From Graphs to Manifolds – Weak and Strong Pointwise Consistency of Graph Laplacians

Matthias Hein[1], Jean-Yves Audibert[2], and Ulrike von Luxburg[3]

[1] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2] CERTIS, ENPC, Paris, France
[3] Fraunhofer IPSI, Darmstadt, Germany

Abstract. In the machine learning community it is generally believed that graph Laplacians corresponding to a finite sample of data points converge to a continuous Laplace operator if the sample size increases. Even though this assertion serves as a justification for many Laplacian-based algorithms, so far only some aspects of this claim have been rigorously proved. In this paper we close this gap by establishing the strong pointwise consistency of a family of graph Laplacians with data-dependent weights to some weighted Laplace operator. Our investigation also includes the important case where the data lies on a submanifold of $\mathbb{R}^d$.

1 Introduction

In recent years, methods based on graph Laplacians have become increasingly popular. In machine learning they have been used for dimensionality reduction [1], semi-supervised learning [12], and spectral clustering (see [11] for references). The usage of graph Laplacians has often been justified by their relations to the continuous Laplace operator. Most people believe that for increasing sample size, the Laplace operator on the similarity graph generated by a sample converges in some sense to the Laplace operator on the underlying space. It is all the more surprising that rigorous convergence results for the setting given in machine learning do not exist. It is only for some cases where the graph has certain regularity properties such as a grid in $\mathbb{R}^d$ that results are known.

In the more difficult setting where the graph is generated randomly, only some aspects have been proven so far. The approach taken in this paper is first to establish the convergence of the discrete graph Laplacian to a continuous counterpart ("variance term"), and in a second step the convergence of this continuous operator to the continuous Laplace operator ("bias term"). For compact submanifolds in $\mathbb{R}^d$ the second step has already been studied by Belkin [1] for Gaussian weights and the uniform measure, and was then generalized to general isotropic weights and general densities by Lafon [7]. Belkin and Lafon show that the bias term converges pointwise for $h \to 0$, where h is the bandwidth of isotropic weights. However, the convergence of the variance term was left open in [1] and [7].

P. Auer and R. Meir (Eds.): COLT 2005, LNAI 3559, pp. 470–485, 2005.

The first work where, in a slightly different setting, both limit processes have been studied together is Bousquet et al. [3]. Using the law of large numbers for U-statistics, the authors studied the convergence of the regularizer $\Omega_n(f) = \langle f, L_n f\rangle$ for sample size $n \to \infty$ (where $f \in \mathbb{R}^n$ and L_n is the unnormalized graph Laplacian on n sample points). Then taking the limit for the bandwidth $h \to 0$ they arrived at a weighted Laplace operator in $\mathbb{R}^d$. The drawback of this approach is that the limits in n and h are not taken simultaneously.

In contrast to this work, in [11] the bandwidth h was kept fixed while the large sample limit $n \to \infty$ of the graph Laplacian (normalized and unnormalized) was considered. In this setting, the authors show strong convergence results of graph Laplacians to certain limit integral operators, which then even imply the convergence of the eigenvalues and eigenfunctions of the graph Laplacian.

The goal of this paper is to surpass the limitations of previous approaches. We study the convergence of both bias and variance term, where the limits $n \to \infty$ and $h \to 0$ are taken simultaneously. The main achievement of this paper is Theorem 3, where the strong pointwise consistency of the normalized graph Laplacian with varying data dependent weights as introduced in [4] is shown. The limit operator is in general a weighted Laplace-Beltrami operator. Based on our analysis we argue against using the unnormalized graph Laplacian.

We would like to mention that after submission of our manuscript, we learned that a result related to a special case of Theorem 2 has been proven independently by Belkin and Niyogi in their parallel COLT paper [2] and has been announced in [8] (see Section 4 for a short discussion).

Theorem 3 is proven as follows. In section 2 we introduce general graph Laplacians. Then in Section 3, we establish the first step of Theorem 3, namely the convergence of the bias term in the general case where the data lies on a submanifold M in $\mathbb{R}^d$. We prove that the difference between the weighted Laplace-Beltrami operator and its kernel-based approximation goes to zero when the bandwidth $h \to 0$. Then in Section 4 we show that the variance term, namely the difference between the normalized graph Laplacian and the kernel-based approximation, is small with high probability if $nh^{d+4}/\log n \to \infty$. Plugging both results together we arrive at the main result in Theorem 3.

2 The Graph Laplacian

In this section we define the graph Laplacian on an undirected graph. To this end one has to introduce Hilbert spaces H_V and H_E of functions on the vertices V resp. edges E, define a difference operator d, and then set the graph Laplacian as $\Delta = d^* d$. This approach is well-known in discrete potential theory and was independently introduced in [13]. In many articles, graph Laplacians are used without explicitly mentioning d, H_V and H_E. This can be misleading since there always exists a whole family of choices for d, H_V and H_E which all yield the same graph Laplacian.

Hilbert Space Structure 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 is a positive, symmetric $n \times n$ similarity matrix, that is $w_{ij} = w_{ji}$ and $w_{ij} \geq 0,\ i, j = 1, \ldots, n$. We say that there is an (undirected) edge from i to j if $w_{ij} > 0$. Moreover, the degree function d is defined as $d_i = \sum_{j=1}^{n} w_{ij}$. We assume here that $d_i > 0,\ i = 1, \ldots, n$. That means that each vertex has at least one edge. The inner products on the function spaces $\mathbb{R}^V$ resp. $\mathbb{R}^E$ are defined as $\langle f, g \rangle_V = \sum_{i=1}^{n} f_i\, g_i\, \chi(d_i)$ and $\langle F, G \rangle_E = \frac{1}{2} \sum_{i,j=1}^{n} F_{ij}\, G_{ij}\, \phi(w_{ij})$, where $\chi : \mathbb{R}_+^* \to \mathbb{R}_+^*$, $\phi : \mathbb{R}_+^* \to \mathbb{R}_+^*$, and $\mathbb{R}_+^* = \{x \in \mathbb{R} | x > 0\}$. By our assumptions on the graph both inner products are well-defined. Let $\mathcal{H}(V, \chi) = (\mathbb{R}_V, \langle \cdot, \cdot \rangle_V)$ and $\mathcal{H}(E, \phi) = (\mathbb{R}^E, \langle \cdot, \cdot \rangle_E)$.

The Difference Operator d and its Adjoint d^*: We define the difference operator $d : \mathcal{H}(V, \chi) \to \mathcal{H}(E, \phi)$ as follows:

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

where $\gamma : \mathbb{R}_+^* \to \mathbb{R}_+^*$. In the case of a finite graph (i.e., $|V| < \infty$) d is always a bounded operator. The adjoint operator d^* is defined by $\langle df, u \rangle_E = \langle f, d^* u \rangle_V$, for any $f \in H(V, \chi), u \in \mathcal{H}(E, \phi)$. It is straightforward to derive

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

The two terms in the right hand side of Equation (1) can be interpreted as the outgoing resp. ingoing flow.

The General Graph Laplacian: The operator $\Delta : \mathcal{H}(V, \chi) \to \mathcal{H}(V, \chi)$ defined as $\Delta = d^* d$ is obviously self-adjoint and positive semi-definite:

$$\langle f, \Delta g \rangle_V = \langle df, dg \rangle_E = \langle \Delta f, g \rangle_V, \quad \langle f, \Delta f \rangle_V = \langle df, df \rangle_E \geq 0.$$

Using the definitions of the difference operator d and its adjoint d^* we can directly derive the graph Laplacian:

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

The following operators are usually defined as the 'normalized' and 'unnormalized' graph Laplacian Δ_{nm} resp. Δ_{unm}:

$$(\Delta_{\mathrm{nm}} f)(i) = f(i) - \frac{1}{d_i} \sum_{j=1}^{n} w_{ij} f(j), \quad (\Delta_{\mathrm{unm}} f)(i) = d_i f(i) - \sum_{j=1}^{n} w_{ij} f(j).$$

We observe that there exist several choices of χ,γ and ϕ which result in Δ_{nm} or Δ_{unm}. Therefore it can cause confusion if one speaks of the 'normalized' or 'unnormalized' graph Laplacian without explicitly defining the corresponding Hilbert spaces and the difference operator. We just note that one can resolve this ambiguity at least partially if one not only asks for consistency of the graph Laplacian but also for consistency of $\mathcal{H}_V$. Unfortunately, due to space restrictions we cannot further elaborate on this topic.

3 The Weighted Laplacian and Its Approximations

The Laplacian is one of the most prominent operators in mathematics. Nevertheless, most books either deal with the Laplacian in $\mathbb{R}^d$ or the Laplace-Beltrami operator on a manifold M. Not so widely used is the weighted Laplacian on a manifold. This notion is useful when one studies a manifold with a measure, in our case the probability measure generating the data, which in the following we assume to be absolutely continuous wrt the natural volume element of the manifold[1]. In this section we show how the weighted Laplacian can be approximated pointwise by using kernel-based averaging operators. The main results are Theorem 1 and Corollary 1.

Approximations of the Laplace-Beltrami operator based on averaging with the Gaussian kernel have been studied in the special case of the uniform measure on a compact submanifold without boundary in Smolyanov et al.[9, 10] and Belkin [1]. Belkin's result was then generalized by Lafon [7] to general densities and to a wider class of isotropic, positive definite kernels. Whereas the proof of Theorem 1 in [7] applies for compact hypersurfaces[2] in $\mathbb{R}^d$, a proof for general compact submanifolds using boundary conditions is stated in [4]. In this section, we will prove Theorem 1 for general submanifolds M, including the case where M is not compact and without the assumptions of positive definiteness of the kernel nor with any boundary conditions[3]. 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. In this section we will use the Einstein summation convention.

Definition 1 (Weighted Laplacian). *Let (M, g_{ab}) be a Riemannian manifold with measure P, where P has a 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. Then we define the s-th weighted Laplacian Δ_s as*

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

In the family of weighted Laplacians there are two cases which are particularly interesting. The first one, $s = 0$, corresponds to the standard Laplace-Beltrami operator. This notion 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 weighted Laplacian $\Delta_1 = \frac{1}{p}\nabla^a(p\nabla_a)$.

---

1 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.

2 A hypersurface is a submanifold of codimension 1.

3 Boundary conditions are hard to transfer to the graph setting.

This operator can be extended to a self-adjoint operator[4] in $L_2(M, p\,dV)$, which is the natural function space on M given $P = p\,dV$.

Let us introduce the following notations: $C^k(M)$ is the set of C^k-functions on M with finite norm[5] given by $\|f\|_{C^k(M)} = \sup_{\sum_{i=1}^m l_i \leq k,\, x \in M} \left| \frac{\partial^{|\sum_{i=1}^m l_i|}}{\partial(x^1)^{l_1} \dots \partial(x^m)^{l_m}} f(x) \right|$. $B(x, \epsilon)$ denotes a ball of radius ϵ. To bound the deviation of the extrinsic distance in $\mathbb{R}^d$ in terms of the intrinsic distance in M we define for each $x \in M$ the *regularity radius* $r(x)$ as

$$r(x) = \sup\{r > 0 \,\big|\, \|i(x) - i(y)\|^2_{\mathbb{R}^d} \geq \frac{1}{2} d_M^2(x,y), \quad \forall\, y \in B_M(x,r)\}. \tag{3}$$

Assumption 1 – *$i : M \to \mathbb{R}^d$ is a smooth, isometric embedding*[6]*,*
- *The boundary ∂M of M is either smooth or empty,*
- *M has a bounded second fundamental form,*
- *M has bounded sectional curvature,*
- *for any $x \in M$, $r(x) > 0$, and r is continuous,*
- *for any $x \in M$, $\delta(x) := \inf_{y \in M \setminus B_M(x, \frac{1}{3}\min\{\mathrm{inj}(x), r(x)\})} \|i(x) - i(y)\|_{\mathbb{R}^d} > 0$, where $\mathrm{inj}(x)$ is the injectivity radius at x* [7]*.*

The first condition ensures that M is a smooth submanifold of $\mathbb{R}^d$ with the metric induced from $\mathbb{R}^d$ (this is usually meant when one speaks of a submanifold in $\mathbb{R}^d$). The next four properties guarantee that M is well behaved. The last condition 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 $\mathbb{R}^d$. 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	$\Delta_s = \sum_{i=1}^d \frac{\partial^2 f}{\partial (z_i)^2} + \frac{s}{p} \sum_{i=1}^d \frac{\partial p}{\partial z^i} \frac{\partial f}{\partial z^i}$

[4] When M is compact, connected and oriented and for any $f, g \in C^\infty(M)$ vanishing on the boundary, by the first Green identity, we have

$$\int_M f(\Delta_s g)\, p^s dV = \int_M f\big(\Delta g + \frac{s}{p}\langle \nabla p, \nabla g\rangle\big) p^s dV = -\int_M \langle \nabla f, \nabla g\rangle\, p^s dV. \tag{2}$$

[5] We refer to Smolyanov et al.[9] for the technical details concerning this definition.

[6] i.e. the Riemannian metric g_{ab} on M is induced by $\mathbb{R}^d$, $g^M_{ab} = i_* g^{\mathbb{R}^d}_{ab}$, where $g^{\mathbb{R}^d}_{ab} = \delta_{ab}$.

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

The kernels used in this paper are always isotropic, that is they can be written as functions of the norm in $\mathbb{R}^d$. Furthermore we make the following assumptions on the kernel function k:

Assumption 2 – *$k : \mathbb{R}_+ \to \mathbb{R}$ is measurable, non-negative and non-increasing,*
- *$k \in C^2(\mathbb{R}_+)$, that is in particular k and $\frac{\partial^2 k}{\partial x^2}$ are bounded,*
- *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 $t \geq A$, $f(t) \leq c e^{-\alpha t}$, where $f(t) = \max\{k(t), |\frac{\partial k}{\partial x}|(t), |\frac{\partial^2 k}{\partial x^2}|\}$.*

Also let us introduce the helpful notation[8] $k_h(t) = \frac{1}{h^m} k\left(\frac{t}{h^2}\right)$, where we call h the bandwidth of the kernel. Let us now define our kernel-based averaging operators similar to Lafon [7][9]. We define the h-averaged density as:

$$p_h(x) = \int_M k_h\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right) p(y) \sqrt{\det g}\, dy.$$

Note that the distance used in the kernel function is the distance in the ambient space $\mathbb{R}^d$. In this paper we use a family of measure-dependent kernels parameterized by $\lambda \geq 0$ introduced in [4] defined as:

$$\tilde{k}_{\lambda,h}\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right) := \frac{k_h\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right)}{\left[p_h(x) p_h(y)\right]^\lambda}.$$

Let $\tilde{d}_{\lambda,h}(x) = \int_M \tilde{k}_{\lambda,h}\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right) p(y) \sqrt{\det g}\, dy$.

Definition 2 (Kernel-based approximation of the Laplacian). *We introduce the following kernel-based averaging operator $A_{\lambda,h}$:*

$$(A_{\lambda,h} f)(x) = \frac{1}{\tilde{d}_{\lambda,h}(x)} \int_M \tilde{k}_{\lambda,h}\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right) f(y) p(y) \sqrt{\det g}\, dy \tag{4}$$

and the approximation of the Laplacian $\Delta_{\lambda,h} f := \frac{1}{h^2}\left(f - A_{\lambda,h} f\right)$.

A very useful tool in the proof of our main theorems is the following Proposition of Smolyanov et al.[9], which locally relates the extrinsic distance in $\mathbb{R}^d$ with the intrinsic distance $d_M(x, y)$ of the manifold.

Proposition 1. *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 $y \in V$:*

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

where Π is the second fundamental form of M and γ the unique geodesic from x to $\Psi(y)$ such that $\dot{\gamma} = y^i \partial_{y^i}$.

---

[8] In order to avoid problems with differentiation the argument of the kernel function will be the squared norm.

[9] But note that we do not require the kernel to be positive definite and we integrate with respect to the natural volume element.

The volume form $dV = \sqrt{\det g_{ij}(y)}dy$ of M satisfies in normal coordinates

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

in particular $(\Delta \sqrt{\det g_{ij}})(0) = -\frac{1}{3}R$, where R is the scalar curvature (i.e., $R = g^{ik} g^{jl} R_{ijkl}$).

The following proposition describes the asymptotic expression of the convolution parts in the averaging operators $A_{\lambda,h}$. This result is interesting in itself since it shows the interplay between intrinsic and extrinsic geometry of the submanifold if one averages locally. The proof is similar to that of [10], but we now use general kernel functions, which makes the proof a little bit more complicated. We define $C_1 = \int_{\mathbb{R}^m} k(\|y\|^2)dy < \infty, \quad C_2 = \int_{\mathbb{R}^m} k(\|y\|^2) y_1^2 dy < \infty$.

Proposition 2. *Let M and k satisfy Assumptions 1 and 2. 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 $h_0(x) > 0$ for any $f \in C^3(M)$ such that for all $h < h_0(x)$,*

$$\begin{aligned}
&\int_M k_h\left(\|i(x) - i(y)\|^2_{\mathbb{R}^d}\right) f(y) p(y) \sqrt{\det g}\, dy = C_1 p(x) f(x) \\
&\quad + \frac{h^2}{4} C_2 \Big(p(x) f(x) \Big[-R + \frac{1}{2} \Big\| \sum_a \Pi(\partial_a, \partial_a) \Big\|^2_{T_{i(x)}\mathbb{R}^d} \Big] + 2(\Delta_M(pf))(x) \Big) + O(h^3),
\end{aligned}$$

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

Proof: See appendix. □

Now we are ready to formulate the asymptotic result for the operator $\Delta_{\lambda,h}$, which extends the result of Lafon mentioned before.

Theorem 1. *Let M and k satisfy Assumptions 1 and 2. Furthermore let k now have compact support on $[0, R^2]$*[10] *and let P have a density p with respect to the natural volume element which satisfies $p \in C^3(M)$ and $p(x) > 0$, for any $x \in M$. Then for any $\lambda \geq 0$, for any $x \in M \backslash \partial M$, there exists an $h_1(x) > 0$ for any $f \in C^3(M)$ such that for all $h < h_1(x)$,*

$$\begin{aligned}
(\Delta_{\lambda,h} f)(x) &= -\frac{C_2}{2\,C_1} \left((\Delta_M f)(x) + \frac{s}{p(x)} \langle \nabla p, \nabla f \rangle_{T_x M} \right) + O(h^2) \\
&= -\frac{C_2}{2\,C_1} (\Delta_s f)(x) + O(h^2),
\end{aligned} \tag{5}$$

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

Proof: The need for compactness of the kernel k comes from the fact that the modified kernel $\tilde{k}$ depends on $p_h(y)$. Now we can use the Taylor expansion of Proposition 2 for $p_h(y)$ only for h in the interval $(0, h_0(y))$. Obviously it can

[10] That means $k(t) = 0$, if $t > R^2$.

happen that $h_0(y) \to 0$ when we approach the boundary. Therefore, when we have to control $h_0(y)$ over the whole space M, the infimum could be zero, so that the estimate holds for no h. By restricting the support of the kernel k to a compact set $[0, R^2]$, it can be directly seen from the proof of Proposition 2 that $h_0(y)$ has the form $h_0(y) = \epsilon(y)/R$, where $\epsilon(y) = \frac{1}{3}\min\{r(y), \mathrm{inj}(y)\}$. Now $h_0(x)$ is continuous since $r(x)$ is continuous by assumption and $\mathrm{inj}(x)$ is continuous on the compact subset $\overline{B(x, 2\epsilon)}$, see [6][Prop. 2.1.10]. Therefore we conclude that since $h_0(y)$ is continuous on $B(x, 2\epsilon)$ and $h_0(y) > 0$, $h_1(x) = \inf_{y \in \overline{B(x,2\epsilon)}} h_0(y) > 0$. Then for the interval $(0, h_1(x))$ the estimate for $p_h(y)$ holds uniformly over the whole ball $B(x, \epsilon)$. That is, using the definition of $\tilde{k}$ as well as Proposition 2 and the expansion $\frac{1}{(a+h^2 b)^\lambda} = \frac{1}{a^\lambda} - \lambda \frac{h^2 b}{a^{\lambda+1}} + O(h^4)$ we get for $h \in (0, h_1(x))$ that

$$\begin{aligned}
&\int_M \tilde{k}_{\lambda,h}\left(\|i(x) - i(y)\|^2\right) f(y) p(y) \sqrt{\det g}\, dy \\
=&\frac{1}{p_h^\lambda(x)} \int_{B(x,\epsilon)} k_h\left(\|i(x) - i(y)\|^2\right) f(y) \\
&\qquad \left[\frac{C_1 p(y) - \lambda/2 C_2 h^2 (p(y) S + \Delta p)}{C_1^{\lambda+1} p(y)^\lambda} + O(h^3)\right] \sqrt{\det g}\, dy
\end{aligned} \tag{6}$$

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

$$\begin{aligned}
\tilde{d}_{\lambda,h}(x) =&\frac{1}{p_h^\lambda(x)} \int_{B(x,\epsilon)} k_h\left(\|i(x) - i(y)\|^2\right) \\
&\qquad \left[\frac{C_1 p(y) - \lambda/2 C_2 h^2 (p(y) S + \Delta p)}{C_1^{\lambda+1} p(y)^\lambda} + O(h^3)\right] \sqrt{\det g}\, dy
\end{aligned} \tag{7}$$

as an estimate for $\tilde{d}_{\lambda,h}(x)$. Now using Proposition 2 again we arrive at:

$$\begin{aligned}
\Delta_{\lambda,h} f =&\frac{f - A_{\lambda,h} f}{h^2} = \frac{1}{h^2} \frac{\tilde{d}_{\lambda,h} f - \tilde{d}_{\lambda,h} A_{\lambda,h} f}{\tilde{d}_{\lambda,h}} \\
=& -\frac{C_2}{2\, C_1}\left(\Delta_M f + \frac{2(1-\lambda)}{p} \langle \nabla p, \nabla f \rangle\right) + O(h^2)
\end{aligned}$$

where all $O(h^2)$-terms are finite on $B(x, \epsilon)$ since p is strictly positive. □

Note that the limit of $\Delta_{\lambda,h}$ 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, so the reader should be aware of the sign convention. From the last lines of the previous proof, it is easy to deduce the following result for the unnormalized case. Let

$$(\Delta'_{\lambda,h} f)(x) = \frac{1}{h^2}\left(\tilde{d}_{\lambda,h}(x) f(x) - \int_M \tilde{k}_{\lambda,h}\left(\|i(x) - i(y)\|^2\right) f(y) p(y) \sqrt{\det g} dy\right). \tag{8}$$

Corollary 1. *Under the assumptions of Theorem 1, for any $\lambda \geq 0$, any $x \in M \backslash \partial M$, any $f \in C^3(M)$ there exists an $h_1(x) > 0$ such that for all $h < h_1(x)$,*

$$(\Delta'_{\lambda,h} f)(x) = -p(x)^{1-2\lambda} \frac{C_2}{2C_1^{\lambda}} (\Delta_s f)(x) + O(h^2), \quad \text{where} \quad s = 2(1-\lambda). \tag{9}$$

This result is quite interesting. We observe that in the case of a uniform density it does not make a difference whether we use the unnormalized or the normalized approximation of the Laplacian. However, as soon as we have a non-uniform density, the unnormalized one will converge only up to a function to the Laplacian, except in the case $\lambda = \frac{1}{2}$ where both the normalized and unnormalized approximation lead to the same result. This result confirms the analysis of von Luxburg et al. in [11], where the consistency of spectral clustering was studied. There the unnormalized Laplacian is in general not consistent since it has a continuous spectrum. Obviously the limit operator $\Delta'_{\lambda,h} = -p^{1-2\lambda} \frac{C_2}{2C_1^{\lambda}} \Delta_s$ has also a continuous spectrum even if Δ_s is compact since it is multiplied with $p^{1-2\lambda}$.

4 Strong Pointwise Consistency of Graph Laplacians

In the last section we identified certain averaging operators $\Delta_{\lambda,h}$ which in the limit $h \to 0$ converge pointwise to the corresponding Laplacian Δ_s, where $s = 2(1-\lambda)$. In this section we will provide the connection to the normalized graph Laplacian $\Delta_{\lambda,n,h}$ with data-dependent weights $\tilde{w}_\lambda(X_i, X_j)$ defined as

$$\tilde{w}_\lambda(X_i, X_j) = \frac{k\left(\left\|i(X_i) - i(X_j)\right\|^2 / h^2\right)}{[d(X_i) d(X_j)]^{\lambda}}, \quad \lambda \geq 0 \tag{10}$$

where $d(X_i) = \sum_{r=1}^{n} k(\left\|i(X_i) - i(X_r)\right\|^2 / h^2)$. Note that the weights are not multiplied with $1/h^m$, as it was usual for the kernel function in the last section. There are two reasons for this. The first one is that this factor would lead to infinite weights for $h \to 0$. The second and more important one is that this factor cancels for the normalized Laplacian. This is very important in the case where the data lies on a submanifold of unknown dimension m, since then also the correct factor $\frac{1}{h^m}$ would be unknown. Note also that for the unnormalized Laplacian this factor does not cancel if $\lambda \neq \frac{1}{2}$. This means that for $\lambda \neq \frac{1}{2}$ the unnormalized Laplacian cannot be consistently estimated if the data lies on a proper submanifold of unknown dimension, since the estimate in general blows up or vanishes. Therefore we will consider only the normalized graph Laplacian in the following and for simplicity omit the term 'normalized'.

The graph Laplacian is defined only for functions on the graph, but it is straightforward to extend the graph Laplacian to an estimator of the Laplacian for functions defined on the whole space by using the kernel function,

$$(\Delta_{\lambda,h,n} f)(x) = \frac{1}{h^2} \left(f - A_{\lambda,h,n} f\right)(x) := \frac{1}{h^2} \left(f(x) - \frac{1}{\tilde{d}_\lambda(x)} \sum_{j=1}^{n} \tilde{w}_\lambda(x, X_j) f(X_j) \right), \tag{11}$$

where $\tilde{d}_\lambda(x) = \sum_{r=1}^n \tilde{w}_\lambda(x, X_i)$. The factor $\frac{1}{h^2}$ comes from introducing an $\frac{1}{h}$-term in the definition of the derivative operator d on the graph. It is natural to introduce this factor since we want to estimate a derivative. Especially interesting is the form of the second term of the graph Laplacian for $\lambda = 0$ where the weights are not data-dependent. In this case, this term can be identified with the Nadaraya-Watson regression estimate. Therefore, for $\lambda = 0$ we can adapt the proof of pointwise consistency of the Nadaraya-Watson estimator of Greblicki, Krzyzak and Pawlak [5] and apply it to the graph Laplacian. The following Lemma will be useful in the following proofs.

Lemma 1. *Let $X_1, \ldots, X_n$ be n i.i.d. random vectors in $\mathbb{R}^d$ with law P, which is absolutely continuous with respect to the natural volume element dV of a submanifold $M \subset \mathbb{R}^d$ satisfying Assumption 1. Let p denote its density, which is bounded, continuous and positive $p(x) > 0$, for any $x \in M$. Furthermore let k be a kernel with compact support on $[0, R^2]$ satisfying Assumption 2. Let $x \in M \backslash \partial M$, define $b_1 = \|k\|_\infty \|f\|_\infty$, $b_2 = C \|k\|_\infty \|f\|_\infty^2$, where C is a constant depending on x, $\|p\|_\infty$ and $\|k\|_\infty$. Then for any $f \in C^3(M)$,*

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n k_h(\|i(x) - i(X_i)\|^2) f(X_i) - \int_M k_h(\|i(x) - i(y)\|^2) f(y) p(y) \sqrt{\det g}\, dy\Big| \geq \epsilon\Big)$$
$$\leq 2 \exp\left(-\frac{n h^m \varepsilon^2}{2b_2 + 2b_1\varepsilon/3}\right)$$

Now the proof of pointwise consistency in the case $\lambda = 0$ is straightforward.

Theorem 2 (Weak and strong pointwise consistency for $\lambda = 0$). *Let $X_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ be random vectors drawn i.i.d. from the probability measure P on $M \subset \mathbb{R}^d$, where M satisfies Assumption 1 and has $\dim M = m$. Furthermore let P be absolutely continuous with respect to the volume element dV with density $p \in C^3(M)$ and $p(x) > 0$, $\forall\, x \in M$, and let $\Delta_{0,h,n}$ be the graph Laplacian in (11) with weights of the form (10), where k has compact support on $[0, R^2]$. Then for every $x \in M \backslash \partial M$ and for every function $f \in C^3(M)$, if $h \to 0$ and $nh^{m+4} \to \infty$*

$$\lim_{n\to\infty} (\Delta_{0,h,n} f)(x) = -\frac{2\,C_1}{C_2} (\Delta_2 f)(x) \quad \textit{in probability.}$$

If even $nh^{m+4}/\log n \to \infty$, then almost sure convergence holds.

Proof: We rewrite the estimator $\Delta_{0,h,n} f$ in the following form

$$(\Delta_{0,h,n} f)(x) = \frac{1}{h^2}\left[f(x) - \frac{(A_{0,h}\, f)(x) + B_{1n}}{1 + B_{2n}}\right] \tag{12}$$

where

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

In Theorem 1 we have shown that

$$\lim_{h\to 0}(\Delta_{0,h} f)(x) = \lim_{h\to 0} \frac{1}{h^2}\left[f(x) - (A_{0,h}\, f)(x)\right] = -\frac{2\, C_1}{C_2}(\Delta_2 f)(x) \tag{13}$$

Let $hR \leq \mathrm{inj(x)}$, then $\mathbb{E}_Z\, k_h(\|i(x)-i(Z)\|^2) \geq K \inf_{y\in B_M(x,hR)} p(y)$, where K is a constant and using Lemma 1 we get with $d_2 = \frac{\|f\|_\infty^2}{(K \inf_{y\in B_M(x,hR)} p(y))^2}$, $d_1 = \frac{\|f\|_\infty}{K \inf_{y\in B_M(x,\epsilon)} p(y)}$:

$$\mathrm{P}(|B_{1n}| \geq h^2 t) \leq \exp\left(-\frac{nh^{m+4}\, t^2}{2\, \|k\|_\infty\, (d_2 + t\, d_1/3)}\right),$$

Note that since p is continuous and p is strictly positive the infimum is achieved and positive. The same analysis can be done for B_{2n}, where we do not have to deal with the $1/h^2$-factor. 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 $nh^{d+4}/\log n \to +\infty$ when $n\to\infty$. □

Under the more restrictive assumption that the data is sampled from a uniform probability measure on a compact submanifold we learned that Belkin and Niyogi have independently proven the convergence of the unnormalized graph Laplacian in [2]. It is clear from Theorem 1 and Corollary 1 that in the case of a uniform measure the limit operators for normalized and unnormalized graph Laplacian agree up to a constant. However, as mentioned before the unnormalized graph Laplacian has the disadvantage that in order to get convergence one has to know the dimension m of the submanifold M, which in general is not the case.

Lemma 2. *Let $X_i \in \mathbb{R}^d$, $i=1,\ldots,n$ be random vectors drawn i.i.d. from the probability measure P on $M \subset \mathbb{R}^d$, where M satisfies Assumption 1 and has* $\dim M = m$*. Furthermore let P be absolutely continuous with respect to the volume element dV with continuous density $p(x)$. Let $k(\|x-y\|^2)$ be a bounded kernel with compact support on $[0,R^2]$. Let $\lambda \geq 0$, $x \in M$ with $p(x) > 0$, $f \in C(M)$ and $n \geq 2$. Then there exists a constant $C > 1$ such that for any $0 < \varepsilon < 1/C$, $0 < h < \frac{1}{C}$ with probability at least $1 - Cne^{-\frac{nh^m\varepsilon^2}{C}}$, we have*

$$|(A_{\lambda,h,n} f)(x) - (A_{\lambda,h} f)(x)| \leq \varepsilon.$$

Proof: For sufficiently large C, the assertion of the lemma is trivial for $\varepsilon < \frac{2\|k\|_\infty}{(n-1)h^m}$. So we will only consider $\frac{2\|k\|_\infty}{(n-1)h^m} \leq \varepsilon \leq 1$. The idea of the proof is to use deviation inequalities to show that the empirical terms, which are expressed as a sum of i.i.d. random variables, are close to their expectations. Then we can prove that the empirical term

$$(A_{\lambda,h,n} f)(x) = \frac{\sum_{j=1}^n k_h(\|i(x)-i(X_j)\|^2) f(X_j) [d(X_j)]^{-\lambda}}{\sum_{r=1}^n k_h(\|i(x)-i(X_r)\|^2) [d(X_r)]^{-\lambda}} \tag{14}$$

is close to the term $(A_{\lambda,h}f)(x)$. Consider the event $\mathcal{E}$ for which we have

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

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 1 (with $N=n-1$ and the conditional probability wrt X_j for a given $1\le j\le d$) to obtain that for $\varepsilon\le 1$,

$$\mathrm{P}\Big(\Big|\tfrac{1}{n-1}\textstyle\sum_{i\ne j}k_h(\|i(x)-i(X_i)\|^2)-p_h(x)\Big|\ge\varepsilon\Big|X_j\Big)\le 2e^{-\frac{(n-1)h^m\varepsilon^2}{C}}.$$

First integrating wrt to the law of X_j and then using an union bound we get

$$\mathrm{P}\Big(\Big|\tfrac{1}{n-1}\textstyle\sum_{i\ne j}k_h(\|i(x)-i(X_i)\|^2)-p_h(x)\Big|\ge\varepsilon\Big)\le 2e^{-\frac{(n-1)h^m\varepsilon^2}{C}}\ \text{and}$$

$$\mathrm{P}\Big(\text{for any } j\in\{1,\dots,n\},\ \Big|\tfrac{d(X_j)}{n-1}-\tfrac{k_h(0)}{n-1}-p_h(X_j)\Big|\le\varepsilon\Big)\ge 1-2ne^{-\frac{(n-1)h^m\varepsilon^2}{C}}.$$

Therefore for $\frac{2\|k\|_\infty}{(n-1)h^m}\le\varepsilon\le 1$ we have[11]

$$\mathrm{P}\Big(\text{for any } j\in\{1,\dots,n\},\ \Big|\tfrac{d(X_j)}{n-1}-p_h(X_j)\Big|\le\varepsilon\Big)\ge 1-2ne^{-\frac{(n-1)h^m\varepsilon^2}{C}}.$$

Similarly we can prove that for $\frac{2\|k\|_\infty}{nh^m}\le\varepsilon\le 1$ with probability at least $1-2e^{-Cnh^m\varepsilon^2}$, the third assertion defining $\mathcal{E}$ holds. For the three last assertions, a direct application of Lemma 1 shows that they also hold with high probability. Finally, combining all these results, we obtain that for $\frac{2\|k\|_\infty}{(n-1)h^m}\le\varepsilon\le 1$, the event $\mathcal{E}$ holds with probability at least $1-Cne^{-\frac{nh^m\varepsilon^2}{C}}$. Let us define

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

and let us now work only on the event $\mathcal{E}$. Let $p_{\min}=p(x)/2$ and $p_{\max}=2p(x)$. By continuity of the density, for C large enough and any $h<1/C$, the density satisfies $0<p_{\min}\le p\le p_{\max}$ on the ball $B_M(x,2hR)$. So for any $y\in B_M(x,hR)$,

[11] We recall that the value of the constant C might change from line to line.

there exists a constant $D_1 > 0$ such that $D_1\, p_{\min} \le p_h(y) \le D_1\, \sqrt{2}\, p_{\max}$. Using the first order Taylor formula of $[x \mapsto x^{-\lambda}]$, we obtain that for any $\lambda \ge 0$ and $a, b > \beta$, $\left|a^{-\lambda} - b^{-\lambda}\right| \le \lambda \beta^{-\lambda-1}|a-b|$. So we can write

$$\begin{aligned}
\left|\hat{\mathcal{B}} - \mathcal{B}\right| &\le \left|\tfrac{1}{n}\textstyle\sum_{j=1}^n k_h(\|i(x)-i(X_j)\|^2)\Big(\big[\tfrac{d(X_j)}{n-1}\big]^{-\lambda} - [p_h(X_j)]^{-\lambda}\Big)\right| \\
&\quad + \left|\tfrac{1}{n}\textstyle\sum_{j=1}^n k_h(\|i(x)-i(X_j)\|^2)[p_h(X_j)]^{-\lambda} - \mathcal{B}\right| \\
&\le \left|\tfrac{d(x)}{n}\right| \lambda\, (D_1 p_{\min})^{-\lambda-1}\varepsilon + \varepsilon \\
&\le \left|\tfrac{d(x)}{n} - p_h(x)\right| \lambda\, (D_1 p_{\min})^{-\lambda-1}\varepsilon + p_h(x)\lambda\, (D_1 p_{\min})^{-\lambda-1}\varepsilon + \varepsilon \\
&\le \lambda (D_1 p_{\min})^{-\lambda-1}\varepsilon + \sqrt{2} D_1 p_{\max} \lambda\, (C_1 p_{\min})^{-\lambda-1}\varepsilon + \varepsilon := C'\varepsilon
\end{aligned}$$

Similarly we prove that $\left|\hat{\mathcal{A}} - \mathcal{A}\right| \le C''\varepsilon$. Let $\zeta := \frac{1}{2}\frac{D_1 p_{\min}}{(\sqrt{2}\, D_1 p_{\max})^{\lambda}}$. We have $\mathcal{B} \ge 2\zeta$. Let us introduce $\varepsilon_0 := \min\{\frac{\zeta}{C'}, 1\}$. For $\frac{2\|k\|_\infty}{(n-1)h^m} \le \varepsilon \le \varepsilon_0$, we have also $\hat{\mathcal{B}} \ge \zeta$. Combining the last three results, we obtain that there exists $D_2 > 0$ such that

$$\left|\frac{\mathcal{A}}{\mathcal{B}} - \frac{\hat{\mathcal{A}}}{\hat{\mathcal{B}}}\right| \le \frac{|\mathcal{A} - \hat{\mathcal{A}}|}{\hat{\mathcal{B}}} + \mathcal{A}\frac{|\mathcal{B} - \hat{\mathcal{B}}|}{\mathcal{B}\hat{\mathcal{B}}} \le \frac{C''\varepsilon}{\zeta} + D_2 p_{\max}(C_1 p_{\min})^{-\lambda}\frac{C'\varepsilon}{2\zeta^2} \le C\varepsilon.$$

Noting that $A_{\lambda,h} f = \mathcal{A}/\mathcal{B}$ and $A_{\lambda,h,n} f = \hat{\mathcal{A}}/\hat{\mathcal{B}}$, we have proved that there exists a constant $C > 1$ such that for any $0 < \varepsilon < 1/C$

$$|(A_{\lambda,h,n} f)(x) - (A_{\lambda,h} f)(x)| \le C\varepsilon$$

with probability at least $1 - Cne^{-\frac{nh^m\varepsilon^2}{C}}$. This leads to the desired result. □

Combining Lemma 2 with Theorem 1 we arrive at our main theorem.

Theorem 3 (Weak and strong pointwise consistency). *Let $X_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ be random vectors drawn i.i.d. from the probability measure P on $M \subset \mathbb{R}^d$, where M satisfies Assumption 1 and has $\dim M = m$. Let P be absolutely continuous with respect to the volume element dV with density $p \in C^3(M)$ and p strictly positive. Let $\Delta_{\lambda,h,n}$ be the graph Laplacian in (11) with weights of the form (10), where k has compact support on $[0, R^2]$ and satisfies Assumption 2. Define $s = 2(1-\lambda)$. Then, for every $x \in M \backslash \partial M$ and for every function $f \in C^3(M)$, if $h \to 0$ and $nh^{m+4}/\log n \to \infty$*

$$\lim_{n\to\infty} (\Delta_{\lambda,h,n} f)(x) = -\frac{2\,C_1}{C_2}(\Delta_s f)(x) \quad \textit{almost surely.}$$

Proof: The proof consists of two steps. By Theorem 1 the bias term converges.

$$\lim_{h\to 0}\left|(\Delta_{\lambda,h} f)(x) - \left[-\left(\frac{2\,C_1}{C_2}\Delta_s f\right)(x)\right]\right| \to 0. \tag{15}$$

Next we consider the variance term $|(\Delta_{\lambda,h,n} f)(x) - (\Delta_{\lambda,h} f)(x)|$. We have

$$|(\Delta_{\lambda,h,n} f)(x) - (\Delta_{\lambda,h} f)(x)| = \frac{1}{h^2}\left|(A_{\lambda,h,n} f)(x) - (A_{\lambda,h} f)(x)\right|.$$

Up to the factor $1/h^2$ this is the term studied in Lemma 2, so that we get under the conditions stated there:

$$\mathrm{P}\Big(\Big|(\Delta_{\lambda,h,n}f)(x) - (\Delta_{\lambda,h}f)(x)\Big| \geq \epsilon\Big) \leq C\,n\,e^{-\frac{nh^{m+4}\,\epsilon^2}{C}}$$

Then, using the same technique as in Theorem 2, one shows complete convergence for $nh^{m+4}/\log n \to \infty$, which implies almost sure convergence. □

This theorem states conditions for the relationship of the sample size n and the bandwidth h for almost sure convergence. It is unlikely that this rate can be improved (up to the logarithmic factor), since the rates for estimating second derivatives in nonparametric regression are the same. Another point which cannot be underestimated is that we show that the rate that one gets only depends on the intrinsic dimension m of the data (that is the dimension of the submanifold M). This means that even if one has data in a very high-dimensional Euclidean space $\mathbb{R}^d$ one can expect to get a good approximation of the Laplacian if the data lies on a low-dimensional submanifold. Therefore, our proof provides a theoretical basis for all algorithms performing dimensionality reduction using the graph Laplacian. Another point is that one can continuously control the influence of the probability distribution with the parameter λ and even eliminate it in the case $\lambda = 1$. The conditions of this theorem are very mild. We only require that the submanifold is not too much twisted and that the kernel is bounded and compact. Note that in large scale practical applications, compactness of the kernel is necessary for computational reasons anyway.

Acknowledgments

We would like to thank Olaf Wittich for his help with [9, 10] and Bernhard Schölkopf for helpful comments and support.

References

[1] M. Belkin. *Problems of Learning on Manifolds.* PhD thesis, University of Chicago, 2003. http://www.people.cs.uchicago.edu/~misha/thesis.pdf.
[2] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. to appear in the Proc. of COLT, 2005.
[3] O. Bousquet, O. Chapelle, and M. Hein. Measure based regularization. In *Advances in Neural Information Processing Systems*, volume 16, 2003.
[4] S. Coifman and S. Lafon. Diffusion maps. Preprint, Jan. 2005, to appear in Applied and Computational Harmonic Analysis, 2005.
[5] W. Greblicki, A. Krzyzak, and M. Pawlak. Distribution-free pointwise consistency of kernel regression estimate. *Annals of Statistics*, 12:1570–1575, 1984.
[6] W. Klingenberg. *Riemannian Geometry.* De Gruyter, 1982.
[7] S. S. Lafon. *Diffusion Maps and Geometric Harmonics.* PhD thesis, Yale University, 2004. http://www.math.yale.edu/~sl349/publications/dissertation.pdf.

[8] P. Niyogi. Learning functional maps on Riemannian submanifolds. Talk presented at IPAM workshop on multiscale structures in the analysis of high-dimensional data, 2004. available at http://www.ipam.ucla.edu/publications/mgaws3/mgaws3_5188.pdf.
[9] 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*, 2000.
[10] O. G. Smolyanov, H. von Weizsäcker, and O. Wittich. Chernoff's theorem and discrete time approximations of Brownian motion on manifolds. Preprint, available at http://lanl.arxiv.org/abs/math.PR/0409155, 2004.
[11] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. Technical Report 134, Max Planck Institute for Biological Cybernetics, 2004.
[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS 16*, 2004.
[13] D. Zhou, B. Schölkopf, and T. Hofmann. Semi-supervised learning on directed graphs. In *NIPS 17*, 2005.

A Appendix: Proof of Proposition 2

The following lemmas are needed in the proof of the asymptotics of $A_{\lambda,h}$.

Lemma 3. *If the kernel $k : \mathbb{R} \to \mathbb{R}$ satisfies the assumptions in Assumption 2,*

$$\int_{\mathbb{R}^m} \frac{\partial k}{\partial x}(\|u\|^2) u^i u^j u^k u^l du = -\frac{1}{2} C_2 \left[\delta^{ij}\delta^{kl} + \delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk} \right]. \tag{16}$$

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

$$\Big| \int_{B(0,r_{\min})} k_h \left(\frac{\|z\|^2 + V_{ijkl} z^i z^j z^k z^l + \beta \|z\|^5)}{h^2} \right) f(z) dz \\ - \left(C_1 f(0) + C_2 \frac{h^2}{2} \Big[(\Delta f)(0) - f(0) \sum_{i,k}^{m} V_{iikk} + V_{ikik} + V_{ikki} \Big] \right) \Big| \leq C h^3. \tag{17}$$

To prove Proposition 2, let $\epsilon = \frac{1}{3} \min\{\mathrm{inj}(x), r(x)\}^{12}$, where ϵ is positive by the assumptions on M. Then we decompose M in $M = B(x, \epsilon) \cup (M \backslash B(x, \epsilon))$ and integrate separately. The integral over $M \backslash B(x, \epsilon)$ can be estimated by using the definition of $\delta(x)$ (see Assumption 1) and the fact that k is non-increasing:

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

[12] The factor 1/3 is needed in Theorem 1.

Since $\delta(x)$ is positive by the assumptions on M 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 $B(x,\epsilon)$. Since ϵ is smaller than the injectivity radius $\mathrm{inj}(x)$, we can introduce normal coordinates $z = \exp^{-1}(y)$ with origin $0 = \exp^{-1}(x)$ on $B(x,\epsilon)$, so that we can write the integral over $B(x,\epsilon)$ as:

$$\int\limits_{B(0,\epsilon)} k_h\left(\frac{\|z\|^2 - \frac{1}{12}\sum_{\alpha=1}^{d}\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^u \partial z^v} z^a z^b z^u z^v + O(\|z\|^5)}{h^2}\right) p(z) f(z) \sqrt{\det g}\, dz \tag{18}$$

by using our assumption that $pf\sqrt{\det g}$ is in $C^3(B(0,\epsilon))$. Therefore we can apply Lemma 4 and compute the integral in (18) which results in:

$$\begin{aligned}&\Big[p(0)f(0)\Big(C_1 + C_2\frac{h^2}{24}\sum_{\alpha=1}^{d}\frac{\partial^2 i^\alpha}{\partial z^a \partial z^b}\frac{\partial^2 i^\alpha}{\partial z^c \partial z^d}\Big[\delta^{ab}\delta^{cd} + \delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc}\Big]\Big)\\ &+ C_2\frac{h^2}{2}\Delta_M(pf\sqrt{\det g})\Big|_0 + O(h^3)\Big],\end{aligned} \tag{19}$$

where we have used that the Laplace-Beltrami operator Δ_M in normal coordinates z^i at 0 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 [10–Proposition 6].

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

where R is the scalar curvature. Plugging this result into (19) and using from Proposition 1, $\Delta_M\sqrt{\det g}\big|_0 = -\frac{1}{3}R$, finishes the proof.