \frac{\lambda_{n+1}(p,\varepsilon)}{V_n(\varepsilon)} + 2\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right) \right]}, \tag{6.10}$$

$$\kappa_\mu^2(\mathrm{Cyl}_p(\varepsilon)) = \frac{n+2}{\varepsilon^2} \left[\frac{n+4}{\varepsilon^2} \left(\frac{\lambda_\mu(p,\varepsilon)}{V_n(\varepsilon)} - \frac{\varepsilon^2}{n+2} \right) - \frac{V_p(\varepsilon)}{V_n(\varepsilon)} + 1 \right], \tag{6.11}$$

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

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

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:

$$|H^2(p) - H^2(\mathrm{Cyl}_p(\varepsilon))| \leqslant \mathcal{O}(\varepsilon^2),$$

$$|\mathcal{R}(p) - \mathcal{R}(\mathrm{Cyl}_p(\varepsilon))| \leqslant \mathcal{O}(\varepsilon^2),$$

$$|\kappa_\mu^2(p) - \kappa_\mu^2(\mathrm{Cyl}_p(\varepsilon))| \leqslant \mathcal{O}(\varepsilon^2).$$

Proof. Solving for the next-to-leading order term in the volume formula 5.16, and for the normal eigenvalue in equation 5.22, 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 5.21, and using $A(\varepsilon) = \sum_\alpha \kappa_\alpha^2$, the last formula is obtained. □

The cylindrical asymptotic ratios are very similar to the spherical component ones but relate the difference of eigenvalues to the difference of the squared principal curvatures.

Corollary 6.2.2. *The tangent eigenvalues of the cylindrical covariance matrix $C(\mathrm{Cyl}_p(\varepsilon))$ satisfy*

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\lambda_\mu(p,\varepsilon) - \lambda_\nu(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{n+4}(\ \kappa_\mu^2(p) - \kappa_\nu^2(p)\ ), \tag{6.12}$$

and the normal eigenvalue has

$$\lim_{\varepsilon\to 0} V_n(\varepsilon)\frac{\lambda_{n+1}(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{4(n+4)}\left(\ 3H^2(p) - 2\mathcal{R}(p)\ \right), \tag{6.13}$$

for any $\mu, \nu = 1, \ldots, n$.

Now, the spherical patch domain descriptors below can be used to determine the relative signs of the principal curvatures, and the cylindrical descriptors can be used to estimate with higher precision their absolute value, since they are guaranteed a better error bound.

Corollary 6.2.3. *Denoting by* $\lambda(p,\varepsilon) \equiv \lambda(D_p(\varepsilon)), V_p(\varepsilon) \equiv V(D_p(\varepsilon))$ *the integral invariants of the spherical hypersurface patch domain, then the corresponding curvature descriptors at scale* $\varepsilon > 0$ *and point* $p \in \mathcal{S}$*, for any* $\mu = 1, \ldots, n$*, are*

$$\mathcal{R}(D_p^+(\varepsilon)) = 2(n+2)^2(n+4)\frac{\lambda_{n+1}(p,\varepsilon)}{n\,\varepsilon^4\,V_n(\varepsilon)} - \frac{8(n+1)(n+2)}{n\,\varepsilon^2}\left(\frac{V_p(\varepsilon)}{V_n(\varepsilon)} - 1\right), \tag{6.14}$$

$$H(D_p^+(\varepsilon)) = (\pm)\sqrt{4(n+2)^2(n+4)\frac{\lambda_{n+1}(p,\varepsilon)}{n\,\varepsilon^4 V_n(\varepsilon)} + \frac{8(n+2)^2}{n\,\varepsilon^2}\left(1 - \frac{V_p(\varepsilon)}{V_n(\varepsilon)}\right)}, \tag{6.15}$$

$$\kappa_\mu(D_p^+(\varepsilon)) = \frac{2(n+2)}{\varepsilon^2 H(D_p^+(\varepsilon))}\left[\frac{V_p(\varepsilon)}{V_n(\varepsilon)} + \frac{n+4}{\varepsilon^2}\left(\frac{\varepsilon^2}{n+2} - \frac{\lambda_\mu(p,\varepsilon)}{V_n(\varepsilon)}\right) - 1\right], \tag{6.16}$$

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 corresponding errors are*

$$|H^2(p) - H(D_p(\varepsilon))^2| \leqslant \mathcal{O}(\varepsilon),$$

$$|\mathcal{R}(p) - \mathcal{R}(D_p(\varepsilon))| \leqslant \mathcal{O}(\varepsilon),$$

$$|\kappa_\mu^2(p) - \kappa_\mu(D_p(\varepsilon))^2| \leqslant \mathcal{O}(\varepsilon).$$

Proof. By solving the second term in equations 4.14 and 4.17, let us define

$$A = \frac{8(n+2)}{\varepsilon^2}\left(\frac{V_p(\varepsilon)}{V_n(\varepsilon)} - 1\right) + \mathcal{O}(\varepsilon),$$

$$B = 2(n+2)(n+4)\frac{\lambda_{n+1}(p,\varepsilon)}{\varepsilon^4 V_n(\varepsilon)} + \mathcal{O}(\varepsilon),$$

so that we have the system of equations $A = H^2 - 2\mathcal{R},\ B = \frac{n+1}{n+2}H^2 - \mathcal{R}$ whose solution is

$$\mathcal{R} = \frac{1}{n}((n+2)B - (n+1)A),$$

$$H^2 = \frac{(n+2)}{n}(2B - A).$$

We can approximate the normal direction and orientation by using $\boldsymbol{e}_{n+1}(p,\varepsilon)$, and since the barycenter 4.15 has normal component with leading order in terms of H, their mutual projection can serve to fix the orientation and overall relative sign of all the principal curvatures. The principal curvatures themselves are then solved from eq. 4.16 substituting the value of H above, resulting in $\kappa_\mu = \dfrac{1}{4H}(A - \Gamma_\mu)$, where

$$\Gamma_\mu = \frac{8(n+2)(n+4)}{\varepsilon^4}\left(\frac{\lambda_\mu(p,\varepsilon)}{V_n(\varepsilon)} - \frac{\varepsilon^2}{n+2}\right) + \mathcal{O}(\varepsilon).$$

The errors follow straightforwardly by the truncation of $A,\ B,\ \Gamma_\mu$. □

For this patch domain the asymptotic ratios are very similar to the spherical component case but multiplied by the mean curvature.

Corollary 6.2.4. *Let $p \in \mathcal{S}$ and consider the hypersurface spherical domain invariants. Then for any $\mu,\nu = 1,\ldots,n$, the first n eigenvalues $\lambda_\mu(p,\varepsilon) \equiv \lambda_\mu(D_p(\varepsilon))$ of the covariance matrix $C(D_p(\varepsilon))$ satisfy the following limit ratio:*

$$\lim_{\varepsilon\to 0^+} V_n(\varepsilon)\frac{\lambda_\mu(p,\varepsilon) - \lambda_\nu(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{2(n+4)}[\kappa_\nu(p) - \kappa_\mu(p)]H(p), \tag{6.17}$$

and the last eigenvalue satisfies:

$$\lim_{\varepsilon\to 0^+} V_n(\varepsilon)\frac{\lambda_{n+1}(p,\varepsilon)}{\lambda_\mu(p,\varepsilon)\lambda_\nu(p,\varepsilon)} = \frac{n+2}{2(n+4)}\left[\frac{n+1}{n+2}H^2(p) - \mathcal{R}(p)\right]. \tag{6.18}$$

The potential benefits of employing the spherical component descriptors lie in the fact that the domain they are computed from is an $(n+1)$-dimensional volume in $\mathbb{R}^{n+1}$, whereas the patch descriptors below are n-dimensional areas of hypersurfaces which in fact are part of the boundary of the aforementioned volume. This makes it reasonable to expect a higher robustness and stability with respect to noise, since intuitively variations of the hypersurface can significantly change the patch while the volume region barely gets distorted in volume and shape. Indeed, numerical computations have confirmed this in the lowest-dimensional cases [60].

Chapter 7

Conclusions and Outlook

Local integral invariants based on Principal Component Analysis have been introduced in the Geometry Processing literature in order to perform shape and feature detection of geometric properties of manifolds, usually from finite samples of points. In particular, principal curvatures and principal directions for surfaces in space have been found in the series expansion of the eigenvalue decomposition of PCA covariance matrices, computed from small kernel domains on the surface. Similar covariance matrices were also introduced with the purpose of finding local frames adapted to the tangent and normal spaces of the invariant manifolds of dynamical systems of large dimension. The core results of these milestones are summarized in [51], [55], which this dissertation generalizes to arbitrary dimension and reproduces as straightforward corollaries.

Indeed, in chapter 2 we have proposed a generalization of traditional PCA integral invariants to regions inside general Riemannian manifolds as ambient space by using the exponential map. The volume, barycenter and eigenvalue decomposition of the covariance matrix of the geodesic coordinates of the region points serve as local invariants that are expected to encode the geometry of the domain. In particular, the covariance matrix measures the statistical correlation among the geodesic distance coordinates of the underlying point set. Two specific kernel domains are proposed: the spherical and cylindrical intersection regions on the submanifold, i.e., the domains given by the intersection of the submanifold with a ball and generalized cylinder of the higher-dimensional ambient manifold. In the case of hypersurfaces, the volume inside a ball delimited by the hypersurface is also considered as a third type of region to study. Since these domains have an intrinsic scale ε, the covariance analysis can be interpreted as the eigenvalue decomposition of matrix-valued functions of scale at every point of the submanifold. Therefore, the eigenvalue asymptotic series with scale is expected to encapsulate geometric information at the point.

Once integral invariants and their kernel domains were defined, we introduced the notion of descriptors at scale as specific approximations to characteristic properties of the submanifold given in terms of integral invariant information. In particular, we are interested in the curvature information encoded by the extrinsic second fundamental form and the intrinsic Riemann curvature tensor. For hypersurfaces all this information is reduced to the principal directions and principal curvatures. In the rest of our work we focus on computing these integral invariants for curves, hypersurfaces and arbitrary embedded Riemannian manifolds in order to obtain corresponding descriptors of their curvatures from integral invariants.

Previous results on the covariance analysis of regular curves were reviewed in chapter 3, where our contribution completed the relationship known between the leading term of the asymptotic series of the covariance eigenvalues and the Frenet-Serret curvatures of the curve. In particular, quotients of successive eigenvalues are proportional to these curvatures squared. The coefficient of proportionality was not explicitly known, for which we needed to use the theory of orthogonal polynomials and its relation to Hankel determinants via the moment problem. This allowed us to obtain explicit recursion relations for a certain family of Hankel determinants that yields precisely the conjectured coefficient in any dimension. This asymptotic relation provides a direct link between the curvatures and the eigenvalues of the local covariance matrix, whose limit eigenvectors also converge to the Frenet-Serret frame. By the existence and uniqueness, up to rigid motion, of Frenet curves given by such information, the covariance integral invariants can be said to completely characterize these curves.

Extending the study of curves to hypersurface in arbitrary dimension, in chapter 4, we were able to obtain all the curvature information from the covariance analysis. We computed the volumes of the three types of domains proposed and showed that they are given in terms of the corresponding ball volume of the same dimension but with second order corrections proportional to the mean curvature and scalar curvature. The appearance of the extrinsic curvature is to be expected, in contrast to the volume of intrinsic geodesic balls, since our domains do depend on the embedding of the hypersurface. The covariance eigenvalue series expansion has one eigenvalue that scales faster than the others, whose eigenvector converges to the normal vector of the hypersurface at the center of the domain; the other eigenvalues have eigenvectors that generically converge to the principal directions at the point. We computed the second order terms in their series expansion and showed that they are completely given by the principal curvatures or their squares, which implies that the covariance matrix is given at second order by the Weingarten operator, establishing a direct link between integral invariants and curvature again.

The most general and complete study of the covariance-curvature relationship was carried out in chapter 5 for embedded Riemannian manifolds of dimension n in $\mathbb{R}^{n+k}$. Here we introduced the generalization of the classical third fundamental form to arbitrary dimension. This tensor is a bilinear form on the tangent space that takes values in the normal endomorphism space. Its components in an orthonormal basis of the normal space are defined by the metric products of the Weingarten maps associated to those normal vectors. Since the Weingarten map at a normal vector is the linear operator on the tangent space associated to the bilinear second fundamental form, this third fundamental form can be interpreted as the tensor associated to the linear operator given by the product of a pair of Weingarten maps at possibly different normal vectors. The geometric meaning of this tensor is given by the classical Ricci equation: the noncommutativity of its components measures the Riemann curvature of the induced connection on the normal bundle of the submanifold. For hypersurfaces it is given by the squares of the principal curvatures. The normal, tangent and total traces of this tensor were computed and showed to be directly related to the Weingarten map at the mean curvature and also the Ricci operator.

Then we obtained the leading order terms of the cylindrical domain integral invariants. In the generic case, the scaling of the eigenvalues singles out the decomposition of the ambient tangent space into the tangent and normal spaces of the submanifold, where the corresponding eigenvectors provide a basis for each of them. In the case of normal cylinders, the second order terms of the eigenvalue series was computed, thanks to a fundamental lemma that we proved in order to show that finding the covariance matrix in an arbitrary basis is enough to determine the eigenvalues from its block structure. In particular, the tangent block shows that the first n eigenvalues scale with ε^{n+2} and have second order corrections given by the eigenvalues of the normal trace of the third fundamental form, thus encoding its curvature information. Moreover the corresponding eigenvectors converge to a basis of the tangent space given by the generalized principal directions of this tensor. The normal block shows that the last k eigenvalues scale as ε^{n+4} and encode the curvature information given by the tangent trace of the third fundamental form and the mean curvature tensor. The corresponding eigenvectors converge to a basis of the normal space given by the eigenvectors of this combined tensor.

Finally, taking into account all the correction terms due to the spherical domain irregularities, we performed the analogous analysis for the ball intersection region. These boundary contributions do introduce significant changes in the tangent eigenvalues and eigenvectors, since now they are directly given at second order by the Weingarten map at the mean curvature vector. Thus, as for hypersurfaces, the tangent eigenvectors of the covariance matrix of the spherical domain converge with scale to the most canonical principal directions one can define in arbitrary codimension, those of the Weingarten map at the mean curvature.

From this, in chapter 6, we solved for the series coefficients to obtain descriptor formulas that give approximations at scale of the curvature information. In particular, this was written as a generalization to higher dimension of the asymptotic ratio formula proved for curves, where now the quotient of differences and products of covariance eigenvalues is proportional to the difference of third fundamental form curvatures. In the case of hypersurfaces, all these ratios are given by differences of the principal curvatures or their squares. By solving for the second order coefficients and truncating the series, the formulas from the previous chapters yield concrete descriptors in terms of the integral invariants. The spherical component descriptors have the best error bound and need no sign choices, whereas the cylindrical descriptors have an error one order better than the one that can be proved for the spherical descriptors.

These results prove a completely general relationship in any dimension between differential-geometric curvature and integral covariance. We can think of the results of our work as a dictionary between differential geometry and local statistical analysis, since it allows for the recovery of manifold curvature information from the statistics of the underlying point set. Therefore, our developments serve as a theoretical basis for more applied and computational implementations geared towards Manifold Learning in arbitrary dimension from big clouds of points. These results provide explicit formulas that relate Riemannian geometry to Geometric Data Analysis via the integral invariant approach.

The natural possible paths to take after our work are multiple. First, obtain local expressions for the induced volume element on a submanifold within an arbitrary Riemannian manifold as ambient space, taking into account the ambient Riemann curvature contributions to the ambient metric in normal coordinates at second order [15]. This would generalize lemma 2.1.1 in order to recompute all integrals with the new ambient space curvature terms, which is fundamental to do covariance analysis on submanifolds of non-embedded matrix manifolds [2], [54] and statistical manifolds [8, 9]. Second, study robustness and stability with respect to noise. In order to do this, the canonical procedure would be to study the first variation of the covariance matrix under the one parameter family of deformations of the submanifold given by the well-known mean normal flow [43], and bound the resulting terms. Third, arrive at a formulation and measurement of submanifold curvature in terms of the principal angles between tangent spaces at nearby points, so that the covariance analysis can be related to a Grassmannian formulation. For this, obtain the EVD of a finite approximation of the parallel transport of a tangent frame moving along a geodesic. Fourth, for specific families of submanifolds whose parameter space is known, try to establish a sampling theorem using curvature descriptors. Fifth, study and implement computationally the most efficient numerical methods to compute the integral invariant descriptors given a big cloud of data points, e.g. using FFT convolutions [51]. Then use this to do Manifold Learning and Geometry Processing, e.g. to do data classification [59] based on curvature profiles.

Therefore, covariance analysis opens a new perspective to look at differential geometry in any setting, both as a theoretical dictionary and as a computational tool.

Bibliography

[1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Applicandae Mathematicae, 80 (2004), pp. 199–220.

[2] ——, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.

[3] P. ALLIEZ, D. COHEN-STEINER, Y. TONG, AND M. DESBRUN, *Voronoi-based variational reconstruction of unoriented point sets*, in Proceedings of the Fifth Eurographics Symposium on Geometry Processing, SGP '07, Aire-la-Ville, Switzerland, 2007, Eurographics Association, pp. 39–48.

[4] J. ÁLVAREZ-VIZOSO, R. ARN, M. KIRBY, C. PETERSON, AND B. DRAPER, *Geometry of Curves in $\mathbb{R}^n$, Singular Value Decomposition, and Hankel Determinants*, arXiv:1511.05008v2.

[5] J. ÁLVAREZ-VIZOSO, M. KIRBY, AND C. PETERSON, *Integral Invariants from Covariance Analysis of Embedded Riemannian Manifolds*, preprint arXiv:1804.10425.

[6] ——, *Manifold Curvature Descriptors from Hypersurface Integral Invariants*, preprint arXiv:1804.04808.

[7] ——, *Manifold Curvature from Covariance Analysis*, 2018 IEEE Workshop on Statistical Signal Processing (SSP).

[8] S. AMARI, *Information Geometry and its Applications*, Springer-Verlag, 2016.

[9] S. AMARI AND H. NAGAOKA, *Methods in Information Geometry*, American Mathematical Society, 2000.

[10] G. ANDREWS, R. ASKEY, AND R. ROY, *Special Functions*, Cambridge University Press, 1999.

[11] J. BERKMANN AND T. CAELLI, *Computation of surface geometry and segmentation using covariance techniques*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (1994), pp. 1114–1116.

[12] D. BROOMHEAD, R. INDIK, A. NEWELL, AND D. RAND, *Local adaptive Galerkin bases for large-dimensional dynamical systems*, Nonlinearity, 4 (1991), p. 159.

[13] F. CAZALS AND M. POUGET, *Estimating differential quantities using polynomial fitting of osculating jets*, Symp. Geometry Processing. Eurographics, (2003), pp. 177–178.

[14] F. CAZALS AND M. P. .-. POUGET, *Molecular shape analysis based upon the Morse-Smale complex and the Connolly function*, Proc. Symp. Comp. Geometry, (2003).

[15] I. CHAVEL, *Eigenvalues in Riemannian Geometry*, Academic Press, 1984.

[16] ——, *Riemannian Geometry*, Cambridge University Press, 2nd ed., 2006.

[17] G. CHEN, A. V. LITTLE, M. MAGGIONI, AND L. ROSASCO, *Wavelets and Multiscale Analysis*, Springer-Verlag, 2011, ch. 10.

[18] S. CHERN, W. CHEN, AND K. LAM, *Lectures on Differential Geometry*, World Scientific, 1999.

[19] U. CLARENZ, M. GRIEBEL, M. RUMPF, M. SCHWEITZER, AND A. TELEA, *Feature sensitive multiscale editing on surfaces*, The Visual Computer, 20 (2004), pp. 329–343.

[20] U. CLARENZ, M. RUMPF, AND A. TELEA, *Robust feature detection and local classification for surfaces based on moment analysis*, IEEE Transactions on Visualization and Computer Graphics, 10 (2003), pp. 516–524.

[21] M. CONNOLLY, *Measurement of protein surface shape by solid angles*, J. Mol. Graphics 4, (1986), pp. 3–6.

[22] C. DAVIS AND W. KAHAN, *The rotation of eigenvectors by a perturbation. III*, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.

[23] G. FOLLAND, *How to integrate a polynomial over a sphere*, The American Mathematical Monthly, 108 (2001), pp. 446–448.

[24] W. GAUTSCHI, *Orthogonal Polynomials, Computation and Approximation*, Oxford University Press, 2004.

[25] A. GRAY, *Tubes*, Springer Basel AG, 2004.

[26] A. GRAY, E. ABBENA, AND S. SALAMON, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, Chapman and Hall/CRC, 3rd ed., 2006.

[27] A. GRAY AND L. VANHECKE, *Riemannian geometry as determined by the volumes of small geodesic balls*, Acta Math., 142 (1979), pp. 157–198.

[28] H. HEINE, *Handbuch der Kugelfunctionen*, vol. I, Berlin, 1878.

[29] H. HOPPE, T. DEROSE, T. DUCHAMP, J. MCDONALD, AND W. STUETZLE, *Surface reconstruction from unorganized points*, SIGGRAPH Comput. Graph., 26 (1992), pp. 71–78.

[30] D. HULIN AND M. TROYANOV, *Mean curvature and asymptotic volume of small balls*, The American Mathematical Monthly, 110 (2003).

[31] T. KATO, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, 1982.

[32] S. KOBAYASHI AND K. NOMIZU, *Foundations of Differential Geometry*, John Wiley & Sons, Inc., 1969.

[33] W. KÜHNEL, *Differential Geometry: curves-surfaces-manifolds*, vol. 16, American Mathematical Society, 2006.

[34] Y.-K. LAI, S.-M. HU, AND T. FANG, *Robust principal curvatures using feature adapted integral invariants*, Proceedings - SPM 2009: SIAM/ACM Joint Conference on Geometric and Physical Modeling, (2009), pp. 325–330.

[35] J. LEE, *Manifolds and Differential Geometry*, vol. 107 of Graduate Studies in Mathematics, American Mathematical Society, 2009.

[36] A. V. LITTLE, *Estimating the Intrinsic Dimension of High-Dimensional Data Sets: A Multiscale, Geometric Approach*, PhD thesis, Duke University, 2011.

[37] A. V. LITTLE, Y.-M. JUNG, AND M. MAGGIONI, *Multiscale estimation of intrinsic dimensionality of data sets*, in Proc. A.A.A.I., 2009.

[38] A. V. LITTLE, J. LEE, Y.-M. JUNG, AND M. MAGGIONI, *Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD*, in Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on, IEEE, 2009, pp. 85–88.

[39] A. V. LITTLE, M. MAGGIONI, AND L. ROSASCO, *Multiscale Geometric Methods for Data Sets I: Multiscale SVD, Noise and Curvature*, tech. rep., MIT, 2012.

[40] H. LIU, U. SIMON, L. VERSTRAELEN, AND C. WANG, *The third fundamental form metric for hypersurfaces in nonflat space forms*, J. geom., 65 (1999), pp. 130–142.

[41] S. MANAY, D. CREMERS, B.-W. HONG, AND A. J. YEZZI, *Integral invariants for shape matching*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28 (2006), pp. 1602–1618.

[42] S. MANAY, B.-W. HONG, A. J. YEZZI, AND S. SOATTO, *Integral invariant signatures*, in Computer Vision - ECCV 2004, T. Pajdla and J. Matas, eds., Berlin, Heidelberg, 2004, Springer Berlin Heidelberg, pp. 87–99.

[43] C. MANTEGAZZA, *Lecture Notes on Mean Curvature Flow*, Birkhäuser, Springer Basel AG, 2011.

[44] Q. MÉRIGOT, *Geometric structure detection in point clouds*, PhD. Thesis, Université Nice Sophia Antipolis, 2009.

[45] Q. MÉRIGOT, M. OVSJANIKOV, AND L. GUIBAS, *Voronoi-based curvature and feature estimation from point clouds*, Visualization and Computer Graphics, IEEE Transactions on, 17 (2011), pp. 743 – 756.

[46] B. O'NEILL, *Semi-Riemannian Geometry*, Academic Press, 1983.

[47] X. PENNEC, *Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements*, in Int. Workshop on Nonlinear Signal and Image Processing (NSIP'99), 1999.

[48] ——, *Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements*, J. Math. Imaging Vis., 25 (2006), pp. 127–154.

[49] M. POSTNIKOV, *Geometry VI - Riemannian Geoemtry*, Springer-Verlag, 2001.

[50] H. POTTMANN, J. WALLNER, Q.-X. HUANG, AND Y.-L. YANG, *Integral invariants for robust geometry processing*, CAGD 26, 1 (2008).

[51] H. POTTMANN, J. WALLNER, Y.-L. YANG, Y.-K. LAI, AND S.-M. HU, *Principal curvatures from the integral invariant viewpoint*, Computer Aided Geometric Design, 24 (2007), pp. 428–442.

[52] F. RELLICH, *Perturbation Theory of Eigenvalue Problems*, Gordon and Breach, 1969.

[53] J. SHOHAT AND J. TAMARKIN, *The Problem of Moments*, American Mathematical Society, 1943.

[54] S. SMITH, *Covariance, Subspace, and Intrinsic Cramér-Rao Bounds*, IEEE Transactions on Signal Processing, 53 (2005), pp. 1610–1630.

[55] F. SOLIS, *Geometric aspects of local adaptive Galerkin bases*, PhD. Thesis, University of Arizona, 1993.

[56] ——, *Geometry of local adaptive Galerkin bases*, Applied Mathematics & Optimization, (2000), pp. 331–342.

[57] M. SPIVAK, *A Comprehensive Introduction to Differential Geometry*, Publish or Perish, Inc., 3rd ed., 1999.

[58] V. TOPONOGOV, *Differential Geometry of Curves and Surfaces. A Concise Guide*, Birkhäuser Boston, 2006.

[59] P. TURAGA, A. VEERARAGHAVAN, A. SRIVASTAVA, AND R. CHELLAPPA, *Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2011), pp. 2273 – 2286.

[60] Y.-L. YANG, Y.-K.-K. LAI, S.-M. HU, AND H. POTTMANN, *Robust principal curvatures on multiple scales*, Proc. Eurographics Symposium on Geometry Processing, (2006), pp. 223–226.

Appendix A

Integration of Monomials over Spheres

Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and denote the sphere and ball of radius ε in $\mathbb{R}^n$ by:

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

where we set $\mathbb{S}^{n-1} = \mathbb{S}^{n-1}(1)$. Using generalized spherical coordinates $(r, \phi_1, \ldots, \phi_{n-1})$, where $r = \|\boldsymbol{x}\|$, $\overline{x}_\mu = x_\mu / r \in \mathbb{S}^{n-1}$, i.e.,

$$\overline{x}_1 = \cos\phi_1, \ \ldots, \quad \overline{x}_{n-1} = \sin\phi_1 \cdots \sin\phi_{n-2}\cos\phi_{n-1}, \quad \overline{x}_n = \sin\phi_1 \cdots \sin\phi_{n-2}\sin\phi_{n-1},$$

the Euclidean measure over the unit sphere and ball of any radius can be written as

$$d\,\mathbb{S}^{n-1} = d\phi_{n-1} \prod_{\mu=1}^{n-2} \sin^{n-1-\mu}(\phi_\mu) d\phi_\mu, \qquad d^n B = dx_1 \cdots dx_n = r^{n-1} dr\, d\,\mathbb{S}^{n-1}. \tag{A.1}$$

Definition A.0.1. For any integers $\alpha_1, \ldots, \alpha_n \in \{0, 1, 2, \ldots\}$, the integrals of the monomials $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ over the unit sphere and the ball of radius ε are denoted by:

$$C^{(n)}_{\alpha_1 \ldots \alpha_n} = \int_{\mathbb{S}^{n-1}} x_1^{\alpha_1} \cdots x_n^{\alpha_n}\, d\,\mathbb{S}^{n-1}, \qquad D^{(n)}_{\alpha_1 \ldots \alpha_n} = \int_{B^n(\varepsilon)} x_1^{\alpha_1} \cdots x_n^{\alpha_n}\, d^n B. \tag{A.2}$$

These can be computed directly in spherical coordinates by collecting factors and separating the integrals into a product of powers of sines and cosines of independent angles which are given in terms of the Beta function. This then telescopes and simplifies. Another shorter proof uses the usual exponential trick, see for example [23], resulting in the following fundamental formula.

Theorem A.0.2. *Denoting $\beta_\mu = \frac{1}{2}(\alpha_\mu + 1)$, the values of the integrals A.2 over spheres are*

$$C^{(n)}_{\alpha_1\ldots\alpha_n} = \begin{cases} 0, & \textit{if some } \alpha_\mu \textit{ is odd,} \\ 2\dfrac{\Gamma(\beta_1)\Gamma(\beta_2)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1+\beta_2+\cdots+\beta_n)}, & \textit{if all } \alpha_\mu \textit{ are even,} \end{cases} \tag{A.3}$$

and the integrals over balls become

$$D^{(n)}_{\alpha_1\ldots\alpha_n} = \frac{\varepsilon^{n+(\alpha_1+\cdots+\alpha_n)}}{n+(\alpha_1+\cdots+\alpha_n)}\, C^{(n)}_{\alpha_1\ldots\alpha_n}. \tag{A.4}$$

Notice that the values of the integrals of these monomials only depend on the combination of powers, not on which particular coordinates have those powers. Using these formulas we compute the relevant integrals that are needed for our work.

Remark A.0.3. Unless integrals over spheres of different dimension appear in the same expression, we shall abbreviate and omit the superscript $^{(n)}$ to be understood from the context.

Example A.0.4. Using the factorial property of the gamma function, $\Gamma(z+1) = z\Gamma(z)$, and the value $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, the integrals of monomials of even powers of order 2, 4 and 6, have the following relations (shortening $d\,\mathbb{S}^{n-1}$ as $d\,\mathbb{S}$):

$$C_2 = \int_{\mathbb{S}^{n-1}} x_1^2 \, d\,\mathbb{S} = 2\frac{\Gamma(\frac{3}{2})\Gamma(\frac{1}{2})^{n-1}}{\Gamma(\frac{3}{2}+\frac{n-1}{2})} = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)},$$

$$C_{22} = \int_{\mathbb{S}^{n-1}} x_1^2 x_2^2 \, d\,\mathbb{S} = \frac{1}{n+2}\, C_2,$$

$$C_4 = \int_{\mathbb{S}^{n-1}} x_1^4 \, d\,\mathbb{S} = \frac{3}{n+2}\, C_2 = 3\, C_{22},$$

$$C_{222} = \int_{\mathbb{S}^{n-1}} x_1^2 x_2^2 x_3^2 \, d\,\mathbb{S} = \frac{1}{(n+2)(n+4)}\, C_2,$$

$$C_{24} = \int_{\mathbb{S}^{n-1}} x_1^2 x_2^4 \, d\,\mathbb{S} = \frac{3}{(n+2)(n+4)}\, C_2 = 3\, C_{222},$$

$$C_6 = \int_{\mathbb{S}^{n-1}} x_1^6 \, d\,\mathbb{S} = \frac{15}{(n+2)(n+4)}\, C_2 = 15\, C_{222}.$$

The value of C_2 is related to the n-dimensional volume of the ball of radius ε, and the $(n-1)$-dimensional area of the unit sphere by

$$V_n(\varepsilon) = \text{Vol}(B^n(\varepsilon)) = \varepsilon^n\, C_2, \qquad S_{n-1} = \text{Area}(\mathbb{S}^{n-1}) = n\, C_2.$$

The integrals over balls needed in our work are:

$$D_2 = \int_{B^n(\varepsilon)} x_1^2 \;\; dx_1 \cdots dx_n = \frac{\varepsilon^{n+2}}{n+2} C_2 = \frac{\varepsilon^2}{n+2} V_n(\varepsilon),$$

$$D_{22} = \int_{B^n(\varepsilon)} x_1^2 x_2^2 \;\; dx_1 \cdots dx_n = \frac{\varepsilon^{n+4}}{(n+2)(n+4)} C_2 = \frac{\varepsilon^4}{(n+2)(n+4)} V_n(\varepsilon),$$

$$D_4 = \int_{B^n(\varepsilon)} x_1^4 \;\; dx_1 \cdots dx_n = \frac{3\, \varepsilon^{n+4}}{(n+2)(n+4)} C_2 = \frac{3\, \varepsilon^4}{(n+2)(n+4)} V_n(\varepsilon).$$

We also need the integral of monomials over half-balls $B^+(\varepsilon)$ (without loss of generality we can consider the half-ball is defined by $x_1 \geqslant 0$). If all the α_i are even then nothing changes in the proof of theorem A.0.2 except that now we integrate over half the domain and an extra factor of $\frac{1}{2}$ is needed. If any α_i is odd for $i \neq 1$, the integration over those variables is still carried out over the same domain so the overall integral is still 0. However, if α_1 is odd the corresponding integral of that coordinate does not cancel out, and the main formula still holds with $\beta_1 = 1$ but without the factor of 2.

Example A.0.5. Using the formula in the mentioned adjusted form, we define and compute

$$D_1^{(n)} = \int_{B^+(\varepsilon)} x_1 \;\; dx_1 \cdots dx_n = \frac{\varepsilon^{n+1}\, \pi^{\frac{n-1}{2}}}{2\Gamma(\frac{n+3}{2})},$$

which gives the constant needed in our main text

$$D_1^{(n+1)} = \int_{B^+(\varepsilon)} x_1 \;\; dx_1 \cdots dx_{n+1} = \frac{\varepsilon^2}{n+2} V_n(\varepsilon).$$

When integrating $\int_{B^+(\varepsilon)} x_1^2 \;\, \text{dVol}$, we shall just write $\frac{D_2}{2}$ to be consistent with our notation above.

In the general Riemannian setting, chart coordinates are often written with superindices, so the integral of a general product of these coordinates depends on the superindices involved which must not be confused with exponents. For instance

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

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 superindices are equal and the nonconnected superindices 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

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