

Diffusion Maps

University of Vienna



Peter Mühlbacher

October 1, 2014

Abstract

Contents

0.1	Dimensionality Reduction	2
0.1.1	The Problem	2
0.1.2	Existing Approaches	2
1	Diffusion Maps	3
1.1	Diffusion Kernels	3
1.1.1	Motivation	3
1.1.2	Construction of a Random Walk on the Data	4
1.1.3	Diffusion Kernels	6
1.1.4	Embedding in the Euclidian Space	7
2	Connection with the Fokker-Planck Equation	10
2.1	A Family of Anisotropic Diffusion Maps	10
2.1.1	Construction of the Family of Diffusions	10
2.1.2	Transition to Diffusion Processes	12
A	Asymptotics for Laplacian Operators	13
A.1	Comparison of the Geodesic and the Local Projection	13
A.2	Infinitesimal Operators for a Family of Graph Laplacians	15

Introduction

The contentual layout follows that of Coifman and Lafon (2006) and ?. Some definitions, theorems and examples will be taken from these works as well.

0.1 Dimensionality Reduction

0.1.1 The Problem

0.1.2 Existing Approaches

Linear PCA

Kernel PCA

Chapter 1

Diffusion Maps

1.1 Diffusion Kernels

1.1.1 Motivation

In this section it will be elaborated on why global structures need not be preserved and how this leads to diffusion processes.

Preservation of Global Structures

Suppose data points $x_i \in \mathbb{R}^k$ are generated by a low dimensional parameter $\theta_i \in \mathbb{R}^{k'}$, $k' \ll k$ given a map $\Phi : \mathbb{R}^{k'} \rightarrow \mathbb{R}^k$. One problem is the (numerical) smoothness of Φ which is not necessarily given. Another problem is that for large k the euclidian distance is no longer a meaningful measure as the volume of the k -dimensional unit ball $\frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)}$ converges to 0 as $\lim_{k \rightarrow \infty}$ which is also known as the *curse of dimensionality*. One may conclude that large distances in the ambient space need not necessarily be preserved as they do not hold much information except that x_i is not “very close” to x_j .

Analogous to Riemannian manifolds (where metric tensors, inducing an inner product on the tangent space and a metric via the exponential map, define the manifold’s geometry) we focus solely on local distances in order to recover intrinsic global structures.

A Dual Approach

From inverse problems in spectral geometry (e.g. “Can One Hear the Shape of a Drum?”) it is known that much¹ of the geometry of a given set Γ can be derived from the analysis of functions defined on Γ .

In this work eigenvalues and eigenfunctions of averaging operators, i.e., operators whose kernel corresponds to transition probabilities of a Markov process, will be studied in order to define a diffusion map which embeds the data into a Euclidian space where the Euclidian distance is just the diffusion metric.

1.1.2 Construction of a Random Walk on the Data

Definitions

Let $(\Gamma, \mathcal{A}, \mu)$ be a measure space. In practical applications Γ is the given data set consisting of finitely many data points and μ is the counting measure to represent the distribution of the points in the data set. In addition, suppose we are given a symmetric kernel $k : \Gamma \times \Gamma \rightarrow \mathbb{R}^+$ which defines the local geometry of Γ .

Examples

Usually Γ is either a subset of the Euclidian space or a weighted graph.

In the first case it seems natural to write k as a function of the Euclidian distance $\nu(\|x - y\|)$.

In the second case let $b(x, y)$ be the associated adjacency matrix, that is, $b(x, y) = 1$ if there is an edge going from x to y , and $b(x, y) = 0$ otherwise. The kernel b defines a notion of neighborhood for each point, and also a non-symmetric distance given by $1 - b(x, y)$. Clearly b is not symmetric in general, but we can consider

$$k_1(x, y) := \int_{\Gamma} b(x, u)b(y, u)d\mu(u)$$

$$k_2(x, y) := \int_{\Gamma} b(u, x)b(u, y)d\mu(u)$$

where $k_1(x, y)$ counts the number of common neighbors to x and y , while $k_2(x, y)$ counts the number of nodes for which x and y are common neighbors.

¹the most famous example being Weyl’s proof of $\#\{\lambda_k : \lambda_k < \lambda\} \approx \frac{\text{area}(\Gamma)}{2\pi}\lambda$ as $\lambda \rightarrow \infty$.
(?)

Normalized Graph Laplacian Construction

Generally, such a kernel represents some notion of affinity between points of Γ and thus one can think of the data points as being the nodes of a symmetric graph whose weight function is specified by k . From the graph defined by (Γ, k) , one can construct a reversible Markov chain on Γ .

To normalize the kernel, define

Definition 1.1.

$$v^2(x) = \int_{\Gamma} k(x, y) d\mu(y)$$

and

Definition 1.2.

$$p(x, y) = \frac{k(x, y)}{v^2(x)}.$$

$p(x, y)$ is no longer symmetric, but inherited the positivity and now satisfies a conservation property:

$$\int_{\Gamma} p(x, y) d\mu(y) = 1$$

As a result the matrix $P := (p(i, j))_{i,j}$ is stochastic and can be interpreted as the transition matrix of a homogeneous Markov process on Γ . In spectral graph theory $\mathbb{I} - P$ is commonly referred to as normalized, weighted graph Laplacian. This naming is justified in A.2.1².

To investigate the spectral properties of the corresponding integral operator P defined by $Pf(x) = \int_{\Gamma} p(x, y)f(y)d\mu(y)$ it is beneficial to examine the symmetric, conjugated Operator A .

Definition 1.3. *Let*

$$a(x, y) = \frac{k(x, y)}{v(x)v(y)} = v(x)p(x, y)\frac{1}{v(y)},$$

then the corresponding diffusion operator A is defined as

$$Af(x) = \int_{\Gamma} a(x, y)f(y)d\mu(y)$$

Notice that by definition of a one obtains a symmetric form and thus a symmetric operator A .

²Let w.l.o.g. $m_0 = 1$ and $m_2 = 2$. Now set $L_{\varepsilon} = \frac{\varepsilon I - G_{\varepsilon}}{\varepsilon}$; for uniformly sampled data the potential term ω vanishes and $\lim_{\varepsilon \rightarrow 0} L_{\varepsilon} = \Delta$

better
double
check that
claim in
the foot-
note

1.1.3 Diffusion Kernels

Theorem 1.1.1 (Spectral Properties of the Diffusion Operator). *The diffusion operator A with kernel a is bounded from $L^2(\Gamma, d\mu)$ into itself, symmetric and positive semi-definite.*

Moreover, its norm is

$$\|A\| = 1$$

and is taken by the eigenfunction

$$Av = v.$$

Proof. Let $f \in L^2(\Gamma, d\mu)$. We have:

$$\langle Af, f \rangle = \int_{\Gamma^2} k(x, y) \frac{f(x)}{v(x)} \frac{f(y)}{v(y)} d\mu(x) d\mu(y). \quad (1.1)$$

Applying the Cauchy-Schwartz inequality we get:

$$\begin{aligned} \left| \int_{\Gamma} k(x, y) \frac{f(y)}{v(y)} d\mu(y) \right| &= \left(\int_{\Gamma} k(x, y) d\mu(y) \right)^{\frac{1}{2}} \left(\int_{\Gamma} k(x, y) \frac{f(y)^2}{v(y)^2} d\mu(y) \right)^{\frac{1}{2}} \\ &= v(x) \left(\int_{\Gamma} k(x, y) \frac{f(y)^2}{v(y)^2} d\mu(y) \right)^{\frac{1}{2}} \end{aligned}$$

Hence:

$$\langle Af, f \rangle \leq \int_{\Gamma} |f(x)| \left(\int_{\Gamma} k(x, y) \frac{f(y)^2}{v(y)^2} d\mu(y) \right)^{\frac{1}{2}} d\mu(x)$$

and by using the Cauchy-Schwartz inequality once again:

$$\langle Af, f \rangle \leq \|f\| \left(\int_{\Gamma^2} \frac{k(x, y)}{v(y)^2} f(y)^2 d\mu(y) d\mu(x) \right)^{\frac{1}{2}} = \|f\|^2$$

by symmetry of the kernel which, in combination with (1.1), also implies the positivity of A .

Plugging in v for f it follows immediately that the eigenvalue 1 is actually obtained and v is an eigenfunction. \square

Theorem 1.1.2 (Spectral Decomposition of the Diffusion Kernel). *Assuming A is compact³ and $A\phi_l = \lambda_l \phi_l$ we may write the kernel as*

$$a(x, y) = \sum_{l \geq 0} \lambda_l \phi_l(x) \phi_l(y)$$

with $\lambda_0 = 1$ and $\lim_{l \rightarrow \infty} \lambda_l = 0$ monotonically.

³which is no constraint in practice since data is finite

try to
general-
ize lemma
3.4 of ?
(guess
it won't
work
though)

Proof. First, note that A being compact implies that the spectrum is discrete and the sum thus is well defined. By A being symmetric and compact the spectral theorem applies and we get that there exists a sequence of real eigenvalues λ_l converging to 0. The corresponding normalized eigenvectors ϕ_l form an orthonormal set and every $f \in L^2(\Gamma, d\mu)$ can be written as

$$f = \sum_{l \geq 0} \langle \phi_l, f \rangle \phi_l + h$$

where $h \in \text{Ker}(A)$.

It follows that

$$Af(x) = \int_{\Gamma} a(x, y) f(y) d\mu(y) = \sum_{l \geq 0} \lambda_l \int_{\Gamma} \phi_l(y) f(y) d\mu(y) \phi_l(x)$$

which, by linearity of the integral and comparison of components, is just what we were looking for. \square

“Komponentenvergleich
auf Englisch
finden

From definition of $a(x, y)$ we see that

$$p(x, y) = \sum_{l \geq 0} \lambda_l \underbrace{\frac{\phi_l(x)}{v(x)}}_{=: \psi_l(x)} \phi_l(y) v(y) \quad (1.2)$$

which enables us to efficiently compute t th powers p_t of p .

There are two ways to interpret p_t :

1. p_t has a probabilistic interpretation as the probability for a Markov chain with transition matrix P to reach y from x in t steps.
2. the dual point of view is that of the functions defined on the data. The kernel p_t can be viewed as a bump or more precisely, if $x \in \Gamma$ is fixed, then $p_t(x, \cdot)$ is a bump function centered at x and of width increasing with t which intuitively captures the idea of diffusion.

1.1.4 Embedding in the Euclidian Space

Definition 1.4. Let

$$D_t(x, y)^2 = \|p_t(x, \cdot) - p_t(y, \cdot)\|_{L^2(\Gamma, d\mu/v)}^2 = \int_{\Gamma} (p_t(x, u) - p_t(y, u))^2 \frac{d\mu(u)}{v(u)}$$

be the family of diffusion distances parameterized by t .

For a fixed value of t D_t defines a distance on the set Γ which is small only if there is a large number of small paths connecting x and y (i.e. if there is a large probability of getting from x to y in t steps). It thus emphasizes the notion of a cluster.

Another property following from the summation over all possible paths is that this distance is very robust to noise perturbation (in contrast to the geodesic distance).

Theorem 1.1.3 (A Numerically Feasible Representation).

$$D_t(x, y) = \left(\sum_{l \geq 0} \left(\lambda_l^t (\psi_l(x) - \psi_l(y)) \right)^2 \right)^{\frac{1}{2}}$$

Proof. $\{\phi_l\}_{l \geq 0}$ forming an orthonormal basis for $L^2(\Gamma, d\mu)$ implies that $\{\phi_l v\}_{l \geq 0}$ is an orthonormal basis for $L^2(\Gamma, d\mu/v)$ and thus for x fixed (1.2) may be seen as orthogonal expansion of the function $y \mapsto p_t(x, y)$ into the basis $\{\phi_l v\}_{l \geq 0}$. The coefficients are given by $\{\lambda_l^t \psi_l(x)\}_{l \geq 0}$. The statement follows directly using the Pythagorean theorem. \square

An immediate consequence is that the diffusion distance is well approximable and that it converges towards a function of (numerical) rank 1 as $t \rightarrow \infty$ because of the vanishing influence of all eigenvectors with eigenvalues < 1 .

One possible interpretation is that $D_t(x, y)$ measures the distance between bumps of “magnitude” t being centered around two points x and y . As t gets larger so does the size of the supports and the number of eigenfunctions needed to calculate $D_t(x, y)$ decreases. This number is related to the minimum number of bumps necessary to cover the set X (like in Weyl’s asymptotic law for the decay of the spectrum).

In order to calculate $D_t(x, y)$ to a preset accuracy $\delta > 0$ with a finite number of terms we set

$$s_t(\delta) = \max\{l \in \mathbf{N} : \lambda_l^t > \delta \lambda_1^t\}$$

so that, up to relative precision δ

$$D_t(x, y) = \left(\sum_{l=0}^{s_t(\delta)} \left(\lambda_l^t (\psi_l(x) - \psi_l(y)) \right)^2 \right)^{\frac{1}{2}}. \quad (1.3)$$

Definition 1.5. Let $\{\Psi_t\}_{t \in \mathbf{N}}$,

$$\Psi_t(x) = \begin{pmatrix} \lambda_1^t \psi_1(x) \\ \lambda_2^t \psi_2(x) \\ \vdots \\ \lambda_{s_t(\delta)}^t \psi_{s_t(\delta)}(x) \end{pmatrix}$$

be the family of diffusion maps. Each component of $\Psi_t(x)$ is termed diffusion coordinate.

According to (1.3) diffusion maps embed data in a Euclidian space in such a way that the Euclidian distance equals the diffusion distance up to a relative error δ .

Chapter 2

Connection with the Fokker-Planck Equation

2.1 A Family of Anisotropic Diffusion Maps

In this section we investigate the probability space (X, \mathcal{A}, μ) and therefore assume an infinite number of data points whose distribution is determined by the probability measure $\mu(x) = e^{-U(x)}$. U may be interpreted as a potential function.

define
FPE, in-
finitesi-
mal gen-
erators,
SDEs, etc.

2.1.1 Construction of the Family of Diffusions

Start with a Gaussian kernel $k_\varepsilon(x, y) = e^{-\frac{\|x-y\|^2}{\varepsilon}}$ and let $\alpha > 0$ being a parameter indexing this family of diffusions.

We can estimate the local density q_ε by

elaborate
on proba-
bility vs.
potential

Definition 2.1.

$$q_\varepsilon(x) = \int_X k_\varepsilon(x, y) d\mu(y)$$

Now consider the family of kernels

Definition 2.2.

$$k_\varepsilon^{(\alpha)}(x, y) = \frac{k_\varepsilon(x, y)}{q_\varepsilon^\alpha(x) q_\varepsilon^\alpha(y)}$$

We compute the normalization factor¹ $d_\varepsilon^{(\alpha)}$

¹Note that this factor is the continuous analogue to definition 1.1.

$$d_\varepsilon^{(\alpha)}(x) = \int_X k_\varepsilon^{(\alpha)}(x, y) d\mu(y)$$

and define the forward and symmetric ($p^{(\alpha)}$ and $a^{(\alpha)}$ respectively) transition probability kernels:

Definition 2.3.

$$p_\varepsilon^{(\alpha)}(x|y) = \frac{k_\varepsilon^{(\alpha)}(x, y)}{d_\varepsilon^{(\alpha)}(y)} = Pr(x(t + \varepsilon) = x \mid x(t) = y)$$

$$a_\varepsilon^{(\alpha)}(x|y) = \frac{k_\varepsilon^{(\alpha)}(x, y)}{\sqrt{d_\varepsilon^{(\alpha)}(x)d_\varepsilon^{(\alpha)}(y)}}$$

Now we define the forward, backward and symmetric Chapman-Kolmogorov operators on functions defined on this probability space, as follows:

Definition 2.4.

$$T_{f,\varepsilon}^{(\alpha)}[\phi](x) = \int_X p_\varepsilon^{(\alpha)}(x|y)\phi(y)d\mu(y)$$

$$T_{b,\varepsilon}^{(\alpha)}[\phi](x) = \int_X p_\varepsilon^{(\alpha)}(y|x)\phi(y)d\mu(y)$$

$$T_{s,\varepsilon}^{(\alpha)}[\phi](x) = \int_X a_\varepsilon^{(\alpha)}(x, y)\phi(y)d\mu(y)$$

If $\phi(x)$ is the probability of finding the system at location x at time $t = 0$, then $T_{f,\varepsilon}^{(\alpha)}[\phi]$ is the evolution of this probability to time $t = \varepsilon$. Similarly, if $\psi(z)$ is some function on the space, then $T_{b,\varepsilon}^{(\alpha)}[\psi](x)$ is the mean value of that function at time ε for a random walk that started at x , and so $\left(T_{b,\varepsilon}^{(\alpha)}\right)^m[\psi](x)$ is the average value of the function at time $t = m\varepsilon$.

By definition, the operators $T_{f,\varepsilon}^{(\alpha)}$ and $T_{b,\varepsilon}^{(\alpha)}$ are adjoint under the inner product with weight μ , while the operator $T_{s,\varepsilon}^{(\alpha)}$ is self adjoint under this inner product.

Moreover, just like in the discrete case, since $T_{s,\varepsilon}^{(\alpha)}$ is obtained via conjugation of the kernel $p^{(\alpha)}$ all three operators share the same eigenvalues. The corresponding eigenfunctions can be found via conjugation by $\sqrt{d_\varepsilon^{(\alpha)}}$; i.e.: If $T_{s,\varepsilon}^{(\alpha)}\phi_s = \lambda\phi_s$, then the corresponding eigenfunctions for $T_{f,\varepsilon}^{(\alpha)}$ and $T_{b,\varepsilon}^{(\alpha)}$ are $\phi_f = \sqrt{q_\varepsilon}\phi_s$ and $\phi_b = \frac{\phi_s}{\sqrt{q_\varepsilon}}$, respectively.

double
check that
section

Since $\sqrt{q_\varepsilon}$ is the first eigenfunction with $\lambda_0 = 1$ of $T_s^{(\alpha)}$, the steady state of the forward operator is simply $q_\varepsilon(x)$, while for the backward operator it is the uniform density, $\phi_b \equiv 1$.

Furthermore, note that the eigenvalues and eigenvectors of the discrete Markov chain described in the previous section are discrete approximations to the eigenvalues and eigenfunctions of these continuous operators.

2.1.2 Transition to Diffusion Processes

Interpreting ε as a timestep² and recalling the asymptotic expansions derived in the appendix it is instructive to look at the limit $\varepsilon \rightarrow 0$. In this case, the transition probability densities of the Markov chain (that is continuous in space, but discrete in time) converge to those of a diffusion process, whose time evolution is described by a differential equation

$$\frac{\partial \phi}{\partial t} = \mathcal{H}_f^{(\alpha)} \phi$$

where $\mathcal{H}_f^{(\alpha)}$ is the infinitesimal generator or propagator of the forward operator, defined as

Definition 2.5.

$$\mathcal{H}_f^{(\alpha)} = \lim_{\varepsilon \rightarrow 0} \frac{I - T_{f,\varepsilon}^{(\alpha)}}{\varepsilon}$$

Similarly, the infinitesimal operator of the backward operator is given by

Definition 2.6.

$$\mathcal{H}_b^{(\alpha)} = \lim_{\varepsilon \rightarrow 0} \frac{T_{b,\varepsilon}^{(\alpha)} - I}{\varepsilon}$$

and as shown in the appendix (theorem A.2.2)

$$\mathcal{H}_b^{(\alpha)} \psi = \Delta \psi - 2(1 - \alpha) \nabla \psi \cdot \nabla U$$

add case
for the
ffPO
and cases
for $\alpha \in$
 $\{0, \frac{1}{2}, 1\}$

²While in the case of a finite amount of data, ε must remain finite so as to have enough neighbors in a ball of radius $\mathcal{O}(\varepsilon^{\frac{1}{2}})$ near each point x , as the number of samples goes to infinity, we can take smaller and smaller values of ε .

Appendix A

Asymptotics for Laplacian Operators

In the following we will deal with a compact manifold \mathcal{M} that is C^∞ . Let x be a fixed point, not on the boundary, $T_x\mathcal{M}$ be the tangent space to \mathcal{M} at x and (e_1, \dots, e_d) be a fixed orthonormal basis of $T_x\mathcal{M}$. Furthermore two systems of local coordinates will be introduced:

1. (*Normal coordinates*) The exponential map \exp_x generates a set of orthogonal geodesics $(\gamma_1, \dots, \gamma_d)$ intersecting at x with initial velocity (e_1, \dots, e_d) . Now every point $y \in \mathcal{M}$ in a sufficiently small neighborhood of x has a set of *normal coordinates* (s_1, \dots, s_d) along these geodesics.
2. (*Tangent coordinates*) Considering the orthogonal projection u of y on $T_x\mathcal{M}$, where $u_i = \langle y - x, e_i \rangle$ in (e_1, \dots, e_d) , we get a system of *tangent coordinates*. The submanifold is now locally parameterized as $y = (u, g(u))$, where $g : \mathbb{R}^d \rightarrow \mathbb{R}^{n-d}$. Since $u = (u_1, \dots, u_d)$ are tangent coordinates, we must have that $\partial g(0) = 0$.

Notice that, locally, any function f on \mathcal{M} may be viewed as \tilde{f} of (s_1, \dots, s_d) and thus we may write $\Delta f(x) = -\sum_{i=1}^d \frac{\partial^2 \tilde{f}}{\partial s_i^2}(0, \dots, 0)$, where Δ is the Laplace-Beltrami operator on \mathcal{M} .

A.1 Comparison of the Geodesic and the Local Projection

In this section we will compute asymptotic expansions for the changes of

write introduction + motivation (later being the justification for naming it graph “Laplacian” earlier)

The following is basically a citation of Coifman and Lafon (2006).

variable $u \mapsto (s_1, \dots, s_d)$ and $u \mapsto y$.

In the following, $Q_{x,m}(u)$ denotes a generic homogeneous polynomial of degree m of the variable $u = (u_1, \dots, u_d)$, whose coefficient depends on x .

Lemma A.1.1. *If $y \in \mathcal{M}$ is in a Euclidean ball of radius $\varepsilon^{\frac{1}{2}}$ around x , then, for ε sufficiently small, there exists:*

$$s_i = u_i + Q_{x,3}(u) + \mathcal{O}(\varepsilon^2) \quad (\text{A.1})$$

Proof. Let γ be the geodesic connection x and y parameterized by arclength. We have $\gamma(0) = x$ and let s be such that $\gamma(s) = y$. If y has normal coordinates (s_1, \dots, s_d) , then we have $s\gamma'(0) = (s_1, \dots, s_d)$. A Taylor expansion yields

$$\gamma(s) = \gamma(0) + s\gamma'(0) + \frac{s^2}{2}\gamma''(0) + \frac{s^3}{6}\gamma^{(3)}(0) + \mathcal{O}(\varepsilon^2).$$

By definition of a geodesic, the covariant derivative of the velocity is zero, which means that $\gamma''(0)$ is orthogonal to the tangent plane at x . Now since the parameter u_i is defined by $u_i = \langle \gamma(s) - \gamma(0), e_i \rangle$, we obtain that $u_i = s_i + \frac{s^3}{6}\langle \gamma^{(3)}(0), e_i \rangle + \mathcal{O}(\varepsilon^2)$. Iterating this equation yields the result. \square

Lemma A.1.2. *Again, let $y \in \mathcal{M}$ be in a Euclidean ball of radius $\varepsilon^{\frac{1}{2}}$ around x ; we have*

$$\|x - y\|^2 = \|u\|^2 + Q_{x,4}(u) + Q_{x,4}(u) + \mathcal{O}(\varepsilon^3) \quad (\text{A.2})$$

and

$$\det \left(\frac{dy}{du} \right) = 1 + Q_{x,2}(u) + Q_{x,3}(u) + \mathcal{O}(\varepsilon^2). \quad (\text{A.3})$$

Proof. The submanifold is locally parameterized as $u \mapsto (u, g(u))$, where $g : \mathbb{R}^d \rightarrow \mathbb{R}^{n-d}$. Writing $g = (g_{i+1}, \dots, g_n)$ and applying Pythagore's theorem, we obtain

$$\|x - y\|^2 = \|u\|^2 + \sum_{i=d+1}^n g_i(u)^2.$$

Using that, by definition, $g_i(0) = 0$ and, as noted before, $\frac{\partial g}{\partial u_i}(0) = 0$. As a consequence $g_i(u) = b_{i,x}(u) + c_{i,x}(u) + \mathcal{O}(\varepsilon^2)$, where $b_{i,x}$ is the Hessian quadratic form of g_i at $u = 0$ and $c_{i,x}$ is the cubic term. This proves (A.2) with

$$Q_{x,4}(u) = \sum_{i=d+1}^n b_{i,x}^2(u) \text{ and } Q_{x,5}(u) = 2 \sum_{i=d+1}^n b_{i,x}(u)c_{i,x}(u).$$

To prove (A.3), observe that $\frac{\partial g}{\partial u_i}(0) = 0$ implies that $\frac{\partial g}{\partial u_i}(0) = \tilde{b}_{i,x}(u) + \tilde{c}_{i,x}(u) + \mathcal{O}(\varepsilon^{\frac{3}{2}})$, where $\tilde{b}_{i,x}(u)$ and $\tilde{c}_{i,x}(u)$ are the linear and quadratic terms in the Taylor expansion of $\frac{\partial g}{\partial u_i}(0)$ at 0. We thus have:

$$\begin{aligned} \frac{\partial y}{\partial u_i}(u) &= \left(v_i, \frac{\partial g}{\partial u_i}(u) \right), \text{ where } v_i = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d \\ &= (v_i, \tilde{b}_{i,x}(u) + \tilde{c}_{i,x}(u) + \mathcal{O}(\varepsilon^{\frac{3}{2}})). \end{aligned}$$

The squared volume generated by these d vectors is the determinant of their Gram matrix, i.e.,

$$\left| \det \left(\frac{dy}{du} \right) \right|^2 = \sum_{i,j=1}^d E_{ij}(u) + \sum_{i,j=1}^d F_{ij}(u) + \mathcal{O}(\varepsilon^2),$$

where

$$E_{ij}(u) = \langle \tilde{b}_{i,x}(u), \tilde{b}_{j,x}(u) \rangle \text{ and } F_{ij}(u) = \langle \tilde{b}_{i,x}(u), \tilde{c}_{j,x}(u) \rangle + \langle \tilde{c}_{i,x}(u), \tilde{b}_{j,x}(u) \rangle.$$

Defining

$$Q_{x,2}(u) = \sum_{i,j=1}^d E_{ij}(u) \text{ and } Q_{x,3}(u) = \sum_{i,j=1}^d F_{ij}(u),$$

we obtain the last result. \square

A.2 Infinitesimal Operators for a Family of Graph Laplacians

In this section we present the calculation of the infinitesimal generators for the different diffusion maps characterized by a parameter α .

To start with we first show an asymptotic expansion for diffusion operators G_ε .

Let $k_\varepsilon(x, y)$ be an isotropic kernel, i.e.:

$$k_\varepsilon(x, y) = h \left(\frac{\|x - y\|^2}{\varepsilon} \right),$$

where h is assumed to have an exponential decay and let G_ε be the corresponding operator

$$G_\varepsilon f(x) = \frac{1}{\varepsilon^{\frac{d}{2}}} \int_{\mathcal{M}} k_\varepsilon(x, y) f(y) dy.$$

The idea is that, using the previous lemmata, for small ε integrating f against the kernel on the manifold is approximately like integrating on the tangent space.

Theorem A.2.1. *Let $f \in C^3(\mathcal{M})$ and let $0 < \gamma < 1/2$. Then we have, uniformly for all $x \in \mathcal{M}$ at distance larger than ε^γ from $\partial\mathcal{M}$,*

$$G_\varepsilon f(x) = m_0 f(x) + \varepsilon \frac{m_2}{2} (\omega(x) f(x) - \Delta f(x)) + \mathcal{O}(\varepsilon^2),$$

where

$$m_0 = \int_{\mathbb{R}^d} h(|u|^2) du \text{ and } m_2 = \int_{\mathbb{R}^d} u_1^2 h(|u|^2) du$$

and ω is a potential term depending on the embedding of \mathcal{M} .

Proof. Because of the exponential decay of h , the domain of integration can be restricted to the intersection of \mathcal{M} with the ball of radius ε^γ around x . In doing so we generate an error of order

$$\begin{aligned} \left| \frac{1}{\varepsilon^{\frac{d}{2}}} \int_{\substack{y \in \mathcal{M} \\ \|x-y\| > \varepsilon^\gamma}} h\left(\frac{\|x-y\|^2}{\varepsilon}\right) f(y) dy \right| &\leq \|f\|_\infty \frac{1}{\varepsilon^{\frac{d}{2}}} \int_{\substack{y \in \mathcal{M} \\ \|x-y\| > \varepsilon^\gamma}} \left| h\left(\frac{\|x-y\|^2}{\varepsilon}\right) \right| dy \\ &\leq \|f\|_\infty \int_{\substack{y \in \mathcal{M} \\ \|y\| > \varepsilon^{\gamma-1/2}}} |h(|y|^2)| dy \\ &\leq C \|f\|_\infty Q(\varepsilon^{1/2-\gamma}) e^{-\varepsilon^{\gamma-1/2}}, \end{aligned}$$

where we have used the exponential decay of the kernel and where Q is a polynomial. Since $0 < \gamma < 1/2$, this term is exponentially small and is bounded by $\mathcal{O}(\varepsilon^{\frac{3}{2}})$. Therefore,

$$G_\varepsilon f(x) = \frac{1}{\varepsilon^{\frac{d}{2}}} \int_{\substack{y \in \mathcal{M} \\ \|x-y\| < \varepsilon^\gamma}} h\left(\frac{\|x-y\|^2}{\varepsilon}\right) f(y) dy + \mathcal{O}(\varepsilon^{\frac{3}{2}}).$$

Now that things are localized around x , we can Taylor-expand the function $(s_1, \dots, s_d) \mapsto f(y(s_1, \dots, s_d))$:

$$f(y) = f(x) + \sum_{i=1}^d s_i \frac{\partial \tilde{f}}{\partial s_i}(0) + \frac{1}{2} \sum_{i,j=1}^d s_i s_j \frac{\partial^2 \tilde{f}}{\partial s_i \partial s_j}(0) + Q_{x,3}(s_1, \dots, s_d) + \mathcal{O}(\varepsilon^2),$$

where $\tilde{f}(s_1, \dots, s_d) = f(y(s_1, \dots, s_d))$. Invoking (A.1), we obtain

$$f(y) = \tilde{f}(0) + \sum_{i=1}^d u_i \frac{\partial \tilde{f}}{\partial s_i}(0) + \frac{1}{2} \sum_{i,j=1}^d u_i u_j \frac{\partial^2 \tilde{f}}{\partial s_i \partial s_j}(0) + Q_{x,3}(u) + \mathcal{O}(\varepsilon^2).$$

Likewise, because of (A.2), the Taylor expansion of the kernel is

$$h\left(\frac{\|x - y\|^2}{\varepsilon}\right) = h\left(\frac{\|u\|^2}{\varepsilon}\right) + \left(\frac{Q_{x,4}(u) + Q_{x,5}(u)}{\varepsilon}\right) h'\left(\frac{\|u\|^2}{\varepsilon}\right) + \mathcal{O}(\varepsilon^2).$$

Using (A.3) to change the variable $s \mapsto u$ in the previous integral defining $G_\varepsilon f(x)$ yields:

$$\begin{aligned} \varepsilon^{\frac{d}{2}} G_\varepsilon f(x) &= \int_{\|u\| < \varepsilon^\gamma} \left[h\left(\frac{\|u\|^2}{\varepsilon}\right) + \left(\frac{Q_{x,4}(u) + Q_{x,5}(u)}{\varepsilon}\right) h'\left(\frac{\|u\|^2}{\varepsilon}\right) \right] \\ &\quad \times \left[\tilde{f}(0) + \sum_{i=1}^d u_i \frac{\partial \tilde{f}}{\partial s_i}(0) + \frac{1}{2} \sum_{i,j=1}^d u_i u_j \frac{\partial^2 \tilde{f}}{\partial s_i \partial s_j}(0) + Q_{x,3}(u) \right] \\ &\quad \times (1 + Q_{x,2}(u) + Q_{x,3}(u)) du + \mathcal{O}\left(\varepsilon^{\frac{d}{2}+2}\right). \end{aligned}$$

This identity can be dramatically simplified by identifying odd functions and setting their integral to zero. One is left with

$$\begin{aligned} \varepsilon^{\frac{d}{2}} G_\varepsilon f(x) &= \tilde{f}(0) \int_{\mathbb{R}^d} h\left(\frac{\|u\|^2}{\varepsilon}\right) du + \frac{1}{2} \left(\sum_{i=1}^d \frac{\partial^2 \tilde{f}}{\partial s_i^2}(0) \right) \int_{\mathbb{R}^d} u_1^2 h\left(\frac{\|u\|^2}{\varepsilon}\right) du \\ &\quad + \tilde{f}(0) \int_{\mathbb{R}^d} \left[\frac{Q_{x,4}(u)}{\varepsilon} h'\left(\frac{\|u\|^2}{\varepsilon}\right) + \tilde{Q}_{x,2}(u) h\left(\frac{\|u\|^2}{\varepsilon}\right) \right] du + \mathcal{O}\left(\varepsilon^{\frac{d}{2}+2}\right), \end{aligned}$$

where the domain of integration has been extended to \mathbb{R}^d (exponential decay of h). Changing the variable according to $u \mapsto \varepsilon^{\frac{1}{2}}u$,

$$\begin{aligned} G_\varepsilon f(x) &= \tilde{f}(0) \int_{\mathbb{R}^d} h(\|u\|^2) du + \frac{\varepsilon}{2} \left(\sum_{i=1}^d \frac{\partial^2 \tilde{f}}{\partial s_i^2}(0) \right) \int_{\mathbb{R}^d} u_1^2 h(\|u\|^2) du \\ &\quad + \varepsilon \tilde{f}(0) \int_{\mathbb{R}^d} (Q_{x,4}(u) h'(\|u\|^2) + Q_{x,2}(u) h(\|u\|^2)) du + \mathcal{O}(\varepsilon^2), \end{aligned}$$

where we have used the homogeneity of $Q_{x,2}$ and $Q_{x,4}$. Finally, observing that

$$\tilde{f}(0) = f(x) \text{ and } \sum_{i=1}^d \frac{\partial^2 \tilde{f}}{\partial s_i^2}(0) = -\Delta f(x),$$

we end up with

$$G_\varepsilon f(x) = m_0 f(x) + \varepsilon \frac{m_2}{2} (\omega(x) f(x) - \Delta f(x)) + \mathcal{O}(\varepsilon^2),$$

where

$$\omega(x) = \frac{2}{m_2} \int_{\mathbb{R}^d} (Q_{x,4}(u) h'(\|u\|^2) + Q_{x,2}(u) h(\|u\|^2)) du.$$

Finally, the uniformity follows from the compactness and smoothness of \mathcal{M} . \square

One can also show the same result holds for all $x \in \mathcal{M}$ at distance smaller than ε^γ from $\partial\mathcal{M}$. For this proof we refer to Coifman and Lafon (2006) as this would go beyond the scope of the discussion.

Suppose that the data set X consists of a Riemannian manifold with a density $p(x) = e^{-U(x)}$ and let $k_\varepsilon(x, y)$ be a Gaussian kernel (which clearly satisfies the requirements for A.2.1).

Theorem A.2.2. *The infinitesimal generator $\mathcal{H}_b\phi$ of the backward operator $T_{b,\varepsilon}^{(\alpha)}\phi = \int_{\Gamma} \frac{k_\varepsilon^{(\alpha)}(x,y)}{d_\varepsilon^{(\alpha)}(x)} \phi(y) p(y) dy$ is $\Delta\phi - 2(1-\alpha)\nabla\phi \cdot \nabla U$.*

Proof. From A.2.1 we see that¹

$$p_\varepsilon(x) = p(x) + \varepsilon(\Delta p(x) + \omega(x)p(x)) + \mathcal{O}(\varepsilon^2)$$

¹w.l.o.g. we assume $m_0 = 1$ and $m_2 = 2$

and consequently,

$$p_\varepsilon^{-\alpha} = p^{-\alpha} \left(1 - \alpha\varepsilon \left(\frac{\Delta p}{p} + \omega \right) \right) (1 + \mathcal{O}(\varepsilon^2)).$$

Let

$$k_\varepsilon^{(\alpha)}(x, y) = \frac{k_\varepsilon(x, y)}{p_\varepsilon^\alpha(x) p_\varepsilon^\alpha(y)}$$

Then, the normalization factor $d_\varepsilon^{(\alpha)}$ is given by

$$d_\varepsilon^{(\alpha)}(x) = \int_\Gamma k_\varepsilon^{(\alpha)}(x, y) p(y) dy = p_\varepsilon^{-\alpha}(x) p_\varepsilon^{1-\alpha}(x) \left[1 + \varepsilon \left((1 - \alpha)\omega - \alpha \frac{\Delta p}{p} + \frac{\Delta p^{1-\alpha}}{p^{1-\alpha}(x)} \right) \right].$$

Therefore, the asymptotic expansion of the backward operator gives

$$T_b^{(\alpha)} \phi = \int_\Gamma \frac{k_\varepsilon^{(\alpha)}(x, y)}{d_\varepsilon^{(\alpha)}(x)} \phi(y) p(y) dy = \phi(x) + \varepsilon \left(\frac{\Delta(\phi p^{1-\alpha})}{p^{1-\alpha}} - \phi \frac{\Delta p^{1-\alpha}}{p^{1-\alpha}} \right)$$

and its infinitesimal generator is

$$\mathcal{H}_b \phi = \lim_{\varepsilon \rightarrow 0} \frac{T_b - I}{\varepsilon} \phi = \frac{\Delta(\phi p^{1-\alpha})}{p^{1-\alpha}} - \phi \frac{\Delta p^{1-\alpha}}{p^{1-\alpha}}.$$

Plugging in $p = e^{-U}$ into the last equation gives the desired result. \square

Similarly, one can show that the infinitesimal generator $\mathcal{H}_f \psi$ of the forward operator $T_{f, \varepsilon}^{(\alpha)} \psi$ is $\Delta \psi - 2\alpha \nabla \psi \cdot \nabla U + (2\alpha - 1)\psi(\nabla U \cdot \nabla U - \Delta U)$.

Bibliography

- Belkin, M. (2003), *Problems of Learning on Manifolds*. Ph.D. thesis, The University of Chicago.
- Canzani, Y. (2013), “Analysis on manifolds via the laplacian.”, URL <http://www.math.harvard.edu/~canzani/math253/Laplacian.pdf>.
- Coifman, Ronald R. and Stéphane Lafon (2006), “Diffusion maps.” *Applied and Computational Harmonic Analysis*, 21, 5 – 30, URL <http://www.sciencedirect.com/science/article/pii/S1063520306000546>. Special Issue: Diffusion Maps and Wavelets.
- Nadler, Boaz, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis (2006), “Diffusion maps, spectral clustering and reaction coordinates of dynamical systems.” *Applied and Computational Harmonic Analysis*, 21, 113 – 127, URL <http://www.sciencedirect.com/science/article/pii/S1063520306000534>. Special Issue: Diffusion Maps and Wavelets.
- Teschl, G. (2014), *Topics in Real and Functional Analysis*. a, URL <http://www.mat.univie.ac.at/~gerald/ftp/book-fa/fa.pdf>.