Journal of Mathematics and Statistics

Original Research Paper

Asymptotic Properties of Local Sampling on Manifold

Yury Aleksandrovich Yanovich

Institute for Information Transmission Problems (Russian Academy of Sciences), Moscow, Russia
Faculty of Computer Science, Higher School of Economics (National Research University), Moscow, Russia

Article history
Received: 30-06-2016
Revised: 24-08-2016
Accepted: 26-08-2016

Email: yyanovich@hse.ru

Abstract: In many applications, the real high-dimensional data occupy only a very small part in the high dimensional 'observation space' whose intrinsic dimension is small. The most popular model of such data is Manifold model which assumes that the data lie on or near an unknown manifold Data Manifold, (DM) of lower dimensionality embedded in an ambient high-dimensional input space (Manifold assumption about high-dimensional data). Manifold Learning is a Dimensionality Reduction problem under the Manifold assumption about the processed data and its goal is to construct a low-di-mensional parameterization of the DM (global low-dimensional coordinates on the DM) from a finite dataset sampled from the DM. Manifold assumption means that local neighborhood of each manifold point is equivalent to an area of low-dimensional Euclidean space. Because of this, most of Manifold Learning algorithms include two parts: 'local part' in which certain characteristics reflecting low-dimensional local structure of neighborhoods of all sample points are constructed and 'global part' in which global low-dimensional coordinates on the DM are constructed by solving certain convex optimization problem for specific cost function depending on the local characteristics. Statistical properties of 'local part' are closely connected with local sampling on the manifold, which is considered in the study.

Keywords: Manifold Learning, Asymptotic Expansions, Large Deviations

Introduction

Many Data Analysis tasks, such as Pattern Recognition, Classification, Clustering, Prognosis, Function reconstruction and others, which are challenging for machine learning problems, 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 learning algorithms for solving these tasks.

Fortunately, in many applications, especially in imaging and medical ones, the real high-dimensional data occupy only a very small part in the high dimensional 'observation space' $\mathbb{R}^p$ whose intrinsic dimension q is small (usually, $q \ll p$) Donoho (2000; Verleysen, 2003). Thus, various Dimensionality Reduction (Feature extraction) algorithms whose goal is a finding of a low-dimensional parameterization of high-dimensional data can be used as a first key step in solutions of such 'high-dimensional' tasks by transforming the data into their low-dimensional representations (features) preserving certain chosen subject-driven data properties (Bengio *et al*., 2013; Bernstein and Kuleshov, 2014; Kuleshov and Bernstein, 2016. Then the low-dimensional features can be used in reduced learning procedures instead of initial high-dimensional vectors avoiding the curse of dimensionality Kuleshov and Bernstein (2014): 'dimensionality reduction may be necessary to discard redundancy and reduce the computational cost of further operations' Lee and Verleysen (2007).

The most popular model of high-dimensional data, which occupy a very small part of observation space $\mathbb{R}^p$, is Manifold model in accordance with which the data lie on or near an unknown manifold (Data manifold, DM) $\mathbb{X}$ of lower dimensionality q<p embedded in an ambient high-dimensional input space $\mathbb{R}^p$ (Manifold assumption Seung and Lee (2000) about high-dimensional data); typically, this assumption is satisfied for 'real-world' high-dimensional data obtained from 'natural' sources. In real examples, a manifold dimension q is usually unknown and can be estimated by a given dataset randomly sampled from the Data manifold Levina and Bickel (2005; Fan *et al*., 2009; Einbeck and Kalantana, 2013; Rozza *et al*., 2011).

Dimensionality Reduction under the Manifold assumption about the processed data is usually referred to as the Manifold Learning Smith *et al.* (2009; Ma and Fu, 2011) whose goal is constructing a low-dimensional parameterization of the DM (global low-dimensional coordinates on the DM) from a finite dataset sampled from the DM.

Manifold assumption means that local neighborhood of each manifold point is equivalent to an area of low-dimensional Euclidean space. Because of this, most of Manifold Learning algorithms include two parts: 'local part' in which certain characteristics reflecting low-dimensional local structure of neighborhoods of all sample points are constructed and 'global part' in which global low-dimensional coordinates on the DM are constructed by solving certain convex optimization problem for specific cost function depending on the local characteristics under some normalization constraints (usually, generalized eigenvalues problem). It is typical structure of certain class of manifold learning algorithms such as Locally Linear Embedding (LLE) Roweis and Saul (2000), ISOmerric MAPping (Isomap) Tenenbaum *et al.* (2000), Laplacian Eigenmaps (LEM) Belkin and Niyogi (2003), Local Tangent Space Alignment (LTSA) Zhang and Zha (2004), Hessian Eigenmaps (HLLE) Donoho and Grimes (2003), Semidefinite Embedding (SDE) Weinberger and Saul (2006) and Diffusion Maps (DFM) Coifman and Lafon (2006).

The radius of the neighborhood should be small enough to achieve small local estimation error. On the other hand, the number of points in the neighborhood should be large enough to get a small statistical error. There are two approaches to choose the ball's size: it consists of the fixed number of neighbors (k nearest), or the radius is set. The first case does not guarantee that the radius would be small so that the local approximation error could be large. The distribution of the k-th neighbor is studied in (Levina and Bickel, 2005; Farahmand *et al.*, 2007; Campadelli *et al.*, 2015). Also in Smith *et al.* (2008) distance to the k-th neighbor assumed to converge to zero and is of the rate of convergence parameter. The second case does not guarantee the large enough number of points if the neighborhood. This question is mentioned in (Levina and Bickel, 2005; Singer and Wu, 2012) but it wasn't specifically discussed. Singer and Wu (2012), it was shown that both local and statistical parts of errors are asymptotically small for a specific statistic and also large deviation error was estimated.

Random variable 'the number of points in the neighborhood of a fixed point on manifold' is considered in the present paper. Such random variable is binomial (sum of Bernoulli) and is well studied in general case Shiryayev (1984): De Moivre-Laplace theorem provides its distribution for a fixed success parameter and Poisson theorem provides its distribution for a fixed finite product of success parameters and sample size. In this study, it is supposed that sample is generated from a good enough continuous measure on a good enough unknown manifold and the parameter slowly tends to zero. Thus, the case between de Moivre-Laplace and Poisson theorems is considered: Success parameter tends to zero but it's product with sample size tends to infinity. Another feature of the work is the measure's support: Unknown and curved manifold. In this study, parametric families of random variables (correlated random fields with neighborhood centers as parameters) are studied and uniform results are obtained.

The paper is organized as follows. It is started with the Manifold Learning problem and typical workflow formulation in Section Common Manifold Learning Problem. In Section Results Description, the main results of the paper are listed and commented. In Section Data Model the data model is defined and all assumptions are listed. Then, Main Results Section contains exact formulations of the main results. In Section Some Definitions and Lemmas are listed useful definitions and lemmas to prove main results. Proof of Main Theorems contains the main proofs. In Section Conclusion the paper summary and future work directions are given. In Appendix A. Definitions and lemmas from differential geometry useful definitions and results from differential geometry are recalled. In Section Appendix B. Lemmas Proofs the lemmas from the Section Some Definitions and Lemmas are proved.

Common Manifold Learning Problem

Manifold Learning as Manifold Embedding

The main results are strictly formulated Consider unknown -dimensional Data manifold:

$$\mathbb{X} = \{X = f(b) \in \mathbb{R}^p : b \in \mathbb{B} \subset \mathbb{R}^q\} \tag{1}$$

Covered by a single coordinate chart $\mathbb{B}$ and embedded in an ambient p-dimensional space $\mathbb{R}^p$, q<p. The chart $\mathbb{B}$ is a one-to-one mapping from open bounded space $\mathbb{B} \subset \mathbb{R}^p$ to manifold $\mathbb{X} = f(\mathbb{B})$ with differentiable inverse map $f^{-1} : \mathbb{X} \to \mathbb{B}$. The manifold intrinsic dimension q is assumed to be known.

Inverse mapping $h_f(X) = f^{-1}(X)$, whose values $b = h_f(X) \in \mathbb{B}$ can be considered as low-dimensional coordinates on the manifold $\mathbb{X}$, gives low-dimensional representations (features) $b = h_f(X)$ of high-dimensional manifold-valued data $\mathbb{X}$.

If the mappings $h_f(X)$ and $f(b)$ are differentiable and $J_f(b)$ is $p \times q$ Jacobi an matrix of mapping $f(b)$ than q-dimensional linear space:

$$T_X(\mathbb{X}) = \mathtt{Span}(J_f(h_f(X))) \tag{2}$$

In $\mathbb{R}^p$ is tangent space to the manifold $\mathbb{X}$ at the point $X \in \mathbb{X}$; hereinafter, *Span*(*H*) is linear space spanned by columns of the arbitrary matrix *H*. The tangent spaces can be considered as elements of the Grassmann manifold *Grass*(*p*,*q*) consisting of all q-dimensional linear subspaces in $\mathbb{R}^p$.

Let $\mathbb{X}_N = \{X_1, \ldots, X_N\} \subset \mathbb{X}$ be a dataset randomly sampled from the DM $\mathbb{X}$ according to certain (unknown) probability measure whose support coincides with $\mathbb{X}$. Common Manifold learning problem is as: Given a sample $\mathbb{X}_N$, construct a low-dimensional parameterization of the DM which produces an Embedding mapping:

$$h : \mathbb{X} \subset \mathbb{R}^p \to \mathbb{Y}_h = h(\mathbb{X}) \subset \mathbb{R}^p \quad (3)$$

From the DM $\mathbb{X}$ to the Feature Space (FS) $\mathbb{Y}_h \subset \mathbb{R}^p, q < p$, which preserves specific properties of the DM.

Manifold Learning as Manifold Embedding

Following Goldberg *et al*. (2008), we consider a class consisting of typical Manifold learning algorithms which recover the underlying structure of the Data manifold from the sample; this class includes so-called 'normalized-output' algorithms Goldberg *et al*. (2008). A common scheme of the considered algorithms is constructed in four steps.

First Step: Neighborhoods Construction

For each sample point X_n, local neighborhood $U_N(X_n) = \{X_n, X_{n,1}, \ldots, X_{n,k(n)}\} \subset \mathbb{X}_N$ consisting of near sample points is constructed. Typical examples: $U_N(X_n) = U_N(X_n, \varepsilon)$ consists of sample points that belong to ε-ball in $\mathbb{R}^p$ centered at X_n, or $U_N(X_n) = U_N(X_n, k)$ in which $k(n) = k$, consists of *k* nearest-neighbors of the considered point X_n.

The constructed neighborhoods determine Sample graph $\Gamma(X_N)$ consisting of *N* vertices $\{X_1, X_2, \ldots, X_N\}$; the vertices X_n and X_j are connected by an edge (X_n, X_j) when $X_n \in U_N(X_j)$ and $X_j \in U_N(X_n)$. These neighborhoods determine 'Euclidean' kernels $K_E(X_n, X_j) = I(X_n \in U_N(X_j); X_j \in U_N(X_n)\}$, in which $I(A)$ is an indicator function of the event *A*, or 'heat-transfer' kernel $K_{h-t}(X_n, X_j) = K_E(X_n, X_j) \cdot \exp(-\varepsilon \cdot \| X_n - X_j \|)$ Belkin and Niyogi (2003).

Second Step: Neighborhoods Descriptions

Chosen descriptions of the neighborhoods (local descriptions of the DM) are computed. Examples of such descriptions:

- Barycentric coordinates $\{w_{n,1}, w_{n,2}, \ldots, w_{n,k}\}$ of the 'central' point X_n with respect to its k nearest-neighbors $\{X_{n,1}, X_{n,2}, \ldots, X_{n,k}\}$ that minimize the reconstruction error functions $\| X_i - \sum_j w_{n,j} X_{n,j} \|^2$ Roweis and Saul (2000); the neighborhood $U_N(X_n, k)$ is used here
- An applying of the Principal Component Analysis (PCA) Jolliffe (2002) to the neighborhood $U_N(X_n, \varepsilon)$ results in an p×q orthogonal matrix $Q_{PCA}(X_n)$ whose columns are the PCA principal eigenvectors corresponding to the q largest PCA eigenvalues (Zhang and Zha, 2004; Bernstein and Kuleshov, 2014). These matrices determine q-dimensional linear spaces $L_{PCA}(X_n) = Span(Q_{PCA}(X_n))$ in the $\mathbb{R}^p$ which, under certain conditions, accurately approximate the tangent spaces $T_{\mathbb{X}}(X_n)$ (2) to the DM $\mathbb{X}$ at the points X_n Singer and Wu (2012).

Third Step: Global Description

Chosen global description of the DM is computed by solving some convex optimization problems under some normalization constraints. Usually, low-dimensional sample features $\mathbb{Y}_N = h(\mathbb{X}_N) = \{y_1, y_2, \ldots, y_N\} \subset \mathbb{Y}_h$ are computed by minimization of chosen cost function $L(\mathbb{Y}_N \mid \mathbb{X}_N)$ over $\mathbb{Y}_N$. Examples of cost functions:

$$L_{LLE}(\mathbb{Y}_N \mid \mathbb{X}_N) = \sum_{n=1}^{N} \| y_n - \sum_j w_{n,j} y_{n,j} \|_F^2$$

$$L_{LE}(\mathbb{Y}_N \mid \mathbb{X}_N) = \sum_{n=1}^{N} \| y_n - y_j \|_F^2$$

$$L_{LTSA}(\mathbb{Y}_N \mid \mathbb{X}_N) = \sum_{n=1}^{N} \| (I_q - Q_{PCA}(X_n) \cdot Q_{PCA}(X_n)^T) \cdot H_n \cdot \mathbb{Y}_{(n)} \|_F^2$$

Are used in the algorithms LLE Roweis and Saul (2000), LE Belkin and Niyogi (2003) and LTSA Zhang and Zha (2004), respectively; here $q \times (k(n)+1)$ matrix Y_n consists of q-dimensional columns $\{y_n, y_{n,1}, y_{n,2}, \ldots, y_{n,k(n)}\}$ in which the same pairs of indices *n*,*j* are uses as in $U_N(X_n)$, $H_n = I_q - (1/k(n)) \times 1 \times 1^T$ is q×q centering matrix in which I_q is q×q unit matrix and q-dimensional vector 1 consists of units. Some normalization constraints on the Feature sample $\mathbb{Y}_N$ are used to avoid the degenerate solutions.

ISOMAP Tenenbaum *et al*. (2000) is based on estimating the geodesic curves on the DM $\mathbb{X}$. Consider the sample weighted graph $\Gamma_W(\mathbb{X}_N)$, in which an edge (X_n, X_j) has weight $\| X_n - X_j \|$. Let $\{D_{nj}\}$ are the lengths of the shortest paths between the vertices X_n and X_j in the graph $\Gamma_W(\mathbb{X}_N)$ that can be computed with using the Dijkstra algorithm; these quantities $\{D_{nj}\}$ accurately estimate geodesic distances between the points X_n and X_j

on the DM $\mathbb{X}$ Tenenbaum *et al*. (2000). After that, the Feature sample $\mathbb{Y}_N$ are computed with use Multi-Dimensional Scaling framework (MDS) Cox and Cox (2008) by minimization over $\mathbb{Y}_N$ the MDS cost function $L_{ISOMAP}(\mathbb{Y}_N \mid \mathbb{X}_N) = \sum_{n,j=1}^{N} \| D_{nj} - \|y_n - y_j\| \|^2$.

Fourth Step: Out-of-Sample Extension

The Feature sample $\mathbb{Y}_N$ gives the values of the Embedding mapping *h*(*X*) (3) only at sample points; a finding of low-dimensional features *h*(*X*) for Out-of-Sample (OoS) points $X \in \mathbb{X} \setminus \mathbb{X}_N$ is usually called OoS-extension problem. The OoS-extension for the algorithms LLE Roweis and Saul (2000), ISOMAP Tenenbaum *et al*. (2000), LE Belkin and Niyogi (2003), has been found in Bengio *et al*. (2003) with use Nyström's eigendecomposition technique Saul *et al*. (2003).

Results Description

The main results are strictly formulated in Section Main Results and all the assumptions, used in the proof, strictly formulated in Section Data Model. Some substantive comments are given here.

Data Model

Data Model consists of assumptions about support (manifold), assumptions about sample distribution and assumptions about neighborhood parameter. The paper deals with 'good enough' manifolds with known dimensionality q. The problem of dimensionality estimation is a problem of the only integer parameter estimation and solutions Campadelli *et al*. (2015) with the rate of error probability ~*exp*(-*C.N*) are known, where N is a sample size C > 0 is constant. Such rate is smaller than the rates in this article. The sample assumed to be independent identically distributed (i.i.d.) with unknown 'good' continuous measure on the manifold. The neighborhood parameter slowly tends to zero.

Main Results

Manifold behaves as a linear subspace in a small neighborhood of a point. Therefore, the intersection of a full dimensional Euclidean ball with a manifold is close to the low dimensional ball. Thus, the number of sampling points, fallen into the neighborhood, should be proportional to the volume of the q-dimensional ball. This result was mentioned a number of times in the works (for example, Levina and Bickel (2005; Einbeck and Kalantana, 2013; Singer and Wu, 2012)). However, in the Theorem 1, we prove that the number of sampling points in the neighborhood, divided by the volume of the neighborhood, is a consistent estimate of the density at the point. The Theorem 2 prove that conditional distribution of sample points in the neighborhood is asymptotically uniform. The Corollary 1 sets that all directions from tangent space are equal for the conditional distribution. So, one could think of conditional distribution as of uniform distribution on the ball in tangent space. In the Theorem 3 asymptotic expansion of the considered statistics is given and in the Theorem 4 assesses the probability of large deviations. The Theorem 5 prove a uniform result of the large deviations probability: If we consider all points of the manifold, which are a little removed from the border, as the centers of the balls, then the minimum over all balls of points in each of them, will be asymptotically infinitely large with a high probability.

The features of the results are the curvature of the unknown sample support, the tendency to zero of the random variable mathematical expectation ('hit one sampling point in a ball with decreases to zero radius'), the need to obtain uniform estimation on the manifold.

The basic ideas used in proofs: Local linearization of the support, a generalization of de Moivre-Laplace theorems to the case of decreasing the probability of success in the Bernoulli scheme, the use of inequalities for the probabilities of large deviations of sums of i.i.d. random variables, the use of finite nets.

Data Model

Let's assume that:

M1. $\mathbb{X} \subset \mathbb{R}^p$ is a q-dimensional manifold covered with a single map. That is, for some open $\mathbb{B} \subset \mathbb{R}^q$ and $f : \mathbb{B} \to \mathbb{R}^p : \mathbb{X} = f(\mathbb{B})$ homeomorphism
M2. q is known
M3. $\mathbb{B}$ is a bounded set
M4. Eigenvalues of q×q Jacobi matrixes product $J_f^T(b) \cdot J_f(b)$ of *f* mapping uniformly separated from 0 and infinity
M5. Hessian of *f* mapping exists and bounded on $\mathbb{X}$
M6. Third order derivatives of *f* existing and are elementwise bounded on $\mathbb{X}$
M7. Condition number $c(\mathbb{X})$ is bounded, where $\mathbb{X}$ is the smallest number such that for each point, which is distant from $\mathbb{X}$ at a distance smaller than $1/c(\mathbb{X})$, the only projection exists on $\mathbb{X}$ Niyogi *et al*. (2008)
M8. Manifold $\mathbb{X}$ is geodesically convex, that is, for any pair of points on $\mathbb{X}$ exists geodesic line and it is the shortest path

Note 1. The assumption M1 is equivalent to the existence of the global dimension coordinate system on the manifold.
Assumption M3 is used to obtain uniform properties of statistics.

Assumptions M4-M6 represent the conditions of smoothness and are used to connect Euclidean distances and volumes with the distances and the volumes on the manifold.
Property M7 means that the manifold does not contain a 'short-circuit': The closeness of the points on the Euclidean distance implies the closeness of points on the manifold. That is, the intersection of the small Euclidean neighborhood of the point creates a small neighborhood of a point on the manifold.
The assumption M8 is a technical simplification for Taylor expansion.

The set $\mathbb{B}$ is bounded and derivatives of f are bounded too, so the manifold $\mathbb{X}$ is also bounded. Let a be an edge of the circumscribed cube of manifold:

$$a = \inf\{a' : a', a_1, \dots, a_p : \mathbb{X} \subset \otimes_{i=1}^{p} [a_i, a_i + a']\} \tag{4}$$

Let C_J and C_j be the minimum and the maximum eigenvalues of the metric tensor $J_f(b)^T J_f(b), b \in \mathbb{B}$ correspondingly:

$$c_J = \inf_{b\in\mathbb{B}} \min_{\lambda\in\sigma\left(J_f(b)^T J_f(b)\right)} \lambda \tag{5}$$

$$C_J = \inf_{b\in\mathbb{B}} \max_{\lambda\in\sigma\left(J_f(b)^T J_f(b)\right)} \lambda \tag{6}$$

Let C_H be the maximum element of Hessian matrix of $(\mathbb{B}, f)$ mapping:

$$C_H = \sup_{X\in\mathbb{X}, i,j=1,\dots,q} \left\| \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \right\| = \sup_{X\in\mathbb{X}, i,j=1,\dots,q} \sqrt{\sum_{k=1}^{p} \left| \frac{\partial^2 f_k(b)}{\partial b_i \partial b_j} \right|^2} \tag{7}$$

where, $X = (f_1(b), \dots, f_p(b))^T$. Let:

$$C_T = \sup_{X\in\mathbb{X}, i,j,m=1,\dots,q} \left\| \frac{\partial^3 f(b)}{\partial b_i \partial b_j \partial b_m} \right\| = \sup_{X\in\mathbb{X}, i,j,m=1,\dots,q} \sqrt{\sum_{k=1}^{p} \left| \frac{\partial^3 f_k(b)}{\partial b_i \partial b_j \partial b_m} \right|^2} \tag{8}$$

Manifold $\mathbb{X}$ is unknown and it is represented by finite random sample $\mathbb{X}_n = \{X_1, \dots, X_N\} \subset \mathbb{X} \subset \mathbb{R}^p$ with N points. Moreover, it is assumed about sample selection that:

S1. Points from $\mathbb{X}_n$ are i.i.d. with some probability measure μ such that $\mathbb{X}$ is its support: $\mathbb{X} = supp\mu$
S2. Measure μ is continuous with respect to Riemannian measure on manifold and its density p_μ is bounded from zero and infinity
S3. Density p_μ is twice smooth on $\mathbb{X}$ and its derivatives are bounded

Note 2. Manifold $\mathbb{X}$ has Riemannian measure dV(X) (volume measure) which is equal to the -dimensional volume in the main term (Section Some Definitions and Lemmas). Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space, than Borelian function, $\mathfrak{B} \to \mathbb{X} : X = X(\omega)$ is called random variable on the manifold. Let's call such induced on $\mathbb{X}$ the measure as μ. If for each Borelian set $\mathfrak{X} \in \mathfrak{B}$ $\mu(X \in \mathfrak{X}) = \int_{\mathfrak{X}} p_\mu(X) dV(X)$, then the function $p_\mu(X), X \in \mathbb{X}$ is a probability density function Pennec (1999).

Let p_{min} and p_{max} be the minimum and maximum values of p_μ (they exist by S2):

$$p_{\min} = \inf_{X\in\mathbb{X}} p_\mu(X) \tag{9}$$

$$p_{\max} = \sup_{X\in\mathbb{X}} p_\mu(X) \tag{10}$$

Let define the bounds for maximum eigenvalues of first and second derivatives (exist by S3):

$$C_{p,1} = \sup_{X\in\mathbb{X}, \theta\in T_X(\mathbb{X}):\|\theta\|=1} \| \nabla_\theta p_\mu(X) \| \tag{11}$$

$$C_{p,2} = \sup_{X\in\mathbb{X}, \theta\in T_X(\mathbb{X}):\|\theta\|=1} \| \nabla_\theta \nabla_\theta p_\mu(X) \| \tag{12}$$

where, ∇_θ is a covariant derivative (Appendix A), which is a kind of directional derivative generalization for the manifold.

Note 3. S1 and S2 are standard assumptions whose guarantee correspondence between sample $\mathbb{X}_N$ and manifold $\mathbb{X}$. Assumption S3 is useful for uniform results.

For neighborhood parameter ε = ε(N) it is assumed that:

- P1. For $N \to \infty : \varepsilon \to 0$
- P2. For $N \to \infty : N\varepsilon^q \to \infty$
- P3. For $N \to \infty : N\varepsilon^{q+4} \to 0$

Note 4. The assumption P1 means that the neighborhood size tends to zero and therefore the expansion of functions at the main term is a term with the lowest degree of length. The assumption P2 provides an infinite number of sample points in the neighborhood despite the *decrease in the size of the neighborhood. Assumption P3 is stronger than P1 and guarantees that the contribution of the bias of the order ε^2 is infinitely small in the results of the central limit theorem for the number of points of order $N\varepsilon^q$.*

Main Results

Let $I_\varepsilon(X_n|X)$ be an indicator of the event 'p-dimensional distance between X_n and X is less of equal to ε', that is $I_\varepsilon(X_n|X)$ for $\| X_n - X \| < \varepsilon$ and $I_\varepsilon(X_n|X) = 0$ otherwise.

Let $N_\varepsilon(X)$ be the number of sample $\mathbb{X}_N$ points in p-dimensional ε-neighborhood of $X \in \mathbb{R}^p$:

$$N_\varepsilon(X) = \sum_{n=1}^{N} I_\varepsilon(X_n \mid X). \tag{13}$$

Let V_q be the volume of q-dimensional ball of a unit radius:

$$V_q = \frac{\pi^{q/2}}{\Gamma(q/2+1)} \tag{14}$$

where, $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, z > 0$ is a gamma function, that is $V_{2m} = \frac{\pi^m}{m!}$ and $V_{2m+1} = \frac{2^m \pi^m}{(2m+1)!!}$ for integer m.

Theorem 1. (weak law of large numbers). If M1-M8, S1-S3 and P1-P2, then for each $X \in \mathbb{X}$ and $N_\varepsilon(X)$ for $N \to \infty$:

$$\frac{N_\varepsilon(X)}{N\varepsilon^q} \xrightarrow{p} p_\mu(X) \cdot V_q$$

where, V_q is a volume of -dimensional ball (14).

Theorem 2. (conditional uniform distribution). If M1-M8, S1-S3 and P1-P2, then for each $X \in \mathbb{X}$ for $N \to \infty$:

$$p(X_n \mid I_\varepsilon(X_n \mid X) = 1) \cdot \varepsilon^q \to 1 / V_q$$

where, V_q is a volume of q-dimensional ball (14).

Corollary 1. If M1-M8, S1-S2 and P1-P2, then for each $X \in \mathbb{X}$ the spectrum of a local covariance matrix $\Sigma = \frac{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X) \cdot (X_n - X) \cdot (X_n - X)^T}{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X)} \in \mathbb{R}^{p \times p}$ tends to ones and p-q zeros as $N \to \infty$. Moreover, ones correspond to vectors from tangent space and zeros corresponds to vectors from cotangent space.

Theorem 3. (central limit theorem for $N_\varepsilon(X)$). If M1-M8, S1-S3 and P1-P3, then for each $X \in \mathbb{X}$ and $N_\varepsilon(X)$ (13) $N \to \infty$:

$$\frac{N_\varepsilon(X) - N\varepsilon^q p_\mu(X) V_q}{\sqrt{N\varepsilon^q p_\mu(X) V_q}} \xrightarrow{d} N(0,1)$$

where, V_q is a volume of q-dimensional ball (14), $N(0,1)$ is standard normal distribution.

Theorem 4. (large deviations of $N_\varepsilon(X)$). If M1-M8, S1-S2 and P1-P3, then for $0<z<1/16$:

$$P\left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q} \le p_{\min} \cdot (1-z) - \varepsilon^2 \cdot \frac{C_{\mathbb{E}} p_{\max}}{V_q p_{\min}} \right) \le \exp\left(-4z^2 \cdot N\varepsilon^q V_q p_{\min}{}^2 / p_{\max}\right)$$

where, V_q is a volume (14), $C_{\mathbb{E}}$ is a constant from Corollary 4, constants p_{min}, p_{max}, C_{Ric}, $C_{p,1}$ and $C_{p,2}$ are defined in (9), (10), Lemma 4, (11), (12) correspondingly.

If also S3, then for 0<z<1/16:

$$P\left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \ge z + \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu(X)} \right) \le \exp\left(-4z^2 \cdot N\varepsilon^q V_q p_\mu(X)\right);$$

$$\le P\left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \le -z - \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu(X)} \right) \le \exp\left(-4z^2 \cdot N\varepsilon^q V_q p_\mu(X)\right).$$

Theorem 5. (uniform large deviation). If M1-M8, S1-S2 and P1-P3, then for $\varepsilon \le \min\left\{ \frac{1}{c(\mathbb{X})}, \sqrt{\frac{24(\sqrt[q]{2}-1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}}, \frac{1}{\sqrt{C_{Ric} C_{p,2}}}, \frac{1}{\sqrt{C_{Ric} C_{p,1} + C_{p,2} + 1}}, \frac{1}{2}, \frac{1}{\sqrt{2C_{\mathbb{E}}}}, \sqrt[q]{\frac{1}{4V_q p_{\min}}}, \sqrt[q]{\frac{1}{4p_{\min}}} \right\}$, and for each point (not too close to the boundary) $X \in \mathbb{X}_\varepsilon$ (17) and:

$$P\left(\inf_{X \in \mathbb{X}_\varepsilon} \frac{N_\varepsilon(X)}{N\varepsilon^q V_q} \le p_{\min} \cdot (1-z) - \varepsilon^2 \cdot \frac{C_{\mathbb{E}} p_{\max}}{V_q p_{\min}} \right) \le \left(\frac{6a\sqrt{p}}{\varepsilon} \right)^p \exp\left(-4z^2 \cdot N\varepsilon^q V_q p_{\min}{}^2 / (9p_{\max})\right)$$

where, V_q is a volume (14), $C_{\mathbb{E}}$ is a constant from Corollary 4, constants p_{min}, p_{max}, C_{Ric}, $C_{p,1}$ and $C_{p,2}$ and are defined in (9), (10), Lemma 4, (11), (12) correspondingly and a>0 is an edge of the circumscribed other cube (4).

DOI: 10.3844/jmssp.2016.157.175

Note 5. If $\varepsilon(N) \geq C_\alpha \cdot N^{-\alpha}, \alpha < 1/q, C_\alpha > 0$*, then the right part of Theorem's 4 equation tends to zero faster than any degree* $N^{-\beta}$*,* $\beta>0$ *as a product of exponential and polynomial terms.*

Some Definitions and Lemmas

Definitions and Lemmas which are used to prove main theorems are listed in this section.

Local Linearization

Manifold $\mathbb{X}$ is close to q-dimensional linear tangent space $T_X(\mathbb{X})$ in a neighborhood of each point. Differentiation on the manifold is slightly different from the differentiation in Cartesian coordinates in $\mathbb{R}^k$. The differences are mainly of a technical nature and the main difference lies in the fact that derivatives are defined only in the directions from $T_X(\mathbb{X})$ (Appendix A).

In a sufficiently small neighborhood of $X \in \mathbb{X}$ there exists a one-to-one mapping between the points of the manifold $\mathbb{X}$ and a subset of the elements of the tangent space $T_X(\mathbb{X})$. In this neighborhood vectors $V \in T_X(\mathbb{X})$ are coordinates. These coordinates are called locally Riemann (Appendix A).

Manifold volume element and q-dimensional volume element of the tangent space are related by the following lemma:

Lemma 1. (Petersen (2006)) Riemannian measure in polar coordinates in the neighborhood of $X \in \mathbb{X}$ has form:

$$dV(\exp_X t\theta) = J(t,\theta)dtd\theta$$

where, $\theta \in T_X\mathbb{X}$, $\|\theta\| = 1, t > 0, \tilde{t} \in [0,t]$, $\exp_X V$ is an exponential map of V at X (Appendix A):

$$J(t,\theta) = t^{q-1} + t^{q+1} Ric_X(\theta,\theta) + O(t^{q+2})$$
$$J(t,\theta) = t^{q-1} + t^{q+1} Ric_{\tilde{X}}(\tilde{\theta},\tilde{\theta})$$

Ric_X (θ,θ) is a Ricci curvature (26) at X, $\tilde{X} \in \{\exp_X \tilde{t}\theta \mid \tilde{t} \in [0,t]\}, \tilde{\theta} \in T_{\tilde{X}}(\mathbb{X})$.

The distance between the closest points of the manifold in q-dimensional space and the distance between them in the q-dimensional Riemannian coordinates Singer and Wu (2012) are related by Lemma:

Lemma 2. (Coifman and Lafon (2006)). For $X, \tilde{X} \in \mathbb{X}$, such that $\tilde{X} = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{X})$ and $\|\theta\| = 1$, for small (and small $\left(\textit{and small } h = \| X - \tilde{X} \|\right)$):

$$t = h + \frac{1}{24} \| II_X(\theta,\theta) \| h^3 + O(h^3)$$
$$t = h + \frac{1}{24} \| II_{\tilde{X}}(\tilde{\theta},\tilde{\theta}) \| h^3$$

where, $\tilde{X} \in \{\exp_X \tilde{t}\theta \mid \tilde{t} \in [0,t]\}$, $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{X})$, $II_X(V,V)$ is a second fundamental form (22).

The values of the right-hand part of the Lemmas 1 and 2 may be bonded under the M1-M8 assumptions:

Lemma 3. For $X \in \mathbb{X}$ and $\theta \in T_X(\mathbb{X})$ and $\|\theta\| = 1$:

$$\| II_X(\theta,\theta) \| \leq C_{II} \equiv \frac{C_H}{c_J} \cdot q$$

where, C_H is a maximal element of Hessian matrix (7), C_J is a minimum eigenvalue of the metric tensor (5):

Lemma 4. For $X \in \mathbb{X}$, $\theta \in T_X(\mathbb{X})$ and $\|\theta\| = 1$:

$$\| Ric_X(\theta,\theta) \| \leq C_{Ric} \equiv 2q^3 \cdot \frac{C_H^2 + C_T\sqrt{C_J}}{c_J^{3/2}}$$
$$+q^5 \cdot \left(18 \cdot C_H^2 \frac{C_J}{c_J} + 4 \cdot C_H \frac{\sqrt{C_J}}{c_J}\right) / c_J$$

where, C_H is a maximum element of Hessian matrix (7), C_J and C_J are a minimum and maximum eigenvalues of the metric tensor (5) and (6), C_T is the maximal norm of the third derivative (8).

The proofs of Lemmas 3 and 4 are in Appendix A.

De Moivre-Laplace Lemmas for Slowly Decreasing Probability

We use local and integral de Moivre-Laplace lemmas for success parameter which slowly tends to zero. In the classical formulation, the parameter is considered to be fixed. However, the proof almost does not change, if we assume success parameter p_n depends on the sample size, but p_n. $N \to \infty$ when $n \to \infty$. The only difference is the functions expansion in a small parameter $1/(p_n \cdot n)$, instead of a small parameter $1/n$:

Lemma 5. (local de Moivre-Laplace for slowly decreasing probability). Let the success probability in a Bernoulli scheme $p_n > 0$ depends on the sample size n and $q_n = 1-p_n$ and also p_n. $n \to \infty$ for $n \to \infty$. Then for $n \to \infty$ and the number of successes k such that $\frac{|k - np_n|}{(np_nq_n)^{2/3}} \to 0$ for $n \to \infty$:

$$P_n(k) \sim \frac{1}{\sqrt{2\pi np_nq_n}} \exp\left(\frac{-(k-np_n)^2}{2np_nq_n}\right)$$

That is:

$$\sup_{\{k:|k-np_n|\le\varphi(np_nq_n)\}}\left|\frac{P_n(k)}{\frac{1}{\sqrt{2\pi np_nq_n}}\exp\left(\frac{-(k-np_n)^2}{2np_nq_n}\right)}-1\right|\to 0$$

where, $\varphi(np_nq_n)$ is an arbitrary function such that $\varphi(np_nq_n)=o\left((np_nq_n)^{2/3}\right)$.

Corollary 2. Lemma is equivalent to the following statement: for each $z\in\mathbb{R}$ such that $z=o(np_nq_n)^{1/6}$ and $np+z\sqrt{np_nq_n}$ is integer from $\{1,.., n\}$:

$$p(np+z\sqrt{np_nq_n})\sim\frac{1}{\sqrt{2\pi np_nq_n}}e^{-z^2/2}$$

That is for $n\to\infty$:

$$\sup_{\{z:|z|\le\psi(n)\}}\left|\frac{p_n(np_n+z\sqrt{np_nq})}{\frac{1}{\sqrt{2\pi np_nq_n}}e^{-z^2/2}}-1\right|\to 0$$

where, $\psi(n)=o(np_nq_n)^{1/6}$.

Lemma 6. (integral de Moivre-Laplace for slowly decreasing probability). Let the success probability in a Bernoulli scheme $p_n > 0$ depends on the sample size n and $q_n = 1-p_n$ and also p_n. $n\to\infty$ for $n\to\infty$. Let $P_n(k)=C_n^kp_n^kq_n^{n-k}$, $P_n(a,b]=\sum_{a<z\le b}P_n(np_n+z\sqrt{np_nq_n})$. Then $\sup_{-\infty\le a<b\le\infty}\left|P_n(a,b]-\frac{1}{\sqrt{2\pi}}\int_a^b\exp(-z^2/2)dz\right|\to 0, n\to\infty$. The proofs of Lemmas 5 and 6 are in Appendix B.

The Probability of Large Deviations for Bounded Random Variables

To estimate the probability of large deviations we will use the following lemma (proved in Appendix B):

Lemma 7. Let $x_1,\dots,x_n$ be i.i.d. random variables, $\mathbb{E}\chi_1<\infty$, $|\chi_1-\mathbb{E}\chi_1|\le C<\infty$ and constants $m_1,\dots,m_n$ and m, such that $\max_{k=1,\dots,n}|m_k-\mathbb{E}\chi_k|\le m$. Let $\bar{\chi}=\frac{1}{n}\sum_{k=1}^{n}(\chi_k-m_k)$ and $\sigma^2=\mathbb{E}(\chi_1-\mathbb{E}\chi_1)^2$.

Then for $H\ge 2C$ and $0\le x\le\frac{1}{H}$:

$$P(\bar{\chi}\ge x\cdot\sigma^2+m)\le\exp(-x^2\sigma^2\cdot n/4)$$
$$P(\bar{\chi}\le -x\cdot\sigma^2-m)\le\exp(-x^2\sigma^2\cdot n/4)$$

and for $x\ge\frac{1}{H}$:

$$P(\bar{\chi}\ge x\cdot\sigma^2+m)\le\exp(-x\sigma^2n/4H)$$
$$P(\bar{\chi}\le -x\cdot\sigma^2-m)\le\exp(-x\sigma^2n/4H)$$

Integration Area Replacement

Let $B_\varepsilon(X)$ be the intersection of p-dimensional Euclidian ε-neighborhood of $X\in\mathbb{X}$ and manifold $\mathbb{X}$:

$$B_\varepsilon(X)=\{\tilde{X}\mid\tilde{X}\in\mathbb{X}\cap\|\tilde{X}-X\|<\varepsilon\}\tag{15}$$

Let $\tilde{B}_\varepsilon(X)$ be the ε-neighborhood of X in locally Riemannian with center $X\in\mathbb{X}$:

$$\begin{aligned}&\tilde{B}_\varepsilon(X)=\{\tilde{X}\mid\exists V\in T_X(\mathbb{X})\\&\tilde{X}=\exp_X(V)\cap\|V\|<\varepsilon\}\end{aligned}\tag{16}$$

Let $\mathbb{X}_\varepsilon$ be the set of internal points which are ε far from manifold:

$$\begin{aligned}&\mathbb{X}_\varepsilon=\{X\in\mathbb{X}\mid\forall V\in T_X(\mathbb{X})\cap\|V\|<\varepsilon\\&\exp_X V\in\mathbb{X}\}\end{aligned}\tag{17}$$

Lemma 8. For each bounded function and $g(X,\tilde{X})=\tilde{g}(X,t,\theta)$, $\varepsilon\le\min\left\{\frac{1}{c(\mathbb{X})},\sqrt{\frac{24\cdot(\sqrt[q]{2}-1)}{\max\{1,C_{II}\}}},\frac{1}{2\sqrt{C_{Ric}}}\right\}$ a nd $X\in\mathbb{X}_\varepsilon,\tilde{X}\in\mathbb{X},\ \theta\in T_X(\mathbb{X})\cap\|\theta\|=1$:

$$\begin{aligned}&\left|\int_{B_\varepsilon(X)}g(X,\tilde{X})dV(\tilde{X})-\int_{\tilde{B}_\varepsilon(X)}g(X,\tilde{X})dV(\tilde{X})\right|\\&\le 8\cdot V_q\cdot\sup_{X,\tilde{X}}|g(X,\tilde{X})|\cdot\varepsilon^{q+2}\end{aligned}$$

where, $c(\mathbb{X})$ is a condition number (M7), V_q is the volume (14), C_{II} and C_{Ric} are constants from Lemmas 3 and 4.

Lemma is proved in Appendix B.

Finite Nets

Additional construction will be used to prove the uniformity of the estimates. δ-net of the metric space $\mathbb{Z}$ is a set $\mathbb{Z}_{net}(\delta)\subset\mathbb{Z}$ such that for each point $Z\in\mathbb{Z}$ exists δ-close point $Z_{net}\in\mathbb{Z}_{net}(\delta)$.

Since the set of $\mathbb{X}_\varepsilon$ (15) far from the border, points is a subset of $\mathbb{X}$ and M2 and M4 hold true, then applying Lemma 10 we get.

Corollary 3. For each δ > 0 and ε > 0 exist finite δ-net $\mathbb{X}_\varepsilon$ with $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ or fewer elements, where $a > 0$ is an edge of p-dimensional hypercube (4).

Proof of Main Theorems

Let χ be the indicator the event 'random point $\tilde{X}$, which is distributed with density p_μ, is in p-dimensional ε-neighborhood of $X \in \mathbb{X}$':

$$\chi = \chi(\tilde{X} \mid X, \varepsilon) = I(\| X - \tilde{X} \| \leq \varepsilon) \tag{18}$$

where, $I(A) = 1$, if occurred and else 0. $N_\varepsilon(X)$ (13) is a sum of Bernoulli random variables:

$$N_\varepsilon(X) = \sum_{n=1}^{N} \chi(X_1 \mid X, \varepsilon) \tag{19}$$

Let's estimate first and second moments of χ:

Lemma 9. Let:

$$\varepsilon \leq \min\left\{\frac{1}{c(\mathbb{X})}, \sqrt{\frac{24 \cdot (\sqrt[q]{2} - 1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}}\right\}$$

$X \in \mathbb{X}_\varepsilon$, then $\mathbb{E}\chi^2 = \mathbb{E}\chi = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2 \delta_{\mathbb{E}})$, where:

$$\begin{aligned} |\delta_{\mathbb{E}}| &\leq 8 \cdot p_{\max} + C_{Ric} \\ &+ \varepsilon^2 \cdot \Big(\big(C_{Ric} \cdot C_{p,1} + C_{p,2} \big) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \Big) \end{aligned} \tag{20}$$

and constants V_q, p_{min}, p_{max}, C_{II} C_{Ric}, $C_{p,1}$ and $C_{p,2}$ are defined in (14), (9), (10), Lemmas 3 and 4, (11) and (12). Also, if S3 is not supposed:

$$\mathbb{E}\chi^2 = \mathbb{E}\chi \geq \varepsilon^q \cdot V_q \cdot p_{\min} - V_q \cdot \varepsilon^{q+2} \cdot (C_{Ric} + 8 p_{\max})$$

Proof. Note that $\chi^2 = \chi$. So $\mathbb{E}\chi^2 = \mathbb{E}\chi$. Let point $\tilde{X} \in \mathbb{X}$ be a random variable with density $p_\mu(\tilde{X})$.

$$\mathbb{E}\chi = \int_{B_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X})$$

Using Lemma 8 for $p_\mu(\tilde{X})$ and (10):

$$\begin{aligned} &\left| \int_{B_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) \right| \\ &\leq 8 \cdot V_q \cdot p_{\max} \cdot \varepsilon^{q+2} \end{aligned}$$

Using expansion:

$$\begin{aligned} p_\mu(\tilde{X}) &= p(X) + t \cdot \nabla_\theta p(X) \\ &+ \frac{t^2}{2} \nabla_{\tilde{\theta}} \nabla_{\tilde{\theta}} p(\tilde{\tilde{X}}), \tilde{\tilde{X}} \in \tilde{B}_\varepsilon(X) \end{aligned}$$

And the symmetry of $\tilde{B}_\varepsilon(X)$:

$$\begin{aligned} &\left| \int_{\tilde{B}_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} p_\mu(X) dV(\tilde{X}) \right| \\ &= \left| \int_{\tilde{B}_\varepsilon(X)} \big(p_\mu(\tilde{X}) - p_\mu(X) \big) dV(\tilde{X}) \right| \\ &= \Big| \int_{S^{q-1}} \int_0^\varepsilon \left(t \cdot \nabla_\theta p(X) + \frac{t^2}{2} \nabla_{\tilde{\theta}} \nabla_{\tilde{\theta}} p(\tilde{\tilde{X}}) \right) \\ &\cdot \Big(t^{q-1} + t^{q+1} Ric_{\tilde{X}_2}(\tilde{\theta}, \tilde{\theta}) \Big) dt d\theta \Big| \\ &\leq \Big| \int_{S^{q-1}} \int_0^\varepsilon t^q \cdot \nabla_\theta p(X) dt d\theta \Big| \\ &+ \Big| \varepsilon^4 \cdot \big(C_{Ric} \cdot C_{p,1} + C_{p,2} \big) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q-1} dt d\theta \Big| \\ &+ \Big| \varepsilon^6 \cdot \big(C_{Ric} \cdot C_{p,2} \big) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q-1} dt d\theta \Big| \\ &= 0 + V_q \cdot \varepsilon^{q+4} \cdot \Big(\big(C_{Ric} \cdot C_{p,1} + C_{p,2} \big) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \Big) \end{aligned}$$

So:

$$\begin{aligned} &\left| \int_{\tilde{B}_\varepsilon(X)} p_\mu(X) dV(\tilde{X}) - V_q \cdot \varepsilon^q p_\mu(X) \right| \\ &= \left| p_\mu(X) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q+1} Ric_{\tilde{X}_2}(\tilde{\theta}, \tilde{\theta}) dt d\theta \right| \\ &\leq p_\mu(X) \cdot \varepsilon^{q+2} \cdot V_q \cdot C_{Ric}. \end{aligned}$$

and for smooth density (with S3):

$$\mathbb{E}\chi = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2 \delta_{\mathbb{E}})$$

Where:

$$\begin{aligned} |\delta_{\mathbb{E}}| &\leq 8 \cdot p_{\max} + C_{Ric} \\ &+ \varepsilon^2 \cdot \Big(\big(C_{Ric} \cdot C_{p,1} + C_{p,2} \big) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \Big) \end{aligned}$$

Without S3 for $\mathbb{E}\chi$:

$$\begin{aligned} &\int_{B_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) \geq \int_{\tilde{B}_\varepsilon(X)} p_{\min} dV(\tilde{X}) \\ &- 8 \cdot V_q \cdot p_{\max} \cdot \varepsilon^{q+2} \geq \varepsilon^q \cdot V_q \cdot p_{\min} - V_q \cdot \varepsilon^{q+2} \cdot (C_{Ric} + 8 p_{\max}) \end{aligned}$$

The Lemma is proved.

Corollary 4. If:

$$\begin{aligned} \varepsilon \leq \min\{ &\frac{1}{c(\mathbb{X})}, \sqrt{\frac{24(\sqrt[q]{2} - 1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}} \\ &\frac{1}{\sqrt{C_{Ric} C_{p,2}}}, \frac{1}{\sqrt{C_{Ric} C_{p,1} + C_{p,2} + 1}} \} \end{aligned}$$

and $X \in \mathbb{X}_\varepsilon$, then in Lemma 8 for $\delta_{\mathbb{E}}$ (20):

$$|\delta_{\mathbb{E}}| \le C_{\mathbb{E}} \equiv 8 \cdot p_{\max} + C_{Ric} + 1.$$

Corollary 5. If:

$$\varepsilon \le \min\{\frac{1}{c(\mathbb{X})}, \sqrt{\frac{24(\sqrt[q]{2}-1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}}$$
$$\frac{1}{\sqrt{C_{Ric}C_{p,2}}}, \frac{1}{\sqrt{C_{Ric}C_{p,1} + C_{p,2} + 1}}\},$$

and $X \in \mathbb{X}_\varepsilon$, then:

$$\mathrm{Var}\chi \le \varepsilon^q V_q \cdot \left(p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4 C_{\mathbb{E}}^2\right)$$
$$\mathrm{Var}\chi \ge \varepsilon^q V_q \cdot \left(p_\mu(X) - C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q - \varepsilon^{q+2}C_{\mathbb{E}} - \varepsilon^4 C_{\mathbb{E}}^2\right)$$

Proof. Using Lemma 9 and Corollary 4 we transform:

$$\mathrm{Var}\chi = \mathbb{E}\chi^2 - (\mathbb{E}\chi)^2 = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}})$$
$$\cdot\left(1 - \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}})\right)$$
$$\le \varepsilon^q V_q \cdot \left(p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4 C_{\mathbb{E}}^2\right)$$

$$\mathrm{Var}\chi = \mathbb{E}\chi^2 - (\mathbb{E}\chi)^2$$
$$= \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}}) \cdot \left(1 - \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}})\right)$$
$$\ge \varepsilon^q V_q \cdot \left(p_\mu(X) - C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q - \varepsilon^{q+2}C_{\mathbb{E}} - \varepsilon^4 C_{\mathbb{E}}^2\right)$$

The Corollary is proved.

Proof of Theorem 1. Using (19):

$$\mathbb{E}N_\varepsilon(X) = N \cdot \mathbb{E}\chi$$

As $\mathbb{B}$ is open set (assumption M1), from Corollary 4 for small ε:

$$\left|\frac{N_\varepsilon(X)}{N} - \varepsilon^q V_q p_\mu(X)\right| \le \varepsilon^{q+2} V_q C_{\mathbb{E}}$$

Using Chebyshev's inequality for $\frac{N_\varepsilon(X)}{N}$ and $\varepsilon^q \cdot \psi(N)$, where (any $\psi(N) = \left(N\varepsilon^q\right)^{-1/4}$ $\psi(N) : N\varepsilon^q \cdot \psi(N)^2 \to \infty$ and ψ(N) → 0 for N → 0 is allowed), we have:

$$P\left(\left|\frac{N_\varepsilon(X)}{N} - \mathbb{E}\chi\right| \ge \varepsilon^q\psi(N)\right) \le \frac{\mathrm{Var}\chi}{N\varepsilon^{2q}\psi(N)^2}$$

So:

$$P\left(\left|\frac{N_\varepsilon(X)}{N} - \varepsilon^q V_q p_\mu(X)\right| \ge \varepsilon^{q+2}V_q C_{\mathbb{E}} + \varepsilon^q\right) \le \frac{\mathrm{Var}\chi}{N\varepsilon^{2q}\psi(N)^2}$$

Note that for N → ∝:

$$\frac{\mathrm{Var}\chi}{N\varepsilon^{2q}\psi(N)^2}$$
$$\le \frac{V_q \cdot \left(p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4 C_{\mathbb{E}}^2\right)}{N\varepsilon^q\psi(N)^2} \to 0$$
$$\varepsilon^{q+2}V_q C_{\mathbb{E}} + \psi(N) \to 0$$

So:

$$\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \xrightarrow{p} 0.$$

The Theorem is proved.

Proof of Theorem 2. As N→0 $I_\varepsilon(X_n \mid X) = 1 \Rightarrow X_n \to X$.

So:

$$p(X_n \mid I_\varepsilon(X_n \mid X) = 1) \cdot \varepsilon^q$$
$$= p_\mu(X_n) \cdot \frac{\varepsilon^q}{P(B_\varepsilon(X))} = \frac{p_\mu(X_n)}{p_\mu(X)} \cdot \frac{\varepsilon^q}{V_q \cdot \varepsilon^q} \to 1/V_q$$

The Theorem is proved.

Proof of Corollary 1. Let be an orthonormal basis in the tangent space $T_X(\mathbb{X})$ and $U_1,\ldots,U_{p-q}$ be an orthonormal basis in the cotangent space $T_X(\mathbb{X})^\perp$. Then $V_1,\ldots,V_q,U_1,\ldots,U_{p-q}$ is a basis in $\mathbb{R}^p$.

From the Theorems 1 and 2 for the elements of local covariance matrix:

$$\frac{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X) \cdot ((X_n - X)^T \cdot V_i) \cdot ((X_n - X)^T \cdot V_j)}{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X)} \to I(i = j)$$

$$\frac{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X) \cdot ((X_n - X)^T \cdot V_i) \cdot ((X_n - X)^T \cdot U_j)}{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X)} \to 0$$

$$\frac{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X) \cdot ((X_n - X)^T \cdot U_i) \cdot ((X_n - X)^T \cdot U_j)}{\sum_{n=1}^{N} I_\varepsilon(X_n \mid X)} \to 0$$

The Corollary is proved.

Proof of Theorem 3. χ (18) has Bernoulli distribution, so using Lemma 8 and it's Corollaries we get the conditions of Lemma 6 for χ:

$\mathbb{E}\chi \to 0$ and $N \cdot \mathbb{E}\chi \to \infty$ for $N \to \propto$. It remains to note that for $N \to \propto$:

$$\frac{\mathbb{E}\chi}{\varepsilon^q V_q p_{\mu(X)}} \to 1$$
$$\frac{\mathtt{Var}\chi}{\varepsilon^q V_q p_{\mu(X)}} \to 1$$

So by Lemma 8:

$$\frac{N_\varepsilon(X) - N\varepsilon^q p_\mu(X) V_q}{\sqrt{N\varepsilon^q p_\mu(X) V_q}} \xrightarrow{d} N(0,1)$$

The Theorem is proved.

Proof of Theorem 4. We verify the conditions of Lemma 7 for χ (18), $m_k = \varepsilon^q$. V_p. p_μ (X), $m = \varepsilon^{q+2} C_{\mathbb{E}}$ (Corollary 4) and:

$$|\chi - \mathbb{E}\chi| \le 1$$
$$|m_k - \mathbb{E}\chi_k| \le \varepsilon^{q+2} C_{\mathbb{E}}$$

For $\varepsilon < \min\left\{\frac{1}{\sqrt{4C_{\mathbb{E}} p_{\min}}}, \sqrt[q]{\frac{1}{4p_{\min}}}\right\}$ by Corollary 4:

$$\begin{aligned}
&\mathtt{Var}\chi \le \varepsilon^q V_q \\
&\cdot\left(p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2 \varepsilon^q + \varepsilon^{q+2} C_{\mathbb{E}} + \varepsilon^4 C_{\mathbb{E}}^2\right) \\
&\le \varepsilon^q V_q \cdot p_\mu(X)\left(1 + 1/4 + 1/4 + 1/4\right) < 2\varepsilon^q V_q \cdot p_\mu(X) \\
&\mathtt{Var}\chi \ge 1/4 \cdot \varepsilon^q V_q \cdot p_\mu(X)
\end{aligned}$$

Thus, the conditions of Lemma 7 fulfilled and for $z \le 1/16$:

$$\begin{aligned}
&P\left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \ge z + \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu}\right) \\
&\le \exp\left(-z^2 \cdot N\varepsilon^q V_q p_\mu(X)/4\right); \\
&P\left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \le -z - \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu}\right) \\
&\le \exp\left(-z^2 \cdot N\varepsilon^q V_q p_\mu(X)/4\right)
\end{aligned}$$

Similarly, if the S3 is not fulfilled, then replacing the evaluation density expansion in Taylor's formula at their rough equivalents.

The Theorem is proved.

Proof of Theorem 5. Indicator function is discontinuous, so $N_\varepsilon(X)$ is a discontinuous function too. However, due to nesting balls $\tilde{B}_{\varepsilon/3}(X) \subset \tilde{B}_\varepsilon(\tilde{X})$ for $X \in \tilde{B}_{\varepsilon/3}(\tilde{X})$ and ε:

$$N_\varepsilon(\tilde{X}) \ge N_{\varepsilon/3}(X)$$

So we get the statement of this Theorem by Corollary 3, Lemma 11 and Theorem 3.

The Theorem is proved.

Conclusion

Random variable 'the number of points in the neighborhood of a fixed point on manifold' is considered in the present paper. Points are assumed to lie on a good enough unknown manifold, the neighborhood is ball shaped and Euclidean, its radius slowly tends to zero with sample size growth.

Asymptotic expansion and uniform large deviation results are obtained for the considered random variable. The problem statement is motivated by manifold learning problems (Roweis and Saul, 2000; Zhang and Zha, 2004; Bernstein and Kuleshov, 2014).

The results of the paper could be used for the manifold learning algorithms analysis and could be generalized to get asymptotic properties of all algorithm steps.

Acknowledgement

The study was performed in the IITP RAS exclusively by the grant from the Russian Science Foundation (project № 14-50-00150).

Ethics

The author hereby declares no conflict of interest with regard to this manuscript.

Appendix A. Definitions and Lemmas from Differential Geometry

In this Section, we introduce the necessary information for further consideration related to the differential geometry.

Covariant Derivative

Differentiation on manifolds refers as usual. We recall that does it mean. It is easy to determine the derivative of a function in a given direction $V \in \mathbb{R}^p \setminus \{0\}$ at a given point $Z \in \mathbb{R}^p$ of the p-dimensional real space $\mathbb{R}^p$. However, for $X \in \mathbb{X}$ even a small displacement in all directions leaves the transfer result from the manifold. More precisely, for the manifold $\mathbb{X}$ covered with the map $(\mathbb{B}, f)$ and full rank Jacobi matrix $J_F(f^{-1}(X))$, at $X_0 = f(b_0)$ manifold locally behaves almost like a q-dimensional linear space tangent space $T_X(\mathbb{X})$, which can be defined as a linear space spanned by the p-dimensional column vectors of the matrix:

$$J_f(b_0) = \left(\frac{\partial f}{\partial b^1}(b_0) \middle| \ldots \middle| \frac{\partial f}{\partial b^q}(b_0) \right)$$

where, the superscript denotes the component of the vector. That is, one can only differentiate in directions $V(X_0) \in T_{X_0}(\mathbb{X})$. Therefore, the tangent space $T_{X_0}(\mathbb{X})$ depends on the point X_0 of the manifold, in which it is defined. Hence, on the manifold one cannot determine the derivative of a scalar function $\varphi(X)$ as the limit of changes in the function φ from point X_0 to point $X_0 + tV(X_0)$ over the length of the $tV(X_0)$, since $X_0 + tV(X_0) \notin \mathbb{X}$ and the value of φ at this point is not defined in the general case. So, instead of $X_0 + Tv(X_0)$ a curve $\gamma(t)$, $t \in (-\varepsilon, \varepsilon)$ is considered on the manifold, such that $\gamma(0) = X_0$ and $\frac{\partial \gamma}{\partial t}(0) = V(X_0)$. The derivative of a scalar function is defined as $\nabla_{V(X_0)}\varphi(X_0) = \lim_{t \to 0} \frac{\varphi(\gamma(t)) - \varphi(\gamma(0))}{t} = \lim_{t \to 0} \frac{\varphi(\gamma(t)) - \varphi(X_0)}{t}$.

Note. The identity covering exists in the case of Euclidean space $\mathbb{R}^m : \mathbb{B} = \mathbb{R}^m$ and $f(b) = b$, i.e., $X = b$. Therefore, as a curve $\gamma(t)$, one can choose $\gamma(t) = X_0 + tV(X_0)$ and the covariant derivative of the function φ coincides with the usual directional derivative.

Consider the restriction of the vector field $V(X) \in T_X(\mathbb{X})$, $X \in \mathbb{X}$ on the curve $\gamma(t)$: $\tilde{V}(t) = V(\gamma(t))$, $t \in (-\varepsilon, \varepsilon)$. The derivative of $\tilde{V}(t)$ is defined as usual:

$$\frac{\partial \tilde{V}}{\partial t}(t) = \lim_{h \to 0} \frac{\tilde{V}(t+h) - \tilde{V}(t)}{t}$$

where, $t \in (-\varepsilon, \varepsilon)$. However, the derivative $\frac{\partial \tilde{V}}{\partial t}(t)$ can be not in the $T_{\gamma(t)}(\mathbb{X})$. Therefore, one can define the covariant derivative $\frac{D\tilde{V}}{dt}(h)$ as a projection $\frac{\partial \tilde{V}}{\partial t}(t)$ on $T_{\gamma(t)}(\mathbb{X})$.

Next, we consider the equation:

$$\begin{cases} \frac{D\tilde{W}}{dt}(t) = 0 \\ \tilde{W}(0) = \tilde{W} \end{cases} \tag{21}$$

where, $W \in T_{\gamma(0)}(\mathbb{X})$. It's solution $\tilde{W}(t)$ exist and is called parallel transport of $\tilde{W}(t)$ over $\gamma(t)$ and is designated as $\tilde{W}(t) = P_{\gamma(t),\gamma(0)}\tilde{W}$. Now, to determine the derivative of the vector field $W(X)$, $X \in \mathbb{X}$ along a curve $\gamma(t)$, $t \in (-\varepsilon, \varepsilon)$. We will transfer $W(\gamma(t))$ from point $X_0 = \gamma(0)$. The result of the parallel transport $P_{\gamma(0),\gamma(t)}W(\gamma(t)) \in T_{\gamma(0)}(\mathbb{X})$, so the difference $P_{\gamma(0),\gamma(t)}W(\gamma(t)) - W(\gamma(0))$ is determined in $T_{\gamma(0)}(\mathbb{X})$ and the covariant derivative is defined for the vector field $W(X)$ on $\mathbb{X}$:

$$\nabla_{V(X_0)}W(X_0) = \lim_{t \to 0} \frac{P_{\gamma(0),\gamma(t)}W(\gamma(t)) - W(\gamma(0))}{t}$$

where, $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{X}$ and $\gamma(0) = X_0 \in \mathbb{X}$, $\gamma'(0) = V(X_0) \in T_{X_0}(\mathbb{X})$.

Note. For the Euclidian space $\mathbb{R}^m$ the solution of (21) is $P_{\gamma(0),\gamma(t)}W(\gamma(t)) = W(\gamma(t)) = W(X_0 + tV(X_0))$ and covariant derivative of the vector field equals to directional derivative:

$$\nabla_{V(X_0)}W(X_0) = \lim_{t \to 0} \frac{W(X_0 + tV(X_0)) - W(X_0)}{t}$$

We will denote the covariant derivative in the direction $V \in T_X(\mathbb{X}) \setminus \{0\}$ at X as ∇_V. The covariant derivative has the usual properties of the directional derivative, for example, linearity $V_1, V_2 \in T_X(\mathbb{X}) \setminus \{0\}$, $\alpha_1, \alpha_2 \in \mathbb{R} \setminus 0$ for $\alpha_1 V_1 + \alpha_2 V_2 \neq 0$:

$$\nabla_{\alpha_1 V_1 + \alpha_2 V_2} = \alpha_1 \nabla_{V_1} + \alpha_2 \nabla_{V_2}$$

Quadratic Forms and Ricci Curvature

In this study, it is assumed that the metric tensor of the manifold $\mathbb{X}$ is generated by embedding at multi-dimensional space $\mathbb{R}^p$. This means that the scalar product of vectors $V, W \in T_X(\mathbb{X})$ is the restriction of the scalar product of $\mathbb{R}^p$ on $T_X(\mathbb{X})$, that is, in an orthonormal basis is expressed by a bilinear form $I_X(V, W) = V^T W$. The first quadratic form of manifold is the length of the vectors of the tangent space and allows to calculate the lengths of curves:

$$I_X(V, V) = V^T V.$$

We write down the coordinates of the vector in the basis induced by parameterization $(\mathbb{B}, f)$:

$$\begin{aligned} &b = f^{-1}(X) \\ &V = J_f(b)\alpha_V \\ &\alpha_V \in \mathbb{R}^q, \alpha_V = \left(J_f(b)^T J_f(b)\right)^{-1} J_f(b)^T V \\ &I_X^c(\alpha_v) = \alpha_v^T \left(J_f(b)^T J_f(b)\right)\alpha_v \end{aligned}$$

The second bilinear form shows the orthogonal to the tangent space component change of vectors from the tangent space along different directions $V,W \in T_X(\mathbb{X})$. We write the expression of the second bilinear form $II_X(V,W)$ in the basis, given parameterization $(\mathbb{B}, f)$ and point $X = f(b)$:

$$II_X(V,W) = \pi(X)^{\perp} \cdot \sum_{i,j=1}^{q} \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{V,i} \alpha_{W,j} \tag{22}$$

where, $\alpha_V = (\alpha_{V,1},\ldots,\alpha_{V,q})^T$ and $\alpha_W = (\alpha_{W,1},\ldots,\alpha_{W,q})^T$ are coordinates of V and W in the parametric basis, $\frac{\partial^2 f(b)}{\partial b_i \partial b_j} \in \mathbb{R}^p$, $\pi(X)^{\perp} = \left(I - J_f(b)\left(J_f^T(b) J_f(b)\right)^{-1} J_f(b)^T\right)$ is a projector on cotangent space $\left(T_X(\mathbb{X})\right)^{\perp}$.

The second fundamental form of $II_X(V, V)$ determines the normal curvature of the manifold.

To determine the Ricci curvature we introduce the notation for the Christoffel symbols. Define:

$$g_{ij} = \frac{\partial f(b)}{\partial b_i} \quad \Gamma_{jk,l} = \frac{1}{2}\left(\frac{\partial g_{jl}}{\partial b_k} + \frac{\partial g_{kl}}{\partial b_j} - \frac{\partial g_{jk}}{\partial b_l}\right) \tag{23}$$

$\Gamma_{jk,l}$ are called Christoffel symbols of the first kind. Let g^{ij} be elements of the inverse matrix $J_F^T(b) J_F(b)$, i.e. the elements of the matrix $\left(J_f^T(b) J_f(b)\right)^{-1}$. The Christoffel symbols of the second kind:

$$\Gamma_{jk}^{i} = \sum_{l=1}^{q} \Gamma_{jk,l} \cdot g^{li} \tag{24}$$

Elements of the Ricci curvature tensor are defined as:

$$R_{ij} = \sum_{k=1}^{q}\left(\frac{\partial \Gamma_{ji}^{k}}{\partial b_k} - \frac{\partial \Gamma_{ki}^{k}}{\partial b_j} + \sum_{l=1}^{q} \Gamma_{kl}^{k}\Gamma_{ji}^{l} - \Gamma_{jl}^{k}\Gamma_{ki}^{l}\right) \tag{25}$$
$$i,j = 1,\ldots,q$$

Let $R = (R_{ij})_{i,j=1}^{q}$ be the matrix of rank q. For $\theta \in T_X(\mathbb{X})$: $\alpha_\theta = \left(J_f(b)^T J_f(b)\right)^{-1} J_f(b)^T \theta$. Ricci curvature in the direction of θ at X:

$$Ric_X(\theta,\theta) = \alpha_\theta^T \cdot R \cdot \alpha_\theta \tag{26}$$

Ricci curvature describes the difference between Euclidean volume element and the manifold volume element.

Locally Riemannian Coordinates and Exponential Map

For a small neighborhood of $X_0 \in \mathbb{X}$ mapping $X \mapsto V \in T_X(\mathbb{X})$: $X = f(\gamma(1))$, where γ: $\gamma(0) = X_0$ and $\frac{\partial \gamma}{\partial t}(0) = V$, defines a one-to-one correspondence between the neighborhood of X_0 and the neighborhood of 0 in $T_{X_0}(\mathbb{X})$. The inverse mapping is called exponential and is denoted $X = \exp_{X_0}(V)$.

In this neighborhood vectors $V \in T_{X_0}(\mathbb{X})$ are coordinates. These coordinates are called locally Riemann coordinates.

> *Note. For the Euclidian space $\mathbb{X} = \mathbb{R}^m$ the tangent space is $T_X(\mathbb{X}) = \mathbb{R}^m$ and $\exp_{X_0}(V) = X_0 + V$ for each $V \in T_X(\mathbb{X})$ (that is, the result is valid for an arbitrary neighborhood of X_0).*

The distance between the points of the manifold in p-dimensional space and the distance between them in the q-dimensional Riemannian coordinates are related with Lemmas 1 and 2. Lemmas 3 and 4 roughly upper bound the second fundamental form and Ricci curvature through a fairly smooth parameterization of the manifold.

We prove these Lemmas:

Proof of Lemma 3. For an arbitrary point $b \in \mathbb{B}$ denote by $V_1,\ldots,V_q$ orthonormal basis of eigenvectors of $J_f(b)^T J_f(b)$. Let $\lambda_1,\ldots,\lambda_q$ be the corresponding to $V_1,\ldots,V_q$ eigenvalues. Let $\theta = \sum_{i=1}^{q} \beta_i V_i$, where $\sum_{i=1}^{q} \beta_i^2 = 1$, be the coordinates $\theta \in T_{f(b)}(\mathbb{X}) : \|\theta\| = 1$ is the basis $V_1,\ldots,V_q$. Using (5):

$$\begin{aligned}
\alpha_\theta &= \left(J_f(b)^T J_f(b)\right)^{-1} J_f(b)^T \theta \\
&= \left(J_f(b)^T J_f(b)\right)^{-1} J_f(b)^T \sum_{i=1}^{q} \beta_i V_i; \\
\|\alpha_\theta\|^2 &= \sum_{i=1}^{q} \frac{1}{\lambda_i}\beta_i^2; \ \|\alpha_\theta\|^2 \le \frac{1}{c_J}.
\end{aligned} \tag{27}$$

Using $X = f(b)$ (7), (27):

$$\|\mathrm{II}_X(\theta,\theta)\| = \left\|\pi^{\perp}(X) \cdot \sum_{i,j=1}^{q} \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{\theta,i}\alpha_{\theta,j}\right\|$$
$$\le \left\|\sum_{i,j=1}^{q} \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{\theta,i}\alpha_{\theta,j}\right\| \le C_H \cdot \|\alpha_\theta\|_1^2 \le C_H \cdot q \cdot \frac{1}{c_J}$$

The Lemma is proved.

Proof of Lemma 4. Using (23) and (24) we estimate (25) by triangle inequality:

$$R_{ij}|=\left|\sum_{k=1}^{q}\left(\frac{\partial\Gamma_{ji}^{k}}{\partial b_k}-\frac{\partial\Gamma_{ki}^{k}}{\partial b_j}+\sum_{l=1}^{q}\Gamma_{kl}^{k}\Gamma_{ji}^{l}-\Gamma_{jl}^{k}\Gamma_{ki}^{l}\right)\right|X_1,\ldots,X_n$$

$$\le 2q\cdot\max_{i,j,k=1,\ldots,q}\left|\frac{\partial\Gamma_{ji}^{k}}{\partial b_k}\right|+2q^2\cdot\left(\max_{j,i,k=1,\ldots,q}\left|\Gamma_{ji}^{l}\right|\right)^2$$

$$\left|\Gamma_{jk}^{i}\right|=\left|\sum_{l=1}^{q}\Gamma_{jk,l}g^{li}\right|=\left|\sum_{l=1}^{q}\frac{1}{2}\left(\frac{\partial g_{jl}}{\partial b_k}+\frac{\partial g_{kl}}{\partial b_j}-\frac{\partial g_{jk}}{\partial b_l}\right)g^{li}\right|$$

$$\le\frac{3q}{2}\cdot\max_{k,l,j=1,\ldots,q}\left|\frac{\partial g_{jl}}{\partial b_k}\right|\cdot\max_{k,l=1,\ldots,q}\left|g^{kl}\right|$$

$$\left|\frac{\partial g_{jl}}{\partial b_k}\right|=\left|\frac{\partial\left(\frac{\partial f(b)}{\partial b_j},\frac{\partial f(b)}{\partial b_l}\right)}{\partial b_k}\right|$$

$$\le 2\cdot\max_{j,k=1,\ldots,q}\left\|\frac{\partial^2 f(b)}{\partial b_j\partial b_k}\right\|\cdot\max_{j=1,\ldots,q}\left\|\frac{\partial f(b)}{\partial b_j}\right\|$$

$$\le 2\cdot C_H\cdot\sqrt{C_J};\max_{k,l=1,\ldots,q}\left|g^{kl}\right|\le 1/\sqrt{c_J}$$

$$\left|\frac{\partial\Gamma_{ji}^{k}}{\partial b_k}\right|=\left|\sum_{l=1}^{q}\frac{\partial\Gamma_{jk,l}}{\partial b_k}g^{li}-\sum_{l_1,i_1=1}^{q}\Gamma_{jk,l}\cdot g^{ll_1}\frac{\partial g_{l_1i_1}}{\partial b_k}g^{i_1i}\right|$$

$$\le q\cdot\max_{j,k,l=1,\ldots,q}\left|\frac{\partial\Gamma_{jk,l}}{\partial b_k}\right|\cdot\max_{k,l=1,\ldots,q}\left|g^{kl}\right|$$

$$+q^3\cdot\max_{j,k,l=1,\ldots,q}\left|\Gamma_{jk,l}\right|\cdot\left(\max_{k,l=1,\ldots,q}\left|g^{kl}\right|\right)^2$$

$$\max_{j,l,k=1,\ldots,q}\left|\frac{\partial g_{jl}}{\partial b_k}\right|$$

$$\le q/\sqrt{c_J}\cdot\max_{j,k,l=1,\ldots,q}\left|\frac{\partial\Gamma_{jk,l}}{\partial b_k}\right|+2q^3\cdot C_H\cdot\sqrt{C_J}/c_J$$

$$\left|\frac{\partial\Gamma_{jk,l}}{\partial b_k}\right|\le 3/2\cdot\max_{i,j,k,m=1,\ldots,q}\left|\frac{\partial^2\left(\frac{\partial f(b)}{\partial b_j},\frac{\partial f(b)}{\partial b_l}\right)}{\partial b_k\partial b_m}\right|$$

$$\le 3\cdot\max_{i,j,k,m=1,\ldots,q}\left|\frac{\partial\left(\frac{\partial^2 f(b)}{\partial b_j\partial b_k},\frac{\partial f(b)}{\partial b_l}\right)}{\partial b_m}\right|\le 3\cdot\left(C_H^2+C_T\cdot\sqrt{C_J}\right)$$

So:

$$|R_{ij}|\le 2q^2\cdot\frac{C_H^2+C_T\sqrt{C_J}}{\sqrt{c_J}}+q^4\cdot\left(18\cdot C_H^2\frac{C_J}{c_J}+4\cdot C_H\frac{\sqrt{C_J}}{c_J}\right)$$

Using (26) and (27):

$$\|Ric_X(\theta,\theta)\|=\|\alpha_\theta^T\cdot R\cdot\alpha_\theta\|\le\max_{i,j=1,\ldots,q}|R_{ij}|\cdot\|\alpha_\theta\|_1^2$$

$$\le 2q^3\cdot\frac{C_H^2+C_T\sqrt{C_J}}{c_J^{3/2}}+q^5\cdot\left(18\cdot C_H^2\frac{C_J}{c_J}+4\cdot C_H\frac{\sqrt{C_J}}{c_J}\right)/c_J$$

The Lemma is proved.

Appendix B. Lemmas Proofs

In this section, Lemmas used in the proof of the main theorems are listed and proved.

De Moivre-Laplace for Slowly Decreasing Success Probability

We prove local and integral de Moivre-Laplace lemmas for slowly tending to zero success parameter.

Proof of Lemma 5. The proof almost repeats the proof of the local limit theorem from Shiryayev (1984) and essentially uses the Stirling formula $n!=\sqrt{2\pi n}\cdot e^{-n}n^n\cdot\left(1+R(n)\right)$, where $R(n)=\frac{1}{12n}+\frac{1}{288n^2}-\frac{139}{51840n^3}+O(\frac{1}{n^4})$ and $\frac{1}{12n}>R(n)>\frac{1}{13n},\ n\ge 1$ could be written via Bernoulli numbers. We will omit subscript n in p_n и q_n for convenience.

For $n\to\infty,k\to\infty,n-k\to\infty$:

$$C_n^k=\frac{n!}{k!(n-k)!}$$

$$=\frac{\sqrt{2\pi n}\cdot e^{-n}n^n}{\sqrt{2\pi k\cdot 2\pi(n-k)}\cdot e^{-k}k^k\cdot e^{-(n-k)}(n-k)^{n-k}}$$

$$=\frac{1}{\sqrt{2\pi n\frac{k}{n}\cdot\left(1-\frac{k}{n}\right)}}\cdot\frac{1+r}{\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k}}$$

Where:

$$r=r(n,k,n-k)=\frac{1+R(n)}{(1+R(k))(1+R(n-k))}$$

Estimate r for $k>1$, n-k>1:

$$1>\frac{1+R(n)}{(1+R(k))(1+R(n-k))}>\left(1+\frac{1}{13n}\right)\left(1-\frac{1}{12k}\right)$$

$$\cdot\left(1-\frac{1}{12(n-k)}\right)>1-\frac{1}{12k}-\frac{1}{12(n-k)}$$

$$|r|\le\frac{1}{12k}+\frac{1}{12(n-k)}$$

So:

$$p(k) = C_n^k p^k q^{n-k} = \frac{1}{\sqrt{2\pi n \frac{k}{n}\left(1-\frac{k}{n}\right)}} \cdot \frac{p^k (1-p)^{n-k}}{\left(\frac{k}{n}\right)^k \left(1-\frac{k}{n}\right)^{n-k}} \cdot (1+r)$$

Denote $\hat{p} = \frac{k}{n}$, $\hat{q} = 1-\hat{p}$. $p - \hat{p} = \hat{q} - q$ because of $p + q = \hat{p} + \hat{q} = 1$. Also, $\frac{p-\hat{p}}{p}$ and $\frac{q-\hat{q}}{q} = -\frac{p-\hat{p}}{q}$ are small parameters as $\frac{p-\hat{p}}{pq} = \frac{np-k}{npq}$ is small from the lemmas assumptions and $0 < p,q < 1$. So:

$$\begin{aligned} p(k) &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \left(\frac{p}{\hat{p}}\right)^k \left(\frac{1-p}{1-\hat{p}}\right)^{n-k} \cdot (1+r) \\ &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\left(k \ln\frac{p}{\hat{p}} + (n-k)\ln\frac{1-p}{1-\hat{p}}\right)^{n-k} \cdot (1+r) \\ &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\left(n \cdot \left(\hat{p}\ln\frac{p}{\hat{p}} + \hat{q}\ln\frac{q}{\hat{q}}\right)\right) \cdot (1+r) \\ &= \frac{1}{\sqrt{2\pi n\hat{p}(1-\hat{p})}} \exp\Bigl(-n \cdot \Bigl(p \cdot \left(1+\frac{\hat{p}-p}{p}\right) \ln\left(1+\frac{\hat{p}-p}{p}\right) \\ &\quad + q \cdot \left(1-\frac{\hat{p}-p}{q}\right) \ln\left(1-\frac{\hat{p}-p}{q}\right)\Bigr)\Bigr) \cdot (1+r) \end{aligned}$$

We use an expansion in Taylor's formula with the remainder term in the Lagrange form:

$$\begin{aligned} &\left(1+\frac{\hat{p}-p}{p}\right)\ln\left(1+\frac{\hat{p}-p}{p}\right) = -\frac{\hat{p}-p}{p} + \frac{1}{2}\cdot\left(\frac{\hat{p}-p}{p}\right)^2 \\ &+ \left(\frac{1}{2} - \frac{1}{3}\cdot\frac{1}{\left(1+\delta_p\right)^3}\right)\cdot\left(\frac{\hat{p}-p}{p}\right)^2 \\ &\left(1-\frac{\hat{p}-p}{q}\right)\ln\left(1-\frac{\hat{p}-p}{q}\right) = \frac{\hat{p}-p}{q} + \frac{1}{2}\cdot\left(\frac{\hat{p}-p}{q}\right)^2 \\ &+ \left(-\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{\left(1-\delta_q\right)^3}\right)\cdot\left(\frac{\hat{p}-p}{q}\right)^2 \end{aligned}$$

$$\begin{aligned} &p\left(1+\frac{\hat{p}-p}{p}\right)\ln\left(1+\frac{\hat{p}-p}{p}\right) + q\left(1-\frac{\hat{p}-p}{q}\right)\ln\left(1-\frac{\hat{p}-p}{q}\right) \\ &= \frac{1}{2}\left(\frac{1}{p}+\frac{1}{q}\right)(\hat{p}-p)^2 + \left(\frac{1}{2} - \frac{1}{3}\cdot\frac{1}{\left(1+\delta_p\right)^2}\right)\cdot\frac{(\hat{p}-p)^3}{p^2} \\ &+ \left(-\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{\left(1-\delta_q\right)^2}\right)\cdot\frac{(\hat{p}-p)^3}{q^2} \end{aligned}$$

where, $\delta_p \in \left[\min\{0, \frac{\hat{p}-p}{p}\}, \max\{0, \frac{\hat{p}-p}{p}\}\right]$ and $\delta_q \in \left[\min\{0, \frac{\hat{p}-p}{q}\}, \max\{0, \frac{\hat{p}-p}{q}\}\right]$. Also $\frac{1}{p} + \frac{1}{q} = \frac{p+q}{pq} = \frac{1}{pq}$ and $\frac{n}{2}\left(\frac{1}{p}+\frac{1}{q}\right)(\hat{p}-p)^2 = \frac{n}{2pq}\left(\frac{k}{n}-p\right)^2 = \frac{(k-np)^2}{2npq}$.

So:

$$\frac{n}{2pq}\left(\frac{k}{n}-p\right)^2 = \frac{(k-np)^2}{2npq}$$

We get:

$$p(k) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) \cdot (1 + \tilde{r}(n,k,n-k))$$

Where:

$$\begin{aligned} &1 + \tilde{r}(n,k,n-k) = (1 + r(n,k,n-k)) \\ &\cdot \exp\Bigl(-n \cdot \Bigl(\left(\frac{1}{2} - \frac{1}{3}\cdot\frac{1}{\left(1+\delta_p\right)^2}\right) \\ &\cdot \frac{(\hat{p}-p)^3}{p^2} + \left(-\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{\left(1-\delta_q\right)^2}\right)\cdot\frac{(\hat{p}-p)^3}{q^2}\Bigr)\Bigr)\sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}} \end{aligned}$$

Finally, $\sup|\tilde{r}(n,k,n-k)| \to 0,\ n \to \infty$ for $k : |k-np|/(npq)^{2/3} \to 0$.

The Lemma is proved.

Proof of Lemma 6. Let $\infty < a \le b < \infty$ $P_n(a,b] = \sum_{a<z\le b} P_n(np_n + z\sqrt{np_n q_n})$, where the sum is over such z that $np_n + z\sqrt{np_n q_n}$ is an integer. From Lemma 5 for each t_k such that $k = np_n + t_k\sqrt{np_n q_n}$ and $|t_k| \le T < \infty$:

$$P_n(np_n + t_k\sqrt{np_n q_n}) = \frac{\delta_k}{\sqrt{2\pi}} \exp\left(-t_k^2/2\right)(1 + \varepsilon(t_k, n))$$

where, $\sup_{|t_k|\le T} |\varepsilon(t_k,n)| \to 0,\ n \to \infty$ and $\delta_k = \frac{1}{\sqrt{np_n q_n}}$.

So for the fixed a and b such that $-T \le a \le b \le T,\ T \le \infty$:

$$\begin{aligned} &\sum_{a<t_k\le b} P_n\left(np + t_k\sqrt{npq}\right) \\ &= \sum_{a<t_k\le b} \frac{\delta_k}{\sqrt{2\pi}} \exp\left(-t_k^2/2\right) + \sum_{a<t_k\le b} \varepsilon(t_k,n)\frac{\delta_k}{\sqrt{2\pi}} \exp\left(-t_k^2/2\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-z^2/2\right)dx + R_n^{(1)}(a,b) + R_n^{(2)}(a,b) \end{aligned}$$

Where:

$$R_n^{(1)}(a,b) = \sum_{a<t_k\le b} \frac{\delta_k}{\sqrt{2\pi}} \exp\left(-t_k^2/2\right) - \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-z^2/2\right)dx;$$

$$R_n^{(2)}(a,b) = \sum_{a<t_k\le b} \varepsilon(t_k,n)\frac{\delta_k}{\sqrt{2\pi}} \exp\left(-t_k^2/2\right).$$

From the properties of the integral sums $\sup_{-T\le a\le b\le b} |R_n^{(1)}(a,b)| \to 0, n\to\infty.$

Given the non-negativity of the integrand:

$$\frac{1}{\sqrt{2\pi}}\int_{-T}^{T} \exp\left(-z^2/2\right)dz \le \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-z^2/2\right)dx$$

That is:

$$\sup_{-T\le a\le b\le T} |R_n^{(2)}(a,b)| \le \sup_{|t_k|\le T} |\varepsilon(t_k,n)| \cdot\left(\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\exp\left(-z^2/2\right)dz + \sup_{-T\le a\le b\le T}|R_n^{(1)}(a,b)|\right) \to 0$$

Denote $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-t^2/2\right)dt.$ Get:

$$\sup_{-T\le a\le b\le T} |P_n(a,b] - (\Phi(b)-\Phi(a))| \to 0,\ n\to\infty. \quad (28)$$

Now for T = ∝: $\forall\varepsilon>0\ \exists T = T(\varepsilon)>0$ $\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\exp\left(-z^2/2\right)dz > 1-\frac{\varepsilon}{4}$.

From (28):

$$\exists N:\ \forall n>N:\ \sup_{-T\le a\le b\le T} |P_n(a,b] - (\Phi(b)-\Phi(a))| < \frac{\varepsilon}{4}$$

So $P_n(-T,T] > 1-\frac{\varepsilon}{2},\quad P_n(-\infty,-T] + P_n[T,\infty) < \frac{\varepsilon}{2},$ where $P_n(-\infty,-T] = \lim_{S\to-\infty} P_n(S,-T]$ и $P_n(T,\infty) = \lim_{S\to-\infty} P_n[T,S)$. Finally $\forall T:\ -\infty\le -T\le a\le b\le\infty$:

$$\begin{aligned}
&\left|P_n(a,b] - \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-z^2/2)dz\right| \\
&\le \left|P_n(-T,T] - \frac{1}{\sqrt{2\pi}}\int_{-T}^{T} \exp(-z^2/2)dz\right| \\
&+\left|P_n(a,-T] - \frac{1}{\sqrt{2\pi}}\int_a^{-T} \exp(-z^2/2)dz\right| \\
&+\left|P_n(T,b] - \frac{1}{\sqrt{2\pi}}\int_T^{b} \exp(-z^2/2)dz\right| \le \frac{\varepsilon}{4} + P_n(-\infty,-T] \\
&+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-T}\exp(-z^2/2)dz + P_n(T,\infty) \\
&+\frac{1}{\sqrt{2\pi}}\int_{T}^{\infty}\exp(-z^2/2)dz \le \frac{\varepsilon}{4}+\frac{\varepsilon}{2}+\frac{\varepsilon}{8}+\frac{\varepsilon}{8} = \varepsilon.
\end{aligned}$$

That is $P_n(a,b] \to \Phi(b)-\Phi(b)$ for all $-\infty\le a<b\le\infty$. The Lemma is proved.

The Probability of Large Deviations for Bounded Random Variables

Theorem 6. (from Petrov (1987)). Let $x_1,\ldots,x_n$ be independent random variables, $\mathbb{E}\chi_k = 0$, $\sigma_k^2 = \mathbb{E}\chi_k^2 < \infty,\ k=1,\ldots,n$, $B=\sum_{k=1}^{n}\sigma_k^2$. Let $H>0$ be such constant that $|\mathbb{E}\chi_k^m| \le \frac{m!}{2}\cdot\sigma_k^2\cdot H^{m-2},\ k=1,\ldots,n$ for all integer $m\ge 2$. Let $S=\sum_{k=1}^{n}\chi_k$ then for $0\le x\le B/H$:

$$P(S\ge x)\le \exp(-x^2/4B)$$
$$P(S\le -x)\le \exp(-x^2/4B)$$

and for $x\ge B/H$:

$$P(S\ge x)\le \exp(-x/4H)$$
$$P(S\le -x)\le \exp(-x/4H)$$

We prove the lemma about the probability of large deviations for bounded random variables:

Proof of Lemma 7. Modify:

$$\begin{aligned}
\overline{\chi} &= \frac{1}{n}\sum_{k=1}^{n}\chi_k - m_k = \frac{1}{n}\sum_{k=1}^{n}(\chi_k - \mathbb{E}\chi_k) + (\mathbb{E}\chi_k - m_k) \\
&= \frac{1}{n}\sum_{k=1}^{n}(\chi_k - \mathbb{E}\chi_k) + \frac{1}{n}\sum_{k=1}^{n}(\mathbb{E}\chi_k - m_k)
\end{aligned}$$

For the first sum multiplied by t, we use Theorem 6. To do this, check the conditions of the theorem:

$$\begin{aligned}
&\mathbb{E}(\chi_k - \mathbb{E}\chi_k) = 0 \\
&\mathbb{E}(\chi_k - \mathbb{E}\chi_k)^2 = \sigma^2 < \infty \\
&\left|\mathbb{E}(\chi_k - \mathbb{E}\chi_k)^m\right| \le \mathbb{E}\left|\chi_k - \mathbb{E}\chi_k\right|^m \\
&\le \mathbb{E}\left|\chi_k - \mathbb{E}\chi_k\right|^2\cdot C^{m-2} \\
&\le \sigma^2\cdot C^{m-2} \le \frac{m!}{2}\sigma^2 C^{m-2}
\end{aligned}$$

Thus, the conditions of Theorem 6 fulfilled for $H\ge 2C$. For $\chi_1 - \mathbb{E}\chi_1,\ldots,\chi_n-\mathbb{E}\chi_n$ using $\left|\frac{1}{n}\sum_{k=1}^{n}(\chi_k - \mathbb{E}\chi_k) - \overline{\chi}\right| \le m$ we get the result. The Lemma is proved.

Integration Area Replacement

Proof of Lemma 8. We transform difference of considered in the lemma integrals for $\varepsilon < \frac{1}{c(\mathbb{X})}$:

$$\int_{B_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X})$$
$$= \int_{B_\varepsilon(X)\setminus\tilde{B}_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X})$$
$$-\int_{\tilde{B}_\varepsilon(X)\setminus B_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X}).$$

From Lemma 1 follows that each element from $B_\varepsilon(X)\Delta\tilde{B}_\varepsilon(X)$ is not far from $\tilde{B}_\varepsilon(X)$ as $\tilde{\varepsilon} = \frac{1}{24} \| \mathrm{II}_{\tilde{X}}(\tilde{\theta},\tilde{\theta}) \| \varepsilon^3$ for some $\tilde{X} \in \mathbb{X}$ and $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{X}) : \| \theta \| = 1$. From Lemma 3: $\tilde{\varepsilon} \le \frac{1}{24} C_{II}\varepsilon^3$.

For $\varepsilon \le \sqrt{\frac{24\cdot(\sqrt[q]{2}-1)}{\max\{1, C_{II}\}}} : (\varepsilon+\tilde{\varepsilon})^q \le 2\cdot\varepsilon^q$.

For $\varepsilon \le \frac{1}{2\sqrt{C_{Ric}}} : 1 + C_{Ric}(\varepsilon+\tilde{\varepsilon})^2 \le 2$.

Using Lemmas 2 and 4:

$$\left|\int_{B_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} g(X,\tilde{X})dV(\tilde{X})\right|$$
$$\le \left|\int_{\tilde{B}_{\varepsilon+\tilde{\varepsilon}}(X)} g(X,\tilde{X})dV(\tilde{X}) - \int_{\tilde{B}_{\varepsilon-\tilde{\varepsilon}}(X)} g(X,\tilde{X})dV(\tilde{X})\right|$$
$$\le \left((\varepsilon+\tilde{\varepsilon})^q - (\varepsilon-\tilde{\varepsilon})^q\right)\cdot V_q$$
$$\cdot \sup_{X,\tilde{X}} |g(X,\tilde{X})| \cdot \left(1+(\varepsilon+\tilde{\varepsilon})^2\cdot C_{Ric}\right)$$
$$\le V_q \cdot \sup_{X,\tilde{X}} |g(X,\tilde{X})| \cdot \tilde{2}\varepsilon \cdot \left(1+4\cdot\varepsilon^2\cdot C_{Ric}\right)$$
$$\cdot \sum_{m=0}^{q-1} (\varepsilon+\tilde{\varepsilon})^m \cdot (\varepsilon-\tilde{\varepsilon})^{q-m-1}$$
$$\le 2\cdot V_q \cdot \sup_{X,\tilde{X}} |g(X,\tilde{X})| \cdot \tilde{\varepsilon}\cdot q\cdot(\varepsilon+\tilde{\varepsilon})^{q-1}\left(1+4\cdot\varepsilon^2\cdot C_{Ric}\right)$$
$$\le \varepsilon^{q+2}\cdot 2\cdot V_q \cdot \left(1+\frac{1}{24}C_{II}\varepsilon^2\right)^q \cdot \left(1+4\cdot\varepsilon^2\cdot C_{Ric}\right)\cdot \sup_{X,\tilde{X}} |g(X,\tilde{X})|$$
$$\le 8\cdot V_q \cdot \sup_{X,\tilde{X}} |g(X,\tilde{X})| \cdot \varepsilon^{q+2}$$

The Lemma is proved.

Finite Nets

Lemma 10. For each δ exists finite δ-net on a manifold $\mathbb{X}$ with $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ or fewer elements, where $a > 0$ is an edge of circumscribed p-dimensional hypercube.

Proof. Thus $\mathbb{X} \subset \mathbb{R}^p$ is bounded it could be placed into a hypercube C_a with edge $a > 0$. Let δ > 0 be a fixed number. Consider a uniform grid G(δ) for the cube with the distances between points along each edge at most $\delta/\sqrt{p}$. The number of points in the grid does not exceed $\left(\frac{a\sqrt{p}}{\delta}\right)^p$. Cube C_a is divided by net G(δ) into small cubes with edges not exceeding $\delta/\sqrt{p}$.

Therefore for every point Z of the cube G(δ) the distance between the and does not exceed the length of the diagonal of the small cube, which it belongs to:

$$d(Z, G(\delta)) \le \sqrt{\sum_{k=1}^{p} \frac{\delta^2}{p}} = \delta$$

where, $d(Z,A) = \inf_{Z'\in A} \| Z - Z' \|$ is a distance between point $Z \in \mathbb{R}^p$ and set $A \subset \mathbb{R}^p$.

Therefore on a manifold G(δ) for every point $X \in \mathbb{X}$ contains a point distant from it by no more than δ. But $G(\delta) \not\subset \mathbb{X}$.

Denote the ball with center X and radius δ as B_δ (X). Denote $\tilde{G}(\delta)$: For each point $X \in G(\delta/2)$, if $\tilde{B}_{\delta/2}(X) = B_\delta(X) \cap \mathbb{X} \ne \varnothing$, we get $\tilde{X}$ from $\tilde{B}_\delta$ and add it to $\tilde{G}(\delta)$. The set $\tilde{G}(\delta)$ is a δ-net for $\mathbb{X}$ and contains not more than $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ points. The Lemma is proved.

We prove the lemma about the joint occurrence of events system:

Lemma 11. Let each of the events A_1, *dots*, A_M occurs with a probability of not less than p. Then they all come together with a probability of at least $P(\cap_{m=1}^M A_m) \ge 1 - M\cdot(1-p)$.

Proof. Let $\bar{A}$ be the complement of a Borel set A to the set of elementary events. Transform:

$$\cap_{m=1}^M A_m \subset \overline{\cap_{m=1}^M \bar{A}_m}; P(\cap_{m=1}^M A_m) \ge P(\overline{\cap_{m=1}^M \bar{A}_m})$$
$$= 1 - P(\cap_{m=1}^M \bar{A}_m) \ge 1 - \sum_{m=1}^M P(\bar{A}_m) \ge 1 - M\cdot(1-p)$$

The Lemma is proved.

References

Belkin, M. and P. Niyogi, 2003. Laplacian eigenmaps for dimensionality reduction and data representation. J. Neural Computation, 15: 1373-1396. DOI: 10.1162/089976603321780317

Bengio, Y., A. Courville and P. Vincent, 2013. Representation learning: A review and new perspectives. IEEE Transactions Pattern Analysis Machine Intelligence, 35: 1798-1828. DOI: 10.1109/TPAMI.2013.50

Bengio, Y., J.F. Paiement and P. Vincent, 2003. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps and spectral clustering. Proceedings of the Advances in Neural Information Processing Systems, (ANI' 03), pp: 177-184.

Bernstein, A. and A. Kuleshov, 2014. Low-Dimensional Data Representation in Data Analysis. In: Artificial Neural Networks in Pattern Recognition, Mana, N., F. Schwenker and E. Trentin, (Eds.), Springer Berlin Heidelberg, Berlin, ISBN-10: 3642332110, pp: 47-58.

Campadelli, P., E. Casiraghi, C. Ceruti and A. Rozza, 2015. Intrinsic dimension estimation: Relevant techniques and a benchmark framework. Mathematical Problems Eng., 2015: 1-21. DOI: 10.1155/2015/759567

Coifman, R.R. and S. Lafon, 2006. Diffusion maps. Applied Comput. Harmonic Analysis, 21: 5-30. DOI: 10.1016/j.acha.2006.04.006

Cox, M.A.A. and T.F. Cox, 2008. Multidimensional Scaling. In: Handbook of Data Visualization, Berlin, Heidelberg: Springer Berlin Heidelberg, pp: 315-347.

Donoho, D.L., 2000. High-dimensional data analysis: The curses and blessings of dimensionality. Proceedings of the AMS Conference on Math Challenges of 21st Century, (CMC' 00), pp: 1-31.

Donoho, D.L. and C. Grimes, 2003. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proc. Nati. Acad. Sci., 100: 5591-5596. PMID: 16576753

Einbeck, J. and Z. Kalantana, 2013. Intrinsic dimensionality estimation for high-dimensional data sets: New approaches for the computation of correlation dimension. J. Emerg. Technol. Web Intelli., 5: 91-97. DOI: 10.4304/jetwi.5.2.91-97

Fan, M., H. Qiao and B. Zhang, 2009. Intrinsic dimension estimation of manifolds by incising balls. Patt. Recognit., 42: 780-787. DOI: 10.1016/j.patcog.2008.09.016

Farahmand, A.M., C. Szepesvári and J.Y. Audibert, 2007. Manifold-adaptive dimension estimation. Proceedings of the 24th International Conference on Machine Learning (ICML '07), New York, pp: 265-272.

Goldberg, Y., A. Zakai, D. Kushnir and Y. Ritov, 2008. Manifold learning: The price of normalization. J. Mach. Learn. Re., 9: 1909-1939.

Jolliffe, I.T., 2002. Principal Component Analysis. 1st Edn., Springer Science and Business Media, New York, ISBN-10: 0387954422, pp: 487.

Kuleshov, A. and A. Bernstein, 2016. Extended regression on manifolds estimation. Proceedings of the 5th International Symposium on Conformal and Probabilistic Prediction with Applications, (ISC' 16), New York, pp: 208-228. DOI: 10.1007/978-3-319-33395-3_15

Kuleshov, A. and A. Bernstein, 2014. Manifold Learning in Data Mining Tasks. In: Machine Learning and Data Mining in Pattern Recognition, Perner, P. (Ed.), Springer, Cham, ISBN-10: 331908979X, pp: 119-133.

Lee, J.A. and M. Verleysen, 2007. Nonlinear Dimensionality Reduction. 1st Edn., Springer Science and Business Media, ISBN-10: 038739351X, pp: 309.

Levina, E. and P.J. Bickel, 2005. Maximum likelihood estimation of intrinsic dimension. Proceedings of the Advances in Neural Information Processing Systems, (ANI' 05), MIT Press, pp: 777-784.

Ma, Y. and Y. Fu, 2011. Manifold Learning Theory and Applications. 1st Edn., CRC Press, Boca Raton, ISBN-10: 1466558873, pp: 314.

Niyogi, P., S. Smale and S. Weinberger, 2008. Finding the homology of submanifolds with high confidence from random samples. Discrete Computational Geometry, 39: 419-441.

Pennec, X., 1999. Probabilities and Statistics on Riemannian Manifolds: Basic Tools For Geometric Measurements. Proceedings of the Workshop on Nonlinear Signal and Image Processing, Jun. 20-23, Antalya, Turkey, pp: 194-198.

Petersen, P., 2006. Riemannian Geometry. Springer New York.

Petrov, V.V., 1987. Limit theorems for sums of independent random variables. Moscow, Nauka.

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

Rozza, A., G. Lombardi, M. Rosa, E. Casiraghi and P. Campadelli, 2011. IDEA: Intrinsic dimension estimation algorithm. Proceedings of the 16th International Conference, Sep. 14-16, Ravenna, Italy, pp: 433-442. DOI: 10.1007/978-3-642-24085-0_45

Saul, L.K., S.T. Roweis and Y. Singer, 2003. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. J. Machine Learning Research, 4: 119-155.

Seung, H.S. and D.D. Lee, 2000. Cognition. The manifold ways of perception. Science, 290: 2268-2269. PMID: 11188725

Shiryayev, A.N., 1984. Probability. 1st Edn., Springer-Verlag, New York.

Singer, A. and H.T. Wu, 2012. Vector diffusion maps and the connection Laplacian. Commun. Pure Applied Math., 65: 1067-1144. DOI: 10.1002/cpa.21395FVAC

Smith, A., X. Huo and H. Zha, 2008. Convergence and Rate of Convergence of a Manifold-Based Dimension Reduction Algorithm. Proceedings of the Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Dec. 8-11, Vancouver, British Columbia, pp: 1529-1536.

Smith, A., H. Zha and X. Wu, 2009. Convergence and Rate of Convergence of a Manifold-Based Dimension Reduction Algorithm. Advances Neural Information Proc. Sys., 21: 1529-1536.

Tenenbaum, J.B., V. de Silva and J. Langford, 2000. A global geometric framework for nonlinear dimensionality reduction. Science, 290: 2319-2323. DOI: 10.1126/science.290.5500.2319

DOI: 10.3844/jmssp.2016.157.175

Verleysen, M., 2003. Learning high-dimensional data. Limitations Future Trends Neural Computation, 186: 141-162.

Weinberger, K.Q. and L.K. Saul, 2006. Unsupervised learning of image manifolds by semidefinite programming. Int. J. Comput. Vision, 70: 77-90.

Zhang, Z. and H. Zha, 2004. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM J. Sci. Comput., 26: 313-338.