Data-based Manifold Reconstruction via Tangent Bundle Manifold Learning

Alexander Bernstein abernstein@hse.ru
National Research University Higher School of Economics (HSE), 109028 Moscow, Russia
Alexander Kuleshov kuleshov@iitp.ru
Kharkevich Institute for Information Transmission Problems RAS (IITP), 127994 Moscow, Russia

Abstract

The goal of Manifold Learning (ML) is to find a description of low-dimensional structure of an unknown q-dimensional manifold embedded in high-dimensional ambient Euclidean space R^p, q < p, from their finite samples. There are a variety of formulations of the problem. The methods of Manifold Approximation (MA) reconstruct (estimate) the manifold but don't find a low-dimensional parameterization on the manifold. The most of Manifold Embedding (ME) methods find a low-dimensional parameterization but don't reconstruct the manifold from a sample. In the paper, the ML is considered as Tangent Bundle Manifold Learning (TBML) problem in which the manifold, its tangent spaces and low-dimensional representation accurately reconstructed from a sample. A new geometrically motivated method for the TBML solution is presented, it also gives a new solution for the MA and ME.

1. Introduction

In general, the goal of Manifold Learning (ML) is to find a description of low-dimensional structure of an unknown q-dimensional manifold **X** embedded in high-dimensional ambient Euclidean space R^p, q < p, from their random samples $\mathbf{X}_n = \{X_1, X_2, \ldots, X_n\}$ (Freedman, 2002). It is usually assumed that **X**, which is further referred to as the Data Manifold (DM), is a 'well-behaved' smooth manifold with positive-radius tube (thus, no self-intersections, no 'short-circuit'), and the DM **X** is 'well-sampled'; this means that the sample size n is sufficiently large. In most studies, the DM is modelled using single coordinate chart. The ML is closely related with Topological Data Analysis problems consisting in finding of the topological structure in data (Carlsson, 2009; Edelsbrunner and Harer, 2010; Balakrishnan et al., 2014).

The term 'to find a description' is not formalized in general, and it has different meaning in different ML-articles. In the articles related to computational geometry this term means 'to approximate (to reconstruct, to estimate) the manifold'. The corresponding Manifold Approximation (MA) problem is as follows: Given a finite dataset $\mathbf{X}_n$ randomly sampled from the DM **X**, to construct (to learn) some set **X*** in R^p that approximates **X** in a suitable sense (Freedman, 2002; Levin, 2003; Kolluri et al., 2004; etc.). The MA-methods are usually based on a decomposition of the DM **X** on small regions (using, for example, Voronoi decomposition or Delaunay triangulation on **X**), and each region is piecewise approximated by some geometrical structure, such as simplicial complex (Freedman, 2002), tangential Delaunay complex (Boissonnat and Ghosh, 2014), finitely many affine subspaces called 'flats' (Karygianni and Frossard, 2012), k-means and k-flats (Canas et al., 2012), etc. In many works, the manifold is locally (piecewise) approximated by its tangent bundle.

However, the MA-methods have a common drawback: they do not find a low-dimensional parameterization on the Data manifold; such parameterization is usually required in the Machine Learning/Data Mining tasks which deal with high-dimensional data.

Data Analysis tasks arising in Machine Learning/Data Mining, such as Pattern Recognition, Classification, Clustering, and other, lead to other formulations of the ML. The Data Analysis algorithms deal with real-world data that are presented in high-dimensional spaces, and the 'curse of dimensionality' phenomena is often an obstacle to the use of many methods for solving these tasks. To avoid these phenomena, various Representation learning algorithms are used as a first key step in solutions of these tasks (Bengio et al., 2012). Representation learning (Feature extraction) algorithms transform the original high-dimensional data into their lower-dimensional representations (or features) so that as much information about the original data required for the considered task is preserved as possible. Representation learning problems that consist in extracting a low-dimensional structure from high-dimensional data can be formulated as various Dimensionality Reduction (DR) problems.

As a rule, high-dimensional real-world data lie on or near some unknown low-dimensional DM embedded in an ambient high-dimensional 'observation' space; in other words, it is satisfied Manifold assumption (Seung and Lee, 2000). The DR under the Manifold assumption are usually referred to as the ML (Cayton, 2005; Narayanan and Mitter, 2010; Ma and Fu, 2011; Rifai et al., 2011; Izenman, 2012; etc.).

The ML problem in Data Analysis is usually formulated as Manifold Embedding (ME) problem, which is as follows: Given dataset $\mathbf{X}_n$ sampled from the DM **X**, construct a low-dimensional parameterization of the DM **X** which produces an Embedding mapping from the DM **X** to an Feature Space (FS) $\mathbf{Y} = h(\mathbf{X}) \subset R^q$, preserving specific geometrical and topological properties of the DM like local data geometry, proximity relations, geodesic distances, angles, etc. Thus, sample features $\mathbf{h}_n = h(\mathbf{X}_n) = \{h_1, h_2, \dots, h_n\}$ must 'faithfully represent' the sample $\mathbf{X}_n$. Note that mapping h must also be defined not only on the sample $\mathbf{X}_n$ but also on new Out-of-Sample (OoS) points $X \in \mathbf{X} / \mathbf{X}_n$.

ME-methods are proposed in many papers, see the surveys (Cayton, 2005; Huo et al., 2007; Ma & Fu, 2011; etc.). The ME methods are also used successfully for solution of the various topological problems, see, for example, (Niyogi et al., 2008; 2011).

ME is usually a first step in various Data Analysis tasks, in which reduced q-dimensional features $y = h(X)$ are used in the reduced learning procedures instead of initial p-dimensional vectors X. If the Embedding mapping h in the ME preserves only specific properties of high-dimensional data, then substantial data losses are possible when using a reduced vector $y = h(X)$ instead of the initial vector X. To prevent these losses, the mapping h must preserve as much available information contained in the high-dimensional data as possible (Lee and Verleysen, 2008; 2009). Thus, it is necessary to find a Reconstruction mapping g from the FS $\mathbf{Y} = h(\mathbf{X})$ to the ambient space R^p with small reconstruction error

$$\delta(X) = |X - g(h(X))|.$$

This possibility is directly required in various Data Analysis tasks such as multidimensional time series prognosis (Chen et al, 2008), data-based approximation of function with high-dimensional inputs, etc. (Bernstein et al., 2011; etc.).

However, the most of popular ME-methods have common drawback: they do not allow reconstructing high-dimensional points X from their low-dimensional features h(X). Thus, it is necessary to formulate ML problem in such a way that its solution does not have the above drawbacks. In other words, corresponding ML procedure must not only find the low-dimensional parameterization of the DM but also reconstruct the high-dimensional manifold points from their low-dimensional features.

In this paper, by the ML we mean a constructing two interrelated mappings h and g from the sample which ensure small reconstruction errors $\delta(X)$ (Bernstein and Kuleshov, 2012; 2013).

The rest of the paper is organized as follows. Section 2 contains a strict definition of the ML in the above statement, which can be formulated as a Manifold Reconstruction problem. Based on some desired properties of the ML solution, in Section 3 we introduce an amplification of the ML, called Tangent Bundle ML problem (TBML), which requires an estimating of tangent spaces to the DM also. The proposed TBML solution is presented in Section 4; some properties of this solution are described shortly in this section.

2. Manifold Reconstruction problem

We will consider the following ML definition: Given a dataset $\mathbf{X}_n$ randomly sampled from the DM $\mathbf{X}$, construct an ML-solution $\theta = (h, g)$ consisting of two interrelated mappings: an Embedding mapping

$$h: \mathbf{X}_h \subset R^p \to R^q, \tag{1}$$

defined on a domain of definition $\mathbf{X}_h \supset \mathbf{X}$, and a Reconstruction mapping

$$g: \mathbf{Y}_g \subset R^q \to R^p, \tag{2}$$

defined on a domain of definition $\mathbf{Y}_g \supset \mathbf{Y}_h = h(\mathbf{X}_h)$, which ensure the approximate equality

$$r_\theta(X) \approx X \quad \text{for all} \quad X \in \mathbf{X}, \tag{3}$$

where $r_\theta(X) \stackrel{\text{def}}{=} g(h(X))$ is the result of successively applying of Embedding and Reconstruction mappings to a vector $X \in \mathbf{X}$.

Note that the ML-solution includes a determination of the domains of definition $\mathbf{X}_h$ and $\mathbf{Y}_g$ also. The Reconstruction error $\delta_\theta(X) = |r_\theta(X) - X|$ is a measure of quality of the ML solution θ at point $X \in \mathbf{X}$.

The solution θ determines a q-dimensional Reconstructed Manifold (RM)

$$\mathbf{X}_\theta = \{X = g(y) \in R^p: y \in \mathbf{Y}_\theta \subset R^q\} \tag{4}$$

embedded in R^p and covered (parameterized) by the single chart g defined on the FS $\mathbf{Y}_\theta = h(\mathbf{X})$. Thus, the approximate equalities (4) can be written in the form

$$\mathbf{X} \approx \mathbf{X}_\theta \equiv r_\theta(\mathbf{X}) \tag{5}$$

and considered as Manifold proximity property.

The above defined ML, in which the ML-solution θ accurately reconstructs the unknown DM $\mathbf{X}$ by the RM $\mathbf{X}_\theta$ and determines the low-dimensional parameterization $y = h(X)$ on the DM $\mathbf{X}$, can be referred to as the Manifold Reconstruction (MR). Note that ME-solution h reconstructs a parameterization on the DM $\mathbf{X}$ only.

There are some (though limited number of) methods for reconstruction of the DM $\mathbf{X}$ from the FS $h(\mathbf{X})$. For a specific linear DM, the reconstruction can be easily made with the Principal Component Analysis (PCA) (Jollie, 2002). For nonlinear DM, the sample-based Auto-Encoder Neural Networks (Kramer, 1991; Hinton and Salakhutdinov, 2006; etc.) determine both the embedding and reconstruction mappings. General method, which reconstructs the DM in the same manner as Locally Linear Embedding (Saul and Roweis, 2000), has been introduced in (Saul and Roweis, 2003). An interpolation-like nonparametric regression reconstruction of the DM is proposed in another embedding method called Local Tangent Space Alignment (Zhang and Zha, 2005).

Above defined MR may be also considered as a Manifold Estimation Problem which is as follows: Given a dataset $\mathbf{X}_n$ randomly sampled from an unknown smooth q-dimensional DM $\mathbf{X}$ in R^p covered by a single chart, estimate the DM. It is natural to evaluate a quality of the estimator $\mathbf{X}_\theta$ (4) (sample-based q-dimensional manifold in R^p also covered by a single chart) by the Hausdorff distance $H(\mathbf{X}_\theta, \mathbf{X})$ between the DM and RM (Genovese et al., 2012); the following relation

$$H(\mathbf{X}_\theta, \mathbf{X}) \le \sup_{X \in \mathbf{X}} \delta_\theta(X). \tag{6}$$

between the qualities of MR and Estimation problems takes a place.

The Reconstruction error $\delta_\theta(X)$ can be directly computed at the sample points $X \in \mathbf{X}_n$; for OoS point X it describes the generalization ability of the considered MR-solution θ at a specific point X. The local lower and upper bounds are obtained for the maximum reconstruction error in a small neighborhood of an arbitrary point $X \in \mathbf{X}$ (Bernstein and Kuleshov, 2013); these bounds are defined in terms of the distance between the tangent spaces L(X) and $L_\theta(r_\theta(X))$ to the DM $\mathbf{X}$ and the RM $\mathbf{X}_\theta$ at the points X and $r_\theta(X)$, respectively. It follows from the bounds that the greater the distances between these tangent spaces, the lower the local generalization ability of the solution θ.

Thus, it is natural to require that the MR-solution ensures not only Manifold proximity (3) but also Tangent proximity

$$L(X) \approx L_\theta(r_\theta(X)) \qquad \text{for all } X \in \mathbf{X} \qquad (7)$$

between these tangent spaces in some selected metric on the Grassmann manifold Grass(p, q) consisting of all the q-dimensional linear subspaces in R^p (the tangent spaces are treated as elements of the Grass(p, q)). The requirement of the Tangent proximity for the MR-solution arises also in various applications in which the MR is an intermediate step for Intelligent Data Analysis problem solution.

A statement of the extended ML problem, which includes an additional requirement of the tangent spaces proximity, has been proposed in (Bernstein and Kuleshov, 2012) and is described in next section.

3. Extended Manifold Reconstruction problem

Before introducing the extended ML, note that the set $T(\mathbf{X}) = \{(X, L(X)): X \in \mathbf{X}\}$ composed of points X of the manifold $\mathbf{X}$ equipped by the tangent spaces L(X) at these points, is known in Topology as tangent bundle of the manifold $\mathbf{X}$. So, proximity between the manifolds and their tangent spaces can be called Tangent bundle proximity, and the amplification of the MR consisting in accurate reconstruction of the tangent bundle $T(\mathbf{X})$ from the sample $\mathbf{X}_n$ may be referred to as the Tangent Bundle Manifold Learning (TBML).

A strict definition of the TBML is as follows: Given dataset $\mathbf{X}_n$ randomly sampled from a q-dimensional DM $\mathbf{X}$ embedded in an ambient p-dimensional space R^p and covered by a single chart, construct a TBML-solution θ = (h, g) (1), (2) which provides both the above defined Manifold proximity (3), (5) and the Tangent proximity (7) where the tangent space $L_\theta(r_\theta(X)) = \mathrm{Span}(J_g(h(X)))$ is spanned by columns of the Jacobian matrix $J_g(y)$ of the mapping g at the point y = h(X).

The TBML-solution θ determines the Reconstructed tangent bundle

$$RT_\theta(\mathbf{X}_\theta) = \{(g(y), \mathrm{Span}(J_g(y))): y \in \mathbf{Y}_\theta\} \qquad (8)$$

of the RM $\mathbf{X}_\theta$, which is close to the $T(\mathbf{X})$, and the q-dimensional submanifold $\mathbf{L}_\theta = \{\mathrm{Span}(J_g(y)): y \in \mathbf{Y}_\theta\}$ of the Grassmann manifold which reconstructs the Tangent Manifold $\mathbf{L} = \{L(X): X \in \mathbf{X}\}$.

4. Reconstruction of tangent bundle of the Data Manifold

Describe TBML-solution called Grassman&Stiefel Eigenmaps (GSE); its early version has been proposed in (Bernstein and Kuleshov, 2012).

The GSE consists of two successively performed main parts: Part I (approximation of the Tangent manifold) and Part II (reconstruction of the DM).

In Part I, the sample-based family $\mathbf{H} = \{H(X), X \in \mathbf{X}\}$ consisting of p×q matrices H(X) smoothly depending on $X \in \mathbf{X}$ is constructed to meet the relations

$$L_H(X) \approx L(X) \qquad \text{for all } X \in \mathbf{X} \qquad (9)$$

here $L_H(X) = \mathrm{Span}(H(X))$ are q-dimensional linear spaces in R^p spanned by columns $\mathbf{H}^{(1)}(X), \mathbf{H}^{(2)}(X), \ldots, \mathbf{H}^{(q)}(X)$ of the matrices H(X).

The mappings h and g will be built in the Part II in such a way as to provide the proximities $g(h(X)) \approx X$ and

$$J_g(h(X)) \approx H(X); \qquad (10)$$

thence, the linear space $L_H(X)$ must be the tangent space to the RM $\mathbf{X}_\theta$, whence comes the Tangent Bundle proximity (5), (7). For possibility of constructing such mappings, the family $\mathbf{H}$ must satisfy the additional property: the tangent vector fields $\mathbf{H}^{(1)}(X), \mathbf{H}^{(2)}(X), \ldots, \mathbf{H}^{(q)}(X) \in L_\theta(r_\theta(X))$ are potential vector fields; thence, they must meet the following relations:

$$\nabla_{\mathbf{H}^{(i)}} \mathbf{H}^{(j)}(X) = \nabla_{\mathbf{H}^{(j)}} \mathbf{H}^{(i)}(X), \; 1 \le i < j \le q, \qquad (11)$$

here $\nabla_{\mathbf{H}}$ denotes a covariant differentiation with respect to the vector field $\mathbf{H}(X) \in L_\theta(r_\theta(X))$.

In Part II, given $\mathbf{L}_H$, the Embedding mapping h(X) is constructed as is follows. Taylor series expansions

$$g(y') - g(y) \approx J_g(y) \times (y' - y) \qquad (12)$$

at near points y and y′, under the desired equalities (3) and (10) for mappings h and g specified further, imply the equalities:

$$X' - X \approx H(X) \times (h(X') - h(X)) \qquad (13)$$

for near points $X, X' \in \mathbf{X}$. Under the already constructed family $\mathbf{H}$, these approximate equalities are used for constructing the embedding mapping h; that, in turn, determines the Feature space $\mathbf{Y}_\theta = h(\mathbf{X})$.

To construct he mapping satisfying the proximities (3) and (10), the p×q matrix G(y) dependent on $y \in \mathbf{Y}_\theta$ is constructed to meet the condition

$$G(h(X)) \approx H(X). \qquad (14)$$

Then, under the already constructed family $\mathbf{H}$, mapping h and matrix G(y), the equations (12) can be written as

$$g(y) \approx X + G(y) \times (y - h(X)). \qquad (15)$$

for near points $y, h(X) \in \mathbf{Y}_\theta$, which are used for constructing the mapping g.

Describe briefly the details of the GSE. Preliminarily, the tangent space L(X) for the points $X \in \mathbf{X}_h$ are estimated by the q-dimensional linear space $L_{PCA}(X)$ which is a result of the PCA applied to sample points from an ε_n-ball in R^p centered at X; the set $\mathbf{X}_h \subset R^p$ consists of the points X in which the q^{th} eigenvalue in PCA is positive.

The data-based kernel K(X, X′), $X', X \in \mathbf{X}_h$, is constructed as a product $K_E(X, X') \times K_G(X, X')$, where K_E is standard Euclidean 'heat' kernel (Belkin and Niyogi, 2002) and $K_G(X, X') = K_{BC}(L_{PCA}(X), L_{PCA}(X'))$ is the Binet-Cauchy kernel (Wolf and Shashua, 2003) on the Grass(p, q); this aggregate kernel reflects not only geometrical nearness between the points X and X′ but also nearness between the linear spaces $L_{PCA}(X)$ and $L_{PCA}(X')$, whence comes a nearness between the tangent spaces L(X) and L(X′).

Approximation of the Tangent manifold. The set $\mathbf{H}_n$ consisting of p×q matrices H_i that meet the constraints $Span(H_i) = L_{PCA}(X_i)$ and satisfy the linear equations are written explicitly, is constructed to minimize the quadratic form

$$\Delta_{H,n}(\mathbf{H}_n) = \frac{1}{2}\sum_{i,j=1}^{n} K(X_i, X_j) \times \|H_i - H_j\|_F^2, \quad (16)$$

under the normalizing condition

$$\sum_{i=1}^{n} K(X_i) \times (H_i^T \times H_i) = I_q$$

required to avoid a degenerate solution, and q(q-1)/2 additional linear conditions, which are obtained from equations (11) by replacing the derivatives in (11) by finite differences, here

$$K(X) = \sum_{j=1}^{n} K(X, X_j) \text{ and } K = \sum_{i=1}^{n} K(X_i).$$

The p×q matrix H(X) for an arbitrary point $X \in \mathbf{X}_h$ is chosen to minimizes the form

$$\Delta_H(H, X) = \sum_{j=1}^{n} K(X, X_j) \times \|H(X) - H_j\|_F^2$$

under the specified linear conditions.

The exact solution of the minimizing problem (16) is obtained as solution of specified generalized eigenvector problems. The matrix H(X) which minimizes the quadratic form $\Delta_H(H, X)$ is written in explicit form.

The cost function (16) is similar to the cost function in the Laplacian Eigenmaps, LE (Belkin and Niyogi, 2002; 2003), but different kernels are used in (16) and LE, and minimization in (16) is over matrices, while minimization in LE is over vectors. The cost function (16) was used in (Goldberg and Ritov, 2009) for aligning the PCA projectors to define a novel measure for a quality of arbitrary Sample embedding.

The problem of estimating the tangent spaces L(X) in the form of a smooth function of the point $X \in \mathbf{X}$ was considered in some previous works. The matrices whose columns approximately span the tangent spaces were constructed using Artificial Neural Networks with one hidden layer (Bengio and Monperrus, 2005) or Radial Basis Functions (Dollár et al., 2006; 2007). Other Persistent Tangent Space Learning method is proposed in (He and Lin, 2011) for constructing the approximations for the tangent spaces, which smoothly varied on the manifold. The constructed linear spaces $\{L_H(X_i)\}$ is the result of an alignment of the PCA-based linear spaces $\{L_{PCA}(X_i)\}$; similar alignment problem was studied in (Zhang and Zha, 2005) with using a cost function, which differs from our cost function (16).

Reconstruction of the DM: **the Embedding mapping.**

First, consider the equations (13) written for near sample points, as regression equations, and compute a preliminary vector set $\mathbf{h}_n = \{h_1, h_2, \dots, h_n\}$ as a standard least squares solution of the regression problem, which minimizes the weighted residual

$$\sum_{i,j=1}^{n} K(X_i, X_j) \times |X_j - X_i - H(X_i) \times (h_j - h_i)|^2$$

under normalizing condition $h_1 + h_2 + \dots + h_n = 0$.

Then, based on $\mathbf{h}_n$, choose a value h(X) for an arbitrary point $X \in \mathbf{X}_h$ by minimizing over h(X) the weighted residual

$$\sum_{j=1}^{n} K(X, X_j) \times |X_j - X - H(X) \times (h_j - h(X))|^2 .$$

Thus, under $\mathbf{h}_n$, the value h(X) for arbitrary point $X \in \mathbf{X}_h$ (including sample points) is written as

$$h(X) = h_{KNR}(X) + v^{-1}(X) \times Q_{PCA}^T(X) \times \tau(X), \quad (17)$$

here $v(X) = Q_{PCA}^T(X) \times H(X)$,

$$\tau(X) = \frac{1}{K(X)}\sum_{j=1}^{n} K(X, X_j) \times (X - X_j), \quad (18)$$

and

$$h_{KNR}(X) = \frac{1}{K(X)}\sum_{j=1}^{n} K(X, X_j) \times h_j \quad (19)$$

is standard Kernel Non-parametric Regression estimator (Wasserman, 2007) for h(X) based on the preliminary values $h_j \in \mathbf{h}_n$ of the vector h(X) at the sample points. Note that the embedding h(X) gives a new solution for the Manifold Embedding problem.

Reconstruction of the DM: **the Reconstruction mapping.**

The data-based kernel k(y, y′) on constructed Feature space $\mathbf{Y}_\theta$ and linear spaces $L^*(y) \in$ Grass(p, q) dependent on $y \in \mathbf{Y}_\theta$ are constructed to provide the equalities

$$k(h(X), h(X')) \approx K(X, X')$$

and

$$L^*(h(X)) \approx L_{PCA}(X)$$

for the points $X \in \mathbf{X}_h$ and $X' \in \mathbf{X}_n$. Denote $\pi^*(y)$ the projector onto the linear space $L^*(y)$, and $\mathbf{Y}_n$ the let dataset consisting of the the features $\{y_i = h(X_i)\}$.

First, we construct the p×q matrix G(y) to meet the condition (14). For this we choose the value G(y) for arbitrary point $y \in \mathbf{Y}_\theta$ by minimizing the form

$$\Delta_G(G, y) = \frac{1}{2}\sum_{j=1}^{n} k(y, y_j) \times \|G - H(X_j)\|_F^2$$

under the constraint Span(G(y)) = L*(y). A solution of this problem in explicit form is as follows:

$$G(y) = \pi^*(y) \times \frac{1}{k(y)}\sum_{j=1}^{n} k(y, y_j) \times H(X_j),$$

here $k(y) = \sum_{j=1}^{n} k(y, y_j)$.

To provide the desired equalities (3) and (10), the function g(y) is chosen to minimize the quadratic form

$$\sum_{j=1}^{n} k(y, y_j) \times \|X_j - g - G(y) \times (y_j - y)\|_F^2$$

under the condition $J_g(y) = G(y)$.

A solution of this problem in explicit form is as follows:

$$g(y) = g_{KNR}(y) + G(y) \times (y - \frac{1}{K(y)}\sum_{j=1}^{n} k(y, y_j) \times y_j), \quad (20)$$

here

$$g_{KNR}(y) = \frac{1}{k(y)}\sum_{j=1}^{n} k(y, y_j) \times X_j , \quad (21)$$

is standard Kernel Non-parametric Regression estimator for g(y) based on the values $X_j \in \mathbf{X}_n$ of the vector g(y) at the sample features $y_j \in \mathbf{Y}_n$. Note, that the relation $G(y) = J_g(y)$ is follows from (20).

The constructed mappings h(X) (17) – (19) and g(y) (20), (21) give a new solution for the MR and ensure the Tangent bundle proximity. Furthermore, under asymptotic $n \to \infty$ and appropriate choice of ball radius $\varepsilon_n \sim O(n^{-1/(q+2)})$, the rate in Tangent Bundle proximity is (Kuleshov et al., 2013) is

$$|X - r_\theta(X)| = O(n^{-2/(q+2)}), \quad (22)$$

$$d_{P,2}(L(X), L_\theta(r_\theta(X))) = O(n^{-1/(q+2)}); \quad (23)$$

it means that the events (22), (23) hold true with high probability, i.e., the probability of these events exceeds the value $(1 - C_\alpha / n^\alpha)$ for any n and $\alpha > 0$, where constant C_α depends only on α; here $d_{P,2}$ is the projection 2-norm metric on the Grassmann manifold Grass(p, q) (Wang, 2006).

The rate in (22) coincides with the asymptotically minimax lower bound for Hausdorff distance between the DM and RM, which was set out in (Genovese et al., 2012). It follows from (22) and (6) that the RM $\mathbf{X}_\theta$ estimates the DM **X** with optimal rate of convergence.

The rate (23) for a deviation of the estimator $L_{PCA}(X)$ from the tangent space L(X) at the reference point X is known (Singer and Wu, 2012; Tyagi et al., 2013).

Acknowledgments

This work is partially supported by RFBR, the research projects 13-01-12447 and 13-07-12111.

References

Balakrishnan, S., Fasy, B.T., Lecci F., Rinaldo, A, Singh, A., Wasserman, L. *Statistical Inference for Persistence homology*. In arXiv preprint: arXiv:1303.7117v12 [math.ST], 28 February, 2014.

Belkin, M., Niyogi, P. *Laplacian eigenmaps and Spectral Techniques for Embedding and Clustering*. In T.G. Dietterich, S. Becker and Z. Ghahramani (eds.), Advances in Neural Information Processing Systems 14, Cambridge, MA: MIT Press, 2002, 585 - 591.

Belkin, M., Niyogi, P. *Laplacian eigenmaps for dimensionality reduction and data representation.* Neural Computation, 2003, 15: 1373 – 1396.

Bengio, Y., Monperrus, M. *Non-local manifold tangent learning*. In Saul L., Weiss, Y., and Bottou, L. (eds), Advances in Neural Information Processing Systems, 17, 129 - 136. Cambridge, MA: MIT Press, 2005.

Bengio, Yoshua, Courville, Aaron, Vincent, Pascal. *Representation Learning: A Review and New Perspectives.* In arXiv preprint: arXiv:1206.5538v2 [cs. LG], 18 Oct. 2012.

Bernstein, A.V., Burnaev, E.V., Chernova, S.S., Zhu, F., Qin, N. *Comparison of Three Geometric Parameterization methods and Their Effect on Aerodynamic Optimization.* In: Proceedings of International Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems (Eurogen 2011, Capua, Italy, September 14 – 16), 2011.

Bernstein, A.V., Kuleshov, A.P. Tangent Bundle Manifold Learning via Grassmann & Stiefel Eigenmaps. In arXiv preprint: arXiv:1212.6031v1 [cs.LG], December 2012, 1 - 25.

Bernstein, A.V., Kuleshov, A.P. *Manifold Learning: generalizing ability and tangent proximity*. International Journal of Software and Informatics, 2013, 7(3), 359 - 390.

Boissonnat, Jean-Daniel, Ghosh, Arijit. *Manifold reconstruction using tangential Delaunay complexes*. Discrete & Computational Geometry, 51(1), 2014, 221 – 267.

Canas, Guillermo D., Poggio, Tomaso, Rosasco, Lorenzo A. *Learning Manifolds with K-Means and K-Flats.* In: Advances in Neural Information Processing Systems, NIPS 2012.

Carlsson, Gunnar. *Topology and Data*. Bull. Amer. Math. Soc., 46, 2009, 255-308.

Cayton, Lawrence. *Algorithms for manifold learning*. Technical Report, Report No CS2008-0923, Univ of California at San Diego (UCSD): Citeseer, June 2005, 541 – 555.

Chen, J., Deng, S.-J., Huo, X. *Electricity price curve modeling and forecasting by manifold learning*. IEEE Transaction on power systems, 2008, 23(3), 877 – 888.

Dollár, P., Rabaud, V., Belongie, S. *Learning to Traverse Image Manifolds*. In Bernhard Schölkopf, John C. Platt and Thomas Hoffman (eds), Advances in Neural Information Processing Systems, 19. Cambridge, MA: MIT Press, 2006, 361 - 368.

Dollár, P., Rabaud, V., Belongie, S. *Non-Isometric Manifold Learning: Analysis and an Algorithm*. In Zoubin Ghahramani (ed.), Proceedings of the 24th International Conference on Machine Learning, Omni Press, 2007, 241 – 248.

Edelsbrunner, H., Harer, J. *Computational Topology: An Introduction.* Amer Mathematical Society, 2010.

Freedman, D. *Efficient simplicial reconstructions of manifold from their samples.* IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(10), 1349-1357.

Genovese, Christopher R., Perone-Pacifico, Marco, Verdinelli, Isabella, Wasserman, Larry. *Minimax Manifold Estimation*. Journal of Machine Learning Research, 13 (2012), 1263-1291.

Goldberg, Y., Ritov, Y. *Local Procrustes for Manifold Embedding: A Measure of Embedding Quality and Embedding Algorithms*. Journal Machine Learning archive 77(1), October 2009, 1 – 25.

He, X., Lin, B.: *Tangent space learning and generalization*. Frontiers of Electrical and Electronic Engineering in China 6(1), 2011, 27-42.

Hinton, G.E., Salakhutdinov, R.R. *Reducing the dimensionality of data with neural networks*. Science, 2006, 313(5786): 504 – 507.

Huo, Xiaoming, Ni, Xuelei (Sherry), Smith, Andrew K. *A Survey of Manifold-based Learning Methods*. In: T. W. Liao and E. Triantaphyllou, eds. Recent Advances in Data Mining of Enterprise Data, Singapore: World Scientific, 2007. 691 – 745.

Izenman, Alan Julian. *Introduction to manifold learning*. Computational Statistics, 2012, 4(5), 439 – 446.

Jollie, T., *Principal Component Analysis*. New-York, Springer, 2002.

Karygianni, Sofia, Frossard, Pascal. *Tangent-based manifold approximation with locally linear models*. In: arXiv:1211.1893v1 [cs.LG], 6 Nov. 2012.

Kolluri, Ravikrishna, Shewchuk, Jonathan, Richard, O'Brien, James F. *Spectral surface reconstruction from noisy point clouds.* In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing, SGP'04, New York, NY, USA, ACM, 2004, 11–21.

Kramer, M. *Nonlinear Principal Component Analysis using autoassociative neural networks.* AIChE Journal, 1991, 37(2), 233 – 243.

Kuleshov, A., Bernstein, A., Yanovich, Yu. *Asymptotically optimal method in Manifold estimation.* In: Márkus, L., Prokaj, V. (eds.) Abstracts of the XXIX-th European Meeting of Statisticians, 20-25 July 2013, Budapest, 325.

Lee, John A., Verleysen, Michel. *Quality assessment of dimensionality reduction based on k-ary neighborhoods*. In: Yvan Saeys, Huan Liu, Iñaki Inza, Louis Wehenkel and Yves Van de Peer, eds. JMLR Workshop and Conference Proceedings. Volume 4: New challenges for feature selection in data mining and knowledge discovery, Antwerpen, Belgium, 2008, 21–35.

Lee, John A., Verleysen, Michel. *Quality assessment of dimensionality reduction*: Rank-based criteria. Neurocomputing, 2009, 72(7-9): 1431–1443.

Levin, David. *Mesh-independent surface interpolation*. In: Hamann Brunnett and Mueller, eds., Geometric Modeling for Scientific Visualization. Springer-Verlag, 2003, 37–49.

Ma, Yunquan, Fu, Yun, eds. *Manifold Learning Theory and Applications*. London: CRC Press, 2011.

Narayanan, H. and Mitter, S. *Sample complexity of testing the manifold hypothesis*. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, Advances in Neural Information Processing Systems 23, 2010, 1786–1794.

Niyogi, P., Smale, S., Weinberger, S. *Finding the Homology of Submanifolds with High Confidence from Random Samples.* Discrete and Computational Geometry, 2008, 39, 419 - 441.

Niyogi, P., Smale, S., Weinberger, S. *A Topological View of Unsupervised Learning from Noisy Data*. SIAM Journal on Computing, 40(3), 2011, 646 – 663.

Rifai, Salah, Dauphin, Yann N., Vincent, Pascal, Bengio, Yoshua, Muller, Xavier. *The manifold Tangent Classifier.* In: J. Shawe-Taylor and R.S. Zemel and P. Bartlett and F. Pereira and K.Q. Weinberger, eds. Advances in Neural Information Processing Systems, 24. Cambridge, MA: MIT Press, 2011.

Saul, L.K., Roweis, S.T. *Nonlinear dimensionality reduction by locally linear embedding*. Science, 2000, 290: 2323 – 2326.

Saul, L.K., Roweis, S.T. *Think globally, fit locally: unsupervised learning of low dimensional manifolds*. Journal of Machine Learning Research, 2003, 4: 119 – 155.

Seung, H. Sebastian, Lee, Daniel D. *The Manifold Ways of Perception.* Science, 290 (5500), 2000, 2268-2269.

Singer, A., Wu, H. *Vector Diffusion Maps and the Connection Laplacian.* Comm. on Pure and App. Math., 2012.

Tyagi, H., Vural, E., Frossard, P. *Tangent space estimation for smooth embeddings of Riemannian manifold.* In arXiv preprint: arXiv:1208.1065v2 [stat.CO] 17 May 2013, pp. 1 – 35 (2013).

Wang, L., Wang, X., Feng, J. *Subspace Distance Analysis with Application to Adaptive Bayesian Algorithm for Face Recognition*. Pattern Recognition 39(3), 2006, 456 – 464.

Wasserman, L. *All of Nonparametric Statistics*. Berlin: Springer Texts in Statistics, 2007.

Wolf, L., Shashua, A. Learning over sets using kernel principal angles. Journal Machine Learning Research, 2003, 4, 913 – 931.

Zhang, Zhenyue, Zha, Hongyuan. *Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment*. SIAM Journal on Scientific Computing, 2005, 26(1), 313 – 338.