Principal Boundary on Riemannian Manifolds

Zhigang Yao & Zhenyue Zhang

Taylor & Francis
Taylor & Francis Group

Check for updates

Principal Boundary on Riemannian Manifolds

Zhigang Yao[a] and Zhenyue Zhang[b]

[a]Department of Statistics and Applied Probability, National University of Singapore, Singapore; [b]School of Mathematical Science, Zhejiang University, Hangzhou, China

ABSTRACT

We consider the classification problem and focus on nonlinear methods for classification on manifolds. For multivariate datasets lying on an embedded nonlinear Riemannian manifold within the higher-dimensional ambient space, we aim to acquire a classification boundary for the classes with labels, using the intrinsic metric on the manifolds. Motivated by finding an optimal boundary between the two classes, we invent a novel approach—the principal boundary. From the perspective of classification, the principal boundary is defined as an optimal curve that moves in between the principal flows traced out from two classes of data, and at any point on the boundary, it maximizes the margin between the two classes. We estimate the boundary in quality with its direction, supervised by the two principal flows. We show that the principal boundary yields the usual decision boundary found by the support vector machine in the sense that locally, the two boundaries coincide. Some optimality and convergence properties of the random principal boundary and its population counterpart are also shown. We illustrate how to find, use, and interpret the principal boundary with an application in real data. Supplementary materials for this article are available online.

1. Introduction

Most of the classification methodology in high-dimensional data analysis is deeply rooted in methods relying on linearity. Modern datasets often consist of a large number of samples, each of which is made up of many more features. Manifold data arise in the sense that the sample space of data is fundamentally nonlinear. Rather than viewing any observation as a point in a high-dimensional Euclidean space, it is more natural to assume the data points lie on an embedded lower-dimensional nonlinear manifold within the higher dimensional space. The lower-dimensional manifold structure can usually be interpreted from at least two scenarios: (1) the physical data space is an actual manifold; (2) the underlying data structure can be approximated by a close manifold. In the former scenario, the data space is usually known, and it can be further seen as data in the shape space (Kendall et al. 1999; Patrangenaru and Ellingson 2015), for example, a seismic event in geophysics, leaf growth pattern, and data with nonlinear constraints. It is thus forced to lie on a manifold. In the latter scenario, the manifold is uncovered from the dataset by a nonlinear dimensionality reduction technique referred to as the manifold learning method (Roweis and Saul 2000; Donoho and Grimes 2003; Zhang and Zha 2004), and is thus considered unknown.

In this work, we consider the classification problem and design the nonlinear methods that perform as a boundary for classifying data sets lying on the *known* manifold. This problem has become increasingly relevant, as many real applications such

as medical imaging (Gerber et al. 2010; Souvenir and Pless 2007) and computer vision (Pennec 2006; Pennec and Thirion 1997; Sen et al. 2008; Tuzel, Porikli, and Meer 2007) produce data in such forms. This encourages researchers to conduct the analysis directly on the manifold. The rationale behind this is that usage of the metric on the manifold is much more reasonable than using the metric in the Euclidean space, if the data resides on manifolds. However, the methodology defined using the manifold space for classification is still lacking. To perform reliable classification for data points on manifolds, a strategy for developing statistical tools, such as the nonlinear classification boundary, in parallel with their Euclidean-counterparts, is significantly relevant.

Though we have seen tremendous efforts in the development of statistical procedures for classification problems, these efforts have mainly been focused on constructing the separating hyperplane between two classes in the Euclidean space. The optimality is essentially built on finding *linear* (affine) hyperplanes that separate the data points as well as possible. Among them, the linear discriminant analysis (LDA), or the slightly different logistic regression method, manifest themselves through the seeking of the hyperplane by minimizing the so-called discriminant function, and are thus able to trace out a linear boundary separating the different classes (Hastie, Tibshirani, and Friedman 2009). Further to the linear boundary, the support vector machine (SVM) finds a seemingly different separating hyperplane; that is, the hyperplane is actually found (up to some loss function)

not in its original feature space, but in an enlarged space, by transforming the feature space into an unknown space via basis functions. Furthermore, locally linear SVM variants (Ladický and Torr 2011; Zhang et al. 2011) learn smooth classifiers from existing descriptions of manifolds that characterize the manifold as a set of piecewise affine charts and show some promise in application. There have been several pieces of works of research on the statistical methods on manifolds over the past decades, centering around finding the main mode of variation (Fréchet 1948; Huckemann, Hotz, and Munk 2010) in the data, or finding a manifold version of the principal components for the data (Jupp and Kent 1987; Fletcher et al. 2004; Huckemann and Ziezold 2006; Kume, Dryden, and Le 2007; Fletcher and Joshi 2007; Kenobi, Dryden, and Le 2010; Jung, Dryden, and Marron 2012; Eltzner et al. 2018), in terms of dimension reduction. Other approaches (Zhu, Ghahramani, and Lafferty 2003; Belkin, Niyogi, and Sindhwani 2006) are more attached to regularization that exploits the geometry of the input space, achieving better generalization error. However, none of them seem to be naturally adaptable for deriving a boundary on a manifold, due to their "noncurve-fitting" nature. A Bayesian framework (Bhattacharya and Dunson 2012) has also been proposed to use the kernel mixture model for the joint distribution of the labels and features, with the kernel expressed in product form and dependence induced through the unknown mixing measure. The principal curve by Hastie and Stuetzle (1989) is a self-consistent curve defined in Euclidean space. Its natural extension to Riemannian manifolds Hauberg (2015) is based on the replacement of the Euclidean conditional mean with the intrinsic mean of the local data on manifolds. Recently, the principal flow (Panaretos, Pham, and Yao 2014) has been proposed as a one-dimensional curve defined on the manifold, such that it attempts to follow the main direction of the data locally, while still being able to accommodate the "curve fitting" characteristic on the manifold. The variational principal flow (Liu et al. 2017) incorporates the level set method to obtain a fully implicit formulation of the problem. The principal submanifold (Yao, Eltzner, and Pham 2017) extends the principal flow to a higher dimensional submanifold. Most recently, A fixed boundary flow (Yao, Xia, and Fan 2019) on the manifold has been carefully studied with its relation to the principal flow. It is natural to raise the question as follows: Is there a way for the data to be separated directly on the manifold?

Inspired by the principal flow, we tackle the problem of finding the classification boundary endowed with a curved metric on Riemannian manifolds. We explore the limitations inherent in the problem when trying to find such a boundary. Our idea is to trace a boundary out of the two summary curves (principal flows) from the two classes, and at the same time retain some canonical interpretation for the boundary. Our intuition is that, as the two principal flows represent the mean trend of the two classes, to classify the points it is enough to separate the two flows in some optimal way. This means that one does not need to consider the data points beyond the two flows on each side, as they are irrelevant to the classification if we can separate the flows well. Naturally, because of the two principal flows, the process of constructing the boundary can be supervised, in the sense that the boundary grows itself by borrowing strength from

the two principal flows. To achieve this, the key insight is the margin, a measure of distance between the target boundary and the two principal flows, subject to the presence of noise originating from each class. In principle, an optimal boundary can be framed by maximizing the margin between the target boundary and the two corresponding principal flows. The optimization involved therein can be relaxed by fine-tuning the subspace of the vector field from the two principal flows, up to their parallel transport on the manifold. From this perspective, the boundary retains the characteristic of being principal, in the sense that at each point of the boundary, it points to the direction calculated over the two directions of the vector field from the two principal flows. This finally results in a classification boundary, which is named the principal boundary to draw a relation to the principal flow.

Figure 1 shows the data of major volcano and significant offshore earthquake activity that occurred around the region of eastern Japan. The principal boundary and misclassified points illustrate the effect of varying flows on the classification performance of the boundary. The transparent pink and transparent yellow areas (Figure 1(b)–(d)) represent the classified regions that correspond to the volcano and earthquake activity, respectively, contrasting with the volcano and earthquake data plotted in red and blue. The principal boundary (in green) exhibits different behavior, as the principal flows vary with respect to their locality parameters: the boundary classifies the two data clouds correctly (Figure 1(b)); there is one misclassified earthquake point in the volcano region (Figure 1(c)); there are three misclassified earthquake points in the volcano region, and one misclassified earthquake point on the boundary (Figure 1(d)). Readers may refer to Section 7 for the analysis. The principal boundary is able to follow the trajectory toward the direction of maximum margin between the flows, while separating the data points as best as it could.

The principal boundary should be a co-dimensional submanifold, that is, a submanifold with dimension $m - 1$ in the m-dimensional manifold. In this article, we focus on two-dimensional manifolds, that is, $m = 2$, for simplicity, and hence, the principal boundary is also a curve in the two-dimensional manifold. This restriction can be extended to higher dimensional manifolds with further efforts. We demonstrate how the problem of obtaining the principal boundary can be transformed to a well-defined integration problem in Section 3.3, with a motivation and an introduction of the margin in Section 3.1, and the optimal boundary between two curves in Section 3.2. The formal definition of the principal boundary is given in Section 3.3. Section 5.1 contains the property of the principal boundary for a finite sample with a detailed analysis of the relation between the boundary and SVM. Some optimality and convergence properties of the random principal boundary, given data with random distribution on manifolds, are provided in Section 5.2. Generally speaking, our formulation of the boundary is feasible for any Riemannian manifold provided that the geodesic is unique, although an unknown geodesic might increase the complexity of computation.

The remaining part of the article is organized as follows. We start with a brief introduction of the principal flow (Sections 2.2 and 2.3) with a modified vector field and a modified principal flow. Section 4 investigates an implementable algorithm used to
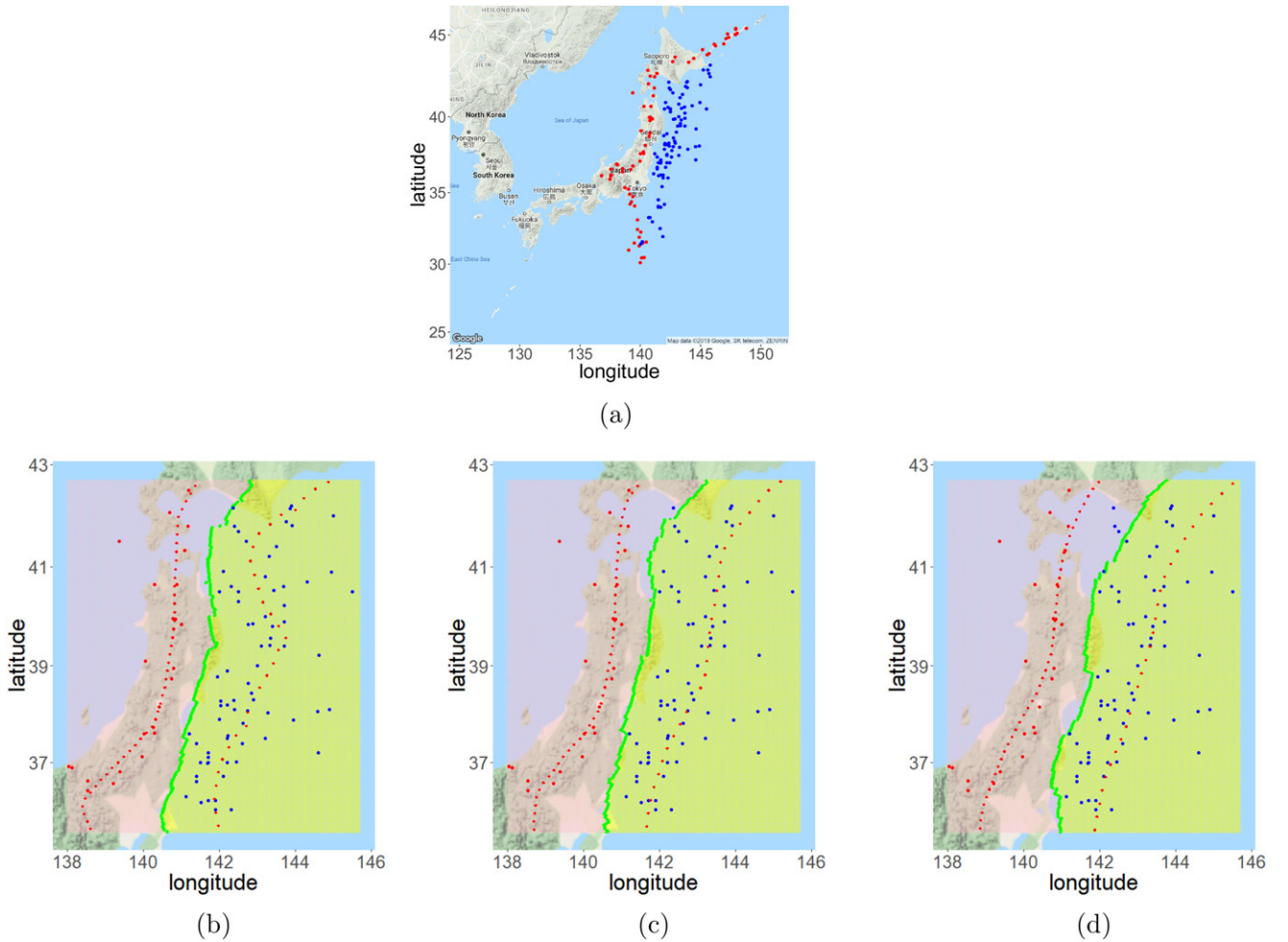
Figure 1. Seismic events data of significant offshore earthquakes with magnitude 7.5 or greater (in blue) between 1900 and 2018, and major volcanoes (in red) in 2001 on a flat world atlas (a). Plots of the principal boundary (in green) between the volcano region (transparent pink) and earthquake region (transparent yellow) in the zoom-in area ($138.0°$–$145.7°$ E, $35.5°$–$42.7°$ N) (b)–(d). The bottom row illustrates the effect of varying flows on the classification performance of the boundary: the two clusters are correctly classified with no misclassified points (b), one misclassified earthquake point (c), and four misclassified earthquake points (d).

determine the principal boundary. In Sections 6 and 7, we illustrate the principal boundary by means of simulated examples and real application. The appendix and supplementary materials provide all the technical details. We end the article with a discussion.

2. The Problem

We are interested in the following problem: let $\mathcal{M}$ be a Riemannian manifold in $\mathbb{R}^d$ with the dimension $m < d$. Let $x_{1,i}$ ($i = 1, \ldots, n_1$) be the data points on $\mathcal{M}$ with label $\ell_{1,i} = +1$ ($i = 1, \ldots, n_1$) and $x_{2,i}$ ($i = 1, \ldots, n_2$) be the data points on $\mathcal{M}$ with label $\ell_{2,i} = -1$ ($i = 1, \ldots, n_2$). We are looking for a classification boundary, say γ, an $m - 1$ submanifold on $\mathcal{M}$, that separates the data points for class 1 and -1 as well as possible.

The heuristic behind the principal boundary is that first we construct the two mean flows γ_1 and γ_2 of the data points $\{x_{1,i}\}$ and $\{x_{2,i}\}$, respectively. Each mean flow represents the principal direction of the data variation for each class on $\mathcal{M}$. Second, the classification problem can now be rephrased as the finding of a

flow γ, lying between γ_1 and γ_2, which separates the two as well as possible.

The two mean flows are in fact called principal flows. Before we continue, let us digress slightly and review the principal flow.

2.1. Preliminaries

Let x_i ($i = 1, \ldots, n$) be n data points on a complete Riemannian manifold $\mathcal{M}$ of dimension m, where $m < d$, embedded in the linear space $\mathbb{R}^d$.

We assume that a differentiable function $F : \mathbb{R}^d \rightarrow \mathbb{R}^m$ always exists, such that

$$\mathcal{M} := \left\{ x \in \mathbb{R}^d : F(x) = 0 \right\}.$$

For each $x \in \mathcal{M}$ the tangent space at x will be denoted by $T_x\mathcal{M}$, then $T_x\mathcal{M}$ is characterized by the equation

$$T_x\mathcal{M} = \left\{ v \in \mathbb{R}^d : \nabla F_x v = 0 \right\}.$$

Thus, $T_x\mathcal{M}$ is in fact a *vector space*, the set of all tangent vectors to $\mathcal{M}$ at x, which essentially provides a local vector space approximation of the manifold $\mathcal{M}$.

By equipping the manifold with the tangent space, we define two mappings back and forth between $T_x\mathcal{M}$ and $\mathcal{M}$: (1) the exponential map, well defined in terms of geodesics, is the map

$$\exp_x : T_x\mathcal{M} \to \mathcal{M} \tag{1}$$

by $\exp_x(v) = \gamma(\|v\|)$ with γ a geodesic starting from $\gamma(0) = x$ with initial velocity $\dot\gamma(0) = v/\|v\|$ and $\|v\| \le \delta$, and (2) the logarithm map (the inverse of the exponential map), is locally defined at least in the neighborhood of x,

$$\log_x : \mathcal{M} \to T_x\mathcal{M}. \tag{2}$$

Here, the exp and log are defined on a local neighborhood of 0 and x such that they are all well-defined, away from the cut locus of x on $\mathcal{M}$.

Let $x, y \in \mathcal{M}$. Denote all (piecewise) smooth curves $\gamma(t) : [0, 1] \to \mathcal{M}$ with endpoints such that $\gamma(0) = x$ and $\gamma(1) = y$. The *geodesic distance* from x to y is defined as

$$d_{\mathcal{M}}(x, y) = \inf \int_{[0,1]} \|\dot\gamma(t)\| \, dt. \tag{3}$$

Minimizing (3) yields the shortest distance within $\mathcal{M}$ between the two points x and y.

2.2. Definition of Principal Flow

Technically, the principal flow incorporates two ingredients: a local covariance matrix and a vector field. For the data point x_j, choose a neighborhood $\mathcal{N}(x_j, h)$ of x_j with a radius h, defined as

$$\mathcal{N}(x_j, h) = \{x_i : d_{\mathcal{M}}(x_i, x_j) \le h\}.$$

Accordingly, the local covariance matrix is defined as

$$\Sigma_h(x_j) = \frac{1}{\sum_i \kappa_h(x_i, x_j)} \sum_i \log_{x_j}(x_i) \otimes \log_{x_j}(x_i) \kappa_h(x_i, x_j),$$

where $y \otimes y := yy^T$, $\kappa_h(x_i, x_j) = K(h^{-1}d_{\mathcal{M}}(x_i, x_j))$ with a smooth nonincreasing univariate kernel K on $[0, \infty)$.

Let $B \subset \mathcal{M}$ be a connected open set covering x_i ($i = 1, \dots, n$) and such that $\log_x y$ is well defined for all $x, y \in B$. Assume that $\Sigma_h(x_j)$ has distinct first and second eigenvalues for all $x \in B$. The vector field is defined in the way that the first eigenvector $e_1(x_j)$ (or eigenvalue $\lambda(x_j)$) of $\Sigma_h(x_j)$ is extended to a vector field $W := \{W(x) \equiv v(x_j) : x \in \mathcal{N}(x_j, h)\}$ where $v(x_j) = e_1(x_j)$; that is, for any $x \in \mathcal{N}(x_j, h)$, we have

$$\Sigma_h(x)W(x) = \lambda(x)W(x) (\text{i.e., } W(x) \in W). \tag{4}$$

In the meantime, it has been proved (Panaretos, Pham, and Yao 2014) that $W : \mathcal{N}(x_j, h) \to \mathbb{R}^d$ is a differentiable mapping with $W(x)$ independent of the local coordinates of the tangent space $T_x\mathcal{M}$.

Definition 2.1. The principal flow γ of x_i ($i = 1, \dots, n$) is defined as the union of the two curves, γ^+ and γ^-, satisfying the following two variational problems, respectively,

$$\gamma^+ = \arg \sup_{\gamma \in \Gamma(x_0, v_0)} \int_\gamma \langle \dot\gamma, W(\gamma) \rangle \, ds, \tag{5}$$

$$\gamma^- = \arg \inf_{\gamma \in \Gamma(x_0, -v_0)} \int_\gamma \langle \dot\gamma, W(\gamma) \rangle \, ds, \tag{6}$$

where x_0 is the starting point and v_0 is a unit vector at x_0. The point x_0 can be chosen as the Fréchet mean $\bar{x}$ of x_i ($i = 1, \dots, n$) such that

$$\bar{x} = \arg\min_{x \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n d_{\mathcal{M}}^2(x, x_i),$$

or any other point of interest. Note that $\Gamma(x_0, v_0)$ is the set of all nonintersecting differentiable curves on $\mathcal{M}$.

It can be seen that the curve γ^+ starts at x_0 and follows the direction of the vector field, and the curve γ^- starts at x_0 and goes in the opposite direction of the vector field. Thus, the integral for γ^- is negative, which explains why the infimum appears in its definition.

Under the principal flow definition, we can define γ_1 of $x_{1,i}$ ($i = 1, \dots, n_1$) as the union of γ_1^+ and γ_1^-, and γ_2 of $x_{2,i}$ ($i = 1, \dots, n_2$) as the union of γ_2^+ and γ_2^-. For convenience, we will only consider the flow γ_1^+ in (1) of Definition (2.1), and rename it as γ_1. By symmetry, the solution to the flow γ_1^- in (2) of Definition (2.1) can be carried out in the same way. Similarly, we will restrict the discussion to γ_2^+ and rename it as γ_2.

2.3. Modified Principal Flow

The principal flow relies heavily on the vector field. However, the original definition (see (4)) of the vector field strictly constructs the direction of every point in the field to solely point to the eigenvector of the local covariance matrix. This definition could be problematic when we need a more delicate field for the flow. To be exact, for each point belonging to a neighborhood where the field is calculated, it can also be in other neighborhoods. It turns out that we will need to modify the vector field for the principal boundary, since the vector field plays a rather crucial role in the problem.

We will equip a vector field for each training sample (say x_j) in the data $\{x_i\}(i = 1, \dots, n)$. For samples in each neighborhood, say $\mathcal{N}(x_i, h)$, determine a locally dominate or principal vector v_i through the local tangent PCA.

A sample x_j can be the neighbor of multiple points. Let I_j be the index set of neighbor sets $\mathcal{N}(x_i, h)$ that holds x_j as a neighbor. The modification of the vector field amounts to the overall effect of holding multiple neighborhoods for a point. To achieve this, it is very natural to equip a vector for x_j as a weighted sum of the locally principal vectors $\{v_i : i \in I_j\}$. Let c_i be the mean of $\mathcal{N}(x_i, h)$ when we determine v_i. Then, we assign the vector $v(x_j)$ for x_j, which is the projection $v_{\mathcal{M}}(x_j)$ of the weighted sum

$$v(x_j) = \sum_{i \in I_j} w_{ij} v_i, \quad w_{ij} = \frac{\exp(-d_{\mathcal{M}}(x_j, c_i))}{\sum_{i \in I_j} \exp(-d_{\mathcal{M}}(x_j, c_i))}$$

onto the manifold at x_j.

As soon as we have a vector field constructed above, a principal flow γ of the given dataset $\{x_i\}$ is defined by

$$\gamma = \arg \sup_{\gamma \in \Gamma(x_0, v_0)} \int_\gamma \left\langle \dot\gamma, \sum_{x_j \in \mathcal{N}(p,h)} v(x_j) \right\rangle ds. \tag{7}$$

That is, at the point $p \in \gamma$, the tangent $\dot\gamma$ should match the vector field of samples in the neighborhood $\mathcal{N}(p, h)$. To differentiate from the principal flow in Definition 2.1, we call γ the modified principal flow.

3. Methodology

Now we return to the original question, that is: Given two principal flows γ_1 and γ_2 determined from two datasets $\{x_{1,i}\}$ and $\{x_{2,j}\}$, can we find a γ, that can be used as a classification boundary between the two classes? In principle, γ should be co-dimension one (of dimension $m - 1$); for simplicity, here we assume that $m = 2$, and hence, γ is also a flow in the two-dimensional manifold.

Strictly speaking, many γ's in $\Gamma(x_0, v_0)$ that separate two classes of data could exist. Therefore, by using the term the classification boundary, we refer to the best one. We present the general idea of constructing such a boundary here, leaving the formal introduction in Section 3.3: let the γ start from $\gamma(0)$ and move infinitesimally in the direction of $\dot{\gamma}(0)$. At this moment, we assume that both $\gamma(0)$ and $\dot{\gamma}(0)$ are carefully chosen so that the flow moves in the correct direction. Once the first move has been made, it may no longer make sense to continue moving in the same direction $\dot{\gamma}(0)$. One may ask, then, in which direction should we move toward? Obviously, we should not move toward γ_1 or γ_2, since this would cause γ to move close to either of the two flows. To update the direction, a natural strategy that plays an important role in building the boundary is to let γ move in a direction supervised by γ_1 and γ_2; that is, we follow the vector field inherited from γ_1 and γ_2, then move by choosing the proportional amount of vector field from γ_1 and γ_2 each time. Indeed, the right amount of vector field to choose for the next move is essentially an optimization problem, the derivation of which will be discussed in the next section. This being said, the intuitive version of such a boundary is not unique in the sense that a parallel curve satisfying the same condition always exists, and this can be seen by varying the initial point. To achieve the classification, let us view the problem slightly differently: note that only the points lying in between γ_1 and γ_2 could influence γ, so a very straightforward approach is to choose the tangent vector for the next move along the direction that creates the biggest margin between the data points for class $+1$ and -1. Under this rationale, iterating the process would approximately trace out an integral curve that is not only proportionally compatible to the vector fields of the two flows at each point, but also, more importantly, separates the margin, therefore allowing it to be considered as a classification boundary.

3.1. Margin

At each point $p \in \gamma$, the tangent vector of γ at p should be the locally principal vector of samples in $\mathcal{N}(p, h)$. Suppose the distinct first and second largest eigenvalues of the local covariance matrix of the centered samples for $\mathcal{N}(p, h)$ exist. The principal vector is the dominant eigenvector in $\mathcal{N}(p, h)$, corresponding to its largest eigenvalue λ_1. The local PCA also determines the second largest eigenvalue value λ_2. The ratio

$$\sigma_\gamma(p) = \frac{\lambda_2}{\lambda_1} h$$

approximately indicates the largest distance of samples in the neighborhood $\mathcal{N}(p, h)$ deviated from the mean along the eigenvector corresponding to λ_2.

The distance of a point q of the manifold to the principal flow γ is defined as

$$d_\mathcal{M}(q, \gamma) = \inf_{p \in \gamma} d_\mathcal{M}(q, p), \tag{8}$$

where $d_\mathcal{M}(p, q)$ is the geodesic distance between p and q on $\mathcal{M}$. We assume that the distance $d_\mathcal{M}(q, \gamma)$ is achievable, that is, there is a point $p \in \gamma$ such that $d_\mathcal{M}(q, p) = d_\mathcal{M}(q, \gamma)$. Such a minimizer p is referred as a projection of q onto γ. We denote by $\mathcal{P}_\gamma(q)$ as the set of all the minimizers.

Hence, if $d_\mathcal{M}(q, \gamma) > \sigma_{q,\gamma} = \sup_{p \in \mathcal{P}_\gamma(q)} \sigma_\gamma(p)$, the gap

$$m_\gamma(q) = d_\mathcal{M}(q, \gamma) - \sigma_{q,\gamma}$$

is a "soft" margin of q for classifying γ in the sense that the neighbor set $\mathcal{N}(p, h)$ locates in a side of q for each $p \in \mathcal{P}_\gamma(q)$, at least locally. More generally, given curve γ' on the same manifold, if we always have a positive value of $m_\gamma(q)$ for each $q \in \gamma'$, then γ' is located on one side. For simplicity, in the following discussion, we always assume that the projection p of q onto γ is unique and denote it as $p = p_\gamma(q)$ as a function of q.

3.2. Optimal Boundary

Before describing the behavior of the best γ in a finite sample setting, let us first consider the phenomenon of an optimal boundary.

Condition 3.1 (Optimal boundary conditions). Consider any curve γ on $\mathcal{M}$ that could be potentially used as a boundary, located between γ_1 and γ_2.

1. For any $q \in \gamma$, the projection of q to the two flows, $d_\mathcal{M}(q, \gamma_1)$ and $d_\mathcal{M}(q, \gamma_2)$, is achievable.
2. All γs are smooth and have equal length.

Definition 3.1. A unit-speed curve $\widetilde{\gamma} \in \Gamma$ is called the optimal boundary if it maximizes the integral of additive distance $d_\mathcal{M}(q, \gamma_1) + d_\mathcal{M}(q, \gamma_2)$ over γ. That is,

$$\widetilde{\gamma} = \arg\max_{\gamma \in \Gamma} \int_\gamma \left(d_\mathcal{M}(q, \gamma_1) + d_\mathcal{M}(q, \gamma_2) \right) ds.$$

$$= \arg\max_{\gamma \in \Gamma} \int_\gamma \min \left\{ d_\mathcal{M}(q, \gamma_1), d_\mathcal{M}(q, \gamma_2) \right\} ds. \tag{9}$$

Remark: Our treatment is abstract, in the sense that γ_1 and γ_2 can best represent the mean field of the two classes. In our view, the optimal boundary is concerned only with γ_1 and γ_2. Note that the maximum of (9) is attained when for any $q \in \gamma$, we have $d_\mathcal{M}(q, \gamma_1) = d_\mathcal{M}(q, \gamma_2)$.

Lemma 3.1. Assume that $\mathcal{M} = \mathbb{R}^d$ and let $\widetilde{\gamma}$ be the optimal boundary for γ_1 and γ_2. Denote that for any $q \in \widetilde{\gamma}$, $p_1 = p_{\gamma_1}(q)$ and $p_2 = p_{\gamma_2}(q)$ are the two projections. Let $T_q \widetilde{\gamma}$ be the gradient line of of $\widetilde{\gamma}$ at q, and let L_{svm} be the straight line determined by the SVM for p_1 and p_2. If Condition 3.1 is satisfied, then $\widetilde{\gamma}$ and L_{svm} are locally equivalent; that is, $T_q \widetilde{\gamma}$ coincides with L_{svm}, where L_{svm} is through the point q.

We remark that, to our knowledge, no SVM has been directly defined on manifolds. It is worthwhile to note that

in the Euclidean space the optimal boundary locally reduces to the usual SVM. Lemma 3.1 is proved in Appendix A (supplementary materials).

3.3. Principal Boundary

Suppose that γ_1 and γ_2 are determined from the two datasets $\{x_{1,i}\}$ and $\{x_{2,j}\}$, respectively. We say γ_1 and γ_2 are separated if there is a curve γ such that γ_1 and γ_2 are on different sides of γ, conditioning on the margins to the two curves; that is, $m_{\gamma_1}(q) > 0$ and $m_{\gamma_2}(q) > 0$ for all $q \in \gamma$. Clearly, such a curve γ can correctly classify the two data sets. The minimum of $m_{\gamma_1}(q)$ and $m_{\gamma_2}(q)$ is referred as the margin of γ to γ_1 and γ_2 at q, and let

$$m_{\gamma_1,\gamma_2}(q) = \min\{m_{\gamma_1}(q),\, m_{\gamma_2}(q)\}.$$

Let Γ be the set of the classification curve with unit speed. A good classification curve should have the margin $m_{\gamma_1,\gamma_2}(q)$ to the two principal flows, as large as possible at each $p \in \gamma$. We call the ideal classification curve that has the largest value of $m_{\gamma_1,\gamma_2}(q)$ the *principal boundary* of the two datasets.

Definition 3.2. A unit-speed curve $\gamma_c \in \Gamma$ is called the principal boundary if it maximizes the integral of the margin $m_{\gamma_1,\gamma_2}(q)$ over γ. That is,

$$\gamma_c = \arg\max_{\gamma \in \Gamma} \int_\gamma \min\{m_{\gamma_1}(\gamma), m_{\gamma_2}(\gamma)\} ds.$$

Remark: Naturally, given the margins, we may define two smooth margin flows, γ_1' and γ_2', such that for any $q_1 \in \gamma_1'$, there is a $q \in \gamma_c$ satisfying $d_{\mathcal{M}}(q, q_1) = m_{\gamma_1}(q)$. Likewise, the same applies for γ_2'. Taking the viewpoint of Definition 3.1, the principal boundary γ_c is equivalent to the optimal boundary of γ_1' and γ_2'. The proof is obvious, so we omit it.

Definition 3.2 seems to suggest optimizing the ideal principal boundary that separates two principal flows by maximizing the margin over a class of flows Γ. Theoretically, the class set Γ contains all the curves γ that lie in between γ_1 and γ_2. This is a much broader class than necessary to sort out a boundary from a classification point of view. Our greater interest lies in achieving such a γ_c over a smaller set, which is more likely to be accessible. Ordinarily, to find such a set one may consider the alignment between the target principal boundary and the principal flows. Hence, we will restrict our attention to a class set Γ', where the correspondence among γ and γ_1 or γ_2 can be explained by a corresponding geodesic.

Assume the projections $p_1 = p_{\gamma_1}(q)$ and $p_2 = p_{\gamma_2}(q)$ are one-to-one for $q \in \gamma_c$; that is, a different q yields a different projection onto γ_1 or γ_2. Hence, the geodesic curve $\mathcal{C}(p_1, p_2)$ between its two projections p_1 and p_2 must go across the principal boundary γ_c at the original q. The maximization in Definition 3.2 means that $q \in \gamma_c$ should also be the middle point of the geodesic curve, that is,

$$q = \arg\max_{q' \in \mathcal{C}(p_1,p_2)} m_{\gamma_1,\gamma_2}(q'),$$

if both p_1 and p_2 have been predetermined. Hence, we can equivalently define the principal boundary in a more direct way, as follows.

Definition 3.3. A curve γ_c on $\mathcal{M}$ is called the principal boundary of two principal flows γ_1 and γ_2, if any $q \in \gamma_c$ satisfies

1. the geodesic curve $\mathcal{C}(p_1, p_2)$ between its two projections $p_1 = p_{\gamma_1}(q)$ and $p_2 = p_{\gamma_2}(q)$ onto γ_1 and γ_2 also contains the point q, and
2. $m_{\gamma_1,\gamma_2}(q) = \max_{q' \in \mathcal{C}(p_1,p_2)} m_{\gamma_1,\gamma_2}(q')$.

3.4. Parameterizations

The condition (2) is equivalent to $m_{\gamma_1}(q) = m_{\gamma_2}(q)$ for any $q \in \mathcal{C}(p_1, p_2)$. Now let us discuss how to obtain such a $\gamma_c = \{q(t) : t \in [0, T]\}$ as a parameterized flow

$$\dot{q}(t) = v(t), \quad t \in [0, T],$$

starting at an initial point $q(0)$ that satisfies the condition (2) in Definition 3.3. The tangent vector $v(t)$ will be carefully chosen, as follows.

Since the projections $p_i(t) = p_{\gamma_i}(q(t))$ can be parameterized as

$$p_1(t) = p_{\gamma_1}(q(t)), \quad p_2(t) = p_{\gamma_2}(q(t))$$

the principal flows can also be parameterized as

$$\gamma_1 = \{p_1(t),\, t \in [0, T]\}, \quad \gamma_2 = \{p_2(t),\, t \in [0, T]\}.$$

Hence, we are equipped with two tangent vectors $v_1(t) = \dot{p}_1(t)$ and $v_2(t) = \dot{p}_2(t)$ at $p_1(t)$ and $p_2(t)$, respectively. Numerically, the tangent vector $v_1(t)$ or $v_2(t)$ can be estimated by the vector field $W(p_1(t))$ of γ_1 at $p_1(t)$ or the vector field $W(p_2(t))$ of γ_2 at $p_2(t)$.

The two tangent vectors $v_1(t)$ and $v_2(t)$ (or their estimates) may not necessarily lie on the same tangent plane at $q(t)$ of γ_c. A natural solution is to move the tangent vectors toward the tangent plane at $q(t)$ under a parallel transport along the geodesic curves. Let $\tilde{v}_1(t)$ and $\tilde{v}_2(t)$ be the transported tangent vectors of $v_1(t)$ and $v_2(t)$, respectively. In Appendix B (supplementary materials), we provide the details of the machinery *Schild's ladder* for an approximate implementation of the parallel transport. As soon as the parallel transport is done at the current $q(t)$, the choice of $v(t)$ is 2-fold: (1) if $v(t)$ lies in the plane spanned by $\tilde{v}_1(t)$ and $\tilde{v}_2(t)$, then there is a $\lambda(t)$ satisfying the equation $v(t) = \lambda(t)\tilde{v}_1(t) + (1 - \lambda(t))\tilde{v}_2(t)$; (2) otherwise, $v(t) \approx \lambda(t)\tilde{v}_1(t) + (1 - \lambda(t))\tilde{v}_2(t)$ with

$$\lambda(t) = \arg\min_\lambda \|v(t) - (\lambda\tilde{v}_1(t) + (1 - \lambda)\tilde{v}_2(t))\|_2,$$

where $\|\cdot\|_2$ is the L_2 norm.

Although the above discussion does not immediately yield an implementable estimation of the true vector $v(t)$ ($\lambda(t)$ is not available), it gives an updating rule of estimating $v(t)$, as follows. Prior to $v(t)$, we choose $v(t - \delta)$ with a small $\delta > 0$ and estimate $\lambda(t)$ as

$$\lambda_\delta(t) = \arg\min_\lambda \|v(t - \delta) - (\lambda\tilde{v}_1(t) + (1 - \lambda)\tilde{v}_2(t))\|_2. \quad (10)$$

Then, we check if the estimate $v_\delta(t) = \lambda_\delta(t)\tilde{v}_1(t) + (1 - \lambda_\delta(t))\tilde{v}_2(t)$ is acceptable by testing whether $q_\delta(t)$, the projection of $q(t - \delta) + \delta v_\delta(t)$ onto the manifold, satisfies the conditions in Definition 3.3 under a given accuracy. If this is not the case, we

slightly tune $\lambda_\delta(t)$ to $\lambda_\delta(t) := \lambda_\delta(t) \pm \epsilon$, and check again until convergence. The updating is

$$\lambda_\delta(t) := \begin{cases} \lambda_\delta(t) + \epsilon & \text{if } m_{\gamma_1}(q_\delta(t)) < m_{\gamma_2}(q_\delta(t)); \\ \lambda_\delta(t) - \epsilon & \text{otherwise.} \end{cases} \quad (11)$$

Initially, when the tangent vector $v(0 - \delta)$ is not available for determining the initial $\lambda_\delta(0)$, we can simply choose $\lambda_\delta(0) = 1/2$. In the next section, we will present a detailed algorithm for computing the principal boundary.

4. Algorithm

In practice, finding the principal boundary can be more challenging than finding a principal flow, in the sense that the former problem is more attached to the picking of the points on the boundary. Recalling that for a point $q \in \gamma_c$,

$$\min\{m_{\gamma_1}(q), m_{\gamma_2}(q)\}$$
$$= \frac{1}{2}\{|m_{\gamma_1}(q) - m_{\gamma_2}(q)| + m_{\gamma_1}(q) + m_{\gamma_2}(q)\}$$

the minimum is achieved if and only if $|m_{\gamma_1}(q) - m_{\gamma_2}(q)|$ is as small as possible. However, one cannot simply identify a sequence of such q's between the two flows to form an approximate boundary. The main reason is that we require the boundary to be a smooth curve. In this respect, we need a much more sophisticated mechanism to guarantee an equal margin on both sides of the boundary, including the choice of the initial point, whereas in the case of principal flow, the picking of the initial point can be very flexible. This is particularly true when the margins differ significantly between the two classes. In these cases, picking the mean or any symmetry of the data points is no longer meaningful. To facilitate the process of generating the boundary, we will require an initial point and then a process of fine-tuning the vector field along the way, which essentially has a direct impact on the principal boundary between the two flows.

We will now present a high-level description of the algorithm for computing the boundary (see Figure 2), the core of which is elaborated as follows.

Step 1 (*Initializing the boundary*): The initialization involves finding a matching pair on γ_1 and γ_2, and calculating an initial point $q(0)$. Arbitrarily choose a point $c \in \gamma_2$, and let $p'_0 = p_{\gamma_1}(c)$ be the projection of c onto γ_1. Consider the geodesic curve $g(c, p'_0)$ between c and p'_0. Obviously, for any point $q \in g(c, p'_0)$, we always have $p'_0 = p_{\gamma_1}(q)$. Let $p''_0 = p_{\gamma_2}(q)$ be the projection of q onto γ_2. Hence, identify $q_0 \in g(c, p'_0)$ such that,

$$d_{\mathcal{M}}(q_0, \gamma_1) - \sigma_{\gamma_1}(p'_0) = d_{\mathcal{M}}(q_0, \gamma_2) - \sigma_{\gamma_2}(p''_0).$$

Here, we call q_0 a warm start. Pick a matching pair as follows

$$p_1(0) = p'_0, \quad p_2(0) = p''_0.$$

Then we can identify a point $q(0) \in g(p'_0, p''_0)$ such that $m_{\gamma_1}(q(0)) = m_{\gamma_2}(q(0))$. Obviously, $q(0)$ satisfies the conditions in Definition 3.3.

Step 2 (*Updating the boundary*): Calculating $q(t)$ for $t > 0$ from the previous point $q(t - \delta)$ with a small $\delta > 0$. Initially, let $\tilde{v}_1(t) = \dot{p}_1(t)$, $\tilde{v}_2(t) = \dot{p}_2(t)$, and set $\lambda_\delta(t)$ as in (10). Then, estimate $v(t)$ by

$$v_\delta(t) = \lambda_\delta(t)\tilde{v}_1(t) + (1 - \lambda_\delta(t))\tilde{v}_2(t), \quad (12)$$

and find the projection as

$$q_\delta(t) = \exp_{q(t-\delta)}(q(t - \delta) + \delta v_\delta(t)). \quad (13)$$

If $q_\delta(0)$ satisfies the conditions in Definition 3.3, let $q(t) = q_\delta(t)$; otherwise, update $\lambda_\delta(t)$ by (11), and recalculate $v_\delta(t)$ in (12) and $q_\delta(t)$ in (13).

The algorithm will be executed for a period of time and will produce a sequence of $\{q(t)\}$. The constructed sequence is indeed the principal boundary since we always have that for every $q(t) \in \gamma_c$, $p_1(t) = p_{\gamma_1}(q(t))$, and $p_2(t) = p_{\gamma_2}(q(t))$,

$$d_{\mathcal{M}}(q(t), \gamma_1) - \sigma_{\gamma_1}(p_1(t)) = d_{\mathcal{M}}(q(t), \gamma_2) - \sigma_{\gamma_2}(p_2(t)).$$

Supposing we may discretize the γ_c as $\gamma_c = (q(0), \dots, q(N))$, $q(i) \in \mathcal{M}$. The length of the principal boundary can be numerically approximated

$$\ell(\gamma_c) \approx \sum_{i=0}^{N-1} d_{\mathcal{M}}(q(i), q(i+1)).$$

5. Property of the Principal Boundary

5.1. Principal Boundary and SVM

This section shows that the local segment of the principal boundary reduces to the boundary given by the SVM. We remark here that the SVM boundary used is a manifold extension of the usual SVM, which is essentially a geodesic curve. The same results hold in the context of Euclidean spaces, where $\mathcal{M}$ is a linear subspace of $\mathbb{R}^d$. By making the notion of "local equivalence" precise, we provide a measure of distance between the principal boundary, obeying Definition 3.3 and the SVM boundary.

To study the relation, we start with a quantitative description of the segment of γ_c on $\mathcal{M}$ by the following proposition.

Proposition 5.1 (Local configuration). Consider a small segment γ_c^* of γ_c on $\mathcal{M}$. Suppose that (Figure 3)

1. The segment is discredited as $\gamma_c^* = \{q_1^*, \dots, q_M^*\}$, where $q_j^* \in \mathcal{M}$.
2. Following the notation in Section 3.1, let $\gamma_1^* = \{p_{1,1}^*, \dots, p_{1,M}^*\}$ be the projections of γ_c^* onto γ_1 and $\mathcal{N}(p_{1,j}^*, h)$ be the set of samples in the h-neighborhood of $p_{1,j}^*$. Clearly, the M local neighborhoods give a configuration of local data points from class $+1$ as

$$\cup_{j=1}^M \mathcal{N}(p_{1,j}^*, h) = \{x_{1,1}, \dots, x_{1,k_1}\}.$$

Similarly, we can define $\{x_{2,1}, \dots, x_{2,k_2}\}$ for the other class -1 based on the projections $\gamma_2^* = \{p_{2,1}^*, \dots, p_{2,M}^*\}$ of γ_c^* onto γ_2.

If γ_c^* is small enough, it is approximately a segment of a straight line L_c^*, and q_j^*s are also approximately located at the line. The SVM on the two classes determines a geodesic curve L_{svm} that separates the two sets γ_1^* and γ_2^* such that the margin of L_{svm}

$$m(L_{\text{svm}}) = \min\left\{\min_k d_{\mathcal{M}}(x_{1,k}, L_{\text{svm}}), \min_k d_{\mathcal{M}}(x_{2,k}, L_{\text{svm}})\right\}$$
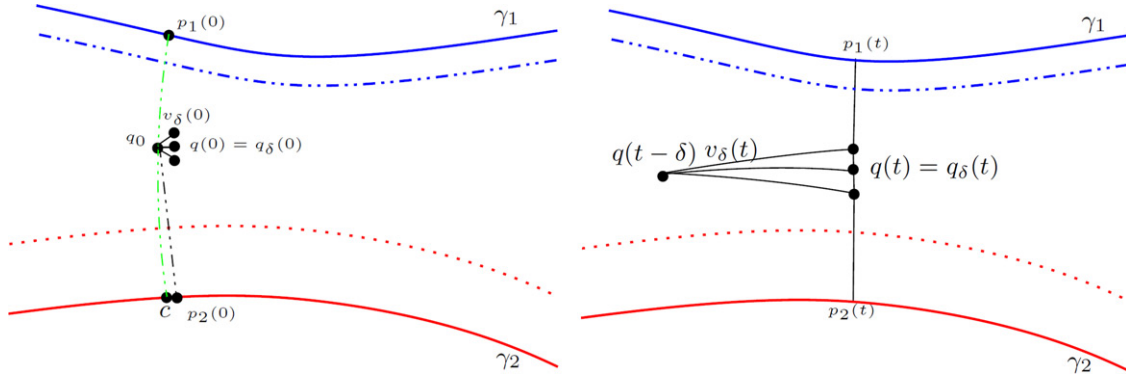
Figure 2. Algorithm. (a) Step 1 (Initializing the boundary): q_0 is the warm start for finding the matching pair $p_1(0)$ and $p_2(0)$; $q(0)$ is the initial point chosen from the projection $q_\delta(0) = \exp_{q_0}(q_0 + \delta v_\delta(0))$ satisfying the conditions of Definition 3.3, via alternating $v_\delta(0)$. (b) Step 2 (Updat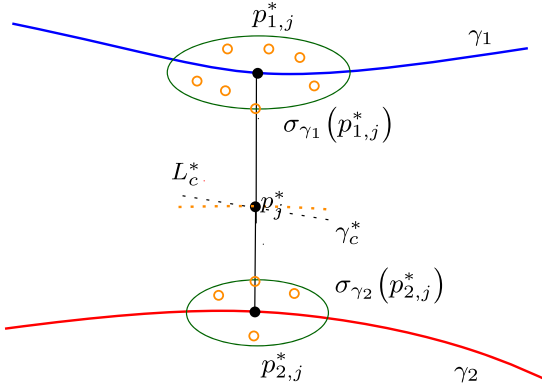ing the boundary): $q(t-\delta)$ is used to find $q(t)$; $q(t)$ is chosen from the projection $q_\delta(t) = \exp_{q(t-\delta)}(q(t-\delta) + \delta v_\delta(t))$ satisfying the conditions of Definition 3.3, via alternating $v_\delta(t)$.

Figure 3. The principal boundary γ_c^* in the local configuration. The covering ellipse ball (in green) with the second radius $\sigma_{\gamma_1}(p_{1,j}^*)$, centered at $p_{1,j}^*$ contains the local points (in "$\circ$") for class $+1$; the shortest distance between q_j^* and the corresponding local configuration of class $+1$ is approximated by $d_{\mathcal{M}}(p_{1,j}^*, q_j^*) - \sigma_{\gamma_1}(p_{1,j}^*)$ (same for class -1).

is maximized. For γ_c^*, we define the margin as

$$m(\gamma_c^*) = \min \left\{ \min_k d_{\mathcal{M}}(x_{1,k}, \gamma_c^*), \min_k d_{\mathcal{M}}(x_{2,k}, \gamma_c^*) \right\}.$$

To quantify the relation of $m(\gamma_c^*)$ and $m(L_{\text{svm}})$, we will basically need three approximations. We remark here that although a careful approximation of γ_c^* to L_{svm} can be bounded under some error assumptions, together with an estimation of $m(\gamma_c^*) \approx m(L_c^*)$, doing so would result in a complicated analysis. To simplify our discussion, we will first sketch an overall review of the local equivalence while highlighting the idea. A refined approximation follows, in terms of Theorem 1.

Let us consider samples in each neighbor set $\mathcal{N}(p_{1,j}^*, h)$. Assume that these neighbors are covered by an ellipse ball of the second radius being $\sigma_{\gamma_1}(p_{1,j}^*)$, centered at $p_{1,j}^*$. Since $m_{\gamma_1}(q_j^*) = d_{\mathcal{M}}(p_{1,j}^*, q_j^*) - \sigma_{\gamma_1}(p_{1,j}^*)$ the quantity

$$\min_{x_{1,k} \in \mathcal{N}(p_{1,j}^*, h)} d_{\mathcal{M}}(x_{1,k}, q_j^*) = m_{\gamma_1}(q_j^*) \tag{14}$$

should approximately hold, up to some degree. Hence,

$$\min_k d_{\mathcal{M}}(x_{1,k}, \gamma_c^*) = \min_j \min_{x_{1,k} \in \mathcal{N}(p_{1,j}^*, h)} d_{\mathcal{M}}(x_{1,k}, q_j^*)$$
$$\approx \min_j m_{\gamma_1}(q_j^*).$$

Similarly, we also have that $\min_k d_{\mathcal{M}}(x_{2,k}, \gamma_c^*) \approx \min_j m_{\gamma_2}(q_j^*)$. Therefore, we have

$$m(\gamma_c^*) \approx \min_j \left\{ m_{\gamma_1}(q_j^*), m_{\gamma_2}(q_j^*) \right\} = \min_j m_{\gamma_1, \gamma_2}(q_j^*).$$

On the other hand, let q_j' be the intersected point of the geodesic curve $\mathcal{C}(p_{1,j}^*, p_{2,j}^*)$ and the straight line L_{svm}. We also have the approximations that

$$\min_{x_{1,k} \in \mathcal{N}(p_{1,j}^*, h)} d_{\mathcal{M}}(x_{1,k}, L_{\text{svm}}) \approx d_{\mathcal{M}}(p_{1,j}^*, q_j') - \sigma_{\gamma_1}(p_{1,j}^*)$$
$$= m_{\gamma_1}(q_j'), \tag{15}$$
$$\min_{x_{2,k} \in \mathcal{N}(p_{2,j}^*, h)} d_{\mathcal{M}}(x_{2,k}, L_{\text{svm}}) \approx d_{\mathcal{M}}(p_{2,j}^*, q_j') - \sigma_{\gamma_2}(p_{2,j}^*)$$
$$= m_{\gamma_2}(q_j'). \tag{16}$$

It follows that the SVM margin

$$m(L_{\text{svm}}) = \min_j \min \left\{ \min_{x_{1,k} \in \mathcal{N}(p_{1,j}^*, h)} d_{\mathcal{M}}(x_{1,k}, L_{\text{svm}}), \right.$$
$$\left. \min_{x_{2,k} \in \mathcal{N}(p_{2,j}^*, h)} d_{\mathcal{M}}(x_{2,k}, L_{\text{svm}}) \right\}$$
$$\approx \min_j \min \left\{ m_{\gamma_1}(q_j'), m_{\gamma_2}(q_j') \right\}$$
$$= \min_j m_{\gamma_1, \gamma_2}(q_j')$$
$$\leq \min_j \max_{q \in \mathcal{C}(p_{1,j}^*, p_{2,j}^*)} m_{\gamma_1, \gamma_2}(q)$$
$$= \min_j m_{\gamma_1, \gamma_2}(q_j^*) = m(\gamma_c^*).$$

Let L_c^* be a geodesic curve most close to γ_c^*. Then $m(\gamma_c^*) \approx m(L_c^*)$, and

$$m(L_{\text{svm}}) \lesssim m(\gamma_c^*) \approx m(L_c^*) \leq m(L_{\text{svm}})$$

by definition, which suggests that L_{svm} approximately coincides with γ_c^*.

Obviously, it can be seen that the local equivalence holds if Approximation (14), plus (15) (or (16)), are satisfactory. We introduce the following condition to guarantee the approximations up to some quantitative degree of uncertainty. This is done by linking the density of sample points in a local neighbor with a probability measure.

Condition 5.1 (Covering ellipse ball). For each $p^*_{1,j}$, consider samples in each neighbor set $\mathcal{N}(p^*_{1,j}, h)$. We assume that,

1. $\mathcal{N}(p^*_{1,j}, h)$ has $k_{1,j}$ neighbors that are covered by an ellipse ball of the second radius $\sigma_{\gamma_1}(p^*_{1,j})$, centered at $p^*_{1,j}$,

2. when $k_{1,j} \to \infty$, with probability of at least $1 - o(\frac{1}{\sqrt{k_{1,j}}})$

$$\left| \min_{x_{1,k} \in \mathcal{N}(p^*_{1,j}, h)} d_{\mathcal{M}}(x_{1,k}, \gamma^*_c) - m_{\gamma_1}(q^*_j) \right| \leq \sqrt{\frac{\log k_{1,j}}{k_{1,j}}};$$

$$\left| \min_{x_{1,k} \in \mathcal{N}(p^*_{1,j}, h)} d_{\mathcal{M}}(x_{1,k}, L_{\text{svm}}) - m_{\gamma_1}(q'_j) \right| \leq \sqrt{\frac{\log k_{1,j}}{k_{1,j}}},$$

3. similar conditions apply to class -1.

Theorem 1. Let γ^*_c and L_{svm} be the separating boundary between the local samples of two classes, derived by the principal boundary and SVM, respectively. Let $k = \min_{i,j} k_{i,j}$ where $i = 1, 2$ and $j = 1, \ldots, M$. Given Proposition 5.1 and Condition 5.1, γ^*_c and L_{svm} are equivalent such that $m(\gamma^*_c) = m(L_{\text{svm}})$, with probability of at least $1 - o(\frac{1}{\sqrt{k}})$, for $k \to \infty$.

Theorem 1 gives an equivalence of γ^*_c and L_{svm} on the curved manifold. The proof of Theorem 1 is given in Appendix A (supplementary materials). Although we have potentially linked L_{svm} with γ^*_c with an interest of interpreting them locally, it does not necessarily mean that Theorem 1 is only valid when the locality is infinitesimal. Instead, it is governed by the spacings of the segment γ^*_c. In fact, when the locality of γ^*_c is not infinitesimal, the equivalence still holds between L^*_c and L_{svm}, provided that L^*_c is close to γ^*_c. In case of a flat manifold, the results still hold, where both L_{svm} and γ^*_c are reduced to straight lines, and γ^*_c is a curve close to them.

5.2. Random Principal Boundary

Definition 5.1 (Random distribution along a curve). Let a unit-speed curve $\gamma^0 = \{q(t) : t \in [0, T]\}$ be the population curve (Figure 4). For simplicity, assume $\dim(\gamma^0) = 1$. The random points along γ^0 are defined by the following procedure: for each $q(t) \in \gamma^0$, we assume that,

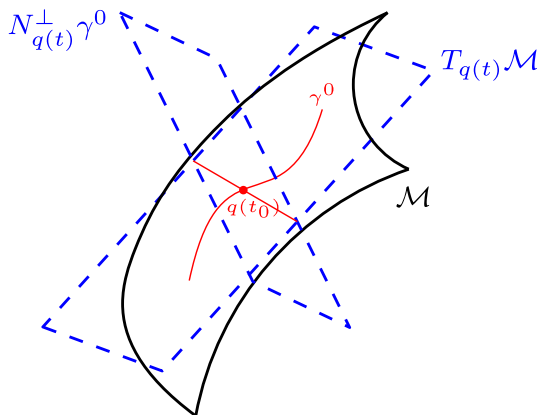1. the normal plane $N^{\perp}_{q(t)}\gamma^0$ exists,

Figure 4. Sketch of random distribution along a curve.

2. $q(t) \sim g$, where g is a uniform distribution on γ^0,
3. the conditional distribution f is well defined on the intersection of the normal plane $N^{\perp}_{q(t)}\gamma^0$ and the tangent plane $T_{q(t)}\mathcal{M}$, with the mean $E(f) = q(t)$ and $\text{cov}(f) = \Sigma_{q(t)}$. For simplicity, let $f \sim N(q(t), \Sigma_{q(t)})$.

Remark 1. Condition 3 allows γ^0 to be extendable to a submanifold ($\dim(\gamma^0) \leq m - 1$). The $\Sigma_{q(t)}$ defined is a $d \times d$ matrix of rank $m - \dim \gamma^0$, that is, $\Sigma_{q(t)}$ is not full rank. Thus, the f defined is degenerate and does not have a density function with respect to the d-dimensional Lebesgue measure in $\mathbb{R}^d$. In the proof, to define the densities and avoid complications, we let $\Sigma_{q(t)}$ be a full rank $(m - \dim(\gamma^0)) \times (m - \dim(\gamma^0))$ matrix in the local coordinates.

Remark 2. The intersection of $N^{\perp}_{q(t)}\gamma^0$ and $T_{q(t)}\mathcal{M}$ is of dimension $m - \dim(\gamma^0)$ in $\mathbb{R}^d$ ($\dim(\gamma^0) = 1$ for principal flow). We assume that γ^0 contains the origin, o, then $T_o\mathcal{M}, N^{\perp}_o\mathcal{M}, T_o\gamma^0$, and $N^{\perp}_o\gamma^0$ are subspaces with $\dim(T_o\mathcal{M}) = m, \dim(N^{\perp}_o\mathcal{M}) = d - m, \dim(T_o\gamma^0) = \dim(\gamma^0)$, and $\dim(N^{\perp}_o\gamma^0) = d - \dim(\gamma^0)$, respectively. Note that $(T_o\mathcal{M} \cap N^{\perp}_o\gamma^0) \oplus N^{\perp}_o\mathcal{M} = N^{\perp}_o\gamma^0$, therefore $\dim(T_o\mathcal{M} \cap N^{\perp}_o\gamma^0) = \dim(N^{\perp}_o\gamma^0) - \dim(N^{\perp}_o\mathcal{M}) = (d - \dim(\gamma^0)) - (d - m) = m - \dim(\gamma^0)$.

Definition 5.2 (Continuous principal flow for a random distribution). Consider the distribution of Definition 5.1 in each neighbor set $D = \mathcal{N}(q(t), h)$, define the continuous principal flow $\tilde{\gamma}^0_f = \tilde{\gamma}^0 = \{\tilde{q}(t) : t \in [0, T]\}$ such that, for each $\tilde{q}(t) \in \tilde{\gamma}^0$,

$$\tilde{q}(t) = E_{g,f}(q),$$

where $q \in D$.

Remark: This definition is for the continuous principal flow with respect to distributions f and g. Although $\dim(\gamma^0)$ is implicitly defined to be 1, Definitions 5.1 and 5.2 still holds for any $\dim(\gamma^0) \leq m - 1$.

Lemma 5.1 (Convergence of continuous principal flow). For a fixed h and t, assume that there exists a one-to-one correspondence between $q(t) \in \gamma^0$ and $\tilde{q}(t) \in \tilde{\gamma}^0$, where γ^0 and $\tilde{\gamma}^0$ are defined in Definitions 5.1 and 5.2 (Figure 5(a)). Then, if we choose $\Sigma_{q(t)}$ such that f is concentrated inside the neighborhood set D with probability of at least $1 - \varepsilon$, then $\|\tilde{q}(t) - q(t)\|_2 = O(\varepsilon)$ for each t.

Remark: The proof in Appendix A (supplementary materials) is given for a two- and three-dimensional manifold. The same proof can be extended up to a $(m - 1)$-dimensional manifold.

Let f be the function that finds the $q_c(t) \in \gamma_c$ given $q_1(t) \in \gamma_1$ and $q_2(t) \in \gamma_2$ and assume that f is Lipschitz continuous with constant L. Assume that $\|\tilde{q}_1(t) - q_1(t)\|_2 = O(\varepsilon_1)$ with probability of at least $1 - \varepsilon_1$ and $\|\tilde{q}_2(t) - q_2(t)\|_2 = O(\varepsilon_2)$ with probability of at least $1 - \varepsilon_2$. Then, with probability of at least $1 - \max(\varepsilon_1, \varepsilon_2)$, $\|\tilde{q}_c(t) - q_c(t)\|_2 = L(O(\varepsilon_1) + O(\varepsilon_2))$ for each t.

Lemma 5.2 (Convergence of continuous principal boundary). Let $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ be the continuous principal flows of two population curves γ_1 and γ_2, respectively (Figure 5(b)). Then, under the same conditions in Definitions 5.1 and 5.2, the continuous

boundary $\tilde{\gamma}_c$ of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ is consistent to the population boundary γ_c of γ_1 and γ_2.

Remark: To fulfill the classification purpose, we need $\dim(\gamma^0) = m - 1$ for the principal boundary. Without loss of generality, we only prove (see Appendix A (supplementary materials)) the case of γ_c being the principal boundary ($m = 2$) with zero margins to γ_1 and γ_2. The proof to nonzero margins would be the same, so we omit it.

6. Simulation

6.1. Principal Boundary and SVM

To illustrate that the principal boundary dependent on the configuration of the datasets, we first generate two sets of noisy data on $S^2 \subset \mathbb{R}^3$. We claim here that choosing the sphere as a test manifold is done only for simplicity, as it stands for the most common manifold that one can work with and compare with other methods. By tuning the vector field along the principal flow of the two datasets, we visualize the evolution of the principal boundary. We remark here that the principal flows are the modified principal flows in (7) for the labeled data. Figure 6 shows the labeled data (Figure 6(a)), the super-imposed principal flows (Figure 6(b)), and the principal boundary (Figure 6(c), (d)). In the step of initializing the boundary, the initial point (in red) has been obtained from the warm start (in blue); both of them are labeled in Figure 6(c). In Figure 6(b), the two green curves are the estimated deviation of the principal flows, providing a measure of the margin of the flow. The boundary enables itself to bend wherever the vector field of the two flows changes. This can be seen from the vectors (in red/black arrows) of Figure 6(d), where they are optimized toward a direction

of achieving the maximum margin between the two flows, a criterion introduced in the step of updating the boundary.

We also contrast the principal boundary with the boundary by the SVM classifier on the same set of data. Among all of the experiments, the principal boundary (in purple) goes through the initial point (in red), maintaining an equal margin between the two labeled datasets. We generate the SVM boundary by performing an SVM classifier with a given kernel function, separating the data in the Euclidean space before mapping it back onto the manifold. Four kernels with suitable parameters have been used in the SVM for training the classification machine: radial basis function (RBF) $K(u, v) = \exp(-\gamma(\|u - v\|^2))$, linear kernel $K(u, v) = u'v$, polynomial kernel $K(u, v) = (\gamma u'v + c_0)^{df}$, and sigmoid kernel $K(u, v) = \tanh(\gamma u'v + c_0)$. Their performances are compared with the principal boundary and the result is reported in Figure 7. The SVM finds the boundary quite differently in all four cases: the best boundary found by the SVM is the one with RBF (in light blue), due to the fact that it is the closest one to the principal boundary, with the exception of part on the left (Figure 7(a)); the linear kernel (in pink) does a decent job in the middle part of the data but clearly could not handle the curvature on the two ends (Figure 7(b)); both the polynomial (in grey) and sigmoid (in orange) kernels fail to achieve a reasonable boundary (Figure 7 (c) or (d)) on the manifold. More experiments on the SVM boundary with varying parameters under different kernels can be found in Appendix B (supplementary materials). This example shows the impact of the resulting SVM boundary from choosing different kernel functions in SVM. If the choice of kernel is done in advance, it can be seen that SVM performs relatively well (i.e., in this case, when the RBF kernel is chosen) compared to the principal boundary. However, very often, one is not given an indication of which kernel to use when performing the classification using
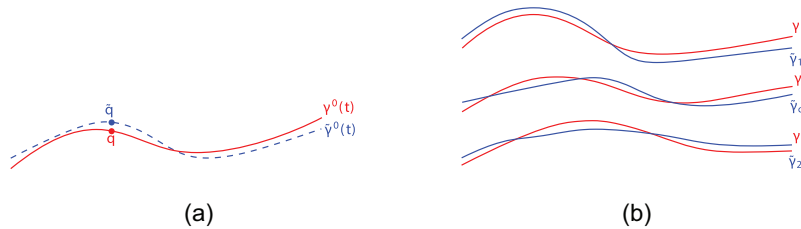

(a)   (b)

Figure 5. Consistency of the principal boundary under random distribution: (a) Population curve (red) and continuous principal flow (blue). (b) Population curves (γ_1, γ_2) and population boundary (γ_c), continuous principal flows ($\tilde{\gamma}_1, \tilde{\gamma}_2$) and continuous boundary ($\tilde{\gamma}_c$).
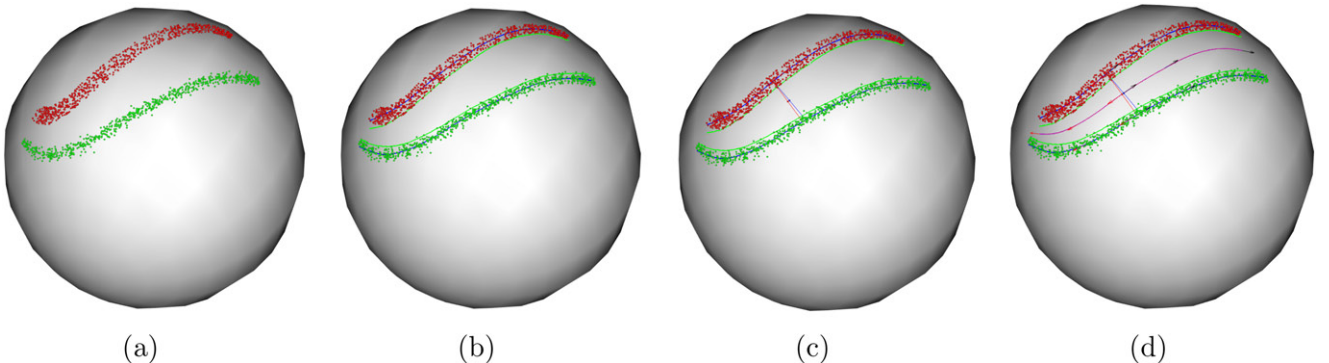

(a)   (b)   (c)   (d)

Figure 6. Plots of the principal boundary: (a) Data points. (b) Principal flow with estimated margin. (c) Initialization of the boundary. (d) Full path of the principal boundary.
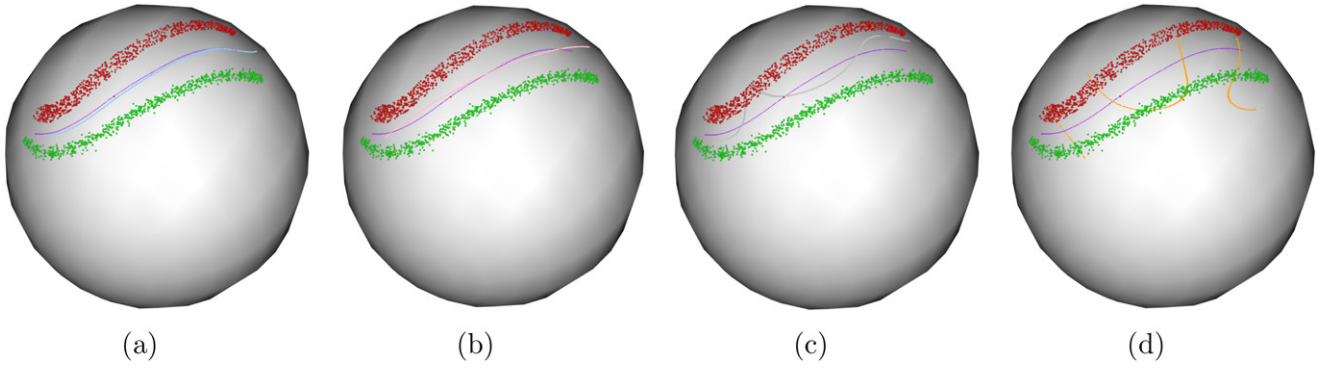
Figure 7. Principal boundary and SVM boundary with different kernel functions: (a) RBF: $K(u, v) = \exp(-\gamma(\|u - v\|^2))$. (b) Linear: $K(u, v) = u'v$. (c) Polynomial: $K(u, v) = (\gamma u'v + c_0)^{\mathsf{df}}$. (d) Sigmoid: $K(u, v) = \tanh(\gamma u'v + c_0)$. All the parameters are set as: bandwidth $\gamma = 1/4$, Lagrange multiplier $C = 1$, df $= 3$, and the coefficient parameter $c_0 = 0$.

the SVM. In this sense, the principal boundary method could be a wise choice when the data clearly has a structure.

Though the SVM boundary above seems to be a reasonable means of understanding the principal boundary, a refined SVM boundary has also been investigated. In line with the setting of local configuration Proposition 5.1, we have obtained a piecewise SVM boundary for this same set of data. This boundary is constructed by performing the SVM process on each pair of the corresponding neighborhoods from the two classes. In this sense, the refined boundary is a local version of the previous SVM boundary. As showed in the proof of Theorem 1, the local SVM is preformed on each paired neighborhood—essentially, it is a geodesic segment or rough line segment; there is no significant difference in choosing the type of kernel. Figure 8 shows the piecewise SVM boundary at each locality parameter $h \in (0.05, 0.1, 0.15, 0.2)$ via linear kernel. It is expected that the piecewise SVM boundary would not necessarily produce a smooth boundary, as we observe that when h increases, the discontinuity of the boundary improves, with a common trend that all SVM segments as a whole are aligned closer to the principal boundary. Although this does not necessarily suggest the existence of an optimal h here, we do see that this behavior matches well with that of the aforementioned Theorem 1.
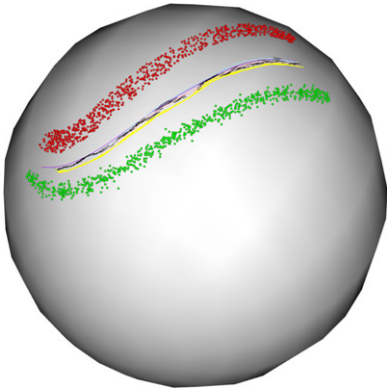
6.2. Principal Boundary in an Overlap Case

To assess the classification performance, we continue to generate data on $S^2 \subset \mathbb{R}^3$ that consists of two classes. The two classes overlap in some regions. There are 500 points in each class. Given two sets, C-type C_1 (in green) and S-type C_2 (in red) in Figure 9, where the points in C_1 are labeled as $+1$ and the points in C_2 are labeled as -1. The goal is to assign the labels of points in the regions near C_1 and C_2.

We find the respective principal flows (in blue) γ_1 and γ_2, and the respective margin flows (in black) $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ ($\tilde{\gamma}_i$ is two-sided along γ_i, $i \in 1, 2$). For a point p on $\mathcal{M}$, we define the distance $d_{\mathcal{M}}(p, C_i)$ of a point p to a set C_i, $i = 1, 2$ as

$$d_{\mathcal{M}}(p, C_i) = m_{\gamma_i}(p).$$

Note that $d_{\mathcal{M}}(p, C_i) = 0$ implies that the point p falls between the two margins of the set C_i, which can be seen to mean that p is inside the set C_i. To classify p, we use the classification rule as follows:

1. If $d_{\mathcal{M}}(p, C_1) < d_{\mathcal{M}}(p, C_2)$, p is classified to C_1; otherwise p is classified to C_2;
2. If $d_{\mathcal{M}}(p, C_1) = d_{\mathcal{M}}(p, C_2) \neq 0$, p is on the classification boundary;
3. If $d_{\mathcal{M}}(p, C_1) = d_{\mathcal{M}}(p, C_2) = 0$, p falls into the overlapping area of sets C_1 and C_2.

Where p falls into the overlapping area of sets C_1 and C_2, we will use the relative gap, the ratio of the geodesic distance

Figure 8. Performance of piecewise SVM boundary and the principal boundary. The piecewise SVM boundary with different h are plotted in black ($h = 0.05$), yellow ($h = 0.1$), pink ($h = 0.15$) and light-blue ($h = 0.2$). The principal boundary (purple) is also superimposed.
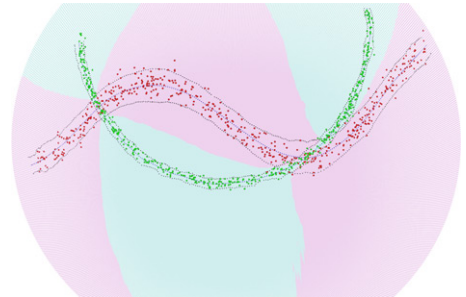

Figure 9. C-S data: C-type points (in green) and S-type points (in red). The principal flows and the two-sided margin flows are plotted in black and blue curves, respectively. The region on S^2 are labeled by pink and light green according to the classification rule.

Table 1. Comparison of principal boundary (column 2) with SVM, random forest and *K*-means (columns 4–6) on classification errors.

	Principal boundary	SVM (R, L, P, S)	Random forest	*K*-means
Error	0.036	0.05, 0.423, 0.075, 0.45	0.039	0.479

NOTES: SVM (R, L, P, S) refers to SVM using RBF, linear, polynomial, and sigmoid kernel.

Table 2. The number of misclassified pairs (m_{vol}, m_{eq}) in the offshore earthquake and volcano data from varying scale parameters h_{vol} and h_{eq}.

$h_{vol}\backslash h_{eq}$	0.020	0.025	0.030	0.035	0.040	0.045	0.050
0.0350	(0,2)	(0,2)	(0,3)	(1,3)	(0,3)	(0,3)	(0,3)
0.0400	(0,3)	(0,3)	(0,3)	(1,3)	(0,3)	(0,3)	(0,3)
0.0450	(0,4)	(0,4)	(0,5)	(0,4)	(1,4)	(1,6)	(1,6)
0.0500	(0,4)	(0,4)	(0,6)	(0,4)	(1,5)	(2,6)	(2,6)
0.0550	(0,4)	(1,5)	(0,8)	(0,6)	(1,6)	(3,6)	(3,6)
0.0600	(0,5)	(1,5)	(0,7)	(0,6)	(0,6)	(2,6)	(2,6)
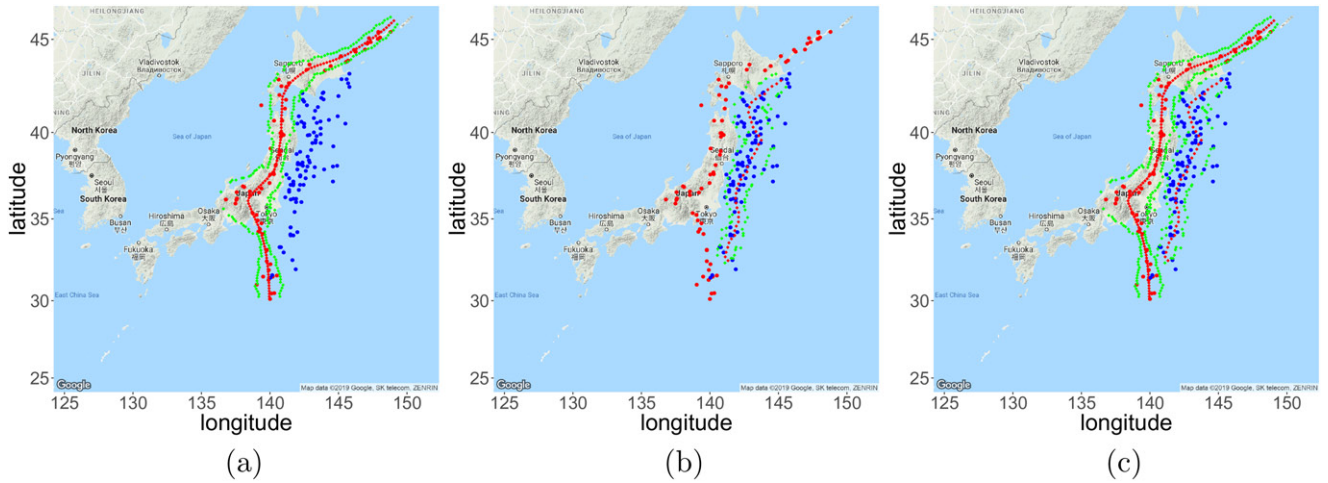
Figure 10. Principal flows (in red) with margins (in green) from the offshore earthquakes ((a), in blue) and volcanoes ((b), in red) data with specific starting points and scale parameters $h_{vol} = 0.035$, $h_{eq} = 0.02$. Misclassified offshore earthquake data ((c), in blue within volcanoes flow margins) from the case with scale parameters $h_{vol} = 0.035$, $h_{eq} = 0.02$.

between p and γ_i to the corresponding margin in C_i, $i = 1, 2$

$$R(p, C_i) = d_{\mathcal{M}}(p, \gamma_i)/\sigma_{\gamma_i}(p). \qquad (17)$$

Note that $R(p, C_i)$ describes the adherences of p to C_i. The larger this ratio is, the less the adherence of p to C_i is. This gives the following classification rule for the overlapping area: p is classified to C_1 when $R(p, C_1) < R(p, C_2)$.

Remark: In general, one can set the ratio as $R(p, C_i) = d_{\mathcal{M}}(p, \gamma_i)^\alpha / \sigma_{\gamma_i}(p)^\beta$. (17) is the case where $\alpha = \beta = 1$. In practice, one can choose to tune α and β within a reasonable range.

We compare the performance of the principal boundary with SVM, random forest and *K*-means. We report an error rate over the C-S data in Table 1. For the principal boundary, the best error rate is calculated by computing the principal flows over $h_1 \in (0.1, 0.25)$ with an increment 0.01 and $h_2 \in (0.05, 0.15)$ with an increment 0.01. The best errors of SVM with different kernels have been computed over the same range of parameters used in Section 6.2. We also include the best error rates from random forest and *K*-means. The supplementary materials contain more details on the implementation procedures for all the methods. In this example, the principal boundary has a much better error rate, even though the dimensionality of the ambient space is not large.

7. Analysis of Seismic Events

We consider a dataset of seismic events involving earthquakes and volcanoes. Volcanoes and earthquakes both result from the movement of tectonic plates and are most likely to occur either on or near plate boundaries. In our study, the volcano dataset contains information about major volcano activity in 2001.[1] The earthquake data contains information about significant earthquakes (measured by damage caused, that is, losses of approximately 1 million US dollars or more, 10 or more deaths, a magnitude of 7.5 or greater, Modified Mercalli Intensity X or greater, or the earthquake generated a tsunami) that happened between 1900 and 2018.[2]

Besides intense volcano activity, Japan is one of the places in the world that is most affected by significant earthquakes. We extract the seismic events that occurred on the eastern side of Japan and consider the significant offshore earthquakes and major volcanoes of eastern Japan. The principal boundary has been applied to the volcano and offshore earthquake dataset, where the sample size of significant offshore earthquake data $n_{eq} = 96$ and the sample size of major volcano data $n_{vol} = 71$,

---

[1]It can be found at *http://legacy.jefferson.kctcs.edu/techcenter/gis%20data/ World/Zip/VOLCANO.zip*.

[2]It is available from the website of National Oceanic and Atmospheric Administration (doi:10.7289/V5TD9V7K).

with two varying sequences of scale parameters h_{vol} and h_{eq}. Table 2 shows the number of misclassified pairs (m_{vol}, m_{eq}) in the offshore earthquake and volcano data from varying h_{vol} and h_{eq}. We calculate the overall error rate $((m_{vol} + m_{eq})/(n_{vol} + n_{eq}))$ and find that the overall error rate achieves its minimum 0.0120, where there are 0 misclassified points for the volcano data and 2 misclassified points for the offshore earthquake data, among all the cases. The supplementary materials contain a similar analysis for the same volcano data and significant earthquakes with magnitude 6 or greater within the same geographic area.

To highlight the results, we plot the principal boundaries with margins from the case with specific scale parameters $h_{vol} = 0.035$, $h_{eq} = 0.02$ in Figure 10(a) and (b). We report that there are 0 misclassified points for the volcano data and two misclassified points for the offshore earthquake data. The misclassified offshore earthquakes are plotted in Figure 10(c). We can see that the principal boundary does a relatively good job on this data, separating the two data clusters well, except for the two instances where the offshore earthquake data is mixed with the volcano data in the south.

8. Conclusion

The classification problem for data points on nonlinear manifolds is a very challenging topic and plays an increasingly important role in real-world problems. With the aim of finding a new method for classifying data points with labels on manifolds, we have proposed a new framework built directly on the manifold. The key techniques we have used involve the tangent space at a given point of the principal flow, which in principle represents the local geometry of the data variation. In other words, the type of local geometric structure we use is the local vector field of the principal flow. We showed the necessity of estimating the margin between the boundary and the flows from a classification perceptive. The combination of fine-tuning the vector fields and the margin provides the alignment of the local data geometry and the global coordinates of the boundary on the manifold. The principal boundary for a finite sample was seen to be interpretable as a local equivalence of the classification boundary by SVM with an analysis. We claim here that although the principal boundary coincides with the SVM locally according to Theorem 1, in practice they appear to be quite different as the basis (kernel) functions that one would use in SVM are usually unknown. We have shown the convergence of the random principal boundary to its continuous counterpart following a probability distribution on the manifold. Examples related to the implemented algorithm of the principal boundary and the numerical comparison with that processed by SVM, random forest and K-means, with the goal of classification, are demonstrated.

The formulation of the principal boundary can be extended to several lines of research. From the theoretical point of view, Condition 5.1 (covering ellipse ball) can essentially be relaxed to suit the needs of the analysis for a finite sample. This, if done with a detailed error analysis, would potentially help us understand the boundary better and improve accuracy. One referee pointed out that the principal boundary may be related to

a kind of medial manifold between the two principal manifolds. Although the boundary is conceptually connected to the medial sheet in Siddiqi and Pizer (2008), the latter is only defined in Euclidean space with a dimension of not more than 3. Moreover, the medial sheet does not necessarily enjoy the smooth property like the boundary does. From the application point of view, this new method has the potential to be a useful tool for real data analysis for manifolds with dimensions greater than 2. In principle, the generalization for higher dimensional manifolds ($m > 2$) is possible by extending both the flows and boundary to high dimension. In practice, the current setting of flows is extendable (Yao, Eltzner, and Pham 2017). Since the submanifolds γ_1 and γ_2 are essentially co-dimension one submanifolds, the extended principal boundary γ_c should be also a submanifold with dimension $m - 1$. Hence, one has to learn the tangent space of γ_c from the tangent space of γ_1 and γ_2. A similar idea applies but the treatment involved will be significantly different from the current one (see Appendix A (supplementary materials)). Certainly, a successful classification also depends on the data configuration and the noise. If the labeled data on the manifold overlaps significantly, one might consider using penalty functions, on top of the adherence used in the current setting. Some of the results in nonparametric regression or machine learning will be helpful in this respect. As this is one of our ongoing works, we will investigate it in the future.

Supplementary Materials

Proofs for main theorems and lemmas, a brief description of the parallel transport implementation, other experiments of SVM boundary with different kernels, additional analysis over C-S data, and more results the analysis of the real data can be found in the online supplementary materials.

Acknowledgments

Funding

References

Belkin, M., Niyogi, P., and Sindhwani, V. (2006), "Manifold Regularization: A Geometric Framework for Learning From Labeled and Unlabeled Examples," *Journal of Machine Learning Research*, 7, 2399–2434. [2]

Bhattacharya, A., and Dunson, D. (2012), "Nonparametric Bayes Classification and Hypothesis Testing on Manifolds," *Journal of Multivariate Analysis*, 111, 369–376. [2]

Donoho, D. L., and Grimes, C. (2003), "Hessian Eigenmaps: New Locally Linear Embedding Techniques for High-Dimensional Data," *Proceedings of the National Academy of Sciences of the United States of America*, 102, 7426–7431. [1]

Eltzner, B., Huckemann, S., Hotz, T., and Mardia, K. (2017), "Torus Principal Component Analysis With Applications to RNA Structure," *Annals of Applied Statistics*, 12, 1332–1359. [2]

Fletcher, P. T., and Joshi, S. (2007), "Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data," *Signal Processing*, 87, 250–262. [2]

Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004), "Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape," *IEEE Transactions on Medical Imaging*, 23, 995–1005. [2]

Fréchet, M. (1948), "Les éléments Aléeatoires de Nature Quelconque Dans un Espace Distancié," *Annales de l'Institut Henri Poincaré*, 10, 215–310. [2]

Gerber, S., Tasdizen, T., Fletcher, P. T., Joshi, S., Whitaker, R., and the Alzheimers Disease Neuroimaging Initiative (ADNI) (2010), "Manifold Modeling for Brain Population Analysis," *Medical Image Analysis*, 14, 643–653. [1]

Hastie, T., and Stuetzle, W. (1989), "Principal Curves," *Journal of the American Statistical Association*, 84, 502–516. [2]

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.), New York: Springer. [1]

Hauberg, S. (2015), "Principal Curves on Riemannian Manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 1915–1921. [2]

Huckemann, S., Hotz, T., and Munk, A. (2010), "Intrinsic Shape Analysis: Geodesic PCA for Riemannian Manifolds Modulo Isometric Lie Group Actions," *Statistica Sinica*, 20, 1–100. [2]

Huckemann, S., and Ziezold, H. (2006), "Principal Component Analysis for Riemannian Manifolds, With an Application to Triangular Shape Spaces," *Advances in Applied Probability*, 38, 299–319. [2]

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568. [2]

Jupp, P. E., and Kent, J. T. (1987), "Fitting Smooth Paths to Spherical Data," *Journal of the Royal Statistical Society*, Series C, 36, 34–36. [2]

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. (1999), *Shape and Shape Theory*, New York: Wiley. [1]

Kenobi, K., Dryden, I. L., and Le, H. (2010), "Shape Curves and Geodesic Modelling," *Biometrika*, 97, 567–584. [2]

Kume, A., Dryden, I. L., and Le, H. (2007), "Shape-Space Smoothing Splines for Planar Landmark Data," *Biometrika*, 94, 513–528. [2]

Ladický, L., and Torr, P. H. (2011), "Locally Linear Support Vector Machines," in *Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA*, pp. 985–992. [2]

Liu, H., Yao, Z., Leung, S., and Chan, T. F. (2017), "A Level Set Based Variational Principal Flow Method for Nonparametric Dimension Reduction on Riemannian Manifolds," *SIAM Journal on Scientific Computing*, 39, 369–376. [2]

Panaretos, V. M., Pham, T., and Yao, Z. (2014), "Principal Flows," *Journal of the American Statistical Association*, 109, 424–436. [2,4]

Patrangenaru, V., and Ellingson, L. (2015), *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*, Boca Raton, FL: CRC Press. [1]

Pennec, X. (2006), "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," *Journal of Mathematical Imaging and Vision*, 25, 127–154. [1]

Pennec, X., and Thirion, J.-P. (1997), "A Framework for Uncertainty and Validation of 3D Registration Methods Based on Points and Frames," *International Journal of Computer Vision*, 25, 203–229. [1]

Roweis, S. T., and Saul, L. K. (2000), "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, 290, 369–376. [1]

Sen, S., Foskey, M., Marron, J., and Styner, M. (2008), "Support Vector Machine for Data on Manifolds: An Application to Image Analysis," in *Proc. Fifth IEEE Int'l Symp. Biomedical Imaging: From Nano to Macro*, pp. 1195–1198. [1]

Siddiqi, K., and Pizer, S. (2008), *Medial Representations—Mathematics, Algorithms and Applications*, Computational Imaging and Vision, Dordrecht: Springer. [13]

Souvenir, R., and Pless, R. (2007), "Image Distance Functions for Manifold Learning," *Image and Vision Computing*, 25, 365–373. [1]

Tuzel, O., Porikli, F., and Meer, P. (2007), "Human Detection via Classification on Riemannian Manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1195–1198. [1]

Yao, Z., Eltzner, B., and Pham, T. (2017), "Principal Sub-manifolds," arXiv no. 1604.04318. [2,13]

Yao, Z., Xia, Y., and Fan, Z. (2019), "Fixed Boundary Flows," arXiv no. 1904.11332. [2]

Zhang, Z., Ladický, L., Torr, P. H. S., and Saffari, A. (2011), "Learning Anchor Planes for Classification," in *Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain*, pp. 1611–1619. [2]

Zhang, Z., and Zha, H. (2004), "Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment," *SIAM Journal on Scientific Computing*, 26, 313–338. [1]

Zhu, X., Ghahramani, Z., and Lafferty, J. (2003), "Semi-supervised Learning Using Gaussian Fields and Harmonic Functions," in *Proceedings of the 20th International Conference on Machine Learning*, pp. 912–919. [2]