Proceedings of Machine Learning Research 60:1–21, 2017 Conformal and Probabilistic Prediction and Applications

Asymptotic Properties of Nonparametric Estimation on Manifold

Yury Yanovich yanovich.yury@ya.ru
Institute for Information Transmission Problems (Kharkevich Institute)
Bolshoy Karetny per. 19, build.1, Moscow 127051, Russia
Faculty of Computer Science, Higher School of Economics (National Research University)
Kochnovsky Proezd 3, Moscow 125319, Russia

Editors: Alex Gammerman, Vladimir Vovk, Zhiyuan Luo, and Harris Papadopoulos

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-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 via nonparametric estimation, and 'global part' in which global low-dimensional coordinates on the DM are constructed by solving the certain convex optimization problem for specific cost function depending on the local characteristics. Both statistical properties of 'local part' and its average over manifold are considered in the paper. The article is an extension of the paper (Yanovich, 2016) for the case of nonparametric estimation.

Keywords: Manifold Learning, Nonparametric Estimation, Asymptotic Expansions, Large Deviations

1. 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 p-dimensional 'observation space' 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) $\mathbb{M}$ of lower dimensionality q 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; massoud Farahmand et al., 2007; Campadelli et al., 2015). Also in (Smith et al., 2009) distance to the k-th neighbor assumed to converge to zero and is 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. In (Singer and Wu, 2012), it was shown that both local and statistical parts of errors are asymptotically small for a specific statistic (elements of the covariance matrix) and also large deviation error was estimated.

Local and global nonparametric statistics with a special form on a manifold are considered in the present paper. The consistency, asymptotic expansion, and large probability are of interest. Such results could be used to study the properties of existing manifold learning algorithms and to propose new statistically motivated ones. The results generalizes the classic results for the nonparametric regression with multivariate input for the case of nonlinear input space. The optimal rate of convergence of the mean absolute deviation for the classic (Henderson and Parmeter, 2015) and considered cases coincides and are equal $O(N^{-\frac{2}{q+4}})$, where N is a sample size.

The paper is organized as follows. In Section **2** the data model is described, and the main results of the paper are listed and discussed. In Section **3** the data model is defined, and all assumptions are listed. Then, Section **4** contains exact formulations of the main results. Section **5** contains the main proofs. In Section **6** the paper summary and future work directions are given. In **Appendix A** supplementary Lemmas proofs are given.

2. Results Description

The main results are strictly formulated in **Section 4**, and the main assumptions, used in the proof, strictly formulated in **Section 3**. Some substantive comments are given here.

2.1. Manifold Learning Data Model

Let $\mathbb{M} \subset \mathbb{R}^p$ be an unknown manifold with dimensionality q and $\mathbb{X}_N = \{X_1, \ldots, X_N\} \subset \mathbb{M}$ be a random sample with size N. Data Model consists of assumptions about support (manifold $\mathbb{M}$) and assumptions about sample distribution. 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 $\sim \exp(-C \cdot N)$ are known, where $C > 0$ is a 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 μ on the manifold $\mathbb{M}$. The strict formulation of the mentioned assumptions are formulated in (Yanovich, 2016) and in Section 3.

2.2. Statistics Form

Consider local statistics

$$F_N(X) = \frac{\sum_{n=1}^{N} K_\varepsilon(X, X_n) \cdot F(X, X_n)}{\varepsilon^d \cdot \sum_{n=1}^{N} K_\varepsilon(X, X_n)}, \tag{1}$$

where $F(X, X_n)$ is a specific function for each problem (Manifold Learning algorithm), $K_\varepsilon(X, X_n)$ is weight (kernel) function for $F(X, X_n)$ such that $K_\varepsilon(X, X_n) > 0$ iff $X_n \in U_{X,\varepsilon}$, where $U_{X,\varepsilon} = \{X' \in \mathbb{R}^p | \ |X' - X| < \varepsilon\}$; $d \in \mathbb{Z}, d \geq 0$ is a parameter, which is defined by function $F(X, X_n)$ behavior for $X \to X_n$. Parameter $\varepsilon = \varepsilon(N)$ is considered as a function of sample size N. The exact assumptions about ε are listed in Section Data Model of (Yanovich, 2016). An example of the kernel which meets assumptions is $\varepsilon = E(N) = C \cdot N^{-\frac{1}{q+2}}$.

The considered global statistics are average values of (1) over the sample

$$F_N = \frac{1}{N} \sum_{n=1}^{N} F_N(X_n). \tag{2}$$

Kernel function examples $K_\varepsilon(X, X')$ from different algorithms:

- indicator function $I(X' \in U_{X,\varepsilon})$ for event 'X' is in $U_{X,\varepsilon}$' is used in Local Linear Embeding and Local Tangent Space Alignment (Roweis and Saul, 2000; Zhang and Zha, 2004);
- heat-kernel from Laplacian Eigenmaps (Belkin and Niyogi, 2003) $\exp(-|X' - X|^2/T) \cdot I(X' \in U_{X,\varepsilon})$, where $T > 0$ is a 'temperature' parameter;
- Epanechnikov kernel from Vector Diffusion Maps (Singer and Wu, 2012) $\left(1 - |X' - X|^2/\varepsilon^2\right) \cdot I(X' \in U_{X,\varepsilon})$;
- kernel from Grassmann&Stiefel Eigenmaps(Bernstein and Kuleshov, 2013, 2014; Bernstein et al., 2015a,b,c) $K_{GSE}(X, X')$, which depends on not only from the distance $|X - X'|$, but also on CauchyBinet between tangent spaces $T_X(\mathbb{M})$ and $T_{X'}(\mathbb{M})$ for $\mathbb{M}$ at points X and X'.

For a fixed point X, let's represent $X' \in U_{X,\varepsilon}$ using its locally Riemannian coordinates: $X' = \exp_X(t\theta)$, where $t \in [0, \infty)$ and $\theta \in S_{q-1} \subset T_X(\mathbb{M})$. The parametric family of smooth kernels with finite support $\{K(X, \theta, t), t \geq 0, \theta \in S_{q-1}\}$, and $K_\varepsilon(X, X')$ are represented as

$$K_\varepsilon(X, X') = K_\varepsilon(X, \exp_X t\theta) = K(X, \theta, t/\varepsilon). \tag{3}$$

The kernels are assumed to be smooth functions of their arguments in the article as in other theoretic papers about Manifold Learning. The kernels from examples above could be represented as (3) but are not continuous or smooth. Fortunately, it is possible to replace these kernels with 'smoothed' counterparts, without any change in algorithms. And it usually doesn't change the calculation procedure quality (Bishop, 2006). An example of the kernel which meets the papers assumptions is $K(X, \theta, t) = \exp\left(\frac{1}{z^2-1}\right) \cdot I\left(t \in (-1, 1)\right)$.

Examples of $F(X, X')$ (1)

1. $F(X, X') = \psi(X')$ leads to $F_N(X)$ (1) which is nonparametric estimate (Wasserman, 2006) for unknown function on manifold $\psi(\dot{)}$ at point X using its values $\psi(X')$ at sample points $X' \in \mathbb{X}_N$;
2. $F(X, X') = (X' - X) \cdot (X' - X)^T$, $X, X' \in \mathbb{R}^p$, $p \times p$ matrix $F_N(X)$ (1) is an estimate for covariance matrix at point $X \in \mathbb{M}$ using points from $U_{X,\varepsilon}$. This statistic is used in the local principal component analysis to estimate tangent space $T_X(\mathbb{M})$ to $\mathbb{M}$ at point X (Singer and Wu, 2012);

3. $F(X, X') = \left|\psi(X) - \psi(\tilde{X})\right|^2$, where ψ is a scalar or vector function. Local statistic $F_N(X)$ (1) is used in such dimension reduction procedures as (Belkin and Niyogi, 2003; Bernstein and Kuleshov, 2013).

The parameter d is defined as $F(X, X') = \Phi(X, \theta, t) = O(t^d)$ for $t \to 0$:

$$\Phi(X, \theta, t) = t^d \cdot \phi(X, \theta, t).$$

It is assumed that d is fixed and is the same for all $X \in \mathbb{M}$ and $\theta \in T_X(\mathbb{M})$. The strict assumptions about kernel family and the function $F(X, X')$ are also listed in Section 3. Denote for $X' = \exp_X(t\theta)$

$$F(X, X') = \tilde{F}(X, \theta, t) = t^d \cdot \varphi(X, \theta, 0) + t^{d+1} \cdot \varphi_1(X, \theta, 0) + \frac{t^{d+2}}{2} \cdot \varphi_2(X, \theta, \tilde{t}), \tag{4}$$

for $t \to 0$ and $\tilde{t} = \tilde{t}(t) \in [0, t]$.

In the article it is shown that statistics (1) and (2) are the consistent estimates for

$$\overline{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,d}(X, \theta) \tilde{F}(X, \theta, 0) d\theta}{\int_{S^{q-1}} \rho_{E,0}(X, \theta) d\theta}, \tag{5}$$

and

$$\overline{F} = \int_{\mathbb{M}} \overline{F}(X) d\mu(X), \tag{6}$$

where S^{q-1} is sphere in q-dimensional space,

$$\rho_{E,m}(X, \theta) = \int_0^1 K(X, \theta, t)\, t^{m+q-1} dt. \tag{7}$$

2.3. 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, local statistics should behave almost as nonparametric estimates in linear tangent q-dimensional space. Similar results where proved in (Singer and Wu, 2012) for a particular statistic (tangent space estimation). However, in Statement 1, it is proved that for the class of local statistics (1) are consistent estimates for the explicitly written values. The Theorem 1 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 to the conditional distribution. So, one could think of conditional distribution as of uniform distribution on the ball in tangent space. In the Theorem 2 asymptotic expansion of the considered statistics is given, and in the Theorem 3 assesses the probability of large deviations. The Theorem 4 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, the need to obtain uniform estimation on the manifold. The basic ideas used in proofs: local linearization of the support, the use of inequalities for the probabilities of large deviations of sums of i.i.d. random variables, the use of finite nets.

3. Data Model

General common assumptions M1-M8 and S1-S3 for Manifold Learning are formulated in **Section Data Model** of (Yanovich, 2016). Here are listed only additional ones.

M9. Exists $\varepsilon_B > 0$ and $C_B > 0$ such that for $\varepsilon < \varepsilon_B$ Riemannian measure $dV(\mathbb{M} \setminus \mathbb{M}_\varepsilon)$, of distant from boundary points ε is bounded

$$dV(\mathbb{M} \setminus \mathbb{M}_\varepsilon) \le \varepsilon \cdot C_B \cdot \varepsilon \cdot dV(\mathbb{M});$$

M10. Exists $\varepsilon_C > 0$ and positive $C_c > 0$ such that $\varepsilon < \varepsilon_c$ and for each point $X \in \mathbb{M}$ Riemannian measure of intersection of p-dimensional ε ball with center at $X \in \mathbb{M}$ and Manifold $\mathbb{M}$ is lower bounded by $C_c \cdot \varepsilon^q$.

For the neighborhood parameter $\varepsilon = \varepsilon(N) > 0$ it is assumed that

P1. For $N \to \infty$: $\varepsilon \to 0$;

P2. For $N \to \infty$: $N \cdot \varepsilon^q \to \infty$;

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

Note. 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 $N \cdot \varepsilon^q$.

For the function $F(X, X')$ for $X' = \exp_X t\theta \in U_{X,\varepsilon}$ it is assumed

F1. Exists $d \in \mathbb{Z}, d \ge 0$ which is the same for all $X \in \mathbb{M}$ such that the d is the main term of the $\tilde{F}(X, \theta, t)$ expansion for $t \to 0$;

F2. The representation (4) is valid for $t \to 0$ and $\tilde{t} \in [0, t]$, where $\varphi(X, \theta, t)$ and $\varphi_1(X, \theta, t)$ are smooth and bounded, $\varphi_2(X, \theta, t)$ is bounded and Lipshitz.

Let

$$C_\varphi = \sup_{X \in \mathbb{M}, t, \theta} |\varphi(X, \theta, t)|; \tag{8}$$

$$C_{\varphi,1} = \sup_{X \in \mathbb{M}, t, \theta} |\varphi_1(X, \theta, t)|; \tag{9}$$

$$C_{\varphi,2} = \sup_{X\in\mathbb{M},t,\theta} |\varphi_2(X,\theta,t)|. \tag{10}$$

For the weight function $K_\varepsilon(X,X')$ for $X' = \exp_X t\theta \in U_{X,\varepsilon}$ it is assumed

K1. $K_\varepsilon(X,X') = K(X,\theta,\frac{|X-X'|}{\varepsilon})$, where $K(X,\theta,z)$ is a fixed nonnegative bounded function with bounded derivative, which is symmetric with respect to θ and z;

K2. Function $K(X,\theta,z)$ has finite support with respect to z: $K(X,\theta,z) = 0$ for $z \geq 1$;

K3. Positive $\rho_{E,0}(X,\theta)$ (7) and $\rho_{V,0}(X,\theta)$ (16) exist;

K4. Function $K(X,\theta,z)$ is Lipshitz with respect to X with constant $\tilde{C}_K$;

K5. $\inf_{X\in\mathbb{M}} \int_{S_q}\int_{1/2}^1 K(X,\theta,z)z^{q-1}d\theta dt \geq C_{K,1/2}$.

Let

$$C_K = \sup_{X,\theta,z} K(X,\theta,z); \tag{11}$$

$$C_{K,1} = \sup_{X,\theta,z} \left|\frac{\partial K(X,\theta,z)}{\partial z}\right|. \tag{12}$$

4. Main Results

The results are formulated under assumptions M1-M10, S1-S3, P1-P3, F1-F2 and K1-K5.

Statement 1 ($F_N(X)$ consistency). For each $X \in \mathbb{M}$ as $N \to \infty$

$$\begin{aligned}\mathbb{E}F_N(X) &\to \overline{F}(X);\\ N\varepsilon^q \cdot \mathtt{Var}F_N(X) &\to d(X);\\ F_N(X) &\to^p \overline{F}(X),\end{aligned}$$

where $m(X)$ and $d(X)$ are functions (19) and (18), $\to^p$ is convergence in probability.

Theorem 1 ($F_N(X)$ asymptotic expansion). For each $X \in \mathbb{M}$ as $N \to \infty$

$$\sqrt{N\varepsilon^q} \cdot \big(F_N(X) - \overline{F}(X)\big) \to^D N(m(X), d(X)),$$

where $m(X)$ and $d(X)$ are functions (19) and (18), $m(X) \equiv 0$ for $d \vdots 2$, $\to^D$ is convergence in distribution.

Theorem 2 (large deviation probability for $F_N(X)$). For each ε-bounded from manifold boundary point $X \in \mathbb{M}_\varepsilon$, and for each $N > N_0$ and $\varepsilon < \varepsilon_0$

$$P\left(\left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot I(d \not\vdots 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2}\right) \leq 4 \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{\sigma^2(X)}\right),$$

where $N_0, \varepsilon_0, C_{LD1}, C_{LD2}$ are positive constants (20), (22), (23), (24),

$$\sigma^2(X) = p_\mu(X) \cdot \int_{S_{q-1}} \rho_{D,2d}(X,\theta)(1+\varphi(X,\theta,0))^2 d\theta, \tag{13}$$

and $\sigma^2(X) \leq C_\sigma$, C_σ is a constant (21).

Theorem 3 (uniform large deviation probability for $F_N(X)$). For each $z \in [0,1]$, for $N > N_0$, $\varepsilon < \varepsilon_0$ for each ε-bounded from manifold boundary point $X \in \mathbb{M}_\varepsilon$

$$P\left(\sup_{X \in \mathbb{M}_\varepsilon} \left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot I(d \not\vdots 2) \cdot C_{LD1} + \varepsilon^2 \cdot \tilde{C}_{LD2}\right) \leq$$
$$\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3}\right)^p \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma}\right) + \left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp\left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2/(9p_{\max})\right),$$

where $N_0, \varepsilon_0, C_{LD1}$ are positive constants (20), (22), (23), $\tilde{C}_{LD2}$ is a positive constant form the proof.

Statement 2 (F_N consistency). For $N \to \infty$:

$$F_N \to^p \overline{F}.$$

Theorem 4 (large deviation probability for F_N). Exist positive constants $N_{U,0}$, $C_{U,1}$, $C_{U,2}$, $C_{U,p}$ such that for $z \in [0,1]$ and $N > N_{U,0}$:

$$P\left(\left|F_N - \overline{F}\right| \geq z + \varepsilon \cdot I(d \not\vdots 2) \cdot C_{U,1} + \varepsilon^2 \cdot C_{U,2}\right) \leq \exp\left(-z^2 \cdot N\varepsilon^q \cdot C_{U,p}\right).$$

5. Proof of Main Theorems

The following Lemma would be used in proofs.

Lemma 1 (random variable moments). Let $X \in \mathbb{M}_{2\varepsilon}$, $\varepsilon \leq C_{int}$ a point $X' \in \mathbb{M}$ is random with density $p_\mu(X')$. Then for $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$

$$\mathbb{E}\xi = p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta + \varepsilon \cdot E_1(X) \cdot I(d \not\vdots 2) + \varepsilon^2 \cdot \alpha_E(X) \cdot C_{E,2}; \quad (14)$$

$$\mathtt{Var}\xi = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta + \varepsilon^{\min\{-q+2,0\}} \cdot \alpha_V(X) \cdot C_{V,2}, \quad (15)$$

where

$$|\alpha_E(X)| \leq 1;$$
$$|\alpha_V(X)| \leq 1;$$
$$E_1(X) \leq C_{E,1},$$

$E_1(X)$ is defined in (27), $C_{E,1}, C_{E,2}, C_{V,2}$ are constants from (28), (29), (30) respectivly, C_{int} is a constant from Lemma 8 from (Yanovich, 2016), $I(d \not\vdots 2)$ is an indicator for even d,

$$\rho_{V,m}(X,\theta) = \int_0^1 K(X,\theta,t)^2 t^{m+q-1} dt. \quad (16)$$

Proof. The Lemma is proved in **Appendix A**.

Consequence 1. Let $X \in \mathbb{M}_{2\varepsilon}$,
$\varepsilon \le C_{int}$ and $X' \in \mathbb{M}$ is random with density $p_\mu(X')$. Then for

$$\xi_{K1,F0} = \frac{1}{\varepsilon^q} \cdot K_\varepsilon(X, X');$$
$$\xi_{K1,F2} = \frac{1}{\varepsilon^{q+2d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')^2;$$
$$\xi_{K2,F1} = \frac{1}{\varepsilon^{2q+d}} \cdot K_\varepsilon(X, X')^2 \cdot F(X, X');$$
$$\xi_{K2,F2} = \frac{1}{\varepsilon^{2q+2d}} \cdot K_\varepsilon(X, X')^2 \cdot F(X, X')^2$$

the moments are

$$|\mathbb{E}\xi_{K1,F0} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta) d\theta| \le \varepsilon^2 \cdot \tilde{C}_{E,2};$$
$$|\mathbb{E}\xi_{K1,F2} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)^2 d\theta| \le \varepsilon^2 \cdot \tilde{C}_{E,2};$$
$$|\mathbb{E}\xi_{K2,F1} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{V,d}(X,\theta)\varphi(X,\theta,0) d\theta| \le \varepsilon \cdot I(d \not\vdots 2) \cdot \tilde{C}_{E,1} + \varepsilon^2 \cdot \tilde{C}_{E,2};$$
$$|\mathbb{E}\xi_{K2,F2} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{V,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta| \le \varepsilon^2 \cdot \tilde{C}_{E,2},$$

where

$$\tilde{C}_{E,1} = V_q \cdot (1 + C_\rho)^2 \cdot \left(p_{max} \cdot (1 + C_{\varphi,1})^2 + C_{p,1} \cdot (1 + C_\varphi)^2\right);$$
$$\tilde{C}_{E,2} = V_q \cdot (1 + C_K + C_{K,1})^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2$$
$$\cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + 8 \cdot V_q \cdot p_{max} \cdot (1 + C_K)^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2,$$

C_{int} is a constant from Lemma 8 from (Yanovich, 2016).

Proof of Statement 1. Using Lemma 5 for $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$ and $\tilde{\xi} = \frac{1}{\varepsilon^q} \cdot K_\varepsilon(X, X') \cdot 1$, using $\varepsilon^{-q} \cdot N^{-1} \to 0$ from P2, from law of large numbers:

$$\frac{1}{N\varepsilon^{q+d}} \cdot \sum_{n=1}^{N} K_\varepsilon(X, X_n) \cdot F(X, X_n) \to^P \int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0) d\theta;$$
$$\frac{1}{N\varepsilon^{q}} \cdot \sum_{n=1}^{N} K_\varepsilon(X, X_n) \cdot F(X, X_n) \to^P \int_{S^{q-1}} \rho_{E,d}(X,\theta) d\theta > 0.$$

From which follows the convergence of the ratio to the ratio of parts limits.

Proof of Theorem 1. Transform

$$F_N(X) - \frac{\mathbb{E}\frac{1}{\varepsilon^{q+d}} K_\varepsilon(X, X_1) \cdot F(X, X_1)}{\mathbb{E}\frac{1}{\varepsilon^q} K_\varepsilon(X, X_1)} = \Big(\frac{1}{N}\sum_{n=1}^{N} \frac{1}{\varepsilon^{2q+d}} \Big(K_\varepsilon(X, X_i) F(X, X_i) \mathbb{E} K_\varepsilon(X, X_n)$$
$$- K_\varepsilon(X, X_n) \left(\mathbb{E} K_\varepsilon(X, X_n) F(X, X_n)\right)\Big)\Big) \Big/ \Big(\mathbb{E}\frac{1}{\varepsilon^q} K_\varepsilon(X, X_1) \cdot \frac{1}{N\varepsilon^q} \sum_{n=1}^{N} K_\varepsilon(X, X_n)\Big). \quad (17)$$

Denominator (17) converges to $\left(p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta\right)^2 > 0$ in probability. Numerator (17) equals to the sum of centered independent (as functions of independent) identically distributed random variables

$$\varsigma_n = \left(\frac{1}{\varepsilon^{q+d}} K_\varepsilon(X,X_i)F(X,X_i)\right) \cdot \left(\frac{1}{\varepsilon^q}\mathbb{E}K_\varepsilon(X,X_n)\right)$$
$$- \left(\frac{1}{\varepsilon^{q+d}}\mathbb{E}K_\varepsilon(X,X_i)F(X,X_i)\right) \cdot \left(\frac{1}{\varepsilon^q} K_\varepsilon(X,X_n)\right),\ n = 1,\dots,N$$

with the second moment from Consequence 1 as $N \to \infty$:

$$\mathbb{E}\varsigma^2 \to d(X),$$

where

$$\tilde{d}(X) = \frac{1}{\varepsilon^q} \cdot p_\mu(X)^3 \cdot \Big(\int_{S^{q-1}} \rho_{V,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta \cdot \big(\int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta\big)^2 +$$
$$+ \int_{S^{q-1}} \rho_{V,2d}(X,\theta)d\theta \cdot \big(\int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta\big)^2 -$$
$$-2 \cdot \int_{S^{q-1}} \rho_{V,d}(X,\theta)\varphi(X,\theta,0)d\theta \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta\Big).$$

Also

$$\mathbb{E}\varsigma_n^2 \le 4 \cdot (1 + C_K)^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2}),$$

where $C_\varphi, C_{\varphi,1}, C_{\varphi,2}, C_K$ are constants (8), (9), (10), (11) respectivly. Which implies that

$$N \cdot \varepsilon^q \cdot \mathtt{Var}\left(F_N(X) - \frac{\mathbb{E}\frac{1}{\varepsilon^{q+d}} K_\varepsilon(X,X_1) \cdot F(X,X_1)}{\mathbb{E}\frac{1}{\varepsilon^q} K_\varepsilon(X,X_1)}\right) \to d(X),$$

where

$$d(X) = N \cdot \varepsilon^q \cdot \tilde{d}(X) / \left(p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta\right)^2$$
$$= p_\mu(X) \cdot \Big(\int_{S^{q-1}} \rho_{V,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta \cdot \big(\int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta\big)^2$$
$$+ \int_{S^{q-1}} \rho_{V,2d}(X,\theta)d\theta \cdot \big(\int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta\big)^2$$
$$-2 \cdot \int_{S^{q-1}} \rho_{V,d}(X,\theta)\varphi(X,\theta,0)d\theta \cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta$$
$$\cdot \int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta\Big) / \left(\int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta\right)^2. \tag{18}$$

As $N \to \infty$:

$$\frac{\mathbb{E}\frac{1}{\varepsilon^{q+d}} K_\varepsilon(X,X_1) \cdot F(X,X_1)}{\mathbb{E}\frac{1}{\varepsilon^q} K_\varepsilon(X,X_1)} - \frac{\int_{S^{q-1}} \rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta}{\int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta}$$
$$= I(d \not\vdots 2) \cdot \frac{E_1(X)}{p_\mu(X)} \cdot \frac{1}{\int_{S^{q-1}} \rho_{E,d}(X,\theta)d\theta} \cdot \varepsilon + O(\varepsilon^2).$$

Denote

$$m(X) = I(d \not\vdots 2) \cdot \frac{E_1(X)}{p_\mu(X)} \cdot \frac{1}{\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta}. \tag{19}$$

Using P3: $N\varepsilon^4 \to 0$ as $N \to \infty$. So from the Central Limit Theorem

$$\sqrt{N\varepsilon^q} \cdot \big(F_N(X) - \overline{F}(X)\big) \to^d N(m(X), d(X)).$$

Moreover, $m(X) \equiv 0$ for $d \vdots 2$.

Proof of Theorem 2. *1.* Denote

$$N_0 = \frac{\max\{1, (C_\varphi + C_{\varphi,1} + C_{\varphi,2})\}}{p_{max} \cdot V_q \cdot \min\{1, C_\varphi^2\} \cdot C_K^{2d-1}}, \tag{20}$$

$$C_\sigma = p_{\max} \cdot C_K^2 \cdot (1 + C_\varphi)^2 \cdot V_q. \tag{21}$$

2. Consider random variables $\chi_n = \xi_n - \mathbb{E}\xi_n$, $\xi_n = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X_n) \cdot F(X, X_n)$, $n = 1, \dots, N$, $\chi = \xi_n - \mathbb{E}\xi_n$, $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$:

$$\texttt{Var}\chi \le 2\varepsilon^{-q} \cdot \sigma^2(X);$$
$$\texttt{Var}\chi \ge 1/2 \cdot \sigma^2(X),$$

from Lemma 1 for

$$\varepsilon \le \varepsilon_0 \equiv \min\left\{\frac{\sigma^2(X)}{2 \cdot C_{V,2}}, 1\right\} \tag{22}$$

Also, when $\varepsilon < 1$ for $m \ge 2$

$$\mathbb{E}|\chi|^m \le \texttt{Var}\chi \cdot \sup_\omega |\chi(\omega)|^{m-2})\texttt{Var}\chi \cdot \varepsilon^{-q(m-2)} \cdot (C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}))^{m-2} \le$$
$$\le \frac{m!}{2} \cdot \texttt{Var}\chi \cdot H^{m-2},$$

where $H = \varepsilon^{-q} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})$. Hence, by Theorem 12 from (Petrov, 1987) for $z \in \left[0, N \cdot \frac{1}{2} \cdot \frac{\sigma^2(X)}{C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})}\right]$:

$$P\left(\left|\frac{\sum_{n=1}^N \chi_n}{N}\right| > z\right) \le 2 \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{2\sigma^2(X)}\right).$$

3. From (13)

$$\sigma^2(X) \le p_{max} \cdot V_q \cdot C_\varphi^2 \cdot C_K^{2d}.$$

So for $N \ge \frac{(C_\varphi + C_{\varphi,1} + C_{\varphi,2})}{p_{max} \cdot V_q \cdot C_\varphi^2 \cdot C_K^{2d-1}}$ item *2* is valid for $z \in [0, 1]$.

4. From Lemma 1

$$\left|\xi_n - p_\mu(X)\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)\varphi(X,\theta,0)d\theta\right| \le \varepsilon\cdot C_{E,1}\cdot I(d \not\vdots 2) + \varepsilon^2\cdot C_{E,2}.$$

5. From items *2-4* for $N \ge \frac{(C_\varphi + C_{\varphi,1} + C_{\varphi,2})}{p_{max}\cdot V_q\cdot C_\varphi^2\cdot C_K^{2d-1}}$ $\varepsilon < \min\{1, C_{int}\}$

$$P\left(\left|\frac{\sum_{n=1}^N \chi_n}{N}\right| > z + \varepsilon\cdot C_{E,1}\cdot I(d \not\vdots 2) + \varepsilon^2\cdot C_{E,2}\right) \le 2\cdot\exp\left(-\frac{z^2\cdot N\varepsilon^q}{2\cdot\sigma^2(X)}\right).$$

6. Repeating items *2-5* for $K_\varepsilon(X,X_n)\cdot 1$ instead of $K_\varepsilon(X,X_n)\cdot F(X,X_n)$ and using Consequence 1 for $N \ge \frac{1}{p_{max}\cdot V_q\cdot C_K^{2d-1}}$ $\varepsilon < \min\{1, C_{int}\}$,

$$P\left(\left|\frac{\sum_{n=1}^N K_\varepsilon(X,X_n)}{N\varepsilon^q} - p_\mu(X)\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta\right| > z + \varepsilon^2\cdot\tilde{C}_{E,2}\right)$$
$$\le 2\cdot\exp\left(-\frac{z^2\cdot N\varepsilon^q}{2\cdot\sigma^2(X)}\right).$$

7. For $\varepsilon \le \sqrt{\frac{1}{2}\cdot\frac{p_{\min}\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta}{\tilde{C}_{E,2}}}$

$$|\mathbb{E}\frac{1}{\varepsilon^q}K_\varepsilon(X,X') - p_\mu(X)\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta| \le \frac{1}{2}\cdot p_{\min}\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta.$$

So

$$|\overline{F}(X)| \le C_\varphi;$$
$$\left|\overline{F}(X) - \frac{\mathbb{E}\frac{1}{\varepsilon^{d+q}}K_\varepsilon(X,X')F(X,X')}{\mathbb{E}\frac{1}{\varepsilon^q}K_\varepsilon(X,X')}\right|$$
$$\le \overline{F}(X)\cdot\frac{2\varepsilon^2\cdot\tilde{C}_{E,2}}{p_\mu(X)\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta} + \frac{2\varepsilon\cdot C_{E,1}\cdot I(d \not\vdots 2) + \varepsilon^2\cdot\tilde{C}_{E,2}}{p_\mu(X)\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta}$$
$$\le \frac{2}{p_{\min}\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta}\cdot\left(\varepsilon\cdot C_{E,1}\cdot I(d \not\vdots 2) + \varepsilon^2\cdot\tilde{C}_{E,2}\cdot(1 + C_\varphi)\right).$$

8. From *5-7*

$$P\left(\left|F_N(X) - \overline{F}(X)\right| \ge z + \varepsilon\cdot I(d \not\vdots 2)\cdot C_{LD1} + \varepsilon^2\cdot C_{LD2}\right) \le 4\cdot\exp\left(-\frac{z^2\cdot N\varepsilon^q}{\sigma^2(X)}\right),$$

where

$$C_{LD1} = \frac{2}{p_{\min}\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta}\cdot C_{E,1}, \tag{23}$$

$$C_{LD2} = \frac{2}{p_{\min}\cdot\int_{S^{q-1}}\rho_{E,d}(X,\theta)d\theta}\cdot\tilde{C}_{E,2}\cdot(1 + C_\varphi). \tag{24}$$

Proof of Theorem 3. *1.* From Lemma 10 from (Yanovich, 2016) and ε-nets properties for $\delta = \varepsilon^3$ exists finite δ-net $\mathbb{N}_\delta$ for $\mathbb{M}_\varepsilon$ with no more than $\left(\frac{2a\sqrt{p}}{\varepsilon^3}\right)^p$ elements (smallness of network ε^3 choosed so that the order of the resulting deviation is equal to the local one from Theorem 2).

2. For $X' = \exp_X(t'\theta')$, where $t' \le \varepsilon^3 \ \ \theta' \in S_{q-1}$ from Taylor formula for $\varepsilon \le 1$:

$$|\overline{F}(X') - \overline{F}(X)| \le \varepsilon^3 \cdot \Bigg(\sup_{X,t',\theta} \left| \frac{\partial \varphi(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| + \frac{1}{\varepsilon} \cdot V_q \cdot C_\varphi \cdot \sup_{X,t',\theta} \left| \frac{\partial K(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| + \frac{1}{\varepsilon} \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot \sup_{X,t',\theta} \left| \frac{\partial K(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| \Bigg) \Big/ \int_{S_{q-1}} \rho_{E,0}(X,\theta) d\theta$$
$$\le \varepsilon^2 \cdot \left(C_{u,\varphi} + V_q C_\varphi \cdot \tilde{C}_K + V_q \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot \tilde{C}_{K,1} \right) / c_\rho.$$

3. Let us upper bound $\frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X', X_n) F(X', X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X', X_n)}$ and $\frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X, X_n) F(X, X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n)}$. From (Theorem 5, (Yanovich, 2016)) for $z = 1/8$ and K4 with probability $1 - \left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp\left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2 / (9 p_{\max})\right)$

$$\inf_{X' \in \mathbb{M}_\varepsilon} \frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n) \ge C_{K,\frac{1}{2}} \cdot \left(V_q \cdot p_{\min} \cdot 7/8 - \varepsilon^2 \cdot \frac{C_E \cdot p_{\max}}{V_q p_{\min}} \right).$$

For $\varepsilon < \sqrt{\frac{3}{4} \cdot V_q \cdot p_{\min} \cdot \frac{1}{\sqrt{C_E \cdot p_{\max}}}}$:

$$\inf_{X' \in \mathbb{M}_\varepsilon} \frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n) \ge C_{K,\frac{1}{2}} \cdot V_q \cdot p_{\min} \cdot 1/2.$$

Whence

$$\left| \frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X', X_n) F(X', X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X', X_n)} - \frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X, X_n) F(X, X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n)} \right| \le 2 \cdot \varepsilon^2 \cdot \frac{\tilde{C}_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) + 2C_K \cdot (\tilde{C}_\varphi + \tilde{C}_{\varphi,1} + \tilde{C}_{\varphi,2})}{C_{K,\frac{1}{2}} \cdot V_q \cdot p_{\min}}.$$

Denote

$$C_{LD,net} = 2 \cdot \frac{\tilde{C}_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) + 2C_K \cdot (\tilde{C}_\varphi + \tilde{C}_{\varphi,1} + \tilde{C}_{\varphi,2})}{C_{K,\frac{1}{2}} \cdot V_q \cdot p_{\min}}. \tag{25}$$

4. From Theorem 2 using Lemma 1 for each element of the finite net $\mathbb{N}_\delta$ for $z \in [0, 1]$

$$P\left(\left| F_N(X) - \overline{F}(X) \right| \ge z + \varepsilon \cdot I(d \not{\vdots} 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2} \right) \le 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3} \right)^p \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma} \right).$$

5. From *3,4* as for each $X \in \mathbb{M}_\varepsilon$ exists $X' \in \mathbb{N}_\delta$

$$P\left(\sup_{X \in \mathbb{M}_\varepsilon} \left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot I(d \not{:} 2) \cdot C_{LD1} + \varepsilon^2 \cdot (C_{LD2} + C_{LD,net})\right)$$
$$\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3}\right)^p \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma}\right) + \left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp\left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2/(9p_{\max})\right).$$

Redenote $\tilde{C}_{LD2} = C_{LD2} + C_{LD,net}$ and get Theorem statement.

Proof of Statement 2. Let $N_{2\varepsilon}$ be the number of sample points which are 2ε far from the boundary. Then From Theorem 3 and Lemma 4

$$F_N = \frac{N_\varepsilon}{N} \sum_{X' \in \mathbb{M} \setminus \mathbb{M}_\varepsilon} F_N(X') + \frac{N - N_\varepsilon}{N} \sum_{X' \in \mathbb{M}_\varepsilon} F_N(X') \to^p 0 + \overline{F}.$$

Proof of Theorem 4. Using the F_N expansion from Statement 2, Theorem 3, Lemmas 4 and 5 obtain, that there are positive $N_{U,0}$, $C_{U,1}$, $C_{U,2}$, $C_{U,p}$, for all $z \in [0,1]$ $N > N_{U,0}$:

$$P\left(\left|F_N - \overline{F}\right| \geq z + \varepsilon \cdot I(d \not{:} 2) \cdot C_{U,1} + \varepsilon^2 \cdot C_{U,2}\right) \leq \exp\left(-z^2 \cdot N\varepsilon^q \cdot C_{U,p}\right).$$

6. Conclusion

The specific for Manifold Learning nonparametric estimation on the manifold and its average are considered in the present paper. Points are assumed to lie on a good enough unknown manifold, the kernels with finite support are used, the radius of support slowly tends to zero with sample size growth. Asymptotic expansion and uniform large deviation results are obtained for the considered nonparametric estimates. And also consistency and large deviations of average is also considered. 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, could be and already are used to get properties of Manifold Learning optimization procedures (Yanovich, 2017).

Acknowledgments

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

Appendix A. Lemmas Proofs

A.1 Moments of Random Variables

Lemma 2. For $d\,\vdots\,2$ function $\varphi(X,\theta,0)$ is even function of θ and function $\varphi_1(X,\theta,0)$ is odd function of θ. For $d \not\vdots\, 2$ function $\varphi(X,\theta,0)$ is odd function of θ and function $\varphi_1(X,\theta,0)$ is even function of θ.

Proof. As

$$\varphi(X,\theta,0) = d! \cdot (\nabla_\theta)^d F(X, \exp_X(t\theta)|_{t=0};$$
$$\varphi_1(X,\theta,0) = (d+1)! \cdot (\nabla_\theta)^{d+1} F(X, \exp_X(t\theta)|_{t=0}.$$

The orders of θ proves Lemma.

Lemma 3. From K3 follows existance of $\rho_{E,m}(X,\theta)$ (7) and $\rho_{V,m}(X,\theta)$ ((16) for $m \in \mathbb{Z}, d \geq 0$.

Proof. Indeed

$$0 \leq \rho_{E,m}(X,\theta) = \int_0^1 K\left(X,\theta,t\right) t^{m+q-1} dt \leq \int_0^1 K\left(X,\theta,t\right) 1^m \cdot t^{q-1} dt = \rho_{E,0}(X,\theta);$$
$$0 \leq \rho_{V,m}(X,\theta) = \int_0^1 K\left(X,\theta,t\right)^2 t^{m+q-1} dt \leq \int_0^1 K\left(X,\theta,t\right)^2 1^m \cdot t^{q-1} dt = \rho_{V,0}(X,\theta).$$

Proof of Lemma 1. *0.* From Lemma 3 $\rho_{E,m}(X,\theta)$ and $\rho_{E,m}(X,\theta)$ exist.

1. From (Lemma 8, (Yanovich, 2016)) for $\xi \cdot \varepsilon^q$ as $\varepsilon \leq C_{int}$:

$$|\mathbb{E}\xi \cdot \varepsilon^q - \int_{\tilde{B}_\varepsilon(X)} K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^d} \cdot p_\mu(X') \cdot dV(X')| \leq$$
$$\leq 8 \cdot V_q \cdot p_{max} \cdot \sup_{X,X'} \left| K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^d} \right| \cdot \varepsilon^{q+2}.$$

Consequently

$$\left| \mathbb{E}\xi - \int_{\tilde{B}_\varepsilon(X)} K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^{q+d}} \cdot p_\mu(X') \cdot dV(X') \right| \leq$$
$$\leq 8 \cdot V_q \cdot p_{max} \cdot \sup_{X,X'} \left| K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^d} \right| \cdot \varepsilon^2.$$

2. Replace in $K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^{d+q}} \cdot p_\mu(X') dV(X')$ X' with its Riemmanian coordinates (θ, t): $X' = \exp_X(\varepsilon \cdot t\theta)$, $t \in [0,1]$ and expand around $\varepsilon = 0$. From (Lemma 2, (Yanovich, 2016))

$$K(X,\theta,\frac{|X'-X|}{\varepsilon}) = K(X,\theta,t) + t^2 \cdot \varepsilon^2 \cdot K'(X,\theta,t_1),$$

where $\tilde{t}_1 \in [0,t]$. Also

$$\frac{F(X,X')}{\varepsilon^d} = t^d \cdot \varphi(X,\theta,0) + \varepsilon \cdot t^{d+1} \cdot \varphi_1(X,\theta,0) + \varepsilon^2 \cdot \frac{t^{d+2}}{2} \cdot \varphi_2(X,\theta,t_2);$$
$$p_\mu(X') = p_\mu(X) + \varepsilon \cdot t \cdot \nabla_\theta p_\mu(X) + \varepsilon^2 \cdot t^2 \cdot (\nabla_{\tilde{\theta}})^2 p_\mu(\tilde{X}),$$

where $t_2 \in [0, t]$, $\tilde{X} = \exp_X(\tilde{t}\theta\varepsilon)$, $\tilde{\theta} \in T_X(\mathbb{M})$. Using (Lemmas 1 and 3, (Yanovich, 2016))

$$K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \cdot p_\mu(X')dV(X') = \left(K(X, \theta, t) + t^2 \cdot \varepsilon^2 \cdot K'(X, \theta, t_1)\right) \\ \cdot \left(t^d \cdot \varphi(X, \theta, 0) + \varepsilon \cdot t^{d+1} \cdot \varphi_1(X, \theta, 0) + \varepsilon^2 \cdot \frac{t^{d+2}}{2} \cdot \varphi_2(X, \theta, t_2)\right) \\ \cdot \left(p_\mu(X) + \varepsilon \cdot t \cdot \nabla_\theta p_\mu(X) + \varepsilon^2 \cdot t^2 \cdot (\nabla_\theta)^2 p_\mu(\tilde{X})\right) \cdot \left(t^{q-1} + t^{q+1} Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta})\right) dt d\theta. \quad (26)$$

Main three terms are

$$\varepsilon^0: \; K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X);$$

$$\varepsilon^1: \; K(X, \theta, t) \cdot t^{d+q} \cdot \varphi_1(X, \theta, 0) \cdot p_\mu(X) + K(X, \theta, t) \cdot t^{d+q} \cdot \varphi(X, \theta, 0) \cdot \nabla_\theta p_\mu(X);$$

$$\varepsilon^2: \; K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi_2(X, \theta, t_2) \cdot p_\mu(X) + K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi(X, \theta, 0) \cdot (\nabla_\theta)^2 p_\mu(X) \\ + t^{d+q+1} \cdot K'(X, \theta, t_1) \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) + t^{d+q+1} \cdot K(X, \theta, t) \cdot \varphi_1(X, \theta, 0) \cdot \nabla_\theta p_\mu(X) \\ + K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) \cdot Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta}).$$

3. As unit ball (t, θ) equals to $[0, 1] \times S_{q-1}$ it is symmetric. So from Lemma 2 integrals of odd powers equals 0. For $d \not\vdots 2$

$$\int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) dt d\theta = 0,$$

and for $d \vdots 2$ integral with ε^1 equals 0.

4. Integrating the item 2 over the unit sphere

$$\int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) dt d\theta = p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta;$$

$$E_1(X) \equiv \int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q} \cdot \varphi_1(X, \theta, 0) \cdot p_\mu(X) \\ + K(X, \theta, t) \cdot t^{d+q} \cdot \varphi(X, \theta, 0) \cdot \nabla_\theta p_\mu(X) dt d\theta \\ = p_\mu(X) \cdot \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi_1(X, \theta, 0) d\theta \; + \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi(X, \theta, 0) \nabla_\theta p_\mu(X) d\theta \quad (27)$$

5. Denote $E_1(X)$:

$$|E_1(X)| \leq |p_\mu(X) \cdot \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi_1(X, \theta, 0) d\theta| \\ + |\int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi(X, \theta, 0) \nabla_\theta p_\mu(X) d\theta| \leq V_q \cdot C_\rho \cdot (p_{max} \cdot C_{\phi,1} + C_{p,1} \cdot C_\phi).$$

Denote

$$C_{E,1} = V_q \cdot C_\rho \cdot (p_{max} \cdot C_{\varphi,1} + C_{p,1} \cdot C_\varphi). \tag{28}$$

sing (26), $\varepsilon \le 1$ and $t \le 1$:

$$|K(X,\theta,t) + t^2 \cdot K'(X,\theta,t_1)| \le C_K + C_{K,1};$$

$$|t^d \cdot \varphi(X,\theta,0) + t^{d+1} \cdot \varphi_1(X,\theta,0) + \frac{t^{d+2}}{2} \cdot \varphi_2(X,\theta,t_2)|$$
$$\le |\varphi(X,\theta,0) + \varphi_1(X,\theta,0) + \varphi_2(X,\theta,t_2)| \le C_\varphi + C_{\varphi,1} + C_{\varphi,2};$$

$$|p_\mu(X) + t \cdot \nabla_\theta p_\mu(X) + t^2 \cdot (\nabla_\theta)^2 p_\mu(\tilde{X})| \le |p_\mu(X) + \nabla_\theta p_\mu(X) + (\nabla_\theta)^2 p_\mu(\tilde{X})|$$
$$\le p_{max} + C_{p,1} + C_{p,2};$$

$$|1 + t^2 Ric_{\tilde{X}}(\tilde{\theta},\tilde{\theta})| \le |1 + Ric_{\tilde{X}}(\tilde{\theta},\tilde{\theta})| \le 1 + C_{Ric}.$$

Using item 1 and

$$8 \cdot V_q \cdot p_{max} \cdot \sup_{X,X'} \left| K_\varepsilon(X,X') \cdot \frac{F(X,X')}{\varepsilon^d} \right| \le 8 \cdot V_q \cdot p_{max} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})$$

get

$$C_{E,2} = V_q \cdot (C_K + C_{K,1}) \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric})$$
$$+ 8 \cdot V_q \cdot p_{max} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \tag{29}$$

6. Repeating 1-5 for ξ^2

$$\mathbb{E}\xi^2 = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta + \varepsilon^{-q+2} \cdot C_{E2,2},$$

where

$$C_{E2,2} = V_q \cdot (C_K + C_{K,1})^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2$$
$$\cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + 8 \cdot V_q \cdot p_{max} \cdot C_K^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2.$$

Finally

$$\mathtt{Var}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X,\theta)\varphi(X,\theta,0)^2 d\theta + \varepsilon^{\min\{-q+2,0\}} \cdot C_{V,2},$$

where

$$C_{V,2} = V_q \cdot (C_K + C_{K,1})^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2$$
$$\cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + 8 \cdot V_q \cdot p_{max} \cdot C_K^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2 + C_{E2,2}. \tag{30}$$

A.2 Random Variables Near the Boundary

Lemma 4. For small $\varepsilon \leq \min\{\varepsilon_B, \varepsilon_c\}$, the number of points $\mathbb{X}_{N,\varepsilon}$ from $\mathbb{X}_N$, near the boundary $\mathbb{M} \setminus \mathbb{M}_\varepsilon$ with high probability

$$P\left(|\mathbb{X}_{N,\varepsilon}| \geq 2 \cdot \varepsilon \cdot N \cdot p_{\max} \cdot dV(\mathbb{M}) \cdot C_B\right) \leq \exp\left(-\tilde{C}_B \cdot N\varepsilon\right),$$

where $\tilde{C}_B$ is a constant, $p_{\max} = \sup_{X \in \mathbb{M}} p_\mu(X)$, C_B is the constant from *M10*.

Proof of Lemma 4. An indicator of event 'point near the boundary' is Bernoulli random variable. The success probability tends to zero slower than $\varepsilon \cdot N \cdot p_{\max} \cdot dV(\mathbb{M}) \cdot C_B$ from assumption *M9*. Using Theorem (Theorem 5, (Yanovich, 2016)) one get Lemma's statement for some $\tilde{C}_B > 0$.

Lemma 5 ($F_N(X)$ near the boundary For all points $X \in \mathbb{M}$, each $N > N_{0,B}$ and $\varepsilon < \tilde{\varepsilon}_{0,B}$ and some positive constants $C_{LD1,B}, C_{LDP,B}$ for $z \in [0,1]$

$$P\left(\sup_{X \in \mathbb{M}} \left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot C_{LD1,B}\right) \leq \exp\left(-z^2 \cdot C_{LDP,B} \cdot N\varepsilon^q\right).$$

Proof of Lemma 5. As in Theorem 3 without using the symmetry of the region of integration and using *M10* and as the polynomial term is slower than the expositional one we get Lemma's statement.

References

Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Journal Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1.1.131.3745. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.131.3745.

Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, aug 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.50. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6472238.

A. Bernstein and A. Kuleshov. Manifold Learning: Generalization Ability and Tangent Proximity. *International Journal of Software and Informatics*, 7(3):359–390, 2013.

A. Bernstein, A. Kuleshov, and Y. Yanovich. Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1 – 9, Paris, 2015a. IEEE. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7344815{&}filter=AND(p{_}Publication{_}Number:7344768).

Alexander Bernstein and Alexander Kuleshov. Low-Dimensional Data Representation in Data Analysis. pages 47–58. 2014. doi: 10.1007/978-3-319-11656-3_5. URL http://link.springer.com/10.1007/978-3-319-11656-3{_}5.

Alexander Bernstein, Alexander Kuleshov, and Yury Yanovich. Manifold Learning in Regression Tasks. In *Lecture Notes in Computer Science*, volume 9047, pages 414–423. 2015b. ISBN 978-3-319-17091-6. doi: 10.1007/978-3-319-17091-6_36. URL http://link.springer.com/chapter/10.1007{%}2F978-3-319-17091-6{_}36http://link.springer.com/10.1007/978-3-319-17091-6{_}36.

Alexander Bernstein, Alexander Kuleshov, and Yury Yanovich. Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–9. IEEE, oct 2015c. ISBN 978-1-4673-8272-4. doi: 10.1109/DSAA.2015.7344815. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7344815.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 2006. ISBN 0387310738.

P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza. Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework. *Mathematical Problems in Engineering*, 2015:1–21, 2015. ISSN 1024-123X. doi: 10.1155/2015/759567. URL http://www.hindawi.com/journals/mpe/2015/759567/.

Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, jul 2006. ISSN 10635203. doi: 10.1016/j.acha.2006.04.006. URL http://linkinghub.elsevier.com/retrieve/pii/S1063520306000546.

D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10): 5591–5596, may 2003. ISSN 0027-8424. doi: 10.1073/pnas.1031596100. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1031596100.

David L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS conference on math challenges of 21st century*, pages 1–31, 2000. URL http://statweb.stanford.edu/{~}donoho/Lectures/CBMS/Curses.pdf.

Jochen Einbeck and Zakiah Kalantana. Intrinsic Dimensionality Estimation for High-dimensional Data Sets: New Approaches for the Computation of Correlation Dimension. *Journal of Emerging Technologies in Web Intelligence*, 5(2), may 2013. ISSN 1798-0461. doi: 10.4304/jetwi.5.2.91-97. URL http://www.jetwi.us/index.php?m=content{&}c=index{&}a=show{&}catid=150{&}id=801.

Mingyu Fan, Hong Qiao, and Bo Zhang. Intrinsic dimension estimation of manifolds by incising balls. *Pattern Recognition*, 42(5):780–787, may 2009. ISSN 00313203. doi: 10.1016/j.patcog.2008.09.016. URL http://linkinghub.elsevier.com/retrieve/pii/S0031320308003804.

D. J. Henderson and C. F. Parmeter. *Applied Nonparametric Econometrics*. Cambridge University Press, New York, 2015. ISBN 978-1-107-01025-3.

Alexander Kuleshov and Alexander Bernstein. Extended Regression on Manifolds Estimation. In *Lecture Notes in Computer Science*, chapter Conformal, pages 208–228.

2016. doi: 10.1007/978-3-319-33395-3_15. URL http://link.springer.com/10.1007/978-3-319-33395-3{_}15.

John A. Lee and Michel Verleysen. *Nonlinear Dimensionality Reduction.* Information Science and Statistics. Springer New York, New York, NY, 2007. ISBN 978-0-387-39350-6. doi: 10.1007/978-0-387-39351-3. URL http://link.springer.com/10.1007/978-0-387-39351-3.

Elizaveta Levina and Peter J. Bickel. Maximum Likelihood Estimation of Intrinsic Dimension. In *Advances in Neural Information Processing Systems*, pages 777–784. MIT Press, 2005. URL https://www.stat.berkeley.edu/{~}bickel/mldim.pdf.

Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications.* CRC Press, Boca Raton, 2011. URL http://www.gbv.de/dms/goettingen/689943164.pdf.

Amir massoud Farahmand, Csaba Szepesvári, and Jean-Yves Audibert. Manifold-adaptive dimension estimation. In *Proceedings of the 24th international conference on Machine learning - ICML '07*, pages 265–272, New York, New York, USA, 2007. ACM Press. ISBN 9781595937933. doi: 10.1145/1273496.1273530. URL http://portal.acm.org/citation.cfm?doid=1273496.1273530.

V. V. Petrov. *Limit Theorems for Sums of Independent Random Variables (in Russian).* Nauka, Moscow, 1987. URL http://gen.lib.rus.ec/book/index.php?md5=78c4ea42de3167ee3e6cfa56f102f950.

Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.3313.

Alessandro Rozza, Gabriele Lombardi, Marco Rosa, Elena Casiraghi, and Paola Campadelli. IDEA: Intrinsic Dimension Estimation Algorithm. pages 433–442. 2011. doi: 10.1007/978-3-642-24085-0_45. URL http://link.springer.com/10.1007/978-3-642-24085-0{_}45.

H. S. Seung and Daniel D. Lee. COGNITION: The Manifold Ways of Perception. *Science*, 290(5500):2268–2269, dec 2000. ISSN 00368075. doi: 10.1126/science.290.5500.2268. URL http://www.sciencemag.org/cgi/doi/10.1126/science.290.5500.2268.

A Singer and H.-T Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, aug 2012. ISSN 00103640. doi: 10.1002/cpa.21395. URL http://doi.wiley.com/10.1002/cpa.21395.

Andrew Smith, Hongyuan Zha, and Xiao-ming Wu. Convergence and Rate of Convergence of a Manifold-Based Dimension Reduction Algorithm. *Advances in Neural Information Processing Systems 21*, pages 1529–1536, 2009.

J. B. Tenenbaum, V. de Silva, and JC. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, dec 2000. ISSN 00368075. doi: 10.1126/science.290.5500.2319. URL http://www.sciencemag.org/cgi/doi/10.1126/science.290.5500.2319.

Michel Verleysen. Learning High-Dimensional Data. *Limitations and Future Trends in Neural Computation*, 186:141–162, 2003. URL https://perso.uclouvain.be/michel.verleysen/papers/nato03mv.pdf.

Larry Wasserman. *All of Nonparametric Statistics*. Springer-Verlag, 2006. ISBN 978-0-387-30623-0.

Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1):77–90, oct 2006. ISSN 0920-5691. doi: 10.1007/s11263-005-4939-z. URL http://link.springer.com/10.1007/s11263-005-4939-z.

Yury Yanovich. Asymptotic Properties of Local Sampling on Manifold. *Journal of Mathematics and Statistics*, 12(3):157–175, 2016. doi: 10.3844/jmssp.2016.157.175. URL http://thescipub.com/abstract/10.3844/jmssp.2016.157.175.

Yury Yanovich. Asymptotic Properties of Eigenvalues and Eigenfunctions Estimates of Linear Operators on Manifolds. *Lobachevskii Journal of Mathematics*, 2017.

Zhenyue Zhang and Hongyuan Zha. Principal Manifolds and Nonlinear Dimensionality Reduction via Local Tangent Space Alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004. ISSN 1064-8275. doi: 10.1137/S1064827502419154. URL http://epubs.siam.org/doi/abs/10.1137/S1064827502419154.