Journal of Machine Learning Research 8 (2007) 1325-1370 Submitted 5/06; Revised 3/07; Published 5/07

Graph Laplacians and their Convergence on Random Neighborhood Graphs

Matthias Hein MH@TUEBINGEN.MPG.DE
Max Planck Institute for Biological Cybernetics
Dept. Schölkopf
Spemannstraße 38
Tübingen 72076, Germany

Jean-Yves Audibert AUDIBERT@CERTIS.FR
CERTIS
Ecole Nationale des Ponts et Chaussées
19, rue Alfred Nobel - Cité Descartes
F-77455 Marne-la-Vallée cedex 2, France

Ulrike von Luxburg ULRIKE.LUXBURG@TUEBINGEN.MPG.DE
Max Planck Institute for Biological Cybernetics
Dept. Schölkopf
Spemannstraße 38
Tübingen 72076, Germany

Editor: Sanjoy Dasgupta

Abstract

Given a sample from a probability measure with support on a submanifold in Euclidean space one can construct a neighborhood graph which can be seen as an approximation of the submanifold. The graph Laplacian of such a graph is used in several machine learning methods like semi-supervised learning, dimensionality reduction and clustering. In this paper we determine the pointwise limit of three different graph Laplacians used in the literature as the sample size increases and the neighborhood size approaches zero. We show that for a uniform measure on the submanifold all graph Laplacians have the same limit up to constants. However in the case of a non-uniform measure on the submanifold only the so called random walk graph Laplacian converges to the weighted Laplace-Beltrami operator.

Keywords: graphs, graph Laplacians, semi-supervised learning, spectral clustering, dimensionality reduction

1. Introduction

In recent years, methods based on graph Laplacians have become increasingly popular in machine learning. They have been used in semi-supervised learning (Belkin and Niyogi, 2004; Zhou et al., 2004; Zhu and Ghahramani, 2002), spectral clustering (Spielman and Teng, 1996; von Luxburg, 2006) and dimensionality reduction (Belkin and Niyogi, 2003; Coifman and Lafon, 2006). Their popularity is mainly due to the following properties of the Laplacian which will be discussed in more detail in a later section:

- the Laplacian is the generator of the diffusion process (label propagation in semi-supervised learning),
- the eigenvectors of the Laplacian have special geometric properties (motivation for spectral clustering),
- the Laplacian induces an adaptive regularization functional, which adapts to the density and the geometric structure of the data (semi-supervised learning, classification).

If the data lies in $\mathbb{R}^d$ the neighborhood graph built from the random sample can be seen as an approximation of the continuous structure. In particular, if the data has support on a low-dimensional submanifold the neighborhood graph is a discrete approximation of the submanifold. In machine learning we are interested in the intrinsic properties and objects of this submanifold. The approximation of the Laplace-Beltrami operator via the graph Laplacian is a very important one since it has numerous applications as we will discuss later.

Approximations of the Laplace-Beltrami operator or related objects have been studied for certain special deterministic graphs. The easiest case is a grid in $\mathbb{R}^d$. In numerics it is standard to approximate the Laplacian with finite-differences schemes on the grid. These can be seen as a special instances of a graph Laplacian. There convergence for decreasing grid-size follows easily by an argument using Taylor expansions. Another more involved example is the work of Varopoulos (1984), where for a graph generated by an ε-packing of a manifold, the equivalence of certain properties of random walks on the graph and Brownian motion on the manifold have been established. The connection between random walks and the graph Laplacian becomes obvious by noting that the graph Laplacian as well as the Laplace-Beltrami operator are the generators of the diffusion process on the graph and the manifold, respectively. In Xu (2004) the convergence of a discrete approximation of the Laplace Beltrami operator for a triangulation of a 2D-surface in $\mathbb{R}^3$ was shown. However, it is unclear whether the approximation described there can be written as a graph Laplacian and whether this result can be generalized to higher dimensions.

In the case where the graph is generated randomly, only first results have been proved so far. The first work on the large sample limit of graph Laplacians has been done by Bousquet et al. (2004). There the authors studied the convergence of the regularization functional induced by the graph Laplacian using the law of large numbers for U-statistics. In a second step taking the limit of the neighborhoodsize $h \to 0$, they got $\frac{1}{p^2}\nabla(p^2\nabla)$ as the effective limit operator in $\mathbb{R}^d$. Their result has recently been generalized to the submanifold case and uniform convergence over the space of Hölder-functions by Hein (2005, 2006). In von Luxburg et al. (2007), the neighborhoodsize h was kept fixed while the large sample limit of the graph Laplacian was considered. In this setting, the authors showed strong convergence results of graph Laplacians to certain integral operators, which imply the convergence of the eigenvalues and eigenfunctions. Thereby showing the consistency of spectral clustering for a fixed neighborhood size.

In contrast to the previous work in this paper we will consider the large sample limit and the limit as the neighborhood size approaches zero simultaneously for a certain class of neighbhorhood graphs. The main emphasis lies on the case where the data generating measure has support on a submanifold of $\mathbb{R}^d$. The bias term, that is the difference between the continuous counterpart of the graph Laplacian and the Laplacian itself has been studied first for compact submanifolds without boundary by Smolyanov, von Weizsäcker, and Wittich (2000) and Belkin (2003) for the Gaussian kernel and a uniform data generating measure and was then generalized by Lafon (2004) to general

isotropic weights and general probability measures. Additionally Lafon showed that the use of data-dependent weights for the graph allows to control the influence of the density. They all show that the bias term converges pointwise if the neighborhood size goes to zero. The convergence of the graph Laplacian towards these continuous averaging operators was left open. This part was first studied by Hein et al. (2005) and Belkin and Niyogi (2005). In Belkin and Niyogi (2005) the convergence was shown for the so called unnormalized graph Laplacian in the case of a uniform probability measure on a compact manifold without boundary and using the Gaussian kernel for the weights, whereas in Hein et al. (2005) the pointwise convergence was shown for the random walk graph Laplacian in the case of general probability measures on non-compact manifolds with boundary using general isotropic data-dependent weights. More recently Giné and Koltchinskii (2006) have extended the pointwise convergence for the unnormalized graph Laplacian shown by Belkin and Niyogi (2005) to uniform convergence on compact submanifolds without boundary giving explicit rates. In Singer (2006), see also Giné and Koltchinskii (2006), the rate of convergence given by Hein et al. (2005) has been improved in the setting of the uniform measure. In this paper we will study the three most often used graph Laplacians in the machine learning literature and show their pointwise convergence in the general setting of Lafon (2004) and Hein et al. (2005), that is we will in particular consider the case where by using data-dependent weights for the graph we can control the influence of the density on the limit operator.

In Section 2 we introduce the basic framework necessary to define graph Laplacians for general directed weighted graphs and then simplify the general case to undirected graphs. In particular, we define the three graph Laplacians used in machine learning so far, which we call the normalized, the unnormalized and the random walk Laplacian. In Section 3 we introduce the neighborhood graphs studied in this paper, followed by an introduction to the so called weighted Laplace-Beltrami operator, which will turn out to be the limit operator in general. We also study properties of this limit operator and provide insights why and how this operator can be used for semi-supervised learning, clustering and regression. Then finally we present the main convergence result for all three graph Laplacians and give the conditions on the neighborhood size as a function of the sample size necessary for convergence. In Section 4 we illustrate the main result by studying the difference between the three graph Laplacians and the effects of different data-dependent weights on the limit operator. In Section 5 we prove the main result. We introduce a framework for studying non-compact manifolds with boundary and provide the necessary assumptions on the submanifold M, the data generating measure P and the kernel k used for defining the weights of the edges. We would like to note that the theorems given in Section 5 contain slightly stronger results than the ones presented in Section 3. The reader who is not familiar with differential geometry will find a brief introduction to the basic material used in this paper in Appendix A.

2. Abstract Definition of the Graph Structure

In this section we define the structure on a graph which is required in order to define the graph Laplacian. To this end one has to introduce Hilbert spaces H_V and H_E of functions on the vertices V and edges E, define a difference operator d, and then set the graph Laplacian as $\Delta = d^*d$. We first do this in full generality for directed graphs and then specialize it to undirected graphs. This approach is well-known for undirected graphs in discrete potential theory and spectral graph theory, see for example Dodziuk (1984), Chung (1997), Woess (2000) and McDonald and Meyers (2002),

and was generalized to directed graphs by Zhou et al. (2005) for a special choice of H_V, H_E and d. To our knowledge the very general setting introduced here has not been discussed elsewhere.

In many articles graph Laplacians are used without explicitly mentioning d, H_V and H_E. This can be misleading since, as we will show, there always exists a whole family of choices for d, H_V and H_E which all yield the same graph Laplacian.

2.1 Hilbert Spaces of Functions on the Vertices V and the Edges E

Let (V,W) be a graph where V denotes the set of vertices with $|V| = n$ and W a positive $n \times n$ similarity matrix, that is $w_{ij} \geq 0$, $i, j = 1, \ldots, n$. W need not to be symmetric, that means we consider the case of a directed graph. The special case of an undirected graph will be discussed in a following section. Let $E \subset V \times V$ be the set of edges $e_{ij} = (i, j)$ with $w_{ij} > 0$. e_{ij} is said to be a directed edge from the vertex i to the vertex j with weight w_{ij}. Moreover, we define the outgoing and ingoing sum of weights of a vertex i as

$$d_i^{out} = \frac{1}{n} \sum_{j=1}^{n} w_{ij}, \quad d_i^{in} = \frac{1}{n} \sum_{j=1}^{n} w_{ji}.$$

We assume that $d_i^{out} + d_i^{in} > 0$, $i = 1, \ldots, n$, meaning that each vertex has at least one in- or outgoing edge. Let $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$ and $\mathbb{R}_+^* = \{x \in \mathbb{R} \,|\, x > 0\}$. The inner product on the function space $\mathbb{R}^V$ is defined as

$$\langle f, g \rangle_V = \frac{1}{n} \sum_{i=1}^{n} f_i g_i \chi_i,$$

where $\chi_i = (\chi_{out}(d_i^{out}) + \chi_{in}(d_i^{in}))$ with $\chi_{out} : \mathbb{R}_+ \to \mathbb{R}_+$ and $\chi_{in} : \mathbb{R}_+ \to \mathbb{R}_+$, $\chi_{out}(0) = \chi_{in}(0) = 0$ and further χ_{out} and χ_{in} strictly positive on $\mathbb{R}_+^*$.

We also define an inner product on the space of functions $\mathbb{R}^E$ on the edges:

$$\langle F, G \rangle_E = \frac{1}{2n^2} \sum_{i,j=1}^{n} F_{ij} G_{ij} \phi(w_{ij}),$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, $\phi(0) = 0$ and ϕ strictly positive on $\mathbb{R}_+^*$. Note that with these assumptions on ϕ the sum is taken only over the set of edges E. One can check that both inner products are well-defined. We denote by $\mathcal{H}(V, \chi) = (\mathbb{R}_V, \langle \cdot, \cdot \rangle_V)$ and $\mathcal{H}(E, \phi) = (\mathbb{R}^E, \langle \cdot, \cdot \rangle_E)$ the corresponding Hilbert spaces. As a last remark let us clarify the roles of $\mathbb{R}^V$ and $\mathbb{R}^E$. The first one is the space of functions on the vertices and therefore can be regarded as a normal function space. However, elements of $\mathbb{R}^E$ can be interpreted as a flow on the edges so that the function value on an edge e_{ij} corresponds to the "mass" flowing from one vertex i to the vertex j (per unit time).

2.2 The Difference Operator d and its Adjoint d^*

Definition 1 *The **difference operator** $d : \mathcal{H}(V, \chi) \to \mathcal{H}(E, \phi)$ is defined as follows:*

$$\forall\, e_{ij} \in E, \qquad (df)(e_{ij}) = \gamma(w_{ij})\,(f(j) - f(i)),$$

where $\gamma : \mathbb{R}_+^* \to \mathbb{R}_+^*$.

Remark 2 *Note that d is zero on the constant functions as one would expect it from a derivative. In Zhou et al. (2004) another difference operator d is used:*

$$(df)(e_{ij}) = \gamma(w_{ij})\Big(\frac{f(j)}{\sqrt{d(j)}} - \frac{f(i)}{\sqrt{d(i)}}\Big). \tag{1}$$

Note that in Zhou et al. (2004) they have $\gamma(w_{ij}) \equiv 1$. This difference operator is in general not zero on the constant functions. This in turn leads to the effect that the associated Laplacian is also not zero on the constant functions. For general graphs without any geometric interpretation this is just a matter of choice. However, the choice of d matters if one wants a consistent continuum limit of the graph. One cannot expect convergence of the graph Laplacian associated to the difference operator d of Equation (1) *towards a Laplacian, since as each of the graph Laplacians in the sequence is not zero on the constant functions, also the limit operator will share this property unless $\lim_{n\to\infty} d(X_i) = c, \forall i = 1,\ldots,n$, where c is a constant. We derive also the limit operator of the graph Laplacian induced by the difference operator of Equation* (1) *introduced by Zhou et al. in the machine learning literature and usually denoted as the normalized graph Laplacian in spectral graph theory (Chung, 1997).*

Obviously, in the finite case d is always a bounded operator. The adjoint operator $d^* : \mathcal{H}(E,\phi) \to \mathcal{H}(V,\chi)$ is defined by

$$\langle df, u\rangle_E = \langle f, d^*u\rangle_V, \quad \forall f \in H(V,\chi), \quad u \in \mathcal{H}(E,\phi).$$

Lemma 3 *The adjoint $d^* : \mathcal{H}(E,\phi) \to \mathcal{H}(V,\chi)$ of the difference operator d is explicitly given by:*

$$(d^*u)(l) = \frac{1}{2\chi_l}\left(\frac{1}{n}\sum_{i=1}^{n}\gamma(w_{il})\,u_{il}\,\phi(w_{il}) - \frac{1}{n}\sum_{i=1}^{n}\gamma(w_{li})\,u_{li}\,\phi(w_{li})\right). \tag{2}$$

Proof Using the indicator function $f(j) = \mathbb{1}_{j=l}$ it is straightforward to derive:

$$\frac{1}{n}\chi_l\,(d^*u)(l) = \langle d\mathbb{1}_{\cdot=l}, u\rangle_E = \frac{1}{2n^2}\sum_{i=1}^{n}\Big(\gamma(w_{il})u_{il}\phi(w_{il}) - \gamma(w_{li})u_{li}\phi(w_{li})\Big),$$

where we have used $\langle d\mathbb{1}_{\cdot=l}, u\rangle_E = \frac{1}{2n^2}\sum_{i,j=1}^{n}(d\mathbb{1}_{\cdot=l})_{ij}u_{ij}\phi(w_{ij})$. ■

The first term of the rhs of (2) can be interpreted as the outgoing flow, whereas the second term can be seen as the ingoing flow. The corresponding continuous counterpart of d is the gradient of a function and for d^* it is the divergence of a vector-field, measuring the infinitesimal difference between in- and outgoing flow.

2.3 The General Graph Laplacian

Definition 4 (graph Laplacian for a directed graph) *Given Hilbert spaces $\mathcal{H}(V,\chi)$ and $\mathcal{H}(E,\phi)$ and the difference operator $d : \mathcal{H}(V,\chi) \to \mathcal{H}(E,\phi)$ the graph Laplacian $\Delta : \mathcal{H}(V,\chi) \to \mathcal{H}(V,\chi)$ is defined as*

$$\Delta = d^*d.$$

Lemma 5 *Explicitly, $\Delta : \mathcal{H}(V,\chi) \to \mathcal{H}(V,\chi)$ is given as:*

$$(\Delta f)(l) = \frac{1}{2\chi_l}\left[\frac{1}{n}\sum_{i=1}^{n}\big(\gamma(w_{il})^2\phi(w_{il}) + \gamma(w_{li})^2\phi(w_{li})\big)\big(f(l)-f(i)\big)\right].$$

Proof The explicit expression Δ can be easily derived by plugging the expression of d^* and d together:

$$\begin{aligned}(d^*df)(l) =& \frac{1}{2\chi_l}\left(\frac{1}{n}\sum_{i=1}^{n}\gamma(w_{il})^2[f(l)-f(i)]\phi(w_{il}) - \frac{1}{n}\sum_{i=1}^{n}\gamma(w_{li})^2[f(i)-f(l)]\phi(w_{li})\right)\\ =& \frac{1}{2\chi_l}\left[f(l)\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_{ij} - \frac{1}{n}\sum_{i=1}^{n}f(i)\widehat{w}_{ij}\right],\end{aligned}$$

where we have introduced $\widehat{w}_{ij} = \big(\gamma(w_{il})^2\phi(w_{il}) + \gamma(w_{li})^2\phi(w_{li})\big)$. ■

Proposition 6 *Δ is self-adjoint and positive semi-definite.*

Proof By definition, $\langle f, \Delta g\rangle_V = \langle df, dg\rangle_E = \langle \Delta f, g\rangle_V$, and $\langle f, \Delta f\rangle_V = \langle df, df\rangle_E \geq 0$. ■

2.4 The Special Case of an Undirected Graph

In the case of an undirected graph we have $w_{ij} = w_{ji}$, that is whenever there is an edge from i to j there is an edge with the same value from j to i. This implies that there is no difference between in- and outgoing edges. Therefore, $d_i^{out} \equiv d_i^{in}$, so that we will denote the degree function by d with $d_i = \frac{1}{n}\sum_{j=1}^{n} w_{ij}$. The same for the weights in H_V, $\chi_{out} \equiv \chi_{in}$, so that we have only one function χ. If one likes to interpret functions on E as flows, it is reasonable to restrict the space $\mathcal{H}_E$ to antisymmetric functions since symmetric functions are associated to flows which transport the same mass from vertex i to vertex j and back. Therefore, as a net effect, no mass is transported at all so that from a physical point of view these functions cannot be observed at all. Since anyway we consider only functions on the edges of the form df (where f is in $\mathcal{H}_V$) which are by construction antisymmetric, we will not do this restriction explicitly.

The adjoint d^* simplifies in the undirected case to

$$(d^*u)(l) = \frac{1}{2\chi(d_l)}\frac{1}{n}\sum_{i=1}^{n}\gamma(w_{il})\phi(w_{il})(u_{il}-u_{li}),$$

and the general graph Laplacian on an undirected graph has the following form:

Definition 7 (graph Laplacian for an undirected graph) *Given Hilbert spaces $\mathcal{H}(V,\chi)$ and $\mathcal{H}(E,\phi)$ and the difference operator $d : \mathcal{H}(V,\chi) \to \mathcal{H}(E,\phi)$ the graph Laplacian $\Delta : \mathcal{H}(V,\chi) \to \mathcal{H}(V,\chi)$ is defined as*

$$\Delta = d^*d.$$

Explicitly, for any vertex l, we have

$$(\Delta f)(l) = (d^*df)(l) = \frac{1}{\chi(d_l)}\left[f(l)\frac{1}{n}\sum_{i=1}^{n}\gamma^2(w_{il})\phi(w_{il}) - \frac{1}{n}\sum_{i=1}^{n}f(i)\gamma^2(w_{il})\phi(w_{il})\right].$$

In the literature one finds the following special cases of the general graph Laplacian. Unfortunately there exist no unique names for the three graph Laplacians we introduce here, most of the time all of them are just called graph Laplacians. Only the term 'unnormalized' or 'combinatorial' graph Laplacian seems to be established now. However, the other two could both be called normalized graph Laplacian. Since the first one is closely related to a random walk on the graph we call it random walk graph Laplacian and the other one normalized graph Laplacian.

The 'random walk' graph Laplacian is defined as:

$$(\Delta^{(\mathrm{rw})} f)(i) = f(i) - \frac{1}{d_i}\frac{1}{n}\sum_{j=1}^{n} w_{ij} f(j), \quad \Delta^{(\mathrm{rw})} f = (\mathbb{1} - D^{-1}W)f,$$

where the matrix D is defined as $D_{ij} = d_i \delta_{ij}$. Note that $P = D^{-1}W$ is a stochastic matrix and therefore can be used to define a Markov random walk on V, see for example Woess (2000). The 'unnormalized' (or 'combinatorial') graph Laplacian is defined as

$$(\Delta^{(\mathrm{u})} f)(i) = d_i f(i) - \frac{1}{n}\sum_{j=1}^{n} w_{ij} f(j), \quad \Delta^{(\mathrm{u})} f = (D - W)f.$$

We have the following conditions on χ, γ and ϕ in order to get these Laplacians:

$$\forall e_{ij} \in E: \quad \mathrm{rw}: \quad \frac{\gamma^2(w_{ij})\phi(w_{ij})}{\chi(d_i)} = \frac{w_{ij}}{d_i}, \quad \mathrm{unnorm}: \quad \frac{\gamma^2(w_{ij})\phi(w_{ij})}{\chi(d_i)} = w_{ij}.$$

We observe that by choosing $\Delta^{(\mathrm{rw})}$ or $\Delta^{(\mathrm{u})}$ the functions ϕ and γ are not fixed. Therefore it can cause confusion if one speaks of the 'random walk' or 'unnormalized' graph Laplacian without explicitly defining the corresponding Hilbert spaces and the difference operator. We also consider the normalized graph Laplacian $\Delta^{(\mathrm{n})}$ introduced by Chung (1997); Zhou et al. (2004) using the difference operator of Equation (1) and the general spaces $H(V,\chi)$ and $\mathcal{H}(E,\phi)$. Following the scheme a straightforward calculation shows the following:

Lemma 8 *The graph Laplacian $\Delta_{\mathrm{norm}} = d^* d$ with the difference operator d from Equation* (1) *can be explicitly written as*

$$(\Delta^{(n)} f)(l) = \frac{1}{n\chi(d_l)\sqrt{d_l}}\left[\frac{f(l)}{\sqrt{d_l}}\frac{1}{n}\sum_{i=1}^{n}\gamma^2(w_{il})\phi(w_{il}) - \frac{1}{n}\sum_{i=1}^{n}\frac{f(i)}{\sqrt{d_i}}\gamma^2(w_{il})\phi(w_{il})\right].$$

The choice $\chi(d_l) = 1$ and $\gamma^2(w_{il})\phi(w_{il}) = w_{il}$ leads then to the graph Laplacian proposed in Chung and Langlands (1996); Zhou et al. (2004),

$$(\Delta^{(n)} f)(l) = \frac{1}{n\sqrt{d_l}}\left[\frac{f(l)}{\sqrt{d_l}} d_l - \frac{1}{n}\sum_{i=1}^{n}\frac{f(i)}{\sqrt{d_i}} w_{li}\right] = \frac{1}{n}\left[f(l) - \frac{1}{n}\sum_{i=1}^{n} f(i)\frac{w_{il}}{\sqrt{d_l\, d_i}}\right],$$

or equivalently

$$\Delta^{(n)} f = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} f = (\mathbb{1} - D^{-\frac{1}{2}} W D^{-\frac{1}{2}})f.$$

Note that $\Delta^{(\mathrm{u})} = D\Delta^{(\mathrm{rw})}$ and $\Delta^{(\mathrm{n})} = D^{-\frac{1}{2}}\Delta^{(\mathrm{u})}D^{-\frac{1}{2}}$.

3. Limit of the Graph Laplacian for Random Neighborhood Graphs

Before we state the convergence results for the three graph Laplacians on random neighborhood graphs, we first have to define the limit operator. Maybe not surprisingly, in general the Laplace-Beltrami operator will *not* be the limit operator of the graph Laplacian. Instead it will converge to the weighted Laplace-Beltrami operator which is the natural generalization of the Laplace-Beltrami operator for a Riemannian manifold equipped with a non-uniform probability measure. The definition of this limit operator and a discussion of its use for different applications in clustering, semi-supervised learning and regression is the topic of the next section, followed by a sketch of the convergence results.

3.1 Construction of the Neighborhood Graph

We assume to have a sample X_i, $i = 1,\ldots,n$ drawn i.i.d. from a probability measure P which has support on a submanifold M. For the exact assumptions regarding P, M and the kernel function k used to define the weights we refer to Section 5.2. The sample then determines the set of vertices V of the graph. Additionally we are given a certain kernel function $k : \mathbb{R}_+ \to \mathbb{R}_+$ and the neighborhood parameter $h \in \mathbb{R}_+^*$. As proposed by Lafon (2004) and Coifman and Lafon (2006), we use this kernel function k to define the following family of data-dependent kernel functions $\tilde{k}_{\lambda,h}$ parameterized by $\lambda \in \mathbb{R}$ as:

$$\tilde{k}_{\lambda,h}(X_i,X_j) = \frac{1}{h^m} \frac{k(\left\|X_i - X_j\right\|^2 / h^2)}{[d_{h,n}(X_i) d_{h,n}(X_j)]^\lambda},$$

where $d_{h,n}(X_i) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h^m} k(\left\|X_i - X_j\right\|^2 / h^2)$ is the degree function introduced in Section 2 with respect to the edge-weights $\frac{1}{h^m} k(\left\|X_i - X_j\right\|^2 / h^2)$. Finally we use $\tilde{k}_{\lambda,h}$ to define the weight $w_{ij} = w(X_i,X_j)$ of the edge between the points X_i and X_j as

$$w_{\lambda,h}(X_i,X_j) = \tilde{k}_{\lambda,h}(X_i,X_j).$$

Note that the case $\lambda = 0$ corresponds to weights with no data-dependent modification. The parameter $h \in \mathbb{R}_+^*$ determines the neighborhood of a point since we will assume that k has compact support, that is X_i and X_j have an edge if $\left\|X_i - X_j\right\| \leq hR_k$ where R_k is the support of kernel function. Note that we will have $k(0) = 0$, so that there are no loops in the graph. This assumption is not necessary, but it simplifies the proofs and makes some of the estimators unbiased.

In Section 2.4 we introduced the random walk, the unnormalized and the normalized graph Laplacian. From now on we consider these graph Laplacians for the random neighborhood graph, that is the weights of the graph w_{ij} have the form $w_{ij} = w(X_i,X_j) = \tilde{k}_{\lambda,h}(X_i,X_j)$. Using the kernel function we can easily extend the graph Laplacians to the whole submanifold M. These extensions can be seen as estimators for the Laplacian on M. We introduce also the extended degree function $\tilde{d}_{\lambda,h,n}$ and the average operator $\tilde{A}_{\lambda,h,n}$,

$$\tilde{d}_{\lambda,h,n}(x) = \frac{1}{n}\sum_{j=1}^n \tilde{k}_{\lambda,h}(x,X_j), \qquad (\tilde{A}_{\lambda,h,n} f)(x) = \frac{1}{n}\sum_{j=1}^n \tilde{k}_{\lambda,h}(x,X_j) f(X_j).$$

Note that $\tilde{d}_{\lambda,h,n} = \tilde{A}_{\lambda,h,n}1$. The extended graph Laplacians are defined as follows:

$$\begin{aligned}
\text{random walk} \qquad & (\Delta^{(\mathrm{rw})}_{\lambda,h,n} f)(x) = \frac{1}{h^2}\left(f - \tfrac{1}{\tilde{d}_{\lambda,h,n}}\tilde{A}_{\lambda,h,n} f\right)(x) = \frac{1}{h^2}\left(\frac{\tilde{A}_{\lambda,h,n} g}{\tilde{d}_{\lambda,h,n}}\right)(x), \qquad (3)\\
\text{unnormalized} \qquad & (\Delta^{(\mathrm{u})}_{\lambda,h,n} f)(x) = \frac{1}{h^2}\left(\tilde{d}_{\lambda,h,n} f - \tilde{A}_{\lambda,h,n} f\right)(x) = \frac{1}{h^2}(\tilde{A}_{\lambda,h,n} g)(x),\\
\text{normalized} \qquad & (\Delta^{(\mathrm{n})}_{\lambda,h,n} f)(x) = \tfrac{1}{h^2\sqrt{\tilde{d}_{\lambda,h,n}(x)}}\left(\tilde{d}_{\lambda,h,n}\tfrac{f}{\sqrt{\tilde{d}_{\lambda,h,n}}} - \left(\tilde{A}_{\lambda,h,n}\tfrac{f}{\sqrt{\tilde{d}_{\lambda,h,n}}}\right)\right)(x)\\
& \qquad = \tfrac{1}{h^2\sqrt{\tilde{d}_{\lambda,h,n}(x)}}(\tilde{A}_{\lambda,h,n} g')(x),
\end{aligned}$$

where we have introduced $g(y) := f(x) - f(y)$ and $g'(y) := \frac{f(x)}{\sqrt{\tilde{d}_{\lambda,h,n}(x)}} - \frac{f(y)}{\sqrt{\tilde{d}_{\lambda,h,n}(y)}}$. Note that all extensions reproduce the graph Laplacian on the sample:

$$(\Delta f)(i) = (\Delta f)(X_i) = (\Delta_{\lambda,h,n} f)(X_i), \quad \forall i = 1,\ldots,n.$$

The factor $1/h^2$ arises by introducing a factor $1/h$ in the weight γ of the derivative operator d of the graph. This is necessary since d is supposed to approximate a derivative. Since the Laplacian corresponds to a second derivative we get from the definition of the graph Laplacian a factor $1/h^2$.

We would like to note that in the case of the random walk and and the normalized graph Laplacian the normalization with $1/h^m$ in the weights cancels out, whereas it does not cancel for the unnormalized graph Laplacian except in the case $\lambda = 1/2$. The problem here is that in general the intrinsic dimension m of the manifold is unknown. Therefore a normalization with the correct factor $\frac{1}{h^m}$ is not possible, and in the limit $h \to 0$ the estimate of the unnormalized graph Laplacian will generally either vanish or blow up. The easy way to circumvent this is just to rescale the whole estimate such that $\frac{1}{n}\sum_{i=1}^n \tilde{d}_{\lambda,h,n}(X_i)$ equals a fixed constant for every n. The disadvantage is that this method of rescaling introduces a global factor in the limit. A more elegant way might be to simultaneously estimate the dimension m of the submanifold and use the estimated dimension to calculate the correct normalization factor, see, for example, Hein and Audibert (2005). However, in this work we assume for simplicity that for the unnormalized graph Laplacian the intrinsic dimension m of the submanifold is known. It might be interesting to consider both estimates simultaneously, but we leave this as an open problem.

We will consider in the following the limit $h \to 0$, that is the neighborhood of each point X_i shrinks to zero. However, since $n \to \infty$ and h as a function of n approaches zero sufficiently slow, the number of points in each neighborhood approaches ∞, so that roughly spoken sums approximate the corresponding integrals. This is the basic principle behind our convergence result and is well known in the framework of nonparametric regression (see Györfi et al., 2004).

3.2 The Weighted Laplacian and the Continuous Smoothness Functional

The Laplacian is one of the most prominent operators in mathematics. The following general properties are taken from the books of Rosenberg (1997) and Bérard (1986). It occurs in many partial differential equations governing physics, mainly because in Euclidean space it is the only linear second-order differential operator which is translation and rotation invariant. In Euclidean space $\mathbb{R}^d$ it is defined as $\Delta_{\mathbb{R}^d} f = \mathrm{div}(\mathrm{grad}\, f) = \sum_{i=1}^d \partial_i^2 f$. Moreover, for any domain $\Omega \subseteq \mathbb{R}^d$ it is a negative-semidefinite symmetric operator on $C_c^\infty(\Omega)$, which is a dense subset of $L_2(\Omega)$ (formally self-adjoint),

and satisfies

$$\int_\Omega f \Delta h \, dx = -\int_\Omega \langle \nabla f, \nabla h \rangle \, dx.$$

It can be extended to a self-adjoint operator on $L_2(\Omega)$ in several ways depending on the choice of boundary conditions. For any compact domain Ω (with suitable boundary conditions) it can be shown that Δ has a pure point spectrum and the eigenfunctions are smooth and form a complete orthonormal basis of $L_2(\Omega)$, see, for example, Bérard (1986).

The Laplace-Beltrami operator on a Riemannian manifold M is the natural equivalent of the Laplacian in $\mathbb{R}^d$, defined as

$$\Delta_M f = \operatorname{div}(\operatorname{grad} f) = \nabla^a \nabla_a f.$$

However, the more natural definition is the following. For any $f, g \in C_c^\infty(M)$, we have

$$\int_M f \Delta h \, dV(x) = -\int_M \langle \nabla f, \nabla h \rangle \, dV(x),$$

where $dV = \sqrt{\det g}\, dx$ is the natural volume element of M. This definition allows easily an extension to the case where we have a Riemannian manifold M with a measure P . In this paper P will be the probability measure generating the data. We assume in the following that P is absolutely continuous wrt the natural volume element dV of the manifold. Its density is denoted by p. Note that the case when the probability measure is absolutely continuous wrt the Lebesgue measure on $\mathbb{R}^d$ is a special case of our setting.

A recent review article about the weighted Laplace-Beltrami operator is Grigoryan (2006).

Definition 9 (Weighted Laplacian) *Let (M, g_{ab}) be a Riemannian manifold with measure P where P has a differentiable and positive density p with respect to the natural volume element $dV = \sqrt{\det g}\, dx$, and let Δ_M be the Laplace-Beltrami operator on M. For $s \in \mathbb{R}$, we define the s-th weighted Laplacian Δ_s as*

$$\Delta_s := \Delta_M + \frac{s}{p} g^{ab} (\nabla_a p) \nabla_b = \frac{1}{p^s} g^{ab} \nabla_a (p^s \nabla_b) = \frac{1}{p^s} \operatorname{div}(p^s \operatorname{grad}).$$

This definition is motivated by the following equality, for $f, g \in C_c^\infty(M)$,

$$\int_M f (\Delta_s g)\, p^s dV = \int_M f \big(\Delta g + \frac{s}{p} \langle \nabla p, \nabla g \rangle \big) p^s dV = -\int_M \langle \nabla f, \nabla g \rangle \, p^s dV, \tag{4}$$

The family of weighted Laplacians contains two cases which are particularly interesting. The first one, $s = 0$, corresponds to the standard Laplace-Beltrami operator. This case is interesting if one only wants to use properties of the geometry of the manifold but not of the data generating probability measure. The second case, $s = 1$, corresponds to the standard weighted Laplacian $\Delta_1 = \frac{1}{p} \nabla^a (p \nabla_a)$.

In the next section it will turn out that through a data-dependent change of the weights of the graph we can get the just defined weighted Laplacians as the limit operators of the graph Laplacian. The rest of this section will be used to motivate the importance of the understanding of this limit in different applications. Three different but closely related properties of the Laplacian are used in machine learning

- The Laplacian generates the diffusion process. In semi-supervised learning algorithms with a small number of labeled points one would like to propagate the labels along regions of high-density, see Zhu and Ghahramani (2002); Zhu et al. (2003). The limit operator Δ_s shows the influence of a non-uniform density p. The second term $\frac{s}{p}\langle \nabla p, \nabla f \rangle$ leads to an anisotropy in the diffusion. If $s < 0$ this term enforces diffusion in the direction of the maximum of the density whereas diffusion in the direction away from the maximum of the density is weakened. If $s < 0$ this is just the other way round.

- The smoothness functional induced by the weighted Laplacian Δ_s, see Equation (4), is given by

$$S(f) = \int_M \|\nabla f\|^2 \, p^s \, dV.$$

  For $s > 0$ this smoothness functional prefers functions which are smooth in high-density regions whereas unsmooth behavior in low-density is penalized less. This property can also be interesting in semi-supervised learning where one assumes especially when only a few labeled points are known that the classifier should be constant in high-density regions whereas changes of the classifier are allowed in low-density regions, see Bousquet et al. (2004) for some discussion of this point and Hein (2005, 2006) for a proof of convergence of the regularizer induced by the graph Laplacian towards the smoothness functional $S(f)$. In Figure 1 this is illustrated by mapping a density profile in $\mathbb{R}^2$ onto a two-dimensional manifold. However,

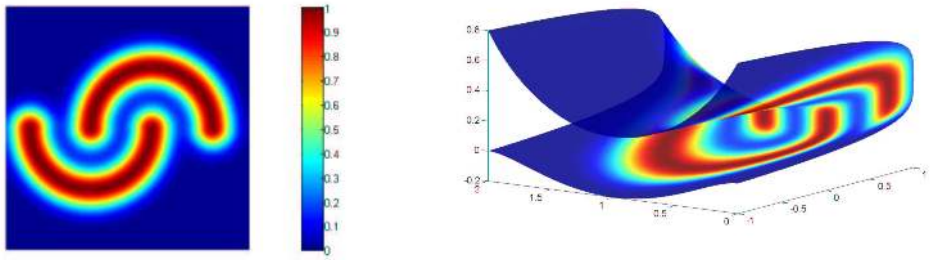

Figure 1: A density profile mapped onto a two-dim. submanifold in $\mathbb{R}^3$ with two clusters.

  also the case $s < 0$ can be interesting. Minimizing the smoothness functional $S(f)$ implies that one enforces smoothness of the function f where one has little data, and one allows the function to vary more where one has sampled a lot of data points. Such a penalization has been considered by Canu and Elisseeff (1999) for regression.

- The eigenfunctions of the Laplacian Δ_s can be seen as the limit partioning of spectral clustering for the normalized graph Laplacian (however, a rigorous mathematical proof has not been given yet, see von Luxburg et al., 2007, for a convergence result for fixed h). If $s = 0$ one gets just a geometric clustering in the sense that irrespectively of the probability measure generating the data the clustering is determined by the geometry of the submanifold. If $s > 0$

the eigenfunction corresponding to the first non-zero eigenvalue is likely to change its sign in a low-density region. This argument follows from the previous discussion on the smoothness functional $S(f)$ and the Rayleigh-Ritz principle. Let us assume for a moment that M is compact without boundary and that $p(x) > 0, \forall x \in M$, then the eigenspace corresponding to the first eigenvalue $\lambda_0 = 0$ is given by the constant functions. The first non-zero eigenvalue can then be determined by the Rayleigh-Ritz variational principle

$$\lambda_1 = \inf_{u \in C^\infty(M)} \left\{ \frac{\int_M \|\nabla u\|^2 \, p(x)^s dV(x)}{\int_M u^2(x) p(x)^s dV(x)} \;\Big|\; \int_M u(x)\, p(x)^s dV(x) = 0 \right\}.$$

Since the first eigenfunction has to be orthogonal to the constant functions, it has to change its sign. However, since $\|\nabla u\|^2$ is weighted by a power of the density p^s it is obvious that for $s > 0$ the function will change its sign in a region of low density.

3.3 Limit of the Graph Laplacians

The following theorem summarizes and slightly weakens the results of Theorem 30 and Theorem 31 of Section 5.

Main Result *Let M be a m-dimensional submanifold in $\mathbb{R}^d$, $\{X_i\}_{i=1}^n$ a sample from a probability measure P on M with density p. Let $x \in M \backslash \partial M$ and define $s = 2(1-\lambda)$. Then under technical conditions on the submanifold M, the kernel k and the density p introduced in Section 5, if $h \to 0$ and $nh^{m+2}/\log n \to \infty$,*

$$\text{random walk:} \qquad \lim_{n\to\infty} (\Delta^{(\mathrm{rw})}_{\lambda,h,n} f)(x) \;\sim\; -(\Delta_s f)(x) \qquad \text{almost surely,}$$

$$\text{unnormalized:} \qquad \lim_{n\to\infty} (\Delta^{(\mathrm{u})}_{\lambda,h,n} f)(x) \;\sim\; -p(x)^{1-2\lambda} (\Delta_s f)(x) \qquad \text{almost surely.}$$

The optimal rate is obtained for $h(n) = O\Big((\log n/n)^{\frac{1}{m+4}}\Big)$. If $h \to 0$ and $nh^{m+4}/\log n \to \infty$,

$$\text{normalized:} \qquad \lim_{n\to\infty} (\Delta^{(\mathrm{n})}_{\lambda,h,n} f)(x) \;\sim\; -p(x)^{\frac{1}{2}-\lambda} \Delta_s \left(\frac{f}{p^{\frac{1}{2}-\lambda}} \right)(x) \qquad \text{almost surely.}$$

where $\sim$ means that there exists a constant only depending on the kernel k and λ such that equality holds.

The first observation is that the conjecture that the graph Laplacian approximates the Laplace-Beltrami operator is only true for the uniform measure, where p is constant. In this case all limits agree up to constants. However, big differences arise when one has a non-uniform measure on the submanifold, which is the generic case in machine learning applications. In this case all limits disagree and only the random walk graph Laplacian converges towards the weighted Laplace-Beltrami operator which is the natural generalization of the Laplace-Beltrami operator when the manifold is equipped with a non-uniform probability measure. The unnormalized graph Laplacian has the additional factor $p^{1-2\lambda}$. However, this limit is actually quite useful, when one thinks of applications of so called label propagation algorithms in semi-supervised learning. If one uses this graph Laplacian as the diffusion operator to propagate the labeled data, it means that the diffusion for $\lambda < 1/2$ is faster in regions where the density is high. The consequence is that labels in regions of high density are propagated faster than labels in low-density regions. This makes sense since under

the cluster assumption labels in regions of low density are less informative than labels in regions of high density. In general, from the viewpoint of a diffusion process the weighted Laplace-Beltrami operator $\Delta_s = \Delta_M + \frac{s}{p}\nabla p \nabla$ can be seen as inducing an anisotropic diffusion due to the extra term $\frac{s}{p}\nabla p \nabla$, which is directed towards or away from increasing density depending on s. This is a desired property in semi-supervised learning, where one actually wants that the diffusion is mainly along regions of the same density level in order to fulfill the cluster assumption.

The second observation is that the data-dependent modification of edge weights allows to control the influence of the density on the limit operator as observed by Coifman and Lafon (2006). In fact one can even eliminate it for $s = 0$ resp. $\lambda = 1$ in the case of the random walk graph Laplacian. This could be interesting in computer graphics where the random walk graph Laplacian is used for mesh and point cloud processing, see, for example, Sorkine (2006). If one has gathered points of a curved object with a laser scanner it is likely that the points have a non-uniform distribution on the object. Its surface is a two-dimensional submanifold in $\mathbb{R}^3$. In computer graphics the non-uniform measure is only an artefact of the sampling procedure and one is only interested in the Laplace-Beltrami operator to infer geometric properties. Therefore the elimination of the influence of a non-uniform measure on the submanifold is of high interest there. We note that up to a multiplication with the inverse of the density the elimination of density effects is also possible for the unnormalized graph Laplacian, but not for the normalized graph Laplacian. All previous observations are naturally also true if the data does not lie on a submanifold but has d-dimensional support in $\mathbb{R}^d$.

The interpretation of the limit of the normalized graph Laplacian is more involved. An expansion of the limit operator shows the complex dependency on the density p:

$$p^{\frac{1}{2}-\lambda}\Delta_s\left(\frac{f}{p^{\frac{1}{2}-\lambda}}\right) = \Delta_M f + \frac{1}{p}\nabla p \nabla f - (\lambda - \frac{1}{2})^2 \frac{f}{p}\|\nabla p\|^2 + (\lambda - \frac{1}{2})\frac{f}{p}\Delta_M p.$$

We leave it to the reader to think of possible applications of this Laplacian.

The discussion shows that the choice of the graph Laplacian depends on what kind of problem one wants to solve. Therefore, in our opinion there is no universal best choice between the random walk and the unnormalized graph Laplacian from a machine learning point of view. However, from a mathematical point of view only the random walk graph Laplacian has the correct (pointwise) limit to the weighted Laplace-Beltrami operator.

4. Illustration of the Results

In this section we want to illustrate the differences between the three graph Laplacians and the control of the influence of the data-generating measure via the parameter λ.

4.1 Flat Space $\mathbb{R}^2$

In the first example the data lies in Euclidean space $\mathbb{R}^2$. Here we want to show two things. First, the sketch of the main result shows that if the data generating measure is uniform all graph Laplacians converge for all values of the reweighting parameter λ up to constants to the Laplace-Beltrami operator, which is in the case of $\mathbb{R}^2$ just the standard Laplacian. In Figure 2 the estimates of the three graph Laplacians are shown for the uniform measure $[-3,3]^2$ and $\lambda = 0$. It can be seen that up to a scaling all estimates agree very well. In a second example we study the effect of a non-uniform data-generating measure. In general all estimates disagree in this case. We illustrate this effect

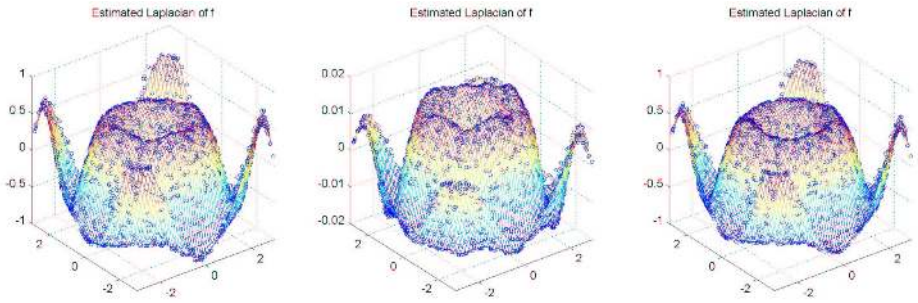

Figure 2: For the uniform distribution all graph Laplacians, $\Delta^{(\mathrm{rw})}_{\lambda,h,n}$, $\Delta^{(\mathrm{u})}_{\lambda,h,n}$ and $\Delta^{(\mathrm{n})}_{\lambda,h,n}$ (from left to right) agree up to constants for all λ. In the figure the estimates of the Laplacian are shown for the uniform measure on $[-3,3]^2$ and the function $f(x) = \sin(\frac{1}{2}\|x\|^2)/\|x\|^2$ with 2500 samples and $h = 1.4$.

in the case of $\mathbb{R}^2$ with a Gaussian distribution $\mathcal{N}(0,1)$ as data-generating measure and the simple function $f(x) = \sum_i x_i - 4$. Note that $\Delta f = 0$ so that for the random walk and the unnormalized graph Laplacian only the anisotropic part of the limit operator, $\frac{1}{p}\nabla p \nabla f$ is non-zero. Explicitly the limits are given as

$$\begin{aligned}
\Delta^{(\mathrm{rw})} \sim \quad & \Delta_s f = \Delta f + \frac{s}{p}\nabla p \nabla f = -s\sum_i x_i, \\
\Delta^{(\mathrm{u})} \sim \quad & p^{1-2\lambda}\Delta_s f = -s\, e^{-\frac{1-2\lambda}{2}\|x\|^2}\sum_i x_i, \\
\Delta^{(\mathrm{n})} \sim \quad & p^{\frac{1}{2}-\lambda}\Delta_{s\frac{f}{p^{\frac{1}{2}-\lambda}}} = -\sum_i x_i - \Big(\sum_i x_i - 4\Big)\Big[(\lambda-\frac{1}{2})(\frac{3}{2}-\lambda)\|x\|^2 - 2(\lambda-\frac{1}{2})\Big].
\end{aligned}$$

This shows that even applied to simple functions there can be large differences between the different limit operators provided the samples come from a non-uniform probability measure. Note that like in nonparametric kernel regression the estimate is quite bad at the boundary. This well known boundary effect arises since at the boundary one does not average over a full ball but only over some part of a ball. Thus the first derivative ∇f of order $O(h)$ does not cancel out so that multiplied with the factor $1/h^2$ we have a term of order $O(1/h)$ which blows up. Roughly spoken this effect takes place at all points of order $O(h)$ away from the boundary, see also Coifman and Lafon (2006).

4.2 The Sphere S^2

In our next example we consider the case where the data lies on a submanifold M in $\mathbb{R}^d$. Here we want to illustrate in the case of a sphere S^2 in $\mathbb{R}^3$ the control of the influence of the density via the parameter λ. In this case we sample from the probability measure with density $p(\phi,\theta) = \frac{1}{8\pi} + \frac{3}{8\pi}\cos^2(\theta)$ in spherical coordinates with respect to the volume element $dV = \sin(\theta)d\theta d\phi$. This density has a two-cluster structure on the sphere, where the northern and southern hemisphere represent one cluster. An estimate of the density p is shown in the Figure 4. We show the results of the

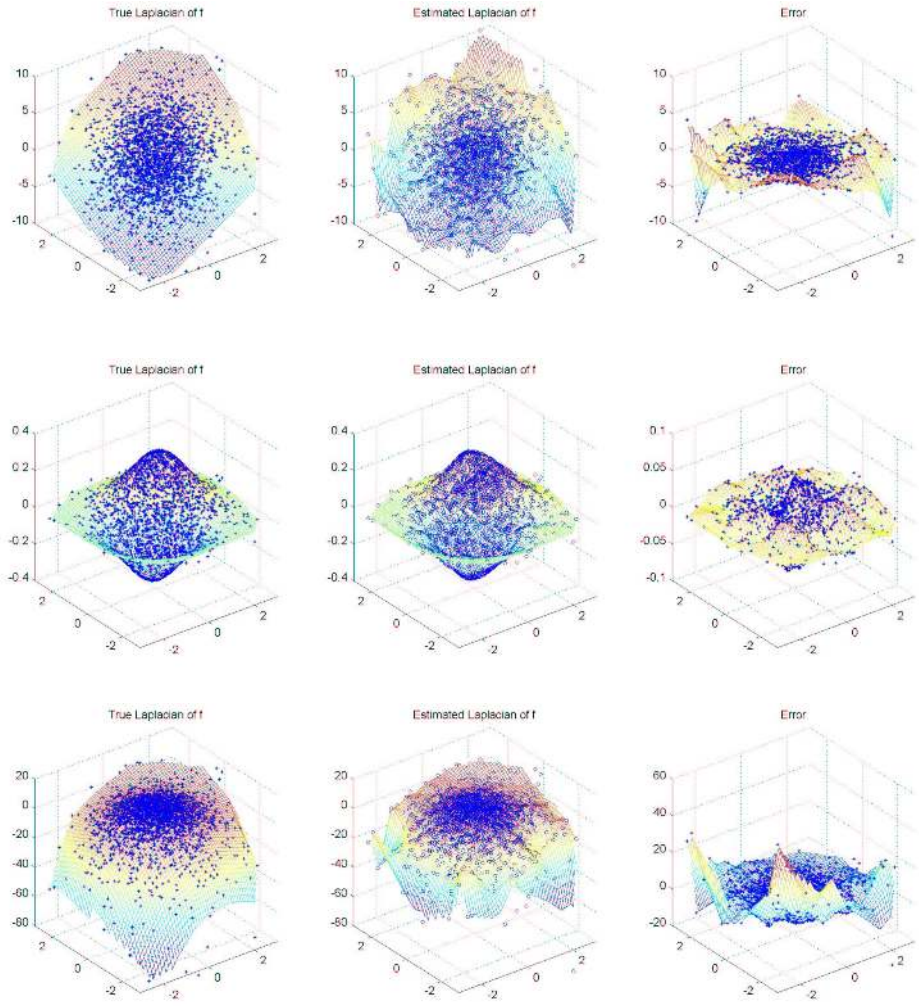

Figure 3: Illustration of the differences of the three graph Laplacians, random walk, unnormalized and normalized (from the top) for $\lambda = 0$. The function f is $f = \sum_{i=1}^{2} x_i - 4$ and the 2500 samples come from a standard Gaussian distribution on $\mathbb{R}^2$. The neighborhood size h is set to 1.2.

random walk graph Laplacian together with the result of the weighted Laplace-Beltrami operator and an error plot for $\lambda = 0, 1, 2$ resulting in $s = -2, 0, 2$ for the function $f(\phi, \theta) = \cos(\theta)$. First one can see that for a non-uniform probability measure the results for different values of λ differ quite a lot. Note that the function f is adapted to the cluster structure in the sense that it does not change much in each cluster but changes very much in region of low density. In the case of $s = 2$ we can see that $\Delta_s f$ would lead to a diffusion which would lead roughly to a kind of step function which changes at the equator. The same is true for $s = 0$ but the effect is much smaller than for $s = 2$. In the case of $s = -2$ we have a completely different behavior. $\Delta_s f$ has now flipped its sign near to the equator so that the induced diffusion process would try to smooth the function in the low density region.

5. Proof of the Main Result

In this section we will present the main results which were sketched in Section 3.3 together with the proofs. In Section 5.1 we first introduce some non-standard tools from differential geometry which we will use later on. In particular, it turns out that the so called manifolds with boundary of bounded geometry are the natural framework where one can still deal with non-compact manifolds in a setting comparable to the compact case. After a proper statement of the assumptions under which we prove the convergence results of the graph Laplacian and a preliminary result about convolutions on submanifolds which is of interest on its own, we then start with the final proofs. The proof is basically divided into two parts, the bias and the variance, where these terms are only approximately valid. The reader not familiar with differential geometry is encouraged to first read the appendix on basics of differential geometry in order to be equipped with the necessary background.

5.1 Non-compact Submanifolds in $\mathbb{R}^d$ with Boundary

We prove the pointwise convergence for non-compact submanifolds. Therefore we have to restrict the class of submanifolds since manifolds with unbounded curvature do not allow reasonable function spaces.

Remark 10 *In the rest of this paper we use the Einstein summation convention that is over indices occurring twice has to be summed. Note that the definition of the curvature tensor differs between textbooks. We use here the conventions regarding the definitions of curvature etc. of Lee (1997).*

5.1.1 MANIFOLDS WITH BOUNDARY OF BOUNDED GEOMETRY

We will consider in general non-compact submanifolds with boundary. In textbooks on Riemannian geometry one usually only finds material for the case where the manifold has no boundary. Also the analysis, for example the definition of Sobolev spaces on non-compact Riemannian manifolds, seems to be non-standard. We profit here very much from the thesis and an accompanying paper of Schick (1996, 2001) which introduces manifolds with boundary of bounded geometry. All material of this section is taken from these articles. Naturally this plus of generality leads also to a slightly larger technical overload. Nevertheless we think that it is worth this effort since the class of manifolds with boundary of bounded geometry includes almost any kind of submanifold one could have in mind. Moreover, to our knowledge, it is the most general setting where one can still introduce a notion of Sobolev spaces with the usual properties.

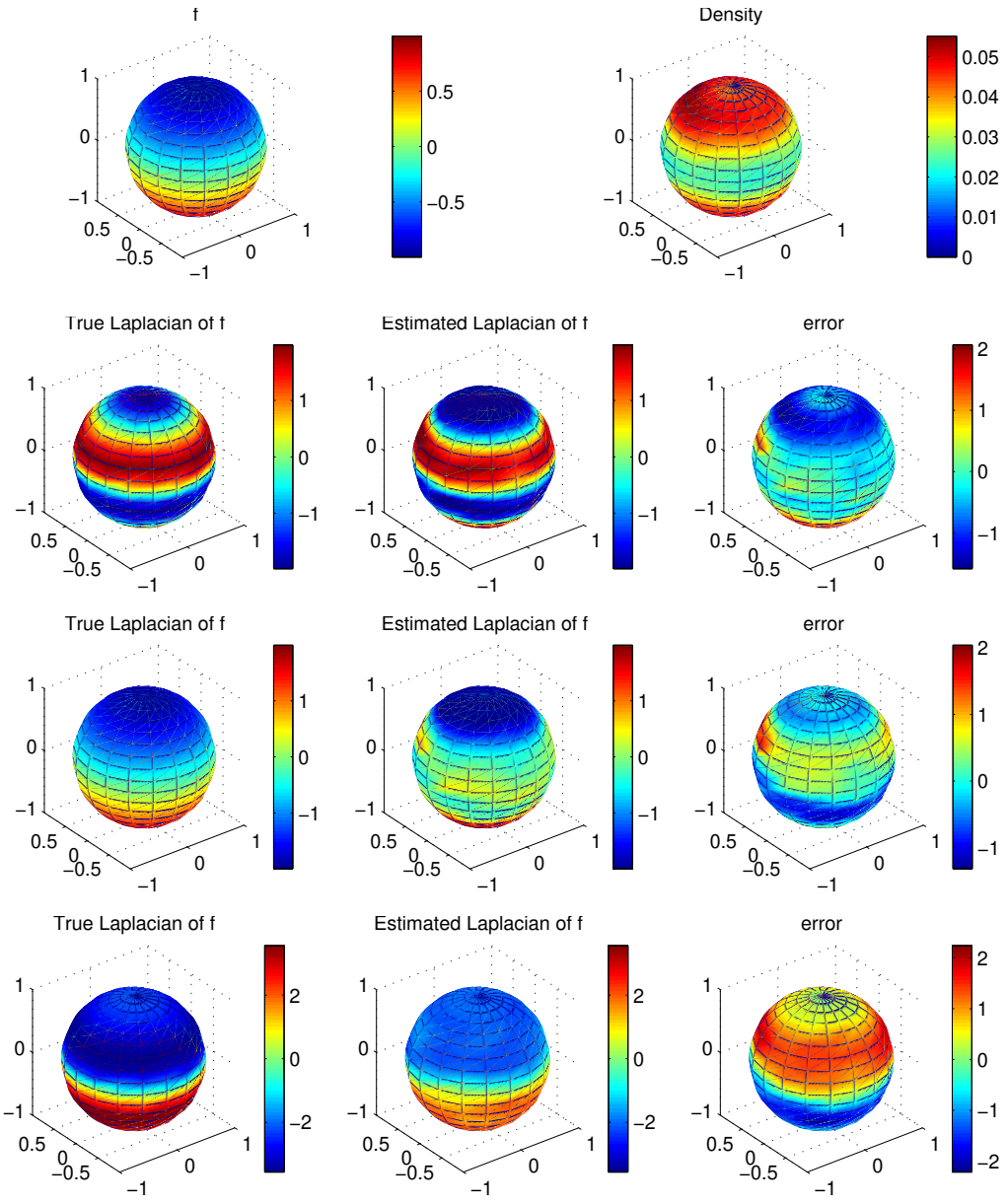

Figure 4: Illustration of the effect of $\lambda = 0, 1, 2$ (row $2-4$) resulting in $s = -2, 0, 2$ for the sphere with a non-uniform data-generating probability measure and the function $f(\theta, \phi) = \cos(\theta)$ (row 1) for the random walk Laplacian with $n = 2500$ and $h = 0.6$

Note that the boundary ∂M is an isometric submanifold of M of dimension $m-1$. Therefore it has a second fundamental form $\overline{\Pi}$ which should not be mixed up with the second fundamental form Π of M which is with respect to the ambient space $\mathbb{R}^d$. We denote by $\overline{\nabla}$ the connection and by $\overline{R}$ the curvature of ∂M. Moreover, let ν be the normal inward vector field at ∂M.

Definition 11 (Manifold with boundary of bounded geometry) *Let M be a manifold with boundary ∂M (possibly empty). It is of bounded geometry if the following holds:*

- *(N) Normal Collar: there exists $r_C > 0$ so that the normal geodesic flow,*

$$K : (x,t) \to \exp_x(t\nu_x),$$

  is defined on $\partial M \times [0, r_C)$ and is a diffeomorphism onto its image (ν_x is the inward normal vector). Let $N(s) := K(\partial M \times [0,s])$ be the collar set for $0 \leq s \leq r_C$.

- *(IC) The injectivity radius $\mathrm{inj}_{\partial M}$ of ∂M is positive.*

- *(I) Injectivity radius of M: There is $r_i > 0$ so that if $r \leq r_i$ then for $x \in M \backslash N(r)$ the exponential map is a diffeomorphism on $B_M(0,r) \subset T_xM$ so that normal coordinates are defined on every ball $B_M(x,r)$ for $x \in M \backslash N(r)$.*

- *(B) Curvature bounds: For every $k \in \mathbb{N}$ there is C_k so that $|\nabla^i R| \leq C_k$ and $\overline{\nabla}^i \overline{\Pi} \leq C_k$ for $0 \leq i \leq k$, where ∇^i denotes the covariant derivative of order i.*

Note that (B) imposes bounds on all orders of the derivatives of the curvatures. One could also restrict the definition to the order of derivatives needed for the goals one pursues. But this would require even more notational effort, therefore we skip this. In particular, in Schick (1996) it is argued that boundedness of all derivatives of the curvature is very close to the boundedness of the curvature alone.

The lower bound on the injectivity radius of M and the bound on the curvature are standard to define manifolds of bounded geometry without boundary. Now the problem of the injectivity radius of M is that at the boundary it somehow makes only partially sense since $\mathrm{inj}_M(x) \to 0$ as $d(x, \partial M) \to 0$. Therefore one replaces next to the boundary standard normal coordinates with normal collar coordinates.

Definition 12 (normal collar coordinates) *Let M be a Riemannian manifold with boundary ∂M. Fix $x' \in \partial M$ and an orthonormal basis of $T_{x'}\partial M$ to identify $T_{x'}\partial M$ with $\mathbb{R}^{m-1}$. For $r_1, r_2 > 0$ sufficiently small (such that the following map is injective) define normal collar coordinates,*

$$n_{x'} : B_{\mathbb{R}^{m-1}}(0,r_1) \times [0,r_2] \to M : (v,t) \to \exp^M_{\exp^{\partial M}_{x'}(v)}(t\nu).$$

The pair (r_1, r_2) is called the width of the normal collar chart $n_{x'}$.

The next proposition shows why manifolds of bounded geometry are especially interesting.

Proposition 13 (Schick 2001) *Assume that conditions $(N), (IC), (I)$ of Definition 11 hold.*

- *(B1) There exist* $0 < R_1 \leq r_{\text{inj}}(\partial M)$, $0 < R_2 \leq r_C$ *and* $0 < R_3 \leq r_i$ *and constants* $C_K > 0$ *for each* $K \in \mathbb{N}$ *such that whenever we have normal boundary coordinates of width* (r_1, r_2) *with* $r_1 \leq R_1$ *and* $r_2 \leq R_2$ *or normal coordinates of radius* $r_3 \leq r_i$ *then in these coordinates,*

$$|D^\alpha g_{ij}| \leq C_K \quad \text{and} \quad |D^\alpha g^{ij}| \leq C_K \quad \text{for all} \quad |\alpha| \leq K.$$

The condition (B) *in Definition 11 holds if and only if* $(B1)$ *holds. The constants* C_K *can be chosen to depend only on* $r_i, r_C, \text{inj}_{\partial M}$ *and* C_k.

Note that due to $g^{ij}g_{jk} = \delta^i_k$ one gets upper and lower bounds on the operator norms of g and g^{-1}, respectively, which result in upper and lower bounds for $\sqrt{\det g}$. This implies that we have upper and lower bounds on the volume form $dV(x) = \sqrt{\det g}\, dx$.

Lemma 14 (Schick 2001) *Let* (M, g) *be a Riemannian manifold with boundary of bounded geometry of dimension m. Then there exists* $R_0 > 0$ *and constants* $S_1 > 0$ *and* S_2 *such that for all* $x \in M$ *and* $r \leq R_0$ *one has*

$$S_1 r^m \leq \text{vol}(B_M(x, r)) \leq S_2 r^m.$$

Another important tool for analysis on manifolds are appropriate function spaces. In order to define a Sobolev norm one first has to fix a family of charts U_i with $M \subset \cup_i U_i$ and then define the Sobolev norm with respect to these charts. The resulting norm will depend on the choice of the charts U_i. Since in differential geometry the choice of the charts should not matter, the natural question arises how the Sobolev norm corresponding to a different choice of charts V_i is related to that for the choice U_i. In general, the Sobolev norms will not be the same. However, if one assumes that the transition maps are smooth and the manifold M is compact then the resulting norms will be equivalent and therefore define the same topology. Now if one has a non-compact manifold this argumentation does not work anymore. This problem is solved in general by defining the norm with respect to a covering of M by normal coordinate charts. Then it can be shown that the change of coordinates between these normal coordinate charts is well-behaved due to the bounded geometry of M. In that way it is possible to establish a well-defined notion of Sobolev spaces on manifolds with boundary of bounded geometry in the sense that any norm defined with respect to a different covering of M by normal coordinate charts is equivalent. Let $(U_i, \phi_i)_{i \in I}$ be a countable covering of the submanifold M with normal coordinate charts of M, that is $M \subset \cup_{i \in I} U_i$, then:

$$\|f\|_{C^k(M)} = \max_{m \leq k} \sup_{i \in I} \sup_{x \in \phi_i(U_i)} \left| D^m (f \circ \phi_i^{-1})(x) \right|.$$

In the following we will denote with $C^k(M)$ the space of C^k-functions on M together with the norm $\|\cdot\|_{C^k(M)}$.

5.1.2 INTRINSIC VERSUS EXTRINSIC PROPERTIES

Most of the proofs for the continuous part will work with Taylor expansions in normal coordinates. It is then of special interest to have a connection between intrinsic and extrinsic distances. Since the distance on M is induced from $\mathbb{R}^d$, it is obvious that one has $\|x - y\|_{\mathbb{R}^d} \sim d_M(x, y)$ for all $x, y \in M$ which are sufficiently close. The next proposition proven by Smolyanov, von Weizsäcker, and Wittich (2000) provides an asymptotic expression of geometric quantities of the submanifold M in

the neighborhood of a point $x \in M$. Particularly, it gives a third-order approximation of the intrinsic distance $d_M(x,y)$ in M in terms of the extrinsic distance in the ambient space X which is in our case just the Euclidean distance in $\mathbb{R}^d$.

Proposition 15 *Let $i : M \to \mathbb{R}^d$ be an isometric embedding of the smooth m-dimensional Riemannian manifold M into $\mathbb{R}^d$. Let $x \in M$ and V be a neighborhood of 0 in $\mathbb{R}^m$ and let $\Psi : V \to U$ provide normal coordinates of a neighborhood U of x, that is $\Psi(0) = x$. Then for all $y \in V$:*

$$\|y\|^2_{\mathbb{R}^m} = d^2_M(x, \Psi(y)) = \|(i \circ \Psi)(y) - i(x)\|^2 + \frac{1}{12} \|\Pi(\dot\gamma, \dot\gamma)\|^2_{T_x\mathbb{R}^d} + O(\|y\|^5_{\mathbb{R}^m}),$$

where Π is the second fundamental form of M and γ the unique geodesic from x to $\Psi(y)$ such that $\dot\gamma = y^i \partial_{y^i}$. The volume form $dV = \sqrt{\det g_{ij}(y)}\, dy$ of M satisfies in normal coordinates,

$$dV = \Big(1 + \frac{1}{6} R_{iuvi} y^u y^v + O(\|y\|^3_{\mathbb{R}^m})\Big) dy,$$

In particular,

$$(\Delta \sqrt{\det g_{ij}})(0) = -\frac{1}{3} R,$$

where R is the scalar curvature (i.e., $R = g^{ik} g^{jl} R_{ijkl}$).

We would like to note that in Smolyanov, von Weizsäcker, and Wittich (2007) this proposition was formulated for general ambient spaces X, that is arbitrary Riemannian manifolds X. Using the more general form of this proposition one could extend the results in this paper to submanifolds of other ambient spaces X. However, in order to use the scheme one needs to know the geodesic distances in X, which are usually not available for general Riemannian manifolds. Nevertheless, for some special cases like the sphere, one knows the geodesic distances. Submanifolds of the sphere could be of interest, for example in geophysics or astronomy.

The previous proposition is very helpful since it gives an asymptotic expression of the geodesic distance $d_M(x,y)$ on M in terms of the extrinsic Euclidean distance. The following lemma is a non-asymptotic statement taken from Bernstein et al. (2001) which we present in a slightly different form. But first we establish a connection between what they call the 'minimum radius of curvature' and upper bounds on the extrinsic curvatures of M and ∂M. Let

$$\Pi_{\max} = \sup_{x \in M} \sup_{v \in T_x M, \|v\| = 1} \|\Pi(v,v)\|, \qquad \overline{\Pi}_{\max} = \sup_{x \in \partial M} \sup_{v \in T_x \partial M, \|v\| = 1} \left\|\overline{\Pi}(v,v)\right\|,$$

where $\overline{\Pi}$ is the second fundamental form of ∂M as a submanifold of M. We set $\overline{\Pi}_{\max} = 0$ if the boundary ∂M is empty.

Using the relation between the acceleration in the ambient space and the second fundamental form for unit-speed curves γ with no acceleration in M ($D_t \dot\gamma = 0$) established in Section A.3, we get for the Euclidean acceleration of such a curve γ in $\mathbb{R}^d$,

$$\|\ddot\gamma\| = \|\Pi(\dot\gamma, \dot\gamma)\|.$$

Now if one has a non-empty boundary ∂M it can happen that a length-minimizing curve goes (partially) along the boundary (imagine $\mathbb{R}^d$ with a ball at the origin cut out). Then the segment c along the boundary will be a geodesic of the submanifold ∂M, see Alexander and Alexander (1981), that

is $\overline{D}_t\dot{c} = \overline{\nabla}_{\dot{c}}\dot{c} = 0$ where $\overline{\nabla}$ is the connection of ∂M induced by M. However, c will not be a geodesic in M (in the sense of a curve with no acceleration) since by the Gauss-Formula in Theorem 41,

$$D_t\dot{c} = \overline{D}_t\dot{c} + \overline{\Pi}(\dot{c},\dot{c}) = \overline{\Pi}(\dot{c},\dot{c}).$$

Therefore, in general the upper bound on the Euclidean acceleration of a length-minimizing curve γ in M is given by,

$$\|\ddot{\gamma}\| = \left\|\overline{\Pi}(\dot{\gamma},\dot{\gamma}) + \Pi(\dot{\gamma},\dot{\gamma})\right\| \leq \overline{\Pi}_{\max} + \Pi_{\max}.$$

Using this inequality, one can derive a lower bound on the 'minimum radius of curvature' ρ defined in Bernstein et al. (2001) as $\rho = \inf\{1/\|\ddot{\gamma}\|_{\mathbb{R}^d}\}$ where the infimum is taken over all unit-speed geodesics γ of M (in the sense of length-minimizing curves):

$$\rho \geq \frac{1}{\overline{\Pi}_{\max} + \Pi_{\max}}.$$

Finally we can formulate the Lemma from Bernstein et al. (2001).

Lemma 16 *Let $x, y \in M$ with $d_M(x,y) \leq \pi\rho$. Then*

$$2\rho\sin(d_M(x,y)/(2\rho)) \leq \|x-y\|_{\mathbb{R}^d} \leq d_M(x,y).$$

Noting that $\sin(x) \geq x/2$ for $0 \leq x \leq \pi/2$, we get as an easier to handle corollary:

Corollary 17 *Let $x, y \in M$ with $d_M(x,y) \leq \pi\rho$. Then*

$$\frac{1}{2}d_M(x,y) \leq \|x-y\|_{\mathbb{R}^d} \leq d_M(x,y).$$

In the given form this corollary is quite useless since we only have the Euclidean distances between points and therefore we have no possibility to check the condition $d_M(x,y) \leq \pi\rho$. In general small Euclidean distance does not imply small intrinsic distance. Imagine a circle where one has cut out a very small segment. Then the Euclidean distance between the two ends is very small however the geodesic distance is very large. We show now that under an additional assumption one can transform the above corollary so that one can use it when one has only knowledge about Euclidean distances.

Lemma 18 *Let M have a finite radius of curvature $\rho > 0$. We further assume that,*

$$\kappa := \inf_{x\in M}\inf_{y\in M\setminus B_M(x,\pi\rho)} \|x-y\|,$$

is non-zero. Then $B_{\mathbb{R}^d}(x,\kappa/2)\cap M \subset B_M(x,\kappa) \subset B_M(x,\pi\rho)$. Particularly, if $x, y \in M$ and $\|x-y\| \leq \kappa/2$, then

$$\frac{1}{2}d_M(x,y) \leq \|x-y\|_{\mathbb{R}^d} \leq d_M(x,y) \leq \kappa.$$

Proof By definition κ is at most the infimum of $\|x-y\|$ where y satisfies $d_M(x,y) = \pi\rho$. Therefore the set $B_{\mathbb{R}^d}(x,\kappa/2)\cap M$ is a subset of $B_M(x,\pi\rho)$. The rest of the lemma then follows by Corollary 17. Figure 5 illustrates this construction. ■

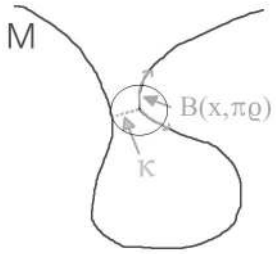

Figure 5: κ is the Euclidean distance of $x \in M$ to $M \backslash B_M(x, \pi\rho)$.

5.2 Notations and Assumptions

In general we work on complete non-compact manifolds with boundary. Compared to a setting where one considers only compact manifolds one needs a slightly larger technical overhead. However, we will indicate how the technical assumptions simplify if one has a compact submanifold with boundary or even a compact manifold without boundary.

We impose the following assumptions on the manifold M:

Assumption 19 *(i) The map $i : M \to \mathbb{R}^d$ is a smooth embedding,*

(ii) The manifold M with the metric induced from $\mathbb{R}^d$ is a smooth manifold with boundary of bounded geometry (possibly $\partial M = \emptyset$),

(iii) M has bounded second fundamental form,

(iv) It holds $\kappa := \inf_{x \in M} \inf_{y \in M \backslash B_M(x, \pi\rho)} \|i(x) - i(y)\| > 0$, where ρ is the radius of curvature defined in Section 5.1.2,

(v) For any $x \in M \backslash \partial M$, $\delta(x) := \inf_{y \in M \backslash B_M(x, \frac{1}{3}\min\{\mathrm{inj}(x), \pi\rho\})} \|i(x) - i(y)\|_{\mathbb{R}^d} > 0$, where $\mathrm{inj}(x)$ is the injectivity radius[1] at x and $\rho > 0$ is the radius of curvature.

The first condition ensures that $i(M)$ is a smooth submanifold of $\mathbb{R}^d$. Usually we do not distinguish between $i(M)$ and M. The use of the abstract manifold M as a starting point emphasizes that there exists an m-dimensional smooth manifold M or roughly equivalent an m-dimensional smooth parameter space underlying the data. The choice of the d features determines then the representation in $\mathbb{R}^d$. The choice of features corresponds therefore to a specific choice of the inclusion map i since i determines how M is embedded into $\mathbb{R}^d$. This means that another choice of features leads in general to a different mapping i but the initial abstract manifold M is always the same. However, in the second condition we assume that the metric structure of M is induced by $\mathbb{R}^d$ (which implies that i is trivially an isometric embedding). Therefore the metric structure depends on the embedding i or equivalently on our choice of features.

The second condition ensures that M is an isometric submanifold of $\mathbb{R}^d$ which is well-behaved. As discussed in Section 5.1.1, manifolds of bounded geometry are in general non-compact, complete Riemannian manifolds with boundary where one has uniform control over all intrinsic curvatures. The uniform bounds on the curvature allow to do reasonable analysis in this general setting. In

1. Note that the injectivity radius $\mathrm{inj}(x)$ is always positive.

particular, it allows us to introduce the function spaces $C^k(M)$ with their associated norm. It might be possible to prove pointwise results even without the assumption of bounded geometry. But we think that the setting studied here is already general enough to encompass all cases encountered in practice.

The third condition ensures that M also has well-behaved extrinsic geometry and implies that the radius of curvature ρ is lower bounded. Together with the fourth condition it enables us to get global upper and lower bounds of the intrinsic distance on M in terms of the extrinsic distance in $\mathbb{R}^d$ and vice versa, see Lemma 18.

The fourth condition is only necessary in the case of non-compact submanifolds. It prevents the manifold from self-approaching. More precisely it ensures that if parts of M are far away from x in the geometry of M they do not come too close to x in the geometry of $\mathbb{R}^d$. Assuming that $i(M)$ is a submanifold, this assumption is already included implicitly. However, for non-compact submanifolds the self-approaching could happen at infinity. Therefore we exclude it explicitly. Moreover, note that for submanifolds with boundary one has $\mathrm{inj}(x) \to 0$ as x approaches the boundary[2] ∂M. Therefore also $\delta(x) \to 0$ as $d(x, \partial M) \to 0$. However, this behavior of $\delta(x)$ at the boundary does not matter for the proof of pointwise convergence in the interior of M.

Note that if M is a smooth and *compact* manifold conditions (ii)-(v) hold automatically.

In order to emphasize the distinction between extrinsic and intrinsic properties of the manifold we always use the slightly cumbersome notations $x \in M$ (intrinsic) and $i(x) \in \mathbb{R}^d$ (extrinsic). The reader who is not familiar with Riemannian geometry should keep in mind that locally, a submanifold of dimension m looks like $\mathbb{R}^m$. This becomes apparent if one uses normal coordinates. Also the following dictionary between terms of the manifold M and the case when one has only an open set in $\mathbb{R}^d$ (i is then the identity mapping) might be useful.

Manifold M	open set in $\mathbb{R}^d$
g_{ij} , $\sqrt{\det g}$	δ_{ij} , 1
natural volume element	Lebesgue measure
Δ_s	$\Delta_s = \sum_{i=1}^d \frac{\partial^2}{\partial (z_i)^2} + \frac{s}{p} \sum_{i=1}^d \frac{\partial p}{\partial z^i} \frac{\partial}{\partial z^i}$

The kernel functions which are used to define the weights of the graph are always functions of the squared norm in $\mathbb{R}^d$. Furthermore, we make the following assumptions on the kernel function k:

Assumption 20 *(i)* $k : \mathbb{R}_+ \to \mathbb{R}$ *is measurable, non-negative and non-increasing on* $\mathbb{R}_+^*$,

(ii) $k \in C^2(\mathbb{R}_+^*)$, *that is in particular* k, $\frac{\partial k}{\partial x}$ *and* $\frac{\partial^2 k}{\partial x^2}$ *are bounded,*

(iii) k, $|\frac{\partial k}{\partial x}|$ *and* $|\frac{\partial^2 k}{\partial x^2}|$ *have exponential decay:* $\exists c, \alpha, A \in \mathbb{R}_+$ *such that for any* $t \geq A$, $f(t) \leq ce^{-\alpha t}$, *where* $f(t) = \max\{k(t), |\frac{\partial k}{\partial x}|(t), |\frac{\partial^2 k}{\partial x^2}|(t)\}$,

(iv) $k(0) = 0$.

The assumption that the kernel is non-increasing could be dropped, however it makes the proof and the presentation easier. Moreover, in practice the weights of the neighborhood graph which are determined by k are interpreted as similarities. Therefore the usual choice is to take weights which

2. This is the reason why one replaces normal coordinates in the neighborhood of the boundary with normal collar coordinates.

decrease with increasing distance. The fourth condition implies that the graph has no loops.[3] In particular, the kernel is not continuous at the origin. All results hold also without this condition. The advantage of this condition is that some estimators become unbiased. Also let us introduce the helpful notation, $k_h(t) = \frac{1}{h^m} k\left(\frac{t}{h^2}\right)$ where we call h the bandwidth of the kernel. Moreover, we define the following two constants related to the kernel function k,

$$C_1 = \int_{\mathbb{R}^m} k(\|y\|^2) dy < \infty, \quad C_2 = \int_{\mathbb{R}^m} k(\|y\|^2) y_1^2 dy < \infty.$$

We also have some assumptions on the probability measure P.

Assumption 21 *(i) P is absolutely continuous with respect to the natural volume element dV on M,*

(ii) the density p fulfills: $p \in C^3(M)$ and $p(x) > 0, \forall\, x \in M \backslash \partial M$,

(iii) the sample X_i, $i = 1, \ldots, n$ is drawn i.i.d. from P,

Note that condition (i) implies $P(\partial M) = 0$, that is the boundary ∂M is a set of measure zero. We will call the Assumptions 19 on the submanifold, Assumptions 20 on the kernel function, and Assumptions 21 on the probability measure P together the **standard assumptions**.

In the following table we summarize the notation used in the proofs:

$k : \mathbb{R}_+ \to \mathbb{R}_+$	kernel function
$h > 0$	neighborhood/bandwidth parameter
$m \in \mathbb{N}$	dimension of the submanifold M
$k_h(t) = \frac{1}{h^m} k\left(\frac{t}{h^2}\right)$	scaled kernel function
$\lambda \in \mathbb{R}$	reweighting parameter
$d_{h,n}(x) = \frac{1}{n} \sum_{i=1}^n k_h(\|x - X_i\|^2)$	degree function associated with k
$\tilde{k}_{\lambda,h}(x, X_i) = \frac{k_h(\|x-X_i\|^2)}{[d_{h,n}(X_i) d_{h,n}(X_j)]^\lambda}$	reweighted kernel
$\tilde{d}_{\lambda,h,n}(x) = \frac{1}{n} \sum_{i=1}^n \tilde{k}_{\lambda,h}(x, X_i)$	degree function associated with $\tilde{k}_{\lambda,h}$
$(\tilde{A}_{\lambda,h,n} f)(x) = \frac{1}{n} \sum_{i=1}^n \tilde{k}_{\lambda,h}(x, X_i) f(X_i)$	empirical average operator $\tilde{A}_{\lambda,h,n}$
$\Delta^{(\mathrm{rw})}_{\lambda,h,n} f = \frac{1}{h^2}\left(f - \frac{1}{\tilde{d}_{\lambda,h,n}} \tilde{A}_{\lambda,h,n} f\right)$	random walk graph Laplacian
$\Delta^{(\mathrm{u})}_{\lambda,h,n} f = \frac{1}{h^2}\left(\tilde{d}_{\lambda,h,n} f - \tilde{A}_{\lambda,h,n} f\right)$	unnormalized graph Laplacian
$\Delta^{(\mathrm{n})}_{\lambda,h,n} f = \frac{1}{h^2}\left(f - \frac{1}{\sqrt{\tilde{d}_{\lambda,h,n}}} \tilde{A}_{\lambda,h,n}\left(\frac{f}{\sqrt{\tilde{d}_{\lambda,h,n}}}\right)\right)$	normalized graph Laplacian
$C_1 = \int_{\mathbb{R}^m} k(\|y\|^2_{\mathbb{R}^m}) dy$, $C_2 = \int_{\mathbb{R}^m} k(\|y\|^2_{\mathbb{R}^m}) y_1^2 dy$	characteristic constants of the kernel
$p_h(x) = \mathbb{E}_Z k_h(\|x - Z\|^2)$	convolution of p with k_h
$(\tilde{A}_{\lambda,h} f)(x) = \mathbb{E}_Z \tilde{k}_{\lambda,h}(x, Z) f(Z)$	average operator $\tilde{A}_{\lambda,h}$
$\Delta^{(\mathrm{rw})}_{\lambda,h}$, $\Delta^{(\mathrm{u})}_{\lambda,h}$, $\Delta^{(\mathrm{n})}_{\lambda,h}$	Laplacians associated with $\tilde{A}_{\lambda,h}$
$\Delta_s = \frac{1}{p^s} \operatorname{div}(p^s \operatorname{grad}) = \frac{1}{p^s} g^{ab} \nabla_a (p^s \nabla_b)$	s-th weighted Laplacian on M

3. An edge from a vertex to itself is called a loop.

5.3 Asymptotics of Euclidean Convolutions on the Submanifold M

The following proposition describes the asymptotic expression of the convolution of a function f on the submanifold M with a kernel function having the Euclidean distance $\|x-y\|$ as its argument with respect to the probability measure P on M. This result is interesting since it shows how the use of the Euclidean distance introduces a curvature effect if one averages a function locally. A similar result has been presented in Coifman and Lafon (2006). We define the density p invariantly with respect to the natural volume element and also explicitly give the second order curvature terms. Our proof is similar to that of Smolyanov, von Weizsäcker, and Wittich (2007) where under stronger conditions a similar result was proven for the Gaussian kernel. The more general setting and the use of general kernel functions make the proof slightly more complicated. In order to emphasize the distinction between extrinsic and intrinsic properties of the manifold we will use the slightly cumbersome notations $x \in M$ (intrinsic) and $i(x) \in \mathbb{R}^d$ (extrinsic).

Proposition 22 *Let M and k satisfy Assumptions 19 and 20. Furthermore, let P have a density p with respect to the natural volume element and $p \in C^3(M)$. Then, for any $x \in M \backslash \partial M$, there exists an $h_0(x) > 0$ such that for all $h < h_0(x)$ and any $f \in C^3(M)$,*

$$\begin{aligned}&\int_M k_h\big(\|i(x)-i(y)\|^2_{\mathbb{R}^d}\big) f(y)p(y)\sqrt{\det g}\,dy \\ =&C_1 p(x)f(x) + \frac{h^2}{2} C_2 \Big(p(x)f(x)S(x) + (\Delta_M(pf))(x)\Big) + O(h^3),\end{aligned}$$

where $O(h^3)$ is a function depending on x, $\|f\|_{C^3(M)}$ and $\|p\|_{C^3(M)}$ and

$$S(x) = \frac{1}{2}\Big[-R\big|_x + \frac{1}{2}\big\|\textstyle\sum_a \Pi(\partial_a,\partial_a)\big\|^2_{T_{i(x)}\mathbb{R}^d}\Big],$$

where R is the scalar curvature and Π the second fundamental form of M .

The following Lemma is an application of Bernstein's inequality. Together with the previous proposition it will be the main ingredient for proving consistency statements for the graph structure.

Lemma 23 *Suppose the standard assumptions hold and let the kernel k have compact support on $[0,R_k^2]$. Define $b_1 = \|k\|_\infty \|f\|_\infty$, $b_2 = K\|f\|^2_\infty$ where K is a constant depending on $\|p\|_\infty$, $\|k\|_\infty$ and R_k. Let $x \in M\backslash\partial M$ and $V_i := k_h(\|i(x)-i(X_i)\|^2) f(X_i)$. Then for any bounded function f,*

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n V_i - \mathbb{E}V\Big| > \varepsilon\Big) \le 2\exp\Big(-\frac{nh^m\varepsilon^2}{2b_2 + 2b_1\varepsilon/3}\Big).$$

Let $W_i = k_h(\|i(x)-i(X_i)\|^2)(f(x)-f(X_i))$. Then for $hR_k \le \kappa/2$ and $f \in C^1(M)$,

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n W_i - \mathbb{E}W\Big| > h\varepsilon\Big) \le 2\exp\Big(-\frac{nh^m\varepsilon^2}{2b_2 + 2b_1\varepsilon/3}\Big).$$

Proof Since by assumption $\kappa > 0$, by Lemma 18, for any $x,y \in M$ with $\|i(x)-i(y)\| \le \kappa/2$, we have $d_M(x,y) \le 2\|i(x)-i(y)\|$. This implies $\forall a \le \kappa/2$, $B_{\mathbb{R}^d}(x,a) \cap M \subset B_M(x,2a)$.

Let $W_i := k_h(\|i(x) - i(X_i)\|^2) f(X_i)$. We have

$$|W_i| \leq \frac{\|k\|_\infty}{h^m} \sup_{y \in B_{\mathbb{R}^d}(x, hR_k) \cap M} |f(y)| \leq \frac{\|k\|_\infty}{h^m} \|f\|_\infty := \frac{b_1}{h^m}.$$

For the variance of W we have two cases. First let $hR_k < s := \min\{\kappa/2, R_0/2\}$. Then we get,

$$\operatorname{Var} W \leq \mathbb{E}_Z k_h^2(\|i(x) - i(Z)\|^2) f^2(Z) \leq \frac{\|k\|_\infty}{h^m} \|f\|_\infty^2 p_h(x) \leq D_2 \frac{\|k\|_\infty}{h^m} \|f\|_\infty^2,$$

where we have used Lemma 46 in the last step. Now consider $hR_k \geq s$, then

$$\operatorname{Var} W \leq \frac{\|k\|_\infty^2}{h^{2m}} \|f\|_\infty^2 \leq \frac{R_k^m \|k\|_\infty^2}{s^m h^m} \|f\|_\infty^2.$$

Therefore we define $b_2 = K \|f\|_\infty^2$ with $K = R_k^m \|k\|_\infty^2 \max\{2^m S_2 \|p\|_\infty, s^{-m}\}$. By Bernstein's inequality we finally get

$$\mathrm{P}\Big(\big| \tfrac{1}{n} \textstyle\sum_{i=1}^n W_i - \mathbb{E} W \big| > \varepsilon \Big) \quad \leq \quad 2e^{-\frac{nh^m \varepsilon^2}{2b_2 + 2b_1 \varepsilon/3}}.$$

Both constants b_2 and b_1 are independent of x. For the second part note that by Lemma 18 for $hR_k \leq \kappa/2$, we have that $\|x-y\| \leq hR_k$ implies $d_M(x,y) \leq 2\|x-y\| \leq 2hR_k$. In particular, for all $x, y \in M$ with $\|x-y\| \leq hR_k$,

$$|f(x) - f(y)| \leq \sup_{y \in M} \|\nabla f\|_{T_y M} d_M(x,y) \leq 2hR_k \sup_{y \in M} \|\nabla f\|_{T_y M}.$$

A similar reasoning as above leads then to the second statement. ■

Note that $\mathbb{E}_Z k_h(\|i(x) - i(Z)\|^2) f(Z) = \int_M k_h(\|i(x) - i(y)\|^2) f(y) p(y) \sqrt{\det g}\, dy$.

5.4 Pointwise Consistency of the Random Walk, Unnormalized and Normalized Graph Laplacian

The proof of the convergence result for the three graph Laplacians is organized as follows. First we introduce the continuous operators $\Delta_{\lambda,h}^{(\mathrm{rw})}$, $\Delta_{\lambda,h}^{(\mathrm{n})}$ and $\Delta_{\lambda,h}^{(\mathrm{u})}$. Then we derive the limit of the continuous operators as $h \to 0$. This part of the proof is concerned with the bias part since roughly $(\Delta_{\lambda,h} f)(x)$ can be seen as the expectation of $\Delta_{\lambda,h,n} f(x)$. Second we show that with high probability all extended graph Laplacians are close to the corresponding continuous operators. This is the variance part. Combining both results we arrive finally at the desired consistency results.

5.4.1 THE BIAS PART - DEVIATION OF $\Delta_{\lambda,h}$ FROM ITS LIMIT

The following continuous approximation of $\Delta_{\lambda,h}^{(\mathrm{rw})}$ was similarly introduced in Lafon (2004) and Coifman and Lafon (2006).

Definition 24 (Kernel-based approximation of the Laplacian) *We introduce the following averaging operator $\tilde{A}_{\lambda,h}$ based on the reweighted kernel $\tilde{k}_{\lambda,h}$,*

$$(\tilde{A}_{\lambda,h} f)(x) = \int_M \tilde{k}_{\lambda,h}(x,y) f(y) p(y) \sqrt{\det g}\, dy,$$

and with $\tilde{d}_{\lambda,h} = (\tilde{A}_{\lambda,h}1)$ *the following continuous operators,*

$$\begin{aligned}
\text{random walk}: \quad & \Delta^{(rw)}_{\lambda,h} f := \frac{1}{h^2}\left(f - \frac{1}{\tilde{d}_{\lambda,h}}\tilde{A}_{\lambda,h} f\right) = \frac{1}{h^2}\left(\frac{\tilde{A}_{\lambda,h} g}{\tilde{d}_{\lambda,h}}\right)(x), \\
\text{unnormalized}: \quad & \Delta^{(u)}_{\lambda,h} f := \frac{1}{h^2}\left(\tilde{d}_{\lambda,h} f - \tilde{A}_{\lambda,h} f\right) = \frac{1}{h^2}(\tilde{A}_{\lambda,h} g)(x), \\
\text{normalized}: \quad & \Delta^{(n)}_{\lambda,h} f := \frac{1}{h^2\sqrt{\tilde{d}_{\lambda,h}}}\left(d_{\lambda,h}\frac{f}{\sqrt{\tilde{d}_{\lambda,h}}} - \tilde{A}_{\lambda,h}\frac{f}{\sqrt{\tilde{d}_{\lambda,h}}}\right) = \tfrac{1}{h^2\sqrt{\tilde{d}_{\lambda,h}(x)}}(\tilde{A}_{\lambda,h} g')(x),
\end{aligned}$$

where we have introduced again $g(y) := f(x) - f(y)$ and $g'(y) := \frac{f(x)}{\sqrt{\tilde{d}_{\lambda,h}(x)}} - \frac{f(y)}{\sqrt{\tilde{d}_{\lambda,h}(y)}}$. The definition of the normalized approximation $\Delta^{(\mathrm{rw})}_{\lambda,h}$ can be justified by the alternative definition of the Laplacian in $\mathbb{R}^d$ sometimes made in physics textbooks:

$$(\Delta f)(x) = \lim_{r\to 0} -\frac{1}{C_d\, r^2}\Big(f(x) - \frac{1}{\mathrm{vol}(B(x,r))}\int_{B(x,r)} f(y)dy\Big),$$

where C_d is a constant depending on the dimension d.

Approximations of the Laplace-Beltrami operator based on averaging with the Gaussian kernel in the case of a uniform probability measure have been studied for compact submanifolds without boundary by Smolyanov, von Weizsäcker, and Wittich (2000, 2007) and Belkin (2003). Their result was then generalized by Lafon (2004) to general densities and to a wider class of isotropic, positive definite kernels for compact submanifolds with boundary. The proof given in Lafon (2004) applies only to compact hypersurfaces[4] in $\mathbb{R}^d$, a proof for the general case of compact submanifolds with boundary using boundary conditions has been presented in Coifman and Lafon (2006). In this section we will prove the pointwise convergence of the continuous approximation for general submanifolds M with boundary of bounded geometry with the additional Assumptions 19. This includes the case where M is not compact. Moreover, no assumptions of positive definiteness of the kernel are made nor any boundary condition on the function f is imposed. Almost any submanifold occurring in practice should be covered in this very general setting.

For pointwise convergence in the interior of the manifold M boundary conditions on f are not necessary. However, for uniform convergence there is no way around them. Then the problem lies not in the proof that the continuous approximation still converges in the right way but in the transfer of the boundary condition to the discrete graph. The main problem is that since we have no information about M apart from the random samples the boundary will be hard to locate. Moreover, since the boundary is a set of measure zero, we will actually almost surely never sample any point from the boundary. The rigorous treatment of the approximation of the boundary respectively the boundary conditions of a function on a randomly sampled graph remains as an open problem.

Especially for dimensionality reduction the case of low-dimensional submanifolds in $\mathbb{R}^d$ is important. Notably, the analysis below also includes the case where due to noise the data is only concentrated around a submanifold.

Theorem 25 *Suppose the standard assumptions hold. Furthermore, let k be a kernel with compact support on* $[0, R_k^2]$. *Let* $\lambda \in \mathbb{R}$, *and* $x \in M \backslash \partial M$. *Then there exists an* $h_1(x) > 0$ *such that for all*

4. A hypersurface is a submanifold of codimension 1.

$h < h_1(x)$ and any $f \in C^3(M)$,

$$(\Delta^{(rw)}_{\lambda,h} f)(x) = -\frac{C_2}{2C_1}\Big((\Delta_M f)(x) + \frac{s}{p(x)}\langle \nabla p, \nabla f\rangle_{T_xM}\Big) + O(h) = -\frac{C_2}{2C_1}(\Delta_s f)(x) + O(h),$$

where Δ_M is the Laplace-Beltrami operator of M and $s = 2(1-\lambda)$.

Proof For sufficiently small h we have $\overline{B_{\mathbb{R}^d}(x, 2hR_k) \cap M} \cap \partial M = \emptyset$. Moreover, it can be directly seen from the proof of Proposition 22 that the upper bound of the interval $[0, h_0(y)]$ for which the expansion holds depends continuously on $\delta(x)$ and $\varepsilon(y)$, where $\varepsilon(y) = \frac{1}{3}\min\{\pi\rho, \mathrm{inj}(y)\}$. Now $h_0(x)$ is continuous since $\mathrm{inj}(x)$ is continuous on compact subsets, see Klingenberg (1982)[Prop. 2.1.10], and $\delta(x)$ is continuous since the injectivity radius is continuous. Therefore we conclude that since $h_0(y)$ is continuous on $\overline{B(x, 2hR_k) \cap M}$ and $h_0(y) > 0$, $h_1(x) = \inf_{y \in \overline{B_{\mathbb{R}^d}(x, 2hR_k) \cap M}} h_0(y) > 0$. Then for the interval $(0, h_1(x))$ the expansion of $p_h(y)$ holds uniformly over the whole set $B(x, 2hR_k) \cap M$. That is, using the definition of $\tilde{k}$ as well as Proposition 22 and the expansion $\frac{1}{(a+h^2b)^\lambda} = \frac{1}{a^\lambda} - \lambda\frac{h^2 b}{a^{\lambda+1}} + O(h^4)$, we get for $h \in (0, h_1(x))$ that

$$\begin{aligned}
&\int_M \tilde{k}_{\lambda,h}\big(\|i(x) - i(y)\|^2\big) f(y) p(y) \sqrt{\det g}\, dy \\
&= \int\limits_{B_{\mathbb{R}^d}(x, hR_k) \cap M} \frac{k_h\big(\|i(x) - i(y)\|^2\big)}{p_h^\lambda(x)} f(y) \left[\frac{C_1 p(y) - \lambda/2 C_2 h^2 (p(y)S + \Delta p)}{C_1^{\lambda+1} p(y)^\lambda} + O(h^3)\right] \sqrt{\det g}\, dy,
\end{aligned}$$

where the $O(h^3)$-term is continuous on $B_{\mathbb{R}^d}(x, hR_k)$ and we have introduced the abbreviation $S = \frac{1}{2}[-R + \frac{1}{2}\|\sum_a \Pi(\partial_a, \partial_a)\|^2_{T_{i(x)}\mathbb{R}^d}]$. Using $f(y) = 1$ we get,

$$\tilde{d}_{\lambda,h}(x) = \int\limits_{B_{\mathbb{R}^d}(x, hR_k) \cap M} \frac{k_h\big(\|i(x) - i(y)\|^2\big)}{p_h^\lambda(x)} \left[\frac{C_1 p(y) - \lambda/2 C_2 h^2 (p(y)S + \Delta p)}{C_1^{\lambda+1} p(y)^\lambda} + O(h^3)\right] \sqrt{\det g}\, dy,$$

as an estimate for $\tilde{d}_{\lambda,h}(x)$. Now using Proposition 22 again, we arrive at:

$$\Delta^{(\mathrm{rw})}_{\lambda,h} f = \frac{1}{h^2}\frac{\tilde{d}_{\lambda,h} f - \tilde{A}_{\lambda,h} f}{\tilde{d}_{\lambda,h}} = -\frac{C_2}{2C_1}\Big(\Delta_M f + \frac{2(1-\lambda)}{p}\langle \nabla p, \nabla f\rangle\Big) + O(h),$$

where all $O(h)$-terms are finite on $B_{\mathbb{R}^d}(x, hR_k) \cap M$ since p is strictly positive. ■

Note that the limit of $\Delta^{(\mathrm{rw})}_{\lambda,h}$ has the opposite sign of Δ_s. This is due to the fact that the Laplace-Beltrami operator on manifolds is usually defined as a negative definite operator (in analogy to the Laplace operator in $\mathbb{R}^d$), whereas the graph Laplacian is positive definite. But this varies through the literature, thus the reader should be aware of the sign convention.

Remark 26 *The assumption of compact support of the kernel k is only necessary in the case of non-compact manifolds M. For compact manifolds a kernel with non-compact support, such as a Gaussian kernel, would work, too. The reason for compact support of the kernel comes from the fact that for non-compact manifolds there exists no lower bound on a strictly positive density. This in turn implies that one cannot upper bound the convolution with the reweighted kernel if one does not impose additional assumptions on the density. In practice the solution of graph-based methods for large-scale problems is usually only possible for sparse neighborhood graphs. Therefore the compactness assumption of the kernel is quite realistic and does not exclude relevant cases.*

With the relations,

$$(\Delta^{(\mathrm{u})}_{\lambda,h,n} f)(x) = \tilde{d}_{\lambda,h,n}(x)(\Delta^{(\mathrm{rw})}_{\lambda,h,n} f)(x),$$
$$(\Delta^{(\mathrm{n})}_{\lambda,h,n} f)(x) = 1/\sqrt{\tilde{d}_{\lambda,h,n}(x)}\,(\Delta^{(\mathrm{u})}_{\lambda,h,n}\Big(f/\sqrt{\tilde{d}_{\lambda,h,n}}\Big))(x),$$

one can easily adapt the last lines of the previous proof to derive the following corollary.

Corollary 27 *Under the assumptions of Theorem 25. Let $\lambda \in \mathbb{R}$ and $x \in M\backslash\partial M$. Then there exists an $h_1(x) > 0$ such that for all $h < h_1(x)$ and any $f \in C^3(M)$,*

$$(\Delta^{(u)}_{\lambda,h} f)(x) = -\frac{C_2}{2C_1^{2\lambda}}\, p(x)^{1-2\lambda}(\Delta_s f)(x) + O(h), \quad \text{where} \quad s = 2(1-\lambda),$$
$$(\Delta^{(n)}_{\lambda,h} f)(x) = -\frac{C_2}{2C_1}\, p(x)^{\frac{1}{2}-\lambda}\Delta_s\left(\frac{f}{p^{\frac{1}{2}-\lambda}}\right)(x) + O(h).$$

5.4.2 THE VARIANCE PART - DEVIATION OF $\Delta_{\lambda,h,n}$ FROM $\Delta_{\lambda,h}$

Before we state the results for the general case with data-dependent weights we now treat the case $\lambda = 0$, that is we have non-data-dependent weights. There the proof is considerably simpler and much easier to follow. Moreover, as opposed to the general case here we get convergence in probability under slightly weaker conditions than almost sure convergence. Since this does not hold for the normalized graph Laplacian in that case we will only provide the general proof.

Theorem 28 (Weak and strong pointwise consistency for $\lambda = 0$) *Suppose the standard assumptions hold. Furthermore, let k be a kernel with compact support on $[0, R_k^2]$. Let $x \in M\backslash\partial M$ and $f \in C^3(M)$. Then if $h \to 0$ and $nh^{m+2} \to \infty$,*

$$\lim_{n\to\infty} (\Delta^{(rw)}_{0,h,n} f)(x) = -\frac{C_2}{2C_1}\,(\Delta_2 f)(x) \qquad \textit{in probability},$$
$$\lim_{n\to\infty} (\Delta^{(u)}_{0,h,n} f)(x) = -\frac{C_2}{2}\, p(x)(\Delta_2 f)(x) \qquad \textit{in probability}.$$

If even $nh^{m+2}/\log n \to \infty$, then almost sure convergence holds.

Proof We give the proof for $\Delta^{(\mathrm{rw})}_{0,h,n}$. The proof for $\Delta^{(\mathrm{u})}_{0,h,n}$ can be directly derived with the second statement of Lemma 23 for the variance term together with Corollary 27 for the bias term. Similar to the proof for the Nadaraya-Watson regression estimate of Greblicki et al. (1984), we rewrite the estimator $\Delta^{(\mathrm{rw})}_{0,h,n} f$ in the following form

$$(\Delta^{(\mathrm{rw})}_{0,h,n} f)(x) = \frac{1}{h^2}\left[\frac{(C_{0,h} f)(x) + B_{1n}}{1 + B_{2n}}\right],$$

where

$$(C_{0,h} f)(x) = \frac{\mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2) g(Z)}{\mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2)},$$
$$B_{1n} = \frac{\frac{1}{n}\sum_{j=1}^n k_h(\left\|i(x) - i(X_j)\right\|^2) g(X_j) - \mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2) g(Z)}{\mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2)},$$
$$B_{2n} = \frac{\frac{1}{n}\sum_{j=1}^n k_h(\left\|i(x) - i(X_j)\right\|^2) - \mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2)}{\mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2)},$$

with $g(y) := f(x) - f(y)$. In Theorem 25 we have shown that for $x \in M \backslash \partial M$,

$$\lim_{h\to 0} (\Delta_{0,h}^{(\mathrm{rw})} f)(x) = \lim_{h\to 0} \frac{1}{h^2} (C_{0,h}\, g)(x) = -\frac{C_2}{2C_1} (\Delta_2 f)(x).$$

Using the lower bound of $p_h(x) = \mathbb{E}_Z\, k_h(\|i(x) - i(Z)\|^2)$ derived in Lemma 46 we can for $hR_k \leq \kappa/2$ directly apply Lemma 23. Thus there exist constants d_1 and d_2 such that

$$\mathrm{P}(\,|B_{1n}| \geq h^2 t\,) \leq 2\exp\left(-\frac{nh^{m+2}t^2}{2\,\|k\|_\infty (d_2 + t\,h\,d_1/3)}\right).$$

The same analysis can be done for B_{2n}. This shows convergence in probability. Complete convergence (which implies almost sure convergence) can be shown by proving for all $t > 0$ the convergence of the series $\sum_{n=0}^{\infty} \mathrm{P}\left(|B_{1n}| \geq h^2 t\right) < \infty$. A sufficient condition for that is $nh^{m+2}/\log n \to \infty$ as $n \to \infty$. ■

The weak pointwise consistency of the unnormalized graph Laplacian for compact submanifolds with the uniform probability measure using the Gaussian kernel for the weights and $\lambda = 0$ was proven by Belkin and Niyogi (2005). A more general result appeared independently in Hein et al. (2005). We prove here the limits of all three graph Laplacians for general submanifolds with boundary of bounded geometry, general probability measures P, and general kernel functions k as stated in our standard assumptions.

The rest of this section is devoted to the general case $\lambda \neq 0$. We show that with high probability the extended graph Laplacians $\Delta_{\lambda,h,n}$ are pointwise close to the continuous operators $\Delta_{\lambda,h}$ when applied to a function $f \in C^3(M)$. The following proposition is helpful.

Proposition 29 *Suppose the standard assumptions hold. Furthermore, let k be a kernel with compact support on $[0, R_k^2]$. Fix $\lambda \in \mathbb{R}$ and let $x \in M\backslash\partial M$, $f \in C^3(M)$ and define $g(y) := f(x) - f(y)$. Then there exists a constant C such that for any $\frac{2\|k\|_\infty}{nh^m} < \varepsilon < 1/C$, $0 < h < \frac{\kappa}{2R_k}$, the following events hold with probability at least $1 - C\,n\,e^{\frac{-nh^m\varepsilon^2}{C}}$,*

$$|(\tilde{A}_{\lambda,h,n} g)(x) - (\tilde{A}_{\lambda,h} g)(x)| \leq \varepsilon\, h, \qquad |\tilde{d}_{\lambda,h,n}(x) - \tilde{d}_{\lambda,h}(x)| \leq \varepsilon.$$

Proof The idea of this proof is to show that several empirical quantities which can be expressed as a sum of i.i.d. random variables are close to their expectation. Then one can deduce that also $(\tilde{A}_{\lambda,h,n} g)(x)$ will be close to $(\tilde{A}_{\lambda,h} g)(x)$. The proof for $\tilde{d}_{\lambda,h,n}$ can then be easily adapted from the following. We consider here only $\lambda \geq 0$, the proof for $\lambda < 0$ is even simpler. Consider the event $\mathcal{E}$ for which one has

$$\begin{cases} \text{for any } j \in \{1,\ldots,n\},\ \left|d_{h,n}(X_j) - p_h(X_j)\right| \leq \varepsilon, \\ \left|d_{h,n}(x) - p_h(x)\right| \leq \varepsilon, \\ \left|\frac{1}{n}\sum_{j=1}^{n} \frac{k_h(\left\|i(x)-i(X_j)\right\|^2)|g(X_j)|}{[p_h(x)\,p_h(X_j)]^\lambda} - \int_M k_h(\|i(x)-i(y)\|^2)|g(y)|\frac{p(y)}{[p_h(x)\,p_h(y)]^\lambda}\sqrt{\det g}\,dy\right| \leq h\varepsilon. \end{cases}$$

We will now prove that for sufficiently large C the event $\mathcal{E}$ holds with probability at least $1 - Cne^{-\frac{nh^m\varepsilon^2}{C}}$. For the second assertion defining $\mathcal{E}$, we use Lemma 23

$$\mathrm{P}(\,|d_{h,n}(x) - p_h(x)| > \varepsilon\,) \leq 2\exp\left(-\frac{nh^m\varepsilon^2}{2b_2 + 2b_1\varepsilon/3}\right),$$

where b_1 and b_2 are constants depending on the kernel k and p. For the first term in the event $\mathcal{E}$ remember that $k(0) = 0$. We get for $\frac{\|k\|_\infty}{nh^m} < \varepsilon/2$ and $1 \le j \le n$,

$$\mathrm{P}\Big(\big| \tfrac{1}{n}\textstyle\sum_{i=1}^n k_h(\big\|i(X_j) - i(X_i)\big\|^2) - p_h(x)\big| > \varepsilon \,\Big|X_j\Big) \le 2\exp\Big(-\tfrac{(n-1)h^m\varepsilon^2}{8b_2+4b_1\varepsilon/3}\Big).$$

This follows by

$$\begin{aligned}\Big|\frac{1}{n}\sum_{i=1}^n k_h(\big\|i(X_j) - i(X_i)\big\|^2) - p_h(X_j)\Big| \le& \Big|\frac{1}{n(n-1)}\sum_{i=1}^n k_h(\big\|i(X_j) - i(X_i)\big\|^2)\Big| \\ &+\Big|\frac{1}{n-1}\sum_{i\neq j} k_h(\big\|i(X_j) - i(X_i)\big\|^2) - p_h(X_j)\Big|,\end{aligned}$$

where the first term is upper bounded by $\frac{\|k\|_\infty}{nh^m}$. First integrating wrt to the law of X_j (the right hand side of the bound is independent of X_j) and then using a union bound, we get

$$\mathrm{P}\Big(\text{for any } j \in \{1,\dots,n\},\ \big|d_{h,n}(X_j) - p_h(X_j)\big| \le \varepsilon\Big) > 1 - 2n\exp\Big(-\tfrac{(n-1)h^m\varepsilon^2}{8b_2+4b_1\varepsilon/3}\Big).$$

Noting that $\frac{1}{p_h(x)p_h(y)}$ is upper bounded by Lemma 46 we get by Lemma 23 for $hR_k \le \kappa/2$ a Bernstein type bound for the probability of the third event in $\mathcal{E}$. Finally, combining all these results, we obtain that there exists a constant C such that for $h \le \frac{\kappa}{2R_k}$ and $\frac{2\|k\|_\infty}{nh^m} \le \varepsilon \le 1$, the event[5] $\mathcal{E}$ holds with probability at least $1 - Cne^{-\frac{nh^m\varepsilon^2}{C}}$. Let us define

$$\begin{cases} \mathcal{B} &:= \int_M k_h(\|i(x) - i(y)\|^2)(f(x) - f(y))[p_h(x)\,p_h(y)]^{-\lambda} p(y)\sqrt{\det g}dy, \\ \hat{\mathcal{B}} &:= \frac{1}{n}\sum_{j=1}^n k_h(\big\|i(x) - i(X_j)\big\|^2)(f(x) - f(X_j))\big[d_{h,n}(x)\,d_{h,n}(X_j)\big]^{-\lambda}, \end{cases}$$

then $(\tilde{A}_{\lambda,h,n}g)(x) = \hat{\mathcal{B}}$ and $(\tilde{A}_{\lambda,h}g)(x) = \mathcal{B}$. Let us now work only on the event $\mathcal{E}$. By Lemma 46 for any $y \in B_{\mathbb{R}^d}(x, hR_k) \cap M$ there exist constants D_1, D_2 such that $0 < D_1 \le p_h(y) \le D_2$. Using the first order Taylor formula of $[x \mapsto x^{-\lambda}]$, we obtain that for any $\lambda \ge 0$ and $a, b > \beta$, $\big|a^{-\lambda} - b^{-\lambda}\big| \le \lambda\beta^{-\lambda-1}|a-b|$. So we can write for $\varepsilon < D_1/2$,

$$\begin{aligned}\Big|\frac{1}{\big(d_{h,n}(x)\,d_{h,n}(X_j)\big)^\lambda} - \frac{1}{\big(p_h(x)\,p_h(X_j)\big)^\lambda}\Big| &\le \lambda(D_1-\varepsilon)^{-2\lambda-2}|d_{h,n}(x)d_{h,n}(X_j) - p_h(x)p_h(X_j)| \\ &\le 2\lambda(D_1-\varepsilon)^{-2\lambda-2}(D_2+\varepsilon)\varepsilon := C\varepsilon.\end{aligned}$$

Noting that for $hR_k \le \kappa/2$ by Lemma 18, $d_M(x,y) \le 2hR_k$, $\forall y \in B_{\mathbb{R}^d}(x, hR_k) \cap M$,

$$\begin{aligned}\big|\hat{\mathcal{B}} - \mathcal{B}\big| &\le \Big|\tfrac{1}{n}\textstyle\sum_{j=1}^n k_h(\big\|i(x) - i(X_j)\big\|^2)|f(x) - f(X_j)|\,C\varepsilon \\ &\quad +\Big|\tfrac{1}{n}\textstyle\sum_{j=1}^n k_h(\big\|i(x) - i(X_j)\big\|^2)(f(x) - f(X_j))[p_h(x)\,p_h(X_j)]^{-\lambda} - \mathcal{B}\Big| \\ &\le 2C\,\|k\|_\infty R_k \sup_{y\in M}\|\nabla f\|_{T_yM}\, h\varepsilon + h\varepsilon.\end{aligned}$$

We have proven that there exists a constant $C > 1$ such that for any $0 < h < \frac{\kappa}{2R_k}$ and $\frac{2\|k\|_\infty}{nh^m} < \varepsilon < 1/C$,

$$\big|(\tilde{A}_{\lambda,h,n}g)(x) - (\tilde{A}_{\lambda,h}g)(x)\big| \le C'''h\varepsilon,$$

5. The upper bound on ε is here not necessary but allows to write the bound more compactly.

with probability at least $1 - Cne^{-\frac{nh^m\varepsilon^2}{C}}$. ■

This leads us to our first main result for the random walk and the unnormalized graph Laplacian.

Theorem 30 (Pointwise consistency of $\Delta^{(\mathbf{rw})}_{\lambda,h,n}$ and $\Delta^{(\mathbf{u})}_{\lambda,h,n}$) *Suppose the standard assumptions hold. Furthermore, let k be a kernel with compact support on $[0,R_k^2]$. Let $x \in M \backslash \partial M$, $\lambda \in \mathbb{R}$. Then for any $f \in C^3(M)$ there exists a constant C such that for any $\frac{2\|k\|_\infty}{nh^{m+1}} < \varepsilon < 1/C$, $0 < h < h_{\max}$ with probability at least $1 - Cne^{\frac{-nh^{m+2}\varepsilon^2}{C}}$,*

$$|(\Delta^{(rw)}_{\lambda,h,n} f)(x) - (\Delta^{(rw)}_{\lambda,h} f)(x)| \leq \varepsilon,$$
$$|(\Delta^{(u)}_{\lambda,h,n} f)(x) - (\Delta^{(u)}_{\lambda,h} f)(x)| \leq \varepsilon.$$

Define $s = 2(1-\lambda)$. Then if $h \to 0$ and $nh^{m+2}/\log n \to \infty$,

$$\lim_{n\to\infty} (\Delta^{(rw)}_{\lambda,h,n} f)(x) = -\frac{C_2}{2C_1}(\Delta_s f)(x) \qquad \textit{almost surely},$$
$$\lim_{n\to\infty} (\Delta^{(u)}_{\lambda,h,n} f)(x) = -\frac{C_2}{2C_1^{2\lambda}}\, p(x)^{1-2\lambda}(\Delta_s f)(x) \qquad \textit{almost surely}.$$

In particular, under the above conditions,

$$\left|(\Delta^{(rw)}_{\lambda,h,n} f)(x) - \big[-\frac{C_2}{2C_1}(\Delta_s f)(x)\big]\right| = O(h) + O\left(\sqrt{\tfrac{\log n}{nh^{m+2}}}\right) \qquad a.s.,$$
$$\left|(\Delta^{(u)}_{\lambda,h,n} f)(x) - \big[-\frac{C_2}{2C_1^{2\lambda}}\, p(x)^{1-2\lambda}(\Delta_s f)(x)\big]\right| = O(h) + O\left(\sqrt{\tfrac{\log n}{nh^{m+2}}}\right) \qquad a.s..$$

The optimal rate for $h(n)$ is $h = O((\log n/n)^{\frac{1}{m+4}})$.

Proof In Equation (3) it was shown that

$$(\Delta^{(\mathrm{rw})}_{\lambda,h,n} f)(x) = \frac{1}{h^2}\left(\frac{\tilde{A}_{\lambda,h,n} g}{\tilde{d}_{\lambda,h,n}}\right)(x), \quad (\Delta^{(\mathrm{u})}_{\lambda,h,n} f)(x) = \frac{1}{h^2}(\tilde{A}_{\lambda,h,n} g)(x),$$

where $g(y) := f(x) - f(y)$. Since f is Lipschitz we can directly apply Proposition 29 so that for the unnormalized Laplacian we get with probability $1 - Cne^{\frac{-nh^{m+2}\varepsilon^2}{C}}$,

$$|(\Delta^{(\mathrm{u})}_{\lambda,h,n} f)(x) - (\Delta^{(\mathrm{u})}_{\lambda,h} f)(x)| \leq \varepsilon.$$

For the random walk Laplacian $\Delta^{(\mathrm{rw})}_{\lambda,h,n}$ we work on the event where $|\tilde{d}_{\lambda,h,n} - \tilde{d}_{\lambda,h}| \leq h\varepsilon$, where $\varepsilon \leq \frac{1}{2h}\tilde{d}_{\lambda,h}$. This holds by Proposition 29 with probability $1 - Cne^{\frac{-nh^{m+2}\varepsilon^2}{C}}$. Moreover, note that by Lemmas 18 and 14 for $hR_k \leq \min\{\kappa/2, R_0\}$, we have

$$\left|\tilde{A}_{\lambda,h} g\right| \leq \frac{2^m R_k^m S_2}{D_1^{2\lambda}} \|p\|_\infty \|k\|_\infty 2L(f) h R_k.$$

Using Proposition 29 for $\tilde{A}_{\lambda,h,n}g$ and the bounds of $p_h(x)$ from Lemma 46,

$$\begin{aligned}
&\left|(\Delta^{(\mathrm{rw})}_{\lambda,h,n}f)(x) - (\Delta^{(\mathrm{rw})}_{\lambda,h}f)(x)\right| = \frac{1}{h^2}\left|\frac{(\tilde{A}_{\lambda,h,n}g)(x)}{\tilde{d}_{\lambda,h,n}(x)} - \frac{(\tilde{A}_{\lambda,h}g)(x)}{\tilde{d}_{\lambda,h}(x)}\right| \\
&\leq \frac{1}{h^2}\left(\frac{|(\tilde{A}_{\lambda,h,n}g)(x) - (\tilde{A}_{\lambda,h}g)(x)|}{\tilde{d}_{\lambda,h,n}(x)} + (\tilde{A}_{\lambda,h}g)(x)\,\frac{|\tilde{d}_{\lambda,h,n}(x) - \tilde{d}_{\lambda,h}(x)|}{\tilde{d}_{\lambda,h,n}(x)\tilde{d}_{\lambda,h}(x)}\right) \\
&\leq \frac{2D_2^{2\lambda}}{D_1}\,\varepsilon + \frac{2^m R_k^m S_2}{D_1^{2\lambda}}\,\|p\|_\infty \|k\|_\infty\, 2L(f)\,R_k\,\varepsilon := C\varepsilon,
\end{aligned}$$

with probability $1 - Cne^{\frac{-nh^{m+2}\varepsilon^2}{C}}$. By Theorem 25 and 27 we have for $s = 2(1-\lambda)$,

$$\left|(\Delta^{(\mathrm{rw})}_{\lambda,h}f)(x) - \left[-\frac{C_2}{2C_1}(\Delta_s f)(x)\right]\right| \leq Ch,$$
$$\left|(\Delta^{(\mathrm{u})}_{\lambda,h}f)(x) - \left[-\frac{C_2}{2C_1^{2\lambda}}\,p(x)^{1-2\lambda}(\Delta_s f)(x)\right]\right| \leq Ch.$$

Combining both results together with the Borel-Cantelli-Lemma yields almost sure convergence. The optimal rate for $h(n)$ follows by equating both order terms. ■

Using the relationship between the unnormalized and the normalized Laplacian the pointwise consistency can be easily derived. However, the conditions for convergence are slightly stronger since the Laplacian is applied to the function $f/\sqrt{\tilde{d}_{\lambda,h,n}}$.

Theorem 31 (Pointwise consistency of $\Delta^{(\mathrm{n})}_{\lambda,h,n}$) *Suppose that the standard assumptions hold. Furthermore, let k be a kernel with compact support on $[0,R_k^2]$. Let $x \in M\backslash\partial M$, $\lambda \in \mathbb{R}$. Then for any $f \in C^3(M)$ there exists a constant C such that for any $\frac{2\|k\|_\infty}{nh^{m+2}} < \varepsilon < 1/C$, $0 < h < h_{\max}$ with probability at least $1 - Cn^2 e^{\frac{-nh^{m+4}\varepsilon^2}{C}}$,*

$$\left|(\Delta^{(n)}_{\lambda,h,n}f)(x) - (\Delta^{(n)}_{\lambda,h}f)(x)\right| \leq \varepsilon.$$

Define $s = 2(1-\lambda)$. Then if $h \to 0$ and $nh^{m+4}/\log n \to \infty$,

$$\lim_{n\to\infty}(\Delta^{(n)}_{\lambda,h,n}f)(x) = -p(x)^{\frac{1}{2}-\lambda}\frac{C_2}{2C_1}\Delta_s\left(\frac{f}{p^{\frac{1}{2}-\lambda}}\right)(x) \quad \textit{almost surely.}$$

Proof We reduce the case of $\Delta^{(\mathrm{n})}_{\lambda,h,n}$ to the case of $\Delta^{(\mathrm{u})}_{\lambda,h,n}$. We work on the event where

$$|\tilde{d}_{\lambda,h,n}(x) - \tilde{d}_{\lambda,h}(x)| \leq h^2\varepsilon, \qquad |\tilde{d}_{\lambda,h,n}(X_i) - \tilde{d}_{\lambda,h}(X_i)| \leq h^2\varepsilon, \quad \forall i = 1,\ldots,n.$$

From Proposition 29 we know that this holds with probability at least $1 - Cn^2 e^{\frac{-nh^{m+4}\varepsilon^2}{C}}$. Working on this event we get by a similar argumentation as in the proof of Theorem 30 that there exists a constant C' such that

$$\begin{aligned}
\left|(\Delta^{(\mathrm{n})}_{\lambda,h,n}f)(x) - \tfrac{1}{\tilde{d}_{\lambda,h}(x)}\left(\Delta^{(\mathrm{u})}_{\lambda,h,n}\tfrac{f}{\sqrt{\tilde{d}_{\lambda,h}}}\right)(x)\right| &= \frac{1}{h^2}\Big|\tilde{d}_{\lambda,h,n}(x)f(x)\left[\tfrac{1}{\tilde{d}_{\lambda,h}(x)} - \tfrac{1}{\tilde{d}_{\lambda,h,n}(x)}\right] \\
&+ \sum_{i=1}^n \tilde{k}_{\lambda,h}(x,X_i)f(X_i)\left[\tfrac{1}{\sqrt{\tilde{d}_{\lambda,h}(x)\tilde{d}_{\lambda,h}(X_i)}} - \tfrac{1}{\sqrt{\tilde{d}_{\lambda,h,n}(x)\tilde{d}_{\lambda,h,n}(X_i)}}\right]\Big| \leq C'\varepsilon.
\end{aligned}$$

Noting that $\frac{f}{\tilde{d}_{\lambda,h}}$ is Lipschitz since f and $\tilde{d}_{\lambda,h}$ are Lipschitz and upper and lower bounded, on $M \cap B_{\mathbb{R}^d}(x, hR_k)$ one can apply Theorem 30 to derive the first statement. The second statement follows by Corollary 27. ■

Acknowledgments

We would like to thank Olaf Wittich and Bernhard Schölkopf for helpful comments.

Appendix A. Basic Concepts of Differential Geometry

In this section we introduce the necessary basics of differential geometry, in particular normal coordinates and submanifolds in $\mathbb{R}^d$, used in this paper. Note that the definition of the Riemann curvature tensor varies across textbooks which can result in sign-errors. Throughout the paper we use the convention of Lee (1997).

A.1 Basics

Definition 32 *A d-dimensional **manifold** X **with boundary** is a topological (Hausdorff) space such that every point has a neighborhood homeomorphic to an open subset of $\mathbb{H}^d = \{(x^1,\dots,x^d) \in \mathbb{R}^d | x_1 \geq 0\}$. A **chart** (or local coordinate system) (U,ϕ) of a manifold X is an open set $U \subset X$ together with a homeomorphism $\phi : U \to V$ of U onto an open subset $V \subset \mathbb{H}^d$. The coordinates $(x^1,\dots,x^d)$ of $\phi(x)$ are called the coordinates of x in the chart (U,ϕ). A C^r**-atlas** $\mathcal{A}$ is a collection of charts,*

$$\mathcal{A} \triangleq \cup\{(U_\alpha,\phi_\alpha), \alpha \in I\},$$

*where I is an index set, such that $X = \cup_{\alpha\in I} U_\alpha$ and for any $\alpha,\beta \in I$ the corresponding **transition map**,*

$$\phi_\beta \circ \phi_\alpha^{-1}\big|_{\phi_\alpha(U_\alpha \cap U_\beta)} : \phi(U_\alpha \cap U_\beta) \to \mathbb{H}^d,$$

*is r-times continuously differentiable. A **smooth manifold with boundary** is a manifold with boundary with a C^∞-atlas.*

For more technical details behind the definition of a manifold with boundary we refer to Lee (2003). Note that the boundary ∂M of M is a $(d-1)$-dimensional manifold without boundary. In textbooks one often only finds the definition of a manifold without boundary which can be easily recovered from the above definition by replacing $\mathbb{H}^d$ with $\mathbb{R}^d$. The interior $M \backslash \partial M$ of the manifold M is a manifold without boundary.

Definition 33 *A subset M of a d-dimensional manifold X is a m-dimensional **submanifold** M **with boundary** if every point $x \in M$ is in the domain of a chart (U,ϕ) of X such that*

$$\phi : U \cap M \to \mathbb{H}^m \times a, \quad \phi(x) = (x^1,\dots,x^m,a^1,\dots,a^{d-m}),$$

*where a is a fixed element in $\mathbb{R}^{d-m}$. X is called the **ambient space** of M.*

This definition excludes irregular cases like intersecting submanifolds or self-approaching submanifolds. In the following it is more appropriate to take the following point of view. Let M be an m-dimensional manifold. The smooth mapping $i : M \to X$ is said to be an immersion if i is differentiable and the differential of i has rank m everywhere. An injective immersion is called embedding if it is an homeomorphism onto its image. In this case $i(M)$ is a submanifold of X. If M is compact and i is an injective immersion, then i is an embedding. This is not the case if M is not compact since $i(M)$ can be self-approaching.

Definition 34 *A **Riemannian manifold** (M,g) is a smooth manifold M together with a tensor[6] of type $(0,2)$, called the metric tensor g, at each $p \in M$, such that g defines an inner product on the tangent space T_pM which varies smoothly over M. The volume form induced by g is given in local coordinates as $dV = \sqrt{\det g}\, dx^1 \wedge \ldots \wedge dx^m$. dV is uniquely determined by $dV(e_1,\ldots,e_m) = 1$ for any oriented orthonormal basis $e_1,\ldots,e_m$ in T_xM.*

The metric tensor induces for every $p \in M$ an isometric isomorphism between the tangent space T_pM and its dual T_p^*M. A submanifold M of a Riemannian manifold (X,g) has a natural Riemannian metric h induced from X in the following way. Let $i : M \to X$ be an embedding so that M is a submanifold of X. Then one can induce a metric h on M using the mapping i, namely $h = i^*g$, where $i^* : T^*_{i(x)}X \to T_x^*M$ is the pull-back[7] of the differentiable mapping i. In this case i trivially is an isometric embedding of (M,h) into (X,g). In the paper we always use on the submanifold M the metric induced from $\mathbb{R}^d$.

Definition 35 *The **Laplace-Beltrami operator** Δ_M of a Riemannian manifold is defined as $\Delta_M = \operatorname{div}(\operatorname{grad})$. For a twice differentiable function $f : M \to \mathbb{R}$ it is explicitly given as*

$$\Delta_M f = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^j} \Big(\sqrt{\det g}\, g^{ij} \frac{\partial f}{\partial x^i} \Big),$$

where g^{ij} are the components of the inverse of the metric tensor $g = g_{ij}\, dx^i \otimes dx^j$.

A.2 Normal Coordinates

Since in the proofs we use normal coordinates, we give here a short introduction. Intuitively, normal coordinates around a point p of an m-dimensional Riemannian manifold M are coordinates chosen such that M looks around p like $\mathbb{R}^m$ in the best possible way. This is achieved by adapting the coordinate lines to geodesics through the point p. The reference for the following material is the book of Jost (2002). We denote by c_v the unique geodesic starting at $c(0) = x$ with tangent vector $\dot{c}(0) = v$ (c_v depends smoothly on p and v).

Definition 36 *Let M be a Riemannian manifold, $p \in M$, and $V_p = \{v \in T_pM,\ c_v \text{ defined on } [0,1]\}$, then, $\exp_p : V_p \to M$, $v \mapsto c_v(1)$, is called the **exponential map** of M at p.*

It can be shown that $\exp_p$ maps a neighborhood of $0 \in T_pM$ diffeomorphically onto a neighborhood U of $p \in M$. This justifies the definition of normal coordinates.

6. A tensor T of type (m,n) is a multilinear form $T_pM \times \ldots T_pM \times T_p^*M \times \ldots \times T_p^*M \to \mathbb{R}$ (n-times T_pM, m-times T_p^*M).

7. T_x^*M is the dual of the tangent space T_xM. Every differentiable mapping $i : M \to X$ induces a pull-back $i^* : T^*_{i(x)}X \to T_x^*M$. Let $u \in T_xM$, $w \in T^*_{i(x)}X$ and denote by i' the differential of i. Then i^* is defined by $(i^*w)(u) = w(i'u)$.

Definition 37 *Let U be a neighborhood of p in M such that $\exp_p$ is a diffeomorphism. The local coordinates defined by the chart $(U, \exp_p^{-1})$ are called **normal coordinates** at p.*

Note that in $T_pM \simeq \mathbb{R}^m \supset \exp_p^{-1}(U)$ we use always an orthonormal basis. The injectivity radius describes the largest ball around p such that normal coordinates can be introduced.

Definition 38 *Let M be a Riemannian manifold. The **injectivity radius** of $p \in M$ is*

$$\text{inj}(p) = \sup\{\rho > 0,\ \exp_p \text{ is defined on } \overline{B_{\mathbb{R}^m}(0,\rho)} \text{ and injective}\}.$$

It can be shown that $\text{inj}(p) > 0, \forall p \in M \backslash \partial M$. Moreover, for compact manifolds without boundary there exists a lower bound $\text{inj}_{\min} > 0$ such that $\text{inj}(p) \geq \text{inj}_{\min}, \forall p \in M$. However, for manifolds with boundary one has $\text{inj}(p_n) \to 0$ for any sequence of points p_n with limit on the boundary. The motivation for introducing normal coordinates is that the geometry is particularly simple in these coordinates. The following theorem makes this more precise.

Theorem 39 *In normal coordinates around p one has for the Riemannian metric g and the Laplace-Beltrami operator Δ_M applied to a function f at $p = \exp_p^{-1}(0)$,*

$$g_{ij}(0) = \delta_{ij}, \quad \frac{\partial}{\partial x^k} g_{ij}(0) = 0, \quad (\Delta_M f)(0) = \sum_{i=1}^{m} \frac{\partial^2 f}{\partial (x^i)^2}(0).$$

The second derivatives of the metric tensor cannot be made to vanish in general. There curvature effects come into play which cannot be deleted by a coordinate transformation. To summarize, normal coordinates with center p achieve that, up to first order, the geometry of M at point p looks like that of $\mathbb{R}^m$.

A.3 The Second Fundamental Form

In this section we assume that M is an isometrically embedded submanifold of a manifold X. At each point $p \in M$ one can decompose the tangent space T_pX into a subspace T_pM, which is the tangent space to M, and the orthogonal normal space N_pM. In the same way one can split the covariant derivative of X at p, $\tilde{\nabla}_U V$ into a component tangent $(\tilde{\nabla}_U V)^\top$ and normal $(\tilde{\nabla}_U V)^\perp$ to M.

Definition 40 *The **second fundamental form** Π of an isometrically embedded submanifold M of X is defined as*

$$\Pi : T_pM \otimes T_pM \to N_pM, \quad \Pi(U,V) = (\tilde{\nabla}_U V)^\perp.$$

The following theorem, see Lee (1997), then shows that the covariant derivative of M at p is nothing else than the projection of the covariant derivative of X at p onto T_pM.

Theorem 41 (Gauss Formula) *Let U, V be vector fields on M which are arbitrarily extended to X, then the following holds along M*

$$\tilde{\nabla}_U V = \nabla_U V + \Pi(U,V),$$

where $\tilde{\nabla}$ is the covariant derivative of X and ∇ the covariant derivative of M.

The second fundamental form connects also the curvature tensors of X and M.

Theorem 42 (Gauss equation) *For any $U,V,W,Z \in T_pM$ the following equation holds*

$$\tilde{R}(U,V,W,Z) = R(U,V,W,Z) - \langle \Pi(U,Z), \Pi(V,W) \rangle + \langle \Pi(U,W), \Pi(V,Z) \rangle ,$$

where $\tilde{R}$ and R are the Riemann curvature[8] *tensors of X and M.*

In this paper we derive a relationship between distances in M and the corresponding distances in X. Since Riemannian manifolds are length spaces and therefore the distance is induced by length minimizing curves (locally the geodesics), it is of special interest to connect properties of curves of M with respect to X. Applying the Gauss Formula to a curve $c(t) : (t_0,t_1) \to M$ yields the following

$$\tilde{D}_t V = D_t V + \Pi(V, \dot{c}),$$

where $\tilde{D}_t = \dot{c}^a \tilde{\nabla}_a$ and $\dot{c}$ is the tangent vector field to the curve $c(t)$. Now let $c(t)$ be a geodesic parameterized by arc-length, that is with unit-speed, then its acceleration fulfills $D_t \dot{c} = \dot{c}^a \nabla_a \dot{c}^b = 0$ (however that is only true locally in the interior of M, globally if M has boundary length minimizing curves may behave differently especially if a length minimizing curve goes along the boundary its acceleration can be non-zero), and one gets for the acceleration in the ambient space

$$\tilde{D}_t \dot{c} = \Pi(\dot{c}, \dot{c}).$$

In our setting where $X = \mathbb{R}^d$ the term $\tilde{D}_t \dot{c}$ is just the ordinary acceleration $\ddot{c}$ in $\mathbb{R}^d$. Remember that the norm of the acceleration vector is inverse to the curvature of the curve at that point (if c is parameterized by arc-length).[9] Due to this connection it becomes more apparent why the second fundamental form is often called the extrinsic curvature (with respect to X).

The following Lemma shows that the second fundamental form Π of an isometrically embedded submanifold M of $\mathbb{R}^d$ is in normal coordinates just the Hessian of i.

Lemma 43 *Let e_α, $\alpha = 1,\dots,d$ denote an orthonormal basis of $T_{i(x)}\mathbb{R}^d$ then the second fundamental form of M in normal coordinates at y is given as:*

$$\Pi(\partial_{y^i}, \partial_{y^j})\Big|_0 = \frac{\partial^2 i^\alpha}{\partial y^i \partial y^j} e_\alpha.$$

Proof Let $\tilde{\nabla}$ be the flat connection of $\mathbb{R}^d$ and ∇ the connection of M. Then by Theorem 41, $\Pi(\partial_{y^i}, \partial_{y^j}) = \tilde{\nabla}_{i^* \partial_{y^i}} (i^* \partial_{y^j}) - \nabla_{\partial_{y^i}} \partial_{y^j} = \partial_{y^i} \left(\frac{\partial i^\alpha}{\partial_{y^j}} \right) e_\alpha = \frac{\partial^2 i^\alpha}{\partial y^i \partial y^j} e_\alpha$, where the second equality follows from the flatness of $\tilde{\nabla}$ and $\Gamma^i_{jk}\Big|_0 = 0$ in normal coordinates. ■

8. The Riemann curvature tensor of a Riemannian manifold M is defined as $R : T_pM \otimes T_pM \otimes T_pM \to T_p^*M$,

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

In local coordinates x^i, $R_{ijk}{}^l \partial_l = R(\partial_i, \partial_j)\partial_k$ and $R_{ijkm} = g_{lm} R_{ijk}{}^l$.

9. Note that if c is parameterized by arc-length, $\dot{c}$ is tangent to M, that is in particular $\|\dot{c}\|_{T_xX} = \|\dot{c}\|_{T_xM}$

Appendix B. Proofs and Lemmas

The following lemmas are needed in the proof of Proposition 22.

Lemma 44 *If the kernel $k : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies Assumptions 20, then*

$$\int_{\mathbb{R}^m} \frac{\partial k}{\partial x}(\|u\|^2) u^i u^j u^k u^l du = -\frac{1}{2} C_2 \left[\delta^{ij}\delta^{kl} + \delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk}\right].$$

Proof Note first that for a function $f(\|u\|^2)$ one has $\frac{\partial f}{\partial \|u\|^2} = \frac{\partial f}{\partial u_i^2}$. The rest follows from partial integration.

$$\int_{-\infty}^{\infty} \frac{\partial k}{\partial u^2}(u^2)\, u^2\, du = \int_0^{\infty} \frac{\partial k}{\partial v}(v)\, \sqrt{v} dv = \left[k(v)\sqrt{v}\right]_0^{\infty} - \int_0^{\infty} k(v)\, \frac{1}{2\sqrt{v}}\, dv = -\frac{1}{2}\int_{-\infty}^{\infty} k(u^2) du,$$

where $[k(v)\sqrt{v}]_0^{\infty} = 0$ due to the boundedness and exponential decay of k.

In the same way one can derive, $\int_{-\infty}^{\infty} \frac{\partial k}{\partial u^2}(u^2)\, u^4\, du = -\frac{3}{2}\int_{-\infty}^{\infty} k(u^2) u^2\, du$. The result follows by noting that since k is an even function only integration over even powers of coordinates will be non-zero. ■

Lemma 45 *Let k satisfy Assumption 20 and let V_{ijkl} be a given tensor. Assume now $\|z\|^2 \geq \|z\|^2 + V_{ijkl} z^i z^j z^k z^l + \beta(z)\|z\|^5 \geq \frac{1}{4}\|z\|^2$ on $B(0, r_{\min}) \subset \mathbb{R}^m$, where $\beta(z)$ is continuous and $\beta(z) \sim O(1)$ as $z \to 0$. Then there exists a constant C and a $h_0 > 0$ such that for all $h < h_0$ and all $f \in C^3(B(0, r_{\min}))$,*

$$\Big| \int_{B(0,r_{\min})} k_h \left(\frac{\|z\|^2 + V_{ijkl} z^i z^j z^k z^l + \beta(z)\|z\|^5)}{h^2} \right) f(z) dz$$
$$- \left(C_1 f(0) + C_2 \frac{h^2}{2} \left[(\Delta f)(0) - f(0) \sum_{i,k}^{m} V_{iikk} + V_{ikik} + V_{ikki} \right] \right) \Big| \leq C h^3,$$

where C is a constant depending on k, $r_{\min}$, V_{ijkl} and $\|f\|_{C^3}$.

Proof As a first step we do a Taylor expansion of the kernel around $\|z\|^2/h^2$:

$$k_h\left(\frac{\|z\|^2 + \eta}{h^2}\right) = k_h\left(\frac{\|z\|^2}{h^2}\right) + \frac{\partial k_h}{\partial x}\Big|_{\frac{\|z\|^2}{h^2}} \frac{\eta}{h^2} + \frac{\partial^2 k_h(x)}{\partial x^2}\Big|_{\frac{\|z\|^2(1-\theta)+\theta\eta}{h^2}} \frac{\eta^2}{h^4},$$

where in the last term $0 \leq \theta(z) \leq 1$. We then decompose the integral:

$$\int_{B(0,r_{\min})} k_h\left(\frac{\|z\|^2 + V_{ijkl} z^i z^j z^k z^l + \beta(z)\|z\|^5}{h^2}\right) f(z) dz$$
$$= \int_{\mathbb{R}^m} \left(k_h\left(\frac{\|z\|^2}{h^2}\right) + \frac{\partial k_h}{\partial x}\Big|_{\frac{\|z\|^2}{h^2}} \frac{V_{ijkl} z^i z^j z^k z^l}{h^2} \right) \left(f(0) + \langle \nabla f|_0, z\rangle + \frac{1}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 z^i z^j \right) dz + \sum_{i=0}^{4} \alpha_i,$$

where we define the five error terms α_i as:

$$\alpha_0 = \int_{B(0,r_{\min})} \frac{\partial k_h}{\partial x}\Big|_{\frac{\|z\|^2}{h^2}} \frac{\beta(z)\|z\|^5}{h^2} f(z)\,dz,$$

$$\alpha_1 = \int_{B(0,r_{\min})} \frac{\partial^2 k_h}{\partial x^2}\Big|_{\frac{\|z\|^2(1-\theta)+\theta\eta}{h^2}} \frac{\left(V_{ijkl}z^i z^j z^k z^l + \beta(z)\|z\|^5\right)^2}{h^4} f(z)dz,$$

$$\alpha_2 = \int_{B(0,r_{\min})} k_h\left(\frac{\|z\|^2 + V_{ijkl}z^i z^j z^k z^l + \beta(z)\|z\|^5}{h^2}\right)\frac{1}{6}\frac{\partial^3 f}{\partial z^i \partial z^j \partial z^k}(\theta z) z^i z^j z^k dz,$$

$$\alpha_3 = \int_{\mathbb{R}^m \setminus B(0,r_{\min})} k_h\left(\frac{\|z\|^2}{h^2}\right)\left(f(0) + \langle \nabla f|_0, z\rangle + \frac{1}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 z^i z^j\right)dz,$$

$$\alpha_4 = \int_{\mathbb{R}^m \setminus B(0,r_{\min})} \frac{\partial k_h}{\partial x}\Big|_{\frac{\|z\|^2}{h^2}} \frac{V_{ijkl}z^i z^j z^k z^l}{h^2}\left(f(0) + \langle \nabla f|_0, z\rangle + \frac{1}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 z^i z^j\right)dz,$$

where in α_1, $\eta = V_{ijkl}z^i z^j z^k z^l + \beta(z)\|z\|^5$. With $\int_{\mathbb{R}^m} k(\|z\|^2) z_i\,dz = 0$, $\forall i$, and $\int_{\mathbb{R}^m} k(\|z\|^2) z_i z_j dz = 0$ if $i \neq j$, and Lemma 44 the main term simplifies to:

$$\begin{aligned}
&\int_{\mathbb{R}^m}\left(k_h\left(\frac{\|z\|^2}{h^2}\right) + \frac{\partial k_h(x)}{\partial x}\Big|_{\frac{\|z\|^2}{h^2}} \frac{V_{ijkl}z^i z^j z^k z^l}{h^2}\right)\left(f(0) + \frac{1}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 z^i z^j\right)dz \\
=&\int_{\mathbb{R}^m}\left(k(\|u\|^2) + h^2 \frac{\partial k(x)}{\partial x}\Big|_{\|u\|^2} V_{ijkl}u^i u^j u^k u^l\right)\left(f(0) + \frac{h^2}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 u^i u^j\right)du \\
=&C_1 f(0) - \frac{h^2}{2}C_2 f(0) V_{ijkl}\left[\delta^{ij}\delta^{kl} + \delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk}\right] + \frac{h^2}{2}C_2 \sum_{i=1}^m \frac{\partial^2 f}{\partial (z^i)^2}\Big|_0 + O(h^4),
\end{aligned}$$

where the $O(h^4)$ term is finite due to the exponential decay of k and depends on k, $r_{\min}$, V_{ijkl} and $\|f\|_{C^3}$. Now we can upper bound the remaining error terms α_i, $i = 0,\ldots,4$. For the argument of the kernel in α_1 and α_2 we have by our assumptions on $B(0,r_{\min})$:

$$\frac{\|z\|^2}{h^2} \geq \frac{\|z\|^2 + V_{ijkl}z^i z^j z^k z^l + \beta(z)\|z\|^5}{h^2} \geq \frac{\|z\|^2}{4h^2}.$$

Note that this inequality implies that β is uniformly bounded on $B(0,r_{\min})$ in terms of $r_{\min}$ and V_{ijkl}. Moreover, for small enough h we have $\frac{r_{\min}}{h} \geq \sqrt{A}$ (see Assumptions 20 for the definition of A) so that we can use the exponential decay of k for α_3 and α_4,

$$|\alpha_0| \leq h^3 \|f\|_{C^3} \int_{B(0,\frac{r_{\min}}{h})} \frac{\partial k_h}{\partial x}\Big|_{\|u\|^2} |\beta(hu)|\,\|u\|^5\,du.$$

Since $\frac{\partial k_h}{\partial x}$ is bounded and has exponential decay, one has $|\alpha_0| \leq K_0 h^3$ where K_0 depends on k, $r_{\min}$ and $\|f\|_{C^3}$.

$$\begin{aligned}
|\alpha_1| \leq& \int_{B(0,r_{\min})} \left|\frac{\partial^2 k_h}{\partial x^2}\left(\frac{\|z\|^2(1-\theta)+\theta\eta}{h^2}\right)\right| \frac{\left(V_{ijkl}z^i z^j z^k z^l + \beta(z)\|z\|^5\right)^2}{h^4} f(z)dz \\
\leq& h^4 \|f\|_{C^3} \int_{B(0,\frac{r_{\min}}{h})} \left|\frac{\partial^2 k}{\partial x^2}\left(\|u\|^2(1-\theta)+\theta\eta\right)\right| \left(m^2 \max_{i,j,k,l}|V_{ijkl}|\,\|u\|^4 + h\|\beta\|_\infty \|u\|^5\right)^2 du.
\end{aligned}$$

First suppose $\frac{r_{\min}}{h} \leq 2\sqrt{A}$ then the integral is bounded since the integrands are bounded on $B(0, \frac{r_{\min}}{h})$. Now suppose $\frac{r_{\min}}{h} \geq 2\sqrt{A}$ and decompose $B(0, \frac{r_{\min}}{h})$ as $B(0, \frac{r_{\min}}{h}) = B(0, 2\sqrt{A}) \cup B(0, \frac{r_{\min}}{h}) \backslash B(0, 2\sqrt{A})$. On $B(0, 2\sqrt{A})$ the integral is finite since $\left|\frac{\partial^2 k}{\partial x^2}\right|$ is bounded and on the complement the integral is also finite since $\left|\frac{\partial^2 k}{\partial x^2}\right|$ has exponential decay since by assumption

$$\|u\|^2 (1 - \theta(hu)) + \theta(hu)\eta(hu) \geq \frac{1}{4}\|u\|^2 \geq A.$$

Therefore there exists a constant K_1 such that $|\alpha_1| \leq K_1 h^4$. Moreover,

$$|\alpha_2| \leq \int_{B(0, r_{\min})} k_h\left(\frac{\|z\|^2}{4h^2}\right) \frac{1}{6} \frac{\partial^3 f}{\partial z^i \partial z^j \partial z^k}(\theta z) z^i z^j z^k dz \leq \frac{m^{3/2}\|f\|_{C^3} h^3}{6} \int_{\mathbb{R}^m} k\left(\frac{\|u\|^2}{4}\right)\|u\|^3 du \leq K_2 h^3,$$

$$\begin{aligned}
|\alpha_3| &\leq \int_{\mathbb{R}^m \backslash B(0, r_{\min})} k_h\left(\frac{\|z\|^2}{h^2}\right)\left(f(0) + \left\langle \nabla f|_0, z\right\rangle + \frac{1}{2}\frac{\partial^2 f}{\partial z^i \partial z^j}\Big|_0 z^i z^j\right) dz \\
&\leq c\|f\|_{C^3} \int_{\mathbb{R}^m \backslash B(0, r_{\min})} e^{-\frac{\alpha\|z\|^2}{h^2}}(1 + mh^2\|z\|^2) dz \leq c e^{-\alpha\frac{r_{\min}^2}{2h^2}} \left(\frac{2\pi}{\alpha}\right)^{\frac{m}{2}} \left(1 + mh^2 \frac{m}{\alpha}\right), \\
|\alpha_4| &\leq K_4 \int_{\mathbb{R}^m \backslash B(0, r_{\min})} \left|\frac{\partial k_h}{\partial x}\left(\frac{\|z\|^2}{h^2}\right)\right| \frac{\|z\|^4 + \|z\|^6}{h^2} dz \leq c K_4 h^2 e^{-\alpha\frac{r_{\min}^2}{2h^2}} \int_{\mathbb{R}^m} e^{-\alpha\|u\|^2}(\|u\|^4 + h^2\|u\|^6) du,
\end{aligned}$$

where K_4 is a constant depending on $\max_{i,j,k,l}|V_{ijkl}|$ and $\|f\|_{C^3}$. Now one has[10]: $e^{-\frac{\xi^2}{h^2}} \leq h^s/\xi^s$ for $h \leq \xi/s$. In particular, it holds $h^3 \geq e^{-\alpha\frac{r_{\min}^2}{2h^2}}$ for $h \leq \frac{1}{3}\sqrt{\frac{\alpha}{2}} r_{\min}$, so that for $h < \min\{\frac{1}{3}\sqrt{\frac{\alpha}{2}} r_{\min}, \frac{r_{\min}}{\sqrt{A}}\} = h_0$ all error terms are smaller than a constant times h^3 where the constant depends on k, $r_{\min}$, V_{ijkl} and $\|f\|_{C^3}$. This finishes the proof. ■

Now we are ready to prove Proposition 22,

Proof Let $\varepsilon = \frac{1}{3}\min\{\mathrm{inj}(x), \pi\rho\}$[11] where ε is positive by the assumptions on M. Then we decompose M as $M = B(x,\varepsilon) \cup (M \backslash B(x,\varepsilon))$ and integrate separately. The integral over $M \backslash B(x,\varepsilon)$ can be upper bounded by using the definition of $\delta(x)$ (see Assumption 19) and the fact that k is non-increasing:

$$\begin{aligned}
\int_M k_h\left(\|i(x) - i(y)\|_{\mathbb{R}^d}^2\right) f(y) p(y) \sqrt{\det g}\, dy &= \int_{B(x,\varepsilon)} k_h\left(\|i(x) - i(y)\|_{\mathbb{R}^d}^2\right) f(y) p(y) \sqrt{\det g}\, dy \\
&\quad + \int_{M \backslash B(x,\varepsilon)} k_h\left(\|i(x) - i(y)\|_{\mathbb{R}^d}^2\right) f(y) p(y) \sqrt{\det g}\, dy.
\end{aligned}$$

Since k is non-increasing, we have the following inequality for the integral over $M \backslash B(x,\varepsilon)$:

$$\int_{M \backslash B(x,\varepsilon)} k_h\left(\|i(x) - i(y)\|_{\mathbb{R}^d}^2\right) f(y) p(y) \sqrt{\det g}\, dy \leq \frac{1}{h^m} k\left(\frac{\delta(x)^2}{h^2}\right)\|f\|_\infty.$$

10. This inequality can be deduced from $e^x \geq x^n$ for all $x \geq 4n^2$.
11. The factor $1/3$ is needed in Theorem 25.

Since $\delta(x)$ is positive by assumption and k decays exponentially, we can make the upper bound smaller than h^3 for small enough h. Now we deal with the integral over $B(x,\varepsilon)$. Since ε is smaller than the injectivity radius $\mathrm{inj}(x)$, we can introduce normal coordinates $z = \exp_x^{-1}(y)$ on $B(x,\varepsilon)$, so that we can rewrite the integral using Proposition 15 as:

$$\int_{B(0,\varepsilon)} k_h\left(\frac{\|z\|^2 - \frac{1}{12}\sum_{\alpha=1}^d \frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^u \partial z^v} z^a z^b z^u z^v + O(\|z\|^5)}{h^2}\right) p(z) f(z)\sqrt{\det g}\, dz. \tag{5}$$

Using our assumptions, we see that $pf\sqrt{\det g}$ is in $C^3(B(0,\varepsilon))$. Moreover, by Corollary 17 one has for $d_M(x,y) \le \pi\rho$, $\frac{1}{2}d_M(x,y) \le \|x-y\| \le d_M(x,y)$. Therefore we can apply Lemma 45 and compute the integral in (5) which results in:

$$p(0)f(0)\Big(C_1 + \frac{h^2 C_2}{24}\sum_{\alpha=1}^d \frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^c \partial z^d}\Big[\delta^{ab}\delta^{cd} + \delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc}\Big]\Big) + \frac{h^2 C_2}{2}\Delta_M(pf\sqrt{\det g})\Big|_0 + O(h^3), \tag{6}$$

where we have used that in normal coordinates z^i at 0 the Laplace-Beltrami operator Δ_M is given as $\Delta_M f\Big|_x = \sum_{i=1}^m \frac{\partial^2 f}{\partial (z^i)^2}\Big|_0$. The second term in the above equation can be evaluated using the Gauss equations, see Smolyanov, von Weizsäcker, and Wittich (2007, Proposition 6).

$$\begin{aligned}
&\sum_{a,b=1}^m\sum_{\alpha=1}^d \frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^c \partial z^d}\Big[\delta^{ab}\delta^{cd} + \delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc}\Big] = \sum_{a,b=1}^m\sum_{\alpha=1}^d \frac{\partial^2 i^\alpha}{\partial (z^a)^2}\frac{\partial^2 i^\alpha}{\partial (z^b)^2} + 2\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\\
&= 2\sum_{a,b=1}^m\sum_{\alpha=1}^d \left(\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b} - \frac{\partial^2 i^\alpha}{\partial (z^a)^2}\frac{\partial^2 i^\alpha}{\partial (z^b)^2}\right) + 3\sum_{a,b=1}^m\sum_{\alpha=1}^d \frac{\partial^2 i^\alpha}{\partial (z^a)^2}\frac{\partial^2 i^\alpha}{\partial (z^b)^2}\\
&= 2\sum_{a,b=1}^m \langle \Pi(\partial_{z^a},\partial_{z^b}), \Pi(\partial_{z^a},\partial_{z^b})\rangle - \langle \Pi(\partial_{z^a},\partial_{z^a}), \Pi(\partial_{z^b},\partial_{z^b})\rangle + 3\Big\|\sum_{a=1}^m \Pi(\partial_{z^a},\partial_{z^a})\Big\|^2_{T_{i(x)}\mathbb{R}^d}\\
&= -2R + 3\Big\|\sum_{j=1}^m \Pi(\partial_{z^j},\partial_{z^j})\Big\|^2_{T_{i(x)}\mathbb{R}^d},
\end{aligned}$$

where R is the scalar curvature and we used Lemma 43 in the third equality. Plugging this result into (6) and using from Proposition 15, $\Delta_M\sqrt{\det g}\big|_0 = -\frac{1}{3}R$, we are done. ■

Lemma 46 *Let k have compact support on $[0,R_k^2]$ and let $0 < h \le h_{\max}$. Then for any $x \in M$ there exist constants $D_1, D_2 > 0$ independent of h such that for any $y \in B_{\mathbb{R}^d}(x, hR_k) \cap M$,*

$$0 < D_1 \le p_h(y) \le D_2.$$

Proof First suppose that $hR_k < s := \min\{\kappa/2, R_0/2\}$. Since $\|y-z\| \le hR_k \le \kappa/2$ we have by Lemma 18: $\frac{1}{2}d_M(y,z) \le \|y-z\| \le d_M(y,z)$. Moreover, since $p(x) > 0$ on M and p is bounded and continuous, there exist lower and upper bounds $p_{\min}$ and $p_{\max}$ on the density on $B_M(x, 4hR_k)$. That implies

$$p_h(y) \le \frac{\|k\|_\infty}{h^m} p_{\max}\int_{B_M(y,2hR_k)} \sqrt{\det g}\, dz \le \|k\|_\infty p_{\max} S_2 2^m R_k^m,$$

where the last inequality follows from Lemma 14. Note further that $d_M(x,y) \leq 2hR_k$ and $d_M(y,z) \leq 2hR_k$ implies $d_M(x,z) \leq 4hR_k$. Since the kernel function is continuous there exists an r_k such that $k(x) \geq \|k\|_\infty /2$ for $0 < x \leq r_k$. We get

$$p_h(y) \geq \frac{\|k\|_\infty}{2h^m} \int\limits_{B_{\mathbb{R}^d}(x,hr_k)\cap M} p(z)\sqrt{\det g}\,dz \geq \frac{\|k\|_\infty}{2h^m} p_{\min} \operatorname{vol}_M(B_M(x,hr_k)) \geq \frac{\|k\|_\infty}{2} p_{\min} S_1 r_k^m.$$

Now suppose $s \leq hR_k$ and $h \leq h_{\max}$. Then $p_h(y) \leq \frac{\|k\|_\infty}{h^m} \leq \|k\|_\infty \left(\frac{R_k}{s}\right)^m$. For the lower bound we get

$$\begin{aligned} p_h(y) &\geq \int_M k_h(d_M(y,z)) p(z)\sqrt{\det g}\,dz \geq \int_{B_M(y,hr_k)} k_h(d_M(y,z)) p(z)\sqrt{\det g}\,dz \\ &\geq \frac{\|k\|_\infty}{2h^m} \mathrm{P}\Big(B_M(y,hr_k)\Big) \geq \frac{\|k\|_\infty}{2h_{\max}^m} \mathrm{P}\Big(B_M(y,s\frac{r_k}{R_k})\Big). \end{aligned}$$

Since p is continuous and $p > 0$, the function $y \to \mathrm{P}\Big(B_M(y,s\frac{r_k}{R_k})\Big)$ is continuous and positive and therefore has a lower bound greater zero on the ball $B_{\mathbb{R}^d}(x,hR_k) \cap M$. ■

References

R. Alexander and S. Alexander. Geodesics in Riemannian manifolds with boundary. *Indiana Univ. Math. J.*, 30:481–488, 1981.

M. Belkin. *Problems of Learning on Manifolds*. PhD thesis, University of Chicago, 2003. http://www.people.cs.uchicago.edu /~misha /thesis.pdf.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comp.*, 15(6):1373–1396, 2003.

M. Belkin and P. Niyogi. Semi-supervised learning on manifolds. *Machine Learning*, 56:209–239, 2004.

M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In P. Auer and R. Meir, editors, *Proc. of the 18th Conf. on Learning Theory (COLT)*, pages 470–485, Berlin, 2005. Springer.

P. H. Bérard. *Spectral Geometry: Direct and Inverse Problems*. Springer, Berlin, 1986.

M. Bernstein, V. de Silva, J. C. Langford, and J.B. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Stanford University, 2001.

O. Bousquet, O. Chapelle, and M. Hein. Measure based regularization. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Adv. in Neur. Inf. Proc. Syst. (NIPS)*, volume 16. MIT Press, 2004.

S. Canu and A. Elisseeff. Regularization, kernels and sigmoid net. unpublished, 1999.

F. Chung. *Spectral Graph Theory*. AMS, Providence, RI, 1997.

F. Chung and R. P. Langlands. A combinatorial Laplacian with vertex weights. *J. Combinatorial Th. Ser. A*, 75:316–327, 1996.

S. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21:5–30, 2006.

J. Dodziuk. Difference equation, isoperimetric inequality and transience of certain random walks. *Trans. Am. Math. Soc.*, 284:787–794, 1984.

E. Giné and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: Large sample results. *High Dimensional Probability*, 51:238–259, 2006.

W. Greblicki, A. Krzyzak, and M. Pawlak. Distribution-free pointwise consistency of kernel regression estimate. *Ann. Stat.*, 12:1570–1575, 1984.

A. Grigoryan. Heat kernels on weighted manifolds and applications. *Cont. Math.*, 398:93–191, 2006.

L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York, 2004.

M. Hein. *Geometrical aspects of statistical learning theory*. PhD thesis, MPI für biologische Kybernetik/Technische Universität Darmstadt, 2005.

M. Hein. Uniform convergence of adaptive graph-based regularization. In G. Lugosi and H. Simon, editors, *Proc. of the 19th Conf. on Learning Theory (COLT)*, pages 50–64, Berlin, 2006. Springer.

M. Hein and J.-Y. Audibert. Intrinsic dimensionality estimation of submanifolds in $\mathbb{R}^d$. In L. De Raedt and S. Wrobel, editors, *Proc. of the 22nd Int. Conf. on Machine Learning (ICML)*, pages 289–296, 2005.

M. Hein, J.-Y. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In P. Auer and R. Meir, editors, *Proc. of the 18th Conf. on Learning Theory (COLT)*, pages 486–500, Berlin, 2005. Springer.

J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer-Verlag, Berlin, third edition, 2002.

W. Klingenberg. *Riemannian Geometry*. De Gruyter, Berlin, 1982.

S. S. Lafon. *Diffusion Maps and Geometric Harmonics*. PhD thesis, Yale University, 2004. http://www.math.yale.edu /˜sl349 /publications /dissertation.pdf.

J. M. Lee. *Riemannian Manifolds*. Springer, New York, 1997.

J. M. Lee. *Introduction to Smooth Manifolds*. Springer, New York, 2003.

P. McDonald and R. Meyers. Diffusions on graphs, Poisson problems and spectral geometry. *Trans. Amer. Math. Soc.*, 354:5111–5136, 2002.

S. Rosenberg. *The Laplacian on a Riemannian Manifold*. Cambridge University Press, 1997.

T. Schick. *Analysis on ∂-manifolds of Bounded Geometry, Hodge-De Rham Isomorphsim and L^2-Index Theorem*. PhD thesis, Universität Mainz, 1996.

T. Schick. Manifolds with boundary of bounded geometry. *Math. Nachr.*, 223:103–120, 2001.

A. Singer. From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.*, 21:128–134, 2006.

O. G. Smolyanov, H. von Weizsäcker, and O. Wittich. Brownian motion on a manifold as limit of stepwise conditioned standard Brownian motions. In *Stochastic Processes, Physics and Geometry: New Interplays, II*, volume 29, pages 589–602, Providence, RI, 2000. AMS.

O. G. Smolyanov, H. von Weizsäcker, and O. Wittich. Chernoff's theorem and discrete time approximations of Brownian motion on manifolds. *Potential Analysis*, 26:1–29, 2007.

O. Sorkine. Differential representations for mesh processing. *Computer Graphics Forum*, 25(4): 789–807, 2006.

D. Spielman and S. Teng. Spectral partioning works: planar graphs and finite element meshes. In *37th Ann. Symp. on Found. of Comp. Science (FOCS)*, pages 96–105. IEEE Comp. Soc. Press, 1996.

N. T. Varopoulos. Brownian motion and random walks on manifolds. *Ann. Inst. Fourier*, 34:243–269, 1984.

U. von Luxburg. A tutorial on spectral clustering. Technical Report 149, Max Planck Institute for Biological Cybernetics, 2006.

U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering, 2007. to appear in Ann. Stat.

W. Woess. *Random Walks on Infinite Graphs and Groups*. Cambridge University Press, Cambridge, 2000.

G. Xu. Discrete Laplace-Beltrami operators and their convergence. *Computer Aided Geometric Design*, 21:767–784, 2004.

D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Adv. in Neur. Inf. Proc. Syst. (NIPS)*, volume 16, pages 321–328. MIT Press, 2004.

D. Zhou, B. Schölkopf, and T. Hofmann. Semi-supervised learning on directed graphs. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Adv. in Neur. Inf. Proc. Syst. (NIPS)*, volume 17, pages 1633–1640. MIT Press, 2005.

X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, 2002.

X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In T. Fawcett and N. Mishra, editors, *Proc. of the 20nd Int. Conf. on Machine Learning (ICML)*, 2003.